新規評価法提案書

ディファインド アプローチによる皮膚感作性評価法

令和 7 年 1 月

国立医薬品食品衛生研究所

新規評価法提案書

令和 7 年 1 月 16 日

No.　2024-01

ディファインド アプローチによる皮膚感作性評価法に関する提案

令和 6 年 11 月 26 日に国立医薬品食品衛生研究所にて開催された新規試験法評価会議（通称：JaCVAM 評価会議）において以下の提案がなされた。

提案内容：本評価法は、AOP に基づく感作性評価という観点から重要な情報を与えてくれる。2o3 DA および ITS DA では、NC と予測された UN GHS 区分 1A 物質はないため、これらの評価法によるハザード予測性は妥当と考える。ITS DA の強度予測性については、限られた参照物質数ではあるが LLNA に劣らない予測性が得られているため、概ね妥当と考える。ただし UN GHS 区分 1B を NC に、UN GHS 区分 1A を 1B にそれぞれ過小評価する物質が一定数存在することに留意する必要がある。

この提案書は、皮膚感作性試験資料編纂委員会によりまとめられた文書を用いて、JaCVAM 評価会議が評価および検討した結果、その有用性が確認されたことから作成された。

以上の理由により、行政当局の安全性評価方法としてディファインド アプローチによる皮膚感作性評価法の使用を提案するものである。

西川秋佳

JaCVAM 評価会議　議長

平林容子

JaCVAM 運営委員会　委員長

JaCVAM 評価会議

西 川 秋 佳 （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　病理部/
名古屋徳洲会総合病院）：座長
石 井 雄 二 （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 病理部）
小 島 幸 一 （一般財団法人 食品薬品安全センター）
中村るりこ （独立行政法人 製品評価技術基盤機構）
西 村 次 平 （独立行政法人 医薬品医療機器総合機構）
西 村 拓 也 （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 毒性部）
平 林 容 子 （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）
松 本 一 彦 （名古屋市立大学大学院）

任期：令和 6 年 4 月 1 日～令和 8 年 3 月 31 日

JaCVAM 運営委員会

平 林 容 子　（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）：委員長
石 井 孝 司　（国立感染症研究所）
田 中 里 依　（厚生労働省 医薬・生活衛生局 医薬品審査管理課 化学物質安全対策室）
豊 田 武 士　（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 病理部）
諫 田 泰 成　（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 薬理部）
北 嶋　聡　（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 毒性部）
杉 山 圭 一　（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 ゲノム安全科学部）
髙 橋 暁 子　（独立行政法人 医薬品医療機器総合機構）
髙 橋 祐 次　（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 毒性部 動物管理室）
束 野 正 明　（厚生労働省 医薬・生活衛生局 医薬品審査管理課 化学物質安全対策室）
西 村 次 平　（独立行政法人 医薬品医療機器総合機構）
本 間 正 充　（国立医薬品食品衛生研究所）
増 村 健 一　（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）
宮 坂 知 幸　（厚生労働省 医薬・生活衛生局 医薬品審査管理課）
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 ゲノム安全科学部 第四室）：事務局
大 野 彰 子　（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 ゲノム安全科学部 第四室）：事務局

JaCVAM statement on the Defined Approach for skin sensitization

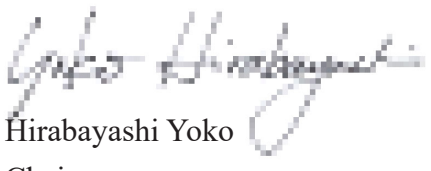
At a meeting held on 26 November, 2024 at National Institute of Health Sciences (NIHS) in Tokyo, Japan, the Japanese Center for the Validation of Alternative Methods (JaCVAM) Regulatory Acceptance Board unanimously endorsed the following statement:

Proposal: This evaluation method provides valuable information from the perspective of sensitization assessment based on the AOP. In both the 2o3 DA and ITS DA approaches, there are no UN GHS Category 1A substances predicted as NC, indicating the hazard predictability is considered appropriate. Regarding the potency predictability of ITS DA, although the number of reference substances is limited, the potency predictability is generally comparable to that of LLNA, making it is generally considered appropriate. However, it is important to note that a certain number of substances tend to be underestimated. Specifically, some substances classified as UN GHS Category 1B are predicted as NC, and some substances classified as UN GHS Category 1A are predicted as 1B.

This statement was released following a review prepared by the skin sensitization test JaCVAM Editorial Committee to acknowledge that the results of the review and study by the JaCVAM Regulatory Acceptance Board have confirmed the usefulness of this evaluation method.

Based on the above, we proposed the Defined Approach for skin sensitization as a useful means for assessing skin sensitization potential and potency during safety assessments by regulatory agencies.

Nishikawa Akiyoshi
Chairperson,
JaCVAM Regulatory Acceptance Board.



Hirabayashi Yoko
Chairperson,
JaCVAM Steering Committee.

January 16, 2025

The JaCVAM Regulatory Acceptance Board was established by the JaCVAM Steering Committee, and is composed of nominees from the industry and academia.

This statement was endorsed by the following members of the JaCVAM Regulatory Acceptance Board:

Nishikawa Akiyoshi (Division of Pathology, Center for Biological Safety and Research: CBSR, NIHS / Nagoya Tokushukai General Hospital) : Chairperson
Hirabayashi Yoko (CBSR, NIHS)
Ishii Yuji (Division of Pathology, CBSR, NIHS)
Kojima Koichi (Food and Drug Safety Center)
Matsumoto Kazuhiko (Nagoya City University)
Nakamura Ruriko (National Institute of Technology and Evaluation)
Nishimura Jihei (Pharmaceuticals and Medical Devices Agency)
Nishimura Takuya (Division of Cellular and Molecular Toxicology, CBSR, NIHS)

Term: From 1st April 2024 to 31st March 2026

This statement was endorsed by the following members of the JaCVAM steering Committee after receiving the report from JaCVAM Regulatory Acceptance Board:

Hirabayashi Yoko (CBSR, NIHS): Chairperson
Honma Masamitsu (NIHS)
Ishii Koji (National Institute of Infectious Diseases)
Kanda Yasunari (Division of Pharmacology, CBSR, NIHS)
Kitajima Satoshi (Division of Cellular and Molecular Toxicology, CBSR, NIHS)
Masumura Kenichi (Division of Risk Assessment, CBSR, NIHS)
Miyasaka Tomohiro (Ministry of Health, Labour and Welfare)
Nishimura Jihei (Pharmaceuticals and Medical Devices Agency)
Sugiyama Keiichi (Division of Genome Safety Science, CBSR, NIHS)
Takahashi Akiko (Pharmaceuticals and Medical Devices Agency)
Tanaka Rie (Ministry of Health, Labour and Welfare)
Taquahashi Yuhji (Animal Management Section of the Division of Cellular and Molecular Toxicology, CBSR, NIHS)
Toyoda Takeshi (Division of Pathology, CBSR, NIHS)
Tsukano Masaaki (Ministry of Health, Labour and Welfare)
Ashikaga Takao (Division of Genome Safety Science, CBSR, NIHS): Secretary
Ohno Akiko (Division of Genome Safety Science, CBSR, NIHS): Secretary

ディファインド アプローチによる皮膚感作性評価法に関する提案

添付資料一覧

添付資料 1

評価会議報告書

ディファインド アプローチによる皮膚感作性評価法

JaCVAM 評価会議

令和 7 年(2025 年)1 月 9 日

JaCVAM 評価会議

西 川 秋 佳 （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　病理部/
名古屋徳洲会総合病院）：座長
石 井 雄 二 （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 病理部）
小 島 幸 一 （一般財団法人 食品薬品安全センター）
中村るりこ （独立行政法人 製品評価技術基盤機構）
西 村 次 平 （独立行政法人 医薬品医療機器総合機構）
西 村 拓 也 （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 毒性部）
平 林 容 子 （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）
松 本 一 彦 （名古屋市立大学大学院）

任期：令和 6 年 4 月 1 日～令和 8 年 3 月 31 日

略語

AOP:	Adverse Outcome Pathway
DA:	Defined Approach
ITS:	Integrated Testing Strategy
KE:	Key Event
LLNA:	Local Lymph Node Assay
NC:	Not classified
OECD:	Organisation for Economic Co-operation and Development
QSAR:	Quantitative Structure-Activity Relationship
TG:	Test Guideline
UN GHS:	United Nations Globally Harmonized System of Classification and Labelling Chemicals
2o3:	2 out of 3

JaCVAM 評価会議は、経済協力開発機構 (OECD) ガイドライン 497[1]および皮膚感作性試験資料編纂委員会により作成された「評価報告書　ディファインド アプローチによる皮膚感作性評価法」[2]をもとに本評価法の科学的妥当性、社会的受け入れ性および行政上の利用性について検討した。

1. 評価法の概要および科学的妥当性

評価法の概要：

当該評価法は、単独の皮膚感作性試験代替法では最終的な評価を下すことが不十分であるとの考えから、複数の代替法試験結果を、ディファインド アプローチ (Defined Approach: DA、定義済み総合判定方式) に従って利用し、LLNA[3]などの動物試験やヒト試験[4]で得られる情報と同等の情報、すなわち有害性の有無や国連の化学品の分類および表示に関する世界調和システム (UN GHS) に利用できる情報を提供するガイドライン497として、OECDが公定化したものである。

ガイドラインに採用された方式のうち、2o3 DA は複数の *in chemico* 試験法および *in vitro* 試験法により化学物質の皮膚感作性 [GHS 区分 1 (感作性物質) または区分に該当しない (NC)]を同定 (ハザード評価) する方法であり、2 種類の ITS DA は複数の *in chemico* 試験法、*in vitro* 試験法、および *in silico* ツールによって化学物質の皮膚感作性 (GHS 区分 1 または NC) を同定 (ハザード評価) する方法、あるいは UN GHS 細区分 1A (強い感作性物質)／1B [その他の (中等度から弱い) 感作性物質] または NC に分類 (強度予測評価) する方法である。

科学的妥当性：

代替法試験のうち、*in chemico* および *in vitro* 試験法としては、皮膚感作性の有害性発現経路 (AOP)[5] に基づく 4 つの Key Event (KE) のうち KE1～KE3 に対応する OECD TG である DPRA (TG 442C)[6]、KeratinoSens™ (TG 442D)[7] および h-CLAT (TG 442E) [8]を用い、*in silico* ツールとしては、構造アラート (Structural alert) を用いて皮膚感作性の予測を提供する Derek Nexus またはリードアクロスによる類推を行う OECD QSAR Toolbox を用いることから、科学的に妥当な手法である。

2. 目的とする物質または製品の皮膚感作性を評価する方法としての社会的受け入れ性および行政上の利用性

社会的受け入れ性：

本評価法は、皮膚感作性に関する既存の *in chemico* および *in vitro* OECD TG や利用可能な *in silico* ツールを用いることで実施可能である。また、本評価法は動物を用いない手法であり、3Rs の精神と合致している。一方、複数の試験法または *in silico* ツールを必要とすることから、LLNA や他の *in vivo* 試験法と比較して、簡便性・経済性の面では必ずしも有用でない場合がある。

行政上の利用性※：

本評価法は、AOP に基づく感作性評価という観点から重要な情報を与えてくれる。2o3 DA および ITS DA では、NC と予測された UN GHS 区分 1A 物質はないため、これらの評価法によるハザード予測性は妥当と考える。

ITS DA の強度予測性については、限られた参照物質数ではあるが LLNA に劣らない予測性が得られているため、概ね妥当と考える。ただし UN GHS 区分 1B を NC に、UN GHS 区分 1A を 1B にそれぞれ過小評価する物質が一定数存在することに留意する必要がある。

※ 皮膚感作性試験資料編纂委員会による評価報告書では、ITS DA の強度予測性の妥当性について言及していないが、LLNA に劣らない予測性が得られていると明記されており、本会議ではガイドライン 497 の記載どおりに ITS DA の強度予測性は妥当と判断した。

参考文献（最終確認日：2024 年 11 月 26 日）

1) OECD (2023). OECD Guidelines for the Testing of Chemicals No.497. Defined Approaches on Skin Sensitisation, Organisation for Economic Cooperation and Development, Paris. Available at: https://www.oecd-ilibrary.org/environment/guideline-no-497-defined-approaches-on-skin-sensitisation_b92879a4-en
2) JaCVAM　皮膚感作性試験資料編纂委員会：　Defined Approach for Skin Sensitisation (DASS)評価報告書(2024 年 9 月 11 日)
3) OECD (2010). OECD Guidelines for the Testing of Chemicals No. 429. The Local Lymph Node Assay (LLNA), Organisation for Economic Cooperation and Development, Paris. Available at: https://www.oecd-ilibrary.org/environment/test-no-429-skinsensitisation_9789264071100-en
4) OECD (2023). Supporting document to the OECD guideline 497 on defined approaches for skin sensitisation Organisation for Economic Cooperation and Development, Paris. Available at: https://one.oecd.org/document/ENV/CBC/MONO(2021)11/En/pdf
5) 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/en/publications/the-adverse-outcome-pathway-for-skin-sensitisation-initiated-by-covalent-binding-to-proteins_9789264221444-en.html
6) OECD (2024). OECD Guideline for the Testing of Chemicals No. 442C. *In Chemico* Skin Sensitisation: Direct Peptide Reactivity Assay (DPRA), Organisation for　Economic Cooperation and Development, Paris. Available at: https://www.oecd-ilibrary.org/environment/test-no-442c-in-chemico-skin-sensitisation_9789264229709-en
7) OECD (2024). OECD Guideline for the Testing of Chemicals No. 442D. *In Vitro* Skin Sensitisation: ARE-Nrf2 Luciferase Test Method. Organisation for Economic Cooperation and Development, Paris. Available at: https://www.oecd-ilibrary.org/environment/test-no-442d-in-vitro-skin-sensitisation_9789264229822-en
8) OECD (2022). OECD Guideline for the Testing of Chemicals No. 442E. *In Vitro* Skin Sensitisation: human Cell Line Activation Test (h-CLAT), Organisation for　Economic Cooperation and Development, Paris. Available at: https://www.oecd-ilibrary.org/environment/test-no-442e-in-vitro-skin-sensitisation_9789264264359-en

添付資料 2

評価報告書

ディファインド アプローチによる皮膚感作性評価法

皮膚感作性試験資料編纂委員会

令和 7 年(2025 年)1 月 9 日

皮膚感作性試験資料編纂委員会

福 山 朋 季　（委員長：麻布大学）
安 達 玲 子　（国立医薬品食品衛生研究所）
大 竹 利 幸　（株式会社資生堂）
笠 原 利 彦　（富士フイルム株式会社）
河 上 強 志　（国立医薬品食品衛生研究所）
小 島 幸 一　（一般財団法人　食品薬品安全センター）
小 島　　肇　（山口東京理科大学/国立医薬品食品衛生研究所）
瀬 崎 拓 人　（三井化学株式会社）
武 吉 正 博　（一般財団法人　化学物質評価研究機構）

用語集

ADRA: Amino acid Derivative Reactivity Assay

AOP: Adverse Outcome Pathway

ARE: Antioxidant Responsive Element

ATP: Adenosine Triphosphate

BA: Balanced Accuracy

BR: Borderline Range

BrdU: Bromodeoxyuridine

CD: Cluster of Differentiation

DA: Defined Approach

DASS: Defined Approach for Skin Sensitisation

DIP: Data Interpretation Procedure

DPRA: Direct Peptide Reactivity Assay

ECHA: European Chemicals Agency

EU: European Union

GL: Guideline

h-CLAT: human Cell Line Activation Test

ITS: Integrated Testing Strategy

KE: Key Event

LLNA: Local Lymph Node Assay

MIT: Minimum Induction Threshold

Nrf2: nuclear factor-erythroid 2-related factor 2

NC: Not classified

OECD: Organisation for Economic Co-operation and Development

QSAR: Quantitative Structure-Activity Relationship

QMRF: QSAR Model Reporting Format

REACH: Registration, Evaluation, Authorisation and Restriction of Chemicals

RFI: Relative Fluorescence Intensity

RI: Radio Isotope

TG: Test Guideline

UN GHS: United Nations Globally Harmonized System of Classification and Labelling of Chemicals

2o3: 2 out of 3

要旨

化学物質の皮膚感作性を評価する代替法の開発が進み、動物を用いない多くの感作性試験代替法が経済協力開発機構 (Organisation for Economic Co-operation and Development: OECD) 試験法ガイドライン (Test Guideline: TG) として承認されている。しかし、各感作性試験代替法単独では、最終的な評価を下すことが不十分であるとの考えから、OECD は複数の代替法試験結果を、ディファインド アプローチ (Defined Approach: DA、定義済み総合判定方式）に従って利用し、動物試験で得られる情報と同等の情報、すなわち有害性の有無や国連の化学品の分類および表示に関する世界調和システム (United Nations Globally Harmonized System of Classification and Labelling of Chemicals: UN GHS) に利用できる情報を提供するガイドライン (Guideline: GL) 497 を公定化した。

本委員会では、OECD が定めた GL497 に含まれる 2 out of 3 ディファインド アプローチ (2o3 DA) による感作性のハザード評価および Integrated Testing Strategy ディファインド アプローチ (ITS DA) による感作性物質のハザード評価は妥当と考え、行政的な受け入れについても利用可能と判断した。

1. 序論

皮膚感作性を評価することは化学物質の安全性評価において重要である。化学物質の接触皮膚炎のリスクを動物で予測する OECD TG としてモルモットを用いる皮膚感作性試験 (OECD TG 406) [1)]やマウスを用いる局所リンパ節試験 (Local Lymph Node Assay: LLNA) がある。LLNA には放射性同位体 (Radio Isotope: RI) の取込量を測定する LLNA-RI 法 (OECD TG 429) [2)]のほか、RI を用いず ATP (Adenosine Triphosphate) 量を測定する LLNA: DA (OECD TG 442A) [3)]や Bromodeoxyuridine (BrdU) 量を測定する LLNA: BrdU-ELISA および LLNA: BrdU-FCM (OECD TG 442B) [4)]がある。

European Union (EU) における欧州化学品規則の一つである Registration、Evaluation、Authorisation and Restriction of Chemicals (REACH) では、安全性評価はコンピューターを用いた定量的構造活性相関 (Quantitative Structure-Activity Relationship: QSAR) モデルや *in vitro* 試験等による代替法が推奨されており、動物実験により安全性が評価された成分を含んだ化粧品の輸入および販売が禁止された (2013 年 3 月全面施行) 。そのため、化学物質の皮膚感作性を評価する代替法の開発が進み、動物を用いない多くの感作性試験代替法が OECD TG として承認されている (OECD TG 442C、TG 442D および TG 442E) [5-7)]。

皮膚感作性に伴う化学的および生物学的機序に関する最新の知見は、有害性発現経路 (Adverse Outcome Pathway: AOP) として提示されている [8)]。皮膚感作性は、その AOP に基づく 4 つの Key Event (KE) を経て成立し、下記のように、KE1～KE3 には動物を用いない代替法が開発され、OECD TG 化されている。

- KE1 化学物質とタンパク質の共有結合：Direct Peptide Reactivity Assay (DPRA)、Amino acid Derivative Reactivity Assay (ADRA) および kinetic DPRA (kDPRA) (OECD TG 442C) [5)]
- KE2 角化細胞活性化に関連する Antioxidant Responsive Element (ARE) nuclear factor-erythroid 2-related factor 2 (Nrf2) ルシフェラーゼの発現：KeratinoSens™および LuSens、抗酸化応答および炎症関連遺伝子の発現：EpiSensA (OECD TG 442D) [6)]
- KE3 特異的細胞表面マーカーの発現およびケモカインやサイトカインの産生を指標と

した樹状細胞の活性化、h-CLAT (human Cell Line Activation Test) 、U-SENS™ 、IL-8 Luc assay および GARD™-skin (OECD TG 442E) [7)]

- KE4 リンパ節における T 細胞の増殖、LLNA (OECD TG 429、TG 442A および TG 442B) [2-4)]

しかし、化学物質の皮膚感作性評価および UN GHS に従った細区分 1A (強感作性物質) 、1B (その他の感作性物質) の情報を提供するためには、動物を用いないいずれの代替法も単独試験での運用は不適当と考えられている。そのため、*in chemico* および *in vitro* の試験法は *in silico* および類似化合物からのリードアクロス等と同様に、複数の結果を併用することが提案された。

動物を用いない試験も複数の試験結果を決められた DA に従って利用することにより、ヒトにおける皮膚感作性予測を動物試験と同等以上に高めることができる。DA は、専門家の判断を必要とせずに予測を導き出すために、決められた情報セットから生成されたデータを既定のデータ解釈手順 (Data Interpretation Procedure: DIP) に適用する。DA は得られた結果の信頼性を高めるために、それぞれの試験系の限界を克服するような方法がとられている。DA の最終的な目標は、動物試験で得られる情報と同等の情報、すなわち有害性の有無や UN GHS 分類に利用できる情報を提供することである。

試験機関は、DA に従って試験を実施する前に、被験物質に関して入手可能な全ての情報について考慮すべきである。例えば被験物質の化学構造や物理化学的特徴などの情報は DA に基づくそれぞれの OECD TG に適用可能かどうかを決定するために利用できる。

OECD GL497 [9)]には 3 つの DA が含まれており、パート I では皮膚感作性の有無の判定のための DA が 1 つ、パート II では皮膚感作性の有無の判定および強度判定のための DA が 2 つ収載されている。その他の DA の追加については、将来の審査、承認後に GL497 に含められる。

2. ガイドラインに含まれる DA

ガイドラインに記載されている DA は次の通りである。

パート I

- 2 out of 3 (2o3) DA：*in chemico* (KE1)、*in vitro* (KE2 および KE3) のデータに基づき、皮膚感作性の有無を判定

  Integrated Testing Strategy DA (ITSv1 DA)：*in chemico* (KE1)、*in vitro* (KE3) および *in silico* (Derek Nexus v6.1.0) のデータに基づき、DIP により皮膚感作性の有無および強度を判定

パート II

- Integrated Testing Strategy DA (ITSv2 DA)：*in chemico* (KE1)、*in vitro* (KE3) および *in silico* (OECD QSAR Toolbox) のデータに基づき、DIP により皮膚感作性の有無および強度を判定

2o3 DA は、OECD TG である DPRA、KeratinoSens™、h-CLAT を用いて作成されている。ITS DA (ITSv1 および ITSv2) では、*in silico* の情報も利用する。ITSv1 DA で使用する Derek Nexus v6.1.0 (以下、Derek Nexus と記す) は、化学構造に含まれる警告構造を用いて、皮膚感作性の有無を予測する専門家の知識に基づくツールである。また、ITSv2 DA で使用する OECD QSAR Toolbox v4.5 は類似化学物質に基づくリードアクロスまたは「Profiler (プロファ

イラ) 」により同定されたタンパク質結合に対する警告構造を用いる計算ツールである。

2.1. 2o3 DA

2o3 DA は、動物を用いた試験を使用せずに、化学物質の皮膚感作性 (GHS 区分 1 または区分に該当しない (Not classified: NC)) を同定する方法である。現在のところ、2o3 DA は皮膚感作性の強度の細区分 (すなわち UN GHS 1A、1B) を判定することはできない。2o3 DA に含まれている試験方法の組み合わせは、皮膚感作性 AOP の KE1～KE3 のうち、少なくとも 2 つをカバーしている。2o3 DA は専門家の判断を必要としない、透明性のあるルールベースの方法である。KE1～KE3 (すなわち DPRA、KeratinoSens™、h-CLAT) の 3 種の *in chemico* あるいは *in vitro* 試験のうち 2 種類の試験で結果が一致した場合に、その結果を支持し、皮膚感作性の有無を判定する (3 種の試験の実施順については定めていない) 。最初の 2 つの試験結果が一致しない場合は、残りの KE の試験を実施する。最終判定は得られた予測の信頼性を考慮に入れた 2 つの一致した結果に基づいて判定される。2o3 DA の個々の試験は、皮膚感作性 KE に基づく OECD TG に含まれる試験 (OECD TG 442C、442D および 442E) であり、方法はそれぞれの TG に詳述されており、各試験法の適用範囲を考慮することが必要である。

- DPRA (TG 442C ; KE1) : 皮膚感作性物質は一般的に求電子性であり、タンパク質の求核部分と反応する。DPRA はシステインまたはリジン残基を含む 2 つのペプチドと化学物質を反応させ、未反応のペプチド量を基に化学物質の反応性を評価する。システインおよびリジン含有ペプチドの平均減少率が 6.38% (リジン含有ペプチドに共溶出がある場合は、システイン含有ペプチドの減少率が 13.89%) を超える場合、その化学物質は陽性と予測される。ペプチド減少率の平均値が 3 – 10%の範囲に含まれる場合、または、システイン単独予測モデルにおけるシステイン含有ペプチドの減少率が 9 – 17% の範囲に含まれる場合には 2 回の試験を実施し、結果が同じ場合は、その結果を最終判定とする。2 回の試験の結果が一致しない場合、3 回目の試験を実施し、多数決で最終判定を決定する。
- KeratinoSens™ (TG 442D ; KE2) : レポーター遺伝子を有するケラチノサイトは Nrf2-Keap1 経路を介して感作性物質に反応する。媒体対照物質と比較して細胞生存率＞70%で 1.5 倍を超えるルシフェラーゼの誘導を引き起こす化学物質は陽性と予測される。独立した 2 回の試験を実施し、結果が同じ場合は、その結果を最終判定とする。2 回の試験の結果が一致しない場合、3 回目の試験を実施し、多数決で最終判定を決定する。
- h-CLAT (TG 442E ; KE3) : 抗原提示細胞が活性化すると CD (Cluster of Differentiation) 86 および/または CD54 の発現が亢進する。h-CLAT は媒体対照と比較して細胞生存率≧50%で CD86 の発現が 1.5 倍を超える場合および／または CD54 の発現が 2 倍を超える場合、その化学物質は陽性と予測される。1 試験は 3 回の反復測定から成る独立した 2 回の試験を実施し、結果が同じ場合は、その結果を最終判定とする。2 回の試験の結果が一致しない場合、3 回目の試験を実施し、多数決で最終判定を決定する。
- ボーダーラインレンジ (Borderline Range: BR)
  試験データは変動する可能性があり、特にカットオフ値に近い場合、つまり BR 内にある場合、これらの変動により試験結果の不確実性が増大する。つまり、信頼性が低い領域を定義するために、2o3 DA の 3 つの KE に対応するアッセイごとに BR が定義されている。各アッセイの具体的な BR は次のとおりである。

- DPRA BR: 平均ペプチド減少率: 4.95 – 8.32%、システイン単独予測モデルのシステイン含有ペプチドの減少率：10.56 – 18.47%
- KeratinoSens™ BR: Imax: 1.35 – 1.67 倍
- h-CLAT BR: RFI CD86: 122 – 184%、RFI CD54: 157 – 255%

2.2. ITS DA

ITS DA は、AOP の KE1 と KE3 の試験方法と皮膚感作性の *in silico* による予測を使用する。KE1 の試験には DPRA、KE3 の試験には h-CLAT を使用し、各試験法の適用範囲を考慮することが必要である。皮膚感作性を予測する *in silico* ツールは Derek Nexus (ITSv1 DA) または OECD QSAR Toolbox (ITSv2 DA) のいずれかが使われる。ITS DA は h-CLAT と DPRA および Derek Nexus または OECD QSAR Toolbox の結果をスコア化し、その合計値によって化学物質を UN GHS 細区分 1A、1B または NC に分類できる (表 1) 。

表 1：ITS DA の概略

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		

DEREK: Derek Nexus

OECD TB: OECD QSAR Toolbox

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.

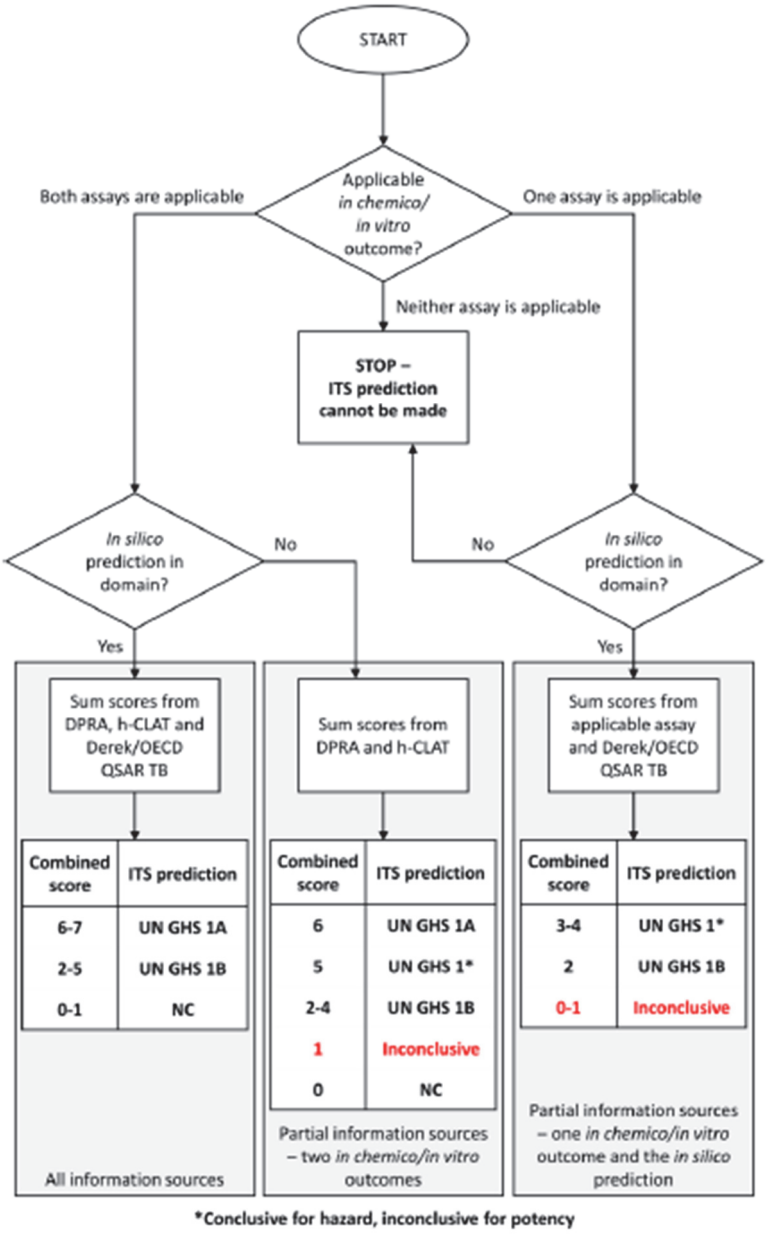


図 1：ITS DA の概略

Derek: Derek Nexus
OECD QSAR TB: OECD QSAR Toolbox

h-CLAT および DPRA の測定結果は、表 1 のように 0 から 3 のスコアに変換され、*in silico* による予測は、陽性の判定はスコア 1 に、陰性の判定はスコア 0 となる。

h-CLAT においては、最小誘導閾値 (Minimum Induction Threshold: MIT) を 10、150 および 5000 µg/mL のカットオフ値に基づいて 0 から 3 のスコアに変換する。試験の用量反応曲線から CD86 および/または CD54 のそれぞれ 1.5 倍および／または 2 倍の発現亢進を誘導する濃度の中央値を計算し、2 つの値の最小値を MIT と定義する。スコアは MIT の値に基づいて表 1 に示されたように割り当てられる。

MIT = min (EC150 CD86、EC200 CD54)

DPRA では、6.38、22.62 および 42.47%のカットオフ値に基づいてシステイン含有ペプチドおよびリジン含有ペプチドの平均減少率を 0 から 3 のスコアに変換する。リジン含有ペプチドに共溶出がある場合は、13.89、23.09 および 98.24%のカットオフ値に基づいてシステイン含有ペプチドの減少率を 0 から 3 のスコアに変換する。スコアはシステイン含有ペプチドおよびリジン含有ペプチドの平均減少率またはシステイン含有ペプチドの減少率に基づいて表 1 に示されたように割り当てられる。

- ITSv1 DA の *in silico* 予測は皮膚感作性を含むいくつかの毒性エンドポイントに関する警告構造を含む専門的な知識ベースのソフトウェアである Derek Nexus から導かれる。Derek Nexus は構造的な特徴、すなわちハプテンが皮膚のタンパク質に直接あるいは代謝／自動酸化後に求電子的に結合する可能性があるかどうかに基づいて警告を発する。
- ITSv2 DA の *in silico* 予測は OECD QSAR Toolbox v4.5 の皮膚感作性予測の自動ワークフローから導かれる。被験物質はタンパク質結合アラートについてプロファイリングされる。さらに自動酸化生成物および皮膚代謝物についてもタンパク質結合アラートについてプロファイリングされる。親化合物またはその生物学的代謝物中にタンパク質結合アラートが同定された場合、同じアラートを持つ化合物で皮膚感作性データがあるものを類似物質とする。タンパク質結合アラートが同定されない場合、構造プロファイラを用いて類似化合物を同定し、適切な類似物質が自動的に同定されない場合にはリードアクロスまたはプロファイラの結果から直接データギャップを埋める。
- 各スコアを合計した総合スコア (0 から 7) を用いて、皮膚感作性の有無 (UN GHS 区分 1) または NC および皮膚感作性強度 (UN GHS 区分 1A、1B または NC) を予測する。皮膚感作性の有無については、総合スコアが 2 以上の場合、その化学物質は皮膚感作性物質と判定される。皮膚感作性強度については、総合スコア 6 から 7 は UN GHS 区分 1A、総合スコア 2 から 5 は UN GHS 区分 1B、総合スコア 0 から 1 は NC と判定される。ITS DA に含まれる *in vitro* 試験は h-CLAT と DPRA である。

3. DASS の予測性

新たな試験法の組み合わせや新たなモデルを OECD GL497 に加える場合には、既存の結果と同等以上になることが求められると推察しており、この予測値をよく理解して開発に取り組むべきと考える。新たな DASS 開発の基準になる現時点で GL497 に採用されている DA を組み合わせたモデルの予測値を以下に示す。

3.1. 2o3 DA

2o3 DA の LLNA の感作性ハザードに対する予測結果を表 2 および Appendix 1 に示す。Accuracy (正確度) は 83%、Sensitivity (感度) は 82%、Specificity (特異度) は 85%、Balanced Accuracy (バランス精度) は 84%であった。

表 2　2o3 DA の予測値 (vs LLNA)

	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.*

LLNA の偽陰性物質 19 物質のリストを Appendix 2 に示している。2o3 DA で偽陰性となった LLNA 陽性物質の中に区分 1A の物質は含まれていなかった。ヒト感作性ハザードに対する予測結果も表 3 および Appendix 1 に示すが、2o3 DA の正確度は 89%、感度は 89%、特異度は 88%、平均正確度は 88%と LLNA のハザードに対する予測結果よりも全体的にわずかに高かった。ヒト皮膚感作性の偽陰性物質 5 物質のリストを Appendix 2 に示している。この中にも UN GHS 区分 1A の評価を誤った物質はなかった。Appendix 1 に示すように、2 試験ずつや 3 試験の組み合わせで予測性を検討したが、2o3 DA に匹敵するバランス精度は得られなかった。

表 3　2o3 DA の予測値 (vs human 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).

3.2. ITS DA

LLNA のハザード結果に対する ITSv1 DA の予測性は、正確度 87%、感度 92%、特異度 70%、平均正確度 81%であった (表 4、5 および Appendix 3) 。皮膚感作性強度の細区分ごとの予測性表 (以後、3×3 表とする) でまとめたところ、各細区分に対する全体の正確度は 71%であった (表 5) 。LLNA の偽陰性物質 11 物質のリストを Appendix 4 に示している。2o3 DA で偽陰性となった LLNA 陽性物質の中に区分 1A の物質は含まれていなかった。一方、ヒトのハザード結果に対する ITSv1 DA の予測性は、正確度は 86%、感度は 93%、特異度は 44%、バランス精度は 69%であった (表 6、7 および Appendix 3) 。表 7 に示すように、3×3 表でまとめたところ、各細区分に対する全体の正確度は 68%であった。ヒト皮膚感作性の偽陰性物質 4 物質のリストを Appendix 4 に示している。2o3 DA で偽陰性となった LLNA 陽性物質の中に区分 1A の物質は含まれていなかった。

GHS 細区分に関しては、UN GHS 区分 1A に相当する物質を誤って NC と評価することはなかったが、1A に相当する物質が 1B に、1B に相当する物質が誤って NC に過小評価される例が一部認められた。

なお、表 4 と表 5 および表 6 と表 7 の陽性物質数の違いは LLNA の結果はあるものの細区分結果が得られていない (EC3 値が算出されなかった) 物質によって起こっている。

表 4. ITSv1 DA の予測値 (vs LLNA)

	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).

表 5. ITSv1 DA の細区分ごとの予測値 (vs LLNA)

	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

表 6. ITSv1 DA の予測値 (vs human)

	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).

表 7.　ITSv1 DA の細区分ごとの予測値 (vs human)

	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

また、LLNA のハザード結果に対する ITSv2 DA の予測性は、正確度 88%、感度 93%、特異度 67%、バランス精度 80%であった (表 8、9 および Appendix 3) 。表 9 に示すように、3×3 表でまとめたところ、各細区分に対する全体の正確度は 71%であった。LLNA の偽陰性物質 9 物質のリストを Appendix 5 に示している。2o3 DA で偽陰性となった LLNA 陽性物質の中に区分 1A の物質は含まれていなかった。

一方、ヒトのハザード結果に対する ITSv2 DA の予測性は、正確度 87%、感度 94%、特異度 44%、バランス精度 69%であり、ITSv1 DA とほぼ同等であった (表 10) 。表 11 に示すように、3x3 表でまとめたところ、各細区分に対する全体の正確度は 70%であった。

ヒト皮膚感作性の偽陰性物質 3 物質のリストを Appendix 5 に示している。2o3 DA で偽陰性となった LLNA 陽性物質の中に区分 1A の物質は含まれていなかった。

GHS 細区分に関しては、UN GHS 区分 1A に相当する物質を誤って NC と評価することはなかったが、1A に相当する物質が 1B に、1B に相当する物質が誤って NC に過小評価される例が一部認められた。

なお、表 8 と表 9 および表 10 と表 11 の陽性物質数の違いは LLNA の結果はあるものの細区分結果が得られていない (EC3 値が算出されなかった) 物質によって起こっている。

表 8.　ITSv2 DA の予測値 (vs LLNA)

	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).

表 9.　ITSv2 DA の細区分ごとの予測値 (vs LLNA)

	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

表 10.　ITSv2 DA の予測値 (vs human)

	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).

表 11.　ITSv2 DA の細区分ごとの予測値 (vs human)

	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

Appendix 4 および 5 に示す偽評価物質に、UN GHS 区分 1A の評価 (2 段階のずれ) を誤った物質はなかった。1 段階で評価がずれた物質は、ITSv1 DA と ITSv2 DA を比較しても大差はなかった。それらの物質リストを Appendix 4 および 5 に示している。

なお、参考までに h-CLAT と DPRA でのみ評価し、*in silico* ツールを用いなかった場合の 3×3 表も Appendix 3 に示した。感度は 91～92%から、85～87%に落ち、特異度は 67～70%が 73%とわずかに上昇した。なお、バランス精度はほぼ同様であった。この場合でも UN GHS 区分 1A に相当する物質を誤って NC と評価することはなかった（あくまで参考として記載したものであり、本ガイドラインにおいて h-CLAT と DPRA でのみ評価可能ということではありません）。

上記の結果の要約を表 12 に示す。2o3 DA のハザード予測に関しては、ヒトおよび LLNA の結果に対する予測値を比較すると、バランス精度は 88%および 84%、感度は 89%および 82%、特異度は 88%および 85%であり、わずかであるが全ての項目でヒトの結果に対する予測値が、LLNA のそれよりも高かった。

ITS DA のハザード予測に関しては、ヒトおよび LLNA の結果に対する予測値を比較すると、バランス精度は 69% (ITSv1 DA と ITSv2 DA) および 81% (ITSv1 DA) と 80% (ITSv2 DA) 、感度は 93% と 94% (ITSv1 DA と ITSv2 DA) および 92% と 93% (ITSv1 DA と ITSv2 DA) 、特異度は 44% (ITSv1 DA と ITSv2 DA) および 70% と 67% (ITSv1 DA と ITSv2 DA) であった。ヒトの結果に対する予測の特異度が LLNA の結果に対するそれより 20%以上低かったため、ヒトの結果に対する予測のバランス精度も LLNA の結果に対するそれより 10%以上低かった。この原因は、ヒトのデータ数が少なく、特に非感作性物質のデータが少ないことによる可能性が考えられる。

ITS DA の感作性強度の予測に関して、ヒトおよび LLNA の結果に対する予測を比較すると、NC に対しては 44% (ITSv1 DA と ITSv2 DA) および 70%と 67% (ITSv1 DA と ITSv2 DA) 、1B に対しては、77%と 80% (ITSv1 DA と ITSv2 DA) および 71% と 72% (ITSv1 DA と ITSv2 DA) 、1A は、65% と 67% (ITSv1 DA と ITSv2 DA) および 74% と 72% (ITSv1 DA と ITSv2 DA) であった。ヒトの NC に対する予測が、LLNA のそれより、20%以上低かったが、1B および 1A に対する予測は、ヒトと LLNA で大きな差はなかった。この原因は、ハザード予測のケースと同じく、ヒトのデータ数が少ない可能性が示唆された。

LLNA のヒトのハザード予測に関しては、バランス精度は 58%、感度は 94%、特異度は 22% であった。ITS DA は、LLNA と比較して、感度は同等であったが、特異度は LLNA より 20%程度高く、それによりバランス精度も 10%程度高かった。これより、ITS DA は、ヒトのデータ数が少ない懸念はあるが、LLNA よりヒトの感作性ハザード予測能が高いことが示された。また、2o3 DA のヒトのハザード予測に関しては、LLNA と比較して、感度は 5%程度低かったが、特異度が 65%以上高く、それによりバランス精度も 30%程度高いことが示された。

LLNA のヒトの感作性強度予測に関しては、NC に対して 25%、1B に対して 74%、1A に対して 56%であった。ITS DA は、LLNA と比較して、1B に対しては同程度であったが、NC に対して 20%程度高く、1A に対しても 10%程度高かった。これより、ITS DA は、LLNA よりヒトの感作性強度予測能が高いことが示された。

表 12.　GL497 に記載された DA の予測性要約

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.

3.3. 本委員会の見解

LLNA のハザードに対する 2o3 DA の感度は表 2 に示すように、82%であり、化学物質の安全性評価目標値である 90%よりも 10%程度低い。解析の結果、UN GHS 区分 1A 物質 (強い感作性物質) を陰性と評価することはないことが分かったが (Appendix 2)、偽陰性が 18%も存在する。ただし、表 3 に示すように、ヒトのハザードに対する 2o3 DA の感度は 89%で、偽陰性率は 11%と低くなり、UN GHS 区分 1A 物質は見落としていない (Appendix 2)。そのため、2o3 DA による NC (区分に該当しない) の判定に、UN GHS 1B に相当する物質が含まれる可能性があることに留意が必要であるが、2o3 DA の感作予測性は妥当な範囲にあると考えた。

一方、LLNA およびヒトの感作性予測において、表 4 および表 6 に示すように ITS DA の感度は 90%より高く、UN GHS 区分 1A 物質は見落とさないことから (Appendix 4)、LLNA およびヒトに対する感作性ハザード予測にも用いることはできると考える。ただし、ITS DA のヒトに対する特異度の低さ (44%) が懸念点と考えられる。この原因の一つは、非感作性物質のデータ数が少ないことであり、これらのデータが増えない限り、偽陽性率が多いことを念頭におかねばならない。なお、ITS　DA に掲載されているヒトの感作性の細区分に対する全体の正確度は表 9 および 11 に示すように 70%程度であり、UN GHS 区分 1B を NC に、UN GHS 区分 1A を 1B に過小評価される物質が一定数存在する。UN GHS は化学物質の有害性情報の伝達を目的としているため、LLNA よりは高いとはいえ、感作性強度の過小評価はヒトの安全性リスクを低く見積もりかねないことに留意する必要がある。

4. DASS 追加の留意点

DASS の開発にあたり留意する点として、2o3 DA で BR が設定されたこと、および、ITS DA において *in silico* ツールが利用されたことの 2 点がある。

まず重要な点は、2o3 DA として使用されるアッセイの結果がアッセイごとに規定された BR 内に入った場合、それらの結果は結論できない (inconclusive) となることである。TG を用いて評価される使用者は注意せねばならない。

次に、GL497 に掲載されていない試験法は TG であっても利用できないことが挙げられる。もちろん、検証されていない試験法は、公表されている試験法であったとしても、行政的には利用できない。*In silico* ツールにおいても 4.2 に示す基準を満たしている必要がある。

今後、新たな *in vitro* 試験や *in silico* を両アプローチに組み込む場合に留意する点を、以下に詳述する。

4.1. ボーダーラインレンジ (BR)

2o3 DA に新たな *in vitro* 試験を組み込む場合には、*in vitro* 試験の不確定要素を減らすため、カットオフ値の付近にある値の採用を再検討する必要がある。そこで、バリデーション試験の結果を解析して、BR を設定することになり、感度が上がることになった。

DPRA に関しては、表 13 に示すようにバリデーション試験のペプチド減少率 (%) およびシステイン含有ペプチドのみの減少率 (%) の結果をもとに、各試験施設の BR が示された。3 つの試験施設における BR の下限値および上限値の平均値が最終的な BR の下限値と上限値である (4.95-8.32 および 10.56-18.47) 。この範囲に入った化合物は inconclusive となり、評価に使われることはない。なお、2023 年 7 月に TG442C が改訂され、DPRA の重量濃度法が TG に掲載されたことから、DPRA は混合物の評価が可能となっており、当委員会では、2o3 DA による混合物の感作性評価は可能であると考えている。

表 13. DPRA の BR の値

Data source	**Mean peptide depletion [%] (cut-off 6.38)**	**Cysteine-only depletion [%] (cut-off 13.89%)**
Validation study lab 1 (n[a]=13)	4.81 - 8.46	10.53 - 18.31
Validation study lab 2 (n[a]=14)	5.49 - 7.42	12.17 - 15.85
Validation study lab 3 (n[a]=14)	4.54 - 9.08	8.97 - 21.25
Validation study mean (lab 1-3)	**4.95 -8.32**	**10.56 - 18.47**
BASF SE historical data[b] (n = 385)	5.29 - 7.69	11.62 - 16.61
BASF SE experimental data[b] (n = 27)	5.45 – 7.31	11.89 – 15.90

[a] n: number of test chemicals out of the 24 chemicals assessed in the validation study for which at least three test runs were available.

[b] Published in Gabbert *et al.* (2020).

For the three laboratories participating in the validation study and only considering test chemicals for which at least three test runs were available (i.e. 13 to 14), the borderline range around the 6.38% mean peptide depletion cut-off varied between 4.54 to 5.49 (lower boundary) and 7.42 to 9.08% (upper boundary). ***The DPRA mean borderline range of all participating laboratories was 4.95 to 8.32%***, which was comparable to the range of 5.29 to 7.69% derived from historical data of a routine testing lab (assessing 385 substances) and to the experimentally determined range of 5.45 to 7.31% when repeatedly testing a single substance (i.e. EGDMA) (the latter two published in Gabbert et al., 2020). Likewise, the different borderline ranges determined for the cysteine-only depletion model were also very comparable with each other.

KeratinoSens™ に関しては、表 14 に示すようにバリデーション試験のルシフェラーゼ活性の誘導倍率の結果をもとに、各試験施設の BR が示された。5 つの試験施設における BR の下限値および上限値の平均値が最終的な BR の下限値と上限値である (1.35-1.67) 。この範囲に入った化合物は inconclusive となり、評価に使われることはない。

表 14.　KeratinoSens™ の BR の値

Data source	**Luciferase induction (cut-off 1.5)**
Validation study lab 1 (n[a] = 28)	1.37 – 1.64
Validation study lab 2 (n[a] = 28)	1.33 – 1.69
Validation study lab 3 (n[a] = 28)	1.33 – 1.69
Validation study lab 4 (n[a] = 28)	1.35 – 1.67
Validation study lab 5 (n[a] = 26)	1.37 – 1.65
Validation study mean (lab 1-5)	**1.35 – 1.67**
Givaudan experimental data on positive control (n = 123)	1.40 – 1. 60

[a] n: number of test chemicals out of the 28 chemicals assessed in the validation study for which at least three test runs were available.

h-CLAT に関しては、表 15 に示すようにバリデーション試験の CD54 の RFI (Relative Fluorescence Intensity) 値および CD86 の RFI 値の結果をもとに、各試験施設の BR が示された。4 つの試験施設における BR の下限値および上限値の平均値が最終的な BR の下限値と上限値である (RFI CD54: 157-255、RFI CD86: 122-184) 。この範囲内の化合物は inconclusive となり、評価に使われることはない。h-CLAT の BR の範囲は非常に広いため、使用者は特に注意すべきである。また、h-CLAT の適用範囲外にあたる logKow>3.5 の難水溶性物質については、陰性評価ができないことに留意が必要である。

表 15.　h-CLAT の BR の値

Table 2.4. h-CLAT borderline ranges determined based on the log pooled median absolute deviations.

Data source	**RFI CD54 (cut-off 200)**	**RFI CD86 (cut-off 150)**
Validation study lab 1 (n[a] = 24)	152 - 264	125 - 181
Validation study lab 2 (n[a] = 24)	153 - 261	125 - 181
Validation study lab 3 (n[a] = 24)	161 - 248	115 - 196
Validation study lab 4 (n[a] = 24)	162 - 247	125 - 180
Validation study mean (lab 1-4)	**157 - 255**	**122 - 184**
BASF SE historical data[b] (n = 136)	170 - 235	132 - 170

[a] n: number of test chemicals of the 24 chemicals assessed in the ring trial for which at least

4.2. ITS DA における *in silico* ツール

ITSv1 DA では *in silico* ツールとして Derek Nexus、ITSv2 DA では、OECD QSAR Toolbox を利用する。Derek Nexus は英国の Lhasa 社が開発した市販ソフトウェア (有償) であり、警告構造 (structural alert) を用いて皮膚感作性の予測を提供する知識ベースの予測ツールである。皮膚感作性については、化学物質が直接または代謝/自動酸化後に皮膚タンパク質への求電子的な結合能を有するかどうかを構造的特徴に基づいて予測する。OECD QSAR Toolbox は、OECD と欧州化学品庁 (European Chemicals Agency: ECHA) の協力により開発され、化学物質の有害性等を予測あるいは類推するための無償ツールである。DASS において OECD QSAR Toolbox で化学物質の感作性を予測する際には、OECD QSAR Toolbox に搭載されている機能を用いて標的化合物の他に自動酸化産物および皮膚代謝産物の予測を行う。これらは、プロファイラと呼ばれる機能を用いてタンパク質結合アラートについてプロファイリングし、同定されたタンパク質結合性のプロファイルを有する物質を検索してアナログベースのリードアクロスによる類推を行う、あるいは適切なアナログが自動的に見つからない場合にはプロファイラの結果をそのまま用いる。

ITS DA で使用される *in silico* ツール (Derek Nexus および OECD QSAR Toolbox) は、(Q)SAR 予測または自動化ワークフロー (OECD QSAR Toolbox 内の Automated workflow) を用いたリードアクロスのいずれかを実行できる。(Q)SAR には、構造活性相関 (SAR) モデル (警告構造やエキスパートシステム) と定量的構造活性相関 (QSAR) モデル (統計ツール) の両方が含まれる。DASS で使用する(Q)SAR モデルは、「OECD PRINCIPLES FOR THE VALIDATION 、 FOR REGULATORY PURPOSES OF (QUANTITATIVE) STRUCTURE-ACTIVITY RELATIONSHIP MODELS (規制目的のための(Q)SAR モデルのバリデーションのための OECD 原則) (以下、OECD QSAR バリデーション原則という) 」を満たす必要があり、既定の QSAR モデル報告様式 (QSAR Model Reporting Format: QMRF) [12] による文書が必要である。OECD QSAR バリデーション原則の 1 つでは、モデルの適用範囲に言及している。適用範囲は、記述子を用いるモデルの場合にはトレーニングデータセット (モデル構築に利用されるデータセット) のうちモデルに使用された記述子の範囲、およびトレーニングデータセットに含まれる部分構造の種類などによって定義され、適用範囲を外れると信頼性の低い予測しか得られない可能性を示している。

DASS では使用する *in silico* ツールの適用範囲の考え方を提供しており (Appendix 6) 、DASS におけるデータの解釈手順によっては、適用範囲外の化学物質に関しては信頼度の低い予測をもたらす可能性があり、inconclusive との判断が下される。*In silico* ツールを含む DASS では、使用者は個々の *in silico* ツールの限界と適用範囲を参照し、結果の判断を行う必要がある。

In silico ツールでは、被験物質の化学構造を予測の情報源として使用するため、予測は化学物質の入力された化学構造に依存しており、描画された化学構造、SMILES 記法や InChI による化学構造の線形表記によって入力できる。また、単一の化学物質は立体構造の違い、塩の違い、混合物中の主成分の違いなどにより、いくつかの CAS または EC 番号で表されるため、正確な構造を特定することが重要である。

ITSv1 DA で用いられる Derek Nexus による予測では、すべての陽性予測物質 (ITSv1 DA では certain、probable、plausible or equivocal の場合を指す) は、適用範囲内にあると考えられる。陰性予測物質 (ITSv1 DA では doubted、improbable、impossible or non-sensitiser の場合) も、Derek Nexus による“Misclassified features”および／または“Unclassified features”の特徴を含ま

ない限り、適用範囲にあると考える。“Misclassified feature”を有する陰性予測は、Derek Nexus において警告構造とはしていないが、開発元である Lhasa 社が有する皮膚感作性に関するデータセットにおいて陽性物質にのみ観察された構造があることを示唆している。また、“Unclassified features”を有する陰性予測は、Lhasa 社が保有する皮膚感作性に関するデータセットに含まれない、未知の構造が存在することを示している[13]。通常、これらの特徴を含む予測には専門家判断が推奨されるが、DASS で必要とされるデータ解釈手順としては、専門家判断を要求していない。

ITSv2 DA で用いられる OECD QSAR Toolbox によるリードアクロスの予測適用範囲の計算は、Toolbox によって自動的に提供され、構造的、パラメトリックおよび機構的の 3 層から構成される。個々の予測のために考慮される適用範囲は、予測の種類と結果に依存し、適用範囲内の予測結果は、DASS において適用可能とみなされる。

5. 結論

本委員会では、OECD が定めた GL497 に含まれる 2o3 DA および ITS DA による感作性物質のハザード評価は妥当性が高いと考え、行政的な受け入れについても利用可能と判断した。

6. 参考文献

1) OECD (2022). OECD Guidelines for the Testing of Chemicals No.406. Skin Sensitisation Guinea Pig Maximisation Test and Buehler Test, Organisation for Economic Cooperation and Development, Paris. Available at: (oecd-ilibrary.org)
2) OECD (2010). OECD Guidelines for the Testing of Chemicals No. 429. Skin Sensitization: Local Lymph Node Assay (LLNA), Organisation for Economic Cooperation and Development, Paris. Available at: (oecd-ilibrary.org)
3) OECD (2010). OECD Guidelines for the Testing of Chemicals No. 442A. Skin Sensitization: Local Lymph Node Assay: DA, Organisation for Economic Cooperation and Development, Paris. Available at: (oecd-ilibrary.org)
4) OECD (2018). OECD Guidelines for the Testing of Chemicals No. 442B. Local lymph mode assay: BrdU-ELISA or -FCM, Organisation for Economic Cooperation and Development, Paris. Available at: (oecd-ilibrary.org)
5) OECD (2024). OECD Guidelines for the Testing of Chemicals No. 442C: *In chemico* Skin Sensitisation Assays addressing the Adverse Outcome Pathway key event on covalent binding to proteins, Organisation for Economic Cooperation and Development, Paris. Available at: (oecd-ilibrary.org)
6) OECD (2024). OECD Key event based test Guideline No. 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)
7) OECD (2024). OECD Key event based test Guideline No. 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)

8) 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: (oecd-ilibrary.org)

9) OECD (2023). OECD Guidelines No.497. Defined Approaches on Skin Sensitisation, Organisation for Economic Cooperation and Development, Paris. Available at: (oecd-ilibrary.org)

10) OECD (2021). Series on Testing and Assessment No. 336. Supporting Document to the OECD Guideline 497 on Defined Approaches for Skin Sensitisation, Organisation for Economic Cooperation and Development, Paris. Available at: (oecd-ilibrary.org)

11) Kolle, S.N., et al. (2021). Assessing Experimental Uncertainty in Defined Approaches: Borderline Ranges for *In Chemico* and *In Vitro* Skin Sensitization Methods Determined from Ring trial Data, *Applied in Vitro Toxicology*, 7(3), 102-101. DOI:10.1089/aivt.2121.0003

12) 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)].

13) 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

Appendix 1　3 of 3, 2 of 2 および 2 of 3 DA の予測性

1) LLNA との比較

	LLNA	
3o3	Non	Sens
Non	10	1
Sens	9	80
Inconclusive	14	53

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

	3o3
(N=100)	
Accuracy (%)	90
Sensitivity (%)	99
Specificity (%)	53
Balanced Accuracy (%)	76

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

2o2(1)	LLNA	
DPRA and KeratinoSens	Non	Sens
Non	19	40
Sens	10	78
Inconclusive	4	17

2o2(2)	LLNA	
h-CLAT and KeratinoSens	Non	Sens
Non	11	26
Sens	9	59
Inconclusive	13	49

DA performance vs. LLNA data	DPRA and Keratino-Sens
(N=147)	
Accuracy (%)	66
Sensitivity (%)	66
Specificity (%)	66
Balanced Accuracy (%)	66

DA performance vs. LLNA data	h-CLAT and Keratino-Sens
(N=105)	
Accuracy (%)	67
Sensitivity (%)	69
Specificity (%)	55
Balanced Accuracy (%)	62

2o2(3)	LLNA	
h-CLAT and DPRA	Non	Sens
NC	12	24
Sens	9	62
Inconclusive	12	48

DA performance vs. LLNA data	DPRA and h-CLAT
(N=107)	
Accuracy (%)	69
Sensitivity (%)	72
Specificity (%)	57
Balanced Accuracy (%)	65

2) Human skin sensitization との比較

	Human	
3o3	Non	Sens
Non	2	1
Sens	3	36
Inconclusive	6	16

	Human	
2o3 DA	Non	Sens
Non	7	5
Sens	1	42
Inconclusive	3	7

	3o3
(N=42)	
Accuracy (%)	90
Sensitivity (%)	97
Specificity (%)	40
Balanced Accuracy (%)	69

DA performance vs. Human data	2o3 DA
(N=55)	
Accuracy (%)	89
Sensitivity (%)	89
Specificity (%)	88
Balanced Accuracy (%)	88

2o2(1)	Human	
DPRA and KeratinoSens	Non	Sens
Non	6	12
Sens	2	34
Inconclusive	3	8

2o2(2)	Human	
h-CLAT and KeratinoSens	Non	Sens
Non	3	12
Sens	3	27
Inconclusive	5	15

DA performance vs. Human data	DPRA and Keratino-Sens
(N=54)	
Accuracy (%)	74
Sensitivity (%)	74
Specificity (%)	75
Balanced Accuracy (%)	74

DA performance vs. Human data	h-CLAT and Keratino-Sens
(N=45)	
Accuracy (%)	67
Sensitivity (%)	69
Specificity (%)	50
Balanced Accuracy (%)	60

2o2(3)	Human	
h-CLAT, DPRA	NC	Sens
NC	2	10
Sens	4	31
Borderline negative, Borderline positive, Inconclusive, NA	5	12

NA: Not Available

DA performance vs. Human data	DPRA and h-CLAT
(N=47)	
Accuracy (%)	70
Sensitivity (%)	76
Specificity (%)	33
Balanced Accuracy (%)	54

Appendix 2　2o3 DA の偽評価物質

1)LLNA との比較 (偽陰性 19 物質、偽陽性 4 物質)

Sort	Curated Chemical name	CASRN	LLNA. GHS.BIN	LLNA. GHS.SUB	2o3 DA Call.Conf
10	3-Aminophenol	591-27-5	1	1B	0
12	alpha-Amylcinnamic alcohol	101-85-9	1	1B	0
13	Anethole	104-46-1	1	NA	0
15	Anisyl alcohol	105-13-5	1	1B	0
30	Benzyl salicylate	118-58-1	1	1B	0
47	Chlorpromazine	50-53-3	1	NA	0
76	DMSO	67-68-5	1	1B	0
117	p-Isobutyl-alpha-methylhydrocinnamaldehyde	6658-48-6	1	1B	0
120	alpha-Isomethylionone	127-51-5	1	1B	0
122	Isopropyl myristate	110-27-0	1	1B	0
129	Linalool	78-70-6	1	1B	0
141	Methyl pyruvate	600-22-6	1	1B	0
160	OTNE	54464-57-2	1	1B	0
176	Pyridine	110-86-1	1	1B	0
177	Resorcinol	108-46-3	1	1B	0
180	Salicylic acid	69-72-7	1	1B	0
181	Sodium lauryl sulfate	151-21-3	1	1B	0
182	Squaric acid	2892-51-5	1	NA	0
187	2,2,6,6-Tetramethylheptane-3,5-dione	1118-71-4	1	1B	0
111	2-Hydroxypropyl methacrylate	923-26-2	NC	NC	1
114	1-Iodohexane	638-45-9	NC	NC	1
135	Methyl 3-bromoproprionate	3395-91-3	NC	NC	1
150	4-Methyl-2-nitroanisole	119-10-8	NC	NC	1

2) Human skin sensitisation との比較(偽陰性 5 物質、偽陽性 1 物質)

Sort	Curated Chemical name	CASRN	HU. GHS.BIN	2 of 3 DA Call.Conf
47	Chlorpromazine	50-53-3	1B	0
176	Pyridine	110-86-1	1B	0
25	Benzyl alcohol	100-51-6	1B	0
123	Kanamycin	59-01-8	1B	0
183	Sulfanilamide	63-74-1	1B	0
106	Hydrocortisone	50-23-7	NC	1

Note: Chemicals highlighted in light green have false negatives and those highlighted in yellow have false positives.
LLNA.GHS.BIN: LLNA Binary hazard reference classification
LLNA.GHS.SUB: LLNA Potency reference subcategorisation
2 of 3 DA Call.Conf: 2 of 3 DA Hazard prediction considering confidence workflow
HU.GHS.BIN: Human Binary hazard reference classification

Appendix 3　ITS DA の予測性

1)ITSv1 DA

ITSv1 DA in comparison to LLNA
（GL497 と同じ）

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

71% correct classification overall for potency
80% balanced accuracy overall for hazard
(Sensitivity 91%　Specificity 70%)

ITSv1 DA Score h-CLAT and DPRA in comparison to LLNA（*in silico* を含まず、Borderline 未使用）

h-CLAT, DPRA	LLNA		
	NC	1B	1A
NC	22	17	0
1B	8	55	20
1A	0	6	18

65% correct classification overall for potency
79% balanced accuracy overall for hazard
(Sensitivity 85%　Specificity 73%)

ITSv1 DA in comparison to Human
（GL497 と同じ）

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

68% correct classification overall for potency
68% balanced accuracy overall for hazard
(Sensitivity 92%　Specificity 44%)

ITSv1 DA Score h-CLAT and DPRA in comparison to Human（*in silico* を含まず、Borderline 未使用）

h-CLAT, DPRA	Human		
	NC	1B	1A
NC	4	7	2
1B	5	22	12
1A	0	1	6

54% correct classification overall for potency
63% balanced accuracy overall for hazard
(Sensitivity　82%　Specificity 44%)

2)　ITSv2 DA

ITSv2 DA in comparison to LLNA
（GL497 と同じ）

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

71% correct classification overall for potency
79% balanced accuracy overall for hazard
(Sensitivity 92%　Specificity 67%)

ITSv2 DA Score h-CLAT and DPRA in comparison to LLNA（*in silico* を含まず、Borderline 未使用）

h-CLAT, DPRA	LLNA (n=141)		
	NC	1B	1A
NC	22	14	0
1B	8	55	18
1A	0	6	18

67% correct classification overall for potency
80% balanced accuracy overall for hazard
(Sensitivity 87%　Specificity73%)

ITSv2 DA in comparison to Human
（GL497 と同じ）

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

70% correct classification overall for potency
69% balanced accuracy overall for hazard
(Sensitivity 94% Specificity 44%)

ITSv2 DA Score h-CLAT and DPRA in comparison to Human（*in silico* を含まず、Borderline 未使用）

h-CLAT, DPRA	Human		
	NC	1B	1A
NC	4	6	1
1B	5	22	11
1A	0	1	6

57% correct classification overall for potency
65% balanced accuracy overall for hazard
(Sensitivity 85% Specificity 44%)

Appendix 4 1TS v 1 DA 偽評価物質

1)LLNA との比較(偽陰性 11 物質、偽陽性 9 物質)

Sort	Curated Chemical name	CASRN	LLNA. GHS.SUB	ITSv1 DA Score	ITSv1 DA Pot.Conf
18	BADGE	1675-54-3	1A	4	1B
60	Dibenzoyl peroxide	94-36-0	1A	4	1B
77	DNBS, sodium salt	885-62-1	1A	5	1B
96	Glyoxal	107-22-2	1A	5	1B
104	HHPA	85-42-7	1A	3	1B
119	Isoeugenol	97-54-1	1A	4	1B
130	Maleic anhydride	108-31-6	1A	5	1B
154	1-Naphthol	90-15-3	1A	4	1B
157	2-Nitro-p-phenylenediamine	5307-14-2	1A	5	1B
171	Phthalic anhydride	85-44-9	1A	3	1B
8	5-Amino-o-cresol	2835-95-2	1B	6	1A
21	1,2-Benzisothiazol-3(2H)-one	2634-33-5	1B	7	1A
51	Citral	5392-40-5	1B	7	1A
58	Diacetyl	431-03-8	1B	6	1A
65	Diethyl maleate	141-05-9	1B	6	1A
80	Ethyl acrylate	140-88-5	1B	6	1A
86	2-Ethylhexyl acrylate	103-11-7	1B	6	1A
98	Hepta-2,4-dienal	5910-85-0	1B	7	1A
100	trans-Hex-2-enal	6728-26-3	1B	7	1A
134	Methyl acrylate	96-33-3	1B	7	1A
179	Safranal	116-26-7	1B	6	1A
189	Thiram	137-26-8	1B	7	1A
5	Allyl phenoxyacetate	7493-74-5	1B	0	NC
15	Anisyl alcohol	105-13-5	1B	1	NC
30	Benzyl salicylate	118-58-1	1B	1	NC
70	Dihydroeugenol	2785-87-7	1B	1	NC
76	DMSO	67-68-5	1B	1	NC
122	Isopropyl myristate	110-27-0	1B	1	NC
141	Methyl pyruvate	600-22-6	1B	1	NC
161	Oxalic acid	144-62-7	1B	1	NC
176	Pyridine	110-86-1	1B	1	NC
180	Salicylic acid	69-72-7	1B	1	NC
181	Sodium lauryl sulfate	151-21-3	1B	0	NC
17	Applelide	478695-70-4	NC	2	1B
44	3-Chloro-p-anisaldehyde	4903-09-7	NC	3	1B
111	2-Hydroxypropyl methacrylate	923-26-2	NC	3	1B
114	1-Iodohexane	638-45-9	NC	4	1B
123	Kanamycin	59-01-8	NC	2	1B
135	Methyl 3-bromoproprionate	3395-91-3	NC	3	1B
146	2-Methyldecanenitrile	69300-15-8	NC	2	1B
150	4-Methyl-2-nitroanisole	119-10-8	NC	2	1B
166	3-Phenoxypropanenitrile	3055-86-5	NC	2	1B

2)Human skin sensitisation との比較(偽陰性 4 物質、偽陽性 5 物質)

Sort	Curated Chemical name	CASRN	HU. GHS.SUB	ITSv1 DA Score	ITSv1 DA Pot.Conf
32	BGE	2426-08-6	1A	3	1B
96	Glyoxal	107-22-2	1A	5	1B
133	1-(4-Methoxyphenyl)pent-1-en-3-one	104-27-8	1A	4	1B
168	Phenylacetaldehyde	122-78-1	1A	5	1B
147	6-Methylhepta-3,5-dien-2-one	1604-28-0	1A	4	1B
68	Diethylenetriamine	111-40-0	1A	2	1B
20	Benzaldehyde	100-52-7	1A	2	1B
189	Thiram	137-26-8	1B	7	1A
80	Ethyl acrylate	140-88-5	1B	6	1A
131	2-Mercaptobenzothiazole	149-30-4	1B	6	1A
176	Pyridine	110-86-1	1B	1	NC
25	Benzyl alcohol	100-51-6	1B	1	NC
183	Sulfanilamide	63-74-1	1B	1	NC
54	Coumarin	91-64-5	1B	1	NC
52	Citronellol	106-22-9	NC	4	1B
101	Hexyl salicylate	6259-76-3	NC	2	1B
120	alpha-Isomethylionone	127-51-5	NC	3	1B
160	OTNE	54464-57-2	NC	3	1B
106	Hydrocortisone	50-23-7	NC	4	1B

Note: Note: Chemicals highlighted in light green have misclassification and those highlighted in yellow have false positives.
LLNA.GHS.SUB: LLNA Potency reference subcategorization
ITSv1 DA Score: ITSv1 DA total score
ITSv1.DA Pot.Conf: ITSv1 DA Hazard prediction considering confidence workflow
HU.GHS.BIN: Human Binary hazard reference classification

Appendix5 1TSv2 DA 評価物質

1)LLNA との比較(偽陰性 9 物質、偽陽性 10 物質)

Sort	Curated Chemical name	CASRN	LLNA. GHS.SUB	ITSv2 DA Score	ITSv2 DA Pot.Conf
18	BADGE	1675-54-3	1A	4	1B
60	Dibenzoyl peroxide	94-36-0	1A	4	1B
77	DNBS, sodium salt	885-62-1	1A	5	1B
96	Glyoxal	107-22-2	1A	5	1B
104	HHPA	85-42-7	1A	3	1B
119	Isoeugenol	97-54-1	1A	4	1B
130	Maleic anhydride	108-31-6	1A	5	1B
154	1-Naphthol	90-15-3	1A	4	1B
157	2-Nitro-p-phenylenediamine	5307-14-2	1A	5	1B
171	Phthalic anhydride	85-44-9	1A	3	1B
8	5-Amino-o-cresol	2835-95-2	1B	6	1A
21	1,2-Benzisothiazol-3(2H)-one	2634-33-5	1B	7	1A
51	Citral	5392-40-5	1B	7	1A
58	Diacetyl	431-03-8	1B	6	1A
65	Diethyl maleate	141-05-9	1B	6	1A
80	Ethyl acrylate	140-88-5	1B	6	1A
86	2-Ethylhexyl acrylate	103-11-7	1B	6	1A
98	Hepta-2,4-dienal	5910-85-0	1B	7	1A
100	trans-Hex-2-enal	6728-26-3	1B	7	1A
134	Methyl acrylate	96-33-3	1B	7	1A
179	Safranal	116-26-7	1B	6	1A
189	Thiram	137-26-8	1B	6	1A
5	Allyl phenoxyacetate	7493-74-5	1B	1	NC
30	Benzyl salicylate	118-58-1	1B	1	NC
70	Dihydroeugenol	2785-87-7	1B	1	NC
76	DMSO	67-68-5	1B	1	NC
122	Isopropyl myristate	110-27-0	1B	1	NC
141	Methyl pyruvate	600-22-6	1B	1	NC
161	Oxalic acid	144-62-7	1B	1	NC
176	Pyridine	110-86-1	1B	1	NC
181	Sodium lauryl sulfate	151-21-3	1B	0	NC
17	Applelide	478695-70-4	NC	2	1B
25	Benzyl alcohol	100-51-6	NC	2	1B
44	3-Chloro-p-anisaldehyde	4903-09-7	NC	3	1B
111	2-Hydroxypropyl methacrylate	923-26-2	NC	2	1B
114	1-Iodohexane	638-45-9	NC	3	1B
123	Kanamycin	59-01-8	NC	2	1B
135	Methyl 3-bromoproprionate	3395-91-3	NC	3	1B
146	2-Methyldecanenitrile	69300-15-8	NC	2	1B
150	4-Methyl-2-nitroanisole	119-10-8	NC	2	1B
166	3-Phenoxypropanenitrile	3055-86-5	NC	2	1B

2)Human skin sensitisation との比較(偽陰性 3 物質、偽陽性 5 物質)

Sort	Curated Chemical name	CASRN	HU. GHS.SUB	ITSv2 DA Score	ITSv2 DA Pot.Conf
32	BGE	2426-08-6	1A	3	1B
96	Glyoxal	107-22-2	1A	5	1B
133	1-(4-Methoxyphenyl)pent-1-en-3-one	104-27-8	1A	4	1B
168	Phenylacetaldehyde	122-78-1	1A	5	1B
147	6-Methylhepta-3,5-dien-2-one	1604-28-0	1A	4	1B
68	Diethylenetriamine	111-40-0	1A	2	1B
189	Thiram	137-26-8	1B	6	1A
80	Ethyl acrylate	140-88-5	1B	6	1A
131	2-Mercaptobenzothiazole	149-30-4	1B	6	1A
176	Pyridine	110-86-1	1B	1	NC
183	Sulfanilamide	63-74-1	1B	0	NC
54	Coumarin	91-64-5	1B	0	NC
52	Citronellol	106-22-9	NC	4	1B
101	Hexyl salicylate	6259-76-3	NC	2	1B
120	alpha-Isomethylionone	127-51-5	NC	3	1B
160	OTNE	54464-57-2	NC	3	1B
106	Hydrocortisone	50-23-7	NC	3	1B

Note: Note: Chemicals highlighted in light green have misclassification and those highlighted in yellow have false positives.
LLNA.GHS.SUB: LLNA Potency reference subcategorization
ITSv2 DA Score: ITSv2 DA total score
ITSv2.DA Pot.Conf: ITSv2 DA Hazard prediction considering confidence workflow
HU.GHS.BIN: Human Binary hazard reference classification

添付資料 3

OCDE/OECD 497

Adopted
14 June 2021
Corrected: 4 July 2023

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 (2.2.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.

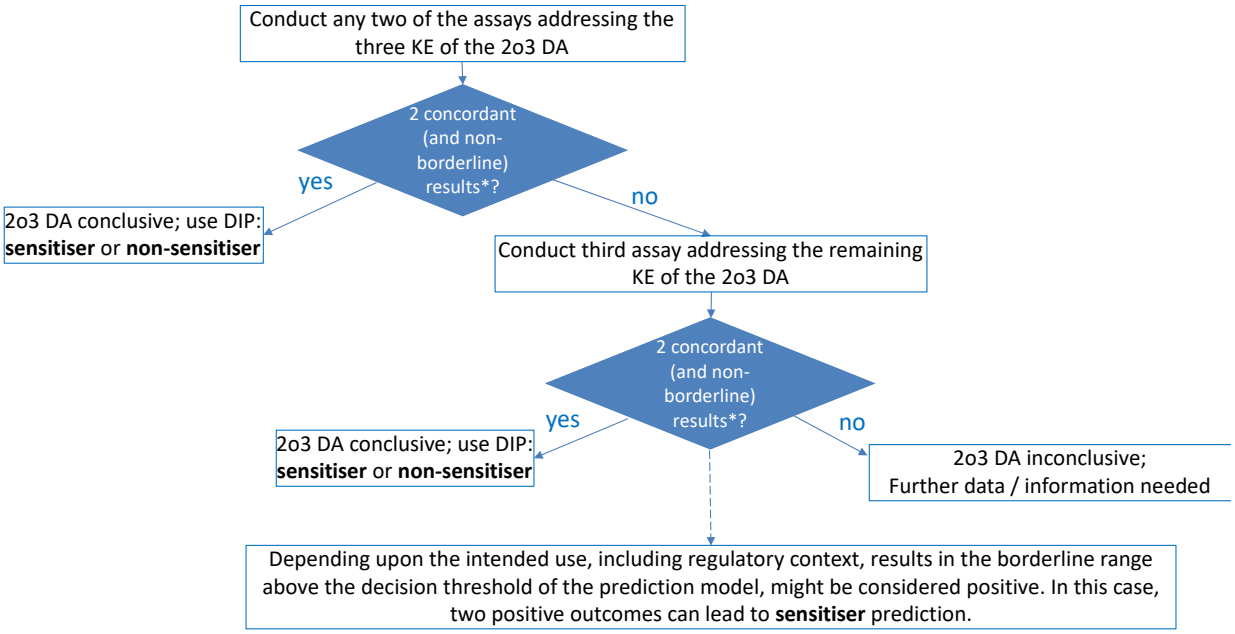
45. Positive and negative test results falling within these BRs as well as inconclusive results due to limitations in the *in chemico/in vitro* test guidelines are of lower confidence. For example, negative h-CLAT results obtained for a chemical with Log P > 3.5 (according to TG 442E (4)) are of lower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

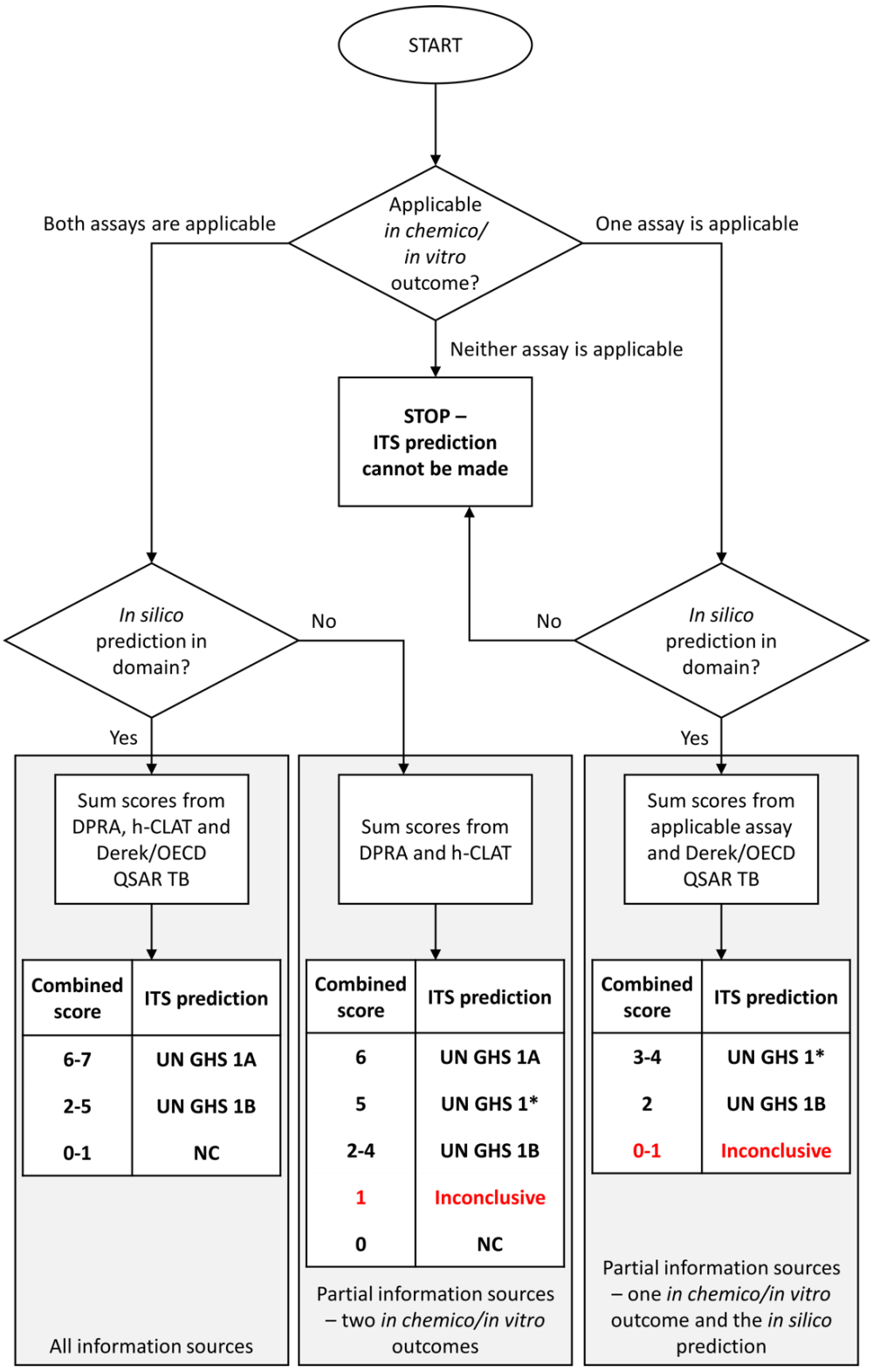
3.1.4. Confidence in the ITS DA predictions

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

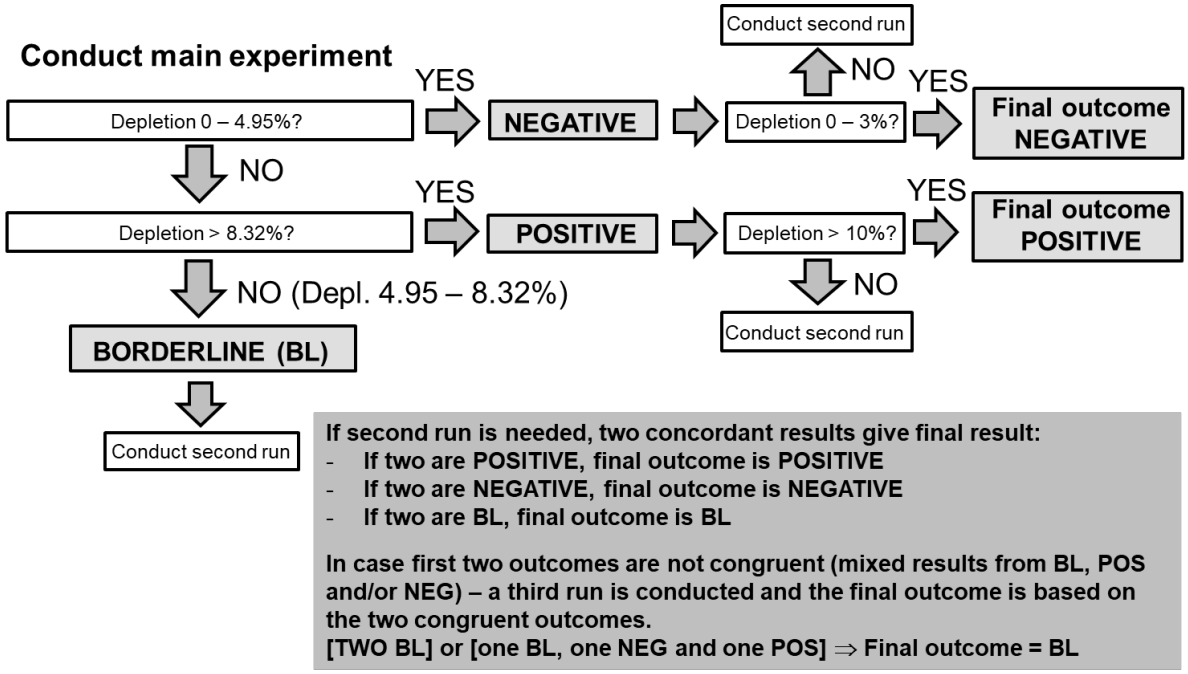
3.2. References

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)]

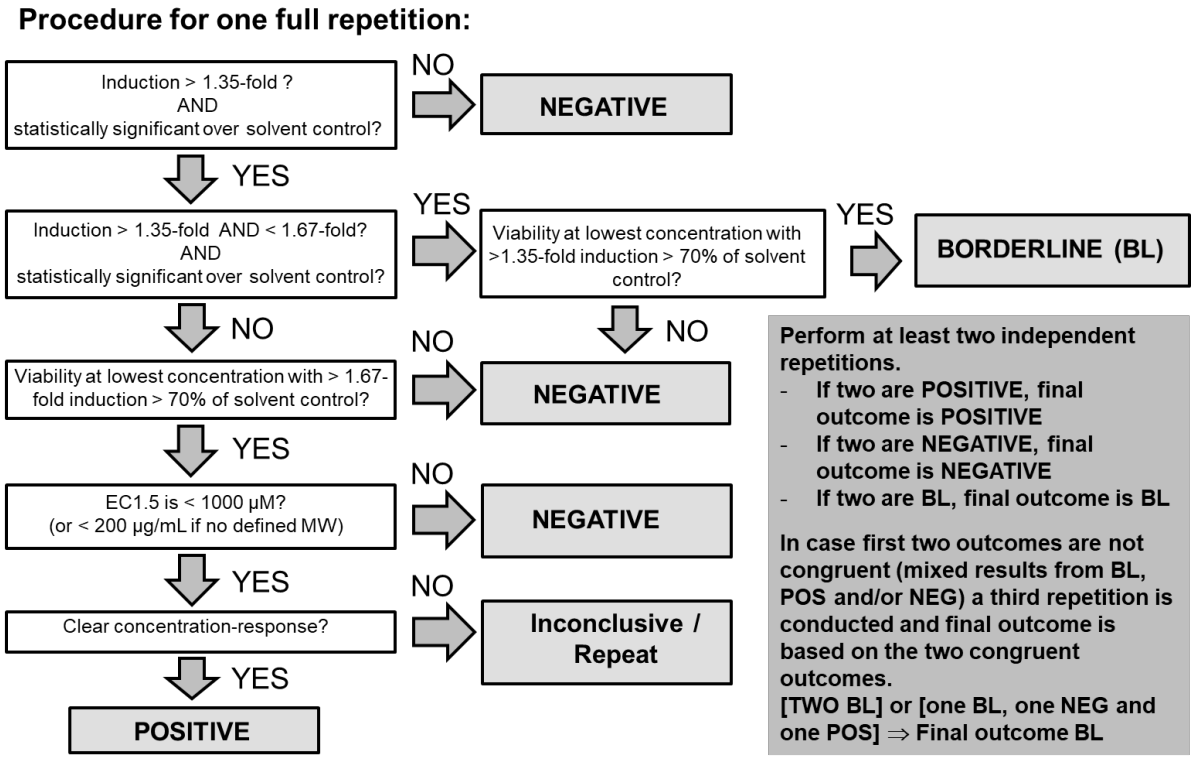
Annex 1: Prediction model for the individual *in chemico*/*in vitro* tests with multiple runs for use in 2o3 DA

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.

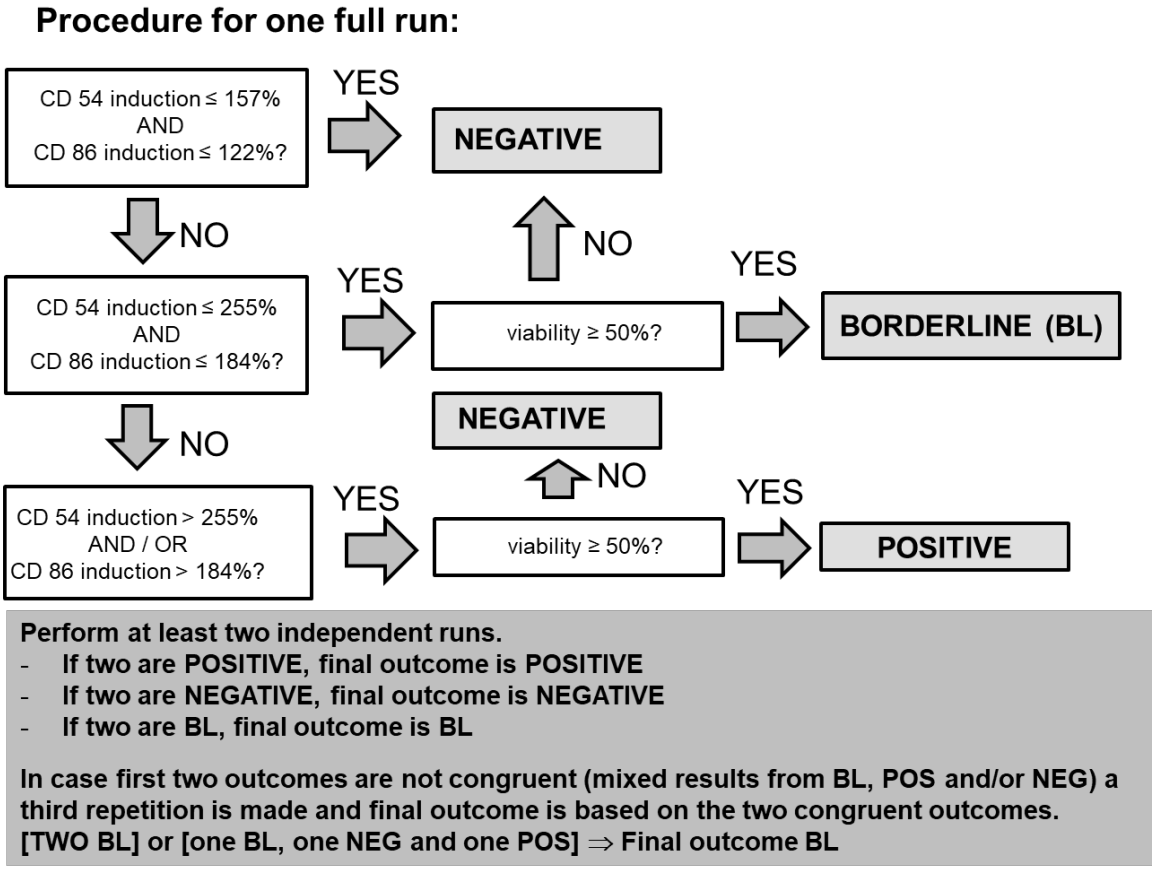
97. For the DPRA, repeated runs are required to be conducted if average depletion is within the range 3 - 10% (9 – 17% in case of Cysteine only depletion model is used). For this adaptation, the flowchart in Figure 1.1 is used to decide on run repetition and borderline assessment within the 2o3 DA.



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]

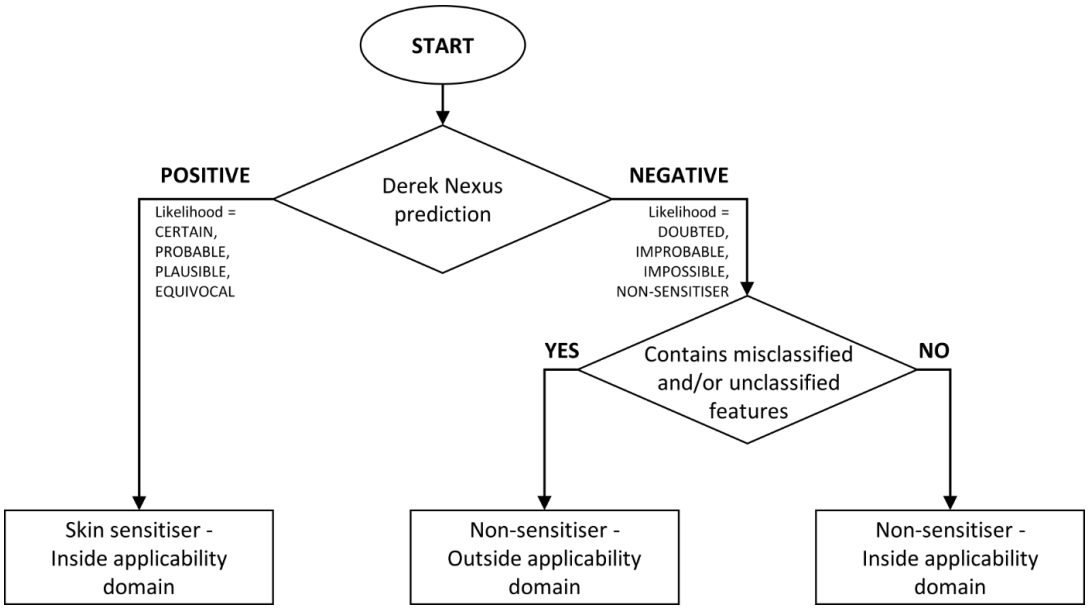
[3] OECD (2017), *Guidance on Grouping of Chemicals, Second Edition*, OECD Series on Testing and Assessment, No. 194, OECD Publishing, Paris, https://doi.org/10.1787/9789264274679-en.

[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**

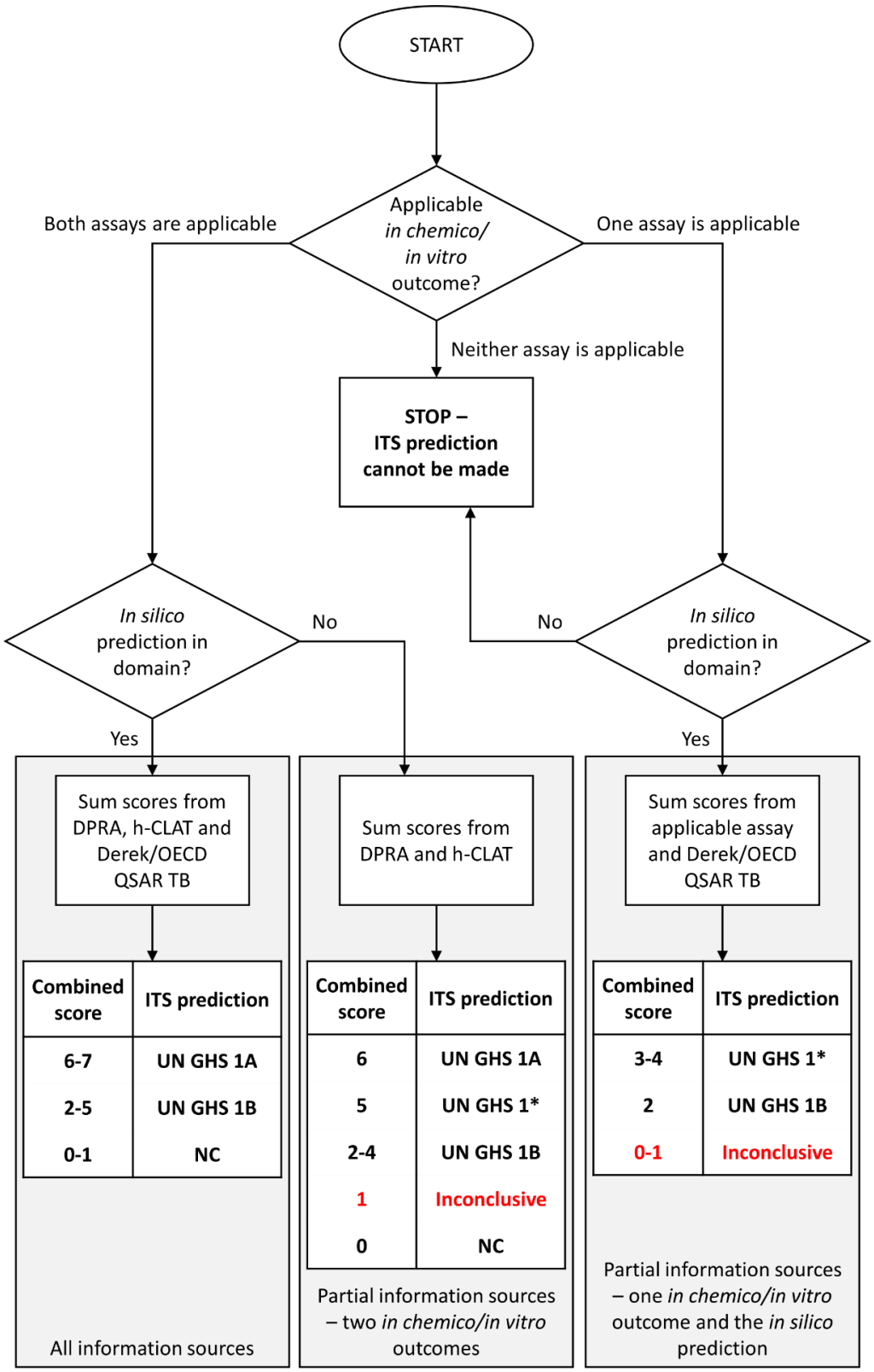
Confidence in ITS predictions

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.