

A Minireview of Available Skin Sensitization (Q)SARs/Expert Systems

Grace Patlewicz^{a*}, Aynur O. Aptula^b, David W. Roberts^c and Eugenio Uriarte^d

^a European Commission, DG Joint Research Centre, Institute for Health and Consumer Protection, European Chemicals Bureau (ECB) TP 582, 21020 Ispra (VA), Italy, E-mail: grace.patlewicz@ec.europa.eu, patlewig@hotmail

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

^c Liverpool John Moores University, Liverpool, UK

^d Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela, 15872 Santiago de Compostela, Spain

Keywords: Expert systems, (Q)SAR models, REACH, Skin sensitization

Received: May 31, 2007; Accepted: September 3, 2007

DOI: 10.1002/qsar.200710067

Abstract

This minireview describes the state-of-the-art of available (Q)SARs/expert systems for skin sensitization and evaluates their utility for potential regulatory use. It is one of a series of minireviews in this journal and is based on an in-depth review performed by the European Chemicals Bureau (ECB) of the European Commission's Joint Research Centre (JRC) in support of the development of technical guidance for the implementation of the Registration, Evaluation, and Authorization of Chemicals (REACH) legislation. Most existing models fall into one of two main categories: either they are local in nature, usually specific to a chemical class or reaction chemical mechanism or else they are global in form, derived empirically using statistical methods. Some global QSARs have been recently characterized in accordance with the OECD validation principles and shown to be of limited use for regulatory applications. These have been evaluated in more detail elsewhere. Expert systems capable of predicting skin sensitization are discussed briefly. The strong mechanistic understanding of skin sensitization has facilitated the development of relative alkylation index (RAI) models. The RAI approach, relying on reactivity and hydrophobicity as the key parameters, actually appears to be the most promising means of deriving robust and mechanistically interpretable models which can be applied in a regulatory context. These are now referred to as Quantitative Mechanistic Models (QMM). Whilst hydrophobicity is conveniently modelled using log *P*, reactivity is often not easily determined from chemical structure. Initiatives are in progress to generate the necessary reactivity data for reactions relevant to skin sensitization but more resources are still required. A strategy of how information from using the RAI approach can be used in the evaluation of skin sensitization potential is described. This is the subject of on-going and future work.

Abbreviations: GPMT, guinea pig maximization test; ITS, integrated testing strategy; LLNA, local lymph node assay; OECD, Organization for Economic Co-operation and Development; OPS, optimal prediction space; QMM, quantitative mechanistic models; (Q)SAR, (quantitative) structure–activity relationship; RAI, relative alkylation index; REACH, registration, evaluation, and authorization of chemicals; TIMES-SS, tissue metabolism simulator for skin sensitization; TOPKAT, toxicity prediction komputer-assisted technology; TOPS-MODE, topological sub-structural molecular design.

1 Introduction

Skin sensitization is a significant environmental and occupational health concern. Workers and consumers being sensitized are a major problem for individuals, for employers and for marketing certain products. Consequently there exists an important need to accurately identify chemicals that have the potential to cause skin sensitization [1–3]. Sensitization is also an endpoint that requires assessment under the new European Union (EU) Registration, Evaluation and Authorization of Chemicals (REACH) programme. Under REACH, sensitizing potential needs

to be assessed for chemicals above the 1 tonne threshold according to Annex VII [4].

Current assessment of skin sensitization still relies on animal tests. Guinea pigs [5] have been the species of choice for predictive sensitization tests for many years; however, under REACH, the test of first choice is now the mouse Local Lymph Node Assay (LLNA) [6]. Further information describing the LLNA and its quantitative output, the EC3, can be found in Ref. [7–13].

Under REACH, all chemicals produced or imported > 1 tonnes *per annum* (tpa) in the EU will need to be assessed for human and environmental hazards. If conducted by means of the present data requirements/test strategy, this assessment will use a huge number of test animals and will be neither resource nor time efficient. To address this challenge, REACH calls for an increased use of alternatives such as *in vitro* methods, (quantitative) structure–activity relationships [(Q)SARs] and chemical grouping approaches (see Annex XI of Ref. [4]). Since no *in vitro* replacement is currently available for skin sensitization, nor expected to be ready in the near future [14, 15], the use of (Q)SAR approaches presents an attractive solution. The application of (Q)SARs, raises with it a number of issues including: the availability of models; the extent to which they are able to play an effective role in the assessment of skin sensitization hazard and potency as well as their validity and applicability.

Several of these issues were first considered in 2002, whereupon a number of activities were initiated in an effort to increase the regulatory acceptance of (Q)SARs. Perhaps the most notable of these was the development of a set of principles for (Q)SAR Validation [16]. These were subsequently discussed and agreed within the Organization for Economic Co-operation and Development (OECD) and are now referred to as the ‘OECD Principles for (Q)SAR Validation’ [17, 18]. The principles provide a practical framework for assessing the scientific validity of a (Q)SAR model. However, (Q)SAR model validity is just one of the factors to consider in the use and acceptance of (Q)SARs. Other considerations concerning (Q)SARs are their application in the context of endpoint specific Integrated Testing Strategies (ITSs). These are discussed in more detail elsewhere [19, 20].

Whilst previous reviews for skin sensitization have been conducted by Karol *et al.* [21], Bashir and Maibach [22] and Rodford *et al.* [23], this minireview will outline the different approaches and philosophies that have been used to develop non-animal predictive models. Furthermore, it will provide an objective assessment of the usability of these models under REACH using the OECD principles as a reference. Some of the challenges in terms of where future research could be directed are also highlighted. As background information, a brief description of the mechanism of skin sensitization induction is provided.

Skin sensitization is the process following epicutaneous application of a substance to the skin which results in a T-

cell-mediated immunological response specific for the substance. Skin sensitization involves two phases, an induction and elicitation phase. Briefly, the process of induction is as follows: a chemical penetrates through the stratum corneum into the viable epidermis where it binds to skin proteins/peptides. The resulting complex is internalized by Langerhans Cells (LCs) resident in the skin. This then begins a maturation and migration process by which the LCs process the complex into a form (an antigen) that can be recognized by T-cells, and travel to the draining lymph nodes where they present it to naïve T-cells. This invokes the stimulation and proliferation of T-cells, some of which form a population of so-called memory T-cells. This process is known as skin sensitization induction. Figure 1 outlines the chain of events. Note, induction exhibits no clinical signs, these manifest themselves upon re-exposure to the same chemical during the elicitation phase. Here, memory T-cells release cytotoxic mediators which result in the local tissue inflammation characteristics of sensitization – redness, swelling and itching. For a more detailed account of the current understanding of the biological and chemical mechanisms of skin sensitization, the reader is referred to several excellent reviews [3, 24–26].

The key determining step in sensitization induction, is the binding of a chemical to skin protein. Binding is believed to be covalent in nature such that the chemical behaves as an electrophile and the protein as a nucleophile. The covalent hypothesis was actually first articulated by Landsteiner and Jacobs [27] in 1936 and followed up by others (*e.g.* Dupuis and Benezra, [28]). Chemicals must either be inherently protein reactive or converted in the skin to a protein reactive species to induce sensitization. Chemicals that are unable to associate effectively with proteins will fail to stimulate an immune response [25].

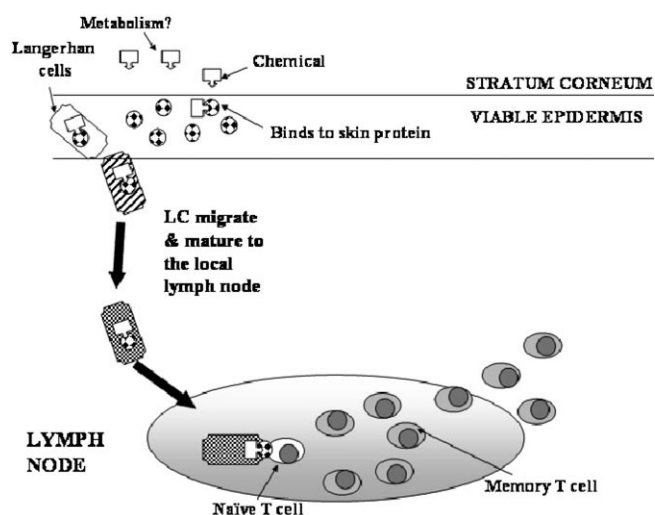


Figure 1. Diagrammatic representation of the molecular events thought to be involved in the process of skin sensitization.

2 (Q) SARs and Expert Systems

A number of different approaches and philosophies have been used to develop (Q)SAR models and expert systems for skin sensitization. Available published (Q)SAR models have been chemical class-based/mechanism-based or derived empirically using statistical approaches. Expert systems for skin sensitization include TOXicity Prediction Komputer-Assisted Technology (TOPKAT) (Accelrys Inc., San Diego, CA, USA) and M-CASE (Computer-Automated Structure Evolution; MultiCASE Inc., Cleveland, Ohio, USA) which are both statistically based, Derek for Windows (DfW, LHASA Ltd, Leeds, UK) and Hazard Expert (Compudrug, Budapest, Hungary) which are knowledge-based and TIssue Metabolism Simulator for Skin Sensitization (TIMES-SS; LMC, University of Bourgas, Bulgaria) which is a hybrid [29, 30].

2.1 Structure–Activity Relationships (SARs)

Early efforts in correlating chemical groups to skin sensitization guinea pig data are exemplified in work by Rao *et al.* [31], Benezra *et al.* [32] and Hausen and coworkers [33–39]. More recent efforts include the development of structural alerts by Barratt *et al.* [40], Payne and Walsh [41], Ashby *et al.* [42] and Gerner *et al.* [43]. In Barratt *et al.* [40], common reaction mechanisms (*e.g.* acylating agents, alkylating agents, Michael electrophiles, aldehydes *etc.*) were identified following the grouping of the chemicals into potency categories. Forty structural alerts were extracted either on the basis of these reaction mechanisms or empirically. In Ref. [41], a set of structural alerts were defined based on the structural requirements for reactivity with skin proteins using existing experimental skin sensitization data. The alerts were classified, where possible, according to anticipated reaction mechanism. Ashby *et al.* [42] suggested broad SARs using a dataset of 106 chemicals that had been tested in the LLNA. The 73 sensitizers within the dataset, were grouped into six main classes, partly based on structural criteria, and partly on reaction mechanistic domains. Gerner *et al.* [43] derived two sets of new structural alerts using a regulatory database; 15 rules for directly acting sensitizers and three rules for chemicals that were sensitizers following metabolic activation. These efforts served to define the types of functional groups that were predictive of skin sensitization. In certain cases, they were incorporated into the expert system, DfW as alerts. These structural alerts provide a useful means of identifying potential sensitizers as a first stage in an ITS. They can also facilitate the discovery of similar analogues for use in a read-across evaluation.

2.2 Statistical QSARs

Probably the first global models for sensitization were those developed by Magee and Hostynek, [44, 45]. They investigated the feasibility of deriving models for fragrance allergens using classification and ranking approaches. The models developed are statistically sound (see equations in Table 1) and to an extent mechanistically interpretable. The classification model predicts the absence or presence of sensitization whilst the rank model addresses potency through the use of a modified version of the Magnusson–Kligman categorization [46]. The models' applicability domain is not systematically defined, aside from the fact that the models are trained on fragrance allergens. It is thought that these models could show promise in screening potential sensitizers as part of an ITS.

Cronin and Basketter [47], used stepwise Linear Discriminant Analysis (LDA) on a set of 259 chemicals [mostly tested in the Guinea Pig Maximization Test (GPMT) in the spirit of the OECD guidelines] to produce a 14 parameter model for the prediction of sensitization. 12 of the 14 descriptors in the final model were structural features associated with reactivity. The other two descriptors were the Shannon index [48], a measure of molecule size and the difference in HOMO and LUMO energies, a measure of reactivity [49] (see Table 1). The model predicted 82.6% of compounds correctly after cross-validation but had a tendency to predict non-sensitizers better than sensitizers (88% correctly predicted *cf.* 76%). The authors of this publication concluded that structural alerts might have been a better means of identifying potential sensitization hazard since the statistical model was poor at discriminating between the two classes.

Devillers [50] used the dataset of Ref. [47] to derive a neural network model relating the same 14 descriptors to sensitization outcome. A three-layer feed forward neural network trained by the back-propagation algorithm previously described [51, 52] was used on a training set of 242 chemicals and a test set of 17 chemicals. The total percentage of correct classifications was 89.19% (*cf.* 82.6% in Cronin and Basketter [47]). This model does appear to be better in terms of its predictive performance though there are notable tradeoffs – greater complexity in terms of the modelling approach used; non-linear *versus* linear and a lack of transferability. These tradeoffs are not insurmountable in terms of demonstrating scientific validity as evidenced in Vracko *et al.* [53] but since Devillers' [50] primary aim was to demonstrate the improved performance of the neural network compared with an LDA rather than trying to predict sensitization potential or its underlying mechanism, the performance statistics alone would be insufficient to merit the use of this model over another less complex one.

Fedorowicz *et al.* [54] developed a logistic regression model for LLNA sensitising potential starting from a LLNA dataset of 54 chemicals (comprising 25 sensitizers

Table 1. Table summarizing selected skin sensitization models.

References	Type of (Q)SAR model	Method	No. of training set	No. of test set	Outliers	Performance characteristics	Algorithm	Comments
Benezra <i>et al.</i> [32]	SARs	Structural alerts					Derived 63 structural alerts for skin sensitization in the Structure-Activity Tree (SAT)	
Hausen [33–39]	SARs	Chemical class specific SARs					A number of correlation studies on gallates, coumarins, sesquiterpene lactones, benzoquinones, naphthoquinones, primin (2-methoxy-6-pentyl-1,4-benzoquinone) analogues, non-terpenoid and diterpenoid phenanthrenequinones (PACs)	
Rao <i>et al.</i> [31]	SARs	SARs					Various chemical groups evaluated	
Payne and Walsh [41]	SARs	Structural alerts					Eight broad structure activity classes	
Ashby <i>et al.</i> [42]	SARs	Structural alerts					Six broad structure activity classes identified for the 73 sensitizers in the dataset of 106 chemicals	
Gerner <i>et al.</i> [43]	SARs	Structural alerts					Derived two sets of new structural alerts using a regulatory database of sensitizers; 15 rules for directly acting sensitizers and three rules for chemicals that were sensitizers following metabolic activation	
Hostynek and Magee [45]	Global statistical model	Discriminant analysis	70 Comprising 36 non-sensitizers, 36 sensitizers		2	$R^2 = 0.75$ $s = 0.275$ $F = 15.81$	Classification score = $0.0116 MR - 0.161 PL + 0.0498 HBA - 0.129 HBD - 0.284 ICO - 0.571 CONJ + 0.161 IX + 0.596 IArOH + 0.503 IArNH_2 + 0.207 IOH + 0.512 IUNIQ + 0.118$ Ranking score = $0.0294 MR - 0.281 PL - 0.0960 HBA - 0.207 HBD - 0.409 ICOOR + 0.547 IOH + 0.519 IX + 1.375 IPOS + 1.39 IQUIN + 1.7$	Developed on fragrance chemicals. Could be useful for a screening purposes as part of an ITS
Hostynek and Magee [45]	Global statistical model	Discriminant analysis	88 Comprising 29 non-sensitizers, 18 weak sensitizers, 15 moderate sensitizers and 26 strong sensitizers			$R^2 = 0.726$ $s = 0.68$ $F = 23$		Developed on fragrance chemicals. Could be useful for screening purposes as part of an ITS
Cronin and Basketter [47]	Global Statistical Model	Discriminant analysis	259 Comprising 120 sensitizers, 139 non-sensitizers			Jackknife classification: 82.6% correct classification, 76% correctly predicted sensitizers, 88% correctly predicted non-sensitizers	Class-sensitizers = $-8.08 Diff of HOMO-LUMO + 7.19 Phenylesters + 5.96 Aldehydes + 47.41 Shannon Index + 5.78 Putative formaldehyde donors + 15.67 Primary aromatic amines + 1.07 Epoxy groups + 4.24 Acid chloride, bromides, azides + 12.87 Acid anhydrides + 15.61 Halogenated aromatics with 2 or more nitro groups + 9.61 alpha,beta-unsat esters + 28.04 Thioesters + 11.98 Hydroquinones, catechols and o-alkyl precursors + 17.58 Benzoquinone or similar - 64.81$	

Table 1. (cont.)

References	Type of (Q)SAR model	Method	No. of training set	No. of test set	Outliers	Performance characteristics	Algorithm	Comments
Devillers [50]	Global Statistical Model	Neural network	242	17		89.19% correct classification, 83.19% correctly predicted sensitizers, 94.29% correctly predicted non-sensitizers	Class-non-sensitizers = $-8.57 \text{ Diff of HOMO-LUMO} + 2.75 \text{ Phenyl esters} + 2.68 \text{ Aldehydes} + 49.41 \text{ Shannon Index} + 1.97 \text{ Putative formaldehyde donors} + 12.31 \text{ Primary aromatic amines} - 4.61 \text{ Epoxy groups} - 1.22 \text{ Acid chloride, bromides, azides} + 7.81 \text{ Acid anhydrides} + 10.72 \text{ Halogenated aromatics with 2 or more nitro groups} + 7.57 \text{ alpha,beta-unsat esters} + 24.03 \text{ thioesters} + 9.93 \text{ Hydroquinones, catechols and o-alkyl precursors} + 14.19 \text{ Benzoquinone or similar} - 69.92$ A three-layer feed forward neural network (14/5/1) trained by the back-propagation algorithm on the same 14 parameters as used in Ref. [46]	Only discriminates between sensitizers and non-sensitizers. Could be useful for screening purposes as part of an ITS
Fedorowicz <i>et al.</i> [54]	Global Statistical Model	Logistic regression	54 Comprising 25 sensitizers, 29 non-sensitizers			83% correct classification, 72% correctly predicted sensitizers, 93% correctly predicted non-sensitizers	Logistic regression model outcome = $-1.63 + 1.46nDB + 183.9GATSom - 5.97HATS6e + 3.02C-003$, where nDB no of Double bonds; $GATSom$ is a mass weighted Geary graph spatial autocorrelation coefficient of the sixth lag; $HATS6e$ is the GETAWAY descriptor weighted by the atomic Sanderson electronegativities; $C-003$ atom centred fragments descriptor indicating the presence of the CHR3 molecular subfragment	No mechanistic interpretation reported. Limited dataset. Large number of screening variables
Fedorowicz <i>et al.</i> [57]	Global Statistical Model	Logistic regression	178 Comprising 132 sensitizers, 46 non-sensitizers	25, 92% correct classification, 100% correctly predicted sensitizers, 33.3% correctly predicted non-sensitizers		83.2% correct classification, 94.7% correctly predicted sensitizers, 50% correctly predicted non-sensitizers	Logistic regression model outcome (LLNA) = $15.2 - 0.345Hy + 0.351Gmin - 7.76RARS - 4.09BELel - 0.259Dipole$, where Hy hydrophobic factor; $Gmin$ minimum E state; RAR R matrix average row sum; $BELel$ lowest eigenvalue, no. of Burden matrix/weighted by atomic Sanderson electronegativities; $Dipole$ dipole moment	Only discriminates between sensitizers and non-sensitizers. No mechanistic interpretation reported. Unbalanced dataset. Very low specificity

Table 1. (cont.)

References	Type of (Q)SAR model	Method	No. of training set	No. of test set	Outliers	Performance characteristics	Algorithm	Comments
Fedorowicz <i>et al.</i> [57]	Global Statistical Model	Logistic regression	105 Comprising 82 sensitizers, 23 non-sensitizers			87.6% correct classification, 98.8% correctly predicted sensitizers, 47.8% correctly predicted non-sensitizers	Logistic regression model outcome (Guinea pig) = $9.53 - 2.01H_{max} - 19.9(R3e+) - 15.8(R3u+) - 12.9HATS1u$, where H_{max} maximum H electropotential state; $R3e+$ maximal autocorrelation of Lag3/weighted by atomic Sanderson electronegativities; $R3u+$ maximal autocorrelation of Lag3/unweighted; $HATS1u$ leverage weighted autocorrelation of Lag1/unweighted	Only discriminates between sensitizers and non-sensitizers. No mechanistic interpretation reported. Unbalanced dataset. Very low specificity.
Li <i>et al.</i> [58]	Global Statistical Model	Classification tree	178 Comprising 132 sensitizers, 46 non-sensitizers			99.2% sensitivity, 34% specificity	Classification tree contained the following 4 descriptors: $Atype$ C-3 count of atom type CHR3; $nssCH$ count of atom type; $SssCH$ sum of atom type E state; Hy hydrophilic factor	Large number of screening variables Only discriminates between sensitizers and non-sensitizers. No mechanistic interpretation reported. Unbalanced dataset. Very low specificity.
Li <i>et al.</i> [58]	Global Statistical Model	Random forest	178 Comprising 132 sensitizers, 46 non-sensitizers			At a threshold of 0.7: 84% sensitivity 83% specificity	The best random forest contained the following ten descriptors: $Gmin$, $HATS1u$, $HATS4u$ leverage weighted autocorrelation of Lag4/unweighted; $Hmaxpos$ max positive H E-state; $Mor3p$ 3D Morse signal 32/weighted by atomic polarizabilities; $R4e+$ maximal autocorrelation of Lag4/weighted by atomic Sanderson electronegativities; $R4u+$ R maximal autocorrelation of Lag4/unweighted; $SEigp$ eigenvalue sum from polarizability weighted distance matrix; $SssCH$; $sumdelI$ sum of delta I values	Only discriminates between sensitizers and non-sensitizers. No mechanistic interpretation reported. Large number of screening variables
Estrada <i>et al.</i> [60]	Global statistical model	Discriminant analysis	N (total) = 93 78	9 (CV test set), 3 1 compound was indeterminate, 7/8 chemicals were correctly classified, 15 (external test validation). Of the 8 strong/moderate chemicals, 1 was incorrectly predicted and a second indeterminate – hence 6/7 strong/		$Wilks - lambda = 0.61$ $F = 3.39$ $D2 = 2.52$ $p < 0.0007$	Class1 model = $(1.331 * \mu1H) - (0.00598 * \mu4H) + (0.0078 * \mu2PS) - (0.00021366 * \mu3PS) + (0.0755 * \mu1MR) + (0.0319 * \mu2MR) - (0.0011133 * \mu2Pol) - (2.3797 * \mu1Ch) + (0.1547 * \mu3Ch) + (0.00425 * \mu6Ch) + (2.0932 * \mu1vDW) - (0.8683 * \mu2vDW) + 0.7954$, where μ denotes the spectral moment to a power and weight by one of the following parameters: H , hydrophobicity; PS , polar surface area; Pol , polarizability;	Predicts a crude potency classification. Some of the mechanistic interpretations are based on erroneous arguments [30,56]. External validation not rigorously designed so good performance might be on chance alone. Evaluation against

Table 1. (cont.)

References	Type of (Q)SAR model	Method	No. of training set	No. of test set	Outliers	Performance characteristics	Algorithm	Comments
Estrada <i>et al.</i> [60]	Global statistical model	Discriminant analysis	N (total) = 42 36	6 (CV test set), 5/6 chemicals correctly classified		$Wilks-lambda = 0.38$, $F = 4.63$, $D2 = 8.76$, $p < 0.001$	MR , molar refractivity; Ch , charges; νDW , van der Waal radii Class2 model = $(0.946 * \mu H) - (0.00468 * \mu H) - (0.894 * \mu PS) + (0.1004 * \mu PS) - (0.0024 * \mu PS) + (0.0057 * \mu PS) - (1.429 * \mu Ch) + (0.0053 * \mu Ch) - (0.00111 * \mu Ch) - 5.309$	a separate dataset found the specificity to be as low as 10% [30] Predicts a crude potency classification. Some of the mechanistic interpretations are based on erroneous arguments [30,56] Use of %EC3 seems flawed [56], molar conversions were considered by the authors but given the success of the resulting model, this was not carried out. Removal of 20 outliers on account of poor statistical fit. Methodology – heuristic algorithm to derive a correlation is not further elaborated. Large number of variables screened relative to number of observations Predicts a crude potency classification
Miller <i>et al.</i> [68]	Global statistical model	Heuristic algorithm	N (total) = 87 (training set 65 and test set 22). Remaining 67 chemicals were analysed after removal of 20 outliers: 50 as a training set and 17 as a test set	17	20	$N = 50$ $R^2 = 0.773$ $R^2_{adj} = 0.763$ $R_{CV} = 0.738$ $F = 79.9$	$EC3 = 9.16$ $FPSA2ESP + 4.29$ $EHOMO-LUMO - 45.89$	count of poor statistical fit. Methodology – heuristic algorithm to derive a correlation is not further elaborated. Large number of variables screened relative to number of observations Predicts a crude potency classification
Miller <i>et al.</i> [68]	Global statistical model	Classification model	50	17, 14/17 were correctly predicted		44/50 correctly predicted	A bin selection method was carried out to predict potency classification	
Li <i>et al.</i> [71]	Global Statistical Model	Stepwise linear & PLS logistic regression	N (total) = 132 91 comprising non-weak chemicals and 41 strong-extreme sensitizers	15, Comprising 10 non-weak and 5 strong-extreme sensitizers, 80–86.7% accuracy, 60% sensitivity, 90–100% specificity		79.5–85.6% accuracy, 68.3–78% sensitivity and 80.2–93.4% specificity depending on cutoffs	Stepwise linear logistic regression (using standardized and und standardized descriptors) and PLS logistic regression Stepwise Logistic regression equation on the basis of standardized descriptors shown only: $Logit(P) = -1.745 + 1.119 \in$ $3(hbd,aro) + 2.686 \in$ $12(hsaro) - 0.9722 \in 18(all,all) + 1.654 \in 2(np,p) - 0.9936 \in 2(all,all) - 0.8841 \in 1(p,-hbd) - 3.966 \in 19(hsall) + 1.466 \in 54(all,all) + 1.388 \in 39(all,all)$	Model is limited to discriminating between strong-extreme and non-sensitizers. Limited mechanistic interpretability
Ren <i>et al.</i> [69]								

Table 1. (cont.)

References	Type of (Q)SAR model	Method	No. of training set	No. of test set	Outliers	Performance characteristics	Algorithm	Comments
	Global Statistical Model	Stepwise forward discriminant analysis	N (total) = 131 88	43 67.44% of the compounds were correctly classified		$F = 9.632$, $Wilks - \lambda = 0.584$, 79.55% of the 88 compounds correctly classified	Class 1 (strong/mod sensitizers) = $-6.587 + 12.7RN + 0.001Gi + 0.712KNInd + 0.23YZS + 6.833PCMax; C + 118.414RIMaxI, C$. Class 2 (non-strong/mod sensitizers) = $-8.345 - 21.8RN + 0.004Gi - 1.384KNInd + 0.427YZS - 50.309PCMax; C + 26.952RIMaxI, C$, where RN Relative number of N atoms Gi gravitation index; $KNInd$ Kier and Hall index (order 3); YZS YZ shadow; $PCMax, C$ Max partial charges for a C atom; $RIMaxI, C$ Max 1-electron react. index for a C atom	Model discriminates between strong/moderate sensitizers and non-strong/moderate sensitizers. Limited mechanistic interpretability. Large number of screening variables
Ren <i>et al.</i> [69]	Global statistical model	SVM	N (total) = 131 88	43, 72.09% of the compounds were correctly classified (non-linear). 67.44% of the compounds were correctly classified (linear)		89.77% correct classification (non-linear). 82.95% correct classification (linear)		Model discriminates between strong/moderate sensitizers and non-strong/moderate sensitizers. Limited mechanistic interpretability. No defined algorithm. Large number of screening variables
DTWs [40, 72]	Expert system	Knowledge-based				Performance is difficult to assess as training sets underpinning alerts are hidden. Several evaluation studies have been performed but results very dependent on the datasets applied	Version 9 contains 64 alerts for skin sensitization where an alert is a toxicophore, a substructure known or thought to be responsible for the toxicity alongside associated literature references, comments and examples	Training set hidden. Applicability domain is on the basis of structural rules which identify sensitizers but are not able to definitively exclude non-sensitizers. No alert identified implies no knowledge NOT that the chemical is safe. Of potential use in an ITS
TOPKAT [77, 78]	Expert system	Statistically based – 2 group discriminant model	Original models were developed on the basis of data for 315 chemicals. Current version 6.2 contains data for an additional 20 chemicals. Model 1 (Aliphatics – Sens/Non) 252.			Several evaluation studies have been performed but results very dependent on the datasets applied. Reported crossvalidated (LOO) by de-	Version 6.2 contains 2 sets of models – one set for discriminating between sensitizers and non-sensitizers and a second set for discriminating between strong and weak/moderate sensitizers. Each set comprises a model based on aliphatic compounds and those containing 1 benzene ring only and a model	Predicts the potency classification of GPMs. Of potential use within an ITS

Table 1. (cont.)

References	Type of (Q)SAR model	Method	No. of training set	No. of test set	Outliers	Performance characteristics	Algorithm	Comments
			Model 2 (Aromatics – Sens/Non) 75. Model 3 (Aliphatics – Strong vs Weak/Mod) 158. Model 4 (Aromatics – Strong vs Weak/Mod) 59			velopers. Model 1: 91% specificity; 93% sensitivity; 3% indeterminate. Model 2: 81% specificity; 95% sensitivity; 1% indeterminate. Model 3: 89% specificity; 85% sensitivity; 6% indeterminate. Model 4: 90% specificity; 92% sensitivity; 3% indeterminate	based on aromatic compounds. An OPS applicability domain is defined for these models. A structural similarity functionality is encoded into the system to retrieve similar compounds with associated toxicity data. Training set is partly accessible in the system	
TIMES – SS [30, 86]	Expert system	Hybrid expert system	729	40, Concordance 83% (diverse representative test set chosen by experimental design and consideration of the applicability domain)			TIMES-SS comprises a series of structure toxicity and structure metabolism relationships through a number of transformations simulating skin metabolism and interaction of the generated reactive metabolites with skin proteins; some of these are underpinned by 3D QSARs	Predicts a potency classification developed by harmonizing the endpoint score from GPMT, BgVV and LLNA datasets. Of potential use within an ITS
Roberts and Williams [87] and related references [88–99, 102]	Local RAI models	RAI approach					RAI approach has been successfully applied to many different datasets including primary alkyl bromides, sulfonates, acrylates, sulphonates, sulphonate esters, unshielded analogues, halo- and pseudohalo-benzenes, aldehydes and diketones	Of limited applicability but relationships developed are robust and mechanistically interpretable
Basketter <i>et al.</i> [88]	Local RAI	Regression	9			$R = 0.97$ $s = 0.11$ $F = 50$	Primary alkyl bromides, $\text{Log}(1/\text{TC3}) = 1.61 \text{Log } P - 0.09 \text{Log } P^2 - 7.4$	Of limited applicability but relationships developed are robust and mechanistically interpretable
Patlewicz <i>et al.</i> [95]	Local RAI	Regression	5			$R^2 = 0.998$ $s = 0.018$ $F = 1181$	Model developed for α, β -unsaturated aldehydes $\text{pEC3} = 0.99\sigma^* + 1.505$	Of limited applicability but relationships developed are robust and mechanistically interpretable
Patlewicz <i>et al.</i> [96]	Local RAI	Regression	9			$R^2 = 0.741$ $s = 0.184$ $F = 4.77$	Michael addition aldehydes $\text{pEC3} = 0.536 + 0.168 \text{Log } P + 0.488 R\sigma^* + 1.313 R'\sigma^*$	Of limited applicability but relationships developed are robust and mechanistically interpretable
Patlewicz <i>et al.</i> [96]	Local RAI	Regression	12			$R^2 = 0.825$ $s = 0.712$ $F = 21.26$		Of limited applicability but relationships developed are robust and mechanistically interpretable

Table 1. (cont.)

References	Type of (Q)SAR model	Method	No. of training set	No. of test set	Outliers	Performance characteristics	Algorithm	Comments
Roberts <i>et al.</i> [94]	Local RAI	Regression	11			$R^2 = 0.944$ $s = 0.11$ $F = 152.1$	Schiff-base formers $pEC3 = 0.245 + 0.278 \text{Log } P + 0.862 R\sigma^*$ 1,2-Diketones $\text{Log (T/C)} = 0.60 RAI + 0.23$	tionships developed are robust and mechanistically interpretable Of limited applicability but relationships developed are robust and mechanistically interpretable
Patlewicz <i>et al.</i> [99]	Local RAI	Regression	13			$R^2 = 0.73$ $R^2_{\text{adj}} = 0.64$ $s = 0.2695$ $F = 8.16$	Michael addition aldehydes $pEC3 = 0.55 + 0.14 \text{Log } P + 0.51 R\sigma^*(\beta) + 1.07 R\sigma^*(\alpha)$	Of limited applicability but relationships developed are robust and mechanistically interpretable
Patlewicz <i>et al.</i> [98]	QMM based on RAI	Regression	16			$R^2 = 0.873$ $s = 0.18$ $F = 44.5$	Aldehydes and 1,2-diketones $pEC3 = 0.16 + 0.32 \text{Log } P + 0.96\sigma^*$	Of limited applicability but relationships developed are robust and mechanistically interpretable
Patlewicz <i>et al.</i> [99]	QMM based on RAI	Regression	14			$R^2 = 0.87$ $R^2_{\text{adj}} = 0.85$ $s = 0.165$ $F = 37.7$	Schiff-base formers $pEC3 = 0.17 + 0.93\sigma^* + 0.30 \text{Log } P$	Of limited applicability but relationships developed are robust and mechanistically interpretable
Roberts <i>et al.</i> [102]	QMM based on RAI	Regression	16			$R^2 = 0.952$ $R^2_{\text{adj}} = 0.945$ $s = 0.12$ $F = 129.6$	Schiff-base formers (11 aliphatic aldehydes, 1 α -ketoester and 4 α,β -diketones) $pEC3 = 1.12(\pm 0.11) \Sigma\sigma^* + 0.42(\pm 0.04) \text{log } P - 0.62(\pm 0.13)$	Of limited applicability but relationships developed are robust and mechanistically interpretable
Aptula <i>et al.</i> [100]	QMM Mech Domains	Structural alerts	41				Evaluation of chemicals on the basis of mechanistic domains to derive structural alert rules	General structural alerts derived for use as part of a ITS
Roberts <i>et al.</i> [104]	QMM Mech Domains	Structural alerts	106				Evaluation of chemicals on the basis of mechanistic domains to refine structural alert rules derived in [96, 99]	General structural alerts derived for use as part of a ITS
Roberts <i>et al.</i> [105]	QMM Mech Domains	Structural alerts	210				Evaluation of chemicals on the basis of mechanistic domains to refine structural alert rules derived in Ref. [96, 99, 100]	General structural alerts derived for use as part of a ITS

and 29 non-sensitizers taken from Ref. [42, 55]) and using a pool of 420 generated descriptors. The final model (Table 1) was based on four structural descriptors and predicted 83% of the responses correctly based on the training set of compounds, and 79% in a cross-validated test set. The relevance of the chosen descriptors to sensitization mechanism was not explained in the article. Roberts *et al.* [56] scrutinized this model in detail and criticized it in terms of, *inter alia*, the tendency of the model to predict non-sensitizers better than sensitizers as well as the lack of a plausible mechanistic rationale.

Using an augmented dataset of chemicals tested in the LLNA taken from the Interagency Coordinating Committee on the Validation of Alternative Method (ICCVAM) dataset [55] of 209 chemicals, Fedorowicz *et al.* [57] then developed logistic regression models using a subset of 178 organic chemicals. The resulting dataset comprised 132 sensitizers and 46 non-sensitizers. Of the same dataset, 105 chemicals had guinea pig data, 82 sensitizers and 23 non-sensitizers. A total of 1777 descriptors, after removal of duplicates, were calculated using Cerius 2 (Accelrys Inc.), Dragon 2.1 (Talete srl, Milan, Italy) and MolconnZ software (eduSoft, LC, Ashland, VA, USA). Stepwise logistic regression was then used to select relevant descriptors and derive models based on the two LLNA and GPMT datasets. Polarity, charge distribution and lipophilicity parameters were associated with the LLNA model whereas polarity and hydrogen bonding were important factors in the guinea pig model. Both models produced correct responses of greater than 80%. Their sensitivity values were greater than 94% but their specificities were very low (see Table 1). This was attributed to the unbalanced datasets, in that there was a 3:1 ratio of sensitizers to non-sensitizers in the LLNA dataset and a 4:1 ratio in the guinea pig dataset. The model performances were additionally compared with DfW (Version 6, LHASA Ltd, Leeds, UK) and TOPKAT (Version 6.2, Accelrys Inc.) and found to perform better (in terms of their correct classification). The performances were 73.3% for TOPKAT, 82.9% for DfW and 87.6% (guinea pig data logistic regression model). The LLNA logistic regression model gave a correct classification of 83.2% compared with 73% for DfW.

Li *et al.* [58] used both random forest and classification tree modelling methods to develop models for skin sensitization. The same dataset taken from Ref. [55] was used. A total of 1380 descriptors (after removal of duplicates and constants) were calculated using Cerius 2 (Accelrys Inc.), Dragon (Talete srl) and MolConnZ (eduSoft) software. The best classification tree contained four descriptors. It possessed a very high sensitivity of 99.2% but a very low specificity of 34%. The best model reported with the random forest approach contained ten descriptors and possessed a specificity of 83% and a sensitivity of 84% at a threshold of 0.7. The random forest approach was additionally exploited to provide a proximity measure defined as the probability of assigning two chemicals to the same

node. A hierarchical clustering was performed to group chemicals according to this index in order to gain insights on how structurally similar chemicals were combined. A number of groups emerged such as halogeno derivatives and simple phenyl derivatives but no detailed interpretation was offered by the authors.

Fedorowicz's models [54, 57, 58] are limited in their scope in that they only predict the absence and presence of sensitization based on a limited and unbalanced dataset. No mechanistic interpretation for the final descriptors in the models is provided and from the large initial pool of descriptors, it is questionable that the choice of final descriptors is meaningful or statistically sound given the high probability of chance correlations [59]. The final model in the series [58] aims solely to compare the performance of a random forest approach against a standard classification approach using statistical performance as a criterion. The models have been characterized to an extent but have questionable robustness (*e.g.* low specificities) suggesting that they could play no meaningful role within a regulatory context but might be useful as crude screening tools for compound discovery scenarios instead.

Estrada *et al.* [60] performed an LDA relating topological descriptors to skin sensitization data as measured in the LLNA. The topological descriptors are so-called TOPS-MODE descriptors or TOPological Substructural MOlecular Design. These descriptors are spectral moments of a bond matrix weighted for different physico-chemical properties (lipophilicity, polarizability, polar surface area, molar refractivity, atomic charges, van der Waals radii) and raised to different power (in this case, the bond matrix was raised to an order of 15). A set of 93 diverse chemicals and their associated LLNA EC3 values were collated. The EC3 values were categorized into five bands of potency [61] which were then amalgamated into two groups for model building. The first model discriminated strong/moderate sensitizers from all other chemicals whereas the second model differentiated weak sensitizers from non-sensitising/extremely weak sensitizers. Model 1 had correct classification of 80%; the overall correct classification of Model 2 was 80.26%. A dataset of 15 compounds unseen by the model and with available skin sensitization data, taken from other available existing public sources, was used to test the predictivity of Model 1. All but one of the strong/moderate sensitizers were classified correctly. As the validation set was not designed rigorously, this good performance might well have been on chance alone. Subsequent evaluation, [29] using another dataset [62], has revealed a specificity to be as low as 10%, with a sensitivity of 85.5% and an overall concordance of 59.5%. No assessment of the representativeness of the external test set relative to the training set (*i.e.* model applicability domain) was conducted at the time. The importance of applicability domain has since been the subject of much debate and recent scientific research [63–66]. The descriptors in these two models do appear quite complex al-

though a mechanistic interpretation was possible for each of the descriptors in turn. Factors such as lipophilicity, polar surface area, van der Waals radii may model partition-related effects in skin sensitization whereas parameters such as charges and polarizability may relate to the reactivity or electrophilicity of the chemicals. Calculating local spectral moments for each of the bonds with respect to skin sensitization highlights which features contribute positively or negatively to it, *i.e.* structural rules for sensitization may be systemically derived. Examples of proposed structural rules were discussed in both Estrada *et al.* [60] and Estrada *et al.* [67]. In Roberts *et al.* [56], the derived alerts were reassessed (*e.g.* alkyl halides, alkyl alkane sulphonates, aromatic nitro compounds, 3-Methyl-1,2,5-thiadiazole-1,1-dioxide (MPT), hydroxylamines, clotrimazole) and some of the proposed insights were thought to be based on erroneous arguments. In light of these issues, we propose that this model would not be appropriate for use in a regulatory context.

Miller *et al.* [68] collected a set of 87 LLNA data and after removal of 20 outliers (based on poor statistical fit), 67 chemicals were analysed: 50 as a training set and 17 as a test set. EC3 values as w/v percent concentrations were used as the dependent variable. Several hundred descriptors were calculated using Comprehensive Descriptors for Structural and Statistical Analysis (CODESSA; SemiChem, Inc., Shawnee Mission, KS, USA) and a heuristic algorithm (not further described in the paper) was used to derive several correlations from these descriptors. The best model was selected based on the statistical parameters (R^2 , adjusted R^2 , F -test, *etc.*) as well as on whether there was a 'chemical sense of the descriptors for the understood mechanism of sensitization'. The performance for the test set of 17 chemicals was expressed in terms of R^2 ($R^2=0.877$ where 14 chemicals out of the 17 chemicals were correctly predicted). On account of biological variability of the EC3 values, 'a bin selection method' was then used to classify compounds into one of three categories; strong, moderate and weak. The cutoffs were chosen to balance the number of chemicals that fell into the strong and moderate categories rather using the potency categories as defined in Kimber *et al.* [12]. The bin selection method was found to predict 44 out of 50 training set chemicals correctly, but if the proposed category boundaries for the LLNA from Ref. [12], been used, than the QSAR model would have been shown to perform much less favourably. Use of the EC3 (as a weight percentage) rather than a molar basis EC3 appears to be flawed. It disguises the fact that the model was poor at discriminating within a given category. Further evaluation of this model has been described in Roberts *et al.* [56]. Owing to the poor performance, the unsound statistics in starting with such a large pool of descriptors [59], the use of the percentage EC3 amongst other things, it is suggested that this model would be of no value in evaluating sensitization robustly.

Ren *et al.* [69] used a Support Vector Machine (SVM) [70] to develop a non-linear binary classification model for skin sensitization for a diverse set of 131 organic compounds tested in the LLNA. Six descriptors were selected by stepwise forward discriminant analysis (LDA) from a starting set of 282 CODESSA molecular descriptors. These six descriptors were thought to reflect the mechanistic relevance to skin sensitization (though this was not explained in the paper) and were used as inputs to the SVM model. The non-linear model developed from SVM (89.77% correct classification) algorithm outperformed the LDA (correct classification 79.55%). This exercise aimed to compare the utility of one statistical technique over a second to judge whether the statistical performance was enhanced. No effort was made to fully characterize this model in the paper. The legitimacy of using such a high number of starting descriptors (282) for the number of observations (131) is also an issue.

The feasibility of using 4D fingerprints derived from a methodology called 4D molecular fingerprint similarity analysis was investigated to develop QSAR models for skin sensitization [71]. An early version of the LLNA dataset [62] comprising data on 219 chemicals was sorted into three categories; non-sensitizers (NS) and weak sensitizers (WS), (101 chemicals); moderate sensitizers (MS; 72 chemicals), and strong (SS) or extreme (ES) sensitizers (46 chemicals). The moderate sensitizers were removed to create a separation gap and the other two classes were retained to develop a model to discriminate between them. Ten NS+WS and five SS+ES sensitizers were randomly chosen to be the test set. The remaining training set comprised 91 NS+WS chemicals and 41 SS+ES sensitizers. Models were initially investigated using discriminant analysis and partial least square discriminant analysis techniques but the performance of these models were as low as 67% accuracy. Two other statistical techniques namely logistic regression and partial least squares coupled with logistic regression were then applied to generate categorical models which did successfully differentiate between NS+WS chemicals and SS+ES sensitizers. However, no prediction of potential moderate sensitizers could be made. The fingerprint descriptors used lack any physical meaning and are extremely difficult to interpret in terms of sensitization mechanism. For these reasons, we propose that this model cannot be reliably used in regulatory skin sensitization risk assessment (*e.g.* under REACH).

Statistical QSARs covering several mechanistic domains have become a growth area recently. Some have apparent high predictive rates, but typically lack any sound mechanistic basis and are usually only able to model a secondary endpoint of skin sensitization presence or absence or else a crude potency classification [29, 56]. The majority of these QSARs use selections from the same set of published LLNA data [62] suggesting an over emphasis on applying new statistical techniques or descriptors rather than trying to rationalize the underlying skin sensitization

mechanism. As a consequence, it is suggested that these models may serve, at best, in a restricted capacity as screening tools within an ITS.

2.3 Expert Systems

2.3.1 Knowledge-Based Systems

DfW 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 [72]. Within DfW, 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 sensitization knowledge base in DfW was initially developed in collaboration with Unilever in 1993 using its historical database of GPMT data for 294 chemicals [40, 73, 74]. The DfW knowledge base is under constant development as new chemical insights and data become available [43, 75, 76]. Version 9.0.0 contains 64 alerts for skin sensitization. Whilst DfW is an extremely valuable tool, the training datasets used to derive the structural alerts are hidden from the end-user. The applicability domain is also insufficiently characterized. This means that the performance from any evaluation exercise using published data (as shown in Ref. [29]) is likely to be inflated. It is worth noting that the system is only able to flag the potential for sensitization, not the absence of sensitization, which limits its practical utility as a stand-alone tool. Since the alert is derived on the basis of a combination of different data sources, it is not possible to relate a given alert to a defined test protocol. DfW is of value in the assessment of skin sensitization but only as part of an ITS framework.

2.3.2 Statistical Expert Systems

TOPKAT comprises two sets of sensitization models that were developed originally by Enslein *et al.* [77]. GPMT data for 315 chemicals was assembled from published collections in Contact Dermatitis and from the dataset published by Cronin and Basketter, [47]. The data were then scaled according to classes defined by Barratt *et al.* [40, 73]. Two sets of models were developed: one for aromatics (excluding chemicals with one-benzene ring) and the other for aliphatics and chemicals with one-benzene ring. The first set of models discriminated between non-sensitizers and sensitizers, the second set of models resolved the potency: weak/moderate *versus* strong. The present version of TOPKAT (Version 6.2) has been supplemented with data for a further 20 chemicals [78]. The training set of TOPKAT is partly accessible and the domain is defined using the patented Optimal Prediction Space (OPS) [79]. A similarity function to identify analogues with associated data is also embedded into TOPKAT which is useful for

verifying the robustness of a given prediction. The lack of a defined unambiguous algorithm and explanation of the descriptors lower the practical usability of the tool. The performance has found to be variable [29] but nonetheless the model may be of some value in providing, through the similarity function, supporting information within an ITS.

CASE methodology and all its variants (*e.g.* MCASE, CASETOX) were developed by Klopman and Rosenkranz [80–82]. There are more than 180 modules covering various areas of toxicology and pharmacology endpoints including skin sensitization currently marketed by MultiCASE Inc. (Cleveland, Ohio, US). The CASE approach uses a probability assessment to determine whether a structural fragment is associated with toxicity. To achieve this, molecules are split into structural fragments up to a certain path length. Probability assessments determine whether fragments significantly promote or inhibit toxicity. To create models, structural fragments are incorporated into a regression analysis. The MCASE modules available for skin sensitization are described further in primary articles [82–85]. The (Q)SAR estimates for the MCASE skin sensitization model are also included in the Danish Environmental Protection Agency (EPA)'s (Q)SAR Database which is hosted on the website of the Joint Research Centre's European Chemicals Bureau (ECB) <http://ecb.jrc.it/QSAR/>. Like the other expert systems already described, the training set is hidden and the algorithm is not fully defined suggesting this tool might at best offer some insights for a weight of evidence skin sensitization assessment.

2.3.3 Hybrid Expert Systems

The Times MEtabolism Simulator platform, marketed by LMC (University 'As Zlatarov', Bourgas, Bulgaria), used for predicting Skin Sensitization (TIMES-SS) is a hybrid expert system which aims to encode 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 catalysed 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. Some of these reactions are additionally underpinned by mechanistically based 3D-QSARs [66, 86]. Recently an external validation activity was undertaken whereby data were generated for 40 new chemicals in the 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 [30]. This model has been characterized with respect to the OECD principles [30].

The tool shows promise but more efforts are required to refine some of the rules, in particular to improve the predictions of weak sensitizers. In addition, since the training set is a mix of three different datasets; the translation required to relate the unified modelled endpoint to one used in regulatory practice is not altogether straightforward. For this reason, TIMES-SS is thought to be best applied in the context of an ITS.

Overall, there are reasonable prospects for the use of expert systems such as DfW, TIMES-SS though only as part of a weight of evidence assessment of skin sensitization (see Table 1).

2.4 Local QSAR Models

The majority of the local models available have been derived by the Relative Alkylation Index (RAI) approach, which was first developed by Roberts and Williams, [87] in 1982. This mathematical model quantified the degree of carrier alkylation with sensitization potential and is expressed in most general form as:

$$\text{RAI} = \log D + a \log k + b \log P \quad (1)$$

where D is the dose of sensitizer, $\log k$ models the reactivity and is the rate constant or relative rate constant k for the reaction of the sensitizer with a model nucleophile and $\log P$, is the octanol–water partition coefficient which models the hydrophobicity. For specific use with LLNA data, an RAI–potency relationship can be described by:

$$\text{pEC3} = a \log k + b \log P + c \quad (2)$$

where p denotes $-\log$, EC3, for this purpose is the molar percentage of chemical sensitizer required to elicit a three-fold increase in lymph node cell proliferation.

The RAI model has been used to evaluate a wide range of different datasets of skin sensitizing chemicals. It has in particular, served well for the development of QSARs for small sets of structurally related chemicals, to substantiate formulated hypothesis. Examples of RAI models include those derived for primary alkyl bromides [88], sultones [87, 89], acrylates [90], sulphonate esters [91], urushiol analogues [92], halo- and pseudohalobenzenes [93], aldehydes and diketones [94–99] (see Table 1 for a selection of illustrative examples).

2.4.1 Quantitative Mechanistic Models (QMM)

The RAI model had been thought up until recently to be only applicable for closely related datasets. This view started to change following work on aldehydes and ketones [94–99] but work in Ref. [100] found that it was possible to classify a selection of diverse compounds [101] into reaction mechanistic applicability domains, *i.e.* grouping according to the mechanisms whereby the compounds could

react with nucleophiles. For the 41 chemicals evaluated, a domain containing a diversity of structures could be identified which was related by common reaction chemistry and in which ranking of compounds in order of reactivity was feasible. A number of reaction mechanistic applicability domains were identified as a result: Michael-type addition domain; bimolecular nucleophilic substitution (S_N2) domain; aromatic nucleophilic substitution (S_NAr) domain; acylation domain; Schiff-base domain. These started to provide the foundation to underpin skin sensitization with sound reaction chemistry principles, which were later used to develop a new QMM model for Schiff-base electrophiles [102]. The principles were hence generalized in [103] and further evaluated using larger datasets in Ref. [104] and [105].

The principles provided the framework for how information from RAI mechanistic approaches could be best exploited in the evaluation of skin sensitization potential and potency. First, the mechanistic domain would need to be assigned based on the principles outlined in Ref. [103] or by using experimental chemistry if necessary. If a QMM existed for this domain, this could be used to estimate the skin sensitization potency, the EC3 value of a chemical. The output would be the same as the regulatory test result thus requiring no transformation. Alternatively, if no QMM existed, a mechanistic read-across could be potentially undertaken, *i.e.* relating mechanistically similar chemicals with available sensitization data (taken for example from Ref. [62]) in order to estimate an EC3 value for the chemical of interest by interpolation. This would still rely on the same inputs as for a QMM, namely reactivity and hydrophobicity information. Hydrophobicity is conveniently modelled using $\log P$. The challenge is in deriving the relevant reactivity information. Approaches to encode reactivity include heats of reaction [106], an activation index [107], substituent coefficients (such as Taft σ^* coefficients) [108] or experimental measures of reactivity (*e.g.* $\log$ of rate constants, k) of chemicals with model nucleophiles [109–114]. The $\log P$ and $\log k$ information for both the chemical of interest and the chemicals of known sensitization potency could then be used to derive an estimate of sensitizing potency for the chemical of interest [113]. Ideally, the generation of a large database of experimental measures of reactivity could also help to eventually develop a convenient and robust *in silico* reactivity parameter. This is the subject of on-going work but more resources are required.

The QMM approach of integrating mechanistic principles for skin sensitization shows great promise for the development of robust models to predict skin sensitization potential and potency in both a standalone mode under REACH or as part of an ITS. A comprehensive reactivity database would greatly facilitate implementation of this strategy.

3 Conclusions

There is a very strong mechanistic understanding of skin sensitization, as summarized earlier [3, 25, 26]. Most published skin sensitization QSARs to date have fallen into one of two main categories: either they are RAI QSARs, typically of high statistical quality but applicable to a narrow range of closely related structures or they are statistical QSARs which aim to be global in their applicability, are variable in their successes, insufficiently characterized and lacking in real mechanistic insight [29, 56]. Whilst these have become a growth area recently, all have been based on selections from the same limited set of published LLNA data suggesting too much focus on applying new statistical techniques or descriptors rather than understanding and rationalising the likely sensitization potential of new compounds. Their lack of characterization makes them less useful for regulatory applications such as in REACH which calls for the use of estimates from valid (Q)SARs.

Expert systems available such as DfW, TOPKAT are more promising alternatives. Their lack of full transparency and in some cases lack of established applicability domain means that their use whilst valuable is probably best limited to supporting information in a weight of evidence assessment as part of an ITS [29].

The covalent hypothesis continues to be the most promising way of developing mechanistically based robust QSARs. The reaction chemistry concepts recently outlined in Aptula and Roberts [103] have changed the perspective regarding RAI QSARs. These QSARs are actually more widely applicable than originally thought [102] where robust mechanistic-based QSAR models (QMMs) can be developed within reaction chemistry domains. Such models were available (an example is Ref. [102]) are well characterized with respect to the OECD principles and are sufficiently robust for standalone purposes under REACH. In addition, the strategy described above demonstrates how RAI mechanistic approaches can currently be used in the evaluation of skin sensitization potency. Routine implementation of this strategy would be greatly facilitated if there existed an adequate parameter to model reactivity *in silico*. Experimental work to generate the necessary reactivity data to realize this is on-going as evidenced by Aptula *et al.* [109] and by others [110–114], but more resources are required [113].

An additional consideration for all types of models, is to establish relevance, in this case REACH relevance. At the current time, this might best be achieved using an inventory such as the European INventory of Existing Commercial chemical Substances (EINECS). This would necessitate the characterization of existing substances (such as EINECS) by a domain [115] and hence an assessment to be performed to determine the extent in which the existing (Q)SARs were representative of that domain. This would also help to identify information gaps where further test-

ing might need to be carried out to ensure that a portfolio of models could span the entire domain. This is an immediate research need.

Disclaimer

The views in this paper belong to the authors, and do not necessarily represent formal views or positions of the European Commission or its Joint Research Centre. Any reference to a software product is for illustrative purpose only and does not imply its endorsement by the European Commission.

References

- [1] W. Steiling, D. A. Basketter, K. Berthold, M. Bultler, J. L. Garrigue, I. Kimber, L. Lea, C. Newsome, R. Roggeband, G. Stropp, S. Waterman, C. Wiemann, *Food Chem. Toxicol.* **2001**, *39*, 293–301.
- [2] I. Kimber, R. J. Dearman, D. A. Basketter, C. A. Ryan, G. F. Gerberick, *Contact Derm.* **2002**, *47*, 315–328.
- [3] I. Kimber, R. J. Dearman, *Ann. Allergy, Asthma Immunol.* **2003**, *90*, 28–31.
- [4] Commission of the European Communities. 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. *Off. J. Eur. Union* L 396/1 of 30.12.2006.
- [5] OECD, Organisation for Economic Cooperation and Development, Guidelines for Testing of Chemicals No. 406, Skin sensitisation, **1992**.
- [6] OECD, Organisation for Economic Cooperation and Development, 2002, Guidelines for Testing of Chemicals No. 429, Skin sensitisation: The Local Lymph Node Assay, **2002**.
- [7] D. A. Basketter, G. F. Gerberick, I. Kimber, S. E. Loveless, *Food Chem. Toxicol.* **1996**, *34*, 985–997.
- [8] D. A. Basketter, L. J. Lea, K. Cooper, J. Stocks, A. Dickens, I. Pate, R. J. Dearman, I. Kimber, *Food Chem. Toxicol.* **1999**, *37*, 1167–1174.
- [9] D. A. Basketter, L. Blaikie, R. J. Dearman, I. Kimber, C. A. Ryan, G. F. Gerberick, P. Harvey, P. Evans, I. R. White, R. J. Rycroft, *Contact Derm.* **2000**, *42*, 344–348.
- [10] D. A. Basketter, P. Evans, R. J. Fielder, G. F. Gerberick, R. J. Dearman, I. Kimber, *Food Chem. Toxicol.* **2002**, *40*, 593–598.
- [11] G. F. Gerberick, C. A. Ryan, I. Kimber, R. J. Dearman, L. J. Lea, D. A. Basketter, *Am. J. Contact Derm.* **2000**, *11*, 3–18.
- [12] I. Kimber, D. A. Basketter, M. Butler, A. Gamer, J. L. Garrigue, G. F. Gerberick, C. Newsome, W. Steiling, H. W. Vohr, *Food Chem. Toxicol.* **2003**, *41*, 1799–1809.
- [13] D. A. Basketter, C. K. Smith Pease, G. Y. Patlewicz, *Clin. Exp. Dermatol.* **2003**, *28*, 218–221.
- [14] D. Basketter, S. Casati, G. F. Gerberick, P. Griem, B. Phillips, A. Worth, *Altern. Lab. Anim.* **2005**, *33*, 83–103.

- [15] I. R. Jowsey, D. A. Basketter, C. Westmoreland, I. Kimber, *J. Appl. Toxicol.* **2006**, 26, 341–350.
- [16] J. S. Jaworska, M. Comber, C. Auer, C. J. van Leeuwen, *Environ. Health Perspect.* **2003**, 22, 1358–1360.
- [17] OECD, 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. http://www.oecd.org/document/23/0,2340,en_2649_34365_33957015_1_1_1_1,00.html.
- [18] OECD, 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>.
- [19] A. P. Worth, A. Bassan, J. de Bruijn, A. Gallegos Saliner, T. I. Netzeva, G. Patlewicz, M. Pavan, I. Tsakovska, S. Eisenreich, *SAR QSAR Environ. Res.* **2006**, 18, 111–125.
- [20] ECB, *Final Report of REACH Implementation Project (RIP) 3.3-2. Technical Guidance Document to Industry on the Information Requirements for REACH*. European Commission-Joint Research Centre, Ispra, Italy **2007**, accessible from: <http://ecb.jrc.it/reach/>.
- [21] M. H. Karol, C. Graham, M. D. Barratt, *Toxicology* **1999**, 7, 59–77.
- [22] S. J. Bashir, H. I. Maibach, in: A. F. Kydonieus, J. J. Wille (Eds.), *Biochemical Modulation of Skin Reactions*, CRC Press LLC, Boca Raton **2000**, pp. 61–64.
- [23] R. Rodford, G. Patlewicz, J. D. Walker, M. P. Payne, *Environ. Toxicol. Chem.* **2003**, 22, 1855–1861.
- [24] D. Basketter, A. Dooms-Goossens, A. T. Karlberg, J.-P. Lepoittevin, *Contact Derm.* **1995**, 32, 65–73.
- [25] C. K. Smith Pease, *Toxicology* **2003**, 192, 1–22.
- [26] D. W. Roberts, A. O. Aptula, *J. Appl. Toxicol.* **2007**, in press. Aug 17 (Epub ahead of print).
- [27] K. Landsteiner, J. L. Jacobs, *J. Exp. Med.* **1936**, 64, 625–639.
- [28] G. Dupuis, C. Benezra, *Allergic Contact Dermatitis to Simple Chemicals: A Molecular Approach*, Dekker, New York and Basel **1982**.
- [29] G. Patlewicz, A. O. Aptula, E. Uriate, D. W. Roberts, P. S. Kern, G. F. Gerberick, C. A. Ryan, I. Kimber, R. Dearman, D. A. Basketter, *SAR QSAR Environ. Res.* **2007**, 18, 515–541.
- [30] G. Patlewicz, S. D. Dimitrov, L. K. Low, P. S. Kern, G. D. Dimitrova, M. I. H. Comber, A. O. Aptula, R. D. Phillips, J. Niemela, C. Madsen, E. B. Wedebye, D. W. Roberts, P. T. Bailey, O. G. Mekenyan, *Reg. Toxicol. Pharmacol.* **2007**, 48, 225–239.
- [31] K. S. Rao, J. E. Betso, K. J. Olson, *Drug Chem. Toxicol.* **1981**, 4, 331–351.
- [32] C. Benezra, C. C. Sigman, L. R. Perry, C. Tucker Helmes, H. I. Maibach, *J. Invest. Dermatol.* **1985**, 85, 351–356.
- [33] B. M. Hausen, W. Beyer, *Contact Derm.* **1992**, 26, 253–258.
- [34] B. M. Hausen; M. Berger, *Contact Derm.* **1989**, 21, 141–147.
- [35] B. M. Hausen, H. W. Schmalle, *Contact Derm.* **1985**, 13, 329–332.
- [36] D. Cremer, B. M. Hausen, H. W. Schmalle, *J. Med. Chem.* **1987**, 30, 1678–1681.
- [37] K. H. Schulz, I. Garbe, B. M. Hausen, M. H. Simatupang, *Arch. Dermatol. Res.* **1977**, 258, 41–52.
- [38] B. M. Hausen, H. Hitsch, B. Borrmann, D. Koch, R. Rathmann, B. Richter, W. A. Konig, *Contact Derm.* **1995**, 33, 12–16.
- [39] B. M. Hausen, B. Elsasser, K. Krohn, U. Looock, *Am. J. Contact Derm.* **2003**, 14, 82–89.
- [40] M. D. Barratt, D. A. Basketter, M. Chamberlain, G. D. Admans, J. J. Langowski, *Toxicol. In Vitro* **1994**, 8, 1053–1060.
- [41] M. P. Payne, P. T. Walsh, *J. Chem. Inf. Comp. Sci.* **1994**, 34, 154–161.
- [42] J. Ashby, D. A. Basketter, D. Paton, I. Kimber, *Toxicology* **1995**, 103, 177–194.
- [43] I. Gerner, M. D. Barratt, S. Zinke, K. Schlegel, E. Schlede, *Altern. Lab. Anim.* **2004**, 32, 487–509.
- [44] P. S. Magee, J. J. Hostynek, H. I. Maibach, *Quant. Struct. Act. Relat.* **1994**, 13, 22–33.
- [45] J. J. Hostynek, P. S. Magee, *Toxicol. In Vitro* **1997**, 11, 377–384.
- [46] B. Magnusson, A. M. Kligman, *J. Invest. Dermatol.* **1969**, 52, 268–276.
- [47] M. T. Cronin, D. A. Basketter, *SAR QSAR Environ. Res.* **1994**, 2, 159–179.
- [48] J. C. Dearden, S. J. A. Bradburne, M. T. D. Cronin, P. Solanki, The Physical Significance of Molecular Connectivity, in: J. E. Turner, M. W. England, T. W. Schultz, N. J. Kwaak (Eds.), *QSAR 88. Proceedings of the Third International Workshop on Quantitative Structure-Activity Relationships in Environmental Toxicology*, USDOE, Oak Ridge, TN **1988**, pp. 43–50.
- [49] Z. Zhou, R. G. Parr, *J. Am. Chem. Soc.* **1990**, 112, 5720–5724.
- [50] J. Devillers, *Toxicol. Methods* **2000**, 10, 181–193.
- [51] J. Devillers, in: J. Devillers (Ed.), *Neural Networks in QSAR and Drug Design*, Academic press, London **1996**, pp. 1–46.
- [52] D. E. Rumelhart, G. E. Hinton, R. J. Williams, *Nature* **1986**, 323, 533–536.
- [53] M. Vracko, V. Bandelj, P. Barbieri, E. Benfenati, Q. Chaudhry, M. Cronin, J. Devillers, A. Gallegos, G. Gini, P. Gramatica, C. Helma, P. Mazzatorta, D. Neagu, T. Netzeva, M. Pavan, G. Patlewicz, M. Randic, I. Tsakovska, A. Worth, *SAR QSAR Environ. Res.* **2006**, 17, 265–284.
- [54] A. Fedorowicz, L. Zheng, H. Singh, E. Demchuk, *Int. J. Mol. Sci.* **2004**, 5, 56–66.
- [55] K. E. Haneke, R. R. Tice, B. L. Carson, B. H. Margolin, W. S. Stokes, *Reg. Toxicol. Pharmacol.* **2001**, 34, 274–286.
- [56] D. W. Roberts, A. O. Aptula, M. T. D. Cronin, E. Hulzebos, G. Patlewicz, *SAR QSAR Environ. Res.* **2007**, 18, 343–365.
- [57] A. Fedorowicz, H. Singh, S. Soderholm, E. Demchuk, *Chem. Res. Toxicol.* **2005**, 18, 954–969.
- [58] S. Li, A. Fedorowicz, H. Singh, S. C. Soderholm, *J. Chem. Inf. Model.* **2005**, 45, 952–964.
- [59] J. G. Topliss, R. P. Edwards, *J. Med. Chem.* **1979**, 22, 1238–1244.
- [60] E. Estrada, G. Patlewicz, M. Chamberlain, D. Basketter, S. Larbey, *Chem. Res. Toxicol.* **2003**, 16, 1226–1235.
- [61] D. A. Basketter, European Research Group on Experimental Contact Dermatitis (ERGECD), Strasbourg **2001**.
- [62] G. F. Gerberick, C. A. Ryan, P. S. Kern, H. Schlatter, R. J. Dearman, I. Kimber, G. Y. Patlewicz, D. A. Basketter, *Dermatitis* **2005**, 16, 157–202.
- [63] T. I. Netzeva, A. P. Worth, T. Aldenberg, R. Benigni, M. T. D. Cronin, P. Gramatica, J. S. Jaworska, S. Kahn, G. Klopman, C. A. Marchant, G. Myatt, N. Nikolova-Jeliazko-

- va, G. Y. Patlewicz, R. Perkins, D. W. Roberts, T. W. Schultz, D. T. Stanton, J. J. M. van de Sandt, W. Tong, G. Veith, C. Yang, *Altern. Lab. Anim.* **2005**, *33*, 155–173.
- [64] J. Jaworska, N. Nikolova-Jeliazkova, T. Aldenberg, *Altern. Lab. Anim.* **2005**, *33*, 445–459.
- [65] N. Nikolova-Jeliazkova, J. Jaworska, *Altern. Lab. Anim.* **2005**, *33*, 461–470.
- [66] S. Dimitrov, G. Dimitrova, T. Pavlov, N. Dimitrova, G. Patlewicz, J. Niemela, O. Mekenyan, *J. Chem. Inf. Model.* **2005**, *45*, 839–849.
- [67] E. Estrada, G. Patlewicz, Y. Gutierrez, *J. Chem. Inf. Comp. Sci.* **2004**, *44*, 688–698.
- [68] M. D. Miller, D. M. Yourtee, A. G. Glaros, C. C. Chappelow, J. D. Eick, A. J. Holder, *J. Chem. Inf. Model.* **2005**, *45*, 924–929.
- [69] Y. Ren, H. Liu, C. Xue, X. Yao, M. Liu, B. Fan, *Anal. Chim. Acta* **2006**, *572*, 272–282.
- [70] Q. Benables, M. Smith, The R Development Core Team, R Manuals, **2003**.
- [71] Y. Li, Y. J. Tseng, D. Pan, J. Liu, P. S. Kern, G. F. Gerberick, A. J. Hopfinger, *Chem. Res. Toxicol.* **2007**, *20*, 114–128.
- [72] D. M. Sanderson, C. G. Earnshaw, *Human Exp. Toxicol.* **1991**, *10*, 261–273.
- [73] M. D. Barratt, D. A. Basketter, M. Chamberlain, G. D. Adams, J. J. Langowski, *Toxicol. In Vitro* **1994**, *8*, 837–839.
- [74] M. D. Barratt, J. J. Langowski, *J. Chem. Inf. Comp. Sci.* **1999**, *39*, 294–298.
- [75] S. Zinke, I. Gerner, E. Schlede, *Altern. Lab. Anim.* **2002**, *30*, 285–298.
- [76] K. Langton, G. Y. Patlewicz, A. Long, C. A. Marchant, D. A. Basketter, *Contact Derm.* **2006**, *55*, 342–347.
- [77] K. Enslein, V. K. Gombar, B. W. Blake, H. I. Maibach, J. J. Hostynek, C. C. Sigman, D. Bagheri, *Food Chem. Toxicol.* **1997**, *35*, 1091–1098.
- [78] Accelrys Inc., TopKat User Guide Version 6.02, San Diego: Accelrys Inc., **2004**.
- [79] Optimum Prediction Space, US Patent 6,036,349 issued 14/3/2000.
- [80] G. Klopman, *J. Am. Chem. Soc.* **1984**, *106*, 7315–7321.
- [81] G. Klopman, *Quant. Struct. Act. Relat.* **1992**, *11*, 176–184.
- [82] H. S. Rosenkranz, G. Klopman, in: M. E. Wolff (Ed.), *Burger's Medicinal Chemistry and Drug Discovery*, Fifth Edn., Volume 1, John Wiley & Sons Inc., New York **1995**, pp. 223–249.
- [83] R. Gealy, C. Graham, N. B. Sussman, O. T. Macina, H. S. Rosenkranz, M. H. Karol, *Human Exp. Toxicol.* **1996**, *15*, 489–493.
- [84] C. Graham, R. Gealy, O. T. Macina, M. H. Karl, H. S. Rosenkranz, *QSAR Comb. Sci.* **1996**, *15*, 224–229.
- [85] R. Johnson, O. T. Macina, C. Graham, H. S. Rosenkranz, G. R. Cass, M. H. Karol, *Environ. Health Perspect.* **1997**, *105*, 986–992.
- [86] S. D. Dimitrov, L. K. Low, G. Y. Patlewicz, P. S. Kern, G. D. Dimitrova, M. H. I. Comber, R. D. Phillips, J. Niemela, P. T. Bailey, O. G. Mekenyan, *Int. J. Toxicol.* **2005**, *24*, 189–204.
- [87] D. W. Roberts, D. L. Williams, *J. Theor. Biol.* **1982**, *99*, 807–825.
- [88] D. A. Basketter, D. W. Roberts, M. Cronin, E. W. Scholes, *Contact Derm.* **1992**, *27*, 137–142.
- [89] D. W. Roberts, *Chem. Res. Toxicol.* **1995**, *8*, 545–551.
- [90] D. W. Roberts, *Contact Derm.* **1987**, *17*, 281–289.
- [91] D. W. Roberts, D. A. Basketter, *Contact Derm.* **2000**, *42*, 154–161.
- [92] D. W. Roberts, C. Benezra, *Contact Derm.* **1993**, *29*, 78–83.
- [93] D. W. Roberts, D. L. Williams, D. Bethell, *Chem. Res. Toxicol.* **2007**, *20*, 61–71.
- [94] D. W. Roberts, M. York, D. A. Basketter, *Contact Derm.* **1999**, *41*, 14–17.
- [95] G. Patlewicz, D. A. Basketter, C. K. Smith, S. A. Hotchkiss, D. W. Roberts, *Contact Derm.* **2001**, *44*, 331–336.
- [96] G. Y. Patlewicz, Z. M. Wright, D. A. Basketter, C. K. Pease, J.-P. Lepoittevin, E. G. Arnau, *Contact Derm.* **2002**, *47*, 219–226.
- [97] D. W. Roberts, G. Patlewicz, *SAR QSAR Environ. Res.* **2002**, *13*, 145–152.
- [98] G. Patlewicz, D. W. Roberts, J. D. Walker, *QSAR Comb. Sci.* **2003**, *22*, 196–203.
- [99] G. Y. Patlewicz, D. A. Basketter, C. K. Pease, K. Wilson, Z. M. Wright, D. W. Roberts, G. Bernard, E. G. Arnau, J.-P. Lepoittevin, *Contact Derm.* **2004**, *50*, 91–97.
- [100] A. O. Aptula, G. Patlewicz, D. W. Roberts, *Chem. Res. Toxicol.* **2005**, *18*, 1420–1426.
- [101] G. F. Gerberick, C. A. Ryan, P. S. Kern, R. J. Dearman, I. Kimber, G. Y. Patlewicz, D. A. Basketter, *Contact Derm.* **2004**, *50*, 274–288.
- [102] D. W. Roberts, A. O. Aptula, G. Patlewicz, *Chem. Res. Toxicol.* **2006**, *19*, 1228–1233.
- [103] A. O. Aptula, D. W. Roberts, *Chem. Res. Toxicol.* **2006**, *19*, 1097–1105.
- [104] D. W. Roberts, A. O. Aptula, G. Patlewicz, *Chem. Res. Toxicol.* **2007**, *20*, 44–60.
- [105] D. W. Roberts, G. Patlewicz, P. S. Kern, F. Gerberick, I. Kimber, R. J. Dearman, C. A. Ryan, D. A. Basketter, A. O. Aptula, *Chem. Res. Toxicol.* **2007**, *20*, 1019–1030.
- [106] P. S. Magee, *Quant. Struct. Act. Relat.* **2000**, *19*, 356–365.
- [107] A. O. Aptula, D. W. Roberts, M. T. Cronin, *Chem. Res. Toxicol.* **2005**, *18*, 324–329.
- [108] D. D. Perrin, B. Dempsey, E. P. Serjeant, *pKa Prediction for Organic Acids and Bases*, Chapman and Hall, London **1981**, pp. 109–126.
- [109] A. O. Aptula, G. Patlewicz, D. W. Roberts, T. W. Schultz, *Toxicol. In Vitro* **2006**, *20*, 239–247.
- [110] G. F. Gerberick, J. D. Vassallo, R. E. Bailey, J. G. Chaney, S. W. Morrall, J.-P. Lepoittevin, *Toxicol. Sci.* **2004**, *81*, 332–343.
- [111] G. F. Gerberick, J. D. Vassallo, L. M. Foertsch, B. B. Price, J. G. Chaney, J. P. Lepoittevin, *Toxicol. Sci.* **2007**, *97*, 417–427.
- [112] T. W. Schultz, R. E. Carlson, M. T. D. Cronin, J. L. M. Hermens, R. Johnson, P. J. O'Brien, D. W. Roberts, A. Siraki, K. B. Wallace, G. D. Veith, *SAR QSAR Environ. Res.* **2006**, *17*, 413–428.
- [113] D. W. Roberts, A. O. Aptula, G. Patlewicz, C. Pease, *J. Appl. Toxicol.* **2007**, in press. Aug 17 (Epub ahead of print).
- [114] A. Natsch, H. Gfeller, M. Rothaupt, G. Ellis, *Toxicol. In Vitro* **2007**, in press. Apr 14 (Epub ahead of print).
- [115] G. Patlewicz, An Investigation into the Feasibility of Developing Categories Using the Topdown Approach Using Automated Workflows, in: A. Worth, A. Bassan, E. Fabjan, A. Gallegos Saliner, T. Netzeva, G. Patlewicz, M. Pavan, I. Tsakovska, The Use of Computational Methods in the Grouping and Assessment of Chemicals – Preliminary Investigations, European Commission report. European Commission Joint Research Centre, Ispra, Italy. Available from ECB website: <http://ecb.jrc.it/QSAR>.