Guidance on information requirements and Chemical Safety Assessment

Chapter R.7a: Endpoint specific guidance

Draft Version 5.0
February 2016
NOTE

Please note that the present document is a proposed amendment to specific extracts only of the Guidance on IR&CSA, Chapter R.7a. This document was prepared by the ECHA Secretariat for the purpose of this consultation and includes only the parts open for the current consultation, i.e. section R.7.3 only.


The numbering and headings of the sub-sections that are displayed in the document for consultation correspond to those used in the currently published guidance document; this will enable the comparison of the draft revised sub-sections with the current text if necessary.

After conclusion of the consultation and before final publication the updated sub-sections will be implemented in the full document.
## Document history

<table>
<thead>
<tr>
<th>Version</th>
<th>Changes</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>[...]</td>
<td>[...]</td>
<td>[...]</td>
</tr>
<tr>
<td>Version 5.0</td>
<td>Full revision addressing the content of Section R.7.3 related to <em>Skin and Respiratory sensitisation</em>. The update includes the following:</td>
<td>XX 2016</td>
</tr>
<tr>
<td></td>
<td>- Modification of Section R.7.3 structure and subdivision by endpoint: Skin sensitisation (Sections R.7.3.2 to R.7.3.7) and Respiratory sensitisation (Sections R.7.3.8 to R.7.3.12).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Update of the information on new/revised EU test methods and OECD test guidelines for skin sensitisation;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Update of the information on respiratory sensitisation;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Update of the information on non-testing methods;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Update of the recommended testing and assessment strategy for skin and respiratory sensitisation in Sections R.7.2.7 and R.7.2.12, respectively;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Replacement of the terms “Integrated Testing Strategy (ITS)” by “testing and assessment strategy” to account for the non-testing part of the evaluation strategy;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Update of the information on Classification and Labelling to reflect changes coming from the 2nd and 4th Adaptations to Technical and Scientific Progress of the CLP Regulation, and to align the text with the revised Section 3.4 Respiratory or skin sensitisation of the <em>Guidance on the Application of the CLP Criteria</em> (version 4.0, November 2013).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Update of quotations from and references to REACH Annex VII, sections 8.3 and 8.3.1 for Skin sensitisation to take into account the revised legal text.</td>
<td></td>
</tr>
</tbody>
</table>
## List of acronyms *(to be added at the beginning of Chapter R.7a before publication)*

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACD</td>
<td>Allergic contact dermatitis</td>
</tr>
<tr>
<td>AOP</td>
<td>Adverse outcome pathway</td>
</tr>
<tr>
<td>Art.</td>
<td>Article</td>
</tr>
<tr>
<td>AUC</td>
<td>Area under the curve</td>
</tr>
<tr>
<td>Cat.</td>
<td>Category</td>
</tr>
<tr>
<td>CCR2</td>
<td>C–C chemokine receptor type 2</td>
</tr>
<tr>
<td>CLP (Regulation)</td>
<td>Regulation (EC) No 1272/2008 on the classification, labelling and packaging of substances and mixtures</td>
</tr>
<tr>
<td>CREM</td>
<td>cAMP responsive element modulator</td>
</tr>
<tr>
<td>DIP</td>
<td>Data interpretation procedure</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DPRA</td>
<td>Direct peptide reactivity assay</td>
</tr>
<tr>
<td>CD</td>
<td>Cluster of differentiation</td>
</tr>
<tr>
<td>DC</td>
<td>Dendritic cell</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
<tr>
<td>DNEL</td>
<td>Derived no effect level</td>
</tr>
<tr>
<td>EC</td>
<td>European commission</td>
</tr>
<tr>
<td>EC3</td>
<td>Estimated concentration three: estimated concentration of a test substance needed to produce a stimulation index (SI) of three.</td>
</tr>
<tr>
<td>ECETOC</td>
<td>European centre for ecotoxicology and toxicology of chemicals</td>
</tr>
<tr>
<td>ECHA</td>
<td>European chemicals agency</td>
</tr>
<tr>
<td>EC JRC</td>
<td>Joint research centre of the European Commission</td>
</tr>
<tr>
<td>EE</td>
<td>Epidermal-equivalent potency assay</td>
</tr>
<tr>
<td>ESR</td>
<td>Endpoint study record</td>
</tr>
<tr>
<td>EU</td>
<td>European union</td>
</tr>
<tr>
<td>EURLECVAM</td>
<td>European union reference laboratory for alternatives to animal testing</td>
</tr>
<tr>
<td>FCA</td>
<td>Freund’s complete adjuvant</td>
</tr>
<tr>
<td>FITC</td>
<td>Fluorescein isothiocyanate</td>
</tr>
<tr>
<td>GARD</td>
<td>Genomic allergen rapid detection</td>
</tr>
<tr>
<td>GCL</td>
<td>Generic concentration limit</td>
</tr>
<tr>
<td>GLP</td>
<td>Good laboratory practice</td>
</tr>
<tr>
<td>GPMT</td>
<td>Guinea pig maximisation test</td>
</tr>
<tr>
<td>h-CLAT</td>
<td>human cell line activation test</td>
</tr>
<tr>
<td>HMT</td>
<td>Human maximisation test</td>
</tr>
<tr>
<td>HPLC</td>
<td>High-performance liquid chromatography</td>
</tr>
<tr>
<td>HRIPT</td>
<td>Human repeat insult patch test</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>IATA</td>
<td>Integrated approach to testing and assessment</td>
</tr>
<tr>
<td>ICCVAM</td>
<td>(US) Interagency coordinating committee on the validation of alternative methods</td>
</tr>
<tr>
<td>Ig</td>
<td>Immunoglobulin</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>IPCS</td>
<td>International programme on chemical safety</td>
</tr>
<tr>
<td>IR&amp;CSA</td>
<td>Information requirements and chemical safety assessment</td>
</tr>
<tr>
<td>ITS</td>
<td>Integrated testing strategy</td>
</tr>
<tr>
<td>IUCLID</td>
<td>International uniform chemical information database</td>
</tr>
<tr>
<td>IVTI</td>
<td><em>in vitro</em> toxicity index</td>
</tr>
<tr>
<td>JaCVAM</td>
<td>Japanese center for the validation of alternative methods</td>
</tr>
<tr>
<td>KE</td>
<td>Key event</td>
</tr>
<tr>
<td>Kow</td>
<td>Octanol-Water partition coefficient</td>
</tr>
<tr>
<td>LLNA</td>
<td>Local lymph node assay</td>
</tr>
<tr>
<td>LLNA: BrdU-ELISA</td>
<td>Local lymph node assay variant based on the measure of 5-bromo-2-deoxyuridine by an enzyme-linked immunosorbent assay method as an indicator of proliferation of lymphocytes</td>
</tr>
<tr>
<td>Log Kow or Log P</td>
<td>Logarithm of the Octanol-Water partition coefficient (Kow)</td>
</tr>
<tr>
<td>MEST</td>
<td>Mouse ear swelling test</td>
</tr>
<tr>
<td>MHC</td>
<td>Major histocompatibility complex</td>
</tr>
<tr>
<td>NICU</td>
<td>Non-immunological contact urticaria</td>
</tr>
<tr>
<td>NOAEL</td>
<td>No observed adverse effect level</td>
</tr>
<tr>
<td>OECD</td>
<td>Organisation for economic co-operation and development</td>
</tr>
<tr>
<td>QMRF</td>
<td>QSAR model reporting format</td>
</tr>
<tr>
<td>QPRF</td>
<td>QSAR prediction reporting format</td>
</tr>
<tr>
<td>QRA</td>
<td>Quantitative risk assessment</td>
</tr>
<tr>
<td>(Q)SAR</td>
<td>(Quantitative) structure activity relationship</td>
</tr>
<tr>
<td>RAAF</td>
<td>Read-across assessment framework</td>
</tr>
<tr>
<td>REACH (Regulation)</td>
<td>Regulation (EC) No 1907/2006 concerning the registration, evaluation, authorisation and restriction of chemicals</td>
</tr>
<tr>
<td>rLLNA</td>
<td>reduced Local lymph node assay</td>
</tr>
<tr>
<td>SCCS</td>
<td>Scientific committee on consumer safety</td>
</tr>
<tr>
<td>SCL</td>
<td>Specific concentration limit</td>
</tr>
<tr>
<td>SI</td>
<td>Stimulation index</td>
</tr>
<tr>
<td>SPSF</td>
<td>Standard project submission form</td>
</tr>
<tr>
<td>TG</td>
<td>Test Guideline</td>
</tr>
<tr>
<td>Th</td>
<td>T-helper lymphocyte</td>
</tr>
<tr>
<td>TM (Regulation)</td>
<td>Test Methods Regulation (EC) No 440/2008</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>TSAR</td>
<td>Tracking system for alternative test methods towards regulatory acceptance</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet (light)</td>
</tr>
<tr>
<td>UVCB</td>
<td>Substance of unknown or variable composition, complex reaction products or biological materials</td>
</tr>
<tr>
<td>WHO</td>
<td>World health organization</td>
</tr>
<tr>
<td>WoE</td>
<td>Weight of evidence</td>
</tr>
</tbody>
</table>
R.7.3 Skin and respiratory sensitisation

R.7.3.1 Introduction

A number of diseases are recognised as being, or presumed to be, allergic in nature. These include asthma, rhinitis, conjunctivitis, allergic contact dermatitis (ACD), urticaria and food allergies (the latter is not discussed in this document). In this Section, the endpoints discussed are those traditionally associated with occupational and consumer exposure to substances. Photosensitisation is potentially important but its mechanism of action is poorly understood, and it is not discussed in this document.

R.7.3.1.1 Definition of skin and respiratory sensitisation

A sensitisers is an agent that will lead to an allergic response in susceptible individuals following skin contact. As a consequence of a secondary - usually organ-specific - subsequent re-exposure, characteristic health effects on the skin (allergic contact dermatitis) or in the respiratory tract may be provoked. After inhalation exposure to a respiratory sensitisers, adverse health effects that may develop include asthma (and related respiratory symptoms such as rhinitis) or extrinsic allergic alveolitis.

Respiratory sensitisers is an agent that will lead to hypersensitivity of the airways following inhalation exposure of that agent. Respiratory sensitisation (or hypersensitivity) is a term that is used to describe asthma and other related respiratory conditions, irrespective of the mechanism (immunological or non-immunological) by which they are caused. In contrast, skin allergy is based on an immunological mechanism. In this sense, it is important to distinguish an ACD e.g. from non-immunological contact urticaria (NICU) or toxic/irritant contact eczema which represent non-immunologically mediated responses of the skin.

When directly considering human data in this document, the clinical diagnostic terms asthma, rhinitis, extrinsic allergic alveolitis and allergic contact dermatitis have been retained.

These definitions are reflected in the criteria for the classification of skin and respiratory sensitisers, which provide a useful tool against which the hazardous properties of a substance can be judged. These criteria are given in the Regulation (EC) No 1272/2008 on the classification, labelling and packaging of substances and mixtures (CLP Regulation).

Classification and labelling under the CLP Regulation

Substances and mixtures causing skin sensitisation and/or respiratory sensitisation can be further characterised by their classification under the CLP Regulation.

Detailed information on the classification and labelling of substances and mixtures can be found in the Guidance on the Application of the CLP criteria and in the CLP Regulation.

a) For skin sensitisation

- Skin sensitisers are classified in Category 1 with the signal word "warning" and the Hazard statement H317 "May cause an allergic skin reaction". Where data are
sufficient, skin sensitisers can be divided into sub-categories. If data are not sufficient for sub-categorisation, Category 1 must be chosen.

- **Sub-category 1A**: Substances showing a high frequency of occurrence in humans and/or a high potency in animals can be presumed to have the potential to produce significant sensitisation in humans. Severity of reaction may also be considered.

- **Sub-category 1B**: Substances showing a low to moderate frequency of occurrence in humans and/or a low to moderate potency in animals can be presumed to have the potential to produce sensitisation in humans. Severity of reaction may also be considered.

**b) For respiratory sensitisation**

- Respiratory sensitisers are classified in Category 1 with the signal word “danger” the Hazard statement H334 "May cause allergy or asthma symptoms or breathing difficulties if inhaled". Where data are sufficient, respiratory sensitisers can be divided into sub-categories. If data are not sufficient for sub-categorisation, Category 1 must be chosen.

  - **Sub-category 1A**: Substances showing a high frequency of occurrence in humans; or a probability of occurrence of a high sensitisation rate in humans based on animal or other tests. Severity of reaction may also be considered.

  - **Sub-category 1B**: Substances showing a low to moderate frequency of occurrence in humans; or a probability of occurrence of a low to moderate sensitisation rate in humans based on animal or other tests. Severity of reaction may also be considered.

**R.7.3.1.2 Objective of the guidance on skin and respiratory sensitisation**

The general objectives of this guidance are:

- to establish whether information from physical/chemical data, from non-testing methods (grouping, QSARs and expert systems), from *in chemico*, *in vitro* or *in vivo* studies or human experience data, provides sufficient evidence that the substance has skin or respiratory sensitisation potential or the lack thereof; or

- to establish whether new information needs to be generated to meet the information requirements under the REACH Regulation by providing a testing and assessment strategy as presented in this document.

Therefore, in the sections on skin sensitisation and respiratory sensitisation, firstly an overview of types of data is given that may provide information on sensitisation, followed

---

1 The testing and assessment strategies are also referred to as Integrated Approaches to Testing and Assessment (IATAs).
by guidance on the process of judging the available data in terms of adequacy, completeness and remaining uncertainty. In Sections R.7.3.6 and R.7.3.11 guidance is given on application of the data to reach a conclusion on suitability for classification and labelling, including potency, if possible. Finally in Sections R.7.3.7 and R.7.3.12 a testing and assessment strategy is presented for skin sensitisation and respiratory sensitisation, respectively.
SKIN SENSITISATION

R.7.3.2 Mechanisms of skin sensitisation

Contact allergens are reactive chemicals (usually organic chemicals or metal ions) of low molecular weight (<500 - 1000 Da) and have a lipophilicity that favours dermal penetration. Skin sensitisation is considered to be a delayed type hypersensitivity (Type IV according to Gell and Coombs) (Karlberg et al., 2008; Martin et al., 2011; Martin, 2014). Some of the mechanisms leading to skin sensitisation (allergic contact dermatitis in humans) are relatively well understood. In 2012 the OECD published an Adverse Outcome Pathway (AOP), which describes the biological mechanisms of skin sensitisation initiated by the covalent binding of substances to skin proteins (OECD, 2012). It should be noted that this AOP does not cover metals or allergens of biological origin, e.g. high molecular weight proteins, but only substances that form a covalent binding to skin proteins. The key events of this skin sensitisation pathway are: 1) covalent binding of the electrophilic substance to skin proteins; 2) release of pro-inflammatory cytokines and induction of cyto-protective pathways in keratinocytes; 3) activation and maturation of dendritic cells, and their migration to the local lymph nodes; 4) presentation of the chemical allergen by the dendritic cells to naïve T-cells, which leads to their differentiation into allergen-specific memory T-cells and their subsequent clonal expansion. Even though not considered as being part of the key events from one to four leading to the adverse outcome, dermal bioavailability (penetration and, if applicable, metabolism) is a prerequisite for a substance to cause skin sensitisation, i.e. the substance needs to reach the viable epidermis in its reactive form.

The mechanisms of metals leading to induction of the innate immune system and, concomitantly, possibly leading to skin sensitisation are not completely understood (Thierse et al., 2005; Martin et al., 2006). Some types of metal ions may act as non-classical haptns, i.e. they do not require stable binding to processed proteins but may directly or indirectly cause structural changes in the major histocompatibility complex (MHC) molecule-peptide complex (by metal-protein complex formation) which then lead to recognition and activation of T-cells via T-cell receptors (Templeton, 2004; Gammerdinger et al., 2003; Lu et al., 2003). Less is known about other types of metal ions such as beryllium (Bowerman et al., 2014). Therefore, skin sensitisation for metals should be evaluated on a case-by-case basis depending on the metal and amount of available information.

Traditionally the development of skin sensitisation has been divided in two phases, i.e. induction and elicitation. In the induction phase the naïve individual becomes sensitised to the allergenic agent, e.g. through the molecular events as described above, leading to the formation of allergen-specific memory T-cells. Those specific memory cells migrate into the dermis and epidermis for the repeated encounter with the specific allergen. In the elicitation phase the memory T-cells, created before in the induction phase, re-encounter the specific allergen which leads to the quick proliferation and activation of those allergen-specific T-cells. The activated cells start secreting specific cytokines, which in turn mobilise other inflammatory cells leading to the clinical outcome of allergic contact dermatitis.
R.7.3.3 Information requirements for skin sensitisation

The information on skin sensitisation that is required to be submitted for registration and evaluation purposes is specified in Annexes VI to XI to the REACH Regulation. According to Annex VI, the registrant should gather and evaluate all existing available information before considering further testing. This includes structural considerations, physico-chemical properties, (Q)SAR ((Quantitative) Structure-Activity Relationship), information from structurally similar substances, in vitro/in chemico data, animal studies, and human data. For classified substances, information on exposure, use and risk management measures should also be collected and evaluated in order to ensure safe use of the substance.

If these data are inadequate for hazard and risk assessment, including classification and labelling, further testing should be carried out in accordance with the requirements of Annex VII (≥1 tpa) to the REACH Regulation.

The standard information requirements at this tonnage level for skin sensitisation (see Section 8.3 in Column 1 of Annex VII) can be fulfilled by following two consecutive steps:

1. an assessment of the available human, animal and alternative data,
2. In vivo testing.

Column 2 of Annex VII lists specific rules according to which the required standard information may be omitted, replaced by other information, or adapted in another way. If the conditions are met under which column 2 of this Annex allows adaptations, the fact and the reasons for each adaptation should be clearly indicated in the registration. For skin sensitisation column 2 reads:

Step 2 does not need to be conducted if:

- the available information indicates that the substance should be classified for skin sensitisation or corrosivity; or
- the substance is a strong acid (pH≤2.0) or base (pH≥11.5); or
- the substance is flammable in air at room temperature (Please note that this rule should actually read: “the substance is spontaneously flammable in air at room temperature”).
- sufficient information is available from non-animal approaches that is adequate for:

---

2 Please note that the information requirements in REACH Annex VII in relation to skin sensitisation are currently under revision. This provisional new text is based on the discussion held at the REACH Committee and the latest revision proposal from the European Commission. This text was added after the end of the PEG consultation and may need to be revised depending on the conclusion of the REACH Committee discussion and a subsequent change of the legal text. Section R.7.3.3 on the information requirements and other relevant parts of R.7.3 will need to be modified and aligned to the revised legal text.
classification, including an assessment whether the substance can be presumed to have the potential to produce significant sensitisation in humans; and

risk assessment, where required.

General provisions for the generation of information on intrinsic properties of substances are contained in REACH Article 13 which states that, in particular for human toxicity, information must be generated whenever possible by means other than vertebrate animal tests, through the use of alternative methods, for example in vitro methods or qualitative or quantitative structure-activity relationship models or from information from structurally related substances (grouping or read across), provided that the conditions specified in Annex XI are met.

In addition to the specific rules of adaptation (column 2), Annex XI 1.2 to 1.5 to the REACH Regulation lays out general rules of adaptation to the standard information requirements. The specific rules for adaptations are: use of non-animal test methods (e.g. in vitro/in chemico) within a Weight-of-Evidence approach (section 1.2), use of (Q)SARs (section 1.3), use of in vitro methods (section 1.4, note these refer to stand-alone methods) or the use of read-across (section 1.5). In the case of Annex XI adaptation as well, this fact and the reasons for each adaptation should be clearly indicated in the registration dossier, i.e. in IUCLID.

Guidance on application of these rules is given in the testing and assessment strategy described in Section R.7.3.7 of this Guidance.

The Murine Local Lymph Node Assay (LLNA) is the first-choice method for in vivo testing. Only in exceptional circumstances should another test be used. This means that in certain cases other in vivo methods may be conducted. In such cases convincing scientific justification for the use of another test must be provided in the registration dossier.

R.7.3.4 Information sources on skin sensitisation

R.7.3.4.1 Non-human data on skin sensitisation

R.7.3.4.1.1 Experimental data available in databases

Registrants need to collect all available information on their substance. It is advised to start by looking at all experimental data for skin sensitisation that may already be available from REACH registration dossiers, or from other sources. ECHA Chem is the primary source for disseminated REACH data. Data can also be found through the OECD QSAR Toolbox or the eChemPortal. Some other tools may also contain data compilations, like TIMES, Ambit, Topkat, Vitic Nexus, and other tools that may provide complete training sets or information for similar compounds. It is very important to assess how the original data have been interpreted in these tools and a consultation with original sources is always recommended, if possible. More details on data sources are given in Appendix R.7.3.1.

R.7.3.4.1.2 Non-testing data on skin sensitisation

The adaptation of standard information requirements can be used if relevant and reliable alternative data can be provided for the substance of interest. As specified in Annex XI to
the REACH regulation, the use of non-testing methods needs to be justified and sufficiently documented. Read-across and (Q)SAR models are non-testing methods that can provide data for skin sensitisation.

6 Read-across

Read-across/chemical categories are described in Sections R.6.1 and R.6.2 of Chapter R.6 of the Guidance on IR&CSA. The scientific basis for building grouping arguments and read-across cases were revisited in the second version of the OECD Guidance Document on grouping of chemicals (OECD, 2014) and in the OECD Guidance Document on the reporting of defined approaches and individual information sources to be used within IATA for skin sensitisation (OECD, 2016a). More detailed advice on the assessment of read across can be found in ECHA’s Read-Across Assessment Framework - RAAF (see http://echa.europa.eu/support/grouping-of-substances-and-read-across). Developing and assessing read-across for skin sensitisation was discussed and exemplified in the work of Patlewicz et al. (2015).

(Q)SAR models

In the case of QSARs and expert systems, registrants need to prepare property predictions by completion of a QSAR Prediction Reporting Format (QPRF). The QPRF is a harmonised template for summarising and reporting substance-specific predictions generated by (Q)SAR models. For filling a data gap under REACH, it is also necessary to provide information on the prediction model employed following a QSAR Model Reporting Format (QMRF) document. The QMRF is a harmonised template for summarising and reporting key information on (Q)SAR model validity, including the results of any validation studies. The information is structured according to the OECD (Q)SAR validation principles (for further information see http://www.oecd.org/env/ehs/risk-assessment/validationofqsarmodels.htm). The JRC QSAR Model Database is an inventory of information on available QMRFs, freely accessible online (https://eurl-ecvam.jrc.ec.europa.eu/databases/jrc-qsar-model-database). More detailed guidance on QSAR models, their use and reporting formats, including the QMRF, is provided in Section R.6.1 of Chapter R.6 of the Guidance on IR&CSA. There is also an initiative started recently (QsarDB) that aims to develop a dynamic repository for QSAR models and datasets for giving access to them and to facilitating predictions from selected literature models that are transparent enough and reproducible.

There are some (Q)SAR models for skin sensitisation reported in the peer-reviewed literature. Available models include local and global (Q)SARs as well as expert systems. If not implemented in a software tool, their use might be restricted due to accessibility issues of technical nature. Exploring the reaction chemistry of compounds forms the basis of most read-across justifications and many of the available skin sensitisation (Q)SARs. According to the OECD AOP for covalent binding to proteins, the skin sensitisation potential of a substance is related in the first place to its ability to react with skin proteins to form covalently linked conjugates and recognition of these by the immune system. In the vast majority of cases, this is dependent on electrophilic reactivity of the substance or a derivative produced by metabolisms or abiotic degradation (Barratt et al., 1997). There are various types of electrophile-nucleophile reactions in skin sensitisation, of which perhaps the most frequently encountered are: Michael-type reactions, $S_N^2$ reactions, $S_N^A$ reactions, acylation reactions and Schiff-base formation. These chemical reaction mechanisms can serve as a means of describing the domain of applicability (the scope) of a (Q)SAR model or form the basis for grouping
substances into chemical categories. Recent work in this area has been described elsewhere (Aptula et al., 2005; Aptula and Roberts, 2006; Roberts et al., 2007, 2011; Schultz et al., 2009; Natsch et al., 2012; Enoch and Roberts, 2013).

Some (Q)SARs models that may be useful for predicting several REACH relevant endpoints, including skin sensitisation, have been included in software packages. A list of available tools was compiled within ANTARES, an EU LIFE project whose results are freely available online (http://www.antares-life.eu/index.php?sec=modellist). The OECD Guidance on grouping of chemicals (OECD, 2014) also provides a summary of tools that might be useful in predicting endpoints of regulatory relevance, including skin sensitisation (see also: http://www.oecd.org/chemicalsafety/risk-assessment/groupingofchemicalschemicalcategoriesandread-across.htm). A non-exhaustive list of available tools is also given in Table R.7.3.1.

More details on data and (Q)SAR models in scientific publications and in in-silico tools are given in Appendix R.7.3.1, while Section R.7.3.5.1 discusses the evaluation of non-testing methods.

**Table R.7.3–1 In silico tools for skin sensitisation prediction.** Note that qualitative models might have been developed using thresholds different from those indicated in REACH and CLP. The user is advised to use as many softwares as possible to gather different pieces of information.

<table>
<thead>
<tr>
<th>Name of the software</th>
<th>Model/module</th>
<th>Model type</th>
<th>Endpoint and/or training set data type</th>
</tr>
</thead>
</table>
| QSAR Toolbox, free [http://www.qsartoolbox.org/](http://www.qsartoolbox.org/) | Protein binding profilers:  
- OASIS v1.3  
- OECD  
- Potency  
- Alerts for skin sensitisation by OASIS v1.3 | Structural alerts | Protein binding |
| "Endpoint" (i.e. Databases):  
- Skin sensitisation  
- Skin sensitisation ECETOC  
- ECHA Chem | Data repositories | Mainly LLNA and GPMT |
| Data gap filling:  
- Read across  
- Trend analysis | Qualitative or quantitative | Mainly LLNA and GMPT |

---

3 E.g. OASIS scale in the QSAR Toolbox: Non-sensitiser (EC3 ≥ 50%), Weak sensitisir (10%≤EC3<50%), Strong sensitisir (EC3 < 10%)
Skin sensitisation model in VEGA: aggregated data from Geberick et al. (2005) (extreme, strong and weak sensitisers coded as positive, NC as negative)

CASE Ultıra model for LLNA: Weak (EC3 < 100%), Moderate (EC3<10%), Strong (EC3<1%).

The outcome of Derek that can be regarded as “positive” alerts are: CERTAIN, PROBABLE, PLAUSIBLE or EQUIVOCAL (in descending order of likelihood), because the query molecule contains a toxicophore that matches one of the alerts that have been coded. EQUIVOCAL is recommended as being a “positive” result in Derek because the alert has still fired for the molecule, but there is an indication that there is another aspect(s) that Derek has considered in its calculation (such as physicochemical properties).

Aggregated data from different sources and tests converted into three categories (Strong, Weak and Non-sensitisers) according to the following scheme: (i) Data from LLNA: extreme, strong and moderate LLNA data converted into “Strong”, weak to “Weak” and non-sensitiser to “Non-sensitiser”; (ii) Data from GPMT: strong and moderate GPMT data converted into “Strong”, weak into “Weak” and non-sensitiser into “Non-sensitiser”; (iii) Data from BfR (Schlede et al., 2003): category A converted into “Strong”, category B into “Weak”, category C into “Non-sensitiser”.

TOPKAT scale (GPMT, % of animals positive): Weak (1-30%), Moderate (30-70%), Strong (70-100%).

<table>
<thead>
<tr>
<th>Tool</th>
<th>Type</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>ToxTree, free</td>
<td>Skin sensitisation reactivity domains</td>
<td>Structural alerts</td>
<td>Reactivity mode of action</td>
</tr>
<tr>
<td>VEGA, free</td>
<td>Skin sensitisation CAESAR</td>
<td>Qualitative⁴</td>
<td>LLNA</td>
</tr>
<tr>
<td>CASE Ultra, commercial</td>
<td>SkinEye Toxicity models</td>
<td>Qualitative⁵</td>
<td>Many (GMPT, LLNA, other)</td>
</tr>
<tr>
<td>Derek Nexus, commercial</td>
<td>Skin sensitisation</td>
<td>Qualitative⁶</td>
<td>GPMT and LLNA</td>
</tr>
<tr>
<td>TIMES, commercial</td>
<td>TIMES Skin Sensitisation</td>
<td>Qualitative⁷ (Metabolic simulator + “local” QSARs)</td>
<td>Mainly GPMT and LLNA</td>
</tr>
<tr>
<td>TOPKAT, commercial</td>
<td>Two models: • Non-sensitisers vs. Sensitisers • Weak/Moderate vs. Strong sensitisers</td>
<td>Qualitative⁸</td>
<td>GPMT</td>
</tr>
</tbody>
</table>

---

⁴ Skin sensitisation model in VEGA: aggregated data from Geberick et al. (2005) (extreme, strong and weak sensitisers coded as positive, NC as negative)

⁵ CASE Ultra model for LLNA: Weak (EC3 < 100%), Moderate (EC3<10%), Strong (EC3<1%).

⁶ The outcome of Derek that can be regarded as “positive” alerts are: CERTAIN, PROBABLE, PLAUSIBLE or EQUIVOCAL (in descending order of likelihood), because the query molecule contains a toxicophore that matches one of the alerts that have been coded. EQUIVOCAL is recommended as being a “positive” result in Derek because the alert has still fired for the molecule, but there is an indication that there is another aspect(s) that Derek has considered in its calculation (such as physicochemical properties).

⁷ Aggregated data from different sources and tests converted into three categories (Strong, Weak and Non-sensitisers) according to the following scheme: (i) Data from LLNA: extreme, strong and moderate LLNA data converted into “Strong”, weak to “Weak” and non-sensitiser to “Non-sensitiser”; (ii) Data from GPMT: strong and moderate GPMT data converted into “Strong”, weak into “Weak” and non-sensitiser into “Non-sensitiser”; (iii) Data from BfR (Schlede et al., 2003): category A converted into “Strong”, category B into “Weak”, category C into “Non-sensitiser”.

⁸ TOPKAT scale (GPMT, % of animals positive): Weak (1-30%), Moderate (30-70%), Strong (70-100%).
<table>
<thead>
<tr>
<th>toxicology.html</th>
</tr>
</thead>
</table>
R.7.3.4.1.3 Testing data on skin sensitisation


Please note that the latest version of an adopted test guideline should always be used when generating new data, independently from whether it is published by EU or OECD.

The testing and assessment strategy developed for skin sensitisation (see Section R.7.3.7 of this Guidance) emphasises the need to evaluate all available information (including structural considerations and physico-chemical properties) before attempting any in vivo testing.

In chemico/in vitro data

Internationally adopted in chemico/in vitro test methods to assess whether a substance is a skin sensitisier (i.e. category 1 under CLP) or not are listed in Table R.7.3–2. In addition this table lists test methods that underwent validation and have been submitted to validation authorities for independent peer-review or are currently in the peer review process. The development of Test Guidelines for some of these methods is under consideration by the OECD. More information on the specific scope and limitations of these tests is provided in Section R.7.3.5.1.2.

In case several EU/OECD adopted test methods are available for a key event, the registrant should select the most appropriate test method available for their substance based on the applicability of the test method.

Table R.7.3–2 Validated and adopted and validated in chemico/in vitro methods for skin sensitisation

<table>
<thead>
<tr>
<th>AOP Key event</th>
<th>Test method</th>
<th>Validation status, regulatory acceptance</th>
<th>EU Test Methods/ OECD test guideline</th>
<th>Classification according to CLP Regulation</th>
<th>EURL ECVAM DB-ALM protocol Nr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skin sensitisation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Key Event 1
Peptide/protein binding

DPRA
Validated and regulatory acceptance

B.59 /TG 442C
Cat. 1 or NC
154

Key Event 2
Keratinocyte

KeratinoSens™
Validated and regulatory

B.60 /TG 442D
Cat. 1 or NC
155

Note: The test methods have each been validated independently, with a limited scope. This means that each has its limitations and cannot be used as a stand-alone test method; however the limitations are assumed to be compensated when used together with additional information within a Weight-of-Evidence approach. Positive results obtained with such methods may be used on their own to classify a substance in Category 1. However, the test methods cannot be used currently on their own to sub-categorise or to predict potency.
### Key Event 3

**Monocytic /Dendritic cell response**

<table>
<thead>
<tr>
<th>response</th>
<th>acceptance</th>
<th>Cat. 1 or NC</th>
<th>N.A.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LuSens</strong>&lt;sup&gt;10&lt;/sup&gt;</td>
<td>Under validation assessment</td>
<td>N.A/N.A</td>
<td>Cat. 1 or NC</td>
</tr>
<tr>
<td><strong>SENS-IS</strong>&lt;sup&gt;11&lt;/sup&gt;</td>
<td>Under validation assessment</td>
<td>N.A/N.A</td>
<td>Cat. 1 or NC</td>
</tr>
<tr>
<td><strong>h-CLAT</strong></td>
<td>Validated and regulatory acceptance</td>
<td>N.A/TG xxx</td>
<td>Cat. 1 or NC</td>
</tr>
<tr>
<td><strong>U-SENS&lt;sup&gt;TM&lt;/sup&gt;</strong>&lt;sup&gt;10&lt;/sup&gt;</td>
<td>Under validation assessment</td>
<td>N.A/N.A</td>
<td>Cat. 1 or NC</td>
</tr>
<tr>
<td><strong>IL-8 Luc Assay</strong>&lt;sup&gt;12&lt;/sup&gt;</td>
<td>Under validation assessment</td>
<td>N.A/N.A</td>
<td>Cat. 1 or NC</td>
</tr>
<tr>
<td><strong>SENS-IS</strong>&lt;sup&gt;11&lt;/sup&gt;</td>
<td>Under validation assessment</td>
<td>N.A/N.A</td>
<td>Cat. 1 or NC</td>
</tr>
</tbody>
</table>

### Key Event 4<sup>13</sup>

**T-cell response**

<table>
<thead>
<tr>
<th>response</th>
<th>acceptance</th>
<th>Cat. 1 or NC</th>
<th>N.A.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N.A</strong></td>
<td>N.A</td>
<td>N.A/N.A</td>
<td>N.A.</td>
</tr>
</tbody>
</table>

**NOTE:** “Validated” means that the test method has gone through a validation process.

**Abbreviations:** N.A. = not available; NC = not classified; DPRA = Direct Peptide Reactivity Assay; h-CLAT = human Cell Line Activation Test; IL = Interleukine; Luc = Luciferase; TG: Test Guideline.

---

<sup>10</sup> The LuSens and the U-SENS<sup>TM</sup> test methods have undergone industry-led validation studies (Ramirez et al., 2016; Alépée et al., 2015). The information generated in the validation studies has been submitted to EURL ECVAM and is currently under evaluation. A standard project submission form (SPSF) for a Test Guideline concerning U-SENS<sup>TM</sup> was submitted to the OECD in 2015. The project has been included in the OECD work programme.

<sup>11</sup> The SENS-IS test method underwent an industry lead validation and has been submitted to EURL ECVAM (Cottrez et al., 2016). An SPSF for the development of a Test Guideline was submitted to the OECD in 2015.

<sup>12</sup> The IL-8 Luc Assay underwent a validation study coordinated by JaCVAM (Kimura et al., 2015). The test method is currently under peer-review. A SPSF for the development of a Test Guideline was submitted to the OECD in 2014. The project has been included in the OECD work programme.

<sup>13</sup> It is important to note that there are currently no validated or adopted *in vitro* test methods to address Key Event 4 (human T-cell activation) of the AOP. However, this is a key step in ACD, reflecting the adaptive immune response, similarly to the LLNA, even though the LLNA does not measure antigen specific T-cell responses *per se* but only cell proliferation (Dietz et al., 2010; Richter et al., 2013).
The test methods indicated in Table R.7.3–2 are either in 

chemico assay(s) (DPRA), or cell-based assays (Keratinosens™, LuSens, h-CLAT, U-SENS™, IL-8 Luc Assay). These test methods were developed to address specific events of the skin sensitisation AOP (OECD, 2012). The AOP for skin sensitisation describes the current understanding of key events linked to skin sensitisation. As each of the test methods only addresses a specific key event of skin sensitisation, they should not be used in isolation to identify a skin sensitisier or a non-sensitisier. More information on how these test methods can be used in the REACH context can be found in Section R.7.3.7.2 of this Guidance.

It is important to note that currently several non-animal test methods are under development or evaluation, and therefore their regulatory use and predictive value has not been assessed yet. It is it is recommended to monitor the status of current developments through the scientific literature. Test methods under evaluation by EURL ECVAM or other international validation bodies can be monitored through the EURL ECVAM Test Method Submission webpage (https://eurl-ecvam.jrc.ec.europa.eu/test-submission/).

Animal data

- Guideline-compliant tests

For new in vivo testing of skin sensitisation potential, the murine local lymph node assay (LLNA) is the REACH Annex VII-endorsed method. This assay has been validated internationally and has been shown to have clear animal welfare benefits and scientific advantages compared with the guinea pig tests described below. The LLNA is designed to detect the potential of substances to induce sensitisation as a function of lymphocyte proliferative responses induced in regional lymph nodes (induction phase). This method is described in EU B.42/OECD TG 429, which contain the standard LLNA method and the rLLNA (in the rLLNA only one dose is used, which does not allow the assessment of potency). In addition, there are different variants of the LLNA adopted by the EU and OECD, i.e. EU B.50/OECD TG 442A (Local Lymph Node Assay: DA) and EU B.51/OECD TG 442B (Local Lymph Node Assay: BrdU-ELISA). The main difference compared to the OECD TG 429 is that these test methods do not use radioactive labelling.

Two further animal test methods for skin sensitisation are described in EU B.6/OECD TG 406: the guinea pig maximisation test (GPMT) and the Buehler test. The GPMT is an adjuvant-type test in which the acquisition of sensitisation is potentiated by the use of Freund’s Complete Adjuvant (FCA) and in which both intradermal and topical exposure are used during the induction phase. The Buehler test is a non-adjuvant method involving for the induction phase topical application only. Both test methods assess the elicitation phase, i.e. the adverse outcome of skin sensitisation.

Both the GPMT and the Buehler tests are able to detect substances with moderate to strong sensitisation potential, including those with relatively weak sensitisation potential. In such methods the activity is measured as a function of challenge-induced dermal hypersensitivity reactions elicited in test animals compared with controls. Since the LLNA is the preferred method for new in vivo testing, the use of the standard guinea pig tests to obtain new data on the skin sensitisation potential of a substance will be acceptable only in exceptional circumstances and will require scientific justification. However, existing data of good quality derived from such tests are acceptable and, if they provide clear results, will preclude the need for further in vivo testing.

ECETOC Monograph 29 (2000) contains a useful discussion of these tests.

- Non-guideline compliant tests and refinements to the standard assays
Existing data may be available from tests that do not have an OECD guideline, for example:

i. other guinea pig skin sensitisation test methods (such as the Draize test, optimisation test, split adjuvant test, open epicutaneous test);

ii. additional tests (such as the mouse ear swelling test, local lymph node cell count method (Basketter et al., 2012)).

Information may also be available from other endpoints, for example, repeated dose dermal studies that show effects indicative of an allergic response, such as persistent erythema and/or oedema. In this case, care must be taken to distinguish allergenic effects from irritancy or non-immunological effects (such as non-immunological contact urticaria).

R.7.3.4.2 Human data on skin sensitisation

Human data on cutaneous (allergic contact dermatitis and urticarial) reactions may come from a variety of sources:

- consumer experience and comments, preferably followed up by medical examinations (e.g. diagnostic patch tests);
- diagnostic clinical studies (e.g. patch tests, repeated open application tests);
- records of workers’ experience, accidents, and exposure studies including medical surveillance;
- case reports in the general scientific and medical literature;
- consumer tests (monitoring by questionnaire and/or medical surveillance);
- epidemiological studies;
- existing human experimental studies such as the human repeat insult patch test (HRIPT) (Stotts, 1980; McNamee et al., 2008) and the human maximisation test (HMT) (Kligman, 1966), although it should be noted that new experimental testing in humans for hazard identification, including HRIPT and HMT, is not acceptable.

R.7.3.5 Evaluation of available information on skin sensitisation

For hazard identification and potency assessment, it is important that the data provided are reliable and relevant. Conclusion on the usefulness of data may rely on one data point or on a Weight-of-Evidence approach, as described in Section R.7.3.7.2 (Section “How to perform and report Weight-of-Evidence analysis”) and in REACH Annex XI Section 1.2. Such a Weight-of-Evidence approach also includes an evaluation of the available data as a whole, i.e. both over or across endpoints, starting with careful evaluation of sensitisation data. However, information obtained from skin corrosion/irritation and/or dermal toxicity test(s) may provide additional useful information for the Weight-of-Evidence approach. For example, skin corrosion data may enable a specific adaptation (column 2, section 8.3 of REACH Regulation) and may help in distinguishing irritation effects from sensitisation.

When a non-animal testing approach, as described in the Testing and Assessment Strategy in Section R.7.3.7 is followed in order to meet the information requirement,
weight should be given to the validated and/or adopted *in chemico/in vitro* methods described below.

The *Weight-of-Evidence* approach provides a basis to decide whether further information is needed on endpoints for which existing data appear inadequate or data are not available, or whether the requirements are fulfilled.

In the following sections some additional remarks are made on the adequacy of the various types of data that may be available.

### R.7.3.5.1 Non-human data on skin sensitisation

#### R.7.3.5.1.1 Non-testing data on skin sensitisation

**Read-across**

The use of read-across requires the use of data from one or more source substance(s) for assessing the hazardous properties of a target substance. The read-across assessment framework (RAAF) document, which is published on the ECHA website ([http://echa.europa.eu/en/support/grouping-of-substances-and-read-across](http://echa.europa.eu/en/support/grouping-of-substances-and-read-across)), describes ECHA’s assessment of the suitability of a read-across. Among other things, the RAAF emphasises the importance of transformation processes (metabolic or abiotic) for the potential activation of substances (in the case of sensitisation: pre- and pro-haptenation). This is only one of the factors that determine the selection of scenario for assessment; some other general considerations for assessing the suitability of selected analogues are:

- the same endpoint is considered;
- there are any additional functional groups or additional substituents that might influence the reactivity and sensitising behaviour (applicability domain considerations);
- the physico-chemical parameters are similar (e.g. LogP, applicability domain considerations);
- there are impurities that influence the sensitisation profile;
- the likely chemical mechanism is the same.

In the case of skin sensitisation, the most robust means of comparing two or more substances is through an evaluation of their likely chemical reactivity. Work in this area has investigated means of encoding reactivity for the different chemical reaction type domains in form of rules (Aptula and Roberts, 2006; Aptula *et al.*, 2006; Schultz *et al.*, 2009; Roberts and Aptula, 2014)\(^\text{14}\). If the chemical reactivity is not known, or cannot be determined through experimentation, then a pragmatic means of identifying similar substances can be through a substructural/analogue search. In this context the RAAF requires on the one hand to specify why the commonalities between two or more analogue structures suggest similar biological action; on the other hand

\(^{14}\) This approach might involve the systematic generation of *in vitro* reactivity data for these different mechanistic domains.
justification needs to be provided as to why structural dissimilarities are not expected
to result in dissimilar biological action or quantitative differences in potency.

(Q)SAR models

When evaluating the reliability of (Q)SAR predictions, the assessment depends on both
the substances of interest and the (Q)SAR model(s) used to make a prediction. General
advice on (Q)SARs including an evaluation of OECD principles for QSAR validation is
described in Section R.6.1.3 of Chapter R.6 of the Guidance on IR&CSA. Clearly there is a
breadth of different (Q)SARs and expert systems available for the estimation of skin
sensitisation hazard. The approaches are quite varied and each has been developed on
different sets of in vivo data (principally GPMT and LLNA). Whilst efforts have been made
to characterise a number of the literature-based models in terms of the OECD principles
for QSAR validation (see Roberts et al., 2007 as an example), further work is still
required for some of the commercial systems (ECETOC, 2003). In addition, in many
cases these models have been demonstrated to be reasonable for predicting skin
sensitisers correctly but they are limited in predicting non-sensitisers correctly (Roberts
et al., 2007; ECETOC, 2003). For this reason, careful interpretation of model predictions
needs to be considered in light of other pieces of information, e.g. analogue read-across
(other similar substances with respect to their mechanistic domain). A good practice is to
use the results from these models as building blocks of weight of evidence (e.g. the
prediction from a reactivity-based model addressing the molecular initiating event of the
AOP could be used the same way as the DPRA, if the prediction is considered reliable).

Further work is needed to encode more knowledge/rules for non-reactive substances as
well as for those substances likely to undergo chemical or metabolic transformation.

Consideration of which model(s) to apply will be dependent on the specific substance(s)
of interest, the underlying training dataset and the applicability domain of the model(s).
These issues are described more fully in Section R.6.1 of Chapter R.6 of the Guidance on
IR&CSA. An example is illustrated here: if the substance falls into a chemistry reactivity
domain that is well characterised, then a local (Q)SAR model developed for this domain
(such as those previously described) may give rise to the most robust prediction of skin
sensitisation. Where the mechanism is not understood or not known a priori one or more
of the expert systems such as TOPKAT, Derek Nexus or the others already described will
be best placed to provide an estimate. These systems may not be fully transparent but
they often provide a reasonable amount of supporting information to enable the
robustness of a prediction to be evaluated.

The prediction needs to be evaluated by taking into account the likely chemical reactivity
and the presence of similar substances within the training set of the model. This type of
information is needed to assess whether the prediction derived is meaningful and
relevant. For global models available in the literature, the training set(s) and the
algorithm(s) are usually available to allow such comparisons to be made.

The QMRF and QPRF were developed to provide templates for including specific model
and prediction information. More details are provided in Section R.6.1 of Chapter R.6 of
the Guidance on IR&CSA.

Other information

Other information such as results in other assays, e.g. the Ames test (a common feature
of genotoxic substances is that they can bind covalently to DNA and cause direct DNA
damage) or aquatic toxicity tests, may provide supporting information about the
electrophilicity of the substance of interest and hence its likely sensitisation ability. It is
notable that in vitro genotoxicity assays do not always address molecular (DNA) binding.
Also, abiotic transformation of the substance in aquatic toxicity test may lead to
differences in the availability of the actually active substance. Some of this work explores
correlations between aquatic toxicants and skin sensitisers (Aptula et al., 2006) and
between experimentally identified mutagens and sensitisers (Wolfreys and Baskettet
2004; Patlewicz et al., 2014). More recently, the use of mutagenicity data was proposed
as part of an integrated approach to testing and assessment (IATA) for skin sensitisation
(Patlewicz et al., 2014). More work is however needed to allow for a sound scientific
evaluation of this approach.

R.7.3.5.1.2 Testing data on skin sensitisation

In chemico/in vitro data

There are several validated test methods for the assessment of skin sensitisation
potential in chemico/in vitro and, for some of them, EU/OECD-adopted test guidelines
are available (see Section R.7.3.4.1). These test methods have not been developed as
stand-alone test methods, but as test methods to be used together with other pieces of
information in a Weight-of-Evidence approach, e.g. by using several in chemico/in vitro
methods together.

Annex VII to the REACH Regulation specifies that the standard information requirement
for skin sensitisation is an in vivo study. However the REACH Regulation gives several
options for adapting this standard information requirement, such as the specific rules for
adaptation in column 2 of Annex VII, section 8.3 or the general rules for adaptations in
Annex XI. As a consequence, data from the tests described below may be accepted to
fulfil Annex VII requirement when used in combination with each other and/or with
other pieces of information (non–validated in vitro, in silico, (Q)SARs, etc.), e.g. within a
Weight-of-Evidence approach according to Annex XI, sections 1.2 (see Section R.7.3.7).
Although this is not a preferred option, positive results obtained with such methods (in
vitro/in chemico) may in principle be used on their own to meet the relevant information
requirement by applying Annex XI, section 1.4. However, additional information, e.g.
from the assessment of (an)other key event(s), can provide more confidence for the
endpoint. The in chemico/in vitro test results should normally be used in combination in
the context of Weight-of-Evidence. In practice, the acceptance depends on the validity of
the test methods and on the suitability of a specific test method for the substance, i.e.
the test substance should fall under the scope of the test method.

In order to fulfil REACH information requirements, the data from non-animal methods
must allow the detection of substances that can be presumed to have the potential to
produce significant (strong/extreme) sensitisation in humans.

The test methods described below are not currently suitable on their own for sub-
categorisation of skin sensitisers into CLP sub-categories 1A and 1B; these sub-
categories are based on skin sensitising potency. Indicators of potency such as the level
of peptide depletion and concentration-responses can be obtained from the existing in
chemico and in vitro tests, respectively. While there is no widely accepted prediction
model (based on a single test) or data interpretation procedure (based on multiple
tests/sources of information) to integrate these indicators into an adequate potency
classification, needed for classification and setting of Specific concentration limits (SCLs),
some approaches have been proposed in the scientific literature. A review of different
approaches is provided in Appendix R.7.3-4. Therefore, the registrant is advised to
carefully follow the recent developments in this area. If non-animal testing methods will
be considered for the purpose of fulfilling the REACH information requirements, the
registrant will need to consider the skin sensitisation potency of the substance by all
means available. However, at this point of time (at the time of publication of this Guidance) no firm guidance can be provided.

It is important to note that, currently, several non-animal test methods are under development or evaluation, however their regulatory use and value have not been assessed yet. It is advised to monitor the status of current developments through the scientific literature (e.g. provide reference to most recent scientific reviews in the area).

Test methods under evaluation by EURL ECVAM or other international validation bodies can be monitored through the EURL ECVAM Test Method Submission webpage (https://eur-ecvam.jrc.ec.europa.eu/test-submission/).

- **Direct Peptide Reactivity Assay (DPRA) - OECD TG 442C**

The DPRA aims to provide information on the molecular initiation event of skin sensitisation i.e. protein binding of low molecular weight substances using synthetic heptapeptides containing cysteine and lysine amino acids. In the assay peptide reactivity is addressed by measuring the depletion of the synthetic heptapeptides by HPLC using UV detection. However, when considering this limitation, it should be kept in mind that the relative percentage of substances reacting preferably with amino acids other than cysteine and lysine is at present unclear and that the cysteine and lysine peptides represent softer to harder model nucleophiles which would cover different reaction mechanisms. More information can be obtained from the EURL ECVAM Recommendation (available at: https://eur-ecvam.jrc.ec.europa.eu/eurl-ecvam-recommendation-on-the-direct-peptide-reactivity-assay-dpra).

The specific limitations of the test method according to the current test guideline are:

- It is only applicable to test substances that are soluble in an appropriate solvent at a final concentration of 100 mM. Substances that are not soluble at this final concentration can still be tested at lower soluble concentrations. In such a case, positive results could still be used to identify a test substance as a sensitiser whereas negative results obtained with concentrations < 100 mM should be considered inconclusive;

- Co-elution (i.e. the substance and the peptide elute at the same time) may hamper the determination of peptide reactivity, therefore appropriate co-elution controls need to be included in the test design;

- It is not applicable to the testing of metals and metal compounds (known to react with proteins with mechanisms other than covalent binding), and to complex mixtures of unknown composition, substances of unknown or variable composition, complex reaction products or biological materials (i.e. UVCB substances) due to the defined molar ratio of the test substance and peptide;

- The test system has no metabolic capacity, therefore pro-haptens (i.e. substances requiring metabolic activation to exert their sensitising activity) may produce false negative results. Pre-haptens (i.e. substances activated by abiotic transformation, e.g. auto-oxidation or hydrolysis) may produce false negative results, especially in case of slow oxidizers. However, identification of slow oxidizers would also fail by using in vivo methods (REF to EURL ECVAM report once published);

- Test substances with exclusive reactivity towards amino-acids other than cysteine or lysine (e.g. nucleophilic sites of histidine) may lead to false negative results;
- Potential over-predictions may be due to substances that do not covalently bind to the peptide but do promote its oxidation (e.g. cysteine dimerisation).

• ARE-Nrf2 Luciferase Test Method (KeratinoSens™) - OECD TG 442D

The Keap1-Nrf2-ARE pathway is considered to be a major regulator of cyto-protective responses to electrophile and oxidative stress by controlling the expression of detoxification, antioxidant and stress response enzymes and proteins. In the assay, induction of the luciferase gene is measured as an indicator of the activity of the pathway. As the majority of substances causing skin sensitisation are electrophiles reacting with nucleophilic centres in skin proteins, this pathway is relevant for skin sensitisation. However, the Keap1-Nrf2-ARE signalling pathway is not only related to keratinocytes but is also detectable in other cell types, and in addition it may also be affected by non-electrophilic modulators (e.g. corrosive/irritating substances) and may hence produce false positive responses (Richardson et al., 2015). More information can be obtained from the EURL ECVAM Recommendation (available at: https://eurl-ecvam.jrc.ec.europa.eu/eurl-ecvam-recommendations/recommendation-keratinosens-skin-sensitisation).

The specific scope and limitations of the test method according to the current test guideline are:

- It is applicable to test substances that are soluble or that form a stable dispersion either in water or DMSO, or another appropriate solvent if its choice is scientifically justified. Test substances that do not fulfil these conditions at the highest final required concentration of 2000 μM may still be tested at lower concentrations. In such a case, positive results could be used to identify a test substance as sensitiser whereas negative results obtained with concentrations < 1000 μM should be considered inconclusive;

- The test system using the human keratinocyte cell line HaCaT has a limited metabolic capacity, therefore pro-haptens (i.e. substances requiring metabolic activation to exert their sensitising activity) may produce false negative results. Pre-haptens (i.e. substances activated by abiotic transformation, e.g. auto-oxidation or hydrolysis) may also produce false negative results, especially in case of slow oxidizers. However, identification of slow oxidizers would also fail by using in vivo methods (REF to EULR ECVAM report once published);

- Test substances with exclusive reactivity towards nucleophiles other than cysteine’s sulfhydryl group (e.g. lysine residues) can produce false negative results in the assay;

- Test substances that do not act as sensitisers but are nevertheless chemical stressors may produce false positive results;

- Highly cytotoxic substances cannot always be reliably assessed;

- Test substances that interfere with the luciferase enzyme can affect its activity by either increasing or inhibiting the luminescence.

• Human Cell Line Activation Test (h-CLAT) - draft OECD TG available (see http://www.oecd.org/env/ehs/testing/section4healtheffects.htm)
The h-CLAT assay aims to provide information on dendritic cell (DC) activation by using a human monocytic leukemia cell line (THP-1) as an alternative model to DCs. The DC activation is measured by analysing the expression (upregulation) of specific cell surface markers known to be linked to DC maturation, i.e. CD86 and CD54, by using flow cytometry. Monocytic human THP1 cells used in this assay may give different signals of the same cellular molecules after stimulation with a specific substance compared to human dendritic cells (Lehtonen et al., 2007).


The specific scope and limitations of the test method according to the current draft test guideline are:

- It is applicable to test substances that are soluble or form a stable dispersion in an appropriate solvent;
- Test substances with Log Kow ≤ 3.5 can be tested whereas substances with Log Kow > 3.5 tend to produce a higher rate of false negative results. For such substances with Log Kow > 3.5 positive results could be used to support the identification of a test substance as a sensitisier. Negative results for substances with Log Kow > 3.5 should be considered inconclusive.
- The test system has a limited metabolic capacity, therefore pro-haptens (i.e. substances requiring metabolic activation to exert their sensitising activity) may produce false negative results. Pre-haptens (i.e. substances activated by abiotic transformation e.g. auto-oxidation or hydrolysis) may also produce false negative results, especially in case of slow oxidizers. However, identification of slow oxidizers would also fail by using in vivo methods (REF to EULR ECVAM report once published);
- Highly cytotoxic substances cannot always be reliably assessed;
- Since it uses a fluorescein isothiocyanate (FITC)-labelled antibody, strong fluorescent test substances emitting at the same wavelength as FITC may interfere with the flow cytometry light-signal acquisition.

Concerning the in chemico/in vitro test methods, any modification made to the adopted test guidelines is not recommended and should only be done in exceptional circumstances and needs to be properly documented and scientifically justified and shown to yield comparable results using the proficiency chemicals listed in the EU/OECD test guideline. The reporting template in Annex II of the OECD Guidance Document on the reporting of defined approaches to be used within IATA (OECD, 2016b) can be used for this purpose and also to document in chemico/in vitro methods for which no adopted test guideline is available (see Appendix R.7.3-2). Proper documentation and justification is needed when the information submitted has been generated by using test methods that have not been formally validated and do not have adopted test guidelines.

Animal data

Well reported studies using internationally acceptable protocols, particularly if conducted in accordance with the principles of GLP, can be used for hazard identification. Other studies (see Section R.7.3.4.1 and below), not fully equivalent to OECD test protocols, can, in some circumstances, provide useful information. Particular attention should be paid to the quality of these tests and the use of appropriate positive and negative
controls. The specificity and sensitivity of all animal tests should be monitored through the inclusion of appropriate positive and negative controls. In this context, positive controls are the 6-monthly sensitivity checks with an appropriate positive control substance, and negative controls are the vehicle-treated control animals included as part of each test.

- Guideline-compliant tests

**Murine Local Lymph Node Assay**

For the conduct and interpretation of the LLNA the following points should be considered:

i. the vehicle in which the test material and controls have been applied;

ii. the concentrations of test material that have been used;

iii. any evidence for local or systemic toxicity, or skin inflammation resulting from application of the test material;

iv. whether the data are consistent with a biological dose-response;

v. the submitting laboratory should be able to demonstrate its competency to conduct the LLNA.

EU B.42/OECD TG 429 provides guidance on the recommended vehicles, number of animals per group, concentrations of test substance to be applied and substances to be used as positive control. A preliminary study or evaluation of existing acute toxicity/dermal irritation data is normally conducted to determine the highest concentration of test substance that is soluble in the vehicle but does not cause unacceptable local or systemic toxicity. The submission of historical control data will demonstrate the ability of the test laboratory to produce consistent responses. Based on the incorporation of radioactive labelling (tritiated (3H)-methyl thymidine), substances that result in a stimulation index (SI) $\geq 3$ at one or more test concentrations are considered to be positive for skin sensitisation. Both positive and negative responses in the LLNA conducted as described in EU B.42/OECD TG 429 meet the data requirements for classification of a substance as a skin sensitisier including potency estimations: no further testing is required.

Alternative vehicles to those listed in EU B.42/OECD TG 429 may be used in the LLNA if sufficient scientific justification is provided. There can be some variability due to the choice of vehicle (vehicle effect) that may enhance or suppress the response in the LLNA (see Section R.7.3.6.1).

The LLNA: DA test method described in EU B.50/OECD TG 442A measures ATP content by luminescence in the proliferating cells and hence does not require the use of radioactive labelling of cells. Substances that result in SI $\geq 1.8$ at one or more testing concentration(s) are considered to be positive for skin sensitisation. In case of borderline positive results ($1.8 \leq SI \leq 2.5$), linked to the sensitivity of the detection method, additional information may be considered such as the dose-response relationship, evidence of systemic toxicity or excessive irritation, and, where appropriate, statistical significance together with SI values to confirm that such results are indeed positives. Currently, there are no criteria available for predicting skin sensitisation potency with this test method.

The LLNA: BrdU-ELISA test method described in EU B.51/OECD TG 442B uses the non-radiolabelled marker 5-bromo-2-deoxyuridine (BrdU) to measure lymphocyte
proliferation. Substances that result in SI ≥1.6 at one or more testing concentration(s) are considered to be positive for skin sensitisation. In case of borderline positive results (1.6 ≤ SI ≤ 1.9), linked to the sensitivity of the detection method, additional information may be considered such as the dose-response relationship, evidence of systemic toxicity or excessive irritation, and, where appropriate, statistical significance together with SI values to confirm that such results are indeed positives. Currently, there are no criteria available for predicting skin sensitisation potency with this test method.

The EU B.50/OECD TG 442A (LLNA: DA) and EU B.51/OECD TG 442B (LLNA: BrdU-ELISA) recommend the use of the same vehicles as in the standard LLNA EU B.42/OECD TG 429.

Limitations of all the above LLNA variants include the following:

- False negative predictions can be obtained with certain metals (e.g. nickel, Schmidt and Goebler, 2015) and false positive predictions may be obtained with certain surfactant type substances (Kreiling et al., 2008; Garcia et al., 2010; Ball et al., 2011) or siloxanes (Petry et al, 2012).
- Low solubility of the substance may interfere with the accuracy of the predictions.
- The choice of vehicle may affect the prediction for certain substances. For instance DMSO as a polar solvent may enhance dermal bioavailability of some test substances and propylene glycol may suppress the proliferative effects of some test substances (e.g. DNCB) (Anderson et al., 2011). Therefore, it is important to properly select the vehicle used in the study.

The updated OECD TG 429 of 2010 includes the reduced LLNA (rLLNA), in which only one concentration is tested and less animals are used. It is recommended to use this refinement method only in case a confirmation of a negative result obtained with another testing method is required. Since only one dose is used in the study design, the rLLNA cannot currently be used for estimating the skin sensitisation potency of a substance (Ezendam et al., 2013), even though a proposal has recently been published for predicting potency from a single dose (Roberts, 2015). The TGs for the LLNA variants, i.e. DA and BrdU-ELISA test methods, do not include the use of the rLLNA study design. In addition, none of these variants have been assessed for their usefulness in potency prediction.

**Guinea pig studies**

New guinea pig studies should only be conducted in exceptional circumstances. In such cases a justification for using a test method other than the LLNA must be provided in the IUCLID dossier (Annex VII, section 8.3, column 2 to the REACH Regulation).

The guinea pig test method described in EU B.6/OECD TG 406, the GPMT (Magnusson et al., 1969; Schlede and Eppler, 1995) and the Buehler test do also provide suitable information for hazard identification. Recommendations on conducting and analysing these methods are provided by Steiling et al. (2001). Particular attention should be paid to the quality of these tests with consideration given to the following points:

i. numbers of test and control guinea pigs;

ii. number or percentage of test and control animals displaying skin reactions;

iii. whether skin irritation was observed at the induction phase;
iv. whether the maximal non-irritating concentration was used in the challenge phase;

v. the choice of an appropriate vehicle (ideally, one that solubilises or gives a stable suspension or emulsion of the test material, is free of allergenic potential, is non-irritating, enhances delivery across the stratum corneum, and is relevant to the usage conditions of the test material, although it is recognised that it will not always be possible to meet all these conditions);

vi. whether there are signs of systemic toxicity (a sighting study should be performed to determine an appropriate induction dose that causes irritation but not systemic toxicity);

vii. staining of the skin by the test material that may obscure any skin reactions (other procedures, such as chemical depilation of the reaction site, histopathological examination or the measurement of skin fold thickness may be carried out in such cases);

viii. results of rechallenge treatments if performed;

ix. checking of strain sensitivity at regular intervals by using an appropriate control substance (as specified in OECD guidelines and EU Test Methods). Currently (at the time of publication of this Guidance), the recommended interval is 6 months.

The investigation of doubtful reactions in guinea pig tests, particularly those associated with evidence of skin irritation following a first challenge, may benefit from rechallenge of the test animals. In cases where reactions may have been masked by staining of the skin, other reliable procedures may be used to assist with interpretation; where such methods are used, the submitting laboratory should provide evidence of their value.

A justification for performing a new guinea pig test instead of an LLNA could be for example that the test substance contains nickel, as it is known that nickel is not correctly predicted in the LLNA.

There are criteria available for predicting skin sensitisation potency based on guinea pig tests. However, due to the test design, it is often not possible to exclude the possibility that a substance identified as skin sensitisier could be a strong/extreme (i.e. Cat. 1A) skin sensitisier.

- **Non-guideline compliant tests and refinements to the standard assays**

The submitted dossier should include scientific justification for conducting any new test that is a modification or deviation from guideline methods. In such cases, it would be advisable to seek appropriate expert advice on the suitability of the assay before testing is begun.

Historically, guinea pig studies that are not fully equivalent to OECD test protocols have been conducted and can provide useful hazard information. These studies include, but are not limited to, the following: Draize test, optimisation test, split adjuvant test, open epicutaneous test and the cumulative contact enhancement test. In the case of positive results the substance may be considered as a potential skin sensitisier. If, taking into account the above quality criteria, especially the positive and negative control data, there is a clear negative result, i.e. no animals displaying any signs of sensitisation reactions, then no further animal testing is required. Where there is a low level of response, the quality of the study is questionable, or where unacceptably low concentrations of the test...
material have been used for induction and/or challenge, further testing may be required. In addition, existing information may already be available e.g. from mouse ear swelling test (MEST), which is a modification of the LLNA. The MEST has been evaluated in inter-laboratory studies, and it was concluded that the MEST could be used for identifying strong skin sensitisers (Dunn et al., 1990: EC, 2004).

R.7.3.5.2 Human data on skin sensitisation

When reliable and relevant human data are available, they can be useful for hazard identification and even preferable over animal data. However, a lack of positive findings in humans does not necessarily overrule positive and good quality animal data.

Well conducted human studies can provide very valuable information on skin sensitisation. However, in some instances (due to lack of information on exposure, a small number of subjects, concomitant exposure to other substances, local or regional differences in patient referral, etc.) there may be a significant level of uncertainty associated with human data. Moreover, diagnostic tests are carried out to see if an individual is sensitised to a specific agent, and not to determine whether the agent can cause sensitisation. Evidence of skin sensitising activity derived from diagnostic testing may reflect the induction of skin sensitisation to the substance tested or cross-reaction with a chemically very similar substance. In both situations, the normal conclusion would be that this provides positive evidence of the skin sensitising activity of the substance used in the diagnostic test.

For evaluation purposes, existing human experience data for skin sensitisation should contain sufficient information about:

- the test protocol used (study design, controls);
- the substance or preparation studied (should be the main, and ideally, the only substance or preparation present which may possess the hazard under investigation);
- the extent of exposure (concentration per square centimeter, frequency and duration);
- the frequency of effects (versus number of persons exposed);
- the persistence or absence of health effects (objective description and evaluation);
- the presence of interfering factors (e.g. pre-existing dermal health effects, medication, presence of other skin sensitisers);
- the relevance with respect to the group size, statistics, documentation;
- the healthy worker effect\textsuperscript{15}.

\textsuperscript{15} Phenomenon observed initially in studies of occupational diseases: workers usually exhibit lower overall death rates than the general population because severely ill and disabled people are excluded from employment.
Human experimental studies on skin sensitisation are not normally conducted and should be avoided. Where human data are available, quality criteria and ethical considerations as presented in ECETOC monograph no 32 (ECETOC, 2002) should be taken into account.

Ultimately, where a very large number of individuals (e.g. $10^5$) have frequent (daily) skin exposure for at least two years and there is an active system in place to pick up complaints and adverse reaction reports (including via dermatology clinics), and where no or only a very few isolated cases of allergic contact dermatitis are observed, then the substance is unlikely to be a significant skin sensitisier. However, information from other sources should also be considered in making a judgement on the substance's ability to induce skin sensitisation.

It is emphasised that testing with human volunteers is strongly discouraged, but when there are good quality data already available they should be used as appropriate in well justified cases.

**R.7.3.6 Conclusions on skin sensitisation**

**R.7.3.6.1 Remaining uncertainty on skin sensitisation**

Data that comply with REACH standard information requirements can be generated from well designed and well conducted studies in animals. However, it should be noted that no toxicological test is perfect and each test method has to balance between the sensitivity (rate of true positives) and specificity (rate of true negatives) of the prediction of the effect in the human population. The use of adjuvant in the GPMT may lower the threshold for irritation and so lead to false positive reactions, which can therefore complicate interpretation (running a pre-test with FCA treated animals can provide helpful information). In international trials, the LLNA has been shown to predict >80% of skin sensitisers when compared to human data but, like the guinea pig tests, it is dependent on the vehicle used. Variability due to the choice of vehicle (vehicle effect) may enhance or suppress the response in the LLNA by one order of magnitude (Anderson et al., 2011; Hoffmann, 2015; Wright et al. 2010), therefore there may be some inherent uncertainty associated to the vehicle selection. It has been claimed that the LLNA may create false positives for (irritating) surfactants (non-specific lymphocyte proliferation) (Garcia et al., 2010; Kreiling et al., 2008). However, Basketter and Kimber (2011) state that, if the study is performed according to the dose selection criterion as specified in the OECD TG 429, no false positives results should be obtained based only on the irritating properties of the substance. Careful consideration should be given to circumstances where exposure may be sub-optimal due to difficulties in achieving a good solution and/or a solution of sufficient concentration. In some circumstances inconsistent results from guinea pig studies, or between guinea pig and LLNA studies, might increase the uncertainty of making a correct interpretation. Finally, for existing human data consideration must be given to whether inter-individual variability is such that it is not scientifically sound to generalise from a limited population.

The non-animal test methods (*in chemico/in vitro*) currently available have no or limited metabolic capacity (Fabian et al., 2013). Therefore substances requiring metabolic activation before becoming sensitisers may not be correctly identified by such test methods. Also, some chemicals requiring abiotic transformation (e.g. auto-oxidation or hydrolysis) before becoming active may not be detected, however this issue is also applicable also to the animal test methods. More information on these limitations can be found in Section R.7.3.7 of this Guidance. The QSAR models also most often do not account for metabolism/abiotic transformation (e.g. auto-oxidation or hydrolysis) by themselves. There are strategies, however, which can facilitate the consideration of
metabolic information (e.g. see Section R.7.3.7.2.2 “how to deal with the lack of or limited metabolic capacity of non-animal test methods?”)

R.7.3.6.2 Concluding on suitability for Classification and Labelling

In order to conclude on an appropriate classification and labelling position with regard to skin sensitisation, the available data should be considered using the criteria according to Annex I to the CLP Regulation (EC) No 1272/2008 (See also Section 3.4 of the Guidance on the Application of the CLP criteria).

R.7.3.6.3 Concluding on suitability for chemical safety assessment: potency and dose-response assessment

The CLP Regulation specifies that skin sensitisers should be allocated into sub-categories (i.e. 1A or 1B) whenever possible. In case the data are not sufficient for sub-categorisation, the substance must be classified in the general Category 1 (for further information, see Section 3.4 of the Guidance on the Application of the CLP criteria).

R.7.3.6.3.1 Measurement of potency

Appropriate dose-response data can provide important information on the potency of the material being tested. This can facilitate the development of more accurate risk assessments. This section refers to potency in the induction phase of sensitisation.

Neither the standard LLNA nor the GPMT/Buehler test is specifically designed to evaluate the skin sensitising potency of test compounds. Instead, they are used to identify the sensitisation potential for classification purposes. However, these tests can all be used to estimate potency to a varying degree.

The relative potency of substances may be indicated by the percentage of positive animals in the guinea pig studies in relation to the intradermal or topical induction concentration(s) tested. Likewise, in the LLNA, the EC3 value (the dose estimated to cause a 3-fold increase in local lymph node proliferative activity) is used as a measure of potency (see the CLP Regulation, tables 3.4.3 and 3.4.4, and the Guidance on the Application of the CLP criteria, Table 3.4.2.f). Often, linear interpolation of a critical effect dose from the EC3 is proposed (ECETOC, 2000), but more advanced statistical approaches basing conclusions on the characteristics of the dose-response curve and variability of the results is also used (Basketter et al., 1999; van Och et al., 2000). The dose-response data generated by the LLNA makes this test more informative than guinea pig assays for the assessment of skin sensitising potency.

EC3 data correlate quite well with HRIPT thresholds derived from historical testing data (Griem, 2003; Schneider and Akkan, 2004; Basketter et al., 2005b; ICCVAM, 2011; Basketter and McFadden, 2012). However, the human data were not derived using a single well-defined protocol and thereby some uncertainty is associated with these comparisons. Furthermore, the thresholds derived cannot be applied directly to the general population. They must be subjected to a rigorous risk assessment process, including the application of several safety assessment factors (Api et al., 2008; Basketter and Safford, 2015a). However a retrospective analysis performed by ICCVAM (2011) of LLNA data compared to human and/or guinea pig data revealed that ca. 48% of strong sensitisers, i.e. Cat 1A according to the CLP criteria, were underclassified based on animal data (both LLNA and GPMT/Buehler tests) and human data.
In the case of the GPMT and Buehler test, due to the dose selection criteria specified in the OECD TG 406, it is usually not possible to firmly conclude that a substance is a Category 1B sensitiser since classification in Category 1A cannot be excluded. Therefore, in case classification in Category 1A cannot be excluded, the general Category 1 classification must be chosen.

Several approaches for potency prediction have been proposed in the scientific literature, and some of these could be useful, on a case-by-case basis, to support identification of strong sensitisers and setting of SCLs. A review of these approaches is given in Appendix R.7.3-4. However, concerning classification according to CLP, currently (at the time of publication of this Guidance), no widely accepted approach and data interpretation procedure based on non-animal data (in chemico, in vitro, in silico) is available.

A lack of potency information and subsequent possibility to sub-categorise and to set SCLs, may result in a lower level of protection of humans, especially if the substance is used in a mixture and appropriate concentration limits are not used, leading to incorrect labelling of the mixture. Therefore, in order to fulfil REACH information requirements, data from non-animal test methods (in chemico/in vitro/in silico) must allow the detection of substances that can be presumed to have the potential to produce significant (strong/extreme) sensitisation in humans. A review of different approaches for assessing skin sensitisation potency is provided in Appendix R.7.3-4. However, as work is still ongoing to address the lack of potency characterisation based on non-animal approaches, the registrant is advised to follow-up the recent and future developments in the field.

**R.7.3.6.3.2 Derivation of a DNEL**

Even though EC3 values obtained from the LLNA (Basketter et al., 2007) can be used for DNEL derivation, the first step should always be to perform a qualitative approach to assess and control the risks that may arise from the exposure to a substance causing skin sensitisation. It should be noted that, currently, a quantitative assessment cannot be performed by using guinea pig data or non-animal testing approaches. Guidance on how to use potency information for a qualitative assessment (see also Section E.3.4.2 of Part E of the Guidance on IR&CSA) and how to derive a DNEL as a second step in the safety assessment of sensitisers is given in Section R.8.6 and Appendix R.8-10 of Chapter R.8 of the Guidance on IR&CSA.

Quantitative risk assessment (QRA) approaches have been proposed for the identification of safe exposure levels for skin sensitising substances. A QRA approach should use all the information available, i.e. human and animal data. Such an approach has been used e.g. by the fragrance industry (Api et al., 2008; 2015). However, this approach has received criticism, especially when evaluated by the Scientific Committee on Consumer Safety (SCCS), as the safe levels identified by using QRA were not supported by existing data (SCCS, 2015a). The SCCS recommends further development of the approach (SCCS, 2015b). In the framework of the IDEA project (http://www.idealproject.info), the original QRA methodology (Api et al., 2008) has been revised (Api et al., 2015). These revisions include a revision of the safety assessment factors and the introduction of a probabilistic approach for aggregate exposure assessment. It should be noted that, despite its name, a QRA approach is not a precision tool.

Deriving the safe use levels for skin sensitisation can be problematic and may be associated with considerable uncertainty. Uncertainty assessment approaches have been published, e.g. by ECHA (see Chapter R.19 of the Guidance on IR&CSA) and IPCS/WHO (2014), and a draft guidance document is also available from EFSA (2015).
R.7.3.6.4 Additional considerations

Chemical allergy is commonly designated as being associated with skin sensitisation (ACD) and/or with sensitisation of the respiratory tract (asthma rhinitis and extrinsic allergic alveolitis). In view of this it is sometimes assumed that allergic sensitisation of the respiratory tract will result only from inhalation exposure to the causative substance, and that skin sensitisation necessarily results only from dermal exposure. This is misleading, and it is important for the purposes of risk management to acknowledge that sensitisation may be acquired by other routes of exposure. Since adaptive immune responses are essentially systemic in nature, sensitisation of skin surfaces may develop from encounter with contact allergens via routes of exposure other than dermal contact.

Similarly, there is evidence from both experimental and human studies which indicate that effective sensitisation of the respiratory tract can result from dermal contact with a chemical respiratory allergen (Redlich, 2010; Kimber et al., 2014c). Thus, in this case, it appears that the immune response necessary for the acquisition of sensitisation of the respiratory tract can be obtained via skin contact with chemical respiratory allergens (Arts and Kuper 2007; Kimber et al., 2002). Such considerations have important implications for risk management. Thus, for instance, there is a growing view that effective prevention of respiratory sensitisation requires protection of both skin and respiratory tracts. This includes the cautious use of known contact allergens in products to which consumers are (or may be) exposed via inhalation, such as sprays. The generic advice for appropriate strategies to minimise the risk of sensitisation to chemical allergens is to provide protection of all relevant routes of exposure.

R.7.3.6.5 Information not adequate

A Weight-of-Evidence approach, comparing available adequate information with the tonnage-triggered REACH information requirements, may result in the conclusion that the requirements are not fulfilled. In order to proceed in further information gathering the testing and assessment strategy given in the next Section R.7.3.7 can be adopted.

R.7.3.7 Testing and assessment strategy for skin sensitisation

R.7.3.7.1 Objective / General principles

The following testing and assessment strategy is recommended for developing adequate and scientifically sound data for the assessment and classification of the skin sensitisation properties of a substance. For existing substances with insufficient data, this strategy can also be used to decide which additional data, besides those already available, are needed. The objective is to collect all available information (including data from test methods and non-testing approaches) in order to assess the risk for skin sensitisation and/or to identify information gaps to be covered by generation of new information. The key principle of this strategy is that all available information is evaluated before another study is initiated. The strategy seeks to ensure that the data requirements are met in the most efficient and humane manner so that animal usage and costs are minimised.

The animal welfare considerations and the information requirements of the REACH Regulation has stimulated research on integrated strategies for skin sensitisation in the past few years. Some of these works give data on individual non-testing or non-animal testing methods e.g. in silico (Teubner et al., 2013) or in vitro information (Martin et al., 2010; Maxwell et al., 2014; Reisinger et al., 2015; Urbisch et al., 2015), while others also make use of different combinations (Basketter et al., 2013; Rorije et al., 2013; Jaworska et al., 2013). Besides the approaches mentioned above, a number of other data integration approaches are documented in the OECD Guidance Document on the...
reporting of defined approaches and individual information sources to be used within IATA for skin sensitisation (OECD, 2016a).

The testing and assessment strategy presented here comprises three parts (see Figure R.7.3-1 and Figure R.7.3-2): Part 1 (elements 1 to 5) is about retrieving existing information, Part 2 (element 6) represents Weight-of-Evidence analysis and expert judgement, in case a conclusion cannot be reached based on a single element listed in Part 1. Part 3 (elements 7 to 9) is about generation of new information by testing, if needed. The elements presented in Figure R.7.3-2 can be rearranged as appropriate, depending on the information needs to conclude on the substance’s potential to cause skin sensitisation. This may be particularly helpful in cases where a conclusion can be drawn from certain elements without having to consider all of them.

The specific rules for adaptation of standard information requirements are described in column 2 of Annex VII to the REACH Regulation, whereas the general rules for adaptation from standard information requirements are given in Annex XI.

The new elements in the strategy are the recently EU/OECD- adopted and/or internationally validated in chemico/in vitro test methods for skin sensitisation (in particular the three test methods specified in Element 5b in Figure R.7.3-2). These methods are based on the mechanistic understanding of the biological key events of skin sensitisation, initiated by the covalent binding of the substance to skin proteins. These key events have been codified in an Adverse Outcome Pathway (AOP) for skin sensitisation approved by the OECD (OECD, 2012). Three of these key events, i.e. peptide/protein reactivity, keratinocyte response and monocytic/dendritic cell response, correspond to elements 5b (existing data), and to elements 7a, 7b and 7c (generation of new data) of Figure R.7.3-2 below.

The strategy aims to help the registrant to find out how these in chemico/in vitro test methods for skin sensitisation can be used in a Weight-of-Evidence approach according to the Annex XI, sections 1.2 – 1.5, to the REACH Regulation to enable hazard identification and appropriate classification decision for a substance. Also other types of data, such as (Q)SAR, read-across and human data should be used in combination with the in chemico/in vitro test results. The key strengths and limitations of the in chemico/in vitro tests and other types of data are addressed below.
**Figure R.7.3–1 Overview of the testing and assessment strategy for skin sensitisation**

*In case suitable information is available from only one information source to conclude on the skin sensitisation potential of a substance, there is no need to build up a Weight-of-Evidence judgement.

**R.7.3.7.2 Application of the Testing and Assessment Strategy**

Due to the recent developments in the field of non-animal test methods for skin sensitisation, and in line with Article 13(1), Annex VI and the introductory paragraph of Annex VII to the REACH Regulation, registrants are advised to investigate whether the information requirement for skin sensitisation can be fulfilled by using (existing) results from the new test methods e.g. in a Weight-of-Evidence approach. As for any adaptation these are assessed on a case-by-case basis and performed at the registrants own discretion and potential economic risks (e.g. it may be necessary to perform the in vivo test in case the non-animal test(s) results remain inconclusive).

It is important to note that it is the responsibility of the registrant to ensure that the chosen test method (e.g. in vitro, in chemico or in silico) is suitable for testing the substance and obtaining adequate information. So before performing a specific non-animal test the registrant should consider whether there are substance-specific limitations that may hinder the performance of the test (e.g. low solubility or log Kow, UVCB nature of the substance for instance the DPRA is not applicable to UVCBs). There may also be some limitations of the test system like the absence of or limited metabolic capacity and hence pro-haptens may not be correctly detected and may give false negative results. Also substances requiring abiotic transformation, e.g. auto-oxidation or hydrolysis, (pre-haptens) may not be correctly indentified. In case the substance does not fall into the applicability domain of the non-animal test methods, an in vivo test (i.e. an LLNA) would need to be performed.
According to Step 1 of Annex VI to the REACH Regulation, all existing available test data should be gathered before any new testing is initiated. In Part 1 (elements 1 to 5c) of this strategy, existing and available information from the literature and databases is gathered and considered. The order of the different elements of Part 1, is only indicative and they may be arranged as appropriate. This may especially be helpful in cases where a reliable conclusion can be drawn from certain element(s) without having to consider all of them. For instance, if there are adequate human data (element 2) available that indicate that the substance should be classified as skin sensitiser according to the CLP Regulation, further testing is not required. At the end of the Part 1, and if no final conclusion can be derived directly from one or several of the available pieces of information, all the information collected should be analysed using a Weight-of-Evidence approach (element 6).

In the information generation part (elements 7 to 9), new information on the skin sensitisation potential of the substance is produced either by means of non-animal test methods or, as a last resort (see Articles 13(1), 25(1) and Annex XI to the REACH Regulation), in vivo testing (element 9). The properties of the substance and existing information determine the need to generate new information, i.e. new data may not need to be generated for all elements under the information generation part as the order of the elements is only indicative.

While it is recommended that this approach be followed, other approaches may be more appropriate and efficient on a case-by-case basis.

Due to the complexity of the skin sensitisation endpoint, a combination of alternative test methods (e.g. in silico, in chemico and in vitro) in a Weight-of-Evidence approach may need to be considered to increase confidence in the final assessment of skin sensitisation. The in vitro and in chemico test methods described in Sections R.7.3.4.1 and R.7.3.5.1 and in Figure R.7.3–2 below (as elements 5 and 7) have not been developed as stand-alone methods, especially when negative results are obtained. The results obtained with in silico methods that aim at predicting the final endpoint (e.g. LLNA outcome) could be used according to the REACH Regulation, if they fulfil the Annex XI, section 1.3 requirements. However, additional evidence such as read-across from analogues or results method(s) may be needed to confirm the reliability of the (Q)SAR prediction, which would be otherwise difficult to assess and accept.

In case no information on skin sensitisation is available for a substance it is recommended to start the assessment by using the OECD QSAR Toolbox (see Section R.7.3.4.1). The Toolbox can be used for many purposes. First, it facilitates the identification of existing in chemico, in vitro and in vivo data already available for the substance of interest. Second, it identifies skin sensitisation specific alerts and protein-binding alerts using profilers. Third, it can be used to predict and characterise metabolic or abiotic transformation (e.g. auto-oxidation or hydrolysis) products of the substance. Fourth, it facilitates the identification of analogues with experimental data for read-across, trend analysis and QSAR model building. In addition, the existing in vivo data for the substance and/or analogue substance(s) may provide useful information on the skin sensitisation potency, e.g. via EC3 values obtained from the existing LLNA studies. Note that the predictions can address the in vivo endpoints as well as in vitro ones (although for the moment there are not many in vitro data included in the Toolbox and the identification of analogues with data can be difficult).

In case all the available existing data, the use of the OECD QSAR Toolbox and/or other in silico tools do not enable to conclude on the skin sensitisation hazard including the sensitising potency of a substance, it is strongly recommended to investigate a sufficient number of key events (e.g. elements 7a, b and c in Figure R.7.3–2) as described in the AOP for skin sensitisation, by providing information from non-animal test methods or by...
other sources of information. The current understanding is that covering different key events by in chemico/in vitro/in silico or other data provides the best predictivity for the endpoint. Therefore, in case coverage of one or two key event(s) is omitted, the registrant would need to justify the approach taken based on the current knowledge. Based on the current knowledge, information obtained from peptide reactivity, whether obtained from in chemico or in silico methods, seems to show the highest predictive power and may provide more weight to the overall assessment of skin sensitisation (Natsch et al., 2012, Urbisch et al., 2015). However, it is important to consider the specific limitations of the approach with respect to the substance under investigation.

The OECD Guidance Document on the reporting of defined approaches to be used within IATA (OECD, 2016b) aims to contribute to a harmonised approach to the reporting of defined approaches used as elements within IATAs (see Annexes I and II of the OECD Guidance Document, and Appendices R.7.3-2 and R.7.3-3 of this Guidance). In the case of an AOP-informed IATA, the different pieces of information would target key events along the defined toxicity pathway and the results used to inform a regulatory decision, as pointed out in Figure R.7.3-2. The registrant is advised to use the template described in Appendix R.7.3.3 to report a defined data interpretation procedure if this is used as part of the testing and assessment strategy described in Figure R.7.3-2.

Several approaches on how to use non-animal testing approaches have been reported in scientific literature, however at this point of time (at the time of publication of this Guidance) no independent assessment on the predictivity of combining in chemico/in vitro/in silico approaches have been performed. The use of positive predictions obtained from in chemico/in vitro test methods tends to be more straightforward than in case negative or conflicting predictions are obtained. Due to the specific limitations of each of the in chemico/in vitro test methods, in case a negative prediction is obtained, it is important to justify in the dossier how a potential false prediction can be ruled out. Supporting information might be the consideration of whether the substance is or is not a pro-hapten and whether metabolism is expected to occur in vivo. Also whether the substance is a pre-hapten and requires abiotic transformation (e.g. auto-oxidation or hydrolysis) to exert its skin sensitisation potential should be considered (note: the issue of not identifying pre-haptens correctly is not solely related to in chemico/in vitro methods, but can also occur with in vivo test methods).
Figure R.7.3–2 Testing and assessment strategy for evaluating the skin sensitisation potential of substances (footnotes a to c are detailed below the figure)

<table>
<thead>
<tr>
<th>Element</th>
<th>Information</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Existing data on physico-chemical properties</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Is the substance a strong acid (pH ≤ 2.0) or base (pH ≥ 11.5), corrosive to the skin or (spontaneously) flammable in air at room temperature?</td>
<td><strong>YES:</strong> No in vivo testing required (Column 2 adaptation of Annexes VII, section 8.3) Note: extreme pH values/corrosive properties do not preclude performing in chemico/in vitro/in vivo test(s) at suitable concentrations and, therefore, it is possible to assess skin sensitisation hazard in sub-corrosive concentrations, if considered necessary.</td>
</tr>
<tr>
<td><strong>Existing human data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Are there adequate existing human data(^a), which provide evidence that the substance is a skin sensitiser?</td>
<td><strong>YES:</strong> Consider classifying according to CLP criteria (Cat 1, 1A or 1B). If not conclusive on its own, use this information for Weight-of-Evidence analysis under point 6.</td>
</tr>
<tr>
<td><strong>Existing animal data from sensitisation studies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Are there data from existing studies on skin sensitisation in laboratory animals (LLNA, GPMT, or Buehler test, EU B.42, B.50, B.51 and B.6/OECD TGs 429, 442A, 442B and 406), which provide sound conclusive evidence that the substance is a sensitiser, or non-sensitiser?</td>
<td><strong>YES:</strong> Consider classifying according to CLP criteria (Cat 1, 1A or 1B) or consider no classification. If not conclusive on its own, use this information for Weight-of-Evidence analysis under point 6.</td>
</tr>
<tr>
<td><strong>Existing (Q)SAR data and read-across</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Do “read-across” from structurally and mechanistically related substances and/or do suitable (Q)SAR predictions reliably indicate skin sensitisation potential or the absence thereof of the substance? (^b)</td>
<td><strong>YES:</strong> Consider classifying according to CLP criteria (Cat 1, 1A or 1B) or consider no classification. If not conclusive on its own, use this information for Weight-of-Evidence analysis under point 6.</td>
</tr>
</tbody>
</table>
## Existing in chemico and in vitro data

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>YES/NO:</th>
</tr>
</thead>
<tbody>
<tr>
<td>5a</td>
<td>Is there evidence/hypothesis of dermal bioavailability based on physico-chemical, in silico, in vitro or in vivo data?</td>
<td>Use this information for Weight-of-Evidence analysis</td>
</tr>
<tr>
<td>5b</td>
<td>Has the substance demonstrated peptide/protein binding properties in an EU/OECD adopted in chemico test (e.g. OECD TG 442c)? (Key event 1 of the AOP), and/or Has the substance demonstrated activation of the Nrf2-Keap1-ARE toxicity pathway in an EU/OECD adopted in vitro test (e.g. OECD TG 442d)? (Key event 2 of the AOP), and/or Has the substance demonstrated induction of the cell surface markers (CD54 and/or CD86) on monocytic cells in a validated in vitro test, e.g. h-CLAT? (Key event 3 of the AOP).</td>
<td>Consider classifying as Skin sensitiser (Cat 1, 1A or 1B) or consider no classification. If not conclusive, use this information for Weight-of-Evidence analysis under point 6.</td>
</tr>
<tr>
<td>5c</td>
<td>Are there data from (a) non-validated in vitro test(s), which provide evidence that the substance may be a skin sensitiser?</td>
<td>Consider classifying as Skin sensitiser (Cat 1, 1A or 1B). If not conclusive, use this information for Weight-of-Evidence analysis under point 6.</td>
</tr>
</tbody>
</table>

## Weight-of-Evidence analysis

| 6 | The “elements” described above may be arranged as appropriate. Taking all existing and relevant data (elements 1-5) into account, is there sufficient information to meet the information requirement of Section 8.3 of Annex VII and to make a decision on whether classification and labelling are warranted? For specific guidance on Weight of Evidence see below. | YES: Classify as Skin Sensitiser Cat. 1, 1A or 1B or consider no classification. NO: Consider the next elements of the strategy. |
### Generation of new non-animal data c

**Note:** Article 13 of the REACH Regulation specifies that information should be generated whenever possible by means other than vertebrate animal test(s) and the use of non-animal test methods is encouraged. The registrant can generate an adaptation according to Annex XI to the REACH Regulation in order to try and fulfil the standard information requirement for skin sensitisation. As for any adaptation these are assessed case-by-case basis and performed at the registrants own discretion and risks. Therefore, before performing *in chemico/in vitro* testing, it is important to consider whether the test method(s) to be used are suitable for the substance i.e. whether the substance fits in the applicability domain of a specific test method. In case the *in chemico/in vitro* tests are not suitable for the substance, an LLNA would need to be performed.

| 7a | Does the substance demonstrate **peptide/protein binding** properties in an EU/OECD adopted *in chemico* test (e.g. OECD TG 442c)? *(Key event 1 of the AOP)*  
*In chemico* test methods that have been validated and are considered scientifically valid but are not yet adopted by the EU and/or OECD may also be used if the provisions defined in Annex XI to the REACH Regulation are met. | **YES/NO:**  
Consider classifying as Skin sensitiser (Cat 1, 1A or 1B).  
If not conclusive, use this information for **Weight-of-Evidence** analysis under point 8. |
|---|---|---|
| 7b | Does the substance demonstrate **activation of the Nrf2-Keap1-ARE toxicity pathway** in an EU/OECD adopted *in vitro* test (e.g. OECD TG 442d)? *(Key event 2 of the AOP)*  
*In vitro* test methods that have been validated and are considered scientifically valid but are not yet adopted by the EU and/or OECD may also be used if the provisions defined in Annex XI to the REACH Regulation are met. | **YES/NO:**  
Consider classifying as Skin sensitiser (Cat 1, 1A or 1B).  
If not conclusive, use this information for **Weight-of-Evidence** analysis under point 8. |
| 7c | Does the substance demonstrate **induction of the cell surface markers** *(CD54 and/or CD86)* of monocytic cells in a validated *in vitro* test (e.g. h-CLAT)? *(Key event 3 of the AOP)*  
*In vitro* test methods that have been validated and are considered scientifically valid but are not yet adopted by the EU and/or OECD may also be used if the provisions defined in Annex XI to the REACH Regulation are met. | **YES/NO:**  
Consider classifying as Skin sensitiser (Cat 1, 1A or 1B).  
If not conclusive, use this information for **Weight-of-Evidence** analysis under point 8. |
| 7d | Is any additional testing/generation of data considered necessary in order to conclude on classification, or e.g. to explain the inconsistent data obtained in previous elements or to address the **Key event 4 of the AOP** *(T-cell proliferation)* with an *in vitro* test? | **YES/NO:**  
Consider performing the test and use this information for **Weight-of-Evidence** analysis. |
**Weight-of-Evidence analysis**

<table>
<thead>
<tr>
<th>8</th>
<th>The “elements” described above may be arranged as appropriate. Taking all existing and relevant data (elements 1-7) into account, is there sufficient information to meet the respective information requirement of Section 8.3 of Annex VII and to make a decision on whether classification and labelling are warranted? For specific guidance on Weight of Evidence see below.</th>
</tr>
</thead>
<tbody>
<tr>
<td>YES:</td>
<td>Classify according to as Skin Sensitiser Cat. 1, 1A or 1B or consider no classification.</td>
</tr>
<tr>
<td>NO:</td>
<td>Consider the next element of the strategy.</td>
</tr>
</tbody>
</table>

**Generation of new in vivo data for sensitisation as a last resort (Annex VII to the REACH Regulation)**

<table>
<thead>
<tr>
<th>9</th>
<th>Does the substance demonstrate sensitising properties in an EU/OECD adopted in vivo test, the LLNA (EU B.42/OECD TG 429, EU B.50/442A or EU B.51/442B“)? →</th>
</tr>
</thead>
<tbody>
<tr>
<td>YES:</td>
<td>Classify according to CLP criteria (Skin Sensitiser Cat. 1, 1A or 1B).</td>
</tr>
<tr>
<td>NO:</td>
<td>No classification needed.</td>
</tr>
</tbody>
</table>

### Notes to the testing and assessment figure on skin sensitisation:

1. **a)** Data from case reports, occupational experience, poison information centres, Human Patch Tests or from clinical studies.

2. **b)** It is worthwhile to apply the OECD QSAR Toolbox (see Section R.7.3.4.1) to check whether there are existing data available for the substance of interest or existing and good quality data available on skin sensitisation for potential analogue substances. It should be noted that in case read-across or a category approach is to be used, adequate justification must be provided (for further information on ECHA’s read-across assessment framework (RAAF) see [http://echa.europa.eu/support/grouping-of-substances-and-read-across](http://echa.europa.eu/support/grouping-of-substances-and-read-across)). The use of available and suitable (Q)SAR models for skin sensitisation is also recommended. In case substance metabolism and/or abiotic transformation (e.g. auto-oxidation or hydrolysis) leads to the generation of new chemical species the use of the QSAR Toolbox may be helpful in finding relevant data that can be used.

3. **c)** In case (a) non-animal testing approach(es) is (are) used, information needs to be generated to address a sufficient number of the key events specified in elements 7a to 7c in order to conclude on the skin sensitisation endpoint (Basketter et al., 2015b).

4. Additional information obtained from e.g. (Q)SARs can be used to support the conclusions made. The information obtained from the assessment of one of the key events may be used to select the next most appropriate in chemico/in vitro test.

5. **d)** At this point of time ([at the time of publication of this Guidance](#)), there are no validated or adopted in vitro test methods available to address T-cell proliferation. Developments may occur in the future in the field of in chemico/in vitro test methods.
that may be able to address the limitations of the currently adopted and/or validated test methods and could provide more confidence in the results already obtained.

e) Note: The LLNA variants, i.e. EU B.50/OECD TG 442A and B.51/442B, are not suitable for sub-categorisation of the skin sensitisation potency, i.e. for differentiation between CLP categories 1A and 1B.

An overview of how to use the information on the different Key events is given in Figure R.7.3-3 below.
Chapter R.7a: Endpoint specific guidance
Draft Version 5.0 – Public – February 2016

Figure R.7.3–3 Snapshot of the Testing and Assessment Strategy - How to use the data on key events. Note: the order of key events is not specified and can be arranged as appropriate.
R.7.3.7.2.1 Predictive capacity of the existing in vivo and non-animal tests when compared to human data

Different approaches have been presented in the scientific literature on the predictivity of non-animal testing approaches (see Section R.7.3.7.2.5 How to report Weight of Evidence). E.g. Urbisch et al. (2015) compared the predictive capacity of the LLNA and that of non-animal (in chemico/in vitro) testing strategies towards skin sensitisers in humans. The authors showed that for LLNA vs. historical predictive testing in humans, the accuracy of prediction was 82%, with a sensitivity (i.e. true positive rate) of 91% and a specificity (i.e. true negative rate) of 64%. For non-animal test methods used in combination the accuracy was 90% with a sensitivity and a specificity of 90% (n~100 chemicals). While the nature of the data is only partly disclosed and these data have not been assessed independently, there is some indication that, when in chemico and in vitro methods are used in combination, their ability to predict human data seems to be comparable to that of the LLNA in the identification of human sensitisers and non-sensitisers (i.e. Cat 1 vs. non-classified substances). However, the individual tests on their own were not as sensitive as the LLNA.

R.7.3.7.2.2 How to deal with the lack of or limited metabolic capacity of non-animal test methods?

The in chemico DPRA does not have any metabolic capacity and the in vitro Keratinosens™ assay and h-CLAT assay have only a limited metabolic capacity in the test systems. Due to the lack of or the limited metabolic capacity, these test methods may not correctly identify sensitisers that would require metabolic activation to exert their sensitisation activity and therefore they may provide false negative results. The above non-animal test methods may also provide false negative results for substances requiring abiotic transformation (e.g. auto-oxidation or hydrolysis), especially in case of slow oxidation rates. The issue of not identifying pre-haptens correctly is not solely related to in chemico/in vitro methods, but can also occur in in vivo tests. Currently, a modification of the in chemico DPRA is under development (Peroxidase Peptide Reactivity Assay (PPRA)) for a better identification of pro-haptens (Merckel et al., 2013), however work is still ongoing to assess the added value of this assay.

An analysis concerning the ability of non-animal testing methods to detect pro- and pre-haptens was performed in occasion of an EURL ECVAM expert meeting held in 2015 and discussed with a group of experts (Casati et al., 2016). The experts noted that many of the thought pro-haptens were actually also pre-haptens (e.g. geraniol). It was also noted that the majorit of the non-direct acting haptens were pre-haptens which were generally identified with the DPRA and cell-based assays. A problem was noted with slow oxidizers, which were not correctly identified, however their identification would also fail by using in vivo test methods. Substances that are exclusively pro-haptens, which were not identified by the DPRA, were generally correctly identified by one of the cell-based assays (the h-CLAT detecting the majority of those). The outcome of the analysis was that, by using non-animal test methods, a comparable prediction of skin sensitisation hazard to the one of the LLNA was obtained. The expert group concluded that, in light of this analysis, unless there is a compelling scientific argument that a substance could be an exclusively metabolically activate pro-hapten, the negative results obtained from non-animal test methods could be seen acceptable. In silico tools such as TIMES-SS or OECD QSAR Toolbox could be used to support such argumentation.

Thus it is strongly recommended to evaluate all available toxicokinetic information (see Section R.7.12 of Chapter R.7c the Guidance on IR&CSA) and to run computational tools.
such as the OECD QSAR Toolbox or TIMES-SS that can partially cover for the lack of metabolite or abiotic transformation (e.g. auto-oxidation or hydrolysis) information. However the user should not rely solely on the results from these tools to exclude that metabolic activation or abiotic transformation (e.g. auto-oxidation or hydrolysis) may take place. These softwares have modules for simulating (skin) metabolism and abiotic transformation (e.g. auto-oxidation or hydrolysis) of substances. TIMES-SS currently does not have a skin sensitisation model that combines for example hydrolysis with the skin sensitisation endpoint. In case the substance is predicted to be a non-sensitiser but the simulated metabolites or products have positive experimental data or trigger skin sensitisation alerts, the latter might be responsible for sensitisation. The simulated metabolites need a specific assessment and might require the generation of new experimental data. Other tools (e.g. Derek Nexus) incorporate the knowledge for metabolic transformation in developing alerts for skin sensitisation (e.g. hydroquinone and precursors).

It has been proposed that experimental data available from other endpoints (e.g. from in vitro mutagenicity) could provide additional information to support the conclusions on skin sensitisation obtained from non-animal test methods. The approach to use in vitro mutagenicity test results for assessing potential metabolite formation has not gone through an independent review and more experience has still to be gained. However, it may provide some useful insights of the potential electrophilic reactivity of the substance under metabolic conditions (Patlewicz et al., 2010). Information obtained from in vitro mutagenicity study may provide useful information in the Weight-of-Evidence assessment when used in combination with the computational tools, which could either support the results obtained from the in chemico/in vitro test methods or trigger further testing needs.

R.7.3.7.2.3 Use of non-animal data (e.g. in vitro methods) to support a category approach

In case a category approach is used to fulfil the REACH information requirements and data are available for some category members only, the generation of data by using e.g. in chemico/in vitro test methods could be used to support the category approach for this endpoint. This is especially the case when similar results on the skin sensitisation potential (or the lack thereof) are obtained from one (or more) non-animal testing method(s). In practice, it may be possible to perform only one or two in chemico/in vitro tests for the target substance of the read-across. In case of conflicting results, it is important to consider why they occurred: the reason might be that the specific substance does not belong to the category because of sensitising properties different from that of category members with good quality animal and/or human data, or that the substance does not fit into the applicability domain of the specific non-animal test. In those cases, in vivo testing may be required to assess the skin sensitisation potential of the substance.

Whenever a category approach is applied, it is essential to always justify why data can be read across from the category member substances to the target substance for which there is no good quality animal and/or human data. This justification also needs to be endpoint specific. Advice on how to build and report a category can be found on ECHA website http://echa.europa.eu/support/grouping-of-substances-and-read-across.
R.7.3.7.2.4 Sub-categorisation

Currently (at the time of publication of this Guidance), there is no widely accepted approach to integrate non-animal data into an adequate sensitisation potency classification. Some approaches have been proposed for potency prediction (see Appendix R.7.3-4). The current first choice method, i.e. the LLNA (EU B.42/OECD TG 429), allows potency estimation and the setting of SCLs, however a retrospective analysis performed by ICCVAM (2011) of LLNA data compared to human revealed that ca. 48 % of strong sensitisers, i.e. Cat 1A according to the CLP criteria, were under-classified based on an EC3<2%. Other in vivo test methods either have their limitations or are unable to predict potency (for further details see Section R.7.3.5.1.2, under “Animal data”).

The current lack of sub-categorisation potential when using non-animal test methods within a Weight-of-Evidence approach may result in a lower level of protection of humans with respect to classification of mixtures containing sensitising substances. This is due to the fact that, depending on the skin sensitisation potency of a substance, different concentration limits are to be applied: i.e. for Cat 1 and Cat 1B the generic concentration limit (GCL) is 1%, for Cat 1A (strong) the GCL is 0.1% and for very strong (extreme) sensitisers an SCL of 0.001% (or lower) is recommended according to the Guidance on the Application of the CLP criteria (for further information, see Section 3.4 of the Guidance on the Application of the CLP criteria). In short, this may lead to the situation that mixtures with potent sensitisers are not classified in a way which reflects the hazards of the mixture, if the general Cat 1 is used and the GCL of 1% is applied instead of 0.1% (or the SCL of 0.001%). This would mean a lowering of the safety level as compared to current provisions, which may lead to an increased incidence of human sensitisation to potent sensitisers. Therefore, in order to fulfil REACH information requirements, the data from non-animal test methods (in chemico/in vitro/in silico) must allow the detection of substances that can be presumed to have the potential to produce significant (strong/extreme) sensitisation in humans. A review of different approaches aiming to provide indications of skin sensitisation potency is provided in Appendix R.7.3-4. However, as work is still ongoing to try and address the lack of potency characterisation based on non-animal approaches (e.g. Reisinger et al., 2015), the registrant is advised to follow-up the recent and future developments in the field. The registrant is also advised to follow any updates to the ECHA webpage concerning Testing methods and alternatives (see: http://echa.europa.eu/support/oecd-eu-test-guidelines).

When reliable and relevant human data are available, they can also be used for the classification of skin sensitising substances into sub-categories according to the CLP Regulation (for further information, see Section 3.4.2.3.1. of the Guidance on the Application of the CLP criteria).

R.7.3.7.2.5 How to perform and report a Weight-of-Evidence analysis

The Weight-of-Evidence based adaptation of the standard information requirement, i.e. the LLNA, is based on the OECD AOP for skin sensitisation and its key events (OECD, 2012). It is recognised that, in the LLNA, key events 1 to 4 are addressed since the biological response, i.e. induction of skin sensitisation, is caused by the cascade of these key events. Therefore, in the Weight-of-Evidence approach these key events should be covered to the extent possible. At present, three in chemico/in vitro tests, each addressing a specific key event of the AOP, have been adopted by the OECD and/or validated by EURL ECVAM. A sufficient number of these key events should be covered either by an in chemico/in vitro test or by other types of information (e.g. (Q)SAR, read-across). It should be noted that, based on current knowledge, the information obtained from peptide reactivity, whether obtained from in chemico or in silico methods, seems to show the highest predictive power and may provide more weight to the overall
assessment of skin sensitisation (Natsch et al., 2012, Urbisch et al., 2015). There is currently no scientifically valid or internationally adopted in vitro method to cover the fourth key event, i.e. lymphocyte proliferation. In case a suitable non-animal test method, e.g. in vitro method, becomes available, this could bring more weight to the overall Weight-of-Evidence approach. However, the available studies on the predictivity of different combinations of in chemico/in vitro methods/other information type seem to show that a good predictivity for hazard identification (Cat. 1 vs. non-sensitiser) can be achieved by covering the first three key events (Hirota et al., 2015; Maxwell et al., 2014; Patlewicz et al., 2014; Takenouchi et al., 2015; Tsujita-Inoue et al., 2014; Urbisch et al., 2015; Van der Veen et al., 2014): the use of the non-animal test methods in combination seems to be comparable to that of the LLNA in the identification of human sensitisers and non-sensitisers (i.e. Cat 1 vs. non-classified). However, the individual tests on their own were not as sensitive as the LLNA.

When in chemico/in vitro studies are used to fulfill the Annex VII information requirement for skin sensitisation by means of the general rules of adaptation as specified in sections 1.2 – 1.5 of Annex XI to the REACH Regulation, the registrant should provide a case-specific justification on why and how the in chemico/in vitro data used within a Weight-of-Evidence approach can cover for the information requirement. In that Weight-of-Evidence justification, e.g. coverage of the key events (see “Testing and assessment strategy for skin sensitisation” above), the quality and reliability of the data, scope and limitations of each test method used and consistency of the results need to be considered. Further provisions on Weight of Evidence can be found in Section R.4.4 of Chapter R.4 of the Guidance on IR&CSA and in Art. 9(3) of the CLP Regulation.

It should be noted that the data used to cover the key events, whether they are in chemico/in vitro results or other data, can be inconsistent. For example it may happen that two tests/data points are negative and one is positive for skin sensitisation. In case of inconsistent or conflicting data, a scientific explanation should be provided. The explanation may be, for example, that the substance needs metabolic activation to become a skin sensitiser and the test system misses the required metabolic competence. It may also be that the test substance does not fall into the applicability domain(s) of one or more of the in chemico/in vitro methods used. If the conflicting information/results cannot be explained, the registrant will need to generate/collect further information in order to support the prediction of the skin sensitisation potential of the substance. If in the end the registrant is not able to conclude on this endpoint due to inconsistent or inconclusive data, there may be a need to perform an LLNA.

As pointed out in elements 6 and 8 (Weight-of-Evidence analysis) of the testing and assessment strategy above, in case the skin sensitisation potential of a substance cannot be properly characterised based on the available data, generation of new data is necessary. This data can be e.g. (Q)SAR, data that are specific to a key event (e.g. in chemico/in vitro), read-across or, as a last resort, the in vivo study, i.e. the LLNA. The LLNA may have to/must be performed in any case if e.g.:

- The registrant may have some existing information from similar substances and/or QSAR(s) indicating that the substance may be a strong or extreme skin sensitiser and cannot conclude on adequate for classification and labelling and/or risk assessment,

- The test substance does not fall into the applicability domain of the in chemico/in vitro tests for skin sensitisation (Note: assessment of the suitability of a test method for a substance should be performed before testing is initiated), or

- The results of the in chemico/in vitro tests are inconsistent and this inconsistency cannot be explained scientifically, or
• Determination of skin sensitisation potency (Cat 1A vs. 1B) is not possible based on non-animal testing approaches and is needed for ensuring the safe use of the substance.

At the end of the testing and assessment strategy, the data obtained, justification of the choice of the test methods, analysis of data consistency, conclusion made on hazard identification and on classification according to the CLP Regulation should be reported clearly and transparently. For the reporting of the approach applied according to a testing and assessment strategy it is recommended to use the template provided in Appendix R.7.3.3 of this Guidance and which is based on Annex I of the OECD Guidance Document on the reporting of defined approaches to be used within IATA (OECD, 2016b).
RESPIRATORY SENSITISATION

R.7.3.8 Mechanisms of respiratory sensitisation

For substances that sensitise via the respiratory tract, there is still uncertainty regarding the exact mechanisms leading to respiratory sensitisation. Based on the current knowledge the induction of respiratory sensitisation can occur via inhalation or dermal exposure to the sensitising chemical (Redlich, 2010; Kimber et al., 2015).

The current hypothesis is that the mechanism favours Th2-type immune responses (skin sensitisation favours Th1-type response), which is characterised by the production of cytokines, such as IL-4 and IL-5, and IgE antibodies. This is supported by studies performed in rodents and by human evidence (Adenuga et al., 2012; Kimber et al., 2014b; Helakoski et al., 2015). Recently, it has been hypothesised that Th17 cells would also play a crucial role in respiratory sensitisation via secretion of IL-17 (Lambrecht and Hammad, 2013). The role of IgE may be the greatest reason for uncertainty, as there are patients who display serum IgE antibodies of the appropriate specificity, whereas in other instances (and particularly with respect to the diisocyanates) there are symptomatic subjects in whom it is not possible to detect these IgE antibodies. It has been hypothesised that either there may be a mechanisms leading to respiratory sensitisation that is IgE-independent, or this is linked to technical difficulties in the accurate measurements of hapten-specific IgE-antibodies (Cochrane et al., 2015).

In addition, an AOP for respiratory sensitisation to low molecular weight substances is currently under development at the OECD. The proposed Key Events for respiratory sensitisation are: Key Event 1 Covalent binding of substances to proteins (note: respiratory sensitisers seem to prefer binding to lysine, whereas skin sensitisers bind both to cysteine and lysine); Key Event 2 Cellular danger signals (activation of inflammatory cytokines and chemokines and cytoprotective gene pathways (Th2)); Key Event 3 Dendritic cell activation and migration (Th2 skewed); Key Event 4 Activation and proliferation of T-cells (Th2) (Mekenyan et al., 2014 and Sullivan et al., 2015).

R.7.3.9 Information sources on respiratory sensitisation

R.7.3.9.1 Non-human data on respiratory sensitisation

R.7.3.9.1.1 Non-testing data on respiratory sensitisation

Attempts to model respiratory sensitisation have been hampered by the lack of a predictive test protocol for assessing chemical respiratory sensitisation. (Q)SAR models are available but these have largely been based on data for substances reported to cause respiratory hypersensitivity in humans. Examples of some structural alerts are shown in Table R.7.3–3.

Agius et al. (1991) made qualitative observations concerning the chemical structure of substances causing occupational asthma. This work drew attention to the large proportion of chemical asthmagens with at least two reactive groups, e.g., ethylene diamine and toluene diisocyanate. The earlier work was followed up by a simple statistical analysis of the occurrence of structural fragments associated with activity, with similar conclusions (Agius et al., 1994 and 2000).
The MCASE group has developed three models for respiratory hypersensitivity (Karol et al., 1996; Graham et al., 1997, Cunningham et al., 2005). The Danish (Q)SAR Database has an in-house model for respiratory hypersensitivity for which estimates can be extracted from the online database (available at http://qsar.food.dtu.dk/). Derek Nexus contains several alerts derived from a set of respiratory sensitisers/asthmagens (Payne et al., 1995).

The structural alerts (SARs) are in principle transparent and easy to apply. Structural alerts related to respiratory sensitisation have been collected and described in the literature (Aigus et al., 1991, 1994 and 2000; Payne et al., 1995). It should be stressed however that these are derived from chemical asthmagens, i.e. substances causing asthma like symptoms with or without immunological mechanisms, and not specifically chemical respiratory allergens. Enoch et al. (2012) developed a set of mechanism-based structural alerts for low molecular weight organic chemicals with the potential to cause respiratory sensitisation. A need still remains to develop new (Q)SARs when a robust predictive test method becomes available.

**Table R.7.3–3 Examples of structural alerts for respiratory sensitisation**

<table>
<thead>
<tr>
<th>Structural Alert Description</th>
<th>Examples of structures</th>
</tr>
</thead>
<tbody>
<tr>
<td>isocyanate</td>
<td>Toluene-2,4-diisocyanate</td>
</tr>
<tr>
<td>cyclic anhydride</td>
<td>maleic anhydride</td>
</tr>
<tr>
<td>diamine</td>
<td>trimellitic anhydride</td>
</tr>
<tr>
<td>β-lactam</td>
<td>Some β-lactam antibiotics from penicillin and the cephalosporin groups</td>
</tr>
</tbody>
</table>

Recent work on the mechanism of respiratory sensitisation in humans and on the identification of structural alerts specific to respiratory sensitisation has been described in Enoch et al. (2009, 2010, 2012 and 2014). In these papers, the authors investigated a
common molecular initiating event and mechanism for low molecular weight respiratory 
sensitisers (found to be the formation of a covalent bond in the lung) and applied their 
findings to predict respiratory sensitisation by read-across. The authors have proposed a 
set of 52 structural alerts which define the chemistry associated with covalent protein 
binding in the lung. Each structural alert is also characterised by a mechanistic domain 
("mechanistic alert") and some data indicating presence of effect. Most of these alerts (a 
total of 41) have been encoded in the OECD QSAR Toolbox (ver. 3.3) profiler 
"Respiratory sensitisation". The full list of the encoded structural alerts for respiratory 
sensitisation is available under the OECD QSAR Toolbox feature “documentation”, 
together with the description, applicability domain, mechanism, set of substances used 
for the profile training set, profile/alert analysis. Some examples of structural alerts are 
di-isocyanates, anhydrides and lactams (e.g. β-lactams). Dik et al. (2014) give some 
recent guidance to the reader on how to value the different models / alerts and how to 
improve the predictivity from SARs with different positive and negative predictivity in a 
tiered approach. The paper includes compiled datasets (QSAR training datasets, literature 
data, occupational asthma data) on substances which are considered respiratory 
sensitisers. Thus, it can provide a source of analogues for grouping and endpoint-specific 
read-across data.

R.7.3.9.1.2 Testing data on respiratory sensitisation

In vitro data

No validated or widely recognised in vitro test methods specific to respiratory 
sensitisation are available yet, owing to the complexity of the mechanisms of the 
sensitisation process. This is most likely due to the fact that there are still many 
uncertainties concerning the underlying immunological mechanisms, in particular with 
respect to the role of IgE antibody.

Some in vitro methods to assess respiratory sensitisation have been published and here 
are a few examples:

- The MucilAir model from Epithelix® that uses 3D human airway epithelial model 
to study multiple endpoints, e.g. cilia beating monitoring, trans-epithelial electrical 
resistance, cytotoxicity and cytokine/chemokine release (Huang et al., 2013);

- The genomic Allergen Rapid Detection (GARD) assay that measures cellular 
expression markers common to myeloid and dendritic cells by using flow 
cytometry (Forreryd et al., 2015);

- The use of in chemico DPRA and PPRA for the identification of respiratory 
sensitisers has also been considered, as the peptide/protein binding step is similar 
for both skin and respiratory sensitisers; however respiratory sensitisers prefer 
binding to lysine residues (Lalko et al., 2012 and 2013).

Other test methods are currently under development. Therefore it is advised to follow the 
recent developments in the field.

Efforts are still needed to identify the most relevant endpoints in the optimisation of 
existing tests. However, a combination of several in vitro tests, covering the relevant 
mechanistic steps of respiratory sensitisation, into a test battery could eventually lead to 
the identification of respiratory sensitisers. There are efforts ongoing to develop an AOP 
for respiratory sensitisation.
Animal data

At present, although a number of in vivo test protocols have been published to detect respiratory allergens of low molecular weight, none of these are validated nor are these widely accepted. Some of the models are briefly discussed below, however the list is not comprehensive.

- One model is based on the LLNA, where mice are exposed via the inhalation route for 3 consecutive days, after which lymphocyte proliferation is measured in the draining (mandibular) lymph nodes. Known respiratory sensitisers, such as anhydrides and diisocyanates have been assessed in this model and were also shown to be positive in this assay. However many of the substances tested showed local toxicity in the lungs, therefore, due to local toxicity, lower doses can be applied in general when compared to skin exposure. Based on the readout, i.e. stimulation indices, the results could inform about the respiratory sensitising potency (Arts et al., 2008).

- In other protocols, an LLNA-based method is used to assess and measure cytokine profiles relevant to respiratory sensitisation.

- In one protocol mice are exposed via the dermal route, after which lymphocytes are collected from the draining (auricular) lymph nodes and further cultured and prepared for cytokine determination (Dearman et al., 2002).

- Another method to assess the cytokine profile uses inhalation exposure in mice, where the animals are exposed for 3 consecutive days via the inhalation route, after which lymphocytes are collected and following ex vivo proliferation are prepared for the cytokine profile determination (De Jong et al., 2009; Johnson et al., 2011). In addition to the assessment of cytokine profiles, the assessment of gene expression profiles in a similar mouse model could be useful in distinguishing respiratory sensitisers from respiratory irritants (Adenuga et al., 2012).

- One model is based on Brown Norway rats, in which elicitation of respiratory sensitisation is assessed. This method has been used to assess known respiratory sensitisers such as diisocyanates. In this model rats are sensitised either via inhalation or dermal exposure, after which the animals are challenged via the inhalation route. The endpoints measured in this model are respiratory irritation, assessment of bronchoalveolar lavage, measurement of nitric oxide exhaled and delayed onset of respiratory response (Pauluhn, 2014).

- Another, relatively simple, approach may serve the purpose to specifically predict sensitisation of the respiratory tract: i.e. increases in total serum IgE antibodies after induction. This method is based on statistically significant increases in total serum IgE (Arts and Kuper, 2007; Kimber et al., 2011; Vandebriel et al., 2011).

There are currently no predictive methods to identify substances that induce asthma through non-immunological mechanisms. However, when performing challenge tests including non-sensitised but challenged controls, information can be obtained on non-immunological effects of these substances.

R.7.3.9.2 Human data on respiratory sensitisation

Human data on respiratory reactions (asthma, rhinitis, and extrinsic allergic alveolitis) may come from a variety of sources:

- consumer experience and comments, preferably followed up by professionals (e.g. bronchial provocation tests, skin prick tests and measurements of specific IgE serum levels);
• records of workers’ experience, accidents, and exposure studies including medical surveillance;
• case reports in the general scientific and medical literature;
• consumer tests (monitoring by questionnaire and/or medical surveillance);
• epidemiological studies.

R.7.3.10 Evaluation of available information on respiratory sensitisation

R.7.3.10.1 Non-human data on respiratory sensitisation

R.7.3.10.1.1 Non-testing data on respiratory sensitisation

The freely downloadable OECD QSAR Toolbox software (http://www.qsartoolbox.org/) encodes a profiler (set of rules and structural domains) specific for respiratory sensitisation. The profiler offers support to the user in grouping substances which share common structural alerts and possibly predict the respiratory sensitisation potential via read-across. The current version of the profiler encodes 41 structural alerts for respiratory sensitisation.

This profiler is intended to be used for the assessment of the respiratory sensitisation potential of low molecular weight substances. The profiler has been developed based on the mechanistic knowledge of the elicitation phase of respiratory sensitisation, and thus identifies substances able to covalently bind to proteins in the lung. Presence of activity could be predicted from positive predictions. Absence of effect however cannot be predicted from the lack of alert because the lack of alert might be due to the lack of effect or lack of knowledge.

This profiler should also be used with caution due to the limited data available for the development of structural alerts. This is due to the lack of a standardised assay (in vivo or in vitro) suitable for identifying potential respiratory sensitisers. The available data are drawn from clinical reports of occupational asthma, which in a number of cases results in structural alerts defined based on a low number of substances. However, all structural alerts have a clear mechanistic rationale associated with them (in terms of covalent protein binding).

Experimental data on respiratory sensitisation can be found in two of the OECD QSAR Toolbox databases: “Skin sensitisation ECETOC” and ECHA Chem. The “Skin sensitisation ECETOC” database as named in the Toolbox contains data for both skin and respiratory sensitisers.

R.7.3.10.1.2 Testing data on respiratory sensitisation

In vitro data

Presently (at the time of publication of this Guidance) there are neither scientifically valid nor adopted in vitro tests available to assess respiratory sensitisation. Several in vitro test methods have been described in the literature; however more work is needed for wider acceptance of a given test method.

Some in vitro test methods are described in Section R.7.3.9.1. The list of in vitro methods is not complete and others exist. However, none of the current test methods
have gone through a validation process and therefore expert judgement and care are
needed when assessing information obtained from such methods and its relevance.

**Animal data**

Although generation of new information for respiratory sensitisation is not a standard
information requirement under the REACH Regulation, existing information should be
assessed. In case animal data are available on respiratory sensitisation those data should
be assessed and included in the IUCLID dossier.

Information based on the LLNA model(s) as described in Section R.7.3.9.1 may provide
valuable information on the possible respiratory sensitisation potential of the substance.
Use of cytokine assessment in the studies could provide useful information by identifying
Th2-type of cytokines. Also assessment of cytokines and gene expressions could be
useful when trying to differentiate between respiratory sensitisation vs. respiratory
irritation (Adenuga et al., 2012).

Information based on the Brown Norway rat model that assesses the elicitation of
respiratory sensitisation (see Section R.7.3.9.1) may provide relevant information for
respiratory sensitisation, especially in case changes supporting sensitisation are noted in
the endpoint measured. In addition, some indication for the NOAEL on the elicitation
threshold in this test system may be obtained (Pauluhn and Poole, 2011; Pauluhn, 2014).

Moreover, measurement of serum IgE-levels in rodent models, even though variability
has been observed in the animal models, can support the identification of respiratory
sensitisers (Arts and Kuper, 2007; Kimber et al., 2011, Vandebriel et al., 2011; Kimber
et al. 2014b).

**R.7.3.10.2 Human data on respiratory sensitisation**

Although predictive models are under validation, there is as yet no internationally
recognised animal method for identification of respiratory sensitisation. Thus human data
are usually evidence for hazard identification. In case existing human data are available
on respiratory sensitisation, those data should be assessed and included in the IUCLID
dossier.

Although human studies may provide some information on respiratory hypersensitivity,
the data are frequently limited and subject to the same constraints as human skin
sensitisation data.

For evaluation purposes, existing human experience data for respiratory sensitisation
should contain sufficient information about:

- the test protocol used (study design, controls);
- the substance or preparation studied (should be the main, and ideally, the only
  substance or preparation present which may possess the hazard under
  investigation);
- the extent of exposure (magnitude, frequency and duration);
- the frequency of effects (versus number of persons exposed);
- the persistence or absence of health effects (objective description and
evaluation);
- the presence of confounding factors (e.g. pre-existing respiratory health effects,
  medication; presence of other respiratory sensitisers);
- the relevance with respect to the group size, statistics, documentation;
• the healthy worker effect.

Evidence of respiratory sensitising activity derived from diagnostic testing may reflect the induction of respiratory sensitisation to that substance or cross-reaction with a chemically very similar substance. In both situations, the normal conclusion would be that this provides positive evidence for the respiratory sensitising activity of the substance used in the diagnostic test.

For respiratory sensitisation, no clinical test protocols for experimental studies exist but evidence can come from clinical history and data from appropriate lung function tests related to exposure to the substance, or data from one or more positive bronchial challenge tests with the substance (See section 3.4.2.1.2.3 of the CLP Regulation). The test should meet the above general criteria, e.g. be conducted according to a relevant design including appropriate controls, address confounding factors such as medication, smoking or exposure to other substances, etc. Furthermore, the differentiation between the symptoms of respiratory irritation and allergy can be very difficult. Thus, expert judgement is required to determine the usefulness of such data for the evaluation on a case-by-case basis.

Where there is evidence that significant occupational inhalation exposure to a substance has not resulted in the development of respiratory allergy, or related symptoms, then it may be possible to draw the conclusion that the substance lacks the potential for sensitisation of the respiratory tract. Thus, for instance, where there is reliable (e.g. supported by medical surveillance reports) evidence that a large cohort of subjects has had opportunity for regular significant inhalation exposure to a substance for a sustained period of time in the absence of respiratory symptoms, or related health complaints, then this will provide reassurance within a Weight-of-Evidence approach regarding the absence of a respiratory sensitisation hazard.

More information on how to apply human data for C&L purposes can be found in Section 3.4.2.1.3.1 of the Guidance on the Application of the CLP criteria.

R.7.3.11 Conclusions on respiratory sensitisation

R.7.3.11.1 Remaining uncertainty on respiratory sensitisation

Major uncertainties remain in our understanding of the factors that determine whether or not a substance is an allergen, and if so, what makes it a respiratory sensitisier. Evidence that a substance can lead to respiratory hypersensitivity will normally be based on human experience. Hypersensitivity is normally seen as asthma, but other hypersensitivity reactions such as rhinitis or extrinsic allergic alveolitis are also considered. The condition will have the clinical character of an allergic reaction. However, immunological mechanisms do not have to be demonstrated according to the CLP criteria. In case there is evidence available that the substance induces asthma-like symptoms by irritation only, these substances should not be considered as respiratory sensitisers.

R.7.3.11.2 Concluding on suitability for Classification and Labelling

The CLP Regulation specifies that respiratory sensitisation should be allocated into sub-categories (i.e. 1A or 1B) whenever possible. In case the data are not sufficient for sub-categorisation, the substance must be classified in the general Category 1 (for further information, see Section 3.4 of the Guidance on the Application of the CLP criteria).
R.7.3.11.3 Concluding on suitability for chemical safety assessment: dose-response assessment and potency

There is evidence that for both skin sensitisation and respiratory hypersensitivity dose-response relationships exist although these are frequently less well defined in the case of respiratory hypersensitivity. The dose of agent required to induce sensitisation in a previously naive subject or animal is usually greater than that required to elicit a reaction in a previously sensitised subject. Therefore the dose-response relationship for the two phases will differ. Little or nothing is known about dose-response relationships in the development of respiratory hypersensitivity by non-immunological mechanisms.

It is frequently difficult to obtain dose-response information from either existing human or animal data where only a single concentration of the test material has been examined. With human data, exposure measurements may not have been taken at the same time as the disease was evaluated, adding to the difficulty of determining a dose response.

R.7.3.11.3.1 Estimation of potency

The estimation of potency for respiratory sensitisation is currently (at the time of publication of this Guidance) solely based on human data (See Section 3.4.2.1 of the Guidance on the Application of the CLP criteria).

R.7.3.11.3.2 Derivation of a DNEL

Even though respiratory sensitisation might be considered to be a thresholded effect (induction and elicitation), currently available methods do not allow the determination of a threshold and establishment of a DNEL. Guidance on how to perform a qualitative safety assessment for respiratory sensitisers can be found in Section E.3.4.2 and Appendix R.8-10 of Chapter R.8 of the Guidance on IR&CSA.

R.7.3.11.4 Additional considerations

Chemical allergy is commonly designated as being associated with sensitisation of the respiratory tract (asthma, rhinitis and extrinsic allergic alveolitis). According to current knowledge respiratory sensitisation can be induced via both dermal and inhalation route. Therefore it is important for the risk management purposes that the exposures to both routes are prevented.

As the evidence for substance leading to respiratory hypersensitivity is normally based on human data it may be difficult to distinguish respiratory sensitisation from respiratory irritation as for both the clinical symptoms are similar.

R.7.3.11.5 Information not adequate

A Weight-of-Evidence approach, comparing available adequate information with the tonnage-triggered information requirements of REACH, may result in the conclusion that the requirements are not fulfilled. In order to proceed in further information gathering the assessment strategy given in Section R.7.3.12 can be adopted.
R.7.3.12 Assessment strategy for respiratory sensitisation

R.7.3.12.1 Objective / General principles

The objective of this assessment strategy is to give guidance on a stepwise approach to hazard identification with regard to the respiratory sensitisation endpoint. A key principle of the strategy is that the results of one study are evaluated before another is initiated. The strategy should seek to ensure that the data requirements are met in the most efficient and humane manner so that animal usage and costs are minimised.

R.7.3.12.2 Preliminary considerations

Careful consideration of existing toxicological data, exposure characteristics and current risk management procedures is recommended to ascertain whether the fundamental objectives of the assessment strategy (see above) have already been met. Other factors that might mitigate data requirements for the endpoint of interest, e.g. possession of other toxic properties, characteristics that make testing technically not possible, should also be considered.

R.7.3.12.3 Recommended approach

The below strategy for respiratory sensitisation assessment (Figure R.7.3–4) can be followed:
Figure R.7.3–4 Assessment strategy for respiratory sensitisation data*

In contrast to tests for skin sensitisation, the performance of tests for respiratory sensitisation is currently not required under REACH. Therefore the present strategy scheme depicts a strategy for evaluating existing data.

** Based on current knowledge there are no low molecular weight respiratory sensitisers that would not cause skin sensitisation (Kimber et al., 2007).

*** There is an indication that enzymes have the potential to cause respiratory sensitisation e.g. the Scientific Committee for Animal Nutrition (SCAN) states the following “Enzyme and microbial additives will be regarded as respiratory sensitisers unless convincing evidence to the contrary is provided” (http://ec.europa.eu/environment/archives/dansub/pdfs/enzymerepcomplete.pdf).

Therefore, it is advised to consider respiratory sensitisation potential in the case of an enzyme, even though the CLP Regulation does not require to classify all enzymes as respiratory sensitisers.

**** No scientifically validated/independently reviewed or adopted test methods are yet available.
R.7.3.13 References


Chapter R.7a: Endpoint specific guidance
Draft Version 5.0 – Public – February 2016


Chapter R.7a: Endpoint specific guidance
Draft Version 5.0 – Public – February 2016

1 Mekenyan O, Patlewicz G, Kuseva C, Popova I, Mehmed A, Kotov S, Zhechev T, Pavlov T,
2 Merckel F, Giménez-Arnau E, Gerberick GF and Lepoittevin JP (2013) In chemico
   evaluation of prohapten skin sensitizers: behavior of 2-methoxy-4-(13C)methylphenol in
   the peroxidase peptide reactivity assay (PPRA) as an alternative to animal testing.
   sensitization potential for benzaldehydes: can Schiff base formation explain everything?
   dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing
   potency based on in vitro data from KeratinoSens and kinetic peptide binding: global
6 OECD (2012) The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent
   Binding to Proteins, Part 1: Scientific Evidence, Series on Testing and Assessment No.168
   (ENV/JM/MONO(2012)10/PART1), available at:
   http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2
   012)10/part1&doclanguage=en
   and Safety Publications, Series on Testing and Assessment No. 194
   (ENV/JM/MONO(2014)4). Available at:
   http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2
   014)4&doclanguage=en
8 OECD (2016a) Guidance Document On The Reporting Of Defined Approaches And
   Individual Information Sources To Be Used Within IATA For Skin Sensitisation
   (ENV/JM/HA(2015)8). Available at: XXX
9 OECD (2016b) Guidance Document On The Reporting Of Defined Approaches To Be Used
   Within IATA (ENV/JM/HA(2015)7). Available at: XXX
10 Patlewicz G, Mekenyan O, Dimitrova G, Kuseva C, Todorov M, Kotov S, Stoeva S and
    Donner EM (2010) Can mutagenicity information be useful in an Integrated Testing
11 Patlewicz G, Kuseva C, Kesova A, Popova I, Zhechev T, Pavlov T, Roberts DW and
    Mekenyan O (2014) Towards AOP application--implementation of an integrated approach
    to testing and assessment (IATA) into a pipeline tool for skin sensitization. Regul Toxicol
    Pharmacol 69:529-45.
   confidence in the development and evaluation of read-across. Regul Toxicol Pharmacol
   72:117-33.
13 Pauluhn J and Poole A (2011) Brown Norway rat asthma model of diphenylmethane-4,4'-
   diisocyanate (MDI): determination of the elicitation threshold concentration of after
14 Pauluhn J (2014) Development of a respiratory sensitization/elicitation protocol of
   toluene diisocyanate (TDI) in Brown Norway rats to derive an elicitation-based
   occupational exposure level. Toxicology 319:10-22.


Chapter R.7a: Endpoint specific guidance
Draft Version 5.0 – Public – February 2016


Appendices R.7.3-1 to 4 to Section R.7.3
Appendix R.7.3-1 Literature models and \textit{in silico} tools for skin sensitisation

Content of Appendix R.7.3-1:

- Data and (Q)SAR models in scientific publications
- Data and models included in \textit{in-silico} tools

DATA AND (Q)SAR MODELS IN SCIENTIFIC PUBLICATIONS

Data

Peer reviewed publications are a valuable source for skin sensitisation data. Cronin and Basketter (1994) published the results of over 270 \textit{in vivo} skin sensitisation tests (mainly from the guinea pig maximisation test). All data were obtained in the same laboratory and represent one of the few occasions when large amounts of information from corporate databases were released into the open literature. A larger database of animal and human studies for 1034 compounds is described by Graham \textit{et al.} (1996), the MCASE database. Schlede \textit{et al.} (2003) reports data on 244 chemicals. These data have been assessed and expert judgement on the potency ranking of 244 chemicals with contact allergenic properties based on data on humans and results of animal tests is provided in the paper.

A comparatively large number of data have been published for the local lymph node assay; examples include publications by Ashby \textit{et al.} (1995), Gerberick \textit{et al.} (2005) and Kern \textit{et al.} (2010).

SARs

Some collections of structural alerts published in the literature have not (yet) been encoded into softwares (e.g. Gerner \textit{et al.} (2004). However, experts in skin sensitisation can still manually use these structural alerts to make considerations on the chemical of interest.

Local (Q)SAR models

Local models are (Q)SAR models developed for specific chemical classes or mode of actions. The majority of local models for skin sensitisation have been developed for direct-acting electrophiles using the relative alkylation index (RAI) approach. This is a mathematical model derived by Roberts and Williams (1982). It is based on the concept that the degree of sensitisation produced at induction, and the magnitude of the sensitisation response at challenge, depends on the degree of covalent binding (haptenation, alkylation) to carrier protein occurring at induction and challenge. The RAI is an index of the relative degree of carrier protein haptenation and was derived from differential equations modelling competition between the carrier haptenation reaction in a hydrophobic environment and removal of the sensitisier through partitioning into polar lymphatic fluids. In its most general form the RAI is expressed as:

\[ RAI = \log D + a \log k + b \log P \]  

Thus the degree of haptenation increases with increasing dose D of sensitisier, with increasing reactivity (as quantified by the rate constant or relative rate constant \( k \) for the reaction of the sensitisier with a model nucleophile) and with increasing hydrophobicity.
(as quantified by log P, P being the octanol/water partition coefficient). This RAI model has been used to evaluate a wide range of different datasets of skin sensitising substances. Examples include sulfonate esters (Roberts and Basketter, 2000), sulfones (Roberts and Williams, 1982), primary alkyl bromides (Basketter et al., 1992), acrylates (Roberts, 1987), aldehydes and diketones (Patlewicz et al., 2001; Patlewicz et al., 2002; Patlewicz et al., 2004; Roberts et al., 1999; Roberts and Patiewicz, 2002; Patlewicz et al., 2003).

This approach has shown that local models tend to be transparent, simple and mechanistically derived but are labour-intensive to develop and restricted to local areas of chemistry (Cronin et al., 2011).

The covalent hypothesis has served and continues to be the most promising way of developing mechanistically based robust QSARs. These are local in that their scope is characterised by a mechanistic reactivity domain as outlined in Aptula et al. (2005), Aptula and Roberts (2006), and Roberts et al. (2007a). An example of this type of mechanistic model has been recently published (Roberts et al., 2006). In the RAI model, log k, has been typically modelled by experimental rate constants, substituents' constants or molecular orbital parameters. More effort is needed to encode reactivity into descriptors, this could be achieved through the systematic generation of in vitro reactivity data as outlined by Aptula and Roberts (2006), Aptula et al. (2006), Schultz et al. (2006), Gerberick et al. (2004) and in the next section.

Global statistical models

Global Statistical models usually involve the development of empirical QSARs by application of statistical methods to sets of biological data and structural descriptors.

These are perceived to have the advantage of being able to make predictions for a wider range of substances. In some cases, the scope/domain of these models are well described, in most other cases a degree of judgement is required in determining whether the training set of the model is relevant for the substance of interest. Criticism often levied at these types of models is that they lack mechanistic interpretability. The descriptors might appear to lack physical meaning or are difficult to interpret from a chemistry perspective. The sorts of descriptors used may encode chemical reactivity/electrophilicity, e.g. LUMO (the energy of the lowest molecular orbital), and partitioning effects, e.g. Log P, but a more common place is that a large number of descriptors are calculated that encode structural, topological and/or geometrical information. A number of models have been reported in the recent literature; examples include those developed using LLNA data (Devillers, 2000; Estrada et al., 2003; Fedorowicz et al., 2004; Fedorowicz et al., 2005; Li et al., 2005; Miller et al., 2005; Ren et al., 2006; Li et al., 2007; Golla et al., 2009; Chaudhry et al., 2010).

DATA AND MODELS INCLUDED IN IN-SILICO TOOLS

Data

There is a number of computational tools and databases available to facilitate the search and retrieval of skin sensitisation data for the target chemical or its analogues. Example of such databases and tools are the OECD QSAR Toolbox (http://www.qsartoolbox.org/), Chemfinder (www.chemfinder.com), ChemID (http://chem.sis.nlm.nih.gov/chemidplus/), NICEATM LNA Database (http://ntp.niehs.nih.gov/pubhealth/evalatm/test-method-evaluations/immunotoxicity/nonanimal/index.html) and DssTox.
Some of the available search engines are linked to databases (through hyperlinks and indexes) whereas others like DssTox provide a repository of available QSAR datasets which can be downloaded for subsequent use in appropriate QSAR/database software tools.

**In-silico tools**

**OECD QSAR Toolbox**

The OECD QSAR Toolbox software ([http://www.qsartoolbox.org/](http://www.qsartoolbox.org/), current version 3.3) encodes several mechanistic and skin sensitisation endpoint specific databases and profilers. They allow the user to group substances that share common structural alerts and to predict their skin sensitisation potential via read-across. ECHA has published illustrative examples on how to make skin sensitisation read-across predictions using the OECD QSAR Toolbox ([https://echa.europa.eu/documents/10162/21655633/illustrative_example_qsar_part2_en.pdf](https://echa.europa.eu/documents/10162/21655633/illustrative_example_qsar_part2_en.pdf)).

The two dedicated databases for skin sensitisation are “Skin sensitisation”, which includes 1 036 substances and 1 573 experimental data points (includes the OASIS skin sensitisation database and the Liverpool John Moores University skin sensitisation database) and “Skin sensitisation ECETOC”, with 39 substances and 42 experimental data points. ECHA Chem database, which collects the information found in REACH dossiers, contains also data on skin sensitisation.

There are four relevant profilers for skin sensitisation. They are all based on protein binding. Three of these profilers can be found under the general mechanistic profiler branch: Protein binding by OASIS v1.3, Protein binding by OECD, Protein binding potency. The fourth profiler is under the endpoint-specific branch: Protein binding alerts for skin sensitisation by OASIS v1.3.

The users can use profilers for the identification of analogues based on mechanistic commonalities and retrieve experimental information from the dedicated databases. Several data gap filling techniques can be used to predict skin sensitisation for the substance of interest: read-across, trend analysis and QSAR models.

The OECD QSAR Toolbox also encodes an Adverse Outcome Pathway (AOP) for skin sensitisation. This is the first attempt in the QSAR Toolbox to allow predictions through AOPs, and at this stage it is premature to advise the use of the AOP functionality within the OECD QSAR Toolbox for predicting skin sensitisation.

**Expert systems**

Software like VEGA or Toxtree are free to download and use. There are also several commercial (Q)SAR models for skin sensitisation available. Examples include TOPKAT, CASE, Derek Nexus (DN), TIMES (TIssue MEtabolism Simulator), Molcode, HazardExpert, and probably others.

- **Statistical Models:**

**Toxtree** allows the user to estimate toxic hazard by applying a decision tree approach. It includes the “Skin sensitisation reactivity domains” plug-in for the identification of mechanisms of toxic action using a SMARTS pattern based approach. It is important to
note that the alerts are meant to provide grouping into reactivity mode of action and do not predict skin sensitisation potential.

**TOPKAT** (included in Discovery Studio package) marketed by BIOVIA Foundation (formerly Accelrys Enterprise Platform ‘AEP’) is a suite of two models: one for Non-sensitisers vs. Sensitisers and the other for Weak/Moderate vs. Strong sensitisers. The first model calculates the probability of a chemical structure to be a sensitiser. If the probability is greater than or equal to 0.7, the substance is predicted to be a sensitiser, a non-sensitiser would have a probability of less or equal to 0.30. The second model applies to structures predicted as sensitisers by the first model and resolves the potency: weak/moderate vs. strong, where a probability of 0.7 or more indicates a strong sensitiser and a probability below 0.30 indicates a weak or moderate sensitiser. Probability values between 0.30 and 0.70 are referred to as indeterminate. An optimum prediction space algorithm ensures that predictions are only made for substances within the applicability domain of the model. Please note that the models are all based on the guinea-pig maximization test (Enslein *et al*., 1997; http://accelrys.com/solutions/scientific-need/predictive-toxicology.html).

**CASE** methodology and all its variants were developed by Klopman and Rosenkranz. There is a multitude of models for a variety of endpoints and hardware platforms. The CASE approach uses a probability assessment to determine whether a structural fragment is associated with toxicity (Cronin *et al*., 2003). The MCASE models (currently CASE Ultra) that have been developed for skin sensitisation are described further in primary articles (Gealy *et al*., 1996; Graham *et al*., 1996; Johnson *et al*., 1997). There are two sensitisations modules available for purchase from MultiCase Inc (Ohio, USA) (http://www.multicase.com/case-ultra-models). In addition the (Q)SAR estimates for one MCASE skin sensitisation model are included in the Danish Environmental Protection Agency (EPA) (Q)SAR database (http://qsar.food.dtu.dk/).

**VEGA** platform, freely available for download (http://www.vegaqsar.eu), incorporates a model (Chaudhry *et al*., 2010) developed using an Adaptive Fuzzy Partition (AFP) algorithm based on eight descriptors. The AFP assigns the substances to two classes, sensitisers and non-sensitisers. An in-depth assessment of the applicability domain of the prediction, mainly based on similarity with substances in the training set of the model, is also provided.

- Knowledge-based systems:

**Derek Nexus** (DN) is a knowledge-based expert system created with knowledge of structure-toxicity relationships and an emphasis on the need to understand mechanisms of action and metabolism. It is marketed and developed by LHASa Ltd (Leeds, UK) a not-for-profit company and educational charity (http://www.lhasalimited.org/index.php).

Within DN (version 9), there are 361 alerts covering a wide range of toxicological endpoints. An alert consists of a toxicophore, a substructure known or thought to be responsible for the toxicity alongside associated literature references, comments and examples. The skin sensitisation knowledge base in DN was initially developed in collaboration with Unilever in 1993 using its historical database of guinea pig maximisation test (GPMT) data for 294 substances and contained approximately forty alerts (Barratt *et al*., 1994). Since that time, the knowledge base has undergone extensive improvements as more data have become available (Payne and Walsh, 1994). The current version contains about seventy alerts for skin sensitisation and some alerts for photoallergenicity (Barratt *et al*., 2000; Langton *et al*., 2006). The predictivity of DN for skin sensitisation was recently assessed by Guesne *et al* (2014). As a reminder, alert-based systems should not be assessed for their specificity and overall accuracy, contrary to discriminant models.
Hybrids:

The **TIssue MEtabolism Simulator (TIMES)** software has been developed to integrate a Skin metabolism Simulator (SS) with 3D-QSARs for evaluating reactivity of substances in order to predict their skin sensitisation potency (Dimitrov *et al.*, 2005). The current version of the simulator (version 2.27.1) contains more than 200 hierarchically ordered spontaneous and enzyme-controlled reactions. Covalent interactions of substances/metabolites with skin proteins are described by 47 alerting groups. 3D-QSARs (COREPA) are applied for some of these alerting groups. Characterisation and evaluation of TIMES-SS can be found in Patlewicz *et al.* (2007) and Roberts *et al.* (2007b), respectively. New research with TIMES includes the work of Patlewicz *et al.* (2014).

The **Danish QSAR database** contains a collection of pre-calculated predictions on allergic contact dermatitis for over 600 000 chemical structures based on models developed by the Danish Technical University. The predictions were obtained with three different software packages for toxicity modelling: CASE Ultra, SciQSAR and Leadscope. The three models include an automatic applicability domain check, so that, for each prediction, the outcome of the model (positive/negative) and whether the substance is within the domain of the model are indicated. In addition, a fourth prediction based on a consensus system between the three other approaches is provided. The consensus model offers increased predictivity and domain coverage. The database is freely available at: http://www.qsar.food.dtu.dk/.

For expert systems such as DN, TOPKAT etc., the training sets and, to some extent, the algorithms or descriptors used are often kept latent within the software. Some supporting information is provided on the robustness and relevance for a given prediction. For example, within DN it is possible to see representative example substances and explanations of the mechanistic basis for the SAR developed.

TOPKAT supports the users in assessing the reliability of the prediction by: 1) evaluating if the substance falls into the applicability domain of the model (based on structural fragments and descriptors), 2) checking if the substance is present in its database, and 3) identifying analogues of the target substance based on chemical similarity. Similar functionalities and features are present in many of the other commercial expert systems available.

Although the main factors driving skin sensitisation (and therefore the (Q)SARs) is the underlying premise of the electrophilicity of a substance, other factors such as hydrophobicity encoded in the octanol/water partition coefficient (log P) may also be considered as playing a role in the modification of the sensitisation response observed. Within DN, an assessment of the likely skin penetration ability is made using the algorithm by Potts and Guy. This relates the $K_p$ value to log P and MW (Potts and Guy, 1992). It is then possible to rationalise the output in terms of bands of penetration potential. Some methods for assessing percutaneous absorption have been described in Howes *et al.* (1996).

**References**


Chapter R.7a: Endpoint specific guidance
Draft Version 5.0 – Public – February 2016


Appendix R.7.3-2 Template for the reporting of the individual information sources for a non-animal test method when using non-validated and/or non-adopted test methods

The following reporting format (Table R.7.3-4) should be considered especially when information is generated by non-animal test methods to fulfil the REACH information requirement for skin sensitisation. The use of this reporting template is very important in case (a) test method(s) is (are) used which has (have) not been considered scientifically valid in a international validation study and/or there is no internationally adopted test guideline available.

In case a test method has an internationally adopted test guideline available this template can also be used, however some of the points described below can already be included in the test guideline itself, hence detailed reporting of such (an) information source(s) is usually not needed. The reporting of each individual information source needs to be included in a separate endpoint study record (ESR) of the IUCLID dossier, i.e. one ESR per individual information source should be filled in.

Note: this reporting template has been modified based on the OECD template for the reporting of individual information sources (OECD, 2016) to be specific to the skin sensitisation endpoint and REACH information requirements.
Table R.7.3-4 Template for the reporting of the individual information sources describing a non-animal test method when using non-validated and/or non-adopted test methods to fulfill the REACH information requirement for skin sensitisation

<table>
<thead>
<tr>
<th>Name of the information source</th>
<th>Provide the name of the information source and the acronym (if applicable)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mechanistic basis including AOP coverage</td>
<td>Describe which key event of skin sensitisation AOP is addressed by the information source. A description of the extent to which the mechanistic basis of the information source relates to the chemical/biological mechanism covered by the (key) event should be provided.</td>
</tr>
<tr>
<td>Description</td>
<td>Provide a short description of the information source including the experimental system used and any relevant aspect of the procedure (e.g. time of exposure of the experimental system with the test substance, number of doses/concentrations tested, number of replicates, concurrent testing of control(s) and vehicle(s), laboratory instruments/techniques used to quantify the response).</td>
</tr>
<tr>
<td>Response(s) measured</td>
<td>Specify the response(s) measured by the information source and its measure (e.g. in chemico binding to synthetic peptides, expressed as % of peptide depletion).</td>
</tr>
<tr>
<td>Prediction model</td>
<td>Indicate whether there is a prediction model associated to the information source and its purpose. Briefly describe the prediction model and provide a reference to a paper or document where the prediction model is described (if available).</td>
</tr>
<tr>
<td>Metabolic competence (if applicable)</td>
<td>Specify whether the information source encompasses any metabolically competent system/step and, to the extent possible, how this relates to the situation in vivo.</td>
</tr>
<tr>
<td>Status of development, standardisation, validation</td>
<td>Indicate whether the information source is: a) an officially adopted (standard) test method (e.g. a test method covered by an OECD Test Guideline); b) a validated but non-standard test method; c) a test method undergoing formal evaluation (e.g. prevalidation, validation, others); d) a non-validated test method widely in use; e) a non-validated test method implemented by a small number of users.</td>
</tr>
<tr>
<td>Technical limitations and limitations with regard to applicability domain</td>
<td>Indicate the substance(s) and/or chemical categories (e.g. based on physico-chemical properties or functional groups) for which the information source has been shown not to be applicable because of technical limitations, e.g. highly volatile chemicals, poorly water soluble chemicals, solid materials, interference of the chemical with the detection system (e.g. coloured or autofluorescent chemicals interfering with spectrophotometric analysis). Indicate whether the information source is technically applicable to the testing of multi constituent-substances, UVCBs and mixtures. In addition indicate the substance(s) and/or chemical categories for which the information source has been experimentally shown to yield incorrect and/or unreliable predictions with respect to the reference classifications (e.g false negative predictions with substances requiring metabolic activation, high false positive rate for alcohols).</td>
</tr>
<tr>
<td>Strengths and Weaknesses</td>
<td>Provide an indication of the strengths and weaknesses of the information source, compared to existing similar non-testing or testing methods, considering among others the following aspects: a) extent of mechanistic information provided and relevance (i.e. measurement of various responses in the same experimental model, limited or good coverage of the mechanisms at the basis of the effect</td>
</tr>
<tr>
<td>being investigated, predictive of responses in humans);</td>
<td></td>
</tr>
<tr>
<td>b) level of information provided (single-point estimate or dose-response information);</td>
<td></td>
</tr>
<tr>
<td>c) level of performance (e.g. higher or lower reproducibility, predictive capacity);</td>
<td></td>
</tr>
<tr>
<td>d) extent of domain of applicability;</td>
<td></td>
</tr>
<tr>
<td>e) number of substances with published information.</td>
<td></td>
</tr>
</tbody>
</table>

**Reliability (within and between laboratories) (if applicable)**
Describe the level of reliability of the information source (i.e. the agreement among results obtained from testing the same substances over time using the same protocol in one or multiple laboratories) and to what extent this has been characterised including the number of substances used for the assessment.

**Predictive capacity (if applicable)**
Describe the extent to which the information source predicts the key event of interest (as reported in scientific publications and as determined in validation studies). Express the predictive capacity in terms of sensitivity, specificity and accuracy if applicable or by other goodness-of-fit statistics (e.g. linear correlation analysis). Include the number of substances used in this assessment and their predictions using the reference method.

**Proposed regulatory use**
Indicate the proposed regulatory use of the information source (e.g. stand-alone full replacement method, partial replacement method, screening method, others).

**Potential role within a Testing and Assessment Strategy**
Indicate the potential weight the information source is expected to carry within a structured approach to data integration (if applicable) and/or within a Testing and Assessment Strategy, and for which specific purpose the information source can potentially be used on its own.

**Reference**

OECD (2016) Guidance Document On The Reporting Of Structured Approaches To Data Integration And Individual Information Sources Used Within IATA For Skin Sensitisation. Available at: XXX
Appendix R.7.3-3 Reporting format for defined approaches to testing and
assessment based on multiple information sources

This template aims to provide advice for the reporting of defined approaches to testing
and assessment to be used within IATA and for the integration of the individual
information sources used to fulfil the REACH information requirement for skin
sensitisation. The reporting of the defined approaches to testing and assessment and the
conclusions obtained from them should be included in the dossier, i.e. as an attachment
to the endpoint summary record of skin sensitisation of the IUCLID dossier.

Note: the reporting template is based on the OECD reporting format for reporting
defined approaches as described in Annex I of the OECD Guidance Document on the
Reporting of Defined Approaches To Be Used Within Integrated Approaches To Testing
and Assessment (IATAs) (OECD, 2016a), however the template has been adapted to
REACH specific purposes.

1 Summary
Summarise the information in the reporting format in order to provide a concise
overview of the proposed defined approach.

2 General information
2.1 Identifier: Provide a short and informative title for the structured approach.
2.2 Reference to main scientific papers: List the main bibliographic references (if
any).

3 Endpoint addressed
Specify the endpoint (here skin sensitisation). Also specify related properties that have
been measured or predicted by the proposed approach and indicate whether these
address (or partially address) an endpoint, or key event being predicted by an existing
test guideline.

4 Definition of the purpose
Default: meeting the REACH information requirement for skin sensitisation (Annex VII,
8.3) and the relevant classification and/or risk assessment obligations.

5 Rationale underlying the construction of the defined approach
Describe the rationale used to construct the defined approach. This should include an
assessment of the linkage of the individual information sources used within the approach
to the known chemical and the key events being predicted. The reason for the choice of
(a) specific information source(s)/test(s) addressing (a) specific key event(s) possibly in
the light of other existing similar information sources should be provided. In case non-
guideline information source for a key event is used, for which an existing test guideline
is available (e.g. EU or OECD), should be justified.

6 Description of the individual information sources used within the approach
(see Annex II of OECD GD XXX (OECD, 2016a) and Appendix R.7.3.2 of this Guidance)
List the information sources employed within the proposed defined approach (e.g.
physico-chemical properties, non-testing (in silico) methods and testing (in chemico, in vitro, in vivo) methods, including the response(s) measured and the respective measure(s) (e.g. in chemico binding to synthetic peptides, expressed as % peptide depletion). Detailed descriptions of each in chemico, in vitro, and in vivo method should be provided using the endpoint study records (ESRs) in IUCLID (i.e. one ESR per individual information source).

In addition, when QSAR models are used the QSAR Model Reporting Format (QMRF) should be provided and individual predictions, if applicable, should be reported using the QSAR Prediction Reporting Format (QPRF) and included in the ESR of the IUCLID. Both reporting formats are accessible at: `https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/qsar_tools/QRF`.

### 7 Data interpretation procedure applied

Describe the data interpretation procedure (DIP) used. Indicate whether the DIP output is qualitative or quantitative. If possible, provide a workflow to illustrate the manner in which the DIP should be applied.

### 8 Substances used to develop and test the DIP

#### 8.1 Availability of training and test sets:
Indicate whether a training set (i.e. chemical data used in the development of the DIP) and test set (i.e. chemical data used to evaluate the DIP) are available (e.g. published in a paper, stored in a database) or appended to this Reporting format. If they are not available, explain why. Example: "It is available and attached"; "It is available and referenced"; "It is not available because the data set is proprietary"; "The data set could not be retrieved".

#### 8.2 Selection of the training set and test set used to assess the DIP:
If the training set and test set are available please describe the rationale for their selection (e.g. availability of high quality in vivo data for the endpoint being predicted, coverage of the range of effects observed in vivo, coverage of diverse physico-chemical properties, coverage of structural diversity, others).

#### 8.3 Supporting information on the training and test sets:
If the training and/or the test sets are available, append them as supporting information, preferably in the form of an Excel table. The following information on both sets should be reported where available and to the extent possible: a) chemical name (common and/or IUPAC); b) CAS and/or EC numbers; c) in case of multi-constituent or UVCBs, report the composition to the extent possible; d) reference data or classifications(s) for each substance (e.g. in vivo data); e) data from the individual information sources used in the defined approach; f) final result/prediction for each substance.

#### 8.4 Other information on training and test sets:
If the training and/or the test sets are not available for inclusion as supporting information, indicate any other relevant information about the training and/or test sets (e.g. number and type of compounds). This will be useful to gain an appreciation of e.g. the chemical coverage.

### 9 Limitations in the application of the defined approach

Indicate the type(s) of substances, in terms of their physico-chemical properties, structures and functional groups, for which the approach is considered not to be applicable because of technical constraints in the testing of those substances (e.g. poor solubility, interference with detection system etc.) or because such substances have been found to give unreliable results (e.g. non-reproducible results when the defined approach is applied multiple times or because of wrong predictions with respect to reference classification).
10 Predictive capacity of the approach

Provide an indication of the extent to which the defined approach predicts the skin sensitisation potential by considering the associated information sources and by excluding chemical types identified in the limitations above. Express the predictive capacity in terms of sensitivity, specificity and concordance, if applicable, or by other goodness-of-fit statistics (e.g. linear correlation analysis). Rationalise to the extent possible potential misclassifications (i.e. substances under-predicted or over-predicted with respect to the reference classification) or unreliable predictions for substances that are considered to be covered by the applicability domain of the approach.

11 Consideration of uncertainties associated with the application of the defined approach

11.1 Sources of uncertainty

Describe the uncertainties which are considered to be associated with the application of the defined approach by capturing the source(s) of uncertainty that result(s) from:

1. The DIP’s structure
   - What are the uncertainties related to the chosen DIP’s structure?
   - How does the DIP’s coverage or weighing of the AOP key events affect your confidence in the overall prediction?
   - How does your confidence in the DIP’s prediction vary between different substances?

2. Information sources used within the defined approach
   - How does the variability in approach information source’s data for a given substance (i.e. reproducibility) affect your confidence in the DIP’s prediction?

3. Benchmark data used
   - How does the inherent variability of the reference data (e.g. LLNA, human) affect your confidence in the DIP’s prediction?

4. Others sources

11.2 Impact of uncertainty on the DIPS’s prediction

Consider how these sources of uncertainty affect the overall uncertainty in the final prediction in the context of the defined approach application.

12 References

List relevant references, weblinks etc., including those describing the structured approach itself (also provided under Section 2 on General Information).

References

OECD (2016a) Guidance Document On The Reporting Of Defined Approaches To Be Used Within IATA. Available at: XXX

OECD (2016b) Guidance Document On The Reporting Of Defined Approaches And Individual Information Sources To Be Used Within IATA For Skin Sensitisation. Available at: XXX
Appendix R.7.3-4 Potency estimation for skin sensitisation

Background

The estimation of potency for skin sensitisers is important for protecting workers and consumers. According to the CLP Regulation skin sensitisers can be divided in two classes i.e. Extreme and strong sensitisers (Cat 1A) and moderate sensitisers (Cat 1B).

Where non-animal testing methods, e.g. in chemico/in vitro test methods, are used to fulfil the information requirements of the REACH Regulation, it should be noted that no widely accepted approach is currently available to assess the potency leading to sub-categories according to the CLP Regulation. In contrast, the current standard requirement under REACH for fulfilling the requirements of Annexes VII is an in vivo test method, i.e. the LLNA, which allows the assessment of skin sensitisation potency and subsequent sub-categorisation of chemicals (Cat 1A vs. Cat 1B) when EU method B.42/OECD TG 429 is used.

Identification of strong and extreme skin sensitisers and subsequent classification into sub-category 1A according to CLP is important for the protection of human health. This is due to the fact that, depending on the skin sensitisation potency, different concentration limits are to be applied, i.e. for Cat 1 and Cat 1B the generic concentration limit (GCL) is 1%, for Cat 1A (extreme or strong) the GCL is 0.1% and for extreme sensitisers a specific concentration limit (SCL) of 0.01% is recommended according to the Guidance on the Application of the CLP criteria (for further information, see Section 3.4 of the Guidance on the Application of the CLP criteria). In short, this may lead to the situation that mixtures containing potent sensitisers are not correctly classified, if general Cat 1 is used and the GCL of 1% is applied instead of 0.1% or 0.01%. This would lead in lowering of the safety levels, which may lead to an increased incidence of human sensitisation to more potent sensitisers.

In this appendix different approaches for assessing skin sensitisation potency are reviewed, in the context of the capabilities and limitations of the current standard in vivo method.

Uncertainty of the LLNA

One of the challenges for alternative (replacement) approaches when trying to predict an apical endpoint is the uncertainty associated with the inherent variability of the animal-based reference data. In the case of skin sensitisation, the variability of the LLNA defines an upper limit for the predictivity of alternative methods, especially when trying to predict potency classes which can be significantly affected by solvent effects (Basketter et al., 2001). Recently this was confirmed by Hoffmann (2015) who analysed potential vehicle-related variability with respect to the five category classification system used by ECETOC (2003). A retrospective analysis by ICCVAM (2011) comparing LLNA data with human data showed that 48% of Cat 1A sensitisers were underclassified according to
LLNA-based CLP criteria, whereas 28% of Cat 1B sensitisers were misclassified, and 59% of non-sensitisers were overclassified\textsuperscript{16}.

In a more recent study of LLNA variability by EURL ECVAM (Dumont et al., submitted), the number of studies predicting skin sensitisation potency (UN GHS / EU CLP hazard sub-categorisation) was considered, analysing the variability per substance (irrespective of the solvent) and per substance-solvent combination. Consistent with the ICCVAM analysis, the results showed that the inherent variability of LLNA is less significant for Cat 1A chemicals, but more significant for Cat 1B and non-sensitising chemicals\textsuperscript{17}.

**Review of approaches for predicting skin sensitisation potency (classes and EC3)**

At this point in time, the results portrayed in the subsequent paragraphs, while published in the peer-reviewed literature, have not been endorsed or validated by bodies such as EURL ECVAM or the OECD. It is the purpose of this section to provide an overview of the available literature, but ECHA cannot take responsibility for the correctness of the reported results, which may or may not be used at the registrant’s own risk. In any case, the reader is advised to screen the recent literature for the latest developments.

Several attempts to combine the EURL-ECVAM validated methods Direct Peptide Reactivity Assay (DPRA) (Gerberick et al., 2004), KeratinoSens\textsuperscript{TM} (Natsch and Emter 2008; Emter et al., 2010), h-CLAT (Ashikaga et al., 2006; Nukada et al., 2011, 2012), and others (Ade et al., 2006; Piroird et al., 2015; Python et al., 2007, Ramirez et al., 2014), have been made (Natsch et al., 2009, 2015; Bauch et al., 2012; Hirota et al., 2013, 2015; Jaworska et al., 2013, 2015; Tsujita-Inoue et al., 2014; Urbisch et al., 2015) to predict skin sensitisation potency. A summary of the most promising approaches follows.

Natsch et al. made one of the first attempts (Natsch et al., 2009) to predict skin sensitisation potency of a dataset of 116 chemicals with a combination of non-animal methods. The model incorporates four descriptors accounting for different key events of the adverse outcome pathway (AOP). They used a system based on scores (Jowsey et al., 2006) from single methods, i.e. peptide reactivity as a surrogate for protein binding, the induction of antioxidant/electrophile responsive element dependent luciferase activity as a cell-based assay, and an in silico prediction (Dimitrov et al., 2005) for skin sensitisation potential (TIMES-SS). The relationship between scores and potency was not sufficient ($R^2=0.423$) to properly distinguish between potency classes. However, extreme

\textsuperscript{16} More specifically, the ICCVAM analysis revealed that 48% of known strong human sensitisers (showing positive responses at an induction dose per skin area of $\leq 500 \mu g/cm^2$, i.e., Cat 1A according to CLP criteria for human data) showed an EC3 $> 2\%$ (41%) or were negative in the LLNA (7%), therefore being underclassified according to LLNA CLP criteria. Furthermore, of the non-strong human sensitisers (Cat 1B according to CLP criteria for human data), 6% were overpredicted as Cat 1A and 22% were underpredicted as non-sensitisers by the LLNA. Finally, 7% and 52% of the NS in humans were overclassified as Cat 1A and Cat 1B, respectively, by the LLNA.

\textsuperscript{17} More specifically, the EURL ECVAM analysis revealed that for chemicals having at least one LLNA study giving a non-sensitiser result, only 66% of all available LLNA studies (performed with the same solvent) identified these chemicals consistently as non-sensitiser. The rest of the studies classified them as either Cat 1B (23%) or Cat 1A (11%). For chemicals having at least one LLNA study giving a Cat 1B result, only 68% of all available LLNA studies (performed with the same solvent) classified them consistently as 1B. The remaining studies (32%) classified the chemicals in equal proportion to non-sensitisers or Cat 1A sensitisers. The classification of chemicals having at least one LLNA study giving a Cat 1A result was found to be less variable, with 79% of the studies (performed with the same solvent) classifying these chemicals as 1A (15% Cat 1B and 6% non-sensitisers).
sensitisers (not strong) could be easily distinguished from weak sensitisers and non-
sensitisers, as 57 out of 59 weak/NS had average predicted EC3 values <3. Due to the
large overlap between strong, moderate and weak sensitisers, the published model is not
recommended for distinguishing between Cat 1A and 1B chemicals.

Other approaches (also leading to over-prediction of moderates as compared to the
LLNA) have been based on different ways of integrating data, i.e. artificial neural
networks (Hirota et al., 2013, 2015; Tsujita-Inoue et al., 2014), decision trees, score-
based models (Nukada et al., 2013; Takenouchi et al., 2015), and mechanistic domain
based regression models (Natsch et al., 2015). Consistent with the variability in the
LLNA, these models, which integrate validated in vitro methods or similar cell based
assays with or without in silico descriptors, have overall accuracies in predicting skin
sensitisation potency categories ranging from 70% to 85%. It is worth noting, though,
that these methods usually try to predict a large number of compounds, 244 chemicals in
the most extreme case (Natsch et al., 2015), which is more difficult than predicting
smaller numbers (≤ 50 chemicals) because the applicability domain is much smaller and
the models are more local.

One of the conclusions that can be drawn from these studies is that some mechanistic
domains are easier to predict than others. For instance, Natsch et al. found better
predictivity for epoxides and nucleophile substitution domains, with $R^2 > 0.80$. Aldehydes
were the worst predicted group, with $R^2 = 0.21$. The parameters with most prediction
power also vary across domains and this is valuable information for the further
development of models: kinetic rate constants were found to be the most prominent
predictor for the $S_N2/S_NAr$ domain; KeratinoSens™ EC3 was the best predictor for Michael
acceptors, and cytotoxicity and vapour pressure for epoxides. Given the large number of
compounds predicted by Natsch et al. method and the fact that it can distinguish Cat 1A
chemicals from the rest (sensitivity=0.70 if Cat 1A are distinguished from 1B & NS,
results not shown), the method could be uses in a Weight-of-Evidence approach or as the
basis for grouping and read-across.

Important information on mechanistic domains was also provided by Urbisch and
colleagues (Urbisch et al., 2015). They observed that ARE based assays, like
KeratinoSens™ and LuSens, did not perform well at predicting the skin sensitisation
potential of acylating agents, and Schiff base could not be well classified with any of the
methods investigated (DPRA, KeratinoSens™, LuSens, h-CLAT, mMUSST). This
information can be of high value for those methods (shown below) that need further
testing.

In contrast to in vitro based models, Dearden et al. developed QSARs (Dearden et al.,
2015) to predict LLNA EC3 values from purely computational descriptors (CODESSA,
MOE, and winMolconn41 software). They divided a dataset of 204 sensitisers with LLNA
EC3 into 10 mechanistic domains and derived QSARs for each domain. They obtained
good predictivities$^{18}$ ($R^2 > 0.83$) for 7 out of 10 domains and $Q^2 > 0.79$ for 6 out of 10
domains – $Q^2$ is the $R^2$ equivalent for the test set. The domains with best predictions
were in this order: oxidation potential ($R^2 = 0.91$), acyl transfer ($R^2 = 0.90$), Michael
acceptor ($R^2 = 0.83$), pro-Michael acceptor ($R^2 = 0.83$), $S_N2$ ($R^2 = 0.82$), and Schiff base +

---

$^{18}$ Dearden et al. transformed the EC3 (g/ml) into an equivalent parameter with molar units named SSP. In principle, this transformation is convenient from the point of view that if EC3 is expressed in g/ml, two substances that are equally sensitisers and have different molecular weight would have significantly different EC3 values and might fall into different potency classes. The SSP corrects for this as it is not dependent on the molecular weight of the substances. In practice, the transformation from EC3 (g/ml) to SSP (M) had no effect on the potency class definitions for the compounds studied and both parameters were strongly correlated ($R^2 = 0.96$).
pro-Schiff base forms (R²=0.82). The Sₜ₁, pro-Sₜ₂, and SₜAr domains contained too few chemicals to develop meaningful QSARs. The publication does not use the predictions to classify into potency classes, but given that the correlation with LLNA-EC3 is so high, a good performance is expected. While this model might not be adequate as a standalone prediction method, it has high potential to be used in a Weight-of-Evidence approach, especially given its computational and, therefore, fast and reproducible character. It could also be used as a way of grouping chemicals for read-across.

Another partially in silico based model with high predictivity is the Bayesian network (BN) proposed by Jaworska et al. (Jaworska, 2011; Jaworska et al., 2013, 2015). The BN model integrates several sources of information and is capable of guiding the tests that need to be conducted to obtain a higher confidence in the prediction. The second version of the model (ITS-2; Jaworska et al., 2013) was trained on 124 chemicals (training set) and tested on 21 chemicals, predicting correctly 95% and 86% of the chemicals in the test set for hazard and LLNA potency classes, respectively. The LLNA classes were reduced to four as strong and extreme sensitisers were merged under strong sensitisers. However, the model could be used to predict CLP categories if weak and moderate are considered as Cat 1B chemicals. The model uses different in silico (TIMES-SS), in vitro (KeratinoSens™, mMUSST U937, skin permeation model readouts), in chemico (DPRA readouts), and octanol-water partition coefficient (Kow) parameters. The model uses all the provided parameters to derive a probability value of belonging to each of the LLNA potency groups. However, the model also works with data gaps, allowing one to estimate how much certainty in the prediction would be gained if specific test data were obtained before performing such test. In general the model performed very well, although better for non-sensitisers and weak compounds. The performances were: non-sensitisers (93, 100), weak (89, 93), moderate (75, 83), and strong/extreme (81, 73). The values in parenthesis correspond to the % area under the curve (AUC) for the training set and test set, respectively (see original publication for further information on the scoring). The model includes a correction factor for Michael acceptors which were systematically over-predicted.

A third improved version of the model (ITS-3) has just been recently published (Jaworska et al., 2015). The system was constructed with an aim to improve precision and accuracy for predicting LLNA potency beyond ITS-2, by improving representation of chemistry and biology. Among novel elements are corrections for bioavailability both in vivo and in vitro as well as consideration of the individual assays’ applicability domains in the prediction process. The three validated alternative assays, DPRA, KeratinoSens and h-CLAT, representing the first three key events of the adverse outcome pathway for skin sensitisation, are all integrated in ITS-3. In this model, skin sensitisation potency prediction is provided as a probability distribution over four potency classes. The probability distribution is converted to Bayes factors to: 1) remove prediction bias introduced by the training set potency distribution and 2) express uncertainty in a quantitative manner, allowing transparent and consistent criteria to accept a prediction. The novel ITS-3 database includes 207 chemicals with a full set of in vivo and in vitro data. The accuracy for predicting LLNA outcomes on the external test set (n = 60) was assessed in three levels, and was found to be high (>90%) following the order: hazard (two classes) > GHS potency classification (three classes) > potency (four classes).

Another in vitro model is the epidermal-equivalent (EE) potency assay (Gibbs et al., 2013). The model uses 3D reconstituted human epidermis, cytotoxicity, and IL-1α (IL-1α2x) fold increase in order to predict potency (2 key event of the AOP). Its reproducibility and predictive capacities were evaluated in an international ring trial for 13 chemicals (Teunis, 2014). The model appears to be able to separate strong/extreme from weak/moderate sensitisers, with sensitivity=69% and specificity=84%. One of the advantages of this method is that it does not have solubility and stability issues in water like most in vitro methods since the chemicals can be tested neat. This model could offer
a means of differentiating between Cat 1A and Cat 1B compounds, but cannot on its own
be used to distinguish non-sensitisers from sensitisers.

The "classic" in vitro and in chemico assays to predict skin sensitisation do not always
predict LLNA potency with high accuracy even when combined with in silico models.
Newer, versions of methods that are gene-based seem more promising. One of these
methods (McKim et al., 2010) is the in vitro toxicity index (IVTI) developed by Cyprotex
in 2010, also known as SenCeeTox (McKim et al., 2012). It is based on a combination of
cell viability, direct and indirect chemical (peptide) reactivity, and ARE/EpRE- mediated
gene expression and was developed for use with human keratinocyte (HaCaT) cells and
human 3D skin models. It predicts four potency classes: extreme/strong, moderate,
weak, and non sensitisers. The model showed high specificity (92%) and sensitivity
(81%) in discriminating between extreme, strong, and moderate sensitisers from weak
and non-sensitisers and it was reported that 4 out of 4 extreme/strong, 70% of the
moderates, 79% of the weaks, and 73% of the non-sensitisers were correctly classified.
The model was tested on 97 compounds (39 in the training set and 58 in the test set)
and seems capable of identifying extreme/strong sensitisers (Cat 1A) compounds and
distinguishing them from non-Cat 1A compounds, even though it is not capable of
distinguishing weak from non-sensitisers.

Another method for predicting LLNA EC3 values is VITOSENS, which is based on a linear
combination of cell cytotoxicity (IC20) caused by the test chemical and the fold change in
the expression of CCR2 (C–C chemokine receptor type 2) and the transcription factor
cAMP responsive element modulator (CREM) (Lambrechts et al., 2010). Cell cultures from
two different cord blood donors (three in case of discordant results) are used. The
authors showed a very high correlation (Pearson R²=0.79, Spearman rank correlation
coefficient=0.91) between the predicted and the EC3 values for 15 chemicals. The
method properly predicts some pro-haptens showing some metabolic capabilities but
might not be adequate for extreme cytotoxicants as two compounds were considered
outliers due to too high cytotoxicity.

The gene allergen rapid detection (GARD) method uses differentially regulated transcripts
of 200 genes in MUTZ-3 cells (as surrogate of primary human dendritic cells) after
exposure to predict skin sensitisation (Johansson et al., 2011, 2013). A support vector
machine model trained on 38 chemicals performs the final skin sensitisation prediction.
The authors did not provide statistics on potency classification performance, but PCA
plots showed clusters of chemicals belonging to the same potency groups. The model has
recently been reported (Forreryd et al., 2015) to predict respiratory sensitisation hazard
for 30 chemicals. However it should be pointed out that the regulated transcripts of the
200 genes in MUTZ-3 cells are not public, similarly to a proteomic assay (with
approximately 110 candidate proteins to reliably identify human skin sensitisers) also
using MUTZ-3 and human keratinocytes (Reisinger et al., 2015; Thierse et al., 2011;
Roggen et al., 2011).

It seems that the evolution of the two methods just described is the so-called SENS-IS
assay (Cottrez et al., 2015a). SENS-IS makes use of a series of over 60 genes that are
modulated during sensitisation either in mice or humans, and uses the modulation of
these genes in reconstructed human epidermis models (Episkin) to predict the skin
sensitisation potential and potency of chemicals. This set of sensitisation biomarker
genes was selected based on a thorough analysis of the genes modulated during the
sensitisation process in mice (LLNA), humans (blisters) or reconstructed human
epidermis, starting from a panel of over 900 target genes identified through data mining.
Fine analysis of their expression pattern indicated that it was the number of modulated
genes rather than the intensity of up-regulation that correlated best with sensitization
potential (Cottrez et al., 2015a). Thus, the model simply consists of determining the
number of genes that are up-regulated after exposure. A threshold determines whether
the chemical is predicted as sensitisier or non-sensitisier. A test chemical is considered to
be a skin sensitiser if it increases the expression of at least seven genes in either the so-
called "SENS-IS" gene set (consisting of 17 genes) or the "ARE" gene set (consisting of 21 genes). A third set of 23 genes is used to identify if the test chemical is irritant (if at least 15 out of these 23 genes are induced). The skin sensitisation potency of a test chemical is determined based on the lowest concentration at which it becomes positive in the SENS-IS assay. The test chemical is therefore considered to be an extreme sensitiser if it positive at 0.1%, strong if it is positive at 1%, moderate if it is positive at 10% or weak if it is positive at 50%.

The SENS-IS is capable of distinguishing skin sensitisers from non-sensitisers with accuracies of 97% based on LLNA data for 150 chemicals (9 extreme, 17 strong, 27 moderate, 36 weak and 61 non-sensitisers) and of 96% based on human data for 130 chemicals (52 non-sensitisers and 78 sensitisers). The performance of the test method in predicting skin sensitising potency reaches an accuracy of 93% when used to distinguish between five different LLNA classes (non-sensitiser, weak, moderate, strong and extreme sensitiser) for the same 150 chemicals (Cottrez et al., 2015b, submitted). A manuscript reporting the data for these 150 chemicals has been submitted for publication and was provided as an attachment to a recent SPSF proposal submitted to the OECD for the development of a new Test Guideline. These data were also presented to the OECD expert group on skin sensitisation in October 2015.

Finally, it should be mentioned that test methods using reconstructed human epidermis models also have the advantage that their test systems possess a metabolic capacity closer to that of native human skin (Hewitt et al., 2013), which is the target organ of interest when assessing the skin sensitisation potential/potency of chemicals.

Conclusions
The prediction of skin sensitisation potency by alternative methods is currently an important objective, and efforts to develop approaches have to be judged in the context of the inherent variability of the LLNA.

While firm recommendations for alternative testing strategies cannot be made, this mini-review shows that some in vitro and in silico methods show promise either for the identification of Cat 1A chemicals or even discrimination between potency classes. In particular, promising results have been reported for several approaches combining different methodologies (e.g. Dearden et al., 2015; Natsch et al., 2015; Takenouchi et al., 2015; Hirota et al., 2015; Jaworska et al. 2015), and some of in vitro gene based methods (e.g. Cottrez et al., 2015a, 2015b; Lambrechts et al., 2010).

References


Chapter R.7a: Endpoint specific guidance
Draft Version 5.0 – Public – February 2016

1 Allergic Contact Dermatitis in Humans. NIH Publication Number 11-7709. Available at:
2 Jaworska J (2011). Integrating non-animal test information into an adaptive testing
   strategy to assess skin sensitization potency: from theory to practice. Journal of Applied
   Toxicology 33(11):1353–64.
   integrated testing strategy (ITS) for skin sensitization potency assessment: a decision
   support system for quantitative weight of evidence and adaptive testing strategy.
   Archives of Toxicology In Press.
   signature can predict skin sensitizers using a cell-based in vitro alternative to animal
   measuring relative skin sensitising potency: a proposal. Journal of Applied Toxicology
8 Lambrechts N, Vanheel H, Nelissen I, Witters H, Van Den Heuvel R, Van Tendeloo V,
   Schoeters G & Hooyberghs J (2010). Assessment of Chemical Skin-Sensitizing Potency by
   an In Vitro Assay Based on Human Dendritic Cells. Toxicological Sciences 116(1):122–
   129.
   sensitizers combining peptide binding with ARE/EpRE-mediated gene expression in
    sensitization using human reconstructed skin models and its applicability to cosmetic,
    pharmaceutical, and medical device safety testing. Cutaneous and Ocular Toxicology
    Dependent Genes: Application to the In Vitro Testing of the Sensitization Potential of
12 Natsch A, Emter R & Ellis G (2009). Filling the Concept with Data: Integrating Data from
    Different In Vitro and In Silico Assays on Skin Sensitizers to Explore the Battery
    121.
    based on in vitro data from KeratinoSens and kinetic peptide binding: global versus
    relationship between CD86/CD54 expression and cytotoxicity in THP-1 cells stimulated
    with contact allergens. Toxicology Letters 36(3):313–324.
15 Nukada Y, Ashikaga T, Miyazawa M, Hirota M, Sakaguchi H, Sasa H & Nishiyama N
    (2012). Prediction of skin sensitization potency of chemicals by human Cell Line
    Activation Test (h-CLAT) and an attempt at classifying skin sensitization potency.
    Toxicology in Vitro 26(7):1150–1160.


