

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 : KeratinoSensTM 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-SENSTM 、IL-8 Luc assay および GARDTM-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相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、KeratinoSensTM、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、KeratinoSensTM、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。
- KeratinoSensTM (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: I_{max}: 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*	<i>In silico</i> (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 sensitizers and UN GHS 1B correspond to other (moderate to weak) sensitizers. Not classified are considered non-sensitizers. *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。

$$\text{MIT} = \min(\text{EC}_{150} \text{ CD86}, \text{EC}_{200} \text{ CD54})$$

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)

2o3 DA	LLNA	
	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)

2 of 3 DA	Human	
	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)

ITSv1 DA	LLNA	
	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)

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

表 6. ITSv1 DA の予測値 (vs human)

ITSv1 DA	Human	
	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)

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

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、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 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)

ITSv2 DA	LLNA	
	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)

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

表 10. ITSv2 DA の予測値 (vs human)

ITSv2 DA	Human	
	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)

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

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は、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)	<i>in vivo</i>	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

^an: number of test chemicals out of the 24 chemicals assessed in the validation study for which at least three test runs were available.

^bPublished 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.

KeratinSens™ 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. KeratinSens™ 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、プロファイル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ル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](https://www.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](https://www.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](https://www.oecd-ilibrary.org))
- 4) OECD (2018). OECD Guidelines for the Testing of Chemicals No. 442B. Local lymph node assay: BrdU-ELISA or -FCM, Organisation for Economic Cooperation and Development, Paris. Available at: ([oecd-ilibrary.org](https://www.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](https://www.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](https://www.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](https://www.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 \(europa.eu\)](https://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	
	Non	Sens
3o3		
Non	10	1
Sens	9	80
Inconclusive	14	53

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

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

	2o3 DA
(N=134)	
Accuracy (%)	83
Sensitivity (%)	82
Specificity (%)	85
Balanced Accuracy (%)	84

2o2(1)	LLNA	
DPRAs 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	DPRAs and KeratinoSens
(N=147)	
Accuracy (%)	66
Sensitivity (%)	66
Specificity (%)	66
Balanced Accuracy (%)	66

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

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

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

2) Human skin sensitization との比較

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

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

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

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

2o2(1)	Human	
DPRAs 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	DPRAs and KeratinoSens
(N=54)	
Accuracy (%)	74
Sensitivity (%)	74
Specificity (%)	75
Balanced Accuracy (%)	74

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

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

NA: Not Available

DA performance vs. Human data	DPRAs 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-bromopropionate	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 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%)

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%)

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 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%)

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% Specificity 73%)

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 ITS 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	Applelde	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-bromopropionate	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 ITSv2 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	Applelde	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-bromopropionate	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