厚生労働行政推進調査事業費補助金（化学物質リスク研究事業）

OECDプロジェクトでの成果物を厚生労働行政に反映させるための研究

令和３年度 分担研究報告書

OECDにおけるAOP, TG及びGDの開発

研究分担者　小島　肇

国立医薬品食品衛生研究所　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

研究要旨

本研究班は、化学物質やその混合物の安全性を評価するための国際的な合意を推進する経済協力開発機構（OECD: Organisation for Economic Co-operation and Development）の試験法ガイドライン（TG: Test Guideline）プログラム各国調整官作業グループ（WNT: Working Group of National Co-ordinators of the TGs programme）において、日本で開発された種々の TG やガイダンス文書（GD: Guidance Document）、毒性発現経路（AOP : Adverse Outcome Pathway）などの世界各国が必要とする成果物を公定化させることが目的の一つである。

本研究班の中で、私はこの AOP, TG および GD の公定化を担当した。昨年度からの継続した活動の中、日本人の開発した AOP として、「カルシニューリン阻害による T 細胞依存的抗体産生抑制：AOP154」が OECD にて正式に承認された。TG に関しては、既存の TG である皮膚感作性試験代替法 ADRA（Amino acid Derivative Reactivity Assay）を含む TG442C の改定をなすことができた。同時に承認された Defined Approach for Skin Sensitisation ガイドライン 497 の開発にも寄与した。

研究協力者

相場節也
　東北大学医学系研究科・医学部
　名誉教授

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 室長

須方督夫
　一般社団法人日本化学工業協会
　化学品管理部長

大石　巧
　日本免疫毒性学会試験法委員会
　AOP 検討小委員会

尾上誠良
　静岡県立大学　薬学部・薬剤学分野
　教授

笠原利彦
　富士フィルム株式会社
　安全性評価センター 技術マネージャー

木村　裕
　元東北大学医学系研究科 医学部
　皮膚科学分野　助教

A. 研究目的

本研究班は、化学物質やその混合物の安全性を評価するための国際的な合意を推進する経済協力開発機構（OECD: Organisation for Economic Co-operation and Development）の試験法ガイドライン（TG: Test Guideline）プログラム各国調整官作業グループ（WNT: Working Group of National Co-ordinators of the TGs programme）において、日本で開発された種々の TG やガイダンス文書（GD: Guidance Document）、毒性発現経路（AOP : Adverse Outcome Pathway）などの世界各国が必要とする成果物を公定化させるとともに、他国が提案する OECD 大型プロジェクトに関与し、その成果物に日本の主張を反映させ、これらから得られた成果を化学物質の審査及び製造等の規制に関する法律（化審法）や毒物及び劇物取締法（毒劇法）などの我が国の厚生労働行政に反映させることを目的としている。

私は、本研究班の中で目的の一つである AOP, TG 及び GD の開発を担当した。本年度、既存の TG である皮膚感作性試験代替法 ADRA や IL-8 Luc アッセイの改定、免疫毒性試験 MITA（Multi-Immuno Toxicity Assay)の TG を開発に努めた。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(Integrated Approaches to Testing and Assessment)や DA の国内での普及に務める。

B. 研究方法

B.1. AOP の開発

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

研究分担者の足利とともに、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 である「カルシニューリン阻害による T 細胞依存的抗体産生抑制 : AOP154」および「IL-1 receptor 結合阻害 : AOP277 の開発を委託している。文献調査の結果に基づいて、MIE（Molecular Initiating Event）、AO（Adverse Outcome）及びその間に介在する KE (Key Event）を定めて、OECD に指定された外部（または scientific）評価者及びコーチの指摘事項に対応することで開発を進めた。

B.2. TG 及び DRP の開発

研究代表者の平林とともに、OECDのTGの開発プロジェクトWNTの進捗に合わせ、班員を支援した。

B.2.1. 皮膚感作性試験

協力研究者の笠原とともに、皮膚感作性試験代替法 *In Chemico* Skin Sensitisation、ADRA（Amino acid Derivative Reactivity Assay）の追加バリデーション報告書及び TG442C の改定案を作成した。

また、協力研究者の相場とともに、IL-8 Luc assay TG442E の改定案を作成した。

足利とともに、Defined Approach for Skin Sensitisation のプロジェクトに参加し、ガイドラインの成立に協力した。

B.2.2. 免疫毒性試験

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

DRP の承認を待って提出することを予定している IL-2 を指標とした免疫毒性試験の TG 案を作成した。

B.2.3. 生殖毒性試験の DRP

国際的な専門家とともに、*in vitro* 生殖発生毒性に関する DRP を作成した。

B.3. 光毒性 IATA

研究分担者の尾上及び JaCVAM 資料編纂委員会の協力を得て、光毒性 IATA 案を作成した。

（倫理面への配慮）

特になし

C. 研究結果

C.1.免疫毒性の AOP

「カルシニューリン阻害による T 細胞依存的抗体産生抑制：AOP154」については、WNT/WPHA（ Working Party of Hazard Assessment）に提出したのち、ドイツからコメントが提示されたという連絡があり、AOP Wiki の改訂作業を行った。ドイツからの主な指摘は、TDAR（T cell Dependent Antibody Response）アッセイの方法を詳細に記載すること、本 AOP のみで免疫毒性試験が免除されることはなく、本 AOP は IATA 開発に利用されるべきであること、本 AOP の EU 地域での規制上の重要性についても記載することなどであった。指摘事項に対応し、OECD 事務局に改訂完了の連絡と著者回答ファイルを提出したところ、OECD 事務局から、本 AOP が WNT/WPHA で承認された。その後 OECD 事務局に著作権譲渡に関する著者全員の署名書類を提出し、令和 3（2021）年 10 月 15 日に OECD Library において公開された。

「IL-1 receptor 結合阻害：AOP277」については、AOP 開発の引継ぎに関する web 会議を実施し、元の AOP 開発者である相場らとコーチとの web 会議が行われ、評価結果について説明を受けた。その後 OECD から scientific review report を受領した。その主な推奨事項は、IL-1R シグナルを阻害するストレッサーに特異抗体だけでなく化合物/医薬品を加えること、AP-1 など NF-kB が関与しない経路も考慮すること、T cell のタイプを明確にすること、増加する感染のタイプを明確にすることなどであった。現在これらの推奨事項に対応しており、来年度早々に修正 AOP を OECD に提出予定である。

C.2. TG 及び DRP の開発

C.2.1. 皮膚感作性試験

協力研究者の笠原とともに、昨年度から検討を続けてきた *In Chemico* Skin Sensitisation、 ADRA の改定に尽力した結果、2021 年 6 月に TG の改定が承認された。引き続き、OECD の専門家から要請を受け、ADRA の国内施設の協力を得て適用濃度を 1mM から 4mM に引き上げたプロトコルを用いた追加バリデーションの報告書及び TG442C の最終改定案を作成し、7 月に OECD に提出した。11 月の専門家会議を経て、改訂 TG 案を OECD に提出した。

また、協力研究者の相場とともに、IL-8 Luc assay TG442E の改定案を作成し、7 月に OECD に提出した。11 月に peer review panel からコメントを受け、まだ合意に至っていない。来年度、再協議を行うことになっている。

足利とともに、昨年度から専門家委員会で検討を続けてきた Defined Approach（DA）for Skin Sensitisation の承認に寄与した結果、令和 3（2021）年 6 月にガイドライン 497 が承認された。

本ガイドラインは新しいタイプの OECD ガイドラインである。DA では化学物質の物性および *in vitro* 試験データで検

証された OECD の組み合わせを使用している。その一つとして、 DA では、化学物質規制に *in silico* データを受け入れることを可能にした最初の事例である。

このガイドラインには以下に示す画期的な点が複数含まれている。

1　初めて試験法の結果を組み合わせて評価する手法が公定化された。
2　初めて *in silico* の利用が組み合わせ評価に利用された。
3　ヒトの感作性が予測できる初のガイドラインである。

C.2.2.　免疫毒性試験

相場及び国際的な専門家とともに、昨年度、*in vitro* 免疫毒性に関する DRP を作成した。OECD が集めた意見に対応する改定版を 9 月に提出したところ、2 次募集において追加意見が集まった。国際的な専門家の協力を得て DRP を改定し、2 月に OECD に提出した。

DRP の承認を待って提出する予定の IL-2 を指標とした免疫毒性試験の TG 案を作成し、3 月に OECD に提出した。

C.2.3. 生殖毒性試験の DRP

国際的な専門家とともに、*in vitro* 生殖発生毒性に関する DRP の作成をこの一年継続して実施している。本来は夏休み前に適切な論文に投稿する予定であったが、著者の中に EPA や企業の専門家が含まれていることから、公表にあたり EPA や企業の同意を得ることに手間取ったこともあり、2022 年 2 月に Current Research Toxicology に投稿し、3 月に改訂の指示を受けた。

C.3.　光毒性 IATA

令和元（2019）年 6 月に OECD TG495 としてガイドライン化された ROS assay を主軸として、新たに光安全性評価のための IATA 構築を進めている。既に OECD の専門家会議にて共有し、専門家より修正に資する重要なコメントを頂いている。ICH S10 において推奨されているストレテジーをベースとし、(i) 被験物質の光化学的特性評価、(ii) 光生物化学的特性評価、そして (iii) 皮膚や眼への移行性・滞留性等体内動態評価の　3 段階のスクリーニングによる tiered approach を案として光毒性を中心とした提示をし、各国の専門家と議論している。

D. 考察

D.1. 免疫毒性の AOP

開発中の AOP については、コーチ及び scientific reviewer のコメントに基づいて修正を行っている。IL-23 の機能の説明のように既存の情報を収集することで対応できるものもあるが、医薬品だけでなく化学物質のストレッサーを示すこと、など情報がないものについては調査したことを示して納得いただく必要がある。また、AO が疾患の憎悪である AOP については、行政活用の観点からその実用性に疑問が投げかけられており、すでに承認された AOP154 のように TDAR に変更するか、開発そのものを見直すことを検討中である。ただし、OECD に承認された AOP154 の KE の一つは IL-2 産生であり、本年度日本から別途 OECD に提案した免疫毒性スクリーニング試験 IL-2 Luc assay の TG 化の理論的基盤になるため、その意義は大きいと考える。

D.2. TG 及び DRP の開発

もとより TG 及び DRP の開発は、提案してから承認まで数年かけて国際合意を

取っていく作業である。免疫毒性や生殖毒性試験などの全身毒性に関する *in vitro* TG の開発は前例がなく、これまで以上に時間を要しており、費用も嵩んでいる。即ち、OECD は、こうした前例のない TG を開発するために、まずは DRP の作成を求めており、数年掛かりで免疫毒性と生殖毒性の GD 作成を進めてきている。さらに、OECD では公定化にあたり、通常、2 回の意見募集を実施するが、これらに対しては、OECD 事務局も通例になく慎重を期している。この分野の先頭を行く日本としては、国際的な専門家の協力を得ながら、焦らず、ことを進めて行く所存である。

D.3.2. 光毒性 IATA

構築した IATA 案については今後当該領域の専門家から頂いたコメントを基に修正していく予定である。特に decision tree の構築を強く求められているので、draft を作成して関係者間で慎重な協議を進めたい。

F. 添付資料

1. 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 (2021)
2. OECD Guideline No.497, Guideline on Defined Approaches for Skin Sensitisation (2021)

G. 研究発表

G.1. 論文発表

1. Imamura M, Wanibuchi S, Yamamoto Y, Kojima H, Ono A, Kasahara T, Fujita M. Improving predictive capacity of the Amino acid Derivative Reactivity Assay test method for skin sensitization potential with an optimal molar concentration of test chemical solution, *J Appl Toxicol*. 2021;41(2): 303-329.
2. Kojima H. Alternatives to animal testing, *Impact*, 2021; 44-45.
3. Fujita M, Yamamoto Y, Wanibuchi S, Watanabe S, Yamaga H, Wakabayashi K, Tahara Y, Horie N, Fujimoto K, Takeuchi K, Kamiya K, Kawakami T, Kojima K, Sozu T, Kojima H, Kasahara T, Ono A. The within- and between-laboratories reproducibility and predictive capacity of Amino acid Derivative Reactivity Assay using 4 mM test chemical solution: Results of ring study implemented at five participating laboratories, *Applied Toxicology*, 2022;42(2):318-333.
4. Imamura M, Yamamoto Y, Fujita M, Wanibuchi S, Nakashima N, Kojima H, Ono A, Kasahara T. Applicability of ADRA (4 mM) for the prediction of skin sensitization by combining multiple alternative methods to evaluate key events, *J Appl Toxicol*. 2022;Jan 6.
5. Yamamoto Y, Fujita M, Watanabe S, Yamaga H, Wakabayashi K, Tahara Y, Horie N, Fujimoto K, Takeuchi K, Kamiya K, Kawakami T, Kojima K, Sozu T, Kojima H, Kasahara T, Ono A. Within- and between-laboratory reproducibility and predictive capacity of amino acid derivative reactivity assay (ADRA) using a 0.5 mg/mL test chemical solution: Results of the study for reproducibility confirmation implemented in five participating laboratories, *J Appl Toxicol*.2022;Jan 18.

6. Fujita M, Yamamoto Y, Wanibuchi S , Watanabe S, Yamaga H, Wakabayashi K, Tahara Y, Horie N, Fujimoto K, Takeuchi K, Kamiya K, Kawakami T, Kojima K, Sozu T, Kojima H, Kasahara T, Ono A. The within- and between-laboratories reproducibility and predictive capacity of Amino acid Derivative Reactivity Assay using 4 mM test chemical solution: Results of ring study implemented at five participating laboratories, *J Appl Toxicol*. 2022;42(2):318-333.

G.2. 学会発表

1. Kojima H. AI, iPSC and MPS Projects for Systemic Toxicity, Annual meeting on advancing 21stCentury Toxicology, The Center for Alternatives to Animal Testing and the Animal-Free Safety Assessment Collaboration（2021.5.12, USA, On line）
2. Kojima H. Asian activities for alternatives to animal experiments, 11th World Congress on Alternatives and Animal Use in the Life Sciences (WC11) （2021.8.23, the Netherlands, On line）
3. Kojima H, Akbarsha MA, Gunatilake M, Kim BH. Marching Towards Asian Federation for Alternatives to Animal Testing (AFAAT) through Harmonization of Asian 3R Centres and Associations for Alternatives, 11th World Congress on Alternatives and Animal Use in the Life Sciences (WC11) （2021.8.27&31, the Netherlands, On line）
4. Kojima H. 21st-Century Toxicology and Regulatory Testing: An Update from Japan, 11th World Congress on Alternatives and Animal Use in the Life Sciences (WC11)（2021.9.1, the Netherlands, On line）
5. 相場節也, 木村裕, 足利太可雄, 小島肇. Multi-Immuno Toxicity Assay とガイダンス化状況 , 第 28 回日本免疫毒性学会学術年会（2021.9.7, On line）
6. Kojima H. Current projects developing new approach methods (NAMs) for systematic toxicology in Japan, 2021 International Symposium on Alternatives to Animal Testing in Taiwan（2021.10.5, On line）
7. 小島肇. ガイドライン化にむけた動き, 第三回 scChemRISC 研究会（2021.10.12, On line）
8. Kojima H. Alternative methods for developmental and reproductive toxicity testing regarding ICH S5(R3), 9thInternational Congress of Asian Society of Toxicology （ ASIATOX-IX) (2021.10.22, Hangzhou, China, On line）
9. Kojima H. Utilization of Endocrine Disrupter Screening for reproductive Toxicity, 6th Asia-Pacific Symposium on Food Safety 2021（2021.11,11, Jeju, Korea, On line）
10. 小島肇. Computational toxicology の国際動向, 日本動物実験代替法学会第 34 回大会（2021.11.11, 沖縄）

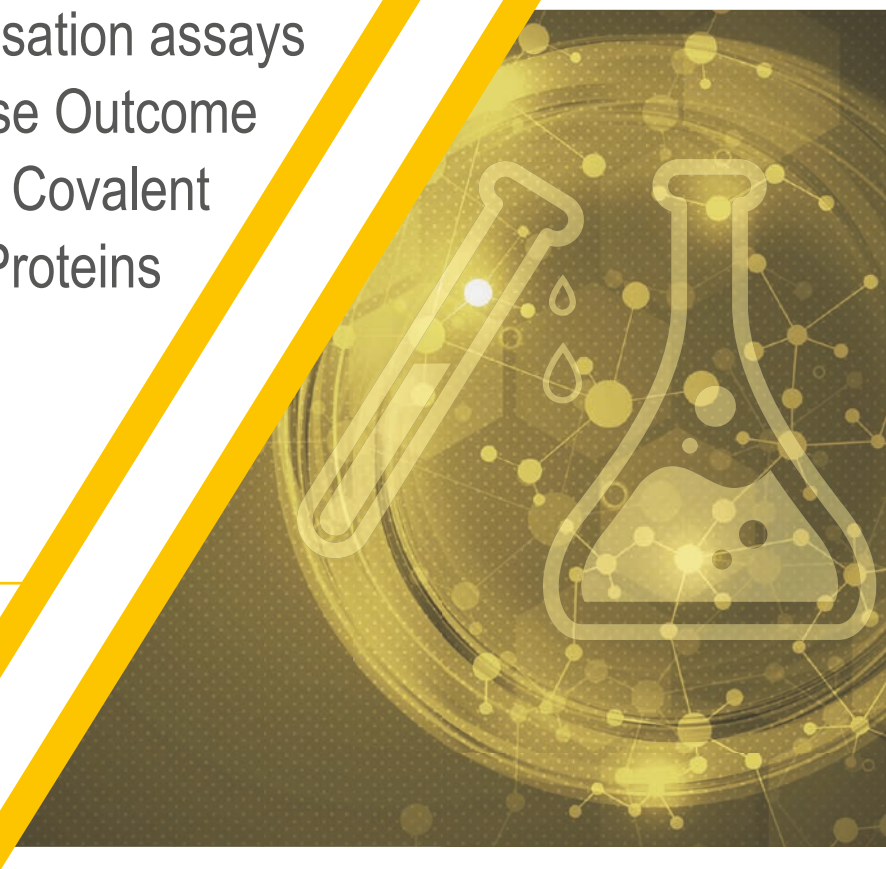
H. 知的所有権の取得状況

1. 特許取得
   特になし
2. 実用新案登録
   特になし
3. その他
   特になし

Section 4
Health effects

Test Guideline No. 442C

Key Event-Based Test Guideline for *in chemico* skin sensitisation assays addressing the Adverse Outcome Pathway Key Event on Covalent Binding to Proteins



14 June 2021

OECD Guidelines for the Testing of Chemicals

OECD
BETTER POLICIES FOR BETTER LIVES

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

INTRODUCTION

Covalent binding to proteins Key Event based Test Guideline.

1. A skin sensitiser refers to a substance that will lead to an allergic response following repeated skin contact as defined by the United Nations Globally Harmonized System of Classification and Labelling of Chemicals (UN GHS) (1). There is general agreement on the key biological events underlying skin sensitisation. The current knowledge of the chemical and biological mechanisms associated with skin sensitisation has been summarised as an Adverse Outcome Pathway (AOP) (2) starting with a molecular initiating event through intermediate events to the adverse effect, namely allergic contact dermatitis. This AOP focuses on chemicals that react with amino-acid residues (i.e. cysteine or lysine) such as organic chemicals. In this instance, the molecular initiating event (i.e. the first key event), is the covalent binding of electrophilic substances to nucleophilic centres in skin proteins. The second key event in this AOP takes place in the keratinocytes and includes inflammatory responses as well as changes in gene expression associated with specific cell signalling pathways such as the antioxidant/electrophile response element (ARE)-dependent pathways. The third key event is the activation of dendritic cells, typically assessed by expression of specific cell surface markers, chemokines and cytokines. The fourth key event is T-cell proliferation.

2. The assessment of skin sensitisation has typically involved the use of laboratory animals. The classical methods that use guinea-pigs, the Guinea Pig Maximisation Test (GPMT) of Magnusson and Kligman and the Buehler Test (OECD TG 406) (11) assess both the induction and elicitation phases of skin sensitisation. The murine tests, such as the LLNA (OECD TG 429) (12) and its three non-radioactive modifications — LLNA:DA (OECD TG 442A) (13), LLNA:BrdU-ELISA, and BrdU-FCM (OECD TG 442B) (14) — all assess the induction response exclusively and have gained acceptance, since they provide an advantage over the guinea pig tests in terms of animal welfare together with an objective measurement of the induction phase of skin sensitisation.

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

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

4. This Test Guideline (TG) describes *in chemico* assays that address mechanisms described under the first key event of the AOP for skin sensitisation, namely covalent binding to proteins (2). The test methods currently included in this Test Guideline are:
    - The Direct Peptide Reactivity Assay (DPRA) (Appendix I),
    - The Amino Acid Derivative Reactivity Assay (ADRA) (Appendix II), and
    - The kinetic Direct Peptide Reactivity Assay (kDPRA) (Appendix III).
5. The test methods are based on *in chemico* covalent binding to proteins and are considered to be scientifically valid. The DPRA has been evaluated in a European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)-lead validation study and subsequent independent peer review by the EURL ECVAM Scientific Advisory Committee (ESAC) (3) (4) (5). The ADRA underwent a validation study coordinated by the Japanese Center for the Validation of Alternative Methods (JaCVAM) (6) (7) (8) (9) followed by an independent peer-review (10). The kDPRA underwent an industry-coordinated validation study followed by an independent peer-review (17).
6. The test methods included in this Test Guideline might differ with regard to the procedures used to generate the data but can each be used to address countries' requirements for test results on protein reactivity, while benefiting from the Mutual Acceptance of Data.
7. The correlation of protein reactivity with skin sensitisation potential is well established (18) (19) (20). Nevertheless, since protein reactivity represents only one key event of the skin sensitisation AOP (2) (21), information generated with test methods developed to address this specific key event may not be sufficient as stand-alone methods to conclude on the presence or absence of skin sensitisation potential of chemicals. Therefore data generated with the test methods described in this Test Guideline are proposed to be used within Integrated Approaches to Testing and Assessment (IATA), together with other relevant complementary information from *in vitro* assays addressing other key events of the skin sensitisation AOP as well as non-testing methods, including *in silico* modelling and read-across from chemical analogues (21). Examples on the use of data generated with these methods within Defined Approaches (DAs), i.e. approaches standardised both in relation to the set of information sources used and in the procedure applied to derive predictions, have been published (21) and can be employed as useful elements within IATA.
8. The DPRA and ADRA described in Appendixes I and II to this Test Guideline, respectively, support the discrimination of skin sensitisers (Category 1) from non-sensitisers. Depending on the regulatory framework, positive results generated with these methods may be used on their own to classify a chemical into UN GHS Category 1. However, these test methods do not allow on their own, the sub-categorisation of skin sensitisers into subcategories 1A and 1B (22), as defined by UN GHS (1) for authorities implementing these two optional subcategories, or potency prediction for safety assessment decisions.
9. In contrast, the kDPRA described in Appendix III of this Test Guideline, allows discrimination of UN GHS subcategory 1A skin sensitisers from those not categorised as subcategory 1A (non-subcategory 1A) i.e., subcategory 1B or no category (1) but does not allow to distinguish sensitisers (Category 1) from non-sensitisers. Depending on the regulatory framework, positive results generated with the kDPRA may be used on their own to classify a chemical into UN GHS subcategory 1A.
10. Definitions are provided in the Annex. Performance Standards for the assessment of proposed similar or modified *in vitro* skin sensitisation DPRA and ADRA test methods have been developed (23).

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 lysine derivatives. J Pharmacol Toxicol Methods. 70, 94-105.

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

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

(9) OECD (2019), 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) Wareing, B., Kolle, S.N., Birk, B., Alépée, N., Haupt, T., Kathawala, R., Kern, P., Nardelli, L., Raabe, H., Rucki, M., Ryan, C., Verkaart, S., Westerink, W., Landsiedel, R., Natsch, A. (2020), 'The kinetic Direct Peptide Reactivity Assay (kDPRA): Intra- and inter-laboratory reproducibility in a seven-laboratory ring trial', ALTEX, preprint, DOI: 10.14573/altex.2004291.

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

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

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

(21) OECD (2016), 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].

(22) 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).

(23) 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). The formula used to derive accuracy is shown under "Calculation" of predictive capacity.

ADRA: Amino acid Derivative Reactivity Assay.

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

Calculation

Calculating depletion of either NAC or NAL

Depletion is calculated as follows:

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

Calculating predictive capacity

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

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

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

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

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.

kDPRA: kinetic Direct Peptide Reactivity Assay.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Sensitivity: The proportion of all positive/active chemicals that are correctly classified by the test method. It is a measure of accuracy for a test method that produces categorical results and is an important

consideration in assessing the relevance of a test method (1). The formula used to derive sensitivity is shown under ”Calculation” of predictive capacity.

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

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

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

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

TFA: Trifluoroacetic acid.

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

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

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

Literature for definitions

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

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

(3) United Nations (UN) (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) (2018), Guidance for Industry: Bioanalytical Method Validation 41 pp. Accessible at: https://www.fda.gov/files/drugs/published/Bioanalytical-Method-Validation-Guidance-for-Industry.pdf

APPENDIX I

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

INITIAL CONSIDERATIONS, APPLICABILITY AND LIMITATIONS

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

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

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

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

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

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

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

PRINCIPLE OF THE TEST

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

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

PROCEDURE

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

Preparation of the cysteine or lysine-containing peptides

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

Preparation of the test chemical

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

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

Preparation of the positive control, reference controls and coelution controls

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

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

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

Preparation of the HPLC standard calibration curve

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

HPLC preparation and analysis

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

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

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

DATA AND REPORTING

Data evaluation

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

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

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

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

Acceptance criteria

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

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

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

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

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

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

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

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

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

Prediction model

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

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

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

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

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

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

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

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

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

24. A single HPLC analysis for both the cysteine and the lysine peptide should be sufficient for a test chemical when the result is unequivocal. However, in cases of results close to the threshold used to discriminate between positive and negative results (i.e. mean percent depletion falls in the range of 3% to 10% for the cysteine 1:10/lysine 1:50 prediction model or cysteine percent depletion falls in the range of 9% to 17% for the cysteine 1:10

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

Test report

25. The test report should include the following information

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

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

Peptides

- Supplier, lot, purity

HPLC instrument setting and analysis

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

System suitability

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

Analysis sequence

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

Proficiency testing

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

Discussion of the results

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

Conclusion

LITERATURE FOR APPENDIX I

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

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

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

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

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

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

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

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

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

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

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

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

APPENDIX I, ANNEX 1

KNOWN LIMITATIONS OF THE DIRECT PEPTIDE REACTIVITY ASSAY

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

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

Test chemicals that do not covalently bind to the cysteine-peptide but promote its oxidation (i.e. cysteine dimerisation)	Could lead to a potential over-estimation of cysteine-peptide depletion, resulting in possible false positive predictions.		DMSO Oxidant
Test chemicals that are only soluble in DMSO	DMSO causes excessive peptide depletion due to cysteine dimerization resulting in high background cysteine depletion.	May lead to false negative results	n/a

APPENDIX I, ANNEX 2

PROFICIENCY SUBSTANCES

In Chemico Skin Sensitisation: Direct Peptide Reactivity Assay

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

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

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

[1]The *in vivo* hazard and (potency) predictions are based on LLNA data (5). The *in vivo* potency is derived using the criteria proposed by ECETOC (8).

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

[3] Ranges determined on the basis of at least 10 depletion values generated by 6 independent laboratories.

APPENDIX I, ANNEX 3

EXAMPLES OF ANALYSIS SEQUENCE

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

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

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

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

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

APPENDIX II

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

INITIAL CONSIDERATIONS, APPLICABILITY AND LIMITATIONS

1. The ADRA is proposed to address the molecular initiating event of the skin sensitisation AOP—namely, protein reactivity—by quantifying the reactivity of test chemicals towards model synthetic amino acid derivatives containing either lysine or cysteine (1) (2) (3). Depletion values of cysteine and lysine derivatives are then used to support the discrimination between skin sensitisers and non-sensitisers (1) (2) (3).

2. 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 (6) (7).

3. The term "test chemical" is used in this Test Guideline to refer to what is being tested and is not related to the applicability of the ADRA to the testing of substances and/or mixtures (see a summary of the known limitations of the ADRA in Annex 1 of this Appendix). 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).

4. The ADRA test method allows testing of poorly soluble chemicals (8). To be tested, a test chemical should be soluble in an appropriate solvent at a final concentration of 1 mM (see paragraph 14). Test chemicals that are not soluble at this concentration may still be tested at lower concentrations. In such cases, a positive result could still be used to support identification of the test chemical as a skin sensitiser but no firm conclusion on the lack of reactivity should be drawn from a negative result.

5. 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 (9).

6. 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 (10) (11). 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 (10) (11). Thus, although it is impossible to define fixed methods in this guideline, which can evaluate multi-constituent substances and mixtures, paragraph 15 describes an evaluation method that is considered to be applicable at the present time for multi-constituent substances or mixtures of known composition (10). 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.

7. 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 (6).

PRINCIPLE OF THE TEST

8. 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).

9. 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 2 of this Appendix.

PROCEDURE

10. This test method is based on the protocol (12) 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 2 of this Appendix.

Quality of NAC and NAL

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

Quality required for NAC and NAL:

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

2) Stability: Using NAC and NAL stock solution, prepare a reference control free of any test chemical and quantify the residual levels of NAC and NAL both immediately after preparation (0 hours) and after a 24 hour incubation. Residual levels of NAC and NAL are to be a minimum of 90% in either case (12). 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 2 and should satisfy the requirement given therein.

Preparation of the NAC and NAL stock solution

12. 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

13. 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 (12). An appropriate solvent should dissolve the test chemical completely. Since the ADRA protocol stipulates that either NAC or NAL are incubated in an excess volume of the test chemical, visual inspection of the clear test chemical solution is considered sufficient to confirm that the test chemical (and all its constituents, if testing a multi-constituent substance or a mixture) is dissolved. 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 (13) and as a result, it may be more difficult to meet the acceptance criteria. If a DMSO-acetonitrile solvent is chosen (5% DMSO in acetonitrile), the test chemical should be dissolved at 20 mM 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.

14. 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 (10). 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.

15. 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.

Moreover, since mixtures consisting of multiple components cannot be evaluated for sensitisation for each chemical, positive result could still be used to support identification of the test chemical as a skin sensitiser but no firm conclusion should be drawn from a negative result.

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

16. Either phenylacetaldehyde (CAS 122-78-1, purity ≥ 90%) or squaric acid diethyl ester (CAS 5231-87-8, purity > 95%) should be used as the positive control (PC) at a concentration of 1 mM in acetonitrile. Phenylacetaldehyde should not be stored for a long time after opening since it may be polymerised and oxidised; it is recommended that it is

purchased as appropriate while paying attention to the NAC and NAL depletion. Squaric acid diethyl ester should be stored away from high temperature or humidity, since it may be hydrolysed. 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 3). The percent NAC and NAL depletion for a test chemical is calculated using an appropriate reference control for that test chemical (see paragraph 23). Also, a co-elution control comprising only the test chemical should be included in the run sequence to detect possible co-elution of the test chemical with either the NAC or NAL.

Incubation of the test chemical with the NAC and NAL solutions

17. 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

18. 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).

19. 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 $R^2 > 0.990$.

20. 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, column size: 3.0 × 150 mm) as preferred column. With this reversed-phase HPLC column, the entire system should be equilibrated for at least 30 minutes at 40ºC with 50% phase A (0.1% (v/v) trifluoroacetic acid in water), 50% phase B (0.1% (v/v) trifluoroacetic acid in acetonitrile) before use. Then, the column is conditioned by running the gradient at least twice before actual use. The HPLC analysis should be performed using a flow rate of 0.30 mL/min and a linear gradient from 30% to 55% acetonitrile for NAC and from 25% to 45% acetonitrile for NAL within 10 minutes, followed by a rapid increase to 100% acetonitrile to remove other materials. Equal volumes of the standard solutions, test chemical solutions, and control solutions should be injected. The column should be re-equilibrated under initial conditions for 6.5 minutes between injections. If a different reversed-phase HPLC column is used, the set-up parameters described above may need to be adjusted to guarantee an appropriate elution and integration of the NAC and NAL, including the injection volume, which may vary according to the system used (typically in the range from 10–20 µL). Importantly, if an alternative HPLC set-up is used, its equivalence to the validated set-up described above should be demonstrated, preferably by testing the proficiency substances in Annex 2. 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 3.

21. 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

22. 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.

23. 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 3) according to the formula described below.

$$\text{Percent NAC or NAL depletion} = \left[1-\left[\frac{\text{NAC or NAL peak area in replicate injection}}{\text{Mean NAC or NAL peak area in reference controls C}}\right]\right]x100$$

Acceptance criteria

24. 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 and the maximum standard deviation (SD) of the three replicates for the positive control (phenylacetaldehyde or squaric acid diethyl ester) should meet the following criteria:

- NAC depletion

Phenylacetaldehyde: 6 - 30%; Squaric acid diethyl ester: 15 - 40 %

- NAL depletion

Phenylacetaldehyde: 75 - 100%; Squaric acid diethyl ester: 40 - 85 %

- Maximum standard deviation (SD)

Both phenylacetaldehyde and squaric acid diethyl ester: < 10%, for both NAC and NAL depletion,

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

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

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

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

b) the mean NAC and NAL concentration of the three Reference Controls C in the appropriate solvent should be 3.2–4.4 µM. The permissible range of the mean NAC concentration of Reference Control C when 5% DMSO in acetonitrile is used as a solvent is 2.8 to 4.0 µM (14).

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

Prediction model

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

Table 1: NAC/NAL prediction model[1]

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

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

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

27. 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 2. When co-elution occurs, it is not possible to integrate the peak of the NAC or NAL, thereby preventing calculation of the percent depletion of NAC or NAL. If co-elution of test chemicals occurs with both the NAC and NAL, or with the NAC only, 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. The 5.6% cut-off value for the percent depletion of NAC was set by using 2 class classification model so that the sensitizer and non-sensitizer could be predicted most appropriately.

Table 2: NAC-only prediction model[1]

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

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

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

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

Test report

29. The test report should include the following information:

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

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

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

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

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

HPLC instrument setting and analysis

- Type of HPLC instrument, HPLC and guard columns, 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

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

Discussion of the results

- Description of any unintended modifications to the test procedure.
- Discussion of the results obtained with the ADRA test method and if it is within the ranges described in paragraph 28.

Conclusion

Literature for Appendix II

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

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

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

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

(5) OECD (2019), 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, Akimoto M, Katsuoka Y, Ono A, Kasahara T (2019) Applicability of amino acid derivative reactivity assay for prediction of skin sensitization by combining multiple alternative methods to evaluate key events. Journal of Toxicological Sciences, 44, 585-600. doi: 10.2131/jts.44.585.

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

(9) Fujita 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.

(10) YamamotoY, Fujita M, Wanibuchi 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

(11) 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.

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

(13) 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

(14) 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 (2020) Oxidation of a cysteine-derived nucleophilic reagent by dimethyl sulfoxide in the amino acid derivative reactivity assay. J Appl Toxicol., 40, 843-854. DOI: 10.1002/jat.3948.

(15) 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.

(16) 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.

(17) 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.

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

APPENDIX II, ANNEX 1

KNOWN LIMITATIONS OF THE AMINO ACID DERIVATIVE REACTIVITY ASSAY

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

Substance class / interference	Reason for potential underprediction or interference	Data interpretation	Example substance
Metals and inorganic compounds	Known to react with proteins via mechanisms other than covalent binding	Should not be tested	Nickel sulphate; 7786-81-4
Pro-haptens	Test Chemicals that require enzymatic bioactivation to exert their skin sensitisation potential ; cannot be detected by the test method unless activation is caused by auto-oxidation to a similar degree as in vivo /in humans. It will however normally not be known whether this will be the case	May lead to false negatives. Negative results obtained with the test method should be interpreted in the context of the stated limitations and in the connection with other information sources within the framework of an IATA	Diethylenetriamine; 111-40-0 (human 1A, LLNA n/a)
Pre-haptens	Chemicals that become sensitisers after abiotic transformation are reported to be in some cases correctly detected by the test method		Linalool: 78-70-6
Test chemicals absorbing significantly at 281 nm and having the same retention time of the NAC or NAL (co-elution)	When co-elution occurs the peak of the NAC or NAL cannot be integrated and the calculation of the percent NAC or NAL depletion is not possible.	The substances that absorb UV in this range of the spectrum are generally limited to those having conjugated double bonds, which significantly lowers the potential for co-elution. If co-elution of such test chemicals occurs with both the NAC and the NAL or with the NAC only, then the analysis should be reported as “inconclusive” and alternative HPLC set up should be considered (see paragraph 27). In cases where co-elution occurs only with the NAL, then the NAC-only prediction model reported in Table 2 can be used.”	Safranal; 116-26-7
Complex mixtures of unknown composition, substances of unknown or variable composition, complex reaction products or biological materials	This is due to the need for defined molar ratio of test chemical and nucleophilic reagent. Limited information is currently available on the applicability of the ADRA	n/a	UVCBs, chemical emissions, products or formulations with variable or not fully known composition
Test chemicals which cannot be dissolved in an appropriate solvent at a final concentration of 1 mM (although it is unlikely to happen since test chemical	Not sure if sufficient exposure can be achieved	The ADRA test method allows testing of poorly soluble chemicals. Test chemicals that are not soluble at this concentration though may still be tested at lower soluble concentrations. In such a case, a positive result could be used to support the	n/a

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

APPENDIX II, ANNEX 2

PROFICIENCY SUBSTANCES

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

Prior to routine use of the test method, laboratories should demonstrate technical proficiency by correctly obtaining the expected ADRA prediction for the 10 proficiency substances recommended in Table 1 and by obtaining NAC and NAL depletion values that fall within the respective reference ranges for 8 out of the 10 proficiency substances. 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	Diphenylcyclopropenone	4886-38-4	Solid	206.24	Sensitiser (strong)	Positive	15-45	≤ 10
3	2-Methyl-2H-isothiazol-3-one	2682-20-4	Solid	115.15	Sensitiser (moderate)	Positive	80-100	≤ 7
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	Isopropanol	67-63-0	Liquid	60.10	Non-sensitiser	Negative	≤ 7	≤ 7
9	Dimethyl isophthalate	1459-93-4	Solid	194.19	Non-sensitiser	Negative	≤ 7	≤ 7
10	Propyl paraben	94-13-3	Solid	180.20	Non-sensitiser	Negative	≤ 7	≤ 7

[1]The *in vivo* hazard (and potency) predictions are based on LLNA data. ((15) (16) (17). The *in vivo* potency is derived using the criteria proposed by ECETOC (18).

[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 3

EXAMPLES OF ANALYSIS SEQUENCE

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

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

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

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

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

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

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

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

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

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

APPENDIX III

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

INITIAL CONSIDERATIONS, APPLICABILITY AND LIMITATIONS

1. The kDPRA is proposed to address the molecular initiating event of the skin sensitisation AOP - namely, protein reactivity - by quantifying the reactivity of test chemicals towards a synthetic model peptide containing cysteine in a time- and concentration dependent manner (1) (2). Kinetic rate constants are calculated and the logarithm of the maximum rate constant (log k_{max} value in $s^{-1}M^{-1}$) for a tested substance is then used to support the discrimination of UN GHS subcategory 1A skin sensitisers (subcategory 1A) from those not categorised as subcategory 1A (non-subcategory 1A) i.e., subcategory 1B or no category according to UN GHS (3). Based on theoretical consideration, the rate constant of the reaction between a test chemical and skin proteins will determine the amount of epitope formed from a given amount of chemical or, vice-versa, determine the dose needed to form the amount of epitope needed for induction of sensitization to occur and it is thus a rate limiting and potency determining step. Based on empirical evidence when evaluating 180 chemicals, the rate constant was shown to be the strongest determinant of potency among all evaluated parameters measured in OECD 442C, 442D and 442E (3).
2. The kDPRA proved to be transferable to laboratories without hands-on training (4). For the 24 test chemicals tested during the validation study, the overall within-laboratory reproducibility of kDPRA for assigning UN GHS subcategory 1A was 96% and the average between-laboratory reproducibility was 88% (4). Results from the validation study (4) as well as from other published studies (3) encompassing 180 test chemicals that fall within kDPRA's applicability domain indicate that kDPRA allows to discriminate UN GHS subcategory 1A skin sensitisers from those not categorised as subcategory 1A (non-subcategory 1A) according to UN GHS with a balanced accuracy of 85%, a sensitivity of 84% (38/45), and a specificity of 86% (116/135) relative to LLNA results (3). Similar performances were obtained when comparing kDPRA outcomes with the OECD LLNA database compiled within the context of the Test Guideline on Defined Approaches for Skin Sensitization (15)[3]. In addition, the prediction for 123 test chemicals (out of the 180) having human skin sensitisation data (5) (6) has a balanced accuracy of 76%, a sensitivity of 64% (21/33), and a specificity of 89% (80/90) (3), although the human reference data are subject to a significant uncertainty[4]. Furthermore, when evaluating non-animal methods for skin sensitisation, it should be kept in mind that the

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

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

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

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

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

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

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

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

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

PRINCIPLE OF THE TEST

8. The kDPRA is a modification of the *in chemico* test method DPRA (described in Appendix I of this Test Guideline). The kDPRA uses the cysteine peptide (Ac-RFAA**C**AA-COOH) also used in the DPRA, while it does not use a lysine containing peptide. The final concentration of the test peptide (0.5 mM) and the reaction medium (25% acetonitrile in phosphate buffer) is identical in the kDPRA and in the DPRA. While the DPRA measures only at one concentration of the test chemical (5 mM for the cysteine peptide) and at one time point (≥ 24 h), the kDPRA performs parallel reactions at five concentrations (5, 2.5, 1.25, 0.625 and 0.3125 mM) and at six time-points (10, 30, 90, 150, 210 and 1440 min) at 25±2.5ºC. Residual concentration of the cysteine peptide after the respective reaction time is measured after stopping the reaction by the addition of monobromobimane (mBrB; CAS 74235-78-2). The highly reactive and non-fluorescent mBrB rapidly reacts with unbound cysteine moieties of the model peptide to form a fluorescent complex which is measured in order to quantify the non-depleted peptide concentration. If the depletion of the highest concentration surpasses the threshold of 13.89% (cut-off used in the DPRA for positivity in the cysteine only prediction model) and this depletion is statistically significant vs. controls with peptide

only, further calculations are performed (otherwise the test chemical is considered to be non-reactive according to the prediction model shown in paragraph 28). The natural logarithm of the non-depleted peptide concentrations is plotted vs. the concentration of the test chemical at each time point. If a linear relationship is observed (correlation coefficient > 0.90), the slope of this curve is determined and divided by the incubation time to calculate the rate constant in [$min^{-1}mM^{-1}$]. This value is transformed to the rate constant in [$s^{-1}M^{-1}$] and the logarithm is calculated. The maximum value observed at any time point is taken as the log k_{max}, and this maximum rate constant is the primary read-out of the test. It gives a quantification of the maximum kinetic rate of the reaction of the test chemical with the test peptide. Kinetic reaction rates of the cysteine peptide depletion are then used to discriminate UN GHS subcategory 1A skin sensitisers from those not categorised as 1A (non-subcategory 1A) according to UN GHS. Chemicals with a log $k_{max} \geq -2.0$ are predicted as UN GHS subcategory 1A. The kinetic rate constant may be further used in integrated approaches such as IATA or DA to assess the skin sensitisation potency of a test chemical in a continuous scale as needed for risk assessment (3) (10).

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

PROCEDURE

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

Preparation of the cysteine-peptide

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

Preparation of the test chemical

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

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

Preparation of controls

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

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

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

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

Incubation of the test chemical with the cysteine peptide solution

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

Fluorescence measurement

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

DATA AND REPORTING

Data evaluation

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

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

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

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

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

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

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

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

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

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

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

Acceptance criteria

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

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

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

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

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

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

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

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

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

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

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

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

Prediction model

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

Table 1: kDPRA prediction model

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

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

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

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

Test report

31. The test report should include the following information

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

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

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

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

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

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

Concentration(s) tested;

Storage conditions and stability to the extent available.

Additional information for positive control

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

Additional information for solvent/vehicle control

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

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

Peptide

Supplier, lot, purity

Fluorescence analysis

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

Proficiency testing

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

Discussion of the results

Description of any unintended modifications to the test procedure.

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

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

Conclusion

LITERATURE FOR APPENDIX III

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

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

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

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

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

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

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

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

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

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

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

(12) Manuscript in prepration: Borderline ranges for *in chemico* and *in vitro* skin sensitization OECD test guideline methods determined from ring trials

(13) ICCVAM, ICCVAM Test Method Evaluation Report: Usefulness and Limitations of the Murine Local Lymph Node Assay for Potency Categorization of Chemicals Causing Allergic Contact Dermatitis in Humans. 2011. NIH Publication 11-7709.

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

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

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

APPENDIX III, ANNEX 1

KNOWN LIMITATIONS OF THE KINETIC DIRECT PEPTIDE REACTIVITY ASSAY

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

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

composition, complex reaction products or biological materials			variable or not fully known composition
Test chemicals which cannot be dissolved in water or acetonitrile or a compatible water-miscible solvent	Not sure if sufficient exposure can be achieved	In such cases, a log k_{max} > -2.0 could still be used to support the identification of the test chemical as a UN GHS subcategory 1A skin sensitiser but no firm conclusion should be drawn in case log k_{max} is < -2.0. Alternative vehicle may be used according to the prescriptions given in paragraph 12.	n/a
Test chemicals which precipitate in reaction solution	Not sure if sufficient exposure can be achieved: If a precipitate is observed immediately upon addition of the test chemical solution to the peptide solution, due to low aqueous solubility of the test chemical, one cannot be sure how much test chemical remained in the solution to react with the peptide.	In such a case, a positive result (i.e. log kmax ≥ -2.0) could still be used, but a negative result (i.e., non-reactive or log kmax < -2.0 outcome) should be interpreted with due care (see also provisions in paragraph 4 for the testing of chemicals not soluble up to a concentration of 20 mM in the kDPRA).	Methyl-2-nonynoate[5]; 111-80-8; LLNA NC
Test chemicals promoting cysteine-peptide oxidation		May lead to a potential over estimation of peptide reactivity.	DMSO

[5] Roberts, D.W. and A. Natsch, *High throughput kinetic profiling approach for covalent binding to peptides: Application to skin sensitization potency of michael acceptor electrophiles.* Chem. Res. Toxicol., 2009. **22**(3): p. 592-603

APPENDIX III, ANNEX 2

PROFICIENCY SUBSTANCES

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

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

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

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

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

[2] Rounded ranges determined on the basis of at least 14 log k_{max} determinations generated by 7 independent laboratories.

[3] Non sensitisers according to the UN GHS.

Section 4
Health effects

Guideline No. 497

Guideline on Defined Approaches for Skin Sensitisation

14 June 2021

OECD Guidelines for the Testing of Chemicals

OECD
BETTER POLICIES FOR BETTER LIVES

OECD GUIDELINE FOR TESTING OF CHEMICALS

Defined Approaches for Skin Sensitisation

Table of Contents

1. Section 1-Introduction

1.1. General Introduction

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 initiated by covalent binding to proteins has been summarised as an Adverse Outcome Pathway (AOP) (2) that begins with a molecular initiating event, leading to intermediate key events, and terminating with the adverse effect, allergic contact dermatitis.

2. The skin sensitisation AOP focuses on chemicals that react with amino acid residues (*i.e.* cysteine or lysine) such as organic chemicals. In this instance, the molecular initiating event (*i.e.* the first key event), is the covalent binding of electrophilic substances to nucleophilic centres in skin proteins. The second key event in this AOP takes place in the keratinocytes and includes inflammatory responses as well as changes in gene expression associated with specific cell signalling pathways such as the antioxidant/electrophile response element (ARE)-dependent pathways. The third key event is the activation of dendritic cells, typically assessed by expression of specific cell surface markers, chemokines and cytokines. The fourth key event is T-cell proliferation, and the adverse outcome is presentation of allergic contact dermatitis.

3. 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) (3) assess both the induction and elicitation phases of skin sensitisation. The murine tests, such as the LLNA (OECD TG 429) (4) and its three non-radioactive modifications — LLNA: DA (OECD TG 442A) (5), LLNA: BrdU-ELISA, and BrdU-FCM (OECD TG 442B) (6) — 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.

4. Mechanistically-based *in chemico* and *in vitro* test methods (OECD TG 442C, 442D, 442E) (7, 8, 9) addressing the first three key events (KE) of the skin sensitisation AOP can be used to evaluate the skin sensitisation hazard potential of chemicals. None of these test methods are considered sufficient stand-alone replacements of animal data to conclude on skin sensitisation potential of chemicals or to provide information for potency sub-categorisation according to the UN GHS (sub-categories 1A and 1B). However, data generated with these *in chemico* and *in vitro* methods addressing multiple KEs of the skin sensitisation AOP are proposed to be used together, as well as with information sources such as *in silico* and read-across predictions from chemical analogues, within integrated approaches to testing and assessment (IATA) or defined approaches (DAs). Results from the individual information sources can only be used in DAs if the substances fall within the applicability domains of the methods (see "Initial Considerations, Applicability and Limitations" sections of respective methods (TG 442C, Appendix 1; TG 442D, Appendix 1A; TG 442E Annex 1) (7, 8, 9).

5. Results from multiple information sources can be used together in DAs to achieve an equivalent or better predictive capacity than that of the animal tests to predict responses in humans. A DA consists of a fixed data interpretation procedure (DIP) (*e.g.* a mathematical model, a rule-based approach) 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 without the need for expert judgment. Individual DAs for skin sensitisation and their respective information sources were originally described in Guidance Document 256, Annex I/II (10) and a preliminary assessment was published in Kleinstreuer et al (11). The DAs use method combinations intended to overcome some of the limitations of the individual, stand-alone methods in order to provide increased confidence in the overall result obtained. The ultimate goal of DAs is to provide information that is equivalent to that provided by animal studies, *i.e.* information that can be used for hazard identification and/or potency categorisation.

6. Testing laboratories should consider all relevant available information on the test chemical prior to conducting the studies as directed by a DA. Such information could include, for example, the identity and chemical structure of the test chemical and its physico-chemical properties. Such information should be considered in order to determine whether the individual OECD test guideline methods under a specific DA are applicable for the test chemical.

7. When performing a hazard evaluation and/or potency sub-categorisation based on the output from an *in vivo* (LLNA or any other) test, from an *in chemico* test, from an *in vitro* test, from an *in silico* approach, from a DA, and any combination thereof, the same principles always apply, *i.e.* all available information relevant to the chemical in question should be taken into consideration as well as toxicological data on structurally related test chemicals if available.

8. This Guideline was developed with the input of an OECD Expert Group on Defined Approaches for Skin Sensitisation (EG DASS) comprised of scientific experts from regulatory agencies, validation bodies, non-governmental organisations, and industry.

9. Three rule-based DAs are included in this Guideline, and are described with respect to their intended regulatory purpose: hazard identification, *i.e.* discrimination between skin sensitisers and non-sensitisers (1.4.Part I), or potency sub-categorisation (1.6.Part II). The DAs included in Part II are also suitable for hazard identification. The evaluation and review of the DAs are described in detail in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (12).

10. A comprehensive dataset of 196 chemicals with DA predictions, data on individual information sources, highly curated LLNA and Human Patch Predictive Test (HPPT) data, and physicochemical properties, was compiled and is attached as **Annex 2** to the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (12). Out of the 196 chemicals, 168 chemicals have LLNA classifications and 66 chemicals have HPPT classifications, which were all agreed upon by the EG DASS and used to evaluate the performance of the DAs. Due to the availability of data, this dataset contains mainly cosmetic ingredients but also other types of chemicals that are used across sectors such as preservatives, dyes, or food ingredients. The dataset is chemically diverse as shown by the physicochemical properties covered by these chemicals: it contains small and large molecules (molecular weight ranges from 30 to 512 g/mol), hydrophobic and hydrophilic substances (Log P ranges from -3.9 to 9.4), solids and liquids (melting point ranges from -122 to 253 ºC), volatile and non-volatile substances (boiling point ranges from -19 to 445 ºC). Further details on the chemical space characterization of the reference

database are available in **Section 4** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (12).

11. Other DAs may be included in this Guideline following future review and approval. DAs able to provide a quantitative measure of sensitisation potency, such as a point of departure which can be used for risk assessment, may be included in a new Part II to this Guideline in the future.

1.2. DAs and Use Scenarios included in the Guideline

12. The DAs currently described in this guideline are:

- The "2 out of 3" (2o3) defined approach to skin sensitisation hazard identification based on *in chemico* (KE1) and *in vitro* (KE2/KE3) data (13, 14). See Part I.
- The integrated testing strategy (ITSv1) for UN GHS potency categorisation based on *in chemico* (KE1) and *in vitro* (KE3) data, and *in silico* (Derek Nexus) predictions (14, 15), with a DIP developed with expert group (EG DASS) input. See Part II Potency Categorisation.
- A modification of the integrated testing strategy (ITSv2) for UN GHS potency categorisation based on *in chemico* (KE1) and *in vitro* (KE3) data, and *in silico* (OECD QSAR Toolbox) predictions, with a DIP developed with expert group (EG DASS) input. See Part II Potency Categorisation.

13. The DAs described in this guideline are based on the use of validated OECD test methods (DPRA, KeratinoSens™, h-CLAT), for which transferability, within- and between-laboratory reproducibility have been characterised in the validation phase (7, 8, 9).

14. The ITS DAs (ITSv1 and ITS v2) also make use of an *in silico* information source; Derek Nexus v6.1.0 (ITSv1), or OECD QSAR Toolbox v4.5 (ITSv2). Derek Nexus (referred to as Derek hereafter) is an expert knowledge-based tool which provides predictions of skin sensitisation potential using structural alerts, and OECD QSAR Toolbox (referred to as OECD QSAR TB hereafter) is a computational tool which uses an analogue-based read-across approach or structural alerts for protein binding identified by profilers to predict whether a chemical will be a sensitiser.

15. All DAs described in this guideline can each be used to address countries' requirements for discriminating between sensitisers (*i.e.* UN GHS Category 1) from non-sensitisers, though they do so with different sensitivities and specificities (detailed in the respective descriptions of each DA).

16. The ITS DAs (ITSv1 and ITS v2) can also be used to discriminate chemicals into three UN GHS potency categories (Category 1A = strong sensitisers; Category 1B = other sensitisers, and No Categorization (NC = not classified).

17. The known limitations and applicability domains of the individual information sources were used to design workflows for assigning confidence to each of the predictions produced by the DAs described in this guideline. In order to have a high confidence prediction, the underlying data must meet criteria in the respective test guidelines (see TG 442C, Appendix 1; TG 442D, Appendix 1A; TG 442E Annex 1 (7, 8, 9)), DA predictions with high confidence for hazard identification and/or potency are considered conclusive. DA predictions with low confidence are considered inconclusive for hazard identification and/or potency (see **Sections 2.1.4** and **3.1.4** for further information). These 'inconclusive'

predictions may nevertheless be considered in a weight-of-evidence approach and/or within the context of an IATA together with other information sources (*e.g.* demonstration of exposure to the test system, existing *in vivo* data, clinical data, read-across, other *in vitro* / *in chemico* / *in silico* data, etc.).

18. The performance of the DAs described in this guideline for discriminating between sensitisers and non-sensitisers was evaluated using 168 (135 GHS Skin Sens. Category 1, and 33 no classification) test chemicals for which DPRA, KeratinoSens™, h-CLAT, Derek, OECD QSAR TB predictions and classifications based on LLNA reference data agreed upon by the EG DASS are available (for additional details see **Section 2.1** and **Annex 3** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation*) (12). For the purpose of evaluating the performance of the ITS DAs for predicting UN GHS classifications based on potency categorization (sub-category 1A, 1B, or "not classified" (NC)), 156 test chemicals (38 1A, 85 1B, and 33 NC) were used because for 12 test chemicals it was not possible to assign with sufficient confidence the potency sub-category 1A or 1B on the basis of LLNA data. Mixtures and botanicals with undefined structural composition were excluded from the curated LLNA reference data.

19. The performance of the three DAs (high confidence predictions only) against the LLNA reference data for predicting skin sensitisation hazard showed balanced accuracies (average of sensitivity and specificity; BA) in the range of 80-84%, with sensitivities of 82-93% and specificities of 67-85% (see **Table 1.1**). Note that specificity measures are more uncertain than sensitivities due to lower number of negative reference chemicals. Detailed performance statistics are reported in Part I (2o3 DA) and Part II (ITS DA). The performance of the ITSv1 and ITSv2 DAs for UN GHS classifications based on potency categorization (high confidence predictions only, sub-category 1A, 1B, or NC) when compared to the LLNA reference data yielded overall accuracies of 71%, overall balanced accuracies of 78% (ITSv1) or 77% (ITSv2), and balanced accuracies within a predicted sub-category or NC ranging from 72-81% (ITSv1) or 71-80% (ITSv2). There were no strong sensitisers (1A) that were incorrectly predicted as being a non-sensitiser (NC) or vice versa. Detailed performance statistics are reported in Part II and in **Section 5** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (12).

20. The performance of the DAs described in this guideline for discriminating between sensitisers and non-sensitisers was also evaluated using a set of 66, or 65 for 2o3, due to lack of assay data for one chemical, test chemicals (55 sensitisers and 11 non-sensitisers) for which classifications based on Human Predictive Patch Test (HPPT) data have been agreed upon by the EG DASS (for additional details see **Section 2.2** and **Annex 4** *of the Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation*) (12). For the purpose of evaluating the performance of the ITS DAs for predicting UN GHS classifications based on potency categorization (sub-category 1A, 1B, or NC), 63 test chemicals were used (21 1A, 31 1B, and 11 NC) because for 3 test chemicals it was not possible to assign with sufficient confidence the potency sub-category 1A or 1B on the basis of human reference data. Mixtures and botanicals with undefined structural composition were excluded from the curated human reference data.

21. The performance of the DAs (high confidence predictions only) against the human reference data for predicting skin sensitisation hazard showed balanced accuracies in the range of 69-88%, with sensitivities of 89-94% and specificities of 44-88% (see **Table 1.1**). Note that specificity measures are more uncertain than sensitivities due to lower number of negative reference chemicals. Detailed performance statistics are reported in Part I (2o3

DA) and Part II (ITS DA). The performance of the ITSv1 and ITSv2 DAs for UN GHS skin sensitisation potency classification (high confidence predictions only, sub-category 1A, 1B and NC) when compared to the human reference data yielded overall balanced accuracies of 72% (ITSv1) or 73% (ITSv2), and balanced accuracies within a predicted sub-category or NC in the range of 68-79% (ITSv1) or 69-79% (ITSv2). Detailed performance statistics are reported in Part II and in **Section 5** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (12).

22. The overlap between the LLNA and human reference datasets was 56 chemicals for hazard and 47 chemicals for skin sensitisation potency categorisation, respectively, and the performance of the LLNA against the human reference data was evaluated using these chemicals as a basis for comparison. The performance of the LLNA against the human reference for predicting skin sensitisation hazard showed a balanced accuracy of 58%, with sensitivity of 94% and specificity of 22%. Note that the specificity measure is more uncertain than the sensitivity due to a lower number of negative reference chemicals. The performance of the LLNA for UN GHS potency classification when compared to the human reference data yielded an overall balanced accuracy of 64%, and balanced accuracies within a predicted sub-category or NC in the range of 59-73% There were no strong skin sensitisers (1A) in the human reference data that were incorrectly predicted by the DAs, or by the LLNA as not being a sensitiser (no classification) or vice versa. Detailed performance statistics are reported Part I and Part II

Table 1.1. Summary of the DAs Included in this Guideline

DA/Method	Information Sources	Capability (Hazard and/or Potency)	Hazard Performance vs. LLNA	Hazard Performance vs. Human	Potency Performance vs. LLNA (Accuracy)	Potency Performance vs. Human (Accuracy)
2o3 DA	DPRA, KeratinoSens™, h-CLAT	Hazard	84% BA, 82% Sens, 85% Spec	88% BA, 89% Sens, 88% Spec	-	-
ITSv1 DA	DPRA, h-CLAT, DEREK Nexus v6.1.0	Hazard, Potency	81% BA, 92% Sens, 70% Spec	69% BA, 93% Sens, 44% Spec	70% NC, 71% 1B, 74% 1A	44% NC, 77% 1B, 65% 1A
ITSv2 DA	DPRA, h-CLAT, OECD QSAR Toolbox v4.5	Hazard, Potency	80% BA, 93% Sens, 67% Spec	69% BA, 94% Sens, 44% Spec	67% NC, 72% 1B, 72% 1A	44% NC, 80% 1B, 67% 1A
LLNA (provided for comparison)	*in vivo*	Hazard, Potency	-	58% BA, 94% Sens, 22% Spec	-	25% NC, 74% 1B, 56% 1A

Note: For hazard performance, sensitivity (Sens) is the true positive rate, specificity (Spec) is the true negative rate, and balanced accuracy (BA) is the average of sensitivity and specificity. Due to the imbalanced nature of the reference data, the measures of specificity are more uncertain than the measures of sensitivity. For potency performance, accuracy reflects correct classification rate within each UN GHS sub-category. Due to the imbalanced nature of the reference data, the measures of accuracy are more uncertain for smaller classes, *e.g.* for NC chemicals. Statistics reflect conclusive DA predictions only. This represents the data available at the time of initial guideline adoption.

1.3. Limitations

23. **Table 1.1** provides an overview of the DAs included in this Guideline, their information sources used, whether they provide hazard and/or potency prediction, and

summarises their performance against the LLNA and human reference data. The LLNA (OECD TG 429) is included in **Table 1.1** as a basis for comparison. More details are provided in Part I and Part II of this Guideline, as well as in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (12).

24. The identified limitations of the DAs and their individual components are summarised below.

1.3.1. Limitations of individual in chemico/in vitro *information sources*

25. Users should refer to the limitations of the individual *in chemico*/*in vitro* test methods as specified in their respective Test Guidelines, which are revised as new data become available and should be consulted regularly. The most up-to-date published version of the respective TGs should always be used. For example, some types of chemicals such as metals, inorganic compounds, UVCBs and mixtures, may not be within the applicability domain for certain test methods. Individual assay results within borderline ranges (**Annex 1**) may yield inconclusive DA predictions. The consideration of limitations of individual *in chemico*/*in vitro* test methods in each DA is detailed in **Section 2.1.4** (**Figure 2.1**) and **Section 3.1.4** (**Figure 3.1**).

1.3.2. Limitations of in silico information sources

26. Some DAs include *in silico* tools as an information source. These tools can either perform automated read-across or (Q)SAR predictions. (Q)SARs include both structure-activity relationship (SAR) models (*i.e.* structural alerts, expert systems) and quantitative structure-activity relationship (QSAR) models (*i.e.* statistical tools). (Q)SAR models should fulfil the OECD Principles for the Validation, for Regulatory Purposes, of (Q)SAR Models and be described in a QSAR Model Reporting Format (QMRF) document (15)[1]. One of the OECD QSAR validation principles refers to a defined domain of applicability. The defined domain of applicability reflects limitations beyond which less reliable predictions may be obtained (*e.g.* training set ranges of descriptors included in the model and types of chemical structures included in the training set). A given *in silico* model may be associated with more than one defined applicability domain, each of which is associated with its own reliability measures as established in the validation. Depending on the DIP, chemicals outside the applicability domain may result in DA predictions of low confidence that are considered inconclusive. Where a DA for skin sensitisation includes an *in silico* tool, users should refer to the limitations and applicability domain of the individual *in silico* tool. Two of the DAs covered in this Guideline, the ITSv1 and the ITSv2, rely upon the *in silico* tools Derek and OECD QSAR TB, respectively, and their specified limitations and applicability domains are detailed in **Annex 2** of this Guideline.

1.3.3. Limitations of DAs

27. The limitations of the DAs are based on the limitations of the individual *in chemico/in vitro/in silico* information sources. Details on using the limitations of individual information sources to determine confidence in DA predictions are provided in **Sections**

[1] The QMRF has been slightly adapted for reporting other *in silico* model predictions in the context of DASS. The adapted QPRF can be found on the OECD site for spreadsheets and software associated with OECD Test Guidelines on Health Effects: https://www.oecd.org/env/ehs/testing/section4software.htm.

2.1.4 and **3.1.4** and in the respective test guidelines (TG 442C, Appendix 1; TG 442D, Appendix 1A; TG 442E, Annex 1) (7, 8, 9).

28. During the evaluation of the DAs covered in this Guideline it was observed that, with respect to LLNA data, the DPRA (TG 442C), KeratinoSens™ (TG 442D), h-CLAT (TG 422E), as well as the proposed DAs, have lower sensitivity for test chemicals with Log P > 3.5 **(**for details see **Section 3.1.4** and **Annex 5** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation*) (12). It was also noted that the LLNA test may produce a higher number of false positive results for these test chemicals when compared with human reference data, and supporting mechanistic information was provided **(**for details see **Section 3.2** and **Annex 6** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation*) (12). Overall, the analyses and the number of reference chemicals with Log P > 3.5 are insufficient to draw firm conclusions. However, according to TG 442E, negative h-CLAT results for substances with Log P > 3.5 should not be considered, and this limitation is applied to the DAs as described in **Sections 2.1.4** and **3.1.4.**

29. For the 2o3 DA, borderline ranges (BRs) have been defined for the individual assays addressing the three KE of the DA, in order to define areas where lower confidence may exist (for details see **Section 2.1.4** and **Annex 1** of this Guideline, and **Section 3.3** and **Annex 7** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation*) (12). Positive and/or negative test results falling within these BRs as well as individual assay limitations, *e.g.* negative h-CLAT results obtained for a chemical with Log P > 3.5 (according to TG 442E), have lower confidence and may result in inconclusive 2o3 DA predictions.

30. Inconclusive DA predictions may nevertheless be considered in a weight-of-evidence approach and/or within the context of an IATA together with other information sources (*e.g.* demonstration of exposure to the test system, existing *in vivo* data, clinical data, read-across, other *in vitro* / *in chemico* / *in silico* data, etc.).

1.4. References

1. United Nations (UN) (2019). Globally Harmonized System of Classification and Labelling of Chemicals (GHS). Eighth revised edition, New York and Geneva, United Nations Publications. Available at: [https://unece.org/ghs-rev8-2019]
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: [https://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm]
3. OECD (1992). OECD Guidelines for the Testing of Chemicals No. 406. Skin Sensitisation. Organisation for Economic Cooperation and Development, Paris. Available at: [https://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm].
4. OECD (2010). OECD Guidelines for Chemical Testing No. 429. Skin sensitisation: Local Lymph Node assay. Organisation for Economic Cooperation and Development, Paris. Available at :(oecd-ilibrary.org).
5. 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:(oecd-ilibrary.org)..
6. 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: (oecd-ilibrary.org).
7. OECD (2020). OECD Guideline for the Testing of Chemicals No. 442C: *In chemico* Skin Sensitisation: Assays addressing the Adverse Outcome Pathway key even on covalent binding to proteins).*In chemico*. Paris, France: Organisation for Economic Cooperation and Development. Available at: (oecd-ilibrary.org).
8. 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: (oecd-ilibrary.org).
9. 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: (oecd-ilibrary.org).
10. 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. Organisation for Economic Cooperation and Development, Paris. Available at: [https://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm].
11. Kleinstreuer N, Hoffmann S, Alepee N, et al. (2018). Non-Animal Methods to Predict Skin Sensitization (II): an assessment of defined approaches. Crit Rev Toxicol Feb 23:1-16. doi: 10.1080/10408444.2018.1429386

12. OECD (2021). Series on Testing and Assessment No. 336: Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation. Organisation for Economic Cooperation and Development, Paris. Available at: [https://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm].

13. Bauch C, Kolle SN, Ramirez T, Eltze T, Fabian E, Mehling A, Teubner W, van Ravenzwaay B, Landsiedel R. (2012). Putting the parts together: combining *in vitro* methods to test for skin sensitizing potentials. Regul Toxicol Pharmacol, 63:489-504.

14. Urbisch D, Mehling A, Guth K, Ramirez T, Honarvar N, Kolle S, Landsiedel R, Jaworska J, Kern PS, Gerberick F, Natsch A, Emter R, Ashikaga T, Miyazawa M, Sakaguchi H. (2015). Assessing skin sensitization hazard in mice and men using non-animal test methods, Regul Toxicol Pharmacol, 71:337-51.

15. ECHA (2008). see "CHAPTER R.6 – QSARS AND GROUPING OF CHEMICALS" *in* Guidance on Information Requirements and Chemical Safety Assessment. European Chemicals Agency [Guidance on Information Requirements and Chemical Safety Assessment - ECHA (europa.eu)].

Part I. – Section 2 - Defined Approaches for Skin Sensitisation Hazard Identification

31. Part I of this guideline applies to DAs that are intended solely for hazard identification, *i.e.* distinguishing between sensitisers and non-sensitisers. A summary of the DAs for hazard identification is provided below; additional detailed information can be found in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1).

2.1. "2 out of 3" Defined Approach

2.1.1. Summary

32. The 2 out of 3 (2o3) DA is intended for the identification of the skin sensitisation hazard of a chemical without the use of animal testing, *i.e.* UN GHS Cat. 1 vs. UN GHS NC. The data interpretation procedure (DIP) is currently not designed to provide information on the potency of a sensitiser.

33. The combination of test methods included in the 2o3 DA covers at least two of the first three KEs of the AOP leading to skin sensitisation as formally described by the OECD: KE1: protein binding (*i.e.* via the direct peptide reactivity assay (DPRA; OECD TG 442C)) (2); KE2: keratinocyte activation (*i.e.* KeratinoSens™; OECD TG 442D) (3); and KE3: dendritic cell activation (*i.e.* via the human cell line activation test (h-CLAT; OECD TG 442E)) (4).

34. The DIP entails that two concordant results obtained from methods addressing at least two of the first three KEs of the AOP determine the final classification. The 2o3 DA was compared to 168 chemicals with curated LLNA reference data agreed upon by the EG DASS and demonstrated an accuracy of 83% and a balanced accuracy of 84% (see **Table 2.1**). The 2o3 DA was also compared to 65 chemicals with curated human reference data agreed upon by the EG DASS and exceeded the accuracy, and balanced accuracy, of the LLNA for hazard identification (see **Tables 2.1-2.2**). It should be noted that due to the imbalanced nature of the reference data (higher numbers of positives than negatives), the measures of balanced accuracy are more uncertain, particularly in the case of the human data comparison.

2.1.2. Data interpretation procedure

35. The data interpretation procedure (DIP) in the 2o3 DA is a transparent, rule-based approach requiring no expert judgment (4, 6, 7). The approach predicts skin sensitisation hazard by sequential testing, in an undefined order, in up to three of the following internationally accepted non-animal assays mapping to KE1-3 (*i.e.* DPRA, KeratinoSens™, h-CLAT). Assays are run for two KEs, and if these assays provide consistent results, then the chemical is predicted accordingly as sensitiser or non-sensitiser. If the first two assays provide discordant results, the assay for the remaining KE is run. The overall result is based on the two concordant findings taking into account the confidence on the obtained predictions as described in **Section 2.1.4**.

36. The performance of the 2o3 DA was found to be impacted by the consideration of borderline ranges for each of the methods, as described below in **Section 2.1.4**, and further

detailed in **Section 3.3** and **Annex 7** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1). A decision tree is provided in **Figure 2.1** of **Section 2.1.4** to derive predictions for the 2o3 DA, with no modification of the 2o3 DA Data Interpretation Procedure.

2.1.3. Description and limitations of the individual information sources

37. The individual information sources in the DA are assays included in OECD KE-based test guidelines for skin sensitisation (OECD TG 442C, 442D, 442E) (2, 3, 4), and the protocols are detailed therein.

38. The following assays from those TGs have been characterised and included in the 2o3 DA.

- Direct Peptide Reactivity Assay (DPRA; OECD TG 442C; KE1) (2): Skin sensitisers are generally electrophilic and react with the nucleophilic moieties of proteins. The DPRA measures depletion of two peptides containing either cysteine or lysine residues due to covalent binding. A test chemical that induces mean peptide depletion of cysteine- and lysine-containing peptide above 6.38% (or in the case of co-elution, cysteine-only depletion above 13.89%) is considered to be positive. In case borderline results are obtained for peptide depletion, additional testing should be conducted, as specified in OECD TG 442C and in **Annex 1**.
- KeratinoSens™ assay (*In vitro* Skin Sensitisation: ARE-Nrf2 Luciferase Test Method; OECD TG 442D; KE2) (3); Keratinocytes harbouring a reporter gene construct react to possible sensitisers via the Nrf2-Keap1 pathway. A test chemical that causes >1.5 fold luciferase induction, at viabilities > 70% when compared to the vehicle control, is considered to be positive. In case borderline results are obtained for luciferase induction, additional testing should be conducted, as specified in **Annex** 1.
- Human cell-line activation test (h-CLAT; OECD TG 442E; KE3) (4): Activation of antigen presenting cells is characterised by the up-regulation of CD86 and/or CD54. The h-CLAT is considered to be positive if CD86 induction exceeds 1.5-fold and/or CD54 exceeds 2-fold at viabilities > 50% when compared to the vehicle control. In case borderline results are obtained for CD54 and/or CD86 induction, additional testing should be conducted, as specified in **Annex 1**.

39. The current limitations of individual *in chemico* and *in vitro* test methods, such as limitations with respect to solubility, are described in the respective test guidelines (TG 442C, Appendix 1; TG 442D, Appendix 1A; TG 442E, Annex 1) and the validation studies cited therein (2, 3, 4).

2.1.4. Confidence in the 2o3 DA predictions

40. The first decision on whether each information element can be used is dictated by the limitations of the *in chemico* and *in vitro* methods (*e.g*. for substances that do not provide conclusive results in the individual methods due to solubility reasons) as found in in the respective test guidelines (TG 442C, Appendix 1; TG 442D, Appendix 1A; TG 442E, Annex 1) (2, 3, 4). Additionally, test results are subject to variation and these variations increase the uncertainty of a test result especially when close to a (classification) cut-off, *i.e.* in the borderline range. In order to define areas where lower confidence in the DA results may exist, borderline ranges (BRs) have been defined for output from the individual assays addressing the three KE of the 2o3 DA, (see **Annex 1** of this document, and **Section**

3.3 and **Annex 7** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation)* (1). The specific borderline ranges for each assay, as derived from their respective validation study data, are:

- DPRA BR: mean peptide depletion: 4.95% – 8.32%, Cys-only depletion (in the case of co-elution with lysine peptide): 10.56% – 18.47%;
- KeratinoSens™ BR: Imax: 1.35-fold – 1.67-fold;
- h-CLAT BR: RFI CD54: 157% – 255%; RFI CD86: 122% – 184%.

41. The incorporation of borderline ranges (BRs) into the prediction models (PM) for each of the individual information sources is are described in **Annex 1** of this guideline.

42. For the data with a single run as reported in the reference database, borderline cases in the DPRA are identified based on the borderline range for the mean peptide depletion or Cys-only depletion as described above. In case repeated runs are conducted, the PM in **Annex 1, Figure 1.1** shall be applied.

43. The prediction model of the KeratinoSens™ assay requires multiple runs. For the assessment of whether the outcome of repeated runs yields a positive, negative or borderline final outcome in KeratinoSens™, the PM in **Annex 1, Figure 1.2** shall be applied (adapted from the PM described in TG 442D to be used within the 2o3 DA to conclude on borderline cases). This prediction model introduces a third outcome (borderline) to be used within the 2o3 DA, based on the same decision cut-offs of the prediction model described in TG 442D. Thus, a negative in the original prediction model can only become negative or borderline, while a positive from the original prediction model can only become positive or borderline.

44. The prediction model of h-CLAT requires multiple runs. For the assessment of whether the outcome of repeated runs yields a positive, negative or borderline final outcome in the h-CLAT, the PM in **Annex 1**, **Figure 1.3** shall be applied (adapted from the PM described in TG 442E to be used within the 2o3 DA to conclude on borderline cases). This prediction model introduces a third outcome (borderline) to be used within the 2o3 DA, based on the same decision cut-offs of the prediction model described in TG 442E. Thus, a negative in the original prediction model can only become negative or borderline, while a positive from the original prediction model can only become positive or borderline.

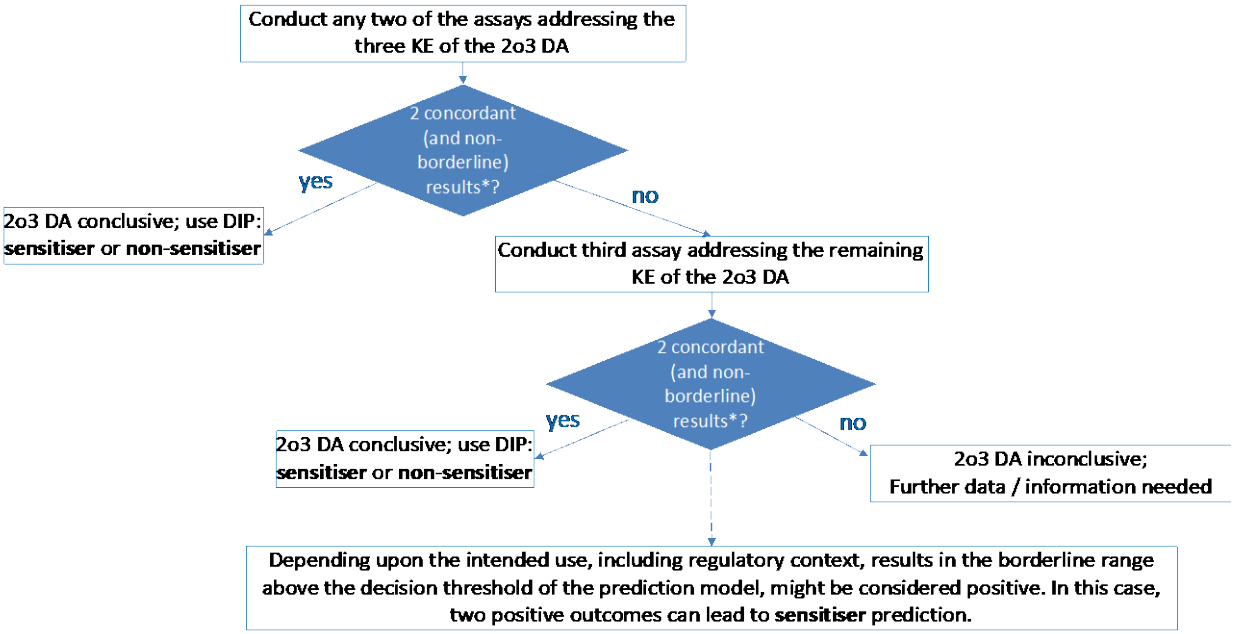
45. Positive and negative test results falling within these BRs as well as inconclusive results due to limitations in the *in chemico/in vitro* test guidelines are of lower confidence. For example, negative h-CLAT results obtained for a chemical with Log P > 3.5 (according to TG 442E (4)) are of lower confidence, and affect the outcome of the 2o3 DA as described below:

- In case the result of one of the 2o3 DA test methods falls into the respective test method's BR, a 2o3 DA prediction can still be made if the outcomes of the other two test methods composing the 2o3 DA are concordant and have high confidence (*i.e.*, results falling outside of the respective BRs).
- Similarly, in case a negative h-CLAT result is obtained for a chemical with Log P > 3.5, a 2o3 DA prediction can still be made if the outcomes of the other two test methods composing the 2o3 DA are concordant and have high confidence (*i.e.*, results falling outside of the respective BRs).

- However, if the result of one of the 2o3 DA test methods falls into the respective test method's BR or a negative h-CLAT result is obtained for a chemical with Log P > 3.5, and the other two methods composing the 2o3 do not provide concordant and high confidence results, the 2o3 DA prediction is considered 'inconclusive'. These inconclusive predictions may nevertheless be considered in a weight-of-evidence approach and/or within the context of an IATA together with other information sources. Depending on the intended use, including regulatory context, results in the borderline range above the decision threshold of the prediction model might still be considered positive; in this case, two positive outcomes can lead to an overall positive (sensitiser) prediction.

46. These borderline considerations and their impact on the confidence of the 2o3 DA predictions are visualized in **Figure 2.1.** DA predictions with high confidence for hazard identification are considered conclusive. DA predictions with low confidence are considered inconclusive for hazard identification. These 'inconclusive' predictions may nevertheless be considered in a weight-of-evidence approach and/or within the context of an IATA together with other information sources.

Figure 2.1. Decision tree to be used for the 2o3 DA, taking into account borderline results



Note: Borderline results are determined based on workflows given in **Annex 1**.
* The use of information elements is dictated by the limitations as found in in the respective test guidelines (TG 442C, Appendix 1; TG 442D, Appendix 1A; TG 442E, Annex 1). For example, in case a negative h-CLAT result is obtained for a chemical with Log P > 3.5 (according to the limitation described in TG 442E (4)), a 2o3 DA prediction can only be made if the outcomes of the other two test methods composing the 2o3 DA are concordant and are non-borderline.

2.1.5. Predictive capacity of the 2o3 DA vs. the LLNA

47. The predictive capacity of the "2o3" DA is reported based on data generated by the LLNA (see **Table 2.1**), curated as agreed upon by the EG DASS (see **Section 2.1** and **Annex 3** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation)*. The borderline range analyses were applied as described above to

assign confidence to the 2o3 DA predictions. Performance statistics are reported for conclusive (high confidence) predictions as compared to LLNA reference data, and inconclusive (low confidence) results are indicated. DA predictions for specific chemicals and further details are available in **Section 5** and **Annex 2** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1).

Table 2.1. Hazard identification performance of the "2o3" DA in comparison to LLNA reference data

	LLNA	
2o3 DA	Non	Sens
Non	22	19
Sens	4	89
Inconclusive	7	27

DA Performance vs. LLNA Data (N=134)	**2o3**
Accuracy (%)	83%
Sensitivity (%)	82%
Specificity (%)	85%
Balanced Accuracy (%)	84%

Note: Accuracy is the correct classification rate, sensitivity is the true positive rate, specificity is the true negative rate, and balanced accuracy is the average of sensitivity and specificity. Performance is reported based on DPRA, KeratinoSens™, and h-CLAT. Statistics reflect conclusive predictions only; inconclusive predictions are shown in grey. Additional performance characterisation is available in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

48. The application of the BR analyses and the designation of high/low confidence for the 2o3 DA predictions is applied as described above in **Section 2.1.4** and **Annex 1,** and further detailed in **Section 3.3** and **Annex 7** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1).

49. Due to the imbalanced nature of the reference data, the measure of specificity (based on 26 LLNA negative chemicals) is more uncertain than the measure of sensitivity (based on 108 LLNA positive chemicals).

2.1.6. Predictive capacity of the 2o3 DA vs. Human Data

50. The predictive capacity of the "2o3" DA is also reported based on Human Predictive Patch Test (HPPT) data (see **Table 2.2**), curated as agreed upon by the EG DASS (see **Section 2.2** and **Annex 4** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1)*)*. The borderline range analyses were applied as described above to assign confidence to the 2o3 DA predictions. Performance statistics are reported for conclusive (high confidence) predictions as compared to human reference data, and inconclusive (low confidence) results are indicated. DA predictions for specific chemicals and further details are available in **Section 5 and Annex 2** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1).

Table 2.2. Hazard identification performance of the "2o3" DA in comparison to human reference data

	Human	
2 of 3 DA	Non	Sens
Non	7	5
Sens	1	42
Inconclusive	3	7

DA Performance vs. Human Data (N=55)	**2o3**
Accuracy (%)	89%
Sensitivity (%)	89%
Specificity (%)	88%
Balanced Accuracy (%)	88%

Note: Accuracy is the correct classification rate, sensitivity is the true positive rate, specificity is the true negative rate, and balanced accuracy is the average of sensitivity and specificity with respect to HPPT data. Performance is reported based on DPRA, KeratinoSens™, and h-CLAT. Statistics reflect conclusive predictions only; inconclusive predictions are shown in grey. Additional performance characterisation is available in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1).

51. The application of the BR analyses and the designation of high/low confidence for the 2o3 DA predictions is applied as described above in **Section 2.1.4** and **Annex 1**, and further detailed in **Section 3.3 and Annex 7** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1).

52. Due to the imbalanced nature of the reference data, the measure of specificity (based on 8 human negative chemicals) is more uncertain than the measure of sensitivity (based on 47 human positive chemicals).

2.1.7. Predictive capacity of the LLNA vs. Human Data

53. To provide a basis for comparison for the DA performance statistics given above, the predictive capacity of the LLNA is reported based on data from the Human Predictive Patch Test (see **Table 2.3**) curated as agreed upon by the EG DASS. Data for specific chemicals and further details are available in **Section 5** and **Annex 2** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1).

Table 2.3. Hazard identification performance of the LLNA in comparison to Human reference data

	Human	
LLNA	Non	Sens
Non	2	3
Sens	7	44

LLNA Performance vs. Human Data (N=56)	**LLNA**
Accuracy (%)	82%
Sensitivity (%)	94%
Specificity (%)	22%
Balanced Accuracy (%)	58%

Note: Accuracy is the correct classification rate, sensitivity is the true positive rate, specificity is the true negative rate, and balanced accuracy is the average of sensitivity and specificity with respect to Human HPPT-based data. Additional performance characterisation is available in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1).

54. The hazard identification performance of the conclusive 2o3 DA predictions vs. human HPPT data was 89% accuracy, 89% sensitivity, 88% specificity, and 88% balanced accuracy, comparable to and/or exceeding the performance of the LLNA vs human HPPT data in every measure.

55. As previously noted, due to the imbalanced nature of the reference data, the measures of specificity are more uncertain than the measures of sensitivity.

2.1.8. Proficiency chemicals

56. The 2o3 DA relies on a simple, rule-based data interpretation procedure and requires no expert judgment. Proficiency chemicals for the individual information sources (KE1-3) are defined in the respective guidelines (2, 3, 4). Proficiency for the individual information sources demonstrates proficiency for the DA.

2.1.9. Reporting of the DA

57. The reporting of the DA application should follow the template described in OECD GD 255 (8), and should include at a minimum the following elements:

- Test chemical identification (*e.g.* chemical name, structural formula, composition, isomers, impurities including their quantities as available, CAS number, batch and lot number, and other relevant identifiers)
- Individual test reports performed per corresponding guideline (OECD TG 442C, 442D, 442E). Note that the chemical identity for each test report should match that above.
- Application of the individual prediction models adapted to be used within the 2o3 DA to determine borderline outcomes, as described in **Annex 1**
- Outcome of the DA application (hazard identification, *i.e.* skin sensitiser or not skin sensitiser or inconclusive result)
- Any deviation from or adaptation of the 2o3 DA

- Conclusion

2.2. References

1. OECD (2021). Series on Testing and Assessment No. 336: Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation. Organisation for Economic Cooperation and Development, Paris. Available at: [https://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm].
2. OECD (2020). OECD Guideline for the Testing of Chemicals No. 442C: *In chemico* Skin Sensitisation: Assays addressing the Adverse Outcome Pathway key even on covalent binding to proteins).*In chemico*. Paris, France: Organisation for Economic Cooperation and Development. Available at: (oecd-ilibrary.org).
3. 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: (oecd-ilibrary.org).
4. 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: (oecd-ilibrary.org).
5. 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.. Organisation for Economic Cooperation and Development, Paris. Available at: [https://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm].
6. Bauch C, Kolle SN, Ramirez T, Eltze T, Fabian E, Mehling A, Teubner W, van Ravenzwaay B, Landsiedel R. (2012). Putting the parts together: combining *in vitro* methods to test for skin sensitizing potential. Regul Toxicol Pharmacol, 63:489-504.
7. Urbisch D, Mehling A, Guth K, Ramirez T, Honarvar N, Kolle S, Landsiedel R, Jaworska J, Kern PS, Gerberick F, Natsch A, Emter R, Ashikaga T, Miyazawa M, Sakaguchi H. (2015). Assessing skin sensitization hazard in mice and men using non-animal test methods, Regul Toxicol Pharmacol, 71:337-51.
8. OECD (2016). Series on Testing & Assessment No. 255: Guidance Document On The Reporting Of Defined Approaches To Be Used Within Integrated Approaches To Testing And Assessment. Organisation for Economic Cooperation and Development, Paris. Available at: [https://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm].

Part II. –SECTION 3 - Defined Approaches for Skin Sensitisation Potency Categorisation

58. Part II of the Guideline includes Defined Approaches that allow the allocation of skin sensitizers into UN GHS sub-category 1A, strong sensitizers, or sub-category 1B for other (moderate to weak) skin sensitizers, following the Globally Harmonised System for Classification and Labeling (GHS). These DAs may also be used for hazard identification, *i.e.* to distinguish between sensitisers (UN GHS Category 1) and non-sensitisers (no classification; NC). Currently the ITSv1 DA and ITSv2 DA are included in this section of the Guideline. Additional detailed information can be found in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1).

3.1. "Integrated Testing Strategy (ITS)" Defined Approach

3.1.1. Summary

59. This defined approach was constructed as an Integrated Testing Strategy (ITS) for prediction of the skin sensitisation hazard potential and potency sub-categorisation according to the UN GHS (sub-categories 1A and 1B) of a chemicals.

60. The ITS DA uses test methods that address key events (KEs) 1 and 3 in the Adverse Outcome Pathway (AOP) and includes an *in silico* prediction of skin sensitisation. Protein binding (KE1) is quantitatively evaluated using the Direct Peptide Reactivity Assay (DPRA; OECD TG 442C) (2). Dendritic cell activation (KE3) is quantitatively evaluated using the human cell line activation test (h-CLAT; OECD TG 442E) (3). The *in silico* prediction of skin sensitisation is provided by either Derek Nexus (ITSv1) or OECD QSAR Toolbox (ITSv2).

61. The ITSv1 DA was evaluated for hazard identification with 167 chemicals and for UN GHS sub-categorisation with 155 chemicals based on LLNA reference data curated as agreed upon by the EG DASS, and achieved accuracies equivalent to the LLNA (see **Tables 3.2-3.3**). The performance of the ITSv1 DA was compared to 64 chemicals with human reference data curated as agreed upon by the EG DASS (see **Tables 3.4-3.5**), and exceeded the accuracy of the LLNA in predicting the same human data for both hazard and potency categorisation.

62. The ITSv2 DA was evaluated for hazard identification for 167 chemicals and for UN GHS sub-categorisation for 153 chemicals based on LLNA reference data curated as agreed upon by the EG DASS, and achieved accuracies equivalent to the LLNA (see **Tables 3.6-3.7**). The performance of the ITSv2 DA was compared to 64 chemicals with human reference data curated as agreed upon by the EG DASS (see **Tables 3.8-3.9**), and exceeded the accuracy of the LLNA in predicting the same human data for both hazard and potency categorisation.

3.1.2. Data interpretation procedure

63. The ITS DIP uses scores assigned to the quantitative results from the h-CLAT (3) and the DPRA (1), and from either Derek Nexus v6.1.0 (2020, Lhasa Limited, https://www.lhasalimited.org/products/derek-nexus.htm) or OECD QSAR TB v4.5 (https://www.oecd.org/chemicalsafety/oecd-qsar-toolbox.htm) to discriminate chemicals

into UN GHS category 1A (strong sensitiser); category 1B (other sensitiser), or Not Classified (non-sensitiser) (**Table 3.1**).

64. The DIP was amended from the original published version of the ITS (4) to change the cut-off for 1A sensitisers from a score of 7 to a score of 6 to optimize the ability of the DA to detect strong sensitisers and to extend the applicability of the ITS to chemicals for which *in silico* predictions cannot be generated. The DIP was also altered from the published version in that it was originally applied to ECETOC categories[2], and is here applied to the UN GHS subcategories.

65. The quantitative results of h-CLAT and DPRA are converted into a score from 0 to 3, as shown in **Table 3.1**. For h-CLAT, the minimum induction threshold (MIT) is converted to a score from 0 to 3 based on the cutoffs of 10 and 150 μg/ml. For DPRA, the mean percent depletion for the cysteine and lysine peptides is converted to a score from 0 to 3, based on the threshold values associated with reactivity classes described in OECD TG 442C (2). In cases where co-elution occurs only with the lysine peptide, the depletion for only cysteine peptides is converted to a score from 0 to 3. For the *in silico* prediction (Derek or OECD QSAR TB), a positive outcome is assigned a score of 1; a negative outcome is assigned a score of 0 (further details on the respective protocols are available in **Annex 2**). When these scores have been assessed, a total battery score ranging from 0 to 7, calculated by summing the individual scores, is used to predict the sensitising potential (hazard identification; UN GHS Cat. 1 vs. UN GHS NC) and potency (UN GHS Cat. 1A, Cat. 1B and NC). The positive criteria for identifying skin sensitisers (UN GHS Cat. 1) are set as a total battery score of 2 or greater. Based on the updated DIP, a total battery score is assigned into three ranks: score of 6-7 is defined as a strong (UN GHS Cat. 1A) sensitiser; score of 2-5 as moderate/weak (UN GHS Cat. 1B) sensitiser; score of 1 or 0, as not classified (*i.e.* a non-sensitiser).

[2] ECETOC Technical Report 087 (2003), Contact Sensitisation: Classification According to Potency. Available at: [https://www.ecetoc.org/publication/tr-087-contact-sensitisation-classification-according-to-potency/]

Table 3.1. Schematic of the ITS defined approach. The DA is a simple score-based system depending on assays from OECD TG 442E and 442C, and an *in silico* structure-based prediction, as shown.

Score	h-CLAT MIT µg/mL	DPRA mean Cysteine and Lysine% depletion	DPRA Cysteine % depletion*	*In silico* (ITSv1: DEREK; ITSv2: OECD TB)
3	≤10	≥42.47	≥98.24	
2	>10, ≤150	≥22.62, <42.47	≥23.09, <98.24	
1	>150, ≤5000	≥6.38, <22.62	≥13.89, <23.09	Positive
0	not calculated	<6.38	<13.89	Negative
	Potency	Total Battery Score		
	UN GHS 1A	6-7		
	UN GHS 1B	2-5		
	Not classified	0-1		

Source: Adapted from Takenouchi (5)

Note: UN GHS 1A correspond to strong sensitisers and UN GHS 1B correspond to other (moderate to weak) sensitisers. Not classified are considered non-sensitisers. *Cysteine-only depletion thresholds are used in the case of co-elution with the lysine peptide.

3.1.3. Description and limitations of the individual information sources

66. The individual *in chemico* and *in vitro* information sources are existing KE-based OECD test guidelines (OECD TG 442C, 442E) (2, 3), and the protocols are detailed therein.

67. The following assays from those TGs have been characterised and included in the ITS DA:

- Human cell-line activation test (h-CLAT; OECD TG 442E; KE3) (3): Activation of antigen presenting cells is characterised by the up-regulation of CD86 and/or CD54. The h-CLAT is considered to be positive if CD86 induction exceeds 1.5-fold and/or CD54 exceeds 2-fold at viabilities > 50% when compared to the vehicle control. From the experimental concentration-response curves, the median concentration(s) inducing 1.5- and/or 2-fold induction of CD86 and/or CD54 are calculated and the lowest of the two values is defined as the minimal induction threshold, MIT:

  $$MIT = \min(EC150\ CD86, EC200\ CD54)$$

  Test chemicals are assigned potency scores based on the MIT thresholds shown in **Table 3.1**.

- Direct Peptide Reactivity Assay (DPRA; OECD TG 442C; KE1) (2): Skin sensitisers are generally electrophilic and react with the nucleophilic moieties of proteins. The DPRA measures depletion of two peptides containing either cysteine or lysine residues due to covalent binding. A test chemical that induces mean peptide depletion of cysteine- and lysine-containing peptide above 6.38% (or in the case of co-elution, cysteine-only depletion above 13.89%) is considered to be positive. In case borderline results are obtained for peptide depletion, additional testing should be conducted, as specified in OECD TG 442C. Test chemicals are assigned potency scores based on the mean peptide depletion thresholds shown in **Table 3.1**.

68. The limitations of the individual *in chemico* and *in vitro* test methods are described in the respective test guidelines and in the respective test guidelines (TG 442C, Appendix 1; TG 442E, Annex 1) (2, 3).

69. The *in silico* information source predictions for ITSv1 are derived from Derek, an expert, knowledge-based software tool comprising alerts on several toxicity endpoints, including skin sensitisation. Derek (Derek Nexus v.6.1.0, 2020, Lhasa Limited) fires alerts based on structural features *i.e.* whether a hapten has potential for electrophilic binding to skin proteins either directly or following metabolism/auto-oxidation. To each alert, a likelihood level is associated. Chemicals firing an alert with a likelihood of certain, probable, plausible, or equivocal are considered to be positive. Chemicals with a negative prediction of 'non-sensitiser with no misclassified or unclassified features' are considered to be negative (https://www.lhasalimited.org/products/skin-sensitisation-assessment-using-derek-nexus.htm#Negative%20Predictions). The approach for characterising the *in silico* applicability domain used in the ITSv1 and the protocol for generating Derek predictions are provided in **Annex 2** of this guideline.

70. The *in silico* information source predictions for ITSv2 are derived from the OECD QSAR TB automated workflow providing skin sensitiser hazard predictions (OECD QSAR TB v4.5). The target compound is profiled for protein binding alerts; auto-oxidation products and skin metabolites are generated and then profiled for protein binding alerts. In case a protein binding alert is identified in the parent or in its (a)biotic metabolites, the same alert is used to identify analogues with experimental skin sensitisation data. If no protein binding alert is identified, then structural profilers are used to identify analogue chemicals and the data gap is filled using read across or directly via profiler outcomes in case no suitable analogues are automatically identified. The approach for characterising the *in silico* applicability domain used in the ITSv2 and the protocol for generating OECD QSAR TB predictions are provided in **Annex 2** of this guideline.

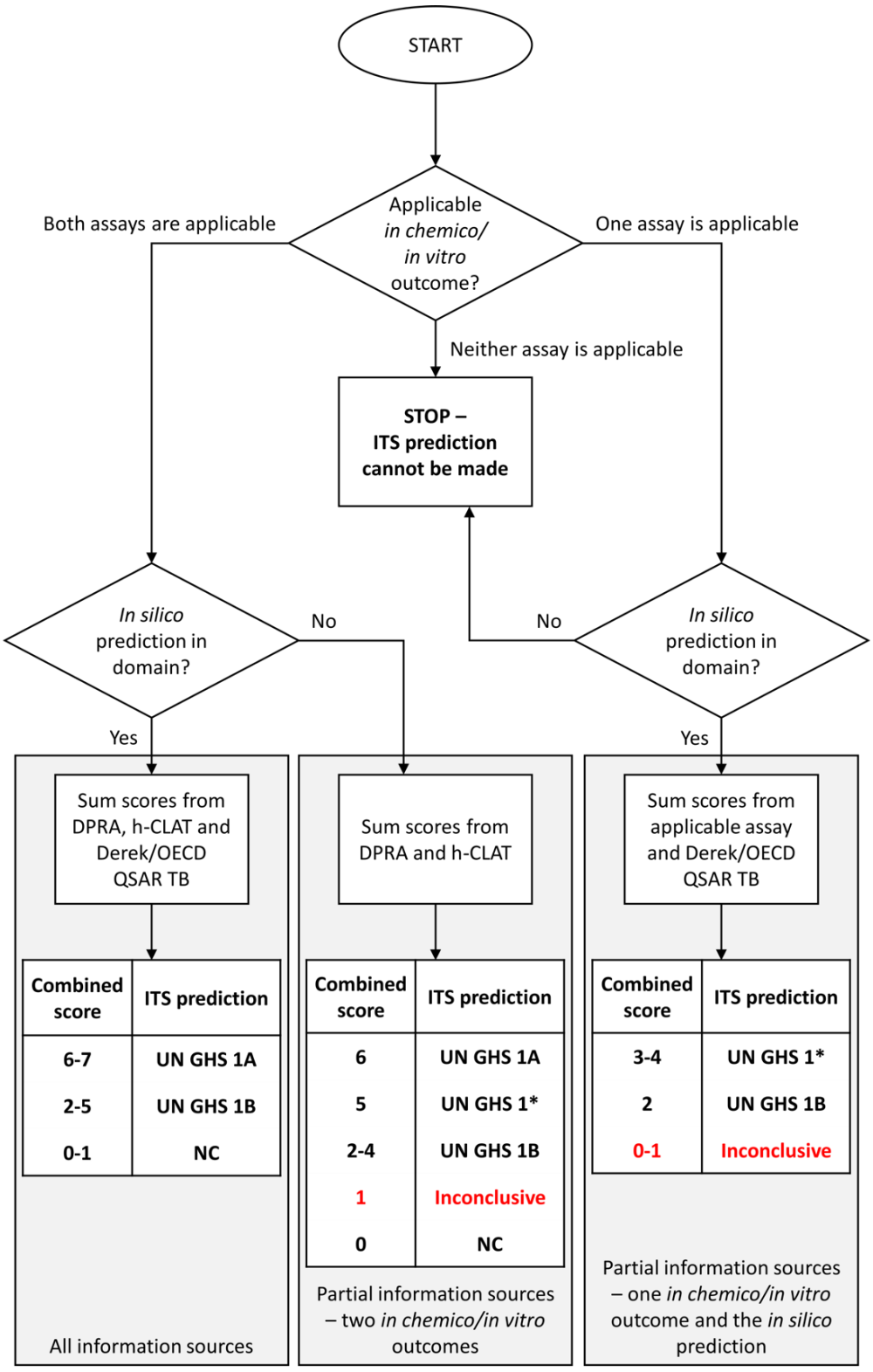
3.1.4. Confidence in the ITS DA predictions

71. The level of confidence of the ITS DA prediction is assigned based on the total DA score and applicability domain of the individual information sources, as shown via the flow chart in **Figure 3.1**. The first decision on whether all information elements can be used is dictated by the limitations of the *in chemico* and *in vitro* methods as found in TG 442C Appendix 1 and TG 442E Annex 1 (3) (*e.g.* for substances that do not provide conclusive results in the individual methods due to limited solubility or negative h-CLAT results for chemicals with Log P > 3.5 which are currently considered unreliable), and by the applicability domain of the *in silico* prediction (**Annex 2**). Partial information sources (*i.e.* two *in chemico/in vitro* outcomes only, or one *in chemico/in vitro* outcome and an *in silico*

prediction) may be used to obtain a DA prediction as shown via the flow chart in **Figure 3.1**.

72. DA predictions with high confidence for hazard identification and potency are considered conclusive. DA predictions with low confidence are considered inconclusive for hazard identification and/or potency. These 'inconclusive' predictions may nevertheless be considered in a weight-of-evidence approach and/or within the context of an IATA together with other information sources. Details including applicability domain and confidence considerations are provided in **Annex 2**.

Figure 3.1. Decision tree for assigning confidence to the ITS DA predictions



***Conclusive for hazard, inconclusive for potency**

3.1.5. Predictive capacity of the ITSv1 DA vs the LLNA

73. The predictive capacity of ITSv1 using Derek is reported based on data from the LLNA (see **Tables 3.2-3.3**), curated as agreed upon by the EG DASS (see **Section 1.1** and **Annex 3** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation)* (1). The workflow shown in **Figure 3.1** was applied to assign confidence to the ITSv1 DA predictions. The designation of conclusive/inconclusive for the ITSv1 DA predictions is further detailed in **Annex 2**. Performance statistics are reported for conclusive predictions as compared to LLNA reference data, and inconclusive results are indicated. DA predictions for specific chemicals and further details are available in **Section 5** and **Annex 2** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1).

Table3.2. Hazard identification performance of the ITSv1 DA in comparison to LLNA reference data

	LLNA	
ITSv1 DA	Non	Sens
Non	21	11
Sens	9	118
Inconclusive	3	6

DA Performance vs. LLNA Data (N=159)	**ITSv1**
Accuracy (%)	87%
Sensitivity (%)	92%
Specificity (%)	70%
Balanced Accuracy (%)	81%

Note: Accuracy is the correct classification rate, sensitivity is the true positive rate, specificity is the true negative rate, and balanced accuracy is the average of sensitivity and specificity with respect to LLNA data. Statistics reflect high confidence predictions only; inconclusive predictions are shown in grey. Additional performance characterisation is available in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1).

74. Due to the imbalanced nature of the reference data, the measure of specificity (based on 30 LLNA negative chemicals) is more uncertain than the measure of sensitivity (based on 129 LLNA positive chemicals).

Table 3.3. Potency categorisation performance of the ITSv1 DA in comparison to LLNA reference data, based on the UN GHS 1A/1B sub-categorisation

	LLNA		
ITSv1 DA	NC	1B	1A
NC	21	11	0
1B	9	55	10
1A	0	12	28
Inconclusive	3	7	0

71% correct classification overall

ITSv1 vs. LLNA reference data: Statistics based on the UN GHS 1A/1B sub-categorisation

Performance (N=146)	NC (N=30)	1B (N=78)	1A (N=38)
Correct classification (%)	70%	71%	74%
Underpredicted (%)	NA	14% (NC)	0% (NC); 26% (1B)
Overpredicted (%)	30% (1B); 0% (1A)	15% (1A)	NA

Note: Statistics reflect high confidence predictions only; inconclusive predictions are shown in grey. For more details on within-class performance (sensitivity, specificity, and balanced accuracy), please see **Section 5** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1).

75. The designation of high/low confidence for the ITSv1 DA predictions is applied as described above in **Figure 3.1** and further detailed in **Annex 2**.

3.1.6. Predictive capacity of the ITSv2 DA vs the LLNA

76. The predictive capacity of ITSv2 using OECD QSAR TB is reported based on data from the LLNA (see **Tables 3.4-3.5**), curated as agreed upon by the EG DASS (see **Section 2.1** and **Annex 3** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation)* (1). The workflow shown in **Figure 3.1** was applied to assign confidence to the ITSv2 DA predictions. The designation of high/low confidence for the ITSv2 DA predictions is further detailed in **Annex 2**. Performance statistics are reported for high confidence predictions as compared to LLNA reference data, and inconclusive results are indicated. DA predictions for specific chemicals and further details are available in **Section 5** and **Annex 2** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1). **Table 3.4**. Hazard identification performance of the ITSv2 DA in comparison to LLNA reference data.

Table 3.4. Hazard identification performance of the ITSv2 DA in comparison to LLNA reference data.

	LLNA	
ITSv2 DA	Non	Sens
Non	20	9
Sens	10	117
Inconclusive	3	9

DA Performance vs. LLNA Data (N=156)	ITSv2
Accuracy (%)	88%
Sensitivity (%)	93%
Specificity (%)	67%
Balanced Accuracy (%)	80%

Note: Accuracy is the correct classification rate, sensitivity is the true positive rate, specificity is the true negative rate, and balanced accuracy is the average of sensitivity and specificity with respect to LLNA data. Statistics reflect conclusive predictions only; inconclusive predictions are shown in grey. Additional performance characterisation is available in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1).

77. Due to the imbalanced nature of the reference data, the measure of specificity (based on 30 LLNA negative chemicals) is more uncertain than the measure of sensitivity (based on 126 LLNA positive chemicals).

Table 3.5. Potency categorisation performance of the ITSv2 DA in comparison to LLNA reference data, based on the UN GHS 1A/1B sub-categorisation

	LLNA		
ITSv2 DA	NC	1B	1A
NC	20	9	0
1B	10	54	10
1A	0	12	26
Inconclusive	3	10	2

71% correct classification overall

ITSv2 vs. LLNA reference data: Statistics based on the UN GHS 1A/1B sub-categorisation

Performance (N=141)	**NC (N=30)**	**1B (N=75)**	**1A (N=36)**
Correct classification (%)	67%	72%	72%
Underpredicted (%)	NA	12% (NC)	0% (NC); 28% (1B)
Overpredicted (%)	33% (1B); 0% (1A)	16% (1A)	NA

Note: Statistics reflect conclusive predictions only; inconclusive predictions are shown in grey. For more details on within-class performance (sensitivity, specificity, and balanced accuracy), please see **Section 5** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1).

78. The designation of conclusive/inconclusive for the ITSv2 DA predictions is applied as described above in **Figure 3.1** and further detailed in **Annex 2**.

3.1.7. Predictive capacity of the ITSv1 DA vs Human Data

79. The predictive capacity of ITSv1 using Derek is reported based on data from the Human Predictive Patch Test (see **Tables 3.6-3.7**), curated as agreed upon by the EG DASS. The designation of high/low confidence for the ITSv1 DA predictions is further detailed in **Annex 2**. Performance statistics are reported for high confidence predictions as compared to human reference data, and inconclusive results are indicated. DA predictions for specific chemicals and further details are available in **Section 5** and **Annex 2** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1).

Table 3.6 Hazard identification performance of the ITSv1 DA in comparison to Human reference data

	Human	
ITSv1 DA	Non	Sens
Non	4	4
Sens	5	51
Inconclusive	2	0

DA Performance vs. Human Data (N=64)	**ITSv1**
Accuracy (%)	86%
Sensitivity (%)	93%
Specificity (%)	44%
Balanced Accuracy (%)	69%

Note: Accuracy is the correct classification rate, sensitivity is the true positive rate, specificity is the true negative rate, and balanced accuracy is the average of sensitivity and specificity with respect to Human HPPT-based data. Statistics reflect conclusive predictions only; inconclusive predictions are shown in grey. Additional performance characterisation is available in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1).

80. Due to the imbalanced nature of the reference data, the measure of specificity (based on 9 Human negative chemicals) is more uncertain than the measure of sensitivity (based on 55 Human positive chemicals).

Table 3.7 Potency categorisation performance of the ITSv1 DA in comparison to Human reference data, based on the UN GHS 1A/1B sub-categorisation

	Human		
ITSv1 DA	NC	1B	1A
NC	4	4	0
1B	5	24	7
1A	0	3	13
Inconclusive	2	0	1

68% correct classification overall

ITSv1 vs. Human reference data: Statistics based on the UN GHS 1A/1B sub-categorisation

Performance (N=60)	**NC (N=9)**	**1B (N=31)**	**1A (N=20)**
Correct classification (%)	44%	77%	65%
Underpredicted (%)	NA	13% (NC)	0% (NC); 35% (1B)
Overpredicted (%)	56% (1B); 0% (1A)	10% (1A)	NA

Note: Statistics reflect conclusive predictions only; inconclusive predictions are shown in grey. For more details on within-class performance (sensitivity, specificity, and balanced accuracy), please see **Section 5** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1).

81. The designation of conclusive/inconclusive for the ITSv1 DA predictions is applied as described above in **Figure 3.1** and further detailed in **Annex 2**.

82. Due to the imbalanced nature of the reference data and the small numbers of chemicals, the measures of accuracy are more uncertain for smaller classes, *e.g.* for NC chemicals.

3.1.8. Predictive capacity of the ITSv2 DA vs Human Data

83. The predictive capacity of ITSv2 using OECD QSAR Toolbox is reported based on data from the Human Predictive Patch Test (see **Tables 3.8-3.9**), curated as agreed upon by the EG DASS. The designation of high/low confidence for the ITSv2 DA predictions is further detailed in **Annex 2**. Performance statistics are reported for conclusive predictions as compared to human reference data, and inconclusive results are indicated. DA predictions for specific chemicals and further details are available in **Section 5** and **Annex 2** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1).

Table 3.8 Hazard identification performance of the ITSv2 DA in comparison to Human reference data

	Human	
ITSv2 DA	Non	Sens
Non	4	3
Sens	5	50
Inconclusive	2	2

DA Performance vs. Human Data (N=62)	**ITSv2**
Accuracy (%)	87%
Sensitivity (%)	94%
Specificity (%)	44%
Balanced Accuracy (%)	69%

Note: Accuracy is the correct classification rate, sensitivity is the true positive rate, specificity is the true negative rate, and balanced accuracy is the average of sensitivity and specificity with respect to Human HPPT-based data. Statistics reflect conclusive predictions only; inconclusive predictions are shown in grey. Additional performance characterisation is available in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation*(1).

84. Due to the imbalanced nature of the reference data, the measure of specificity (based on 9 Human negative chemicals) is more uncertain than the measure of sensitivity (based on 53 Human positive chemicals).

Table 3.9. Potency categorisation performance of the ITSv2 DA in comparison to Human reference data, based on the UN GHS 1A/1B sub-categorisation

	Human		
ITSv2 DA	NC	1B	1A
NC	4	3	0
1B	5	24	6
1A	0	3	12
Inconclusive	2	1	3

70% correct classification overall

ITSv2 vs. Human reference data: Statistics based on the UN GHS 1A/1B sub-categorisation

Performance (N=57)	NC (N=9)	1B (N=30)	1A (N=18)
Correct classification (%)	44%	80%	67%
Underpredicted (%)	NA	10% (NC)	0% (NC); 33% (1B)
Overpredicted (%)	56% (1B); 0% (1A)	10% (1A)	NA

Note: Statistics reflect conclusive predictions only; inconclusive predictions are shown in grey. For more details on within-class performance (sensitivity, specificity, and balanced accuracy), please see **Section 5** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1).

85. The designation of conclusive/inconclusive for the ITSv2 DA predictions is applied as described above in **Figure 3.1** and further detailed in **Annex 2**.

86. Due to the imbalanced nature of the reference data and the small numbers of chemicals, the measures of accuracy are more uncertain for smaller classes, *e.g.* for NC chemicals.

3.1.9. Predictive capacity of the LLNA vs. Human Data

87. To provide a basis for comparison for the DA performance, the predictive capacity of the LLNA is reported based on data from the Human Predictive Patch Test (see **Tables 3.10-3.11**) curated as agreed upon by the EG DASS. Data for specific chemicals and further details are available in **Section 5** and **Annex 2** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1).

Table 3.10 Hazard identification performance of the LLNA in comparison to Human reference data

	Human	
LLNA	Non	Sens
Non	2	3
Sens	7	44

LLNA Performance vs. Human Data (N=56)	LLNA
Accuracy (%)	82%
Sensitivity (%)	94%
Specificity (%)	22%
Balanced Accuracy (%)	58%

Note: Accuracy is the correct classification rate, sensitivity is the true positive rate, specificity is the true negative rate, and balanced accuracy is the average of sensitivity and specificity with respect to Human HPPT-based data. Additional performance characterisation is available in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1).

88. The hazard identification performance of the conclusive ITSv1 DA predictions vs. human data was 86% accuracy, 93% sensitivity, 44% specificity, and 69% balanced accuracy, comparable to and/or exceeding the performance of the LLNA in every measure.

89. The hazard identification performance of the conclusive ITSv2 DA predictions vs. human data was 87% accuracy, 94% sensitivity, 44% specificity, and 69% balanced accuracy, comparable to and/or exceeding the performance of the LLNA in every measure.

90. As previously noted, due to the imbalanced nature of the reference data, the measures of specificity are more uncertain than the measures of sensitivity.

Table 3.11 Potency categorisation performance of the LLNA in comparison to Human reference data, based on the UN GHS 1A/1B sub-categorisation

Additional performance characterisation is available in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1).

	Human		
LLNA	NC	1B	1A
NC	2	3	0
1B	6	17	7
1A	0	3	9

60% correct classification overall

LLNA vs. Human reference data: Statistics based on the UN GHS 1A/1B sub-categorisation

Performance (N=47)	**NC (N=8)**	**1B (N=23)**	**1A (N=16)**
Correct classification (%)	25%	74%	56%
Underpredicted (%)	NA	13% (NC)	0% (NC); 44% (1B)
Overpredicted (%)	75% (1B); 0% (1A)	13% (1A)	NA

Note: For more details on within-class performance (sensitivity, specificity, and balanced accuracy), please see **Section 5** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation* (1).

91. The performance of the conclusive ITSv1 DA predictions vs. human data for potency sub-categorisation showed 68% correct classification overall, with accuracies of 44% for NC, 77% for 1B, and 65% for 1A, comparable to and/or exceeding the performance of the LLNA in every measure.

92. The performance of the conclusive ITSv2 DA predictions vs. human data for potency sub-categorisation showed 70% correct classification overall, with accuracies of 44% for NC, 80% for 1B, and 67% for 1A, comparable to and/or exceeding the performance of the LLNA in every measure.

93. As previously noted, due to the imbalanced nature of the reference data and the small numbers of chemicals, the measures of accuracy are more uncertain for smaller classes, *e.g.* for NC chemicals.

3.1.10. Proficiency chemicals

94. The ITS DA relies on a simple, rule-based data interpretation procedure and no expert judgment is required. Proficiency chemicals for the individual *in chemico* and *in vitro* information sources (KE1 and KE3) are defined in the respective guidelines (OECD TG 442C, 442E) (2, 3). The protocol details for the *in silico* information source options, Derek and OECD QSAR Toolbox, are included in **Annex 2** of this guideline. Proficiency has been demonstrated for Derek Nexus v6.1.0 and OECD QSAR Toolbox v4.5, and these

are the software versions that are intended for use in the ITSv1 and ITSv2 DAs, respectively. Proficiency for the individual information sources demonstrates proficiency for the DA.

3.1.11. Reporting of the DA

95. The reporting of the ITS DA should follow the template described in OECD GD 255 (6), and should include at a minimum the following elements:

- Test chemical identification (*e.g.* chemical name, structural formula, composition, isomers, impurities including their quantities as available, CAS number, batch and lot number, and other relevant identifiers)
- Individual test reports for the individual tests performed per corresponding guideline (OECD TG 442C, 442E). Note that the chemical identity for each test report should match that above.
- Description of protocol used for *in silico* prediction **(Annex 2)** and outcome, *e.g.* reported via a QPRF (7).
- Outcome of the DA application (hazard identification and potency categorisation according to UN GHS categories, or inconclusive result)
- Any deviation from the ITS DA
- Conclusion

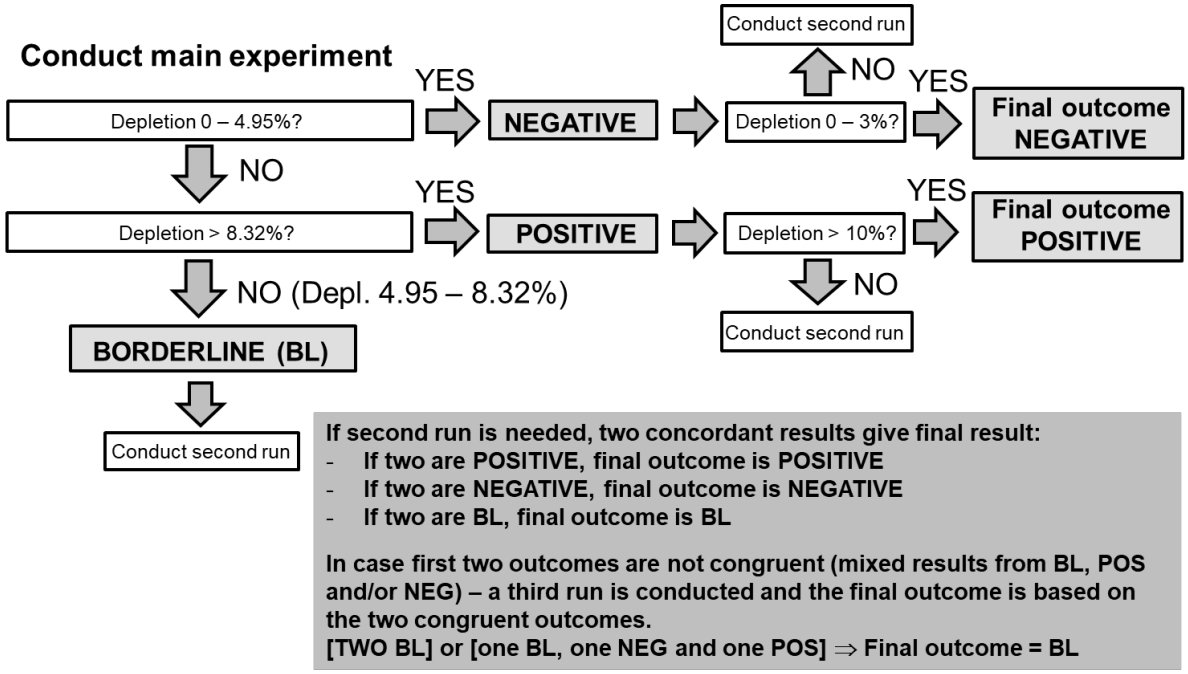
3.2. References

1. OECD (2021). Series on Testing and Assessment No. 336: Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation. Organisation for Economic Cooperation and Development, Paris. Available at: [https://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm].
2. OECD (2020). OECD Guideline for the Testing of Chemicals No. 442C: *In chemico* Skin Sensitisation: Assays addressing the Adverse Outcome Pathway key even on covalent binding to proteins). *In chemico*. Paris, France: Organisation for Economic Cooperation and Development. Available at:(oecd-ilibrary.org).
3. 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: (oecd-ilibrary.org).
4. 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. Organisation for Economic Cooperation and Development, Paris. Available at: [https://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm].
5. Takenouchi O, Fukui S, Okamoto K, Kurotani S, Imai N, Fujishiro M, Kyotani D, Kato Y, Kasahara T, Fujita M, Toyoda A, Sekiya D, Watanabe S, Seto H, Hirota M, Ashikaga T, Miyazawa M. (2015). Test battery with the human cell line activation test, direct peptide reactivity assay and DEREK based on a 139 chemical data set for predicting skin sensitizing potential and potency of chemicals. J Appl Toxicol, 35:1318-32.
6. OECD (2016). Series on Testing & Assessment No. 255: Guidance Document On The Reporting Of Defined Approaches To Be Used Within Integrated Approaches To Testing And Assessment. ENV/JM/HA(2016)28. Organisation for Economic Cooperation and Development, Paris. Available at: [https://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm].
7. ECHA (2008). see "CHAPTER R.6 – QSARS AND GROUPING OF CHEMICALS" *in* Guidance on Information Requirements and Chemical Safety Assessment. European Chemicals Agency [Guidance on Information Requirements and Chemical Safety Assessment - ECHA (europa.eu)]

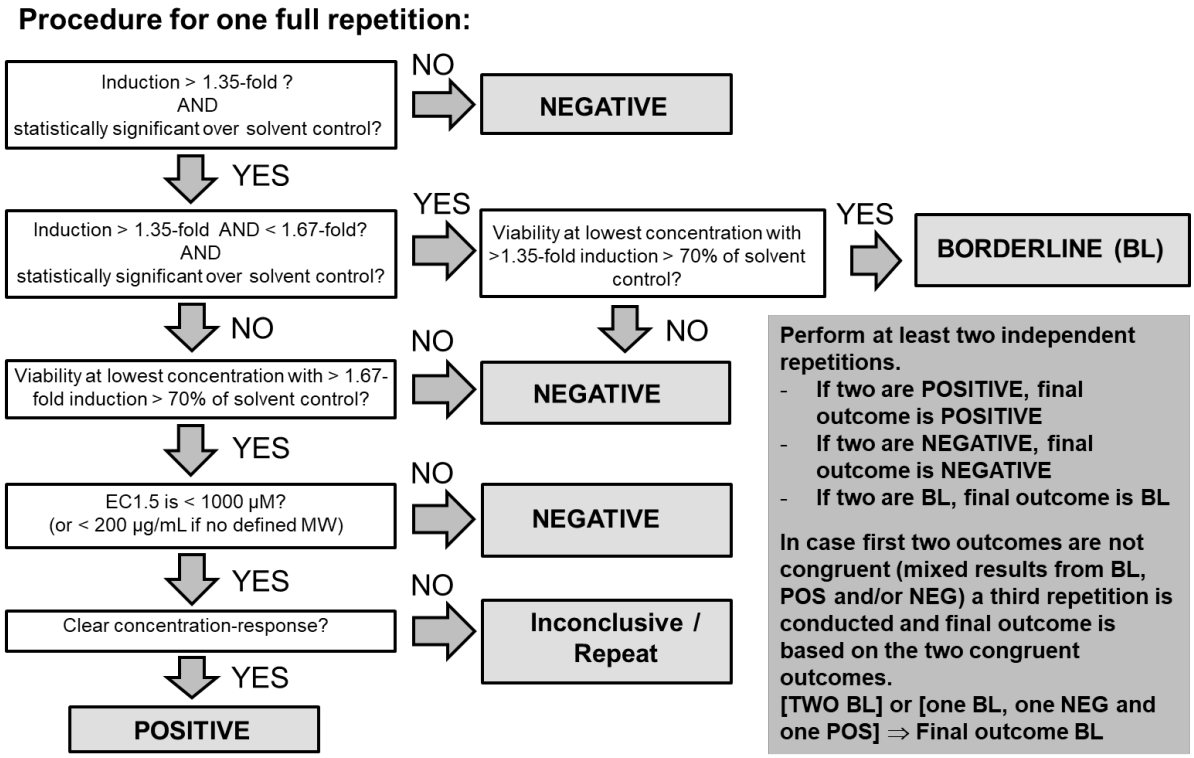
Annex 1: Prediction model for the individual *in chemico*/*in vitro* tests with multiple runs for use in 2o3 DA

96. The individual prediction models of h-CLAT and KeratinoSens™ require multiple runs (independent repetitions). An adaptation of the prediction model was used to determine borderline cases in the individual runs for the purpose of making predictions within the 2o3 DA. These adaptations (Figures 1.2. and 1.3) below should be used in these methods to come to the final conclusion of the individual tests.

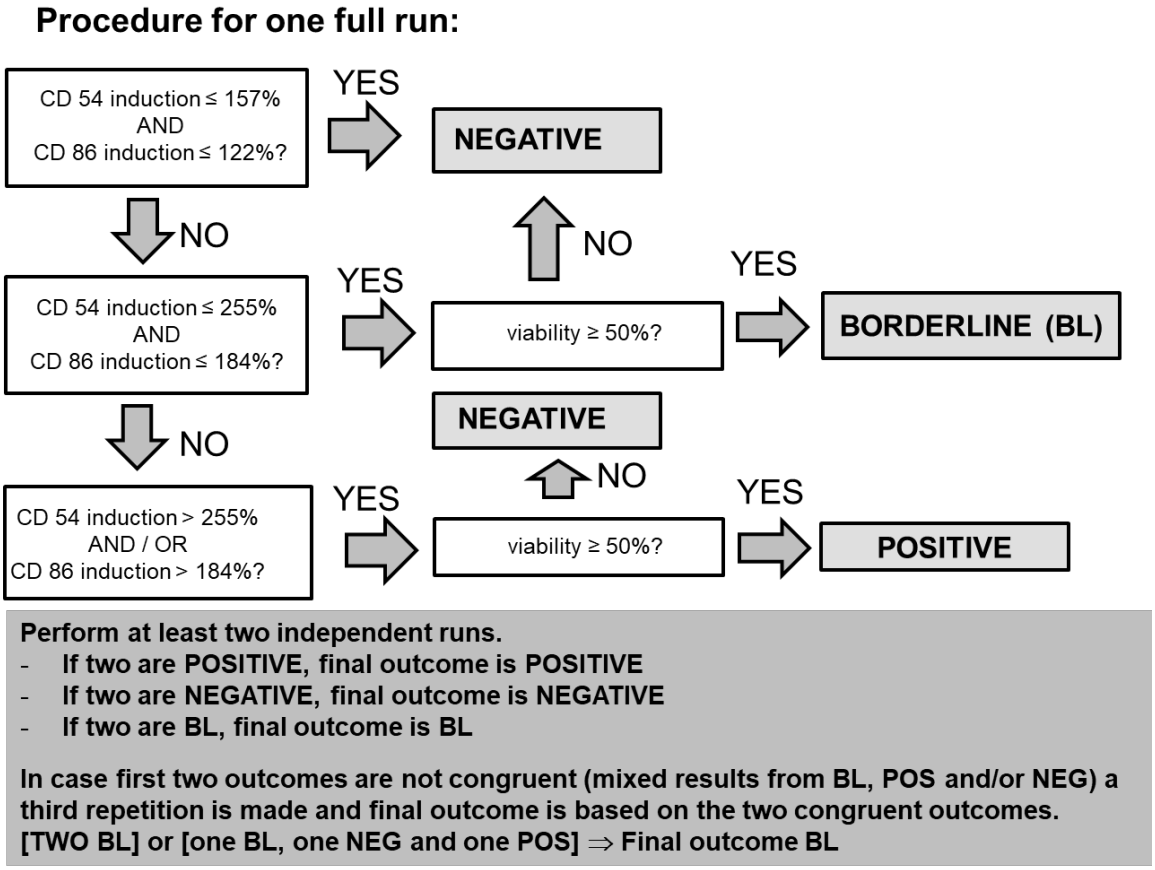
97. For the DPRA, repeated runs are required to be conducted if average depletion is within the range 3 - 10% (9 – 17% in case of Cysteine only depletion model is used). For this adaptation, the flowchart in Figure 1.1 is used to decide on run repetition and borderline assessment within the 2o3 DA.



Annex 1, Figure 1.1. Flow-chart of the DPRA prediction model (mean depletion) taking into borderline ranges and multiple runs conclude on borderline results within the 2o3 DA. The original threshold for a positive classification is 6.38%, and the statistically derived borderline range around this threshold is 4.95% - 8.32%. The same flowchart applies to the cysteine-only prediction model, whereby the following thresholds apply: 9% instead of 3%, >17 % instead of >10%, 10.56 % instead of 4.95% and > 18.47 % instead of >8.32%.



Annex 1, Figure 1.2. Flow-chart of the KeratinoSens™ prediction model taking into account borderline ranges and multiple runs to conclude on borderline results within the 2o3 DA. The original threshold for a positive classification is 1.5-fold induction, and the statistically derived borderline range around this threshold is 1.35 – 1.67-fold. Note: An independent run is referred to as 'repetition' in 442D, while it is called a 'run' in 442C and 442E; these nomenclatures do mean the same thing.



Annex 1, Figure 1.3. Flow-chart of the h-CLAT prediction model taking into account borderline ranges and multiple runs to conclude on borderline results within the 2o3 DA. The original threshold for a positive classification is 150% induction of CD86 with a statistically derived borderline range around this threshold of 122 – 184% and 200% induction of CD54 with a statistically derived borderline range around this threshold of 157 – 255%.

Annex 2: Defining the applicability domain and assessing confidence in DASS ITS predictions and protocols for generating *in silico* predictions

Introduction

98. As described in **Section 3.1** of the ***Guideline for Defined Approaches for Skin Sensitisation*** the ITS defined approaches (DAs) are based on three information sources: two *in chemico/in vitro* assays (DPRA; OECD TG 442C (OECD, 2015) and h-CLAT; OECD TG 442E (OECD, 2018)) and one *in silico* tool (prediction from either Derek Nexus (ITSv1) or OECD QSAR Toolbox (ITSv2) (referred to hereafter as *in silico*)). For each information source a score is given depending on the outcome of the individual assay and/or prediction, that is then summed to obtain the DA prediction.

Applicability domain of the individual information sources

In chemico/in vitro *information source (DPRA and h-CLAT)*

99. A test chemical is considered to be within the *in chemico*/*in vitro* domain (i.e. applicable) of DPRA and/or h-CLAT if it can be tested according to the individual protocols, taking into account the technical and chemical type limitations of each assay (as defined in the respective test guidelines OECD TG 442C and OECD TG 442E (OECD, 2015, 2018)). The *in chemico*/*in vitro* results are considered applicable, in case there are no technical or chemical space specific limitations and no reason why the results obtained from the assay cannot be considered.

In silico *information source*

100. The ITS DAs use *in silico* information sources that are based on chemical structures. These *in silico* sources rely on molecular representation of the chemicals: input usually by drawing the chemical structure, or by entering the Simplified Molecular-Input Line-Entry System (SMILES) or the IUPAC International Chemical Identifier (InChi). As a single chemical can be represented by several CAS or EC numbers (due to differences in composition e.g. stereochemical differences, present as varied salt forms, present as the main component in a mixture), it is important to specify the exact structure if possible. Resources such as the US EPA CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard) or NIH PubChem (https://pubchem.ncbi.nlm.nih.gov/) may be useful in mapping chemical names or structures to SMILES or InChi format. Available guidance can be consulted regarding minimum purity level of substances used in *in silico* predictions based on molecular structure.[34]

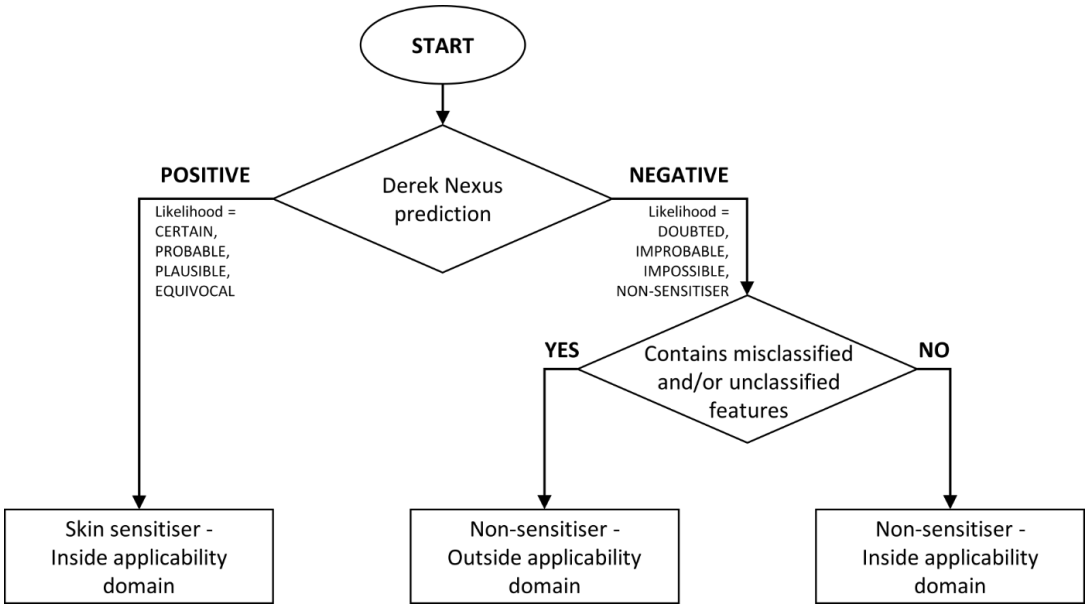
[3] OECD (2017), *Guidance on Grouping of Chemicals, Second Edition*, OECD Series on Testing and Assessment, No. 194, OECD Publishing, Paris, https://doi.org/10.1787/9789264274679-en.

[4] ECHA (2008) CHAPTER R.6 – QSARS AND GROUPING OF CHEMICALS *in* Guidance on Information Requirements and Chemical Safety Assessment. European Chemicals Agency [Guidance on Information Requirements and Chemical Safety Assessment - ECHA (europa.eu)

Derek Nexus (ITSv1)

101. Skin sensitisation predictions from Derek Nexus v6.1.0 are used in ITSv1. The protocol for running Derek Nexus (Derek) predictions is defined in **Appendix 1** of this document. All positive predictions (likelihood = certain, probable, plausible or equivocal) are considered to be inside the applicability domain. Negative predictions (likelihood = doubted, improbable, impossible or non-sensitiser) are also considered to be in the applicability domain unless they contain misclassified and/or unclassified features. A prediction of non-sensitiser with misclassified features indicates the presence of a fragment that has been observed exclusively in known sensitisers which Derek fails to alert for. A prediction of non-sensitiser with unclassified features indicates the presence of a fragment that has not been observed in publicly available data (although Derek may have seen this in proprietary data) (Chilton et al., 2018). Usually expert review is recommended for predictions containing these features but as a fixed data interpretation procedure, required in a DA, does not permit expert review these are best considered as out of domain for use in ITSv1 (**Figure A2.1**).

Figure A2.0.1. Applicability domain for Derek Nexus skin sensitisation predictions used in ITSv1.



QSAR Toolbox (ITSv2)

102. Skin sensitisation predictions from the QSAR Toolbox automated workflow "Skin sensitisation for defined approaches" (Yordanova et al., 2019) are used in ITS v2. The protocol for running QSAR Toolbox predictions is defined in **Appendix 2** of this document.

103. The calculation of the applicability domain of the predictions is automatically provided by Toolbox when running DASS AW predictions and consists of three layers: structural, parametric and mechanistic. The applicability domain layers considered for each individual prediction depend on the type and outcome of the prediction, as summarised in Table A2.1. A detailed description of the three layers and the rationale for their selection is

explained in **Appendix 3** of this document. Toolbox results within applicability domain are considered as applicable in the DA.

Table A2.1. Applicability domain layers for the QSAR Toolbox automated workflow "Skin sensitisation for defined approaches" predictions.

Toolbox DASS AW outcome		Applicability domain layer		
		Structural	Parametric	Mechanistic
Positive	Read-across	Not considered	Not considered	**Considered**
	Profiling	Not considered	Not considered	**Met by definition**
Negative	Read-across	Not considered	Not considered	**Considered**
	Profiling	**Considered**	**Considered**	**Met by definition**

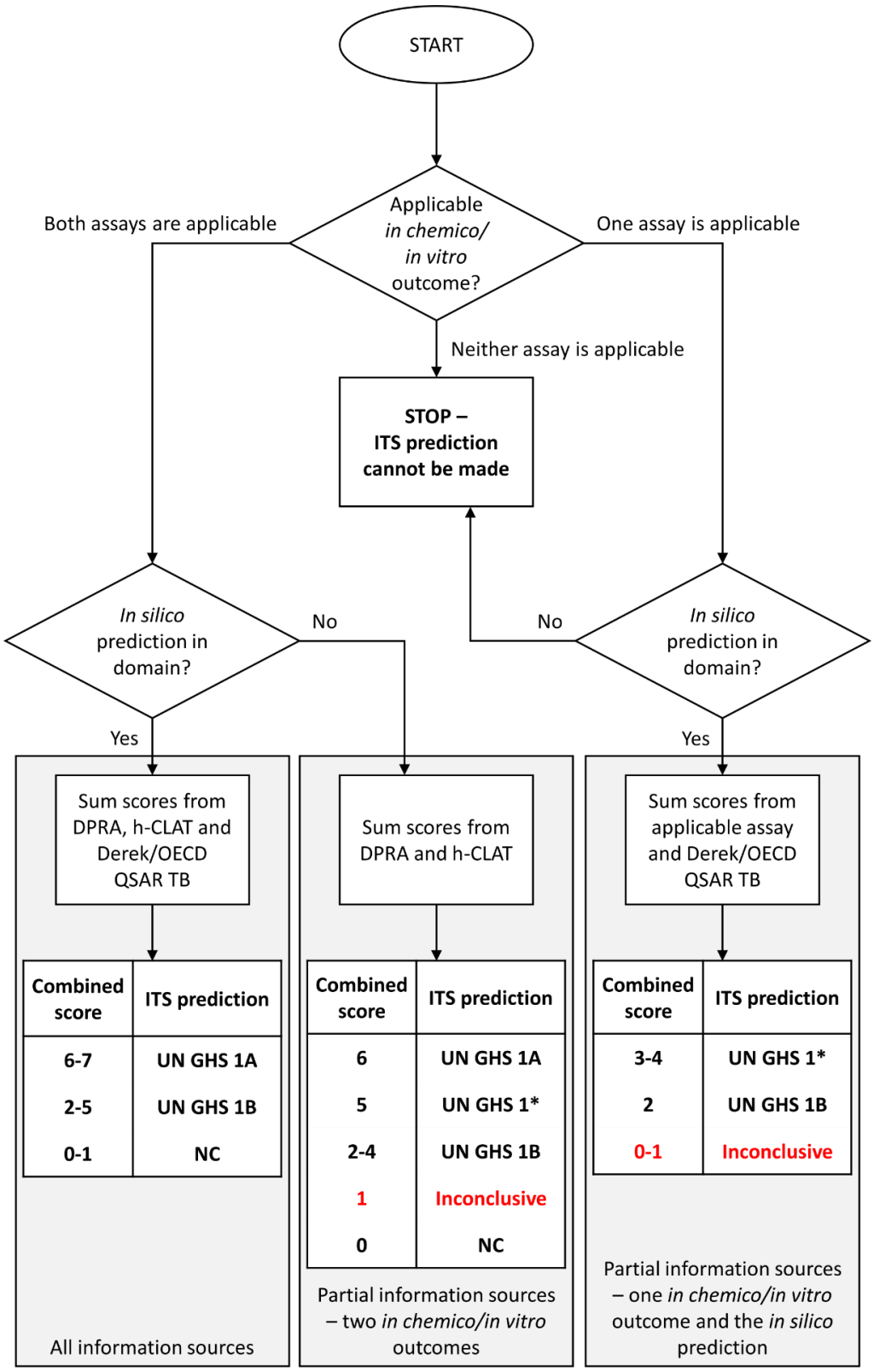
Confidence in ITS predictions

104. The applicability domain of the individual information sources used in the ITS DA are assessed and this determines whether the ITS predictions can be considered conclusive (i.e. high confidence) or inconclusive (i.e. low confidence) for hazard identification and/or potency.

How to apply the data interpretation procedure (DIP) for the ITS

105. The ITS was originally developed to use three information sources (DPRA, h-CLAT, and an *in silico* tool (Derek Nexus or OECD QSAR Toolbox)). Where all three information sources are applicable, a conclusive ITS prediction can be made. In some cases, a conclusive ITS prediction can be made, if there are two information sources with applicable results (**Figure A2.2**).

Figure A2.0.2. Workflow for data interpretation procedure for the ITS.



***Conclusive for hazard, inconclusive for potency**

106. Depending on the applicability of the individual information sources, three different scenarios for the ITS DA are possible (see Figure A2.2 and Table A2.2). In Scenario 1, all three information sources are applicable. In Scenarios 2 and 3, only two information sources are applicable. Details are provided below:

107. Scenario 1: all of the information sources i.e. *in chemico/in vitro* outcomes are applicable and can be considered (as prescribed in each individual assay) and the *in silico* prediction is in domain. The obtained ITS DA prediction is conclusive and of high confidence

108. Scenario 2: *in silico* prediction out of domain, however *in chemico*/*in vitro* methods are in domain and provide conclusive predictions (i.e. *in chemico*/*in vitro* methods are applicable).

- Combined DA score of 0, 2, 3, 4 or 6, *in silico* prediction out of *in silico* domain: DA conclusion is possible based on the two *in chemico/in vitro* outcomes. Conclusive prediction as the *in silico* prediction would not lead to a different DA prediction.
- Combined DA score of 5, *in silico* prediction out of *in silico* domain: DA conclusion possible for hazard identification (conclusive positive DA prediction for hazard identification). DA conclusion not possible for potency (inconclusive DA prediction for potency).
- Combined DA score of 1, *in silico* prediction out of *in silico* domain: DA conclusion not possible. Inconclusive DA prediction for hazard identification and potency.

109. Scenario 3: one *in chemico/in vitro* method out of domain or the result of that method cannot be considered (inapplicable):

- Combined DA score of 2 based on one *in chemico/in vitro* and *in silico* prediction: DA conclusion possible. Conclusive DA prediction as UN GHS 1B, as the outcome of the other *in chemico/in vitro* method would not to a different DA prediction.
- Combined DA score of 3 or 4, based on one *in chemico/in vitro* and *in silico* prediction: DA conclusion possible for hazard identification (conclusive positive DA prediction for hazard identification). DA conclusion not possible for potency (inconclusive DA prediction for potency).
- Combined DA score of 0 or 1, one *in chemico/in vitro* and *in silico* prediction: DA conclusion not possible. Inconclusive prediction for hazard identification and potency.

Table A2.2. Applicability domain and confidence of the ITS.

Scenario	Combined score[5]	ITS prediction	Confidence	DA prediction including confidence considerations
1	**0-1**	NC	High	Conclusive prediction Not Classified (NC).
	2-5	UN GHS 1B	High	Conclusive prediction UN GHS 1B.
	6-7	UN GHS 1A	High	Conclusive prediction UN GHS 1A.
2	**0**	NC	High	Conclusive prediction NC.
	1	Inconclusive	Low	Inconclusive prediction whether positive or negative.
	2-4	UN GHS 1B	High	Conclusive prediction UN GHS 1B.
	5	UN GHS 1	High	Conclusive positive prediction for hazard identification.
			Low	Inconclusive prediction for potency.
	6	UN GHS 1A	High	Conclusive prediction UN GHS 1A.
3	**0-1**	Inconclusive	Low	Inconclusive prediction whether positive or negative.
	2	UN GHS 1B	High	Conclusive prediction UN GHS 1B.
	3-4	UN GHS 1	High	Conclusive positive prediction for hazard identification.
			Low	Inconclusive prediction for potency.

[5]Total scores calculated only from information sources that are applicable/in domain.

References

Chilton, M. L., Macmillan, D. S., Steger-Hartmann, T., Hillegass, J., Bellion, P., Vuorinen, A., Etter, S., Smith, B. P. C., White, A., Sterchele, P., De Smedt, A., Glogovac, M., Glowienke, S., O'Brien, D., & Parakhia, R. (2018). Making reliable negative predictions of human skin sensitisation using an in silico fragmentation approach. *Regulatory Toxicology and Pharmacology*, *95*, 227–235. https://doi.org/10.1016/j.yrtph.2018.03.015

OECD. (2015). *Test No. 442C: In Chemico Skin Sensitisation: Direct Peptide Reactivity Assay (DPRA).* OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264229709-en

OECD. (2018). *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*. OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris.

Yordanova, D., Schultz, T. W., Kuseva, C., Tankova, K., Ivanova, H., Dermen, I., Pavlov, T., Temelkov, S., Chapkanov, A., Georgiev, M., Gissi, A., Sobanski, T., & Mekenyan, O. G. (2019). Automated and standardized workflows in the OECD QSAR Toolbox. *Computational Toxicology*, *10*, 89–104. https://doi.org/10.1016/j.comtox.2019.01.006

Appendix 1: Protocol for Derek Nexus predictions

110. The following protocol may be used to generate predictions for skin sensitisation hazard using Derek Nexus v.6.1.0 with Derek Knowledge Base (KB) 2020 1.0 to be used as the *in silico* information source for the ITSv1 defined approach.

Protocol for generating predictions for skin sensitisation hazard using Derek Nexus v.6.1.0 with Derek KB 2020 1.0

Single chemical

1. Open Nexus
2. Input structure using one of the following options:
    a. Input structure manually by drawing on the canvas
    b. Go to File>Open Structure(s) to input a single structure from a file (.mol, .sdf, .smi, .csv, .cdx (file list not exhaustive))
    c. Go to File>Type Chemistry to enter or paste SMILES, InChi or MOL file
    d. Go to File>New Structure to input structure by drawing a structure
3. Set up prediction
    a. Go to Prediction>Derek Prediction>Derek Prediction Setup
4. Apply processing constraints
    a. Knowledge Bases
        i. For Nexus v6.1.0, ensure Derek KB 2020 1.0 is selected
        ii. For newer releases, use the default Derek KB supplied
    b. Perception
        i. Ensure 'Perceive tautomers' and Perceive mixtures' are selected
        ii. Ensure 'Match alerts without rules' is unselected
    c. Species
        i. Select 'mammal'
    d. Endpoints
        i. Click 'Deselect all' then expand 'Skin sensitisation (ALL)' to view 'Photoallergenicity' and 'Skin sensitisation'. Select 'Skin sensitisation'
    e. Structure properties
        i. Ensure the 'Overwrite' box(es) for logP, logKp, and average molecular mass are unselected to use the values calculated by Derek Nexus, otherwise, check the 'Overwrite' box(es) to input own values.
5. Generate prediction
    a. Click 'Start Prediction'
    b. If an alert is fired: Knowledge base, endpoint, species, reasoning level, alert fired, EC3 prediction (if applicable), and example matched (if applicable) are shown in the prediction navigator.
        i. Click the likelihood (certain, probable, plausible, equivocal) to view the reasoning rules leading to the likelihood level.

ii. Click the Alert in the prediction navigator to view alert match(es), description image, comments, validation comments, endpoint, references, patterns, and examples associated with the alert.

c. If no alert is fired, a negative prediction is generated: Knowledge base, endpoint, species and negative prediction reasoning (non-sensitiser) and negative prediction overview (absence or presence of misclassified and/or unclassified features) are shown in the prediction navigator.
    i. Click the negative prediction overview ('No misclassified or unclassified features', 'Contains misclassified/unclassified features') to view information about the negative prediction. Similar nearest neighbours are available to view for misclassified features.

d. Use the Derek likelihood to classify each compound as positive or negative (alert fired with certain, probable, plausible, or equivocal is classified as positive, alert fired with doubted, improbable, impossible, or a negative prediction of non-sensitiser with no misclassified or unclassified features is classified as negative).
    i. Negative predictions of non-sensitiser with misclassified and/or unclassified features are of lower confidence and are not used in ITSv1.
    ii. In cases where more than one alert is fired or structures in a mixture generate different likelihoods, the most conservative classification is applied (positive > negative).
    iii. A positive outcome from Derek is scored as 1 in the ITSv1 and a negative outcome is scored as 0.

Multiple chemicals

1. Open Nexus
2. Input structures
    a. Go to File>Open Structure(s) to input a file containing multiple structures (.mol, .sdf, .smi, .csv, .cdx (file list not exhaustive))
    b. Select the fields from the file which will be mapped to structure properties used during the prediction (Name, Average Molecular Mass, LogP, LogKp). If left unchanged then the values set by Derek will be used.
3. Set up batch prediction
    a. Go to Prediction>Derek Prediction>Derek Batch Setup
4. Apply processing constraints
    a. Knowledge Bases
        i. For Nexus v6.1.0, ensure Derek KB 2020 1.0 is selected
        ii. For newer releases, use the default Derek KB supplied
    b. Perception
        i. Ensure 'Perceive tautomers' and Perceive mixtures' are selected
        ii. Ensure 'Match alerts without rules' is unselected
    c. Species
        i. Select 'mammal'
    d. Endpoints

i. Click ‘Deselect all’ then expand ‘Skin sensitisation (ALL)’ to view ‘Photoallergenicity’ and ‘Skin sensitisation’. Select ‘Skin sensitisation’

e. Report configuration
   i. Directory - Leave as default directory or map to preferred location.
   ii. Pick type - Select report for batch (left side icon)
   iii. Pick format - Select desired file type (e.g. Excel)
   iv. Pick design - Select desired design (e.g. Tabular Report)
   v. Filename - input desired filename

f. Report display options
   i. Ensure ‘Show predictions of at least impossible’ is selected
   ii. Select ‘Show Negative Predictions’
   iii. Select ‘Filter All Nearest Neighbours by Misclassified Features’
   iv. Select ‘Show Open Likelihood’
   v. Select ‘Show Rapid Prototypes’

5. Generate batch prediction
   a. Click ‘Start Batch Prediction’
      i. Once the batch prediction is finished, select the ‘Open Report Directory’ when prompted
   b. Use the Derek likelihood to classify each compound as positive or negative (alert fired with certain, probable, plausible, or equivocal is classified as positive, alert fired with doubted, improbable, impossible, or a negative prediction of non-sensitiser with no misclassified or unclassified features is classified as negative).
      i. Negative predictions of non-sensitiser with misclassified and/or unclassified features are of lower confidence and are not used in ITSv1.
      ii. In cases where more than one alert is fired or structures in a mixture generate different likelihoods, the most conservative classification is applied (positive > negative).
   c. A positive outcome from Derek is scored as 1 in the ITSv1 and a negative outcome is scored as 0.

Appendix 2: Protocol for OECD QSAR Toolbox predictions

111. The following protocol may be used to generate predictions for skin sensitisation hazard using OECD QSAR Toolbox v.4.5 with the automated workflow for defined approaches for skin sensitisation (DASS AW) to be used as the in silico information source for the ITSv2 defined approach.

Protocol for generating predictions for skin sensitisation hazard using DASS AW in Toolbox 4.5.

Step 1: Input the chemical in the "Input module". SMILES is the preferred way to input the structure. (If other identifiers such as the CAS number are used as input, the Toolbox will assign the SMILES based on its internal database. In this case, the user needs to make sure that Toolbox identifies and consequently uses for the prediction the correct structure.)

Step 2: Go to the "Data gap filling module" and click on "Automated" button. Select "EC3 from LLNA or Skin sensitization from GPMT assays for defined approaches" and click OK. The scheme with the implemented logic will be shown.

Step 3: Click the Run button -  or press F5 key of the keyboard and confirm with "Yes". The workflow will run automatically.

Step 4: If a substance is predicted "positive" or "negative" as a result of read-across, the prediction will appear on the data matrix with "R" in front of the result (e.g. "R: Negative). If a substance is predicted "positive" or "negative" as a result of profiling, then the result will appear next to the name of the customized profiler "Skin sensitization for DASS".

Step 5: Affiliation of the substance to the domain of the automated workflow for DASS will be automatically determined and presented.

Appendix 3: Information on applicability domain for OECD QSAR Toolbox

Technical aspects

112. The Toolbox prediction used by DA ITS v.2 is calculated using the DASS automated workflow (DASS AW) included in OECD QSAR Toolbox v.4.5. The workflow also includes the automatic calculation of the applicability domain of Derek skin described below.

Calculation of the in silico domain of Toolbox

113. Applicability domain of the QSAR Toolbox Skin sensitisation predictions for use in the ITS defined approach approaches automated workflow (DASS AW) is defined by based on the training set substances of the same automated workflow. The training set (TS) consists of 2268 substances having LLNA and/or GPMT skin sensitisation experimental data[6](the full list of substances can be consulted in the QSAR Toolbox). The TS substances are part of the following OECD QSAR Toolbox databases:

- Skin sensitisation;
- REACH Skin sensitisation (normalized) databases.

114. Based on the correctly predicted training set substances, three layers of applicability domain are automatically calculated by the Toolbox: 1) parametric; 2) structural and 3) mechanistic layers. Depending on the Toolbox prediction approach (read-across or profiling predictions) and prediction outcomes (positive or negative), one or more of these layers are taken into account to establish the overall Toolbox domain of the specific prediction.

115. The applicability domain layers considered for different types of Toolbox predictions are summarised in the table here:

Toolbox DASS AW outcome		Applicability domain layer		
		Structural	Parametric	Mechanistic
Positive	Read-across	Not considered	Not considered	**Considered**
	Profiling	Not considered	Not considered	**Met by definition**
Negative	Read-across	Not considered	Not considered	**Considered**
	Profiling	**Considered**	**Considered**	**Met by definition**

116. Explanation and rationale for the use of different domain layers:

1. Positive predictions (both by read-across and profiling): the presence of an alert (which is the requirement for positive Toolbox prediction to be considered within in the mechanistic domain) is sufficient to consider the prediction to be within the Toolbox domain. Substances triggering an alert are considered as in domain because they contain the toxicophore that has been observed experimentally in skin sensitisers. No further checks are needed in this context to consider the prediction within the Toolbox *in silico* domain.

[6] In case of multiple data points for one substance, the most conservative scenario is taken into account.

2. Negative predictions by read-across: the structural and parametric domains are not taken into account because the Toolbox has already ensured some level of similarity with other substances in its training set that met the requirements to be selected as suitable analogues for read-across (these requirements are explained in detail in the DASS AW description).
3. Negative prediction by profiling predictions: all domain layers are taken into account to ensure the highest possible reliability level for the Toolbox prediction. Stricter requirements are needed mainly for two reasons: 1. lack of alerts is not equal to proof of lack of sensitisation potential and 2. to apply a cautious approach since acceptance of negative predictions may lower the human health protection level risk in case of a false negative predictions.

Calculation of applicability domain layers

1. Parametric layer

Four physico-chemical parameters of the substances are taken into consideration: log Kow, molecular weight, vapour pressure and water solubility[7]. The ranges of variation for the selected parameters are defined based on the training set substances that are correctly predicted by the DASS AW.

A substance is considered within the parametric domain of the DASS AW if its physico-chemical parameter values as calculated by the QSAR Toolbox fall into the ranges of variation given in the table below. It is noted that the ranges include parametric values calculated using EPISuite models implemented in Toolbox that in some cases are wider than that covered by existing test methods.

Physico-chemical parameter	Calculated Parameter range
Log Kow	-9.66 ÷ 18.6
Molecular weight	16 Da ÷ 2290 Da
Vapour pressure*	0 Pa ÷ 3.45 x 10^{7} Pa
Water solubility	2.48 x 10^{-15} mg/L ÷ 1.00 x 10^{6} mg/L

*EPIWIN Vapor Pressure (Antoine method) is used for calculation

2. Structural layer

The structural layer is defined based on the atom centred fragments (ACF) derived from the structural characteristics of the TS substances that are correctly predicted[8] by the DASS AW.

The ACF are defined according to the following Toolbox default values for ACF:

- Any atom distance = 1

[7] QSAR Toolbox is used for the calculation of the physico-chemical properties.

[8] All ACF that are extracted from the correctly predicted TS test chemicals “good space”. The “bad space” is formed from the ACF present in the incorrectly predicted test chemicals. The default QSAR Toolbox settings for ACF are used. Supplementary file with the ACF forming the good and the bad space are available.

- Heteroatom distance = 1
- Extract C (sp^3) fragments = YES
- Include whole aromatic rings = NO

For each substance, the following values are calculated:

- % Correct fragments: percentage of ACF occurring in correctly predicted structures in the training set
- % incorrect fragments: percentage of ACF occurring in incorrectly predicted structures in the training set
- % unknown fragments: percentage of ACF not occurring in the training set.

A substance is considered within the structural domain of the DASS AW if 100% of its ACF belong to the correct fragments.

3. Mechanistic layer

The predicted capability of a substance to interact with the skin proteins without and after (a)biotic activation is taken into consideration. The Toolbox endpoint-specific profiler *Protein binding for skin sensitization by OASIS* and two metabolic simulators – *Autoxidation simulator* and *Skin metabolism simulator* are used to predict such interaction.

A positive prediction is considered within the mechanistic domain if the substance triggers "*Protein binding for skin sensitization by OASIS*" alerts without or after (a)biotic activation.

A negative prediction is considered within the mechanistic domain if the substance does not permit expert review these are best considered as out of domain for use in the ITS "*trigger Protein binding for skin sensitization by OASIS*" without or after (a)biotic activation.

117. Note that predictions obtained by profiling results will meet the mechanistic layer requirements by definition because positive Toolbox predictions by profiler are triggered exactly by the presence of alert. If the test chemical cannot be tested or the outcome/prediction cannot be considered in at least two of the information sources (*in chemico/in vitro and/or in silico)* then the DA cannot be applied.