

TIMES-SS—A promising tool for the assessment of skin sensitization hazard. A characterization with respect to the OECD validation principles for (Q)SARs and an external evaluation for predictivity

Grace Patlewicz^{a,*}, Sabcho D. Dimitrov^b, Lawrence K. Low^c, Petra S. Kern^d,
Gergana D. Dimitrova^b, Mike I.H. Comber^e, Aynur O. Aptula^f, Richard D. Phillips^e,
Jay Niemelä^g, Charlotte Madsen^g, Eva B. Wedebye^g, David W. Roberts^h,
Paul T. Baileyⁱ, Ovanes G. Mekenyan^b

^a *European Chemicals Bureau (ECB), TP 582, Institute for Health and Consumer Protection, Joint Research Centre, European Commission, 21020 Ispra (VA), Italy*

^b *Laboratory of Mathematical Chemistry (LMC), University “Prof. As. Zlatarov”, 8010 Bourgas, Bulgaria*

^c *ExxonMobil Biomedical Sciences, 1545 Route 22 East, P.O. Box 971, Annandale, NJ 08801-0971, USA*

^d *Procter & Gamble Eurocor, Temselaan 100, B-1853 Strombeek-Bever, Belgium*

^e *ExxonMobil Biomedical Sciences, Hermeslaan 2, 1831 Machelen, Belgium*

^f *Safety Environmental Assurance Centre (SEAC), Unilever Colworth, Colworth House, Sharnbrook, Bedford MK44 1LQ, UK*

^g *Department of Toxicology and Risk Assessment, National Food Institute, Technical University of Denmark, Mørkøj Bygade 19, DK-2860 Søborg, Denmark*

^h *Liverpool John Moores University, Liverpool, UK*

ⁱ *Bailey & Associates, 31 Covered Bridge Road, Neshanic Station, NJ 08853, USA*

Received 2 February 2007

Available online 25 March 2007

Abstract

The Times MEtabolism Simulator platform used for predicting Skin Sensitization (TIMES-SS) is a hybrid expert system that was developed at Bourgas University using funding and data from a Consortium comprising industry and regulators. The model was developed with the aim of minimizing animal testing and to be scientifically valid in accordance with the OECD principles for (Q)SAR validation. TIMES-SS encodes structure–toxicity and structure–skin metabolism relationships through a number of transformations, some of which are underpinned by mechanistic 3D QSARs. Here, we describe the extent to which the five OECD principles are met and in particular the results from an external evaluation exercise that was recently carried out. As part of this exercise, data were generated for 40 new chemicals in the murine local lymph node assay (LLNA) and then compared with predictions made by TIMES-SS. The results were promising with an overall good concordance (83%) between experimental and predicted values. Further evaluation of these results highlighted certain inconsistencies which were rationalized by a consideration of reaction chemistry principles for sensitization. Improvements for TIMES-SS were proposed where appropriate. TIMES-SS is a promising tool to aid in the evaluation of skin sensitization hazard under legislative programs such as REACH.

© 2007 Elsevier Inc. All rights reserved.

Keywords: (Q)SARs; OECD principles for (Q)SAR Validation; TIMES-SS; Skin sensitization; REACH

* Corresponding author. Fax: +39 0332 786717.

E-mail address: grace.patlewicz@jrc.it (G. Patlewicz).

1. Introduction

Under current European Union (EU) legislation, substances are distinguished as “existing” or “new” chemicals based on a simple cut-off date. “Existing” substances (ca. 100,000 in total) are those that had been introduced in the market before 1981; “new” substances are those introduced after 1981 which require notification and testing in production volumes as low as 10 kg per year. No such testing provisions are required for existing substances. REACH (Registration, Evaluation, and Authorisation of Chemicals) puts an end to the artificial distinction made between “new” and “existing” chemicals by calling for equivalent information requirements for all chemicals manufactured or imported at greater than 1 ton per year (EC, 1999; EU, 2006). It is estimated that 30,000 existing chemicals will need to be assessed resulting in additional information requirements. This presents a significant challenge both in terms of cost and resources. Similar implications are likely to arise from the 7th Amendment to the Cosmetics Directive which poses a ban on the animal testing of cosmetic ingredients by 2009, and for some endpoints by 2013 (EU, 2003; Ruet Rossignol, 2005).

For reasons of practicality, cost effectiveness and animal welfare, it is envisaged that *in vitro* alternatives and non-testing approaches such as (Quantitative) structure activity relationships [(Q)SARs] will play a large role in the assessment of chemicals under these programs.

Skin sensitization is a significant issue from an occupational health and consumer product development perspective and an endpoint impacted by both pieces of legislation as a large number of additional data will need to be generated. To date, there are no non-animal replacements available to definitively identify skin sensitization hazard or evaluate risk (Basketter et al., 2005). *In vivo* methods are still used for the regulatory assessment of sensitization potential. The guinea pig maximization test (GPMT) and the Buehler test, described in Organisation for Economic Co-operation and Development (OECD) guideline 406 (OECD, 1992) and EU Annex V (B.6) (<http://ecb.jrc.it/testing-methods/annex5/>), were for many years the tests of choice. The GPMT is an adjuvant test in which the acquisition of sensitization is potentiated by the use of Freund's complete adjuvant (FCA) for intravenous exposure, combined with intradermal exposure. The Buehler test is a non-adjuvant method requiring topical application for the induction phase. Both the GPMT and the Buehler test have been demonstrated to detect chemicals with moderate to strong sensitization potential as well as those with relatively weak sensitization potential. For classification purposes, a response of at least 30% of the animals is considered positive for the GPMT and at least 15% in the Buehler test (Kimber et al., 2003).

More recently, the murine local lymph node assay (LLNA) as described in OECD guideline 429 (OECD, 2002) has emerged as the preferred choice method for measuring skin sensitization potential. In fact, under REACH,

the LLNA is stated to be the first choice method for *in vivo* testing. In the LLNA, skin sensitization hazard is defined as a function of the ability of the test chemical to provoke immune activation (lymphocyte proliferation) in lymph nodes draining the site of topical application (Gerberick et al., 2000; Basketter et al., 1996, 2002). A substance is classified as a sensitizer if it induces a threefold stimulation index (SI) or greater at one or more test concentrations (Kimber et al., 1994; Basketter et al., 2000, 2003). The LLNA is also able to provide a reliable measure of relative skin sensitizing potency. Potency is measured by derivation of an estimated concentration of substance required to induce a threefold SI value (EC3) as compared with concurrent vehicle controls (Basketter et al., 1999). EC3 values are usually considered to be accurate within a factor of ca. 2 (Basketter et al., 2000).

The use of (Q)SARs provides a promising approach in the evaluation of skin sensitization hazard and risk. For a chemical to induce skin sensitization, it must overcome a number of hurdles including skin penetration and formation of a stable association with a skin protein. This latter hurdle is thought to be the rate determining step of induction. Early work carried out by Landsteiner and Jacobs (1936) followed up by others, e.g., Dupuis and Benezra (1982) led to the hypothesis that this stable association was a covalent one whereby the skin protein behaved as a nucleophile and the chemical as an electrophile. Work has been predominantly focused on identifying the electrophilic characteristics of a chemical and using these to derive SARs or to develop QSAR models. Some have been derived empirically through analyzing sensitization data and looking for trends between the chemicals tested. Others have been identified through an understanding of chemical reactivity and assigning chemicals into mechanistic domains i.e., using common electrophilic–nucleophilic reaction pathways such as S_N2 , Michael addition, Schiff base, etc. (examples using this mechanistic approach include the QSAR work of Aptula et al., 2005; Aptula and Roberts, 2006; Dimitrov et al., 2005a; Gerner et al., 2004; Payne and Walsh, 1994; Roberts et al., 2007).

There are still relatively few (Q)SAR models/expert systems available for the prediction of skin sensitization. (Q)SAR models are usually categorized as global or local with nothing between, but in the present context we distinguish, in addition to expert systems, three categories, which we refer to as local models, global models, and universal models. Local models are typically characterized by a chemical class [e.g., Hansch approach (Hansch et al., 1962; Hansch and Fujita, 1964)]. Global models are based on a single chemical mechanism of action e.g., Relative Alkylation Index (RAI) (Roberts and Williams, 1982). Universal models are those based on skin sensitization datasets comprising diverse chemistries corresponding to a number of mechanisms of action (e.g., Estrada et al., 2003; Miller et al., 2005; Fedorowicz et al., 2004). Expert systems are built upon experimental toxicity results with rules derived from the data. The rules are either based on

statistical inference and take the form of QSARs, else they are based on expert judgment (i.e., knowledge-based) and take the form of SARs describing reactive chemistry or they are based on a combination of these two approaches, i.e., hybrids. Examples include TOPKAT (TOxicity Prediction Komputer-Assisted Technology) and M-CASE (Computer-Automated Structure Evolution) which are both statistically based, Derek for Windows (DfW) and Hazard Expert are knowledge-based and Tissue METabolism Simulator for Skin Sensitization (TIMES-SS) is a hybrid. These are described elsewhere in more detail (see Cronin et al., 2003; Rodford et al., 2003; Smith Pease et al., 2003).

TIMES-SS can be termed a hybrid expert system since it has been developed in principle to make predictions for a wide range of chemicals using a framework of interconnected models. It was developed by the Laboratory of Mathematical Chemistry (LMC) (University “Prof. As. Zlatarov,” Bourgas, Bulgaria) using funding and data from a Consortium comprising Industry (ExxonMobil, P&G, Unilever) and a Regulatory Agency (DK-EPA). TIMES-SS was developed to incorporate structure–toxicity and structure–metabolism relationships through a number of transformations simulating skin metabolism and interaction of the generated reactive metabolites with skin proteins. The skin metabolism simulator mimics metabolism using 2D structural information. Metabolic pathways are generated based on a set of 236 hierarchically ordered principal transformations including spontaneous reactions and enzyme catalyzed biotransformation reactions (phase I and II). The covalent reactions with proteins are described by 47 alerting groups. The associated mechanisms are in accordance with the existing knowledge on electrophilic interaction mechanisms of various structural functionalities. The integral model is essentially based on a set of submodels, associated with each of the reactive groups. Some of these reactions are additionally underpinned by mechanistically based 3D QSARs to predict the intrinsic reactivity of compounds that have substructural features associated with reactivity. These 3D-QSAR models depend on both the structural alert and the rate of skin sensitization. Steric effects around the active site, molecular size, shape, solubility, lipophilicity and electronic properties are all taken into account. These models may involve combinations of molecular descriptors which trigger the alerting group. The TIMES-SS model has already been described previously (Dimitrov et al., 2005a).

The primary objective of this article is to describe the scientific validity of TIMES-SS with reference to the OECD principles for (Q)SAR validation (OECD, 2004) and with particular emphasis on the external evaluation study that was recently conducted. In addition, we present TIMES-SS as a promising model that could be employed as part of an “integrated testing strategy” to provide information to evaluate skin sensitization under legislative programs such as REACH.

1.1. Regulatory use of (Q)SARs and the OECD principles for (Q)SAR validation

Under REACH, QSARs are described in Annex XI (EU, 2006). This states that the results from valid (Q)SARs may be used in place of animal testing for the purposes of identifying the absence or presence of certain dangerous properties as long as a number of conditions are met. These include: (1) the (Q)SAR is scientifically valid, (2) the substance falls within the applicability domain of the (Q)SAR model, (3) the QSAR result is suitable for classification and labelling and/or risk assessment purposes, (4) adequate reliable documentation is provided. The REACH text refers to the need to demonstrate the validity of the (Q)SAR used. It is anticipated that validity will make reference to the internationally agreed OECD principles for (Q)SAR validation (OECD, 2004). It also emphasises the principle that information generated by valid (Q)SARs may be used instead of experimental data, provided a number of conditions are met. The availability of a (Q)SAR for an endpoint of interest is necessary but to replace or partially replace an experimental result, the (Q)SAR data will need to be assessed for its reliability (i.e., inherent quality of the model within its applicability domain), relevance (scientific relevance and regulatory relevance) and adequacy (assessment of completeness of information in making a regulatory decision). Thus for a (Q)SAR to be adequate for a given regulatory purpose, the following conditions should be fulfilled: (1) estimate generated by a valid model; (2) model applicable to the chemical of interest with sufficient reliability; and (3) model endpoint should be relevant for the regulatory purpose (Worth et al., 2007).

In the ideal situation, (Q)SAR results should only be used *on their own* for regulatory purposes if they are considered relevant, reliable and adequate for the purpose. In practice, there will be uncertainty in one or more of these aspects, but this should not preclude the use of the (Q)SAR estimate in the context of a weight-of-evidence approach (Worth et al., 2007).

The OECD principles for (Q)SAR Validation were first drafted during the Setubal workshop, organized by the International Council of Chemical Associations (ICCA) and the European Chemical Industry (CEFIC) in March 2002 (Jaworska et al., 2003). Since then they have been discussed within the OECD (Q)SAR Group and agreed by the Joint Meeting of Chemicals Committee and Working Party on Chemicals, Pesticides & Biotechnology. They are now referred to as the OECD Validation principles for (Q)SARs (OECD, 2004). These state that in order “to facilitate the consideration of a (Q)SAR model for regulatory purposes, it should be associated with the following information:

- (1) A defined endpoint.
- (2) An unambiguous algorithm.
- (3) A defined domain of applicability.
- (4) Appropriate measures of goodness-of-fit, robustness and predictivity.

(5) A mechanistic interpretation, if possible.”

The principles have been summarized briefly in a previous publication (OECD, 2004). An OECD Guidance document on (Q)SAR Validation has just been finalized (OECD, 2007). The basis for this OECD guidance, was a preliminary guidance on the characterization of (Q)SARs written by Worth et al. (2005).

The principles for (Q)SAR validation identify the types of information that are considered useful for the assessment of (Q)SARs for regulatory purposes and to an extent represent what would be the gold standard or best practice. They do not in themselves provide criteria for the regulatory acceptance of (Q)SARs. Since no formal adoption procedure is likely under REACH, the concordance with the OECD principles is thought to be best evidenced by documenting the characteristics of a model using a reporting format, so-called QSAR Model Reporting Format (QMRF). (see http://ecb.jrc.it/QSAR/home.php?CON-TENU=/QSAR/qsar_tools/qsar_tools_qrf.php).

1.2. Skin sensitization—a defined endpoint

According to the OECD Validation Principle 1 (see above), a (Q)SAR should be associated with a “defined endpoint”, where endpoint refers to any physicochemical, biological or environmental effect that can be measured and therefore modelled. The intent of this principle is to ensure transparency in the endpoint being predicted by a given model.

Skin sensitization is a defined toxicological endpoint of regulatory significance. Current approaches to concretely assess skin sensitization potential, is carried out through *in vivo* testing. Test Guidelines 429 and 406 describe the LLNA and GPMT/Buehler tests that are available (OECD, 1992, 2002). The TIMES-SS model was derived from a dataset compiled from chemicals tested in the LLNA (258 chemicals) (see below for references), the GPMT (360 chemicals) (Cronin and Basketter, 1994; Barratt et al., 1994) as well as from the BgVV list (264 chemicals) (Schlede et al., 2003). The LLNA and the GPMT have already been described with reference to their respective test guidelines (see earlier). A group of experts including dermatologists from universities and representatives of the chemical industry and from regulatory authorities was established by the German Federal Institute for Health Protection of Consumers and Veterinary Medicine (BgVV) in 1985 (Schlede et al., 2003). The aim of the project was to bring together a broad range of expertise to collect and evaluate data from the literature on substances with documented contact allergenic properties in humans (from clinical data and experimental studies) and from animal experiments. The evaluation resulted in a publication where chemicals were listed as belonging to one of three categories (A–C) where Category A represented significant contact allergens; B, a solid-based indication for contact

allergenic potential; and C, insignificant or questionable contact-allergenic potential.

The LLNA data available were provided from Unilever and Procter & Gamble’s in-house archives, who are both consortium members in this QSAR project. The data are of high quality, generated to Good Laboratory Practice (GLP) and in the spirit of current OECD test guidelines. The study reports and conclusions were carefully reviewed and quality assured. The majority of the data if not all has been published in peer reviewed publications in Toxicology (e.g., Ashby et al., 1995), Contact Dermatitis (e.g., Gerberick et al., 2004; Patlewicz et al., 2004), Chemical Research in Toxicology (e.g., Estrada et al., 2003) and elsewhere. The most recent and extensive compilation was that published by Gerberick et al. (2005) which included a dataset of 211 chemicals. This dataset was included as part of the set of LLNA data in the present QSAR study. For the available LLNA data, the EC3 was derived in each case and used to group chemicals into bands of potency in accordance with the proposals outlined by Kimber et al. (2003). The GPMT was originally taken from Unilever, and has been published (Barratt et al., 1994; Cronin and Basketter, 1994). Potency categories described in the same publication (Barratt et al., 1994) for the GPMT were also used. One of the aims of the TIMES-SS model was that it would go some way into predicting skin sensitization potency rather than just its potential to occur (i.e., the presence or absence of sensitization). For this reason the full dataset (882 entries in total) needed to be collated and the sensitization data was reassessed using a unified potency scale which summarized the data into one of three categories which corresponded to those of the LLNA, GPMT, and BgVV as far as possible. Given that the classification schemes were different, some of the Category C BgVV data were reviewed by the authors and reclassified from weak-sensitizers to non-sensitizers. A scale of significant, weak or non-sensitizing, where significant comprised both strong and moderate skin sensitizers was developed. Here, extreme, strong or moderate in the LLNA or strong/moderate in the GPMT or Category A in the BgVV corresponded to a significant score in this unified scale.

The unifying scale was derived through evaluating the correlation and concordance of those chemicals that existed in all three datasets. These chemicals were effectively the yardstick for setting the thresholds for the scale and ensuring consistency between the three sources of data.

Seventeen chemicals in total existed in all three datasets; 34 chemicals in both the LLNA and BgVV; 44 chemicals were common to the GPMT and BgVV; and 41 in both the LLNA and GPMT. The strength of the association between the different datasets was estimated using an adjusted Pearson correlation coefficient (see Dimitrov et al., 2005a). The values demonstrated a well-defined correlation between observed sensitization potency of chemicals in the different datasets. For chemicals whose potency was assessed by more than one method, the LLNA was accepted as representative for the chemical and where

no LLNA existed, the GPMT was taken as the representative outcome. After elimination of overlapping chemicals, the training set was reduced to 729 chemicals. It should be mentioned that the initial number of chemicals used for the development of skin sensitization model was 633. As a second iteration in model building, a set of 96 chemicals were collected and used to examine the performance of the model.

The compilation of the three datasets is not entirely ideal since different data sources are combined rather than making use of a single experimental test protocol such as the LLNA and because the BgVV is not a hazard identification scheme but takes into account prevalence and clinical relevance. However in the interests of compiling as much data on the widest range of chemicals to develop a model that would be as broadly applicable as possible, this strategy was both a pragmatic and practical solution.

The dataset is transparent and clearly matches a regulatory requirement of assessing skin sensitization hazard i.e., the primary endpoint is the same as that used in regulatory practice. A secondary endpoint score was adopted for the purposes of developing the TIMES-SS model. TIMES-SS predicts likely skin sensitization potency as defined by its derived unified classification score i.e., significant, weak or non-sensitizing.

On this basis, OECD Validation Principle 1 “defined endpoint” is proposed as being satisfactorily met.

1.3. Skin sensitization—an unambiguous algorithm

According to OECD Validation Principle 2, a (Q)SAR should be expressed in the form of an unambiguous algorithm. The intent of this principle is to ensure transparency in the description of the model algorithm.

As TIMES-SS comprises an array of interconnected models—alerts, transformations and 3D-QSAR models, a brief description of how the model functions is provided here.

TIMES-SS integrates a simulator of skin metabolism together with a number of QSAR models for assessing the reactivity of specific alerts. Since there is a real paucity of reported skin metabolism data, the simulator transformations were developed on the basis of empirical and theoretical knowledge. The transformation probabilities (defining the priority of their execution) were parameterized to reproduce skin sensitization data. The reliability of transformations was then assigned into one of five confidence levels after review by experts. The highest confidence level of 1 was assigned where the transformation was well established and supported by literature references. A level of 0.75 was set where the transformation was well known but no literature references were available. A level of 0.5 for where the transformation was expected but the enzymatic system was unknown and a level of 0.25 where only expert judgement supported its feasibility. A level of 0 was assigned where there was no knowledge of feasibility,

however, these transformations were needed to reproduce skin allergy data. The simulator comprises a total of 236 transformations which can be divided into two main types—rate-determining and non-rate-determining. The rate-determining transformations are the Phase I (comprising 121 reactions) and II (comprising nine reactions) including those such as C-hydroxylation, ester hydrolysis, oxidation, glucuronidation, and sulfonation. The non-rate determining transformations consist of hydrolysis of salts, molecular transformations of highly reactive intermediates as well as covalent reactions with skin proteins (i.e., the structure–toxicity rules). This set of covalent reactions also mimic the mechanistic domains described by Aptula and Roberts (2006), i.e., the common electrophilic–nucleophilic reaction pathways such as S_N2 , Michael addition, and Schiff Base, etc.

Examples of transformation are shown in Table 1. A more extensive list of examples can be found in our previous publication (Dimitrov et al., 2005a).

All non-rate-determining reactions and skin covalent interactions are assigned with a reliability score of 1. Interactions with skin proteins are grouped into three types; significant, weak, and interactions requiring 3D-QSAR models to quantify the potency of sensitization of the alerting group. These Common Pattern REcognition (COREPA) models were derived for ketones, sulfuric/sulfonic/phosphonic acid esters, conjugated double bonds containing O, S or N atoms and aldehydes. COREPA is a probabilistic technique for identifying common stereoelectronic (reactivity) patterns of structurally diverse chemicals which may exert similar or differential biological effects (see Mekenyan et al., 1997,1999; Bradbury et al., 2000).

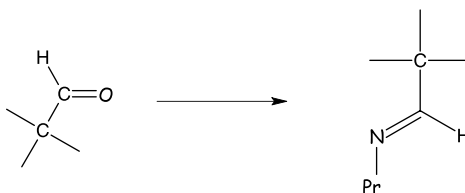
Due to limited data, chemicals only having one of these alerts were combined to form training sets. For ketones, acid esters, and aldehydes, QSARs were only trained to discriminate between significant (S) and non-sensitizers (N) [or between significant (S) and weak sensitizers (W)] owing to insufficient data for W [or N] chemicals. For example, the model for aldehydes uses HOMO–LUMO energy gap (E_{GAP}), log of the octanol/water partition coefficient ($\log K_{ow}$) and the charge of the aldehyde C-site (Q) to discriminate between significant and weak sensitizers. Table 2 lists the COREPA models for reactivity.

The way in which the TIMES-SS works is as follows: a compound is introduced into TIMES-SS and matched against the list of 236 hierarchical transformations. For all the matches identified, the reactive species or metabolic species and their respective protein adducts are then generated.

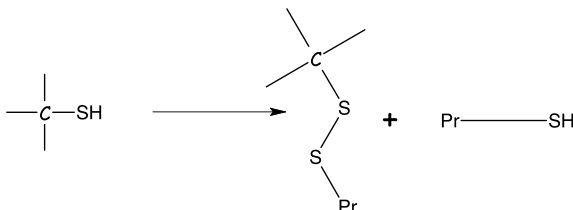
For example, for diethyl maleate TIMES-SS identifies one match in its transformation table. This is namely the potential of the chemical to form an adduct via a Michael addition reaction. This pathway is subsequently underpinned by a 3D-QSAR model to assign potency. The COREPA-M model relates skin sensitization score to the highest occupied molecular orbital (E_{HOMO}), molecular weight (MW) and polarizability. As a result the outcome

Table 1
Examples of typical transformations

Schiff base formation
with aldehydes



Thiol exchange



Michael addition on conjugated
systems containing O, N or S
atoms where R = O, N, and S

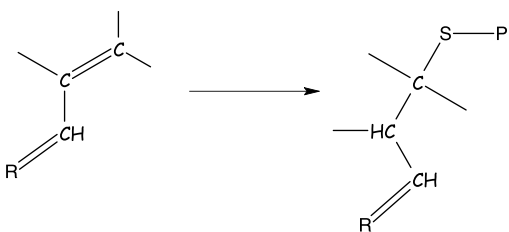
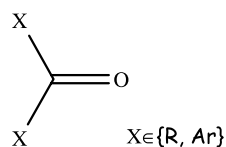


Table 2
COREPA models of reactivity

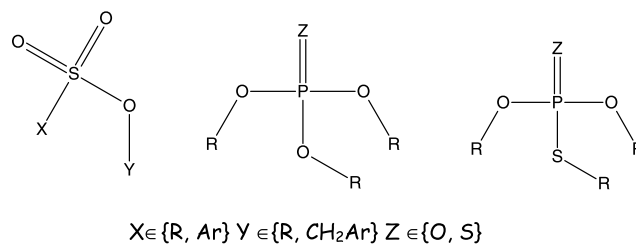
COREPA models

Structural features

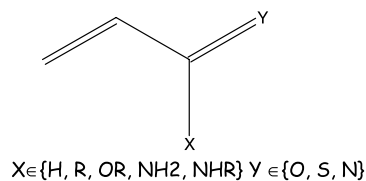
Ketones



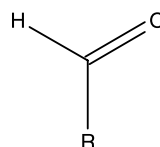
Sulfuric, sulfonic or phosphonic acid esters



Conjugated systems containing O, S, and N atoms



Aldehydes



for diethyl maleate according to TIMES-SS is that it will be a strong sensitizer.

As another example, *p*-aminobenzoic acid is predicted to be metabolized to a reactive species which accounts for a potential weak sensitization response.

TIMES-SS is a complex system comprising many interconnected transformation pathways underpinned with 3D QSARs where relevant. All are transparent and accessible by the end-user (see Scheme 1 for an illustration of the different interconnections). For this reason, the second OECD Validation principle is met.

1.4. Skin sensitization—applicability domain

The OECD Validation Principle 3 states that “a (Q)SAR should be associated with a defined domain of applicability”. This need is based on the fact that (Q)SARs are inevitably associated with limitations in terms of the types of chemical structures, physicochemical properties, and mechanisms of action for which the models can generate reliable predictions.

Recently, Netzeva et al. (2005) have defined the applicability domain (AD) of a (Q)SAR model as “the response and chemical structure space in which the model makes predictions with a given reliability”.

In this definition, chemical structure can be expressed by information on physicochemical properties and/or structural fragments, and the response can be any physicochemical, biological or environmental effect that is being predicted.

TIMES-SS utilizes a four-stage *applicability domain*. This has been recently described by Dimitrov et al. (2005b). The first stage comprises a *global requirements domain* with simple cut-offs for physicochemical parameters such as log- K_{ow} , molecular weight and water solubility. TIMES-SS specifically includes upper and lower thresholds for log K_{ow} and MW based on the training set information.

The next stage consists of defining the *structural domain* which is used to gauge the extent to which a chemical is structurally similar to those in the training set of chemicals

for which the model performs correctly (within a user-defined accuracy threshold). The structural domain can be characterized by the atom centered fragments of these chemicals. A series of rules was proposed to reflect the effect of different neighbors on a specified atom. Hydrogen atoms were treated as an inherent part of the atoms to which they were bound, whereas the first, second, and so on neighbors were used to characterize the specified atom. In practice this means that specific substructures are extracted from a set of compounds to characterize the structural domain for a given atom, e.g., a set of N-containing compounds to define a set of substructures accounting for the first or second neighbors of the N atom. An example is shown in Table 3.

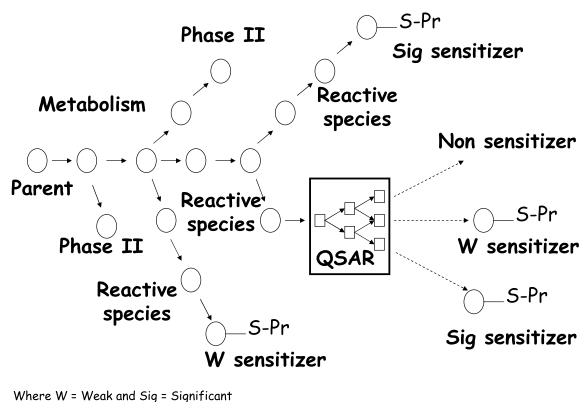
The conservativeness of structural domain is defined by the atomic neighborhood taken into account for atom-centered fragments. This affects the chemicals which are encompassed in the domain and the correctness of the predictions. When the structural domain was taken into account for the first or second significant neighbors, the statistics of the model predicting the original training set chemicals changed as shown in Table 4.

The *mechanistic domain* comprises several steps and is the most challenging. Within TIMES-SS, a set of subdomains have been proposed. The first explores the domain of functional or reactive groups, i.e., the performance of the alerts with respect to the interactions with protein. A second is through interpolating the model explanatory variables. In this instance, a QSAR model's descriptors underpinning one of the alerts would form the basis of the interpolation. The next step would explore the reliability of a prediction within the interpolated space accounting for the predictivity within regions of densely or sparsely populated space. That is to say if an interpolated space was simply defined by the ranges of the descriptors—there would be some regions within those ranges that would be densely populated by chemicals and the robustness of the predictions in those regions would provide a measure of reliability in both interpolated space and predictivity.

The *metabolic simulator* domain takes into account the reliability of the generated metabolites predicted to interact with protein. The reliability of metabolites is determined as a product of reliabilities of transformations used to generate these metabolites.

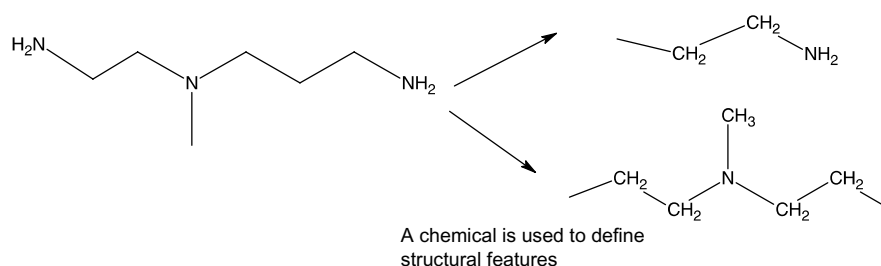
An overall call for the domain status is reported as the total domain. If one of the domains fails for a given chemical—the overall outcome is reported as “outside of domain”.

TIMES-SS provides an applicability domain definition for each of its component models which are applied in a sequential manner. The advantage of processing query chemicals through all four stages is the increased reliability of prediction for those chemicals that satisfy all four conditions for inclusion in the AD. The cost of applying this rigorous approach is that the number of chemicals for which reliable predictions are eventually made is reduced but this increases confidence in reliability of the final prediction.

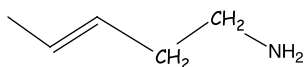


Scheme 1. A schematic to highlight the many interconnected models within TIMES-SS that result in a given prediction.

Table 3
Structural domain accounting for the second neighbors of N atoms



Example: Structure belonging to structural domain of N atom



Example: Structure not belonging to structural domain of N atom

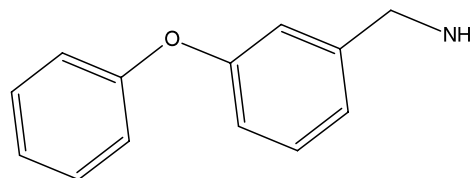


Table 4
Effect of the significant neighbors of centered atoms on model statistics

Statistics	Significant neighbors of centered atoms	
	First	Second
*C	0.8	0.95
Sensitivity (%)	100	100
Specificity (%)	59	87
False negatives (%)	0	0
False positives (%)	41	13

*C, adjusted Pearson contingency coefficient.

Sensitivity, the probability that the chemical belongs to the model domain, given that the chemical is truly an element of the model domain.

Specificity, the probability that the chemical does not belong to the model domain given that the chemical is truly not an element of the model domain.

False negatives, the probability that the chemical will be classified out of the structural domain given the chemical is truly in the domain.

False positives, the probability that the chemical will be classified in the structural domain given the chemical is truly out of the domain.

Thus, OECD Validation Principle 3 is satisfactorily met for TIMES-SS.

1.5. Skin sensitization—goodness-of-fit, robustness, and predictivity

The OECD Validation Principle 4, states that models should be associated with appropriate measures of goodness-of-fit and robustness (internal performance) and predictivity (external performance).

Here, we will first describe the goodness-of-fit and robustness. The external performance was evaluated by a separate study and is described separately.

1.5.1. Goodness-of-fit

TIMES-SS is a classification-based model (CM), such that its goodness-of-fit can be conveniently assessed in

terms of its Cooper statistics (Cooper et al., 1979). The statistics collectively express the performance of a CM, provided they measure its ability to detect known active compounds (sensitivity), non-active compounds (specificity), and all chemicals in general (concordance or accuracy). The false-positive and false-negative rates can be calculated from the complement of specificity and sensitivity, respectively. The positive and negative “predictivities” focus on the effects of individual chemicals, since they act as conditional probabilities. Thus, the positive “predictivity” is the probability that a chemical classified as active is really active, while the negative “predictivity” gives the probability that a classified non-active chemical is really non-active.

In our original paper (Dimitrov et al., 2005a) an external validation was carried out for 96 chemicals not used in the model building process. These chemicals also were used to examine the usefulness of the definition of the model domain. The predictivity of the model was 87% when the applicability domain was taken into account for the 96 chemicals. It should be mentioned that ignoring the model domain significantly reduced the predictivity of the model to 52%. In order to update and refine the original model the external chemicals were added to the training set. As a result for the 729 chemicals considered, the QSAR model was able to predict correctly 82% of the significant sensitizers, 35% of the weak sensitizers and 74% of the non-sensitizers, i.e., an overall performance of 65%.

The failure of the model for predicting the weak sensitization could be due to model insufficiency (lack of appropriate transformations or lack of specificity of the transformations). However, we assume that the most significant reason for this is the accuracy of experimental data and inconsistency of used data. Thus, merging various databases increased structural domain of the model at the cost of inconsistency of test protocols. For example, the

concordance between LLNA and GPMT for overlapping 41 chemicals in both inventories was 83%.

1.5.2. Skin sensitization predictivity: an external evaluation performance exercise

The first step of the model evaluation exercise involved characterizing the applicability domain (as described in Section 1.4). Hence the model was locked from any further development.

As TIMES-SS was developed in light of potential use under REACH, the (European INventory of Existing Commercial chemical Substances) EINECS inventory (compiled by LMC previously) comprising 60,311 chemicals was used as an inventory of REACH relevant chemicals. This inventory was screened through TIMES-SS to identify the subset of chemicals which fell within the applicability domain of TIMES-SS. While the applicability domain of TIMES-SS is actually a 4-step process (as described in Section 1.4), for reasons of pragmatism and practicality a decision was made to stop at step 2 i.e., the structural domain. The 2D structural domain to the nearest first neighbors was used as a working domain for compound selection of which 6000 chemicals were found to belong to this structural model domain.

From this filtered inventory, a set of 160 chemicals that covered a similar chemical distribution as the original training set was chosen using the ChemPick software. This software is an experimental design application developed by LMC (see <http://omega.btu.bg/?section=software&swid=7>) and comprising different statistical algorithms for the selection of diverse and representative subsets of chemicals. To maintain independence, the Danish EPA within the Consortium performed the chemical selection.

A further filter was carried out such that only commercially available chemicals remained; 40 chemicals were then picked at “random” to ideally result in a reasonable balance of predicted sensitizers and non-sensitizers. Of the 40 chemicals chosen, 12 were predicted by TIMES-SS to be sensitizers and 28 to be negatives.

The 40 chemicals were all sourced from Sigma–Aldrich and with at least a purity of 97%. (See Figs. 1–3 for distributions of the purity, $\log K_{ow}$ and molecular weight).

Note it could be argued that a better use of resources would be to understand the boundaries of TIMES-SS more fully and seek to revise and extend these but this study was conducted with the intention of demonstrating model performance based on its true scope and limitations. Lessons learned were identified after testing and these will be discussed in more detail in a separate forthcoming publication.

LLNA testing was chosen for a number of reasons. The LLNA is the assay of first choice for REACH; it is able to provide information on relative skin sensitization potency and requires the use of fewer animals than comparative guinea pig methods. Syngenta Central Toxicology Laboratory (CTL) was identified as the testing laboratory since it met a number of the criteria in terms

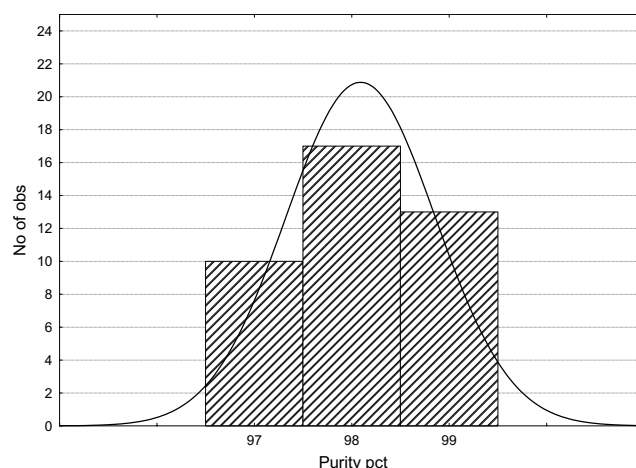


Fig. 1. Distribution of purity (%).

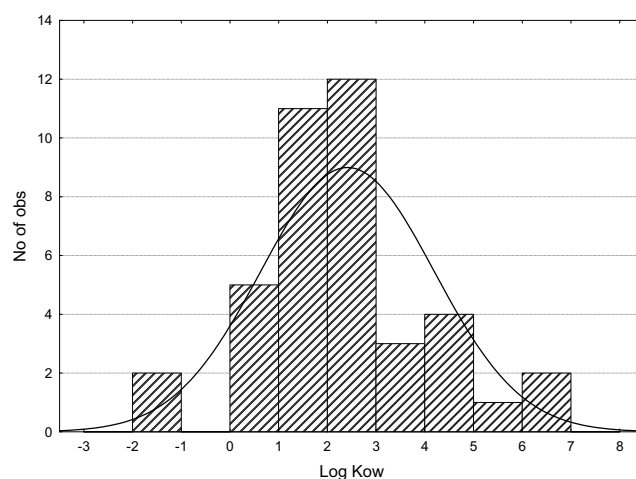


Fig. 2. Distribution of $\log K_{ow}$.

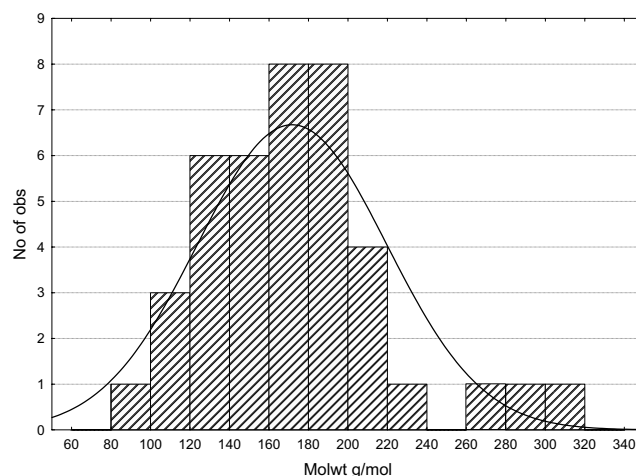


Fig. 3. Distribution of molecular weight.

of excellence. CTL was not the cheapest but appeared to have the most convincing expertise in addition to Good Laboratory Practice (GLP) status and availability of

testing slots. The LLNA as described in OECD 429 (OECD, 2002) was performed.

Test concentrations were selected based on the outcomes of specific range finding studies and to an extent on expert opinion. Range finding studies were conducted for all but five chemicals. Three dose groups were used as standard as described in the OECD guideline. While selecting more test concentrations would have led to better elucidation of dose–response relationships for all compounds and hence a greater likelihood of deriving EC3s for a greater number of chemicals, the three dose groups chosen were at least likely to ensure a conclusive outcome of either positive or negative sensitization response. Testing in three test concentrations for all but two chemicals provided a conclusive outcome. For two chemicals (allyl acetate [CASRN 591-87-7] and

2-fluoro-5-nitroaniline [CASRN 69-36-8]), additional testing was required.

Acetone:olive oil, AOO [4:1 (v/v)] is the standard vehicle according to OECD 429 test guidelines. In case of solubility issues, a list of alternative vehicles is highlighted in the guideline. In the present study, 36 out of the 40 chemicals were tested using AOO as the vehicle. The four remaining compounds were tested in the following vehicles, respectively: tetradecane-1,14-diol [CASRN 19812-64-7] in a 7:3 ratio w/v of ethanol:water. Phenol, 2,2-azobis [CASRN 2050-14-8] and 3-hydroxy-4-nitrobenzoic acid [CASRN 619-14-7] were tested in DMSO (dimethylsulfoxide) and 4-amino-2-nitro phenol [CASRN 119-34-6] in dimethyl formamide (DMF).

Table 5 lists the compounds that were selected together with the vehicles used, predicted TIMES-SS and final

Table 5
Summary of chemicals tested, LLNA classifications and TIMES-SS predicted skin sensitization potential

Chemical name	Registry No. (CASRN)	Vehicle	EC3 %	LLNA classification	TIMES-SS est. activity
Butanoic acid, 3-methyl-, 2-phenylethyl ester	140-26-1	AOO	39.5	Pos	Neg
2-Chloro-6-nitrotoluene	83-42-1	AOO	NC	Neg	Neg
Allyl acetate	591-87-7	AOO	NC	Neg	Pos
2-(Hexadecyloxy)ethanol	2136-71-2	AOO	8.8	Pos	Neg
2-Fluoro-5-nitroaniline	69-36-8	AOO	NC	Neg	Pos
6-Methoxynaphthalene-2-carbaldehyde	3453-33-6	AOO	NC	Neg	Neg
Tetradecane-1,14-diol	19812-64-7	EtOH:H ₂ O (7:3)	NC	Neg	Neg
Phenol, 2,2-azobis-	2050-14-8	DMSO	27.9	Pos	Pos
2-Ethylaminoethyl dimethylamine	123-83-1	AOO	11.3	Pos	Pos
3,4-Dinitrophenol	577-71-9	AOO	NC	Pos	Neg
Benzene, (2-bromoethyl)-	103-63-9	AOO	6.2	Pos	Neg
1,2-Bis(2-chloroethoxy)-ethane	112-26-5	AOO	NC	Neg	Pos
1-Chloro-3-iodopropane	6940-76-7	AOO	NC	Neg	Neg
Cyclooctanol	696-71-9	AOO	NC	Neg	Neg
2,6-Dimethylbenzoic acid	632-46-2	AOO	NC	Neg	Neg
4-Amino-2-nitrophenol	119-34-6	DMF	NC	Pos	Pos
2-Chloro-3',4'-dihydroxyacetophenone	99-40-1	AOO	NC	Pos	Pos
3-Methyl-1-butanol	123-51-3	AOO	NC	Neg	Neg
Butylbenzylphthalate	85-68-7	AOO	NC	Neg	Neg
3-Chloro- <i>p</i> -anisaldehyde	4903-09-7	AOO	NC	Neg	Neg
3-Hydroxy-4-nitrobenzoic acid	619-14-7	DMSO	NC	Neg	Neg
Propanoyl chloride, 2,2-dimethyl-	3282-30-2	AOO	NC	Pos	Pos
Propanoic acid, 3-bromo-, methyl ester	3395-91-3	AOO	NC	Neg	Neg
Phenol, 3-(diethylamino)-	91-68-9	AOO	NC	Pos	Pos
(<i>S</i>)-2-Hydroxy-2-methylsuccinic acid	6236-09-5	AOO	NC	Neg	Neg
Cyclohex-3-ene-1-carboxylic acid	4771-80-6	AOO	9.3	Pos	Pos
<i>p</i> -Chlorophenethyl alcohol	1875-88-3	AOO	31.3	Pos	Neg
Tridecane	629-50-5	AOO	48.4	Pos	Neg
3-Chloropropionyl chloride	625-36-5	AOO	NC	Pos	Pos
Benzene, 1,3,5-trimethoxy-	621-23-8	AOO	NC	Neg	Neg
2,2'-Oxybisethanol	111-46-6	AOO	NC	Neg	Neg
1,2-Benzenedicarboxylic acid, dibutyl ester	84-74-2	AOO	NC	Neg	Neg
Dipiperidinomethane	880-09-1	AOO	NC	Pos	Pos
<i>N,N</i> -dibutylaniline	613-29-6	AOO	19.6	Pos	Neg
Benzenepropanol, acetate	122-72-5	AOO	NC	Neg	Neg
Cyclopentanepropanol	767-05-5	AOO	NC	Neg	Neg
Ethanol, 2-butoxy-, acetate	112-07-2	AOO	NC	Neg	Neg
Benzene, 1-methoxy-4-methyl-2-nitro-	119-10-8	AOO	NC	Neg	Neg
3-Phenoxypropionitrile	3055-86-5	AOO	NC	Neg	Neg
Octanenitrile	124-12-9	AOO	NC	Neg	Neg

Where CASRN, CAS registry number; NC, not calculable; Neg, negative; Pos, positive; EC3, estimated concentration needed to induce a 3-fold increase in the stimulation index (SI) in the LLNA, AOO = [4:1 (v/v)] acetone:olive oil; DMSO, dimethylsulfoxide; DMF, dimethyl formamide; and EtOH:H₂O (7:3), a 7:3 ratio w/v of ethanol:water.

experimental outcome call. It is worth noting here that the final experimental outcome calls are consensus agreements achieved through discussion between experts taken in conjunction with the study outcome reported by the testing laboratory.

In total, there were 24 non-sensitizers and 16 sensitizers. For all 16 sensitizers, an estimate of relative skin sensitizing potency could be made in accordance with the bands outlined by Kimber et al. (2003). For 10 of these sensitizers, an actual EC₃ could be determined which gave a more robust indication of relative potency. Six chemicals were weak sensitizers, and four were moderate sensitizers. The remaining six sensitizers all had SI values in excess of 3 at either 5% or 10% dosage concentrations which was either indicative of strong (four sensitizers) or moderate (two sensitizers) relative potency. For the 24 non-sensitizers, no dose–response relationship was seen at any of the concentrations tested.

The overall concordance of TIMES-SS was found to be 75% (30/40). This equated to 7/16 false negatives and 3/24 false positives. TIMES-SS predicted 87.5% (21/24) non-sensitizers correctly (specificity) and 56% (9/16) of the sensitizers correctly (sensitivity). See Table 6 for the performance scores.

Evaluating the results more closely identified three possible scenarios for explaining the inconsistencies between predicted and experimental outcomes. These were as follows: shortcomings in the transformations and rules within TIMES-SS, uncertainty in the experimental data or borderline sensitization observed.

Three chemicals were predicted to be sensitizers but showed a negative response in the LLNA as shown in Table 7.

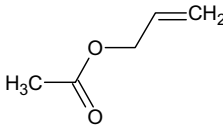
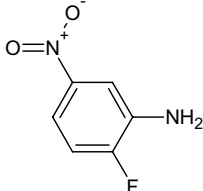
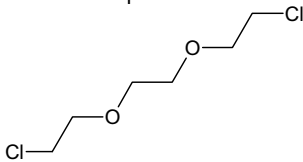
Allyl acetate [CASRN 591-87-7] tested negative in the LLNA yet the prediction in TIMES-SS suggested it was a sensitizer. The alert within TIMES-SS highlighted the potential for epoxide formation. Inspection of the chemistry suggested that this was unlikely and that in fact the rule might simply be too general in scope. Further work would need to be carried out to determine if and how an exclusion rule could help to tighten the scope of this epoxide rule.

1,2-Bis(2-chloroethoxy)-ethane, [CASRN 112-26-5] also tested negative although TIMES-SS identified an alert for mustard-like structures as being sensitizers. Here, the rule might be too general in scope and need refinement to exclude mustards with oxygens as opposed to sulfurs or nitrogens.

2-Fluoro-5-nitroaniline [CASRN 69-36-8] tested negative but TIMES-SS predicted it as positive based on the

Table 7

Three chemicals predicted to be sensitizers but which showed a negative response in the LLNA

Structure	Name	CASRN
	Allyl acetate	591-87-7
	2-Fluoro-5-nitroaniline	69-36-8
	1,2-Bis(2-chloroethoxy)-ethane	112-26-5

potential formation of quinone or iminoquinone metabolites. The mechanistic principles for aromatic hydroxyl and amino compounds are still not completely clear. A number are found to be experimentally sensitizers and a number are not (Roberts et al., 2007). More work is required to determine what refinements are required to this existing transformation. Additional testing or more data on substituted aromatic amines might provide new insights to help rationalize this further.

The TIMES-SS predictions for seven chemicals as non-sensitizers were not in agreement with the experimental LLNA data. Table 8 lists the structures affected.

There was one borderline case for which sensitization effects (EC₃ > 30%) were only seen at very high concentrations, e.g. *p*-chlorophenethyl alcohol [CASRN 1875-88-3] but it was considered as positive in the LLNA study.

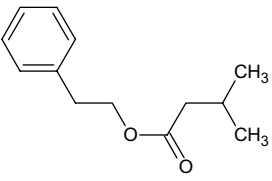
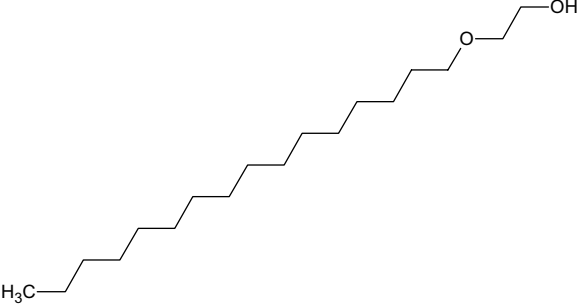
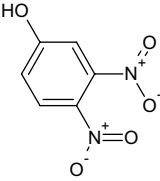
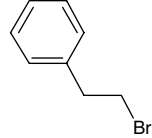
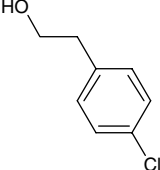
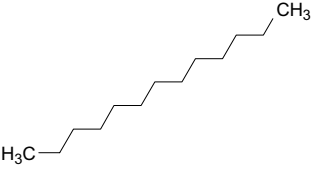
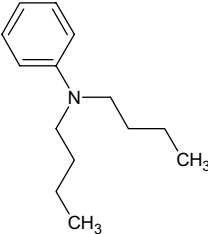
In three instances, the inconsistencies could have been due to the experimental data being a source of error through the presence of an impurity, e.g., tridecane [CASRN 629-50-5], butanoic acid, 3-methyl-, 2-phenylethyl ester [CASRN 140-26-1], 2-(hexadecyloxy)ethanol [CASRN 2136-71-2].

Tridecane [CASRN 629-50-5] gave a response indicative of a weak sensitizer, with an EC₃ of 48%. Review of the chemical structure reveals no electrophilic features which would be indicative of the potential of sensitization to occur. The LLNA result was thought to be incorrect probably as a result of possible impurities or the potential skin-irritating effects of tridecane. Other investigators have reported non-specific lymph node proliferation in the LLNA with tetradecane that has been principally attributable to its skin irritancy potential rather than a skin sensitizing effect (Woolhiser et al., 1998). These investigators had used a combined murine ear-swelling irritancy-LLNA assay to distinguish the skin irritancy potential of

Table 6
Performance scores

TIMES-SS	Performance scores
Specificity	88% (21/24)
Sensitivity	56% (9/16)
Concordance	75% (30/40)

Table 8
Seven chemicals that were predicted to be non-sensitizers but which showed a positive response in the LLNA

Structure	Name	CASRN
	Butanoic acid, 3-methyl-, 2-phenylethyl ester ^a	140-26-1
	2-(Hexadecyloxy)ethanol ^a	2136-71-2
	3,4-Dinitrophenol	577-71-9
	Benzene, (2-bromoethyl)-	103-63-9
	<i>p</i> -chlorophenethylic alcohol ^a	1875-88-3
	Tridecane ^a	629-50-5
	<i>N,N</i> -Dibutylaniline	613-29-6

^a See text for further explanation and discussion for these four chemicals with regards to the LLNA data.

tetradecane. In addition, tetradecane has been reported to cause ear edema following topical application to the ears of mice at 50% concentration which is also suggestive of potential skin irritation (Moloney and Teal, 1988).

Butanoic acid, 3-methyl, 2-phenylethyl ester [CASRN 140-26-1] classified as a weak sensitizer with an EC3 of 39%. The structure was not indicative of a sensitizer and the result might possibly be due to the presence of an impurity.

A spontaneous (auto-oxidation) transformation may explain why 2-(hexadecyloxy)ethanol [CASRN 2136-71-2] tested positive in the LLNA with an EC3 of 9% despite the lack of any electrophilic characteristics within its structure. Long-chain fatty alcohols or ethoxylated alcohols or surfactants have been shown to undergo auto-oxidation and then give positive results in the LLNA (Bodin et al., 2003) despite showing no sensitization in humans.

For *N,N*-dibutylaniline [CASRN 613-29-6], tested positively with an EC3 of 19%, i.e., a weak sensitizer. The mechanistic principles for aromatic hydroxyl and amino compounds are still not completely clear and more work is still required to elucidate the potential mechanisms and hence incorporate these into TIMES-SS (Roberts et al., 2007).

Benzene, (2-bromoethyl) [CASRN 103-63-9], tested positive in the LLNA with an EC3 of 6% yet was predicted to be negative by TIMES-SS. The current rule within TIMES-SS is perhaps oversimplified and requires further refinement.

If the statistics are re-evaluated without considering the four chemicals that had associated uncertainty in experimental data (i.e., tridecane [CASRN 629-50-5], butanoic acid, 3-methyl, 2-phenylethyl ester [CASRN 140-26-1], 2-(hexadecyloxy)ethanol [CASRN 2136-71-2]) or were borderline in sensitization (*p*-chlorophenethylic alcohol [CASRN 1875-88-3]), then the concordance increases to 83.33%. False negatives are now 3/12 and the false positives are now 3/24.

These cases together with a mechanistic rationale for all 40 chemicals tested will be discussed more fully in a separate and forthcoming publication. The insights and suggested improvements are currently being implemented for inclusion into the next version of the TIMES-SS model.

It remains sufficient in the present paper to explain that an external evaluation for TIMES-SS was carried out and carefully scrutinized to identify where the software did require improvement and where the animal data appear to give erroneous results.

The performance of TIMES-SS was assessed on the basis on sound design. The outcomes were carefully evaluated in light of expert opinion and reaction chemistry insight. Hence OECD Validation Principle 4 is fully met.

1.6. Skin sensitization—mechanistic interpretation

The intent of OECD Validation Principle 5 is to ensure that an assessment of the mechanistic association between the descriptors used in the model and the endpoint being predicted has at least been considered. A mechanistic inter-

pretation may be carried out prior to modelling or after modelling. The term *mechanistic basis* refers to the development of a mechanistic hypothesis before modelling (*a priori*). The mechanistic hypothesis is proposed to determine which factors/properties are likely to be significant and thus which descriptors will be best to model that property. In contrast, *mechanistic interpretation* refers to the assignment of physical/chemical/biological meaning to the descriptors after modelling (*a posteriori*). Thus, the descriptors used in the 3D-QSAR models embedded in TIMES-SS are interpreted as to their relevance to the toxicity under consideration.

TIMES-SS contains 236 hierarchically ordered transformations accounting for spontaneous reactions, phase I and II enzyme catalysed biotransformation reactions and reactions with skin proteins. The latter is representative of reaction chemistry that is thought to underpin skin sensitization as described previously by other investigators (Payne and Walsh, 1994; Barratt et al., 1994; Aptula and Roberts, 2006).

The transformations were defined on the basis of empirical and theoretical knowledge and peer reviewed by human experts. This demonstrates a strong mechanistic perspective.

A second component of TIMES-SS are the COREPA 3D-QSARs that model the intrinsic reactivity of some of the alerting groups. These QSARs use transparent and mechanistically interpretable descriptors such as energy HOMO_LUMO, $\log K_{ow}$, charges (*Q*) and MW which are the sort of descriptors typically used to encode penetration, reactivity, and hydrophobicity (Flynn, 1990; Aptula and Roberts, 2006).

On this basis, OECD Validation Principle 5 is clearly met.

2. Discussion

The OECD principles are a valuable means of characterizing a given (Q)SAR model. From the evaluation described so far, the OECD principles have proved useful in describing and documenting the scientific validity of TIMES-SS. As a next step, the information here described will be translated into a QMRF for consideration and potential inclusion into the (Q)SAR inventory currently being developed by the ECB (see: <http://ecb.jrc.it/qsar/qsar-tools/>).

The scientific validity however is just one consideration in the potential application of a given (Q)SAR for regulatory use. The utility of any model's prediction will be context-dependent on the chemical under assessment as well as the regulatory application (hazard identification or risk characterization). This will determine the extent to which the model estimate can be relied upon i.e., as a standalone piece of information or as supporting information as part of an integrated testing strategy. The 4-step applicability domain algorithm and the strong mechanistic basis within TIMES-SS in particular form a robust means of assessing this scientific and regulatory relevance. A challenge for TIMES-SS as well as for some other (Q)SARs surrounds

the translation of the modelled endpoint to one which is more commonly used in a regulatory context. While, the endpoint within TIMES-SS is explicit and hence defined, the potency score predicted still needs to be cross referenced back to an appropriate starting dose descriptor. This will be dependent on the chemical under evaluation and the availability of data for mechanistically related chemicals, for example, the potency score could to be converted into an EC3 range, or the estimate for the new chemical could be compared with other mechanistically related chemicals in order for a read-across to be conducted either through using one of TIMES-SS 3D QSARs or from interpolation with an EC3 value if available. Ideally, TIMES-SS would be underpinned by experimental data from one standardized protocol rather than relying on three different data sources with different categorization schemes. In the future, it is hoped that more LLNA data will become available to realise this and in combination with strategic reactivity testing (see Schultz et al., 2006 for more information), new rules can be developed and existing ones refined so that the performance of TIMES-SS across all potency categories improves.

3. Conclusions

This is the first time a consortium has organized itself to collaborate to develop a model including performing an external evaluation study with the primary aim of fulfilling the OECD principles as far as possible. The 40 chemicals tested have resulted in new LLNA data which should be useful for the development and evaluation of alternative methods in general, but in particular for other sensitization QSAR models. It also represents another example where industry has made available a significant amount of sensitization data to encourage and promote the development of alternative method(s) for assessing this important toxicological endpoint.

The results from the external evaluation study reveals TIMES-SS to be a model that shows great promise in helping to identify skin sensitization potential. The study highlighted several areas for improvements, e.g., minor refinements to the existing transformations (short-term), refinement and extension of reaction chemistry pathways using *in chemico* reactivity data (long-term). Minor refinements are currently being implemented for inclusion into the next version of TIMES-SS.

We propose that this model is adequately characterized and evaluated and could play a useful and important role in the evaluation of sensitization hazard identification as part of a weight of evidence approach to skin sensitization safety assessment.

4. Disclaimer

This article reflects the views and understanding of the corresponding author at the time of writing, and does not necessarily reflect the position of the European Commission.

References

- Aptula, A.O., Patlewicz, G., Roberts, D.W., 2005. Skin sensitization: reaction mechanistic applicability domains for structure–activity relationships. *Chem. Res. Toxicol.* 18, 1420–1426.
- Aptula, A.O., Roberts, D.W., 2006. Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity. *Chem. Res. Toxicol.* 19, 1097–1105.
- Ashby, J., Basketter, D.A., Paton, D., Kimber, I., 1995. Structure activity relationships in skin sensitisation using the murine local lymph node assay. *Toxicology* 103, 177–194.
- Barratt, M.D., Basketter, D.A., Chamberlain, M., Admans, G.D., Langowski, J.J., 1994. An expert system rulebase for identifying contact allergens. *Toxicol. In Vitro* 8, 1053–1060.
- Basketter, D.A., Gerberick, G.F., Kimber, I., Loveless, S.E., 1996. The local lymph node assay: a viable alternative to currently accepted skin sensitization tests. *Food Chem. Toxicol.* 34, 985–997.
- Basketter, D.A., Lea, L.J., Cooper, K., Stocks, J., Dickens, A., Pate, I., Dearman, R.J., Kimber, I., 1999. Threshold for classification as a skin sensitizer in the local lymph node assay: a statistical evaluation. *Food Chem. Toxicol.* 37, 67–74.
- Basketter, D.A., Blaikie, L., Dearman, R.J., Kimber, I., Ryan, C.A., Gerberick, G.F., Harvey, P., Evans, P., White, I.R., Rycroft, R.J., 2000. Use of the local lymph node assay for the estimation of relative contact allergenic potency. *Contact Dermatitis* 42, 344–348.
- Basketter, D.A., Evans, P., Fielder, R.J., Gerberick, G.F., Dearman, R.J., Kimber, I., 2002. Local lymph node assay—validation, conduct and use in practice. *Food Chem. Toxicol.* 40, 3–8.
- Basketter, D.A., Smith Pease, C.K., Patlewicz, G.Y., 2003. Contact allergy: the local lymph node assay for the prediction of hazard and risk. *Clin. Exp. Dermatol.* 28, 8–21.
- Basketter, D., Casati, S., Gerberick, G.F., Griem, P., Philips, B., Worth, A., 2005. Skin sensitisation. *Altern. Lab. Anim.* 33 (Suppl. 1), 83–103.
- Bodin, A., Linnerborg, M., Nilsson, J.L., Karlberg, A.T., 2003. Structure elucidation, synthesis, and contact allergenic activity of a major hydroperoxide formed at autoxidation of the ethoxylated surfactant C12E5. *Chem. Res. Toxicol.* 16, 575–582.
- Bradbury, S.P., Mekenyan, O.G., Kamenska, V.B., Schmieder, P., Ankley, G.T., 2000. A computationally based identification algorithm for estrogen receptor ligands. Part I. Predicting hER binding affinity. *Toxicol. Sci.* 58, 253–269.
- Cooper, J.A., Saracci, R., Cole, P., 1979. Describing the validity of carcinogen screening tests. *Br. J. Cancer* 39, 87–89.
- Cronin, M.T., Basketter, D.A., 1994. Multivariate QSAR analysis of a skin sensitisation database. *SAR QSAR Environ. Res.* 2, 159–179.
- Cronin, M.T., Jaworska, J.S., Walker, J.D., Comber, M.H., Watts, C.D., Worth, A.P., 2003. Use of QSARs in international decision-making frameworks to predict health effects of chemical substances. *Environ. Health Perspect.* 111, 1391–1401.
- Dimitrov, S.D., Low, L.K., Patlewicz, G.Y., Kern, P.S., Dimitrova, G.D., Comber, M.H., Phillips, R.D., Niemela, J., Bailey, P.T., Mekenyan, O.G., 2005a. Skin sensitization: modeling based on skin metabolism simulation and formation of protein conjugates. *Int. J. Toxicol.* 24, 189–204.
- Dimitrov, S., Dimitrova, G., Pavlov, T., Dimitrova, N., Patlewicz, G., Niemela, J., Mekenyan, O., 2005b. A stepwise approach for defining the applicability domain of SAR and QSAR models. *J. Chem. Inf. Model.* 45, 839–849.
- Dupuis, G., Benezra, C., 1982. Allergic Contact Dermatitis to Simple Chemicals: A Molecular Approach. Marcel Dekker, New York and Basel.
- Estrada, E., Patlewicz, G., Chamberlain, M., Basketter, D., Larbey, S., 2003. Computer-aided knowledge generation for understanding skin sensitisation mechanisms: the TOPS-MODE approach. *Chem. Res. Toxicol.* 16, 1226–1235.

- EC, 1999. Proposal for a Regulation of the European Parliament and the Council concerning the Registration, Evaluation, Authorisation and Restrictions of Chemicals (REACH), establishing a European Chemical Agency and amending Directive 1999/45/EC and Regulation (EC). See website: <<http://europa.eu.int/comm/environment/chemicals/whitepaper.htm> <http://europa.eu.int/comm/enterprise/reach/overview.htm>>.
- EU, 2003. Directive 2003/15/EC of the European Parliament and of the Council of 27 February 2003 Amending Council Directive 76/768/EEC on the Approximation of the Laws of the Member States Relating to Cosmetic Products. Official Journal of the European Union L66, pp. 26–35.
- EU, 2006. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 Concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), Establishing a European Chemicals Agency, Amending Directive 1999/45/EC and Repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Official Journal of the European Union L396, pp. 374–375.
- Fedorowicz, A., Zheng, L., Singh, H., Demchuk, E., 2004. Structure–activity models for contact sensitisation. *Int. J. Mol. Sci.* 5, 56–66.
- Flynn, G.L., 1990. Physicochemical determinants of skin absorption. In: Gerrity, T.R., Henry, C.J. (Eds.), *Principles of Route-to-Route Extrapolation for Risk Assessment*. Elsevier, New York, pp. 93–127.
- Gerberick, G.F., Ryan, C.A., Kimber, I., Dearman, R.J., Lea, L.J., Basketter, D.A., 2000. Local lymph node assay: validation assessment for regulatory purposes. *Am. J. Contact Dermat.* 11, 3–18.
- Gerberick, G.F., Ryan, C.A., Kern, P.S., Dearman, R.J., Kimber, I., Patlewicz, G.Y., Basketter, D.A., 2004. A chemical dataset for evaluation of alternative approaches to skin-sensitization testing. *Contact Dermatitis* 50, 274–288.
- Gerberick, G.F., Ryan, C.A., Kern, P.S., Schlatter, H., Dearman, R.J., Kimber, I., Patlewicz, G.Y., Basketter, D.A., 2005. Compilation of historical local lymph node data for evaluation of skin sensitisation alternative methods. *Dermatitis* 16, 157–202.
- Gerner, I., Barratt, M.D., Zinke, S., Schlegel, K., Schlede, E., 2004. Development and prevalidation of a list of structure–activity relationship rules to be used in expert systems for prediction of the skin-sensitising properties of chemicals. *Altern. Lab. Anim.* 32, 487–509.
- Hansch, C., Maloney, P.P., Fujita, T., Muir, R.M., 1962. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. *Nature* 194, 178–180.
- Hansch, C., Fujita, T., 1964. Rho–sigma–pi analysis. A method for the correlation of biological activity with chemical structure. *J. Am. Chem. Soc.* 86, 1616–1626.
- Jaworska, J.S., Comber, M., Auer, C., van Leeuwen, C.J., 2003. Summary of a workshop on regulatory acceptance of (Q)SARs for human health and environmental endpoints. *Environ. Health Perspect.* 111, 1358–1360.
- Kimber, I., Dearman, R.J., Scholes, E.W., Basketter, D.A., 1994. The local lymph node assay: developments and applications. *Toxicology* 93, 13–31.
- Kimber, I., Basketter, D.A., Butler, M., Gamer, A., Garrigue, J.L., Gerberick, G.F., Newsome, C., Steiling, W., Vohr, H.W., 2003. Classification of contact allergens according to potency: proposals. *Food Chem. Toxicol.* 41, 1799–1809.
- Landsteiner, K., Jacobs, J.L., 1936. Studies on the sensitisation of animals with simple chemicals III. *J. Exp. Med.* 64, 625–639.
- Mekenyan, O.G., Ivanov, J.M., Karabunaliev, S., Bradbury, S.P., Ankley, G.T., Karcher, W., 1997. A computationally based hazard identification algorithm that incorporates ligand flexibility. I. Identification of potential androgen receptor ligands. *Environ. Sci. Technol.* 31, 3702–3711.
- Mekenyan, O.G., Nikolova, N., Karabunaliev, S., Bradbury, S., Ankley, G.T., Hausen, B., 1999. New developments in a hazard identification algorithm for hormone receptor ligands. *Quant. Struct.-Act. Relat.* 18, 139–153.
- Miller, M.D., Yourtee, D.M., Glaros, A.G., Chappelow, C.C., Eick, J.D., Holder, A.J., 2005. Quantum mechanical structure–activity relationship analyses for skin sensitisation. *J. Chem. Inf. Model.* 45, 924–929.
- Moloney, S.J., Teal, J.J., 1988. Alkane-induced edema formation and cutaneous barrier dysfunction. *Arch. Dermatol. Res.* 280, 375–379.
- Netzeva, T.I., Worth, A., Aldenberg, T., Benigni, R., Cronin, M.T., Gramatica, P., Jaworska, J.S., Kahn, S., Klopman, G., Marchant, C.A., Myatt, G., Nikolova-Jeliazkova, N., Patlewicz, G.Y., Perkins, R., Roberts, D., Schultz, T., Stanton, D.W., van de Sandt, J.J., Tong, W., Veith, G., Yang, C., 2005. Current status of methods for defining the applicability domain of (quantitative) structure–activity relationships. The report and recommendations of ECVAM Workshop 52. *Altern. Lab. Anim.* 33, 155–173.
- OECD, 1992. Organisation for Economic Cooperation and Development. Guidelines for Testing of Chemicals No. 406. Skin sensitisation <<http://www.oecd.org>>.
- OECD, 2002. Organisation for Economic Cooperation and Development. Guidelines for Testing of Chemicals No. 429. Skin sensitisation: The Local Lymph Node Assay <<http://www.oecd.org>>.
- OECD, 2004. The Report from the Expert Group on (Quantitative) Structure–Activity Relationships [(Q)SARs] on the Principles for the Validation of (Q)SARs. ENV/JM/TG(2004)27/REV. Organisation for Economic Cooperation and Development, Paris.
- OECD, 2007. Guidance Document on the Validation of (Quantitative) Structure–Activity Relationships [(Q)SAR] Models. ENV/JM/MONO(2007)2. Organisation for Economic Cooperation and Development, Paris <<http://www.oecd.org/dataoecd/55/35/38130292.pdf>>.
- Patlewicz, G.Y., Basketter, D.A., Pease, C.K., Wilson, K., Wright, Z.M., Roberts, D.W., Bernard, G., Arnau, E.G., Lepoittevin, J.P., 2004. Further evaluation of quantitative structure–activity relationship models for the prediction of the skin sensitization potency of selected fragrance allergens. *Contact Dermatitis* 50, 91–97.
- Payne, M.P., Walsh, P.T., 1994. Structure–activity relationships for skin sensitization potential: development of structural alerts for use in knowledge-based toxicity prediction systems. *J. Chem. Inf. Comput. Sci.* 34, 154–161.
- Roberts, D.W., Williams, D.L., 1982. The derivation of quantitative correlations between skin sensitisation and physico-chemical parameters for alkylating agents and their application to experimental data for sulfones. *J. Theo. Biol.* 99, 807–825.
- Roberts, D.W., Aptula, A.O., Patlewicz, G., 2007. Electrophilic chemistry related to skin sensitization. Reaction mechanistic applicability domain classification for a published dataset of 106 chemicals tested in the mouse local lymph node assay. *Chem. Res. Toxicol.* 20, 44–60.
- Rodford, R., Patlewicz, G., Walker, J.D., Payne, M.P., 2003. Quantitative structure–activity relationships for predicting skin and respiratory sensitization. *Environ. Toxicol. Chem.* 22, 1855–1861.
- Ruet Rossignol, M., 2005. The 7th Amendment to the Cosmetics Directive. *Altern. Lab. Anim.* 33 (Suppl. 1), 19–20.
- Schlede, E., Aberer, W., Fuchs, T., Gerner, I., Lessmann, H., Maurer, T., Rossbacher, R., Stropp, G., Wagner, E., Kayser, D., 2003. Chemical substances and contact allergy—244 substances ranked according to allergenic potency. *Toxicology* 193, 219–259.
- Schultz, T.W., Carlson, R.E., Cronin, M.T.D., Hermens, J.L.M., Johnson, R., O'Brien, P.J., Roberts, D.W., Siraki, A., Wallace, K.B., Veith, G.D., 2006. A conceptual framework for predicting the toxicity of reactive chemicals: modelling soft electrophilicity. *SAR QSAR Env. Res.* 17, 413–428.
- Smith Pease, C.K., Basketter, D.A., Patlewicz, G.Y., 2003. Contact allergy: the role of skin chemistry and metabolism. *Clin. Exp. Dermatol.* 28, 177–183.
- Woolhiser, M.R., Hayes, B.B., Meade, B.J., 1998. A combined murine local lymph node and irritancy assay to predict sensitization and irritancy potential of chemicals. *Toxicol. Methods* 8, 245–256.
- Worth, A.P., Bassan, A., Gallegos, A., Netzeva, T.I., Patlewicz, G., Pavan, M., Tsakovska, I., Vracko, M., 2005. The Characterisation of (Quantitative) Structure–Activity Relationships: Preliminary Guidance. EUR 21866 EN <<http://ecb.jrc.it/>>.
- Worth, A.P., Bassan, A., de Bruijn, J., Gallegos Saliner, A., Netzeva, T., Patlewicz, G., Pavan, M., Tsakovska, I., Eisenreich, S., 2007. The role of the European Chemicals Bureau in promoting the regulatory use of (Q)SAR methods. *SAR QSAR Env. Res.* 18, 111–125.