EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Skin sensitisation (Q)SARs/Expert systems: From past, present to future

Grace Patlewicz, Aynur O. Aptula, David W. Roberts, Eugenio Uriarte

IHCP

2007 EUR 22872 EN

EUR 22872 EN

The mission of IHCP is to provide scientific support to the development and implementation of EU policies related to health and consumer protection.
The IHCP carries out research to improve the understanding of potential health risks posed by chemicals, physical and biological agents from various sources to which consumers are exposed.

- As a Research Institute, the IHCP contributes to the improvement of scientific knowledge on health care methods and consumer issues.
- As a European Institution, the IHCP participates, at an international level and in collaboration with Member States Authorities, in R&D and regulatory actions intended to improve consumer tutelage.
- As a European Commission Service, the IHCP acts as an independent advisor for the implementation of risk assessment, monitoring and validation of procedures aimed to insure EU citizens of the use of health and consumer services or products.

European Commission
Directorate-General Joint Research Centre
Institute IHCP

Contact information Grace Patlewicz
Address: European Chemicals Bureau, TP582
E-mail: grace.patlewicz@ec.europa.eu
Tel.: +39 0332 789616
Fax: +39 0332 786717

http://ecb.jrc.it/QSAR
http://www.jrc.cec.eu.int

EUR 22872 EN

ISSN 1018-5593

Luxembourg: Office for Official Publications of the European Communities

Printed in Italy

Quality control insert

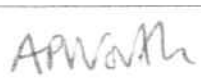
	Name	*Signature*	*Date*
Report Prepared by:	**Grace Patlewicz**		**31-05-07**
Reviewed by: (Scientific level)	**Andrew Worth**		**19-06-07**
Approved by: (Head of Unit)	**Steven Eisenreich**		20.06.07
Final approval: (IHCP Director)	**Elke Anklam**		20/6/07

Table of Contents

Skin sensitisation (Q)SARs/Expert systems: From past, present to future

Grace Patlewicz[1], Aynur O. Aptula[2], David W. Roberts[3], Eugenio Uriarte[4]

[1]European Chemicals Bureau (ECB), TP 582, Institute for Health and Consumer Protection, Joint Research Centre, European Commission, 21020 Ispra (VA), Italy

[2]Safety Environmental Assurance Centre (SEAC), Unilever Colworth, Colworth House, Sharnbrook, Bedford, MK44 1LQ, UK

[3]Liverpool John Moores University, Liverpool, UK

[4]Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela, 15872 Santiago de Compostela, Spain.

Abstract

This review describes the state of (Q)SARs/expert systems for skin sensitisation from the early work on structural alerts to the evolvement of QSAR models through to the development of a mechanistically based QSARs. The review considers the merits of the different approaches with particular focus on their applicability for potential regulatory use. There is a strong mechanistic understanding with respect to skin sensitisation which has facilitated the development of a wide range of models. The majority of the existing models fall into one of two main categories – either they are local in nature, typically specific to a chemical class or else they are global in form, derived empirically using statistical methods. Several of the published global QSARs have been recently characterised and evaluated in accordance with the OECD principles. There are also several expert systems capable of predicting skin sensitisation, these are briefly described. Recently, a new perspective regarding the development of mechanistic skin sensitisation QSARs so-called Quantitative Mechanistic Modelling (QMM) was proposed, where reactivity and hydrophobicity, are used as the key parameters in mathematically modelling skin sensitisation. This approach appears to be the most promising means of deriving robust and mechanistically interpretable models that are of value in a regulatory context. Hydrophobicity can be conveniently modelled using log P, the octanol-water partition coefficient. This is readily determined from chemical structure. No such surrogate for reactivity exists, rate constants are best derived from experimental studies. Initiatives are in progress to generate reactivity data for reactions relevant to skin sensitisation but more resources are required to realise a comprehensive set of reactivity data. This is a fundamental and necessary requirement for the future assessment of skin sensitisation. A framework of how information from using RAI approaches can be used in the evaluation of skin sensitisation is described.

Introduction

Under the European Union (EU) Registration, Evaluation and Authorisation of CHemicals (REACH) programme, all chemicals produced or imported > 1 tonnes per annum (tpa) in the EU 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. REACH calls for an increased use of alternatives such as *in vitro* methods, (quantitative) structure-activity relationships ((Q)SARs) and chemical grouping approaches to help in the assessment of chemicals (see Annex XI of (EC, 2006)).

Allergic contact dermatitis is a significant environmental and occupational health concern. Workers and consumers being sensitised are a major problem for individuals, for employers and for marketing certain products. Consequently there exists an important need to identify chemicals that have the potential to cause skin sensitisation accurately (Steiling et al., 2001; Kimber et al., 2002; Kimber and Dearman, 2003). In REACH, sensitising potential needs to be assessed for chemicals above the 1 tonne threshold according to Annex VII (EC, 2006). No *in vitro* replacement is currently available, nor expected to be ready in the near future (Basketter et al., 2005; Jowsey et al., 2006).

(Quantitative) structure-activity relationships ((Q)SARs) do show promise but regulators have scarcely used human health (Q)SARs in decision making under existing European Union (EU) legislation. Since the publication of the REACH White paper in 2001, several activities were initiated to increase acceptance of (Q)SARs. The first of these included a Workshop organised by European Chemical Industry Council (CEFIC)/The International Council of Chemical Associations (ICCA) in Setubal, Portugal in 2002 (Jaworska et al., 2003) which identified a number of principles for evaluating the validity of (Q)SARs. These were then evaluated by the Organisation for Economic Co-operation and Development (OECD) (as part of the *Ad hoc* group for (Q)SARs) and are now referred to as the 'OECD principles' which read as follows:

"To facilitate the consideration of a (Q)SAR model for regulatory purposes, it should be associated with the following information:

- a defined endpoint
- an unambiguous algorithm
- a defined domain of applicability
- appropriate measures of goodness-of-fit, robustness and predictivity
- a mechanistic interpretation, if possible

The principles have been summarised briefly in an OECD publication (OECD, 2004). A preliminary guidance document on the characterisation of (Q)SARs was then written (Worth et al., 2005), the basis of which was used in the development of the recently finalised OECD Guidance document on (Q)SAR Validation (OECD, 2007).

The guidance developed to date, describes how (Q)SAR models may be evaluated in accordance with the OECD principles and to an extent what best practice should be when developing new (Q)SAR models. A major challenge remains to identify available (Q)SARs for the wide range of endpoints that will need to be assessed under REACH and to highlight which show greatest promise for hazard and risk assessment purposes. Some efforts to address this need were undertaken as part of the REACH Interim Strategy through a number of REACH Implementation Projects (RIPs) (http://ecb.jrc.it/REACH) (in particular RIP 3.3). In addition, there have been a number of publications where existing (Q)SAR models have been characterised and evaluated in accordance with the OECD principles – examples include those for skin sensitisation (Aptula et al., 2006; Roberts et al., 2007a), acute aquatic toxicity (Vracko et al., 2006; Pavan et al., 2006), estrogenicity (Saliner et al., 2006; Liu et al., 2006), percutaneous penetration (Bouwman et al., 2006) as well as mutagenicity and carcinogenicity (OECD, 2004).

This review paper aims to describe what (Q)SARs including expert systems are currently available for skin sensitisation and to illustrate which of these might be of potential value in a regulatory setting. An overview of skin sensitisation, its mechanism and the current testing methods are provided as background.

Biological Mechanisms of Skin Sensitisation

Skin sensitisation (also called delayed contact hypersensitivity, contact hypersensitivity, contact allergy or allergic contact dermatitis) 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 sensitisation involves two phases, an induction and elicitation phase.

During induction, the sensitising chemical penetrates the *stratum corneum* to the viable epidermis and binds to skin proteins/peptides to create an immunogenic complex. This complex is then recognised and processed by Langerhans Cells (LCs) in the epidermis. Upon exposure to the immunogenic complex, the LCs begin a maturation process in which the LCs internalise and process the immunogenic complex to a form that will be recognised by T-cells. These cells then migrate from the epidermis to the lymph nodes

where they present the modified immunogenic complex to naïve T-cells with receptors that are able to specifically recognise the immunogenic moiety and are stimulated to proliferate and circulate throughout the body.

Upon subsequent exposure to the same sensitiser, protein binding and processing of the immunogenic complex by the LCs occurs after which the immunogenic complex is recognised by circulating T-cells triggering a cascade of biochemical and cellular processes which produce the clinical sensitisation response i.e. elicitation. (Basketter et al., 1995; Kimber and Dearman, 2003; Smith Pease, 2003).

Chemical Mechanisms of Skin Sensitisation

Landsteiner and Jacobs (1936) first explored the issue of why some chemicals trigger this chain of events and others did not. This was followed up by others (e.g Dupuis and Benezra, 1982) from which the electrophilic theory of covalent interaction between chemical sensitisers and skin proteins was hypothesised. For effective sensitisation, a chemical must either be inherently protein reactive or be converted (chemically or metabolically) to a protein reactive species. Chemicals that are unable to associate effectively with proteins will fail to stimulate an immune response. There are various types of electrophile-nucleophile reactions such as Michael-type reactions; S_N2 reactions; S_NAr reactions; acylation reactions and Schiff-base formation. These have been described with reference to skin sensitisation by a number of workers including Payne and Walsh, (1999), Smith Pease (2003), Gerner et al., (2004), Dimitrov et al., (2005a) and more recently by Aptula and Roberts, (2006).

Skin sensitisation predictive testing: *in vivo* methods

Currently, animal tests are the only means of definitely assessing skin sensitisation for regulatory purposes. For many years, guinea pigs were the species of choice for predictive sensitisation tests. Two types of tests were developed; adjuvant tests in which sensitisation is potentiated by the injection of Freund's Complete Adjuvant (FCA) and non adjuvant tests. OECD Test Guideline 406 (OECD, 1992) describes these tests as the Guinea Pig Maximisation Test (GPMT) and the Buehler test respectively. In the GPMT, a test substance is regarded as a sensitiser when at least 30% of the animals show a positive response (EEC, 1967). In the Buehler test, a test substance is regarded as a sensitiser when at least 15% of the animals show a positive response (EEC, 1967).

The test of first choice under REACH for skin sensitisation is the Local Lymph Node Assay (LLNA) as described in OECD Test guideline No. 429 (OECD, 2002). The LLNA is based upon the characteristics of induced proliferative responses in draining lymph nodes following topical exposure of chemicals to

mice. The endpoint is the stimulation index (SI), which gives a ratio of thymidine incorporation in lymph nodes from dosed animals compared to the incorporation in lymph nodes from vehicle-treated control animals. The test is positive when the stimulation index (SI) is greater than 3 for any of the dose concentrations. The EC3 value, interpolated from the dose response curve, is the effective concentration of the test substance required to produce a threefold increase in the stimulation index compared to vehicle-treated controls. It can be used as a quantitative measure of the relative sensitising potency. Further information describing the derivation and the use of the EC3 in risk assessment can be found elsewhere (Basketter et al., 1996; Basketter et al., 1999; Basketter et al., 2000; Basketter et al., 2002; Gerberick et al., 2000; Kimber et al., 2003; Basketter et al., 2003).

Data from other *in vivo* tests, such as the mouse ear swelling test (MEST) or human studies (Human Repeat Insult Patch Test (HRIPT) and (Human Maximisation Test (HMT)) may also be useful in evaluating skin sensitizers. More information can be found in Basketter et al., (2005) and cited references.

Regulatory classifications

Under existing EU legislation, substances and preparations are classified as sensitising and assigned the symbol 'Xi', the indication of danger 'Irritant' and the risk phrase R43 if practical experience shows the substance or preparation to be capable of inducing a sensitisation by skin contact in a substantial number of persons, or where there are positive results from animal tests.

Under the forthcoming Globally Harmonised System (GHS) of Classification and Labelling of chemicals, substances will be classified as contact sensitisers (Category 1) if there is evidence in humans that the substance can induce sensitisation by skin contact in a substantial number of persons, or if there are positive results from an appropriate animal test (http://www.unece.org/trans/danger/publi/ghs/ghs_welcome_e.html).
These types of regulatory classification are important considerations when trying to relate the modelled endpoints within (Q)SAR models to their subsequent regulatory application.

Sources of Skin sensitisation data

Systematic development of comprehensive databases designed to assist in the detection or prediction of skin sensitisation potential of chemicals was first started over 30 years ago. Notable amongst the early efforts were those by Ziegler (INPRET – Information on Predictive Tests) (Ziegler et al., 1989) and CSC Associated (CADES – Contact Allergens Database Evaluation Systems) (Sigman et al., 1994).

Other datasets include Cronin and Basketter, (1994) who published the results of over 270 *in vivo* skin sensitisation tests obtained in the same laboratory (mainly from the guinea pig maximisation test). A compilation of summary (animal and human) data for a range of chemicals was included in Smith and Hotchkiss, (2001). A larger database of animal and human studies for 1034 compounds has been described by Graham et al., (1996) though the dataset is not present in this manuscript. The BgVV list of 264 chemicals was compiled and evaluated by a group of experts including dermatologists from universities and representatives of the chemical industry and from regulatory authorities that was established by the German Federal Institute for Health Protection of Consumers and Veterinary Medicine (BgVV) in 1985 (Schlede et al., 2003). Data from the literature on substances with documented contact allergenic properties in humans (from clinical data and experimental studies) and from animal experiments were evaluated resulting in a publication where chemicals were listed as belonging to one of 3 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.
A large number of data have been published for the local lymph node assay including full test results for 106 chemicals by Ashby et al., (1995), data for 41 chemicals (Gerberick et al., 2004a), EC3 values for 93 chemicals (Estrada et al., 2003) and full data for 211 chemicals (Gerberick et al., 2005).

(Quantitative) Structure-Activity Relationship models and expert systems

A number of different approaches and philosophies are used to develop (Q)SAR models and expert systems. In the area of skin sensitisation, available published (Q)SAR models fall into one of two main categories; either they are chemical class based/mechanism-based (local models) or they are derived empirically using statistical approaches (global models). Some of the available SARs are chemical class based whereas others are mechanistically based and some of these have been encoded into expert systems (Payne and Walsh, 1994; Barratt et al., 1994a).

True mechanism-based QSARs take the form of an equation relating toxicity and physicochemical parameters, derived mathematically from the known mechanism of action, and by applying established principles of chemistry and physics. For the purposes of this review, we shall define these mechanism-based QSARs as Quantitative Mathematical Models (QMMs).

An empirical QSAR (quantitative structure-activity relationship) is an equation relating (a) quantitative parameter(s) derived from chemical structure to a quantitative measure of biological activity. This type of QSAR model is typically derived by applying statistical approaches firstly to select appropriate

descriptors from a large pool of calculated parameters and secondly to derive an algorithm (e.g. through using linear regression).

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 which are knowledge based and TIssue MEtabolism Simulator for Skin Sensitisation (TIMES-SS) which is a hybrid.

Whilst previous reviews have been conducted for skin sensitisation, (see Karol et al., (1999); Bashir and Maibach, (2000) and Rodford et al., (2003) as examples), this review aims to provide both a historical overview and a proposal for how skin sensitisation (Q)SARs could be more widely applied in a regulatory context bearing in mind some of the main challenges ahead.

Published SARs

2200 test results were extracted from Contact Dermatitis and used to derive and encode 63 structural alerts for skin sensitisation in the Structure-Activity tree (SAT) as developed by Benezra et al., (1985). Despite much internet searching, no further information as to the current existence of SAT or its associated database PROPHET (Rindone and Kush, 1980) has been found.

A number of simple correlative studies relating sensitising potency to chemical properties (e.g. chain length, substitution position, reactivity) for a range of chemical series have been undertaken by Hausen and co-workers. Examples studied included gallates (Hausen and Beyer, 1992), coumarins (Hausen and Beyer, 1989), sesquiterpene lactones (Hausen and Schmalle, 1985), benzoquinones (Cremer et al., 1987), napthoquinones (Schulz et al., 1977), Primin (2-methoxy-6-pentyl-1,4-benzoquinone) analogues (Hausen et al., 1995) and nonterpenoid and diterpenoid phenanthrenequinones (PACs) (Hausen et al., 2003).

Rao et al., (1981) evaluated various chemical groups for their skin sensitisation potential in the guinea pig. Chemicals with a proven capability to cause skin sensitisation included sulphur containing compounds: thienyl iodonium salts, sulphones, thiomethane sulphonates, nitrogen containing compounds: amines, acetanilides, diisocyanate, pyridines and piperidine derivatives. Positive responses also occurred with benzyl chlorides, some halo-methyl and hydroxyl-methyl diphenyloxides, some phosphoric acid derivatives and some epoxide containing structures. The structural features mimic those in systems such

as DfW. Nonetheless a re-evaluation taking into account likely reaction chemistry might be useful to establish whether this information sheds any new insights to help refine the domains established in Aptula et al., (2005a).

A set of structural alerts based on the structural requirements for reactivity with skin proteins assessed from existing experimental skin sensitisation data were defined by Payne and Walsh, (1994). The alerts were classified as far as possible according to anticipated reaction mechanism and included: alkylating agents (haloalkanes activated towards nucleophilic reactions by conjugation with unsaturated groups. Other examples included dialkyl sulphonates, dialkyl sulphates and epoxides); arylating agents (nucleophilic substitution at an appropriately activated aromatic centres); acylating/sulphonating agents (e.g. phenyl esters); Michael addition electrophiles and precursors, thiol exchange compounds, free radical generators and metabolism to reactive electrophiles. The structural alerts devised were then tested using a set of 93 chemicals with published skin sensitisation taken from the European Community's New Substances Regulations. The insights and findings were incorporated into the expert system DfW (LHASA Ltd, Leeds, UK) where possible.

Ashby et al., (1995) suggested broad structural activity relationships (SARs) for 106 chemicals tested in the LLNA. The 73 sensitisers were grouped into six main classes – potential electrophiles after metabolism (e.g. epoxide formers and aromatic nitro/amino compounds), electrophiles, Michael reactive agents (e.g. alpha,beta-unsaturated esters, amides and aldehydes), benzoylating agents (e.g. benzoyl chloride and phthalic anhydrides, phenyl esters and phenyl benzoates), ionic chemicals (e.g. acid groups such as carboxylic or sulphonic acids) and miscellaneous agents. The electrophiles were further segregated into eight major subgroups – alkyl halides, sulphonates, sulphates, nitrosoamides and nitrosoguanidines, aromatic alkyl halides, aromatic alkylating agents, acylating agents, miscellaneous including isothiocyanate groups and strained lactones. Miscellaneous agents included compounds containing a sulphur atom that might participate in a disulphide bond formation with sulphydryl (-SH groups). Whilst Ashby et al., (1995) went some way toward a mechanism-based classification, their groupings were part based on structural criteria, and part based on reaction mechanistic domains. A re-evaluation (Roberts et al., 2007b) was subsequently conducted with respect to reaction mechanisms as defined in Aptula et al., (2005a).

Zinke et al., (2002) evaluated the 40 original DfW structural alerts published in Barratt et al., (1994a) against the database developed in the German Federal Institute for Health Protection of Consumers and Veterinary Medicine (BgVV). The evaluation revealed eight could be used without any further refinement whereas ten of the alerts were thought to need further specifications or refinements in order to avoid false

positive predictions. There were insufficient substances to evaluate a further 10 of the alerts and for the remaining 12 structural alerts no comparative result could be obtained as these rules did not fire for any of the examined chemicals.

Gerner et al., (2004) derived two sets of new structural alerts using a regulatory database (hence unavailable) of sensitisers; 15 rules for directly acting sensitisers and 3 rules for chemicals that were sensitisers following metabolic activation. The alerts are extensively documented and proposed as being suitable for regulatory use. The predictivity rates of the rules were found to be reasonably high and the rules were substantiated with examples from the BgVV list now held at the BfR. The rules developed include those for hydrazines, hydrazonium salts and precursors, acid halides, aromatic sulphinic acids, acid anhydrides whereas the rules developed for the pro-haptens included catechols, resorcinols and hydroquinones, their precursors and diamines.

Statistical QSAR Models

Statistical QSAR models are those developed empirically by application of various statistical methods to sets of biological data and structural, topological and/or geometrical information descriptors. Often known as global QSARs, they are purported to be able to make predictions for a wide range of chemicals, covering a wide range of different mechanisms of action. A number have been reported in the recent literature of which some have been analysed in more detail in accordance with the OECD principles (see Roberts et al., (2007a) as an example).

Probably the first global models for sensitisation were those developed by Magee and Hostynek, (1994, 1997). They investigated the feasibility of deriving models for fragrance allergens using classification and ranking approaches. They developed a statistical discriminant QSAR model based on a dataset of 72 chemicals – comprising 36 sensitisers and 36 non-sensitisers. The model was based on a combination of continuous transport/binding descriptors and dichotomous descriptors reflecting protein reactivity by a count of toxophores. Molar refraction (*MR*), polarisable molecular volume were used to model transport whereas London forces in non-specific binding, *PL*, the lipophilic contribution of Log P (which models the interaction of compounds with proteins and membranes) were continuous descriptors used to model binding properties. In addition, the number of Hydrogen Bond Acceptors (*HBA*) and Hydrogen Bond Donors (*HBD*) as well as a number of substructures; *ICOOR* (simple aliphatic esters and models esterase degradation), *ICONJ* (conjugated olefins activated by groups that cause Michael addition of SH and NH2), *IX* (reactive aliphatic or aromatic halides), *IArOH* (easily metabolised phenols to reactive quinines), *IArNH2* (easily metabolised anilines to reactive quinines), *IOH* (easily oxidised primary alcohols to reactive aldehydes), *IUNIQ* (strong nucleophiles or unusual reactive electrophiles), *IPOS*

(Michael reactants, Reactive Nucleophiles and Electrophiles) were all used to model protein reactivity. In these discriminant models, predicted sensitisers were those scoring a value greater than 0.5, non-sensitisers scored below 0.5, and values between 0.4-0.6 were designated as indeterminate (i.e. no prediction was possible). Early evaluation studies found the discriminating power to be of the order of 79-88% correct classification where non-sensitisers were predicted somewhat better than sensitisers. The resulting model took the form (Equation (1)):

$$\text{Classification score} = 0.0116MR - 0.161PL + 0.0498HBA - 0.129HBD - 0.284ICOOR + 0.57ICONJ + 0.16IX + 0.596IArOH + 0.503IArNH2 + 0.207IOH + 0.512\ IUNIQ + 0.118 \quad (1)$$

$$N = 70 \text{ (2 outliers) } R^2 = 0.75\ s = 0.275\ F = 15.81$$

The classification model was unable to predict 12 chemicals (i.e. indeterminate), for the 62 remaining chemicals, the model predicted 57 correctly (92% concordance). The model overall performed better on non-sensitisers (12/12, 100% specificity) than on sensitisers (45/50, 90% sensitivity).

The second of these discriminant models was a rank model based on a set of 89 chemicals using largely the same descriptors. The method of ranking was based on non, weak, moderate and strong responses where a numerical scale corresponded to the classification of potency scores. These scale was a modification of the Magnusson-Kligman scale. The specificity was 60% (9/15) and sensitivity 95% (56/59) with an overall 88% (65/74) model performance. The final model (Equation (2)) is given below.

$$\text{Ranking score} = 0.0294MR - 0.281PL - 0.0960HBA - 0.207HBD - 0.409ICOOR + 0.547IOH + 0.519IX + 1.375IPOS + 1.39IQUIN + 1.7 \quad (2)$$

$$N = 29 \text{ (non)}, N = 18 \text{ (weak)}, N = 15 \text{ (moderate)}, N = 26 \text{ (strong)}$$

$$N = 88\ R^2 = 0.726\ s = 0.68\ F = 23$$

Cronin and Basketter, (1994) used stepwise discriminant analysis on a set of 259 chemicals (mostly tested in the GPMT in the spirit of the OECD guidelines) to produce a 14 parameter model for the prediction of sensitisation. 12 of the 14 descriptors in the final model were structural features associated with reactivity. The other two descriptors were the Shannon index (Dearden et al., 1988) which considers a measure of molecule size (and in turn is probably related to penetration) and the difference in HOMO and LUMO energies, thought to be related to reactivity (a large difference between HOMO and LUMO energies implies high stability and thus low reactivity (Zhou and Parr, 1990) – the sensitisers in this dataset did indeed have lower energy differences). The model predicted 82.6% of compounds correctly after cross-validation but had a tendency to predict non-sensitisers better than sensitisers (88% correctly predicted

c.f. 76%). Phenols and acetate derivatives were predicted poorly in particular. The concluding remarks of the authors was that structural alerts might have been a better means of identifying potential sensitisation hazard as the statistical model was poor at discriminating between the two classes. Whilst the compiled dataset included potency classifications (Barratt et al., 1994a) for the GPMT, the statistical model was only trained to discriminate between presence and absence of sensitisation hazard.

Devillers (2000) used the dataset of (Cronin and Basketter, 1994) to derive a neural network model relating the same 14 descriptors to sensitisation outcome. A three-layer feed forward neural network trained by the back-propagation algorithm previously described (Devillers, 1996; Rumelhart et al., 1986) was used on a training set of 242 chemicals and a test set of 17 chemicals. After more than 200 trials, the optimal architecture and parameters of the ANN were determined. The total percentage of correct classifications was 89.19% (cf. 82.6% in Cronin and Basketter, 1994). 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 v.s. linear and a lack of transferability. These tradeoffs are not insurmountable in terms of demonstrating scientific validity as evidenced in Vracko et al., (2006) but since Devillers' (2000) primary aim was to demonstrate the utility of the neural network approach in discriminating between sensitisers and non-sensitisers rather than trying to predict sensitisation potential or its underlying mechanism, the performance statistics alone would perhaps be insufficient to select this model over another simpler one.

Federowicz et al., (2004) developed a logistic regression model that related *nDB* (number of double bonds), *C-003* (number of CHR3 molecular subfragments), *GATS6M* (autocorrelation coefficient) and *HATS6m* (GETAWAY descriptor) to LLNA data (comprising 25 sensitisers and 29 non-sensitisers taken from (Ashby et al., 1995; Haneke et al., 2001). The Dragon software (Talete srl, Milan, Italy) was used to calculate 1024 descriptors. The statistical analysis was carried out using SAS 8.2 statistical package. Overall, 420 out of the 1024 descriptors were found to be statistically significant at the *p*-level of 0.05. Four main classes of descriptors were chosen; Radial Distribution Function descriptors which are based on the distance distribution in the geometrical representation of the molecule, topological descriptors which are based on molecular graphs as a source of different probability distributions to which information theory definitions are applied, GETAWAY class descriptors (Todeschini and Consonni, 2000) which try to match the three dimensional molecular geometry provided by the molecular influence matrix and atom relatedness by molecular topology with chemical information by using different atomic weight schemes and BCUT, a class of molecular descriptors defined as eigenvalues of a modified connectivity matrix (Burden, 1989; Burden, 1997). The descriptors were chosen to have an association with sensitisation activity but the relevance of these descriptors to mechanism was not explained in the

article. The QSAR predicted 83% of the responses correctly based on the training set of compounds and 79% in a cross validated test set. A characterisation of this model with respect to the OECD principles is described in Roberts et al., (2007a) where the lack of plausible mechanistic rationale is scrutinised in particular.

Using an augmented dataset of chemicals tested in the LLNA taken from the ICCVAM dataset (Haneke et al., 2001) of 209 chemicals, Fedorowicz et al., (2005) then developed logistic regression models using a subset of 178 organic chemicals. Three organic chemicals (sodium lauryl sulphate, benzalkonium chloride and streptozotocin) were removed from the dataset citing possible interference between skin sensitisation and skin irritation responses. The resulting dataset comprised 132 sensitisers and 46 non-sensitisers. Of the same dataset, 105 chemicals had guinea pig data, 82 sensitisers and 23 non-sensitisers. 262 descriptors were calculated using Cerius 2 software (Accelrys Inc, San Diego, CA, USA), 1204 descriptors were calculated with Dragon 2.1 software (Talete srl, Milan, Italy) and a further 747 with MolconnZ software (eduSoft, LC, Ashland, VA, USA). After removal of repeated and constant variables, 1777 descriptors were left. Stepwise logistic regression was used to select descriptors from the set of 1777 descriptors generated. The following models based on LLNA and GPMT datasets respectively were derived (Equations (3) and (4)):

$$\text{Logistic regression model outcome (Guinea pig)} = 9.53 - 2.01Hmax - 19.9(R3e+) - 15.8(R3u+) - 12.9HATS1u \quad (3)$$

$$\text{Logistic regression model outcome (LLNA)} = 15.2 - 0.345Hy + 0.351Gmin - 7.76RARS - 4.09BELe1 - 0.259Dipole \quad (4)$$

The molecular descriptors were not interpreted at a molecular structure level but their definitions were: *Hmax* maximum H electrotopological state; *R3e+* maximal autocorrelation of Lag3/weighted y atomic Sanderson electronegativities; *R3u+* maximal autocorrelation of lag 3/unweighted; *HATS1u* leverage weighted autocorrelation of lag 1/unweighted; *Hy* hydrophilic factor; *Gmin* minimum E state; *RAR* R matrix average row sum; *BELe1* lowest eigenvalue, no. of Burden matrix/weighted by atomic Sanderson electronegativities; *Dipole* dipole moment.

The descriptors underpinning the guinea pig model suggested polarity and hydrogen bonding were important factors in skin sensitisation whereas polarity, charge distribution and lipophilicity were associated with the LLNA model.

Both models produced correct responses of greater than 80%. Their sensitivity values were greater than 94% though their specificities were very low. This was attributed to the unbalanced datasets, in that there was a 3:1 ratio of sensitisers to non-sensitisers in the LLNA dataset and a 4:1 ratio in the guinea pig dataset. If the cutoffs in the specificity and sensitivity curves were changed, a better classification performance was feasible. This in itself showed how sensitive the logistic modelling approach was to the size and structure of the dataset.

The model performances were additionally compared with DfW (Version 6, LHASA Ltd, Leeds, UK) and Topkat (Version 6.2, Accelrys Inc, San Diego, CA, USA). The first QSAR model based on guinea pig data only was compared with DfW and Topkat. The second model based on LLNA data only was compared with DfW (since Topkat is based entirely on guinea pig data).

The logistic regression models were shown to perform better (in terms of their correct classification) than the two expert systems, DfW and Topkat. 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.

As a followup to publications by Federowicz et al., (2004) and Federowicz et al., (2005), Li et al., (2005) used random forest and classification tree modelling methods to develop models to predict skin sensitisation activity. The same dataset of 178 organic chemicals taken from the ICCVAM report (Haneke et al., 2001) was used as the training set. This dataset comprised 131 sensitisers and 47 non-sensitisers. 262 molecular descriptors were then calculated using Cerius 2 (Accelrys Inc, San Diego, CA, USA), 1204 Dragon (Talete srl, Milan, Italy) descriptors were calculated as well as 747 descriptors using MolConnZ (eduSoft, LC, Ashland, VA, USA). After removal of duplicates and constants, 1380 descriptors remained. A random forest technique and a standard classification technique were then used to develop models. A classification tree with 4 descriptors was developed with a very high sensitivity of 99.2% but a very low specificity of 34%. The descriptors in this model were *Atype C3* (count of atom types CHR3), *nsssCH* (count of atom types), *SsssCH* (sum of atom type E-state), *Hy* (Hydrophilic factor).

A random forest approach was then deployed firstly using all available 1380 descriptors, before half of those descriptors with the lowest importance were dropped from the model. The remaining descriptors were used as the input set for the formation of a new random forest and so on until only a few descriptors remained in the model. This could be further fine tuned using a backward elimination procedure using the penultimate random forest such that a minimum of two descriptors remained in the final model. Models were created with 10, 7 and 5 descriptors. The sorts of descriptors found in these models included

minimum E state (*Gmin*), *Hmaxpos* (maximum H E-state), *R4e+*, R maximal autocorrelation of Lag 4/weighted by Sanderson electronegativities, sum of delta-I values (*sumdelI*), *SEigp* (eigenvalue sum from polarisability weighted distance matrix. The best model reported was the random forest with 10 descriptors. At a threshold of 0.7, the specificity was 83% and the sensitivity was 84%.

In addition to classifying chemicals as active or inactive, the random forest approach was also 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. This was applied to the 10-descriptor Random forest model to give some insights into how the random forests combined structurally similar chemicals and how these shared similar molecular patterns were associated with skin sensitisation activity. A number of groups emerged from the clustering routine including halogeno derivatives, simple phenyl derivatives, structurally diverse surfactant like chemicals, alkyl and aliphatic acid halides and polycyclic aromatic hydrocarbons, aldehydes, ether, methylene and peroxide or lactone groups, aliphatic halides, catechols, low weight highly hydrophobic chemicals with 4-8 carbon atoms in their structures, surfactants with long hydrophobic aliphatic chain of 18-19 carbon atoms etc. A detailed interpretation of each of these groups was not performed by the authors. However, the random forest approach had largely separated sensitisers from non-sensitisers and each group was made up of structurally similar chemicals.

Estrada et al., (2003) performed a linear discriminant analysis relating topological descriptors to skin sensitisation 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 physicochemical properties and raised to different power (in this case, the bond matrix was raised to an order of 15). The 6 weighting properties were lipophilicity, polarisability, polar surface area, molar refractivity, atomic charges and van der Waals radii. A set of 93 diverse chemicals and their associated LLNA EC3 values were collated. The EC3 values were categorised into bands of potency (Basketter et al., 2002). A chemical with an EC3 value less than 1% was defined as strong sensitiser, one with an EC3 between 1 and 10%, moderate, 10-30% weak, 30-50% extremely weak and greater than 50% non-sensitising. These classifications were further amalgamated into two groups. The 93 compounds were divided at random into a test set (15 compounds) and training set (78 compounds). A two model classification system was then developed; the first model discriminated strong/moderate sensitisers (EC3<10%) (class 1) from all other chemicals. The second model differentiated weak sensitisers (10%<EC3<30%) (class 2) from non-sensitising/extremely weak sensitisers (EC3>30%) (class 3). Model 1 was able to correctly classify 80% of the 75 compounds in class 1. Thirty-one of the 39 strong/moderate sensitisers were correctly predicted. Eight compounds were predicted to be false negatives. Three compounds were found to be statistical outliers and one compound

from class 2/3 could not be classified by Model 1 (Equation (5)). This resulted in a 80% (28/35) correct classification.

Class1 model = (1.331*$\mu 1H$) - (0.00598*$\mu 4H$) + (0.0078*$\mu 2PS$) -(0.00021366*$\mu 3PS$) + (0.0755*$\mu 1MR$) + (0.0319*$\mu 2MR$) - (0.0011133*$\mu 5Pol$) - (2.3797*$\mu 1Ch$) + (0.1547*$\mu 3Ch$) + (0.00425*$\mu 6Ch$) + (2.0932*$\mu 1vDW$) - (0.8683*$\mu 2vDW$) + 0.7954 (5)

Wilks - lamba = 0.61 F = 3.39 $D2$ = 2.52 p<0.0007

36 compounds were used to develop Model 2. 80.5% of the training set compounds in Model 2 (Equation (6)) were predicted correctly: 29 were classified correctly and the other seven were false positives. The overall correct classification of this model was 80.26%.

Class2 model = (0.946*$\mu 1H$) - (0.00468*$\mu 7H$) - (0.894*$\mu 1PS$) + (0.1004*$\mu 2PS$) - (0.0024*$\mu 3PS$) + (0.0057*$\mu 3Pol$) - (1.429*$\mu 1Ch$) + (0.0053*$\mu 8Ch$) - (0.00111*$\mu 9Ch$) - 5.309

(6)

Wilks - lamba = 0.38 F = 4.63 $D2$ = 8.76 p<0.001

A dataset of 15 compounds unseen by the model and with available skin sensitisation data were taken from other available existing public sources formed the basis of the external validation set to test the robustness and predictivity of model 1. Owing to the inavailability of further data, an external validation was carried out only for Model 1. All but one of the strong/moderate sensitisers were classified correctly. Since the validation set was not designed rigorously, this good performance might well have been on the basis of chance. No assessment of the representativeness of the external test set relative to the training set was conducted at the time. The importance of representativeness of test sets relative to training sets is linked to the issues of applicability domain that are described within the OECD principles (OECD, 2004; OECD, 2007). This has been the subject of much debate (Netzeva et al., 2005; Jaworska et al., 2005) and scientific research has been undertaken to develop such approaches (see Dimitrov et al., 2005b; Nikolova-Jeliazkova and Jaworska, 2005).

The descriptors in these two models do appear quite complex although a mechanistic interpretation is 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 sensitisation whereas as the parameters such as charges and polarisability may relate to the reactivity or electrophilicity of the chemicals. Subsequent work in (Roberts et al., 2007b) has re-evaluated the extent to which descriptors such as charges and polarisability are able to model electrophilicity per se. Modelling reactivity remains a challenge (Schultz et al., 2006)

and formed one of the rationales for conducting the work as described in Aptula et al., (2006). This will be discussed in more detail later.

The unique feature of the TOPS-MODE descriptors is that in addition to using the whole descriptors for QSAR modelling, local spectral moments for each of the bonds could also be calculated. These enabled features which contribute positively or negatively to skin sensitisation to be identified, leading to structural rules or alerts that could be later embedded into expert systems such as DfW. Examples of structural rules derived in this way are discussed in both Estrada et al., (2003) and Estrada et al., (2004). In Roberts et al., (2007a) these insights were thoroughly assessed (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.

Miller et al., (2005) collected a set of 87 LLNA data and after removal of 20 outliers (the basis for removal being the poor fit between calculated and experimental EC3 values) 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 CODESSA (Comprehensive Descriptors for Structural and Statistical Analysis (Codessa), Semichem, Inc.: Shawnee Misson, KS, USA). These descriptors may be categorised as being topological, constitutional, geometric, electrostatic, thermodynamic and quantum chemical in nature. 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 sensitisation". The aim of the QSAR developed was that it should cover a broad range of structures, including halogenated and aromatic compounds, alcohols, aldehydes and ketones. Compounds "with iodine, sulphur and anhydride rings, adjacent ketones and alkene aldehydes which are not Michael reactants" were cited as being outside the scope the QSAR. These 20 compounds were removed from the training set on the basis of a poor fit between calculated and experimental values. The resulting QSAR (Equation (7)) derived was:

$$EC3 = 9.16\ FPSA2ESP + 4.29\ EHOMO\text{-}LUMO - 45.89 \qquad (7)$$

$$N = 50\ R^2 = 0.773\ R^2\ adj = 0.763\ R_{CV} = 0.738\ F = 79.9$$

where

FPSA2ESP – fractional positively charged surface area descriptor based on electrostatic potential charge (though to be related to partition)

EHOMO-LUMO – is the energy gap between Highest Occupied Molecular Orbital (HOMO) and Lowest Unoccupied Molecular Orbital (LUMO) (thought to be associated with reactivity)

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. These were defined as: Strong, EC3 < 6 %; moderate 6 % < EC3 < 13 %; weak EC3 >13 %. These cutoffs were chosen to balance the number of chemicals that fell into the strong and moderate categories. The bin selection method was found to predict 44 out of 50 training set chemicals correctly. Use of the EC3 (as a weight percentage) rather than a molar basis EC3 appears to be a flawed choice. It disguises the fact that the model is poor at discriminating within a given category. For example with N-Methyl-N'-nitrosourea, the literature value is 0.05 and the predicted value is 3.27, a residual of 3.22 which appears to be a reasonable agreement. A comparison on the basis of molar EC3 reveals the prediction to be underestimating the potency by a factor of 65. In addition, had the proposed category boundaries for the LLNA been used as defined in (Kimber et al., 2003), than the QSAR model would have been shown to perform much less favourably. Further evaluation of this model has been described extensively in Roberts et al., (2007a).

A support vector machine (SVM) (Benables and Smith, 2003) was used to develop a nonlinear binary classification model for skin sensitisation 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 sensitisation (though this was not explained in the paper) and were used as inputs of the SVM model. The descriptors were Relative number of Nitrogen atoms, Gravitation index, Kier and Hall index, YZ Shadow, Maximum partial charge for a C atom, Maximum 1-electron reactivity index for a C atom. The nonlinear model developed from SVM (89.77% correct classification) algorithm outperformed the LDA (correct classification 79.55%) (Ren et al., 2006).

The feasibility of using 4D fingerprints derived from a methodology called 4D molecular fingerprint similarity analysis was investigated to develop QSAR models for skin sensitisation (Li et al., 2006). An early version of the LLNA dataset published in Gerberick et al., (2005) comprising data on 219 chemicals was sorted into 3 categories; non-weak sensitisers (101 chemicals), moderate sensitisers (72 chemicals) and strong-extreme sensitisers (46 chemicals). The moderate sensitisers were removed to create a separation gap and the other two classes were retained to develop a model to discriminate between them. Ten non-weak and 5 strong-extreme sensitisers were randomly chosen to be the test set. The remaining training set comprised 91 non-weak chemicals and 41 strong-extreme sensitisers. 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 square coupled with logistic regression were then applied to generate categorical models to differentiate between sensitisers and non-sensitisers.

The 4D fingerprints are eigenvalues of molecular similarity eigenvectors determined for a molecule from its set of absolute molecular similarity main distance dependent matrices. The descriptors aim to encode information about the composition of a chemical, its size, shape and conformational flexibility. 720 descriptors were calculated in total. Two QSARs both composed of nine 4D-fingerprints were built using stepwise linear logistic regression on both standardised and unstandardised descriptors. The model (Equation (8)) using standardised descriptors is given below.

$$\text{Logit}(P) = -1.745 + 1.119 \in 3(\text{hbd,aro}) + 2.686 \in 12(\text{hs,aro}) - 0.9722 \in 18(\text{all,all}) + 1.654 \in 2(\text{np,p-}) - 0.9936 \in 2(\text{all,all}) - 0.8841 \in 1(\text{p-,hbd}) - 3.966 \in 19(\text{hs,all}) + 1.466 \in 54(\text{all,all}) + 1.388 \in 39(\text{all,all}) \quad (8)$$

The performance scores were invariant to the standardisation state of the descriptors. The performance scores ranged from 79.5-85.6% accuracy, 68.3-78% sensitivity and 80.2-93.4% specificity for the training sets depending on the classification cut offs. Cross validation accuracies varied between 77.3%-78%.

A stepwise logistic regression was then performed where 266 significant descriptors were picked from the original 720 descriptors resulting in Equation (9). Here the performance scores were similar with accuracy varying between 79.5-88.6%, sensitivity at 80.5% and specificity between 79.1-92.3% again depending on the classification cut off used. Cross validation accuracies varied between 75.8%-80.3%.

$$\text{Logit}(P) = -1.576 + 2.877 \in 3(\text{hbd,aro}) + 1.387 \in 12(\text{hs,aro}) - 4.911 \in 18(\text{all,all}) + 1.017 \in 2(\text{np,p-}) - 0.7601 \in 2(\text{all,all}) - 0.9919 \in 12(\text{hs,hs}) + 4.589 \in 15(\text{np,np}) - 1.890 \in 2(\text{np,hbd}) + 0.8954 \in 2(\text{hs,p+}) + 0.6735 \in 3(\text{np,hba}) \quad (9)$$

Although the classification accuracies of equations (9) and equation (8) were similar, the latter QSAR model was deemed to be a better model by virtue of its more favourable goodness of fit statistics and because it had fewer descriptors (9 vs. 10).

An interpretation of the descriptors used in equation (8) was conducted to rationalise what role they played in sensitisation mechanism. $\in 2(\text{np,p-})$, $\in 3(\text{hbd,aro})$, $\in 12(\text{hs,aro})$ all with positive regression coefficients suggested that non-polar and aromatic atoms of a molecule would increase the lipophilicity and hence promote absorption into the skin thus modelling the “penetration” hurdle in skin sensitisation. The covalent interaction between chemical and protein was thought to be encapsulated in fingerprints

$\in 2$(np,p-) and $\in 3$(hbd,aro) since atom types p- and hbd are indicative of the polarity and hydrogen extraction behaviour of the molecule. The positive regression coefficients of np and p- descriptors may indicate that an increase in a molecular polarity corresponded to an increase in reactivity and the probability that the molecule may be a strong-extreme skin sensitiser. The $\in 1$(p-,hbd) descriptor may indicate a need for a particular spatial distribution p- and hbd atom types over a molecule in the process of sensitisation. Since this descriptor had the smallest regression coefficient, it was thought that it might actually be a correction term. The significant 4D fingerprints in equation (8) including $\in 2$(all,all), $\in 18$(all,all) and the correction terms $\in 39$(all, all) , $\in 54$(all,all) and $\in 19$(hs, all) were thought to capture the important role that molecular shape, size and steric interactions play in skin sensitisation. It was further postulated that since the regression coefficients of $\in 2$(all,all) and $\in 18$(all,all) were negative whereas $\in 39$(all,all) and $\in 54$(all,all) were positive, a medium sized molecule might have the highest probability of being a non-weak sensitiser whereas a large molecule with 54 or more atoms was most likely to be a strong-extreme sensitiser.

A model was also developed using partial least square coupled with logistic regression. The best model (Equation (10)) derived was as follows

$$\text{Logit(P)} = -3.564\ 0.498\text{xscr4}+0.407\text{xscr4}+0.636\text{xscr5}+0.513\text{xscr8}+0.415\text{xscr10}+0.7\text{xscr13}+0.882\text{xscr20} \quad (10)$$

Where xscr^{ith} denoted the ^{ith}PLS component extracted from the 720 non-scaled 4D fingerprints. The performance scores for equation (10) were actually more impressive than for equation (8) (Accuracy 90.2-97% vs. 79.5-85.6%, Sensitivity 85.4-95.1% vs. 68.3-78%, Specificity 90.2%-97.8% vs. 80.2-93.4% and Cross validated accuracies between 87.1-89.4% vs. 77.3-78%. Equation (10) also had a higher goodness of fit, predictive power and contained fewer descriptors. Equation (8) was thought to be more stable to variations in cut off classifications and its descriptors could be more readily interpreted. No mechanistic insights were offered by the authors for equation (10) since the PLS component factors being linear combinations of the original 4D fingerprints were more difficult to interpret.

The models developed appeared to be able to differentiate between weak-non chemicals and strong-extreme sensitisers reasonably well but no prediction of potential moderate sensitisers could be made. The fingerprint descriptors used lack physical meaning and are extremely difficult to interpret in terms of skin sensitisation mechanism.

Statistical QSARs covering several mechanistic domains have become a growth area recently. Whilst some have high predictive rates, they often lack a sound mechanistic basis. The majority of these QSARs use selections from the same set of published LLNA data (Gerberick et al., 2005) suggesting an over emphasis on applying new statistical techniques or descriptors rather than trying to rationalise the underlying skin sensitisation mechanism. As a consequence, of the models described here; their application is thought to be best limited to screening.

Expert systems

According to the definition proposed in Dearden et al., (1997), "An expert system for predicting toxicity is considered to be any formalised system, not necessarily computer-based, which enables a user to obtain rational predictions about the toxicity of chemicals. All expert systems for the prediction of chemical toxicity are built upon experimental data representing one or more toxic manifestations of chemicals in biological systems (the database), and/or rules derived from such data (the rulebase)." Expert systems can also be characterised according to the nature of the rules in their rulebase. The systems that are available for the prediction of skin sensitisation comprise knowledge-based systems (e.g. Derek for Windows (DfW)), statistical systems (e.g. TopKat, MCASE) and hybrid of the two (e.g. Tissue Metabolism Simulator (TIMES)).

Knowledge-based systems

Derek for Windows (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 (Sanderson and Earnshaw, 1991). It is marketed and developed by LHASA Ltd (Leeds, UK) a not-for-profit company and educational charity (http://www.lhasalimited.org/index.php).

Within 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 sensitisation knowledge base in DfW was initially developed in collaboration with Unilever in 1993 using its historical database of guinea pig maximisation test (GPMT) data for 294 chemicals (Barratt et al., 1994a; Barratt et al., 1994b). Common reaction mechanisms (e.g. acylating agents, akylating agents, Michael electrophiles, aldehydes etc) were identified following grouping the chemicals in potency categories – strong sensitisers (70% - 100%), moderate sensitisers (30-70%), weak sensitisers (up to 30%), and non sensitisers 0%. For the purposes of deriving structural alerts, chemicals were grouped on the basis of these reaction mechanisms or by empirical derivation. Forty structural alerts were extracted from the groups including acid halides, acid azides, sulphonyl halides, phenyl esters, isocyanates, epoxides. Payne and Walsh, (1994) also identified reaction mechanisms to propose a set of alerts that could be incorporated into DfW which they evaluated

using a set of 93 chemicals. The DfW knowledge base is under constant development as new chemical insights and data become available. In Barratt and Langowski, (1999) the predictive ability of the sensitisation rules were assessed by processing 84 chemicals in the list of contact allergens issued by the German Federal Institute for Health Protection of Consumers (BgVV). The exercise helped to identify refinements and new rules (organic hydroperoxides, hydroxylamine and precursors, photoallergens) for incorporation into DfW. Suggestions for improvements have also come from work by Zinke et al., (2002), Gerner et al., (2004) and Langton et al., (2006). The current version (version 9.0.0) contains sixty-four alerts for skin sensitisation.

Statistical Expert Systems

Toxicity Prediction by Komputer Assisted Technology (TOPKAT) (http://www.accelrys.com/products/topkat/) marketed by Accelrys Inc (San Diego, USA) comprises two sets of sensitisation models that were developed originally by Enslein et al., (1997). Guinea pig maximisation data for 315 chemicals was assembled from published collections in Contact Dermatitis and from the dataset published by Cronin and Basketter, (1994). The data was then scaled according to classes defined by Barratt et al., (1994b). 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. A variety of descriptors (Enslein et al., 1994; Gombar and Enslein, 1995) were computed for the chemicals selected, and stepwise two-group discriminant analysis was used to identify relevant descriptors to build the models. The descriptors consisted of electronic attributes as embodied in electrotopological state values (Hall et al., 1991) on 1 and 2-atom fragments; transport attributes represented by the counts of the 1 and 2-atom fragments, topological kappa descriptors (Gombar and Jain, 1987; Kier, 1986) and molecular symmetry indices (Gomber and Jain, 1987), bulk attributes as calculated from molecular weight and atom size-corrected E values. The first set of models discriminated between non-sensitisers and sensitisers, the second set of models resolved the potency: weak/moderate vs. strong. In the first set of models, a computed probability greater than 0.7 indicated that a chemical was a predicted sensitiser whereas a probability below 0.30 signified a chemical to be a predicted non-sensitiser. In the weak/moderate vs. strong sensitiser model, a probability of 0.7 or more indicated a predicted strong sensitiser whereas a probability below 0.30 indicated a predicted weak or moderate sensitiser. For both models, probability values between 0.30 and 0.70 referred to an indeterminate region where a prediction was not reliable. An optimum prediction space (OPS) algorithm was also incorporated into the model to ensure predictions were made only for chemicals within the model domain. This domain check comprised two steps, whether the chemical belonged to the structural fragment and descriptor space of the model as well as a similarity analysis to extract to compare similar compounds by reference to the QSTR similarity

distance and their associated predicted and experimental GPMT data. The present version of TOPKAT (Version 6.2) has been supplemented with data for a 20 chemicals (Accelrys, 2002).

CASE methodology and all its variants (e.g. MCASE, CASETOX) were developed by Klopman and Rosenkranz (Klopman, 1984; Klopman, 1992; Rosenkranz and Klopman, 1995). There are more than 180 modules covering various areas of toxicology and pharmacology endpoints including skin sensitisation 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 sensitisation are described further in primary articles (Graham et al., 1996; Gealy et al., 1996; Johnson et al., 1997).

The original model developed was that by Graham et al., (1996) who applied the MultiCASE approach to a large database of chemicals causing allergic contact dermatitis (ACD) compiled from reports of animal and human studies. These reports were identified through a search on Medline from 1966-1994 and included primary articles and monographs prepared by RIFM and information from Aron Products Inc Chemicals so long as the certain criteria was met namely: an epicutaneous exposure to the chemical was used during the challenge phase of test and an in vivo dermal reaction was observed e.g. erythema, the dose used for challenge was non-irritating, the purity of the chemical was known, control animals were used in the study and found to be unresponsive to the challenge dose, in human testing at least 3 cases had to have been cited. In addition, the chemical tested had to have a carbon chain length of greater than one and all chemicals had to be non-metals. The total database consisted of a total of 1024 chemicals of which 317 were classified as sensitisers, 22 had marginal activity and 695 were inactive. MCASE identified 49 biophores consisting of nitrogen double bonded to a carbon or nitrogen, substituted aromatic structures, thiol and disulfide containing fragments and electrophilic moieties. Modulators of each biophore which either augmented or decreased potency included additional structural fragments, 2D distance descriptors as well as physicochemical parameters (HOMO, LUMO, MW and Log P). Internal evaluation of the database revealed a sensitivity of 99.6% and specificity of 99.4%.

Further work was undertaken by Gealy et al., (1996), though the primary purpose of this study was to evaluate the utility of clinical case report data. The original database (from Graham et al., 1996) was supplemented with additional case report studies to result in a total of 1170 chemicals – 447 actives, 28 marginal and 695 inactives. An additional condition to those stipulated in the previous study was that for a chemical to be classified as a non-sensitiser, it had to have been tested in either a human or guinea pig

maximisation test and yield no response in at least 10 guinea pigs or 20 human subjects. A series of different MultiCASE models were developed using subsets (learning sets) of the database as training sets with differing stringency requirements as far as how much or how little clinical case report data was included. In the most stringent subset, all clinical case report data was excluded. The model performances were then compared in order to assess the value of including case report data. The predictive ability of the models was similar for all models regardless of the learning subsets. The predictions remained fairly constant for all learning sets producing a mean sensitivity of approximately 64% and mean specificity of 92%. However the specificity of the learning set model which included only one or two case reports did have a lower mean specificity (85%) when compared with the other learning sets suggesting that it is best to exclude isolated clinical case reports when modelling ACD.

In Johnson et al., (1999), human activity data compiled from the literature on 458 compounds was used to develop CASE/MCASE models. Chemicals were selected if standard human maximisation tests (HMT) had been performed or if three or more case reports of positive patch tests were cited. The dataset was split equally between 229 sensitisers and 229 non sensitisers. In reality there were a total of 695 inactives. Three test models were created first with different non overlapping subsets of 229 chemicals and the resulting structural alerts identified were the same in each case. One model was then selected at random to continue the evaluation. Each chemical in this model was assigned a potency based on the concentration of chemical used for the challenge dose and in HMTs, on the sensitisation rate. This was then used to predict the likelihood of 238 gas phase air pollutants compounds to be sensitisers. 21 of the 238 chemicals were predicted to be contact allergens with potencies ranging from mild to very strong. The compounds came from chemical classes including chlorinated aromatics and chlorinated hydrocarbons, N-containing compounds, phenols and alkenes. The emission rates or measured airborne concentrations as an indication of extent of use together with the predicted sensitising potencies provided an efficient means of prioritising which organic compounds detected as gas phase air pollutants merited experimental sensitisation testing.
The (Q)SAR estimates for the MCASE skin sensitisation model are included in the Danish Environmental Protection Agency (EPA)'s (Q)SAR Database which is hosted as on the European Chemicals Bureau (ECB) website http://ecb.jrc.it/QSAR/.

Hybrid Expert Systems

The TImes MEtabolism Simulator platform used for predicting Skin Sensitisation (TIMES-SS) is a hybrid expert system that was developed by the Laboratory of Mathematical Chemistry (University of Bourgas, Bulgaria) using funding and data from a Consortium comprising Industry (ExxonMobil, Procter and Gamble, Unilever) and a Regulatory Agency (Danish Environmental Protection Agency). TIMES-SS

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. 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 sensitisation. 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 (Dimitrov et al., 2005a; Dimitrov et al., 2005b). Recently an external validation activity was undertaken whereby data were generated for 40 new chemicals in the 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 (Patlewicz et al., 2007a; Roberts et al., 2007c)

Overall there are reasonable prospects for the development and improvement of expert systems such as DfW, TIMES-SS but further refinement of their underlying rules is still required (Patlewicz et al., 2007a, b). Owing to the limited accessibility of training datasets, applicability domains not always being well established, or mechanistic interpretability being unclear; it is proposed that whilst expert systems can play a valuable role in evaluating skin sensitisation; their use should be restricted to the provision of supporting information as part of an overall weight of evidence assessment.

Local QSAR Models

Roberts and Williams, (1982) developed the first mathematical model of the *in vivo* alkylation process, known as the Relative Alkylation Index (RAI) model which quantified the degree of carrier alkylation with sensitisation potential. The model can be summarised as follows:

- The carrier protein is in a lipid environment (e.g. cell membrane) which is washed by a polar fluid (e.g. lymph or blood)
- Alkylation kinetics are first order.

Based on these assumptions, rate equations were set up for alkylation and disappearance of the test compound from the system (by alkylation and partition into the polar fluid) and from these the RAI was derived.

$$RAI = \log(kD/P+P^2) \qquad (11)$$

where k is the alkylation rate constant measured against a standard nucleophile, D is the molar dose and P is the partition coefficient measured between a standard polar/non polar solvent e.g. log P. In its most general form the RAI, an index of the relative degree of carrier protein haptenation, is expressed as:

$$RAI = \log D + a \log k + b \log P \qquad (12)$$

Thus the degree of haptenation increases with increasing dose D of sensitiser, with increasing reactivity (as quantified by the rate constant or relative rate constant k for the reaction of the sensitiser with a model nucleophile) and with increasing hydrophobicity.

The RAI model has been used to evaluate a wide range of different datasets of skin sensitising chemicals. It has in particular served well for the development of QSARs for small sets of structurally similar chemicals – i.e. local models which are chemical class specific. Examples include primary alkyl bromides (Basketter et al., 1992), sultones (Roberts and Williams, 1982), acrylates (Roberts, 1987), sulfonate esters (Roberts and Basketter, 2000), urushiol analogues (Roberts and Benezra, 1993), aldehydes and diketones (Roberts et al., 1999; Patlewicz et al., 2001; Roberts and Patlewicz, 2002) amongst others. A selection of these examples are described in more detail.

The RAI approach has withstood the test of time and evolved in light of greater understanding of the chemical processes and mechanisms underpinning skin sensitisation and now the modeling approach has found itself to be applicable to groups of chemicals that are related by virtue of their similarity to chemical mechanism rather than merely just their structural similarity.

In Basketter et al., (1992), LLNA pEC3 values (the log of the molar EC3 concentration) were found to correlate with a quadratic function of log P, for a series of linear primary alkyl bromides from C4 up until C18. Log P values were calculated by the computerised Hansch and Leo method (1979), Clog P (CLOGP™ program version 4.0, Biobyte Corp., 201 West 4th St. Suite 204, Claremont, USA). The correlation equation was obtained by multiple linear regression of pEC3 against log P and $(\log P)^2$. EC3 values are usually determined from dose response plots though the exact details of how the EC3 values were obtained were not described. Two compounds, 1-bromobutane and 1-bromohexane, classed as non-sensitisers, were not included in the regression. The regression equation (13) reported was:

$$pEC3 = 1.61 \log P - 0.09 (\log P)^2 - 7.4 \qquad (13)$$

$$N = 9\ R = 0.97\ s = 0.11\ F = 50.0$$

In this set, the longest chain lengths demonstrated decreased sensitisation activity rather in the manner of what is termed an overload effect. The dose response data for the individual bromoalkanes showed that in each case (except for C4) the sensitising activity increased with concentration. This supported the argument that none of the bromoalkanes were in fact in the overload region where activity would be inversely related to dose but that skin penetration was becoming rate-determining. A reanalysis has been conducted by Aptula and Roberts (unpublished, 2006) where it was proposed that it would have been better to have treated C4 and C6 as negative outliers (rationalisable in terms of penetration becoming a sensitisation-determining factor) and to do the regression for the remaining compounds using only the log P descriptor.

In Roberts and Basketter, (1990), hybrid dose-response/QSAR regression analyses were performed for guinea pig sensitisation and cross-challenge data on a series of alkyl alkanesulphonates and alkenesulphonates. Sensitisation responses were correlated with RAIi and RAIc, these being the Relative Alkylation Indices corresponding to induction and challenge respectively. In Roberts and Basketter, (1997) further sensitisation and cross-challenge data were compared against predictions from the equations developed in Roberts and Basketter, (1990). The TES value was obtained by summing the erythema scores for each animal. RAI values were calculated from the dose D (ppm units divided by MW), calculated (manually by the Hansch and Leo method (1979)) log P values and log k values which are taken as 1.48 for methyl alkanesulphonates and 0 for higher alkyl alkanesulphonates, by the formula:

$$RAI = \log k + \log D + 0.48 \log P \qquad (14)$$

Where RAI values model the degree of covalent binding of the sensitiser to carrier protein.
The dataset on which the quantitative analyses were based consisted of compounds of general formula RSO3R' where the Rs are alkyl or alkenyl. By current thinking as described in Aptula et al., (2005a), this dataset would now be expressed as the domain of non-ionic S_N2 electrophiles with at least one H-polar substituents.

The equation given in the paper was:

$$PR = 2.24\ RAIi - 0.36 RAIi^2 + 0.54 RAIc + 2.23 \qquad (15)$$

$$N = 27\ R^2 = 0.81\ s = 0.48\ F = 33.1$$

PR is the probit value corresponding to TES.
The negative $RAIi^2$ term models the overload effect.

Whilst a regression, this was not a "pure" QSAR as 27 data points come from 6 compounds. In the external validation presented in Roberts and Basketter, (1997), extra chemicals were tested and cross-challenges were performed. Results of these cross-challenges, in which animals sensitised to one compound were challenged with a different compound were presented. No statistical measurements were given, but predicted values were compared against the observed values. The results were quite good, particularly when log P values of the two compounds involved in cross-challenges were similar. Where there were large differences, the sensitisations responses were overpredicted.

In Roberts and Basketter, (2000), a 20-point hybrid dose-response/QSAR regression analysis was performed for LLNA data on a series of six alkyl alkanesulphonates and alkenesulphonates. Sensitisation responses were correlated with the RAI, corresponding to the single dose used in the LLNA. EC20 was deemed a more appropriate basis for the potency index than EC3 because of the potential large extrapolation errors with the latter. Using pEC20 as the dependent variable, regression analysis against RAI gave a six-point QSAR. RAI values were calculated with a negative log P coefficient, based on the concept that, for the compounds studied, *stratum corneum* penetration was negatively correlated with log P and was sensitisation determining. The SI value was obtained by the standard protocol, and its use to estimate EC20 values was clearly explained. RAI values were calculated from the dose D (ppm units divided by MW), calculated (manually by the Hansch and Leo method (1979)) log P values and log krel values which are taken as 1.48 for methyl alkanesulphonates and 0 for higher alkyl alkanesulphonates, by the formula:

$$RAI = \log C + 2 \log krel - 0.5 \log P \qquad (16)$$

where C is the concentration, in ppm moles, applied topically. RAI values model the degree of covalent binding of the sensitiser to carrier protein. This is different from the RAI formula used in the guinea pig study for the same compounds (in Roberts and Basketter, 1990; 1997), where the log P coefficient was positive.
RSP (relative sensitisation parameter) was calculated in the same way as RAI and used as the independent variable for the regression.

Based on more recent insights (Roberts et al., 2007b), the domain would be potentially defined as non-ionic S_N2 electrophiles with at least one H-polar substituent and with log P not less than 5.49 (below which value at some point the log P dependence will change from negative to positive).

The dataset on which the quantitative analyses were based consisted of compounds of general formula RSO3R' where the Rs are alkyl or alkenyl.

The hybrid dose-response/QSAR equation was:

$$\log SI = 0.39\ (\pm 0.05)\ RAI + 0.69\ (\pm 0.08) \qquad (17)$$
$$N = 20\ R^2 = 0.930\ s = 0.15\ F = 240$$

note: ± figures are 95% confidence limits, based on 2 standard deviations

The "pure QSAR" equation was:

$$pEC20 = 0.74\ (\pm 0.06)\ RSP - 0.61\ (\pm 0.14) \qquad (18)$$
$$N = 6\ R^2 = 0.994\ s = 0.10\ F = 702$$

note: ± figures are 95% confidence limits, based on 2 standard deviations

No external validation was performed owing to lack of suitable test data.

20 S_NAr electrophiles, whose GP skin sensitisation test results were reported by Landsteiner and Jacobs (1936) on a +/- basis, were analysed using calculated reactivity descriptors to discriminate between positive and negatives. In the original paper by Landsteiner and Jacobs, (1936) used an experimental reactivity descriptor, namely observation or not of reaction with aniline under defined experimental conditions.
In Roberts, (1996) the calculated descriptors were based on Hammett and Taft substituent constants (Perrin et al., 1981); in Mekenyan et al., (1997) they were based on molecular orbital (MO) parameters. The resulting Discriminant Analysis (DA) equation was used to predict sensitisation potential of a further set of seven S_NAr electrophiles. The analyses undertaken in these papers gave very good internal performance in terms of discriminating 100% between active and inactive chemicals. In Roberts, (1996)

seven chemicals were all predicted correctly and Mekenyan et al., (1997) the six additional compounds were all predicted correctly.

A set of 4 alpha, beta-diketones were tested in the murine local lymph node assay (LLNA) and a RAI QSAR developed using Taft substituent constants for the reactive carbonyl group as the electrophilicity parameter. Whilst there was only 4 compounds, a QSAR/dose response relationship could be derived. The equation was

$$\text{Log(T/C)} = 0.5\text{RAI} + 0.24 \qquad (19)$$

$$N = 11\ R^2 = 0.963\ s = 0.09\ F = 235$$

where RAI values were calculated using the relationship RAI = Log D + RP + 0.5 Log P
Here D is the % concentration applied divided by the molecular weight, Log P is the partition coefficient and RP is the reactivity parameter. The RP was defined as the sum of the Taft σ* constants for the 2 groups attached to the most electron deficient carbonyl group (Roberts et al., 1999).

Patlewicz et al., (2001) evaluated a selection of 17 aldehydes (13 sensitising and 4 non-sensitising) all of which possessed a benzene ring. The sensitising compounds were classified as strong, moderate and weak on the basis of in vivo data (GPMT and LLNA). Based on a detailed consideration of the mechanistic chemistry, it was found possible to classify each of the aldehydes in accordance with the reactivity of their carbonyl groups. The four subcategories were termed aryl-substituted aliphatic, aryl, aryl with special features that could undergo metabolism and alpha, beta-unsaturated aldehydes. Aryl substituted aldehydes were those likely to undergo a Schiff base formation with a primary amine. Aryl aldehydes comprised those compounds where the aldehyde functionality was directly attached the benzene ring. Aryl aldehydes with special features included vanillin and ethyl vanillin where the presence of the additional features enhanced the sensitisation potency. Alpha,beta-unsaturated aldehydes were thought to all be able to utilise the conjugated double bond to act via Michael addition.

Using the LLNA data available for the 5 alpha, beta-unsaturated aldehydes, a RAI QSAR was developed relating the log of the molar EC3 value (pEC3) to the Taft σ* substituent of the alkyl group on the α carbon which modelled reactivity. The following equation was derived using linear regression.

$$\text{pEC3} = 0.99\sigma^* + 1.50 \qquad (19)$$

$$N = 5\ R^2 = 0.998\ s = 0.018\ F = 1181$$

The good correlation derived reaffirmed the belief of the authors that consideration of probable reaction mechanisms was a necessary requirement when attempting to relate chemical structure to skin sensitisation rather than merely exploring structural similarity.

The ideas and insights derived in Patlewicz et al., (2001) became the starting point for further work in Patlewicz et al., (2002). Here the reaction chemistry principles were evaluated further by development of QSARs specific for Michael addition and Schiff base fragrance aldehydes. The following regression equations were developed:

$$\text{pEC3} = 0.536 + 0.168\text{LogP} + 0.488\text{R}\sigma^* + 1.313\ \text{R}'\sigma^* \qquad (20)$$

$$N = 9,\ R^2 = 0.741\ s = 0.184\ F = 4.77$$

$$\text{pEC3} = 0.245 + 0.278\text{LogP} + 0.862\text{R}\sigma^* \qquad (21)$$

$$N = 12,\ R^2 = 0.825\ s = 0.712\ F = 21.26$$

The σ^* values taken from Perrin et al., (1981) were used to model the electrophilic reactivity of the carbonyl group. Here $\text{R}'\sigma^*$ and $\text{R}\sigma^*$ reflect the Taft constants for the substituents on the alpha and beta carbons.

Patlewicz et al., (2003) continued with the theme of reaction chemistry and sought to investigate the extent of applicability of the Schiff base QSAR developed in Patlewicz et al., (2002) and whether it could be applied for the diketones evaluated in Roberts et al., (1999). The study was prompted since one of the aldehydes tested in Patlewicz et al., (2002), glyoxal (a 1,2-dial) was found to fit the QSAR model well. This also suggested that the site of attack, i.e. the electrophilic centre for the diketones tested in Roberts et al., (1999) was contrary to what had been previously proposed. Accordingly the QSAR was modified to the following;

$$\text{Log (T/C)} = 0.60\text{RAI} + 0.23 \qquad (23)$$

$$N = 11\ R^2 = 0.944\ s = 0.11\ F = 152.1$$

This modified QSAR equation was then used to calculate the pEC3 for four 1,4-diketones, whereupon a good agreement was found. These findings inferred that both aldehydes and 1,2-diketones sensitise via the same rate determining step forming Schiff bases. A new regression equation was developed for the combined set of compounds which took the form of:

$$\text{pEC3} = 0.16 + 0.32\ \text{LogP} + 0.96\sigma^* \qquad (24)$$

$$N = 16\ R^2 = 0.873\ s = 0.18\ F = 44.5$$

The final paper in this series of work was an evaluation of the predictive power of the QSARs developed in Patlewicz et al., (2002) with an additional set of compounds. LLNA data was generated on 10 newly selected compounds that contained a carbonyl group. The data demonstrated that the existing QSARs were fairly accurate but still required improvement. The new data was used to form two new equations.

The new QSAR for Michael addition aldehydes was:

$$pEC3 = 0.55 + 0.14LogP + 0.51R\sigma^*(\beta) + 1.07R'\sigma^*(\alpha) \qquad (25)$$

$$N = 13\ R^2 = 0.73\ R^2adj = 0.64\ s = 0.2695\ F = 8.16$$

For the Michael addition QSAR, agreement was fair. Where there discrepancies e.g. trans-2-methyl-2-butenal [497-03-0], 5-methyl-2-phenyl-2-hexenal [21834-92-4] and β-phenyl cinnamic aldehyde [13702-35-7] being overpredicted; plausible rationalisations in terms of reaction chemistry were proposed.
The agreement for the newly tested Schiff base compounds appeared poorer. The two 1,2-diketones tested were poorly estimated by the model but this was rationalised on account of the resonance effects of the benzene rings. Two aromatic aldehydes with strong withdrawing groups were poorly predicted but the finding helped to refine the scope of the "aryl aldehyde" group as defined in Patlewicz et al., (2001); thus not even a p-NO2 group diminished the resonance effect sufficiently to lead to significant sensitisation potential. In contrast, hexanal [66-25-1] and methyl pyruvate [600-22-6] were well predicted, the latter supporting the premise that other carbonyl-containing compounds in addition to aldehydes reacted in the same way and hence the QSAR model was applicable for these types of chemicals also. The pEC3 values of these two compounds were then combined into a new QSAR which took the form of.

$$pEC3 = 0.17 + 0.93\sigma^* + 0.30LogP \qquad (26)$$

$$N=14\ R^2= 0.87\ R^2adj= 0.85\ s = 0.165\ F =37.7$$

The knowledge generated from this work and previous publications was subsequently incorporated as refined rules in DfW (Langton et al., 2006).

Quantitative Mechanistic Models (QMM)

Up until recently RAI skin sensitisation QSARs, had only been developed for closely related sets of chemicals (e.g. sulphonate esters, saturated sultones) and it had been considered that the RAI model was only applicable for such closely related data sets. This view started to change following work on aldehydes and ketones but the term "reaction chemistry domains" or "mechanistic applicability domains" was not described until Aptula et al., (2005a). Here a data set of 41 LLNA EC3 values for a diverse range of compounds, which had been selected for internal consistency (Gerberick et al., 2004a) were analysed.

Compounds had been classified (Gerberick et al., 2004a) according to chemical class, e.g. ketone, aliphatic aldehyde, halogenated compound etc., with no clear relationship between structure and sensitisation potential. The work here re-classified the compounds into reaction mechanistic applicability domains, i.e. grouping according to the mechanisms whereby the compounds could react with nucleophiles. Each domain contained a diversity of structures, related by their common reaction chemistry rather than by common structural features. Within the domains it was also possible to rank the compounds in order of reactivity, applying established mechanistic organic chemistry principles, and in this way find clear trends within each domain of sensitisation potential increasing with increasing reactivity. The major reaction mechanistic applicability domains identified were: Michael-type addition domain; S_N2 domain; S_NAr domain; acylation domain; Schiff-base domain. Included in the Michael-type domain were pro-Michael acceptors, these being compounds which are not themselves Michael-reactive but are easily converted (e.g. by in vitro or in vivo oxidation to Michael acceptors). There was also an "unreactive" domain; compounds in this domain are expected to be non-sensitisers. Whilst these are not the only mechanistic applicability domains, they did provide a foundation to underpin skin sensitisation with sound reaction chemistry principles.

The conceptual framework in the area of reactive toxicity came about during the 1st Knoxville workshop, organised by Schultz and Veith in May 2005, (Schultz et al., 2006). The framework, which is described in detail in Schultz et al., (2006) outlined a number of sequential events – starting from the chemical to metabolism to an array of so-called molecular initiating events through to biological effects and final adverse outcomes. Chemical reactivity was thought to be the key molecular initiating event. From this starting point, it was clear that the mechanistic ability of chemicals to initiate these events was linked to the similar mechanism of action and this became the central reference when looking at similar chemicals thus many effects could be attributed to reactive toxicity. Approaches to encode reactivity such as using heats of reaction (Magee, 2000), an activation index (Aptula et al., 2005b), Taft coefficients (Perrin, 1981)) or experimental measures of reactivity of chemicals with model nucleophiles could be then used as correlates to a variety of toxic effects including skin sensitisation.

The insights derived in Aptula et al., (2005a) and the discussions from Knoxville (Schultz et al., 2006) were subsequently generalised in Aptula and Roberts, (2006) culminating in a set of rules based on organic reaction mechanistic principles with particular emphasis on reactive toxicity. The rules were exemplified using the protein binding reaction mechanisms associated with skin sensitisation, in part to provide a set of guidelines for the domain classification of reactive toxicants as well as to outline an *in chemico* approach for determining the relevant chemical properties required for toxicity. The rules themselves have subsequently been used to help evaluate the dataset (Ashby et al., 1995) in Roberts et al.,

(2007b). Early efforts in realising the *in chemico* approach came from investigating the correlation between thiol reactivity and LLNA for a set of Michael acceptors (Aptula et al., 2006). The thiol reactivity index was based on glutathione (GSH), pEC(50) thiol (EC(50) being defined as the concentration of the test substance which gave a 50% depletion of free thiol under standard conditions) in combination with a measure of cytotoxicity (pIGC(50)) to *Tetrahymena pyriformis* (TETRATOX). Thiol reactivity was found to discriminate sensitizers from non-sensitizers according to the rule: pEC(50) thiol>-0.55 indicating that the compound would be a skin sensitiser. However, because of metabolic activation a pEC(50) thiol<-0.55 does not necessarily mean that the compound will be a non-sensitiser. Excess toxicity to *T. pyriformis* (i.e. the extent of toxic potency over that expected by non-polar narcosis) was determined in order to assess biological reactivity. The best discrimination based on excess toxicity in the TETRATOX assay was given by the "rule": excess toxicity>0.50 indicating that the compound would be a skin sensitiser. These approaches became more powerful (23 of the 24 compounds were predicted correctly) when used in combination. The approach was promising but not the definitive answer since thiol reactivity is only one measure of the ability of chemicals to form adducts with proteins. Efforts in measuring protein reactivity in vitro is a field rapidly gaining momentum, other groups (such as Gerberick et al., 2004b) have been investigating approaches using a glutathione tripeptide or three synthetic peptides containing cysteine, lysine or histidine residues. In Gerberick et al., (2004b), the reactivity of 38 chemicals with varying skin sensitising potencies were investigated. The results revealed a correlation between skin sensitising potency and depletion of glutathione and binding with the lysine and cysteine synthetic peptides. Further work by Gerberick et al. has since been undertaken (2007).

A number of publications on the skin sensitisation of aldehydes and ketones have already been presented. The work described in Roberts et al., (2006) reanalysed the data from Patlewicz et al., (2004) with the aim of developing a new mechanistic QMM for Schiff bases. A QSAR was derived relating reactivity and hydrophobicity to the log of the molar EC3 (pEC3) where reactivity was modelled by $\Sigma\sigma^*$, the sum of the Taft σ^* substituent constants for the two groups attached to the carbonyl group.

A final QSAR based on 16 compounds (11 aliphatic aldehydes, 1 α-ketoester and 4 α,β -diketones) was developed.

$$\text{pEC3} = 1.12(\pm 0.07)\ \Sigma\sigma^* + 0.42(\pm 0.04)\ \log P - 0.62(\pm 0.13) \quad (27)$$

$$N = 16\ R^2 = 0.952\ R^2adj = 0.945\ s = 0.12\ F = 129.6$$

The predictive performance of this QSAR was then explored across the wider Schiff Base mechanistic applicability domain, using further LLNA data for 1,3-dicarbonyl compounds taken from (Gerberick et al., 2005b).

Two of the 1,3-dicarbonyl compounds of the further test set were classed as non-sensitisers (EC3 not reached at 40%). One of these was calculated to have EC3 = 59%; the other was calculated to have an EC3 = 39%. For the remaining seven 1,3-dicarbonyl compounds, the agreement between calculated and observed pEC3 values was good.

The RAI models have been shown to be robust, well characterised and mechanistically interpretable thus could play a strong role in the assessment of skin sensitisation as standalone tools. They are however limited in terms of their coverage. Considerable efforts would need to be invested in generating a database of reactivity data which would enable the rules in Aptula and Roberts, (2006) to be refeind and for new models covering the remaining domains to be developed.

Conclusions

There is clearly a very strong mechanistic understanding of skin sensitisation, which we have tried to summarise in the early part of this paper. Most published skin sensitisation QSARs have fallen into one of two main categories: either they are mechanistically model based 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 characterised (as evidenced in (Roberts et al., 2007a; Patlewicz et al., 2007b) and lacking in real mechanistic insight.
The covalent hypothesis continues to be the most promising way of developing mechanistically based robust QSARs. The reaction chemistry concepts recently outlined in Aptula et al., (2005a) and Schultz et al., (2006) have changed the perspective regarding RAI QSARs. These QSARs are actually more widely applicable than originally thought. Clear chemistry-activity trends can be seen within mechanistic applicability domains leading to robust mechanistic based QSAR models (or as defined earlier QMMs e.g. Roberts et al., 2006) as well as new insights for inclusion into expert systems such as TIMES-SS and DfW. These QMMs use reactivity and hydrophobicity as the key parameters in mathematically modelling skin sensitisation. Whilst hydrophobicity can be conveniently modelled using log P, the octanol-water partition coefficient; reactivity is less readily determined from chemical structure. Initiatives are in progress to generate reactivity data for reactions relevant to skin sensitisation but more resources are required to realise a comprehensive set of reactivity data (Gerberick et al., 2004b, Gerberick et al., 2007, Schultz et al., 2006, Aptula et al., 2006). This data is a necessary requirement to facilitate the derivation of an *in silico* reactivity index. It also serves as a surrogate piece of information to undertake for mechanistic read across evaluations. More details on this approach can be found in Roberts et al., (2007d).

Disclaimer

Any views in this report belong to the authors, and do not necessarily represent formal views or positions of the European Commission or its Joint Research Centre.

References

Accelrys Inc., TopKat User Guide Version 6.02, San Diego: Accelrys Inc., **2004**.

Aptula AO, Patlewicz G, Roberts DW. Skin sensitization: reaction mechanistic applicability domains for structure-activity relationships. *Chemical Research in Toxicology*, **2005a**, 18(9): 1420-1426.

Aptula AO, Roberts DW, Cronin MT. From experiment to theory: molecular orbital parameters to interpret the skin sensitization potential of 5-chloro-2-methylisothiazol-3-one and 2-methylisothiazol-3-one. *Chemical Research in Toxicology*, **2005b**, 18(2): 324-329.

Aptula AO, Patlewicz G, Roberts DW, Schultz TW. Non-enzymatic glutathione reactivity and in vitro toxicity: a non-animal approach to skin sensitization. *Toxicology In Vitro*, **2006**, 20(2): 239-247.

Aptula AO, Roberts DW. Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity. *Chemical Research in Toxicology*, **2006**, 19(8): 1097-1105.

Ashby J, Basketter DA, Paton D, Kimber I. Structure activity relationships in skin sensitisation using the murine local lymph node assay. *Toxicology*, **1995**, 103: 177-194.

Barratt MD, Basketter DA, Chamberlain M, Admans GD, Langowski JJ. Development of an expert system rulebase for identifying contact allergens. *Toxicology In Vitro*, **1994a**, 8(4): 837-839.

Barratt MD, Basketter DA, Chamberlain M, Admans GD, Langowski JJ. An expert system rulebase for identifying contact allergens. *Toxicology In Vitro*, **1994b**, 8(5): 1053-1060.

Barratt MD, Langowski JJ. Validation and subsequent development of the Derek Skin Sensitisation Rulebase by analysis of the BgVV list of contact allergens. *Journal of Chemical Information and Computer Science*, **1999**, 39: 294-298.

Bashir SJ; Maibach HI. Quantitative structure analysis relationships in the prediction of skin sensitization potential. Editor(s): Kydonieus, Agis F.; Wille, John J. *Biochemical Modulation of Skin Reactions*. **2000**, pp 61-64. Publisher: CRC Press LLC, Boca Raton.

Basketter DA, Roberts DW, Cronin M, Scholes EW. The value of the local lymph node assay in quantitative structure-activity investigations. *Contact Dermatitis*, **1992**, 27: 137-142.

Basketter D, Dooms-Goossens A, Karlberg AT, Lepoittevin JP. The chemistry of contact allergy: why is a molecule allergenic? *Contact Dermatitis*, **1995**, 32(2): 65-73.

Basketter DA, Gerberick GF, Kimber I, Loveless SE. The local lymph node assay: a viable alternative to currently accepted skin sensitization tests. *Food and Chemical Toxicology*, **1996**, 34(10): 985-997.

Basketter DA, Lea LJ, Cooper K, Stocks J, Dickens A, Pate I, Dearman RJ, Kimber I. Threshold for classification as a skin sensitizer in the local lymph node assay: a statistical evaluation. *Food and Chemical Toxicology*, **1999**, 37(12): 1167-1174.

Basketter DA, Blaikie L, Dearman RJ, Kimber I, Ryan CA, Gerberick GF, Harvey P, Evans P, White IR, Rycroft RJ. Use of the local lymph node assay for the estimation of relative contact allergenic potency. *Contact Dermatitis*, **2000**, 42: 344-348.

Basketter DA, Evans P, Fielder RJ, Gerberick GF, Dearman RJ, Kimber I. 2002. Local lymph node assay - validation, conduct and use in practice. *Food and Chemical Toxicology*, **2002**, 40(5): 593-598.

Basketter DA, Smith Pease CK, Patlewicz GY. Contact allergy: the local lymph node assay for the prediction of hazard and risk. *Clinical and Experimental Dermatology*, **2003**, 28(2): 218-221.

Basketter D, Casati S, Gerberick GF, Griem P, Philips B, Worth A. Skin sensitisation. *Alternatives to Laboratory Animals*, **2005**, 33(Suppl 1): 83-103.

Benables Q, Smith M, **2003**. The R Development Core Team, R Manuals.

Benerza C, Sigman CC, Perry LR, Tucker Helmes C, Maibach HI. A systematic search for structure-activity relationships of skin contact sensitizers: methodology. *The Journal of Investigative Dermatology*, **1985**, 85: 351-356.

Bouwman T, Bessems JGM, Cronin MTD, Van de Sandt HJM. Evaluation of published QSARs for percutaneous penetration. *Toxicology Letters*, **2006**, 164(Suppl 1): S322.

Burden FR. Molecular identification number for substructure searches. *Journal of Chemical Information and Computer Science*, **1989**, 29: 225-227

Burden FR. A Chemically Intuitive Molecular Index Based on the Eigenvalues of a Modified Adjacency Matrix. *Quantitative Structure-Activity Relationships*, **1997**, 16(4): 309-314

Commission of the European Communities. Council Directive 67/548/EEC of 27 June 1967 on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances.

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.

Cremer D; Hausen BM; Schmalle HW. Toward a rationalization of the sensitizing potency of substituted p-benzoquinones. Reaction of nucleophiles with p-benzoquinones. *Journal of Medicinal Chemistry*, **1987**, 30(9): 1678-1681.

Cronin MT, Basketter DA. Multivariate QSAR analysis of a skin sensitisation database. *SAR and QSAR in Environmental Research*, **1994**, 2(3): 159-179.

Dearden JC, Bradburne SJA, Cronin MTD, Solanki P. The physical significance of molecular connectivity. In: *QSAR 88. Proceedings of the Third International Workshop on Quantitative Structure-Activity Relationships in Environmental Toxicology* (JE Turner, MW England, TW Schultz, NJ Kwaak Eds.) USDOE, Oak Ridge, TN. **1988**, pp 43-50.

Dearden JC, Barratt MD, Benigni R, Bristol DW, Combes RD, Cronin MTD, Judson PN, Payne MP, Richard AM, Tichy M, Worth AP, Yourick JJ. The Development and Validation of Expert Systems for Predicting Toxicity: The Report and Recommendations of ECVAM Workshop 24. *Alternatives to Laboratory Animals*, **1997**, 25: 223-252.

Devillers J. A Neural Network SAR Model for Allergic Contact Dermatitis *Toxicology Methods*, **2000**, 10(3): 181-193.

Devillers J. Strengths and weaknesses of the backpropagation neural network in QSAR and QSPR studies. In: *Neural Networks in QSAR and Drug Design*, ed J Devillers, **1996**, pp 1-46. London: Academic press.

Dimitrov SD, Low LK, Patlewicz GY, Kern PS, Dimitrova GD, Comber MHI, Phillips RD, Niemela J, Bailey PT, Mekenyan OG. Skin sensitization: Modeling Based on Skin Metabolism Simulation and Formation of Protein Conjugates. *International Journal of Toxicology*, **2005a**, 24: 189-204.

Dimitrov S, Dimitrova G, Pavlov T, Dimitrova N, Patlewicz G, Niemela J, Mekenyan O. A stepwise approach for defining the applicability domain of SAR and QSAR models. *Journal of Chemical Information and Modeling*, **2005b**, 45(4): 839-849.

Dupuis G and Benezra C. *Allergic Contact Dermatitis to Simple Chemicals: A Molecular Approach*. Dekker, New York and Basel, **1982**.

Enslein K, Gombar VK, Blake BW, Maibach HI, Hostynek JJ, Sigman CC, Bagheri D.. A quantitative structure-toxicity relationships model for the dermal sensitization guinea pig maximization assay. *Food and Chemical Toxicology*, **1997**, 35: 1091-1098.

Enslein K, Gombar VK, Blake BW. Use of SAR in computer assisted prediction of carcinogenicity and mutagenicity of chemicals by the TOPKAT program. *Mutation Research*, **1994**, 305: 47-61.

Estrada E, Patlewicz G, Chamberlain M, Basketter D, Larbey S. Computer-aided knowledge generation for understanding skin sensitisation mechanisms: The TOPS-MODE approach. *Chemical Research in Toxicology*, **2003**, 16: 1226-1235.

Estrada E, Patlewicz G, Gutierrez Y. From knowledge generation to knowledge archive. A general strategy using TOPS-MODE with DEREK to formulate new alerts for skin sensitization. *Journal of Chemical Information and Computer Science*, **2004**, 44(2): 688-698.

Fedorowicz A, Zheng L, Singh H, Demchuk E. Structure-Activity Models for Contact Sensitisation. *International Journal of Molecular Science*, **2004**, 5: 56-66.

Fedorowicz A, Singh H, Soderholm S, Demchuk E. Structure-activity models for contact sensitisation. *Chemical Research in Toxicology*, **2005**, 18: 954-969.

Gealy R, Graham C, Sussman NB, Macina OT, Rosenkranz HS, Karol MH. Evaluating clinical case report data for SAR modelling of allergic contact dermatitis. *Human Experimental Toxicology*, **1996**, 15(6): 489-493.

Gerberick GF, Ryan CA, Kimber I, Dearman RJ, Lea LJ, Basketter DA. Local lymph node assay: validation assessment for regulatory purposes. *American Journal of Contact Dermatitis*, **2000**, 11(1): 3-18.

Gerberick GF, Ryan CA, Kern PS, Dearman RJ, Kimber I, Patlewicz GY, Basketter DA. A chemical dataset for evaluation of alternative approaches to skin-sensitization testing. *Contact Dermatitis*, **2004a**, 50(5): 274-288.

Gerberick GF, Vassallo JD, Bailey RE, Chaney JG, Morrall SW, Lepoittevin JP. Development of a peptide reactivity assay for screening contact allergens. *Toxicological Sciences*, **2004b**, 81: 332-343.

Gerberick GF, Ryan CA, Kern PS, Schlatter H, Dearman RJ, Kimber I, Patlewicz GY, Basketter DA. Compilation of historical local lymph node data for evaluation of skin sensitisation alternative methods. *Dermatitis*, **2005**, 16(4): 157-202.

Gerberick GF, Vassallo JD, Foertsch LM, Price BB, Chaney JG, Lepoittevin JP. Quantification of chemical Peptide reactivity for screening contact allergens: a classification tree model approach. Toxicological Sciences, **2007**, 97(2): 417-427.

Gerner I, Barratt MD, Zinke S, Schlegel K, Schlede E. 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. *Alternatives to Laboratory Animals*, **2004**, 32(5): 487-509.

Gombar VK, Enslein K. Use of predictive toxicology in the design of new chemicals. In: *Computer aided Molecular Design*. Edited by CH Reynolds, MK Holloway and HK Cox. **1995**, pp 236-249. American Chemical Society, Washington, DC.

Gombar VK, Jain DVS. Quantification of molecular shape and its correlation with physicochemical properties. *Indian Journal of Chemistry*, **1987**, 26A: 554-555.

Graham C, Gealy R, Macina OT, Karl MH, Rosenkranz HS. QSAR for allergic contact dermatitis. *QSAR and Combinatorial Science*, **1996**, 15(3): 224-229.

Hall HL, Mohney M, Kier LB. The electrotoplogical state: structure information at the atomic level for molecular graphs. *Journal of Chemical Information and Computer Science*, **1991**, 21: 76-82

Haneke KE, Tice RR, Carson BL, Margolin BH, Stokes WS. ICCVAM evaluation of the murine local lymph node assay. Data analyses completed by the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods. *Regulatory Toxicology and Pharmacology*, **2001**, 34(3): 274-286.

Hansch C, Leo A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*, Wiley, New York, **1979**.

Hausen, BM.; Schmalle HW.Structure - activity aspects of 4 allergenic sesquiterpene lactones lacking the exocyclic α -methylene at the lactone ring. *Contact Dermatitis*, **1985**, 13(5): 329-332.

Hausen BM; Berger M. The sensitizing capacity of coumarins (III). *Contact Dermatitis*, **1989**, 21(3): 141-7.

Hausen BM, Beyer W. The sensitizing capacity of the antioxidants propyl, octyl, and dodecyl gallate and some related gallic acid esters. *Contact Dermatitis*, **1992**, 26(4): 253-8.

Hausen BM, Hitsch H, Borrmann B, Koch D, Rathmann R, Richter B, Konig WA. Structure-activity relationships in allergic contact dermatitis. (I). Studies on the influence of side-chain length with derivatives of primin. *Contact Dermatitis*, **1995**, 33: 12-16.

Hausen BM, Elsasser, Krohn K, Loock U. Structure-activity relationships in allergic contact dermatitis. Part III. The Sensitising capacity of substituted phenanthrenequinones: A Quantum-Mechanical Approach. *American Journal of Contact Dermatitis*, **2003**, 14(2): 82-89.

Hostynek JJ, Magee PS. Fragrance Allergens: Classification and Ranking by QSAR. *Toxicology in Vitro*, **1997**, 11: 377-384.

Jaworska JS, Comber M, Auer C, van Leeuwen CJ. Summary of a workshop on regulatory acceptance of (Q)SARs for human health and environmental endpoints. *Environmental Health Perspectives*, **2003**, 22: 1358-1360.

Jaworska J, Nikolova-Jeliazkova N, Aldenberg T. QSAR applicabilty domain estimation by projection of the training set descriptor space: a review. *Alternatives to Laboratory Animals*, **2005**, 33(5): 445-459.

Jowsey IR, Basketter DA, Westmoreland C, Kimber I. A future approach to measuring relative skin sensitising potency: a proposal. *Journal of Applied Toxicology*, **2006**, 26(4): 341-350.

Johnson R, Macina OT, Graham C, Rosenkranz HS, Cass GR, Karol MH. Prioritising testing of organic compounds detected as gas phase air pollutants: structure-activity study for human contact allergens. *Environmental Health Perspectives*, **1997**, 105(9): 986 – 992.

Karol MH; Graham C; Barratt MD. Structure - activity relationship studies of hypersensitivity. *Toxicology*, **1999**, 7(1): 59-77.

Kier LB Shape indexes of order one and three from molecular graphs. *Quantitative Structure-Activity Relationships*, **1986**, 5: 1-7.

Klopman G. Artificial intelligence approach to structure-activity studies. Computer automated structure evaluation of biological activity of organic molecules. *Journal of the American Chemical Society*, **1984**, 106: 7315-7321.

Klopman G. MULTICASE 1. A hierarchical computer automated structure evaluated program. *Quantitative Structure-Activity Relationships*, **1992**, 11: 176-184.

Kimber I, Dearman RJ, Basketter DA, Ryan CA, Gerberick GF. The local lymph node assay: past, present and future. *Contact Dermatitis*, **2002**, 47: 315-328.

Kimber I, Dearman RJ. What makes a chemical an allergen? *Annals of Allergy, Asthma and Immunology*, **2003**, 90(Suppl): 28-31.

Kimber I, Basketter DA, Butler M, Gamer A, Garrigue JL, Gerberick GF, Newsome C, Steiling W, Vohr HW. Classification of contact allergens according to potency: proposals. *Food and Chemical Toxicology*, **2003**, 41(12): 1799-1809.

Landsteiner K, Jacobs JL. Studies on the sensitisation of animals with simple chemicals III. *Journal of Experimental Medicine*, **1936**, 64: 625-639.

Langton K, Patlewicz GY, Long A, Marchant CA, Basketter DA. Structure-activity relationships for skin sensitization: recent improvements to Derek for Windows. *Contact Dermatitis*, **2006**, 55(6): 342-347.

Li S, Fedorowicz A, Singh H, Soderholm SC. Application of the Random Forest method in Studies of Local Lymph Node Assay Based skin sensitisation data. *Journal of Chemical Information and Modeling*, **2005**, 45: 952-964.

Li Y, Tseng YJ, Pan D, Liu J, Kern PS, Gerberick GF, Hopfinger AJ. 4D-Figerprint Categorical QSAR Models for skin sensitisation based on the classification of Local Lymph Node assay measures. *Chemical Research in Toxicology*, **2007**, 20(1): 114-128.

Liu H, Papa E, Gramatica P. QSAR prediction of estrogen activity for a large set of diverse chemicals under the guidance of OECD principles. *Chemical Research in Toxicology*, **2006**, 19(11): 1540-1548.

Magee PS, Hostynek JJ, Maibach MI. A classification model for allergic contact dermatitis. *Quantitative Structure-Activity Relationships*, **1994**, 13: 22-33.

Magee PS Exploring the Potential for Allergic Contact Dermatitis *via* Computed Heats of Reaction of Haptens with Protein End-groups Heats of Reaction of Haptens with Protein End-groups by Computation. *Quantitative Structure-Activity Relationships*, **2000**, 19(4): 356-365.

Mekenyan O, Roberts DW, Karcher W. Molecular orbital parameters as predictors of skin sensitization potential of halo- and pseudohalobenzenes acting as SNAr electrophiles. *Chemical Research in Toxicology*, **1997**, 10(9): 994-1000.

Miller MD, Yourtee DM, Glaros AG, Chappelow CC, Eick JD, Holder AJ. Quantum Mechanical Structure-Activity Relationship Analyses for Skin Sensitisation. *Journal of Chemical Information and Modeling*, **2005**, 45: 924-929.

Netzeva TI, Worth AP Aldenberg T, Benigni R, Cronin MTD, Gramatica P, Jaworska JS, Kahn S, Klopman G, Marchant CA, Myatt G, Nikolova-Jeliazkova N, Patlewicz GY, Perkins R, Roberts DW, Schultz TW, Stanton DT, van de Sandt JJM, Tong W, Veith G, Yang C. Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure-Activity Relationships. The Report and Recommendations of ECVAM Workshop 52. *Alternatives to Laboratory Animals*, **2005**, 33: 155-173.

Nikolova-Jeliazkova N, Jaworska J. An approach to determining applicability domains for QSAR group contribution models: an analysis of SRC KOWWIN. *Alternatives to Laboratory Animals*, **2005**, 33(5): 461-470.

OECD. **1992**. Organisation for Economic Cooperation and Development. Guidelines for Testing of Chemicals No. 406. Skin sensitisation.

OECD. **2002**. Organisation for Economic Cooperation and Development. 2002. 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, 17pp. http://www.oecd.org/document/23/0,2340,en_2649_34365_33957015_1_1_1_1,00.html

OECD. **2004**. ENV/JM/MONO/(2004)24 http://appli1.oecd.org/olis/2004doc.nsf/linkto/env-jm-mono(2004)24

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, Dimitrov SD, Low LK, Kern PS, Dimitrova GD, Comber MIH, Aptula AO, Phillips RD, Niemela J, Madsen C, Wedebye EB, Roberts DW, Bailey PT, Mekenyan OG TIMES-SS – A promising

tool for the assessment of skin sensitization hazard. A characterisation with respect to the OECD Validation Principles for (Q)SARs. *Regulatory Toxicology and Pharmacology*, **2007a**, 48(2): 225-239.

Patlewicz G, Aptula AO, Uriate E, Roberts DW, Kern PS, Gerberick GF, Ryan CA, Kimber I, Dearman R, Basketter DA. An Evaluation of selected Global (Q)SARs/Expert systems for the prediction of skin sensitisation potential. *SAR and QSAR in Environmental Research*, **2007b**, 18(5): 515-541.

Patlewicz G, Basketter DA, Smith CK, Hotchkiss SA, Roberts DW. Skin-sensitization structure-activity relationships for aldehydes. *Contact Dermatitis*, **2001**, 44(6): 331-336.

Patlewicz GY, Wright ZM, Basketter DA, Pease CK, Lepoittevin JP, Arnau EG. Structure-activity relationships for selected fragrance allergens. *Contact Dermatitis*, **2002**, 47(4): 219-226.

Patlewicz G, Roberts DW, Walker JD. QSARs for the skin sensitisation potential of aldehydes and related compounds. *QSAR and Combinatorial Science*, **2003**, 22: 196-203.

Patlewicz GY, Basketter DA, Pease CK, Wilson K, Wright ZM, Roberts DW, Bernard G, Arnau EG, Lepoittevin JP. Further evaluation of quantitative structure--activity relationship models for the prediction of the skin sensitization potency of selected fragrance allergens. *Contact Dermatitis*, **2004**, 50(2): 91-97.

Pavan M, Netzeva TI, Worth AP. Validation of a QSAR model for acute toxicity. *SAR and QSAR in Environmental Research*, **2006**, 17(2): 147-171.

Payne MP, Walsh PT. Structure-activity relationships for skin sensitisation potential: development of structural alerts for use in knowledge-based toxicity prediction systems. *Journal of Chemical Information and Computer Science*, **1994**, 34: 154-161.

Perrin DD, Dempsey B, Serjeant EP. *pKa Prediction for Organic Acids and Bases*. London: Chapman and Hall, **1981**, pp 109-126.

Rao KS, Betso JE, Olson KJ. A collection of guinea pig sensitisation test results – grouped by chemical class. *Drug and Chemical Toxicology*, **1981**, 4(4): 331-351.

Ren Y, Liu H, Xue C, Yao X, Liu M, Fan B. Classification study of skin sensitizers based on support vector machine and linear discriminant analysis. *Analytica Chimica Acta*, **2006**, 572(2): 272-282.

Rindone WP, Kush T (eds): PROPHET Molecules – A User's Guide to the Molecule Facilities of the PROPHET System. Prepared by Bolt, Beranek and Newman Inc. under NIH Contract N01-RR-8-2118, **1980**.

Rosenkranz HS, Klopman G. The application of structureal concepts to the prediction of the carcinogenicity of therapeutical agents. In: *Burger's Medicinal Chemistry and Drug Discovery*, 5th Edition, Volume 1 Wolff ME, (ed.) New York: John Wiley & Sons, Inc., **1995**; pp. 223-249.

Roberts DW, Williams DL. The derivation of quantitative correlations between skin sensitisation and physico–chemical parameters for alkylating agents and their application to experimental data for sultones. *Journal of Theoretical Biology*, **1982**, 99: 807-825.

Roberts DW. Structure-activity relationships for skin sensitisation potential of diacrylates and dimethacrylates. *Contact Dermatitis*, **1987**, 17: 281-289.

Roberts DW, Basketter DA. A quantitative structure activity/dose relationship for contact allergenic potential of alkyl group transfer agents. *Contact Dermatitis*, **1990**, 23: 331-335.

Roberts DW, Benezra C. Quantitative structure-activity relationships for skin sensitization potential of urushiol analogues. *Contact Dermatitis*, **1993**, 29(2): 78-83.

Roberts DW. Linear free energy relationships for reactions of electrophilic halo- and pseudohalobenzenes, and their application in prediction of skin sensitization potential for SNAr electrophiles *Chemical Research in Toxicology*, **1995**, 8(4): 545-551.

Roberts DW, Basketter DA. Further evaluation of the quantitative structure activity/dose relationship for contact allergenic potential of alkyl group transfer agents. *Contact Dermatitis*, **1997**, 37: 107-112.

Roberts DW, York M, Basketter DA. Structure-activity relationships in the murine local lymph node assay for skin sensitization: alpha,beta-diketones. *Contact Dermatitis*, **1999**, 41(1):14-17.

Roberts DW, Basketter DA. Quantitative structure-activity relationships: sulfonate esters in the local lymph node assay. *Contact Dermatitis*, **2000**, 42: 154-161.

Roberts DW, Patlewicz G. Mechanism based structure-activity relationships for skin sensitisation--the carbonyl group domain. *SAR and QSAR in Environmental Research,* **2002**, 13(1): 145-152.

Roberts D W, Aptula A O, Patlewicz G. Mechanistic applicability domains for non-animal based toxicological endpoints. QSAR analysis of the Schiff Base applicability domain for skin sensitization. *Chemical Research in Toxicology*, **2006**, 19(9): 1228-1233.

Roberts D W, Aptula A O, Cronin M T D, Hulzebos E, Patlewicz G. Global (Q)SARs for skin sensitization – assessment against OECD principles, *SAR and QSAR in Environmental Research,* **2007a**, 18(3): 343-365.

Roberts D.W, Aptula A.O, Patlewicz G. 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. *Chemical Research in Toxicology*, **2007b**, 20(1) 44-60.

Roberts DW, Patlewicz G, Dimitrov S, Low LK, Aptula AO, Kern PS, Dimitrova GD, Comber MI, Phillips RD, Niemelä J, Madsen C, Wedebye EB, Bailey PT, Mekenyan OG.TIMES-SS – A mechanistic evaluation of an external validation study using reaction chemistry principles. *Chemical Research in Toxicology***, 2007c**, *in press*

Roberts DW, Aptula AO, Patlewicz G, Pease C. Chemical Reactivity Indices and Mechanism based read-across for non-animal based assessment of skin sensitization potential. *Journal of Applied Toxicology*, **2007d**, *in press*

Rodford R, Patlewicz G, Walker JD, Payne MP. Quantitative structure-activity relationships for predicting skin and respiratory sensitization. *Environmental Toxicology and Chemistry*, **2003**, 22(8): 1855-1861.

Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back propagating errors. *Nature*, **1986**, 323: 533-536.

Ryan CA, Hulette BC, Gerberick GF. Approaches for the development of cell-based in vitro methods for contact sensitization. *Toxicol In Vitro*, **2001**, 15(1): 43-55.

Saliner AG, Netzeva TI, Worth AP. Prediction of estrogenicity: validation of a classification model. *SAR QSAR in Environmental Research*, **2006**, 17(2): 195-223.

Sanderson DM, Earnshaw CG. Computer prediction of possible toxic action from chemical structure; the DEREK system. *Human and Experimental Toxicology*, **1991**, 10(4): 261-273.

Schulz KH; Garbe I; Hausen BM; Simatupang MH. The sensitizing capacity of naturally occurring quinones. Experimental studies in guinea pigs. I. Naphthoquinones and related compounds. *Archives of Dermatological Research*, **1977**, 258(1): 41-52.

Schlede E, Aberer W, Fuchs T, Gerner I, Lessmann H, Maurer T, Rossbacher R, Stropp G, Wagner E, Kayser D. Chemical substances and contact allergy--244 substances ranked according to allergenic potency. *Toxicology*, **2003**, 193, 219-259.

Schultz TW, Carlson RE, Cronin MTD, Hermens JLM, Johnson R, O'Brien PJ, Roberts DW, Siraki A, Wallace KB, Veith GD. A conceptual framework for predicting the toxicity of reactive chemicals" modelling soft electrophilicity. *SAR and QSAR in Environmental Research*, **2006**, 17(4): 413-428.

Sigman CC, Bagheri D, Maibach MI: Approaches to structure-activity relationships in skin sensitisation using CADES: in Rougier A, Goldberg AM, Maibach HI (eds): *In vitro skin toxicology: irritation, phototoxicity, sensitisation*. New York, Mary Ann Liebert, **1994**, pp 271-280.

Smith CK, Hotchkiss SAM. *Allergic Contact Dermatitis: Chemical and Metabolic Mechanisms*. **2001**, Taylor & Francis, London.

Smith Pease CK. From xenobiotic chemistry and metabolism to better prediction and risk assessment of skin allergy. *Toxicology*, **2003**, 192: 1-22.

Steiling W, Basketter DA, Berthold K, Bultler M, Garrigue JL, Kimber I, Lea L, Newsome C, Roggeband R, Stropp G, Waterman S, Wiemann C. Skin sensitisation testing – new perspectives and recommendations. *Food and Chemical in Toxicology*, **2001**, 39: 293-301.

Todeschini R, Consonni V. *Handbook of Molecular Descriptors*, **2000**, Wiley-VCH, Weinheim, Germany.

Vracko M, Bandelj V, Barbieri P, Benfenati E, Chaudhry Q, Cronin M, Devillers J, Gallegos A, Gini G, Gramatica P, Helma C, Mazzatorta P, Neagu D, Netzeva T, Pavan M, Patlewicz G, Randic M, Tsakovska I, Worth A. Validation of counter propagation neural network models for predictive toxicology according to the OECD principles: a case study. *SAR and QSAR in Environmental Research*, **2006**, 17(3): 265-284.

Worth AP, Bassan A, Gallegos A, Netzeva TI, Patlewicz G, Pavan M, Tsakovska I, Vracko M. The Characterisation of (Quantitative) Structure-Activity Relationships: Preliminary Guidance. **2005**, EUR 21866 EN.

Zhou Z, Parr RG. Activation hardness: New index for describing the orientation of electrophilic aromatic substitution. *Journal of the American Chemical Society*, **1990**, 112: 5720-5724.

Ziegler V, Richter C, Ziegler B, Haustein U-F: INPRET-Database on Predictive Tests (Allergy). Seminars in Dermatology, **1989**; 8: 80-82.

Zinke S, Gerner I, Schlede E. Evaluation of a rule base for identifying contact allergens by using a regulatory database: Comparison of data on chemicals notified in the European Union with "structural alerts" used in the DEREK expert system. *Alternatives to Laboratory Animals*, **2002**, 30(3): 285-298.

Pag. 3 ----> Pag. 53

European Commission

EUR 22872 EN – DG Joint Research Centre, Institute IHCP
Title: Skin sensitisation (Q)SARs/Expert systems: From past, present to future
Authors: Grace Patlewicz, Aynur O Aptula, David W Roberts, Eugenio Uriarte
Luxembourg: Office for Official Publications of the European Communities
2007 – 53 pp. – 21 x 29.7 cm
EUR - Scientific and Technical Research series; **ISSN 1018-5593**

Abstract
This review describes the state of (Q)SARs/expert systems for skin sensitisation from the early work on structural alerts to the evolvement of QSAR models through to the development of a mechanistically based QSARs. The review considers the merits of the different approaches with particular focus on their applicability for potential regulatory use. There is a strong mechanistic understanding with respect to skin sensitisation which has facilitated the development of a wide range of models. The majority of the existing models fall into one of two main categories – either they are local in nature, typically specific to a chemical class or else they are global in form, derived empirically using statistical methods. Several of the published global QSARs have been recently characterised and evaluated in accordance with the OECD principles. There are also several expert systems capable of predicting skin sensitisation, these are briefly described. Recently, a new perspective regarding the development of mechanistic skin sensitisation QSARs so-called Quantitative Mechanistic Modelling (QMM) was proposed, where reactivity and hydrophobicity, are used as the key parameters in mathematically modelling skin sensitisation. This approach appears to be the most promising means of deriving robust and mechanistically interpretable models that are of value in a regulatory context. Hydrophobicity can be conveniently modelled using log P, the octanol-water partition coefficient. This is readily determined from chemical structure. No such surrogate for reactivity exists, rate constants are best derived from experimental studies. Initiatives are in progress to generate reactivity data for reactions relevant to skin sensitisation but more resources are required to realise a comprehensive set of reactivity data. This is a fundamental and necessary requirement for the future assessment of skin sensitisation. A framework of how information from using RAI approaches can be used in the evaluation of skin sensitisation is described.

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

Publications Office
Publications.eu.int