ILSI-HESI, ECB, JRC-ECB and JRC-ECVAM workshop, 5-6 May 2006

Reporting formats from the QSAR Experience project
Theo Traas and Betty Hakkert, Chemical Substances Bureau RIVM

Introduction

The QSAR experience project is an initiative from regulators in the European Commission to gain experience with the use of QSARs in risk assessment of chemicals. The project is currently under the guidance of the European Chemicals Bureau (ECB) and a subcommittee of the EU technical committee for new and existing substances.

In chemicals risk assessment, large-scale regulatory programs are underway such as the OECD HPVC program, the Canadian DSL program and the new European Chemicals legislation (REACH). It is expected that in the near future, alternatives for in vivo-testing such as *in silico* and *in vitro* methods will be used much more frequently in risk assessment. Both industry (as responsible entities or registrants) and regulators will need to deal with the question how the results of these alternative methods need to be interpreted, how these are reported and how they can be evaluated (and weighted).

As part of the experience project, reporting formats were suggested to exchange experience between regulators on the use and interpretation of QSAR models in risk assessment. It became clear that reporting on the use and outcome of alternative methods but can be placed in a wider context. If the results of alternative methods are not reported consistently, it will be very difficult to evaluate if the methods used are valid for a specific risk assessment context, if they have been applied correctly and if they have been interpreted correctly. Therefore, we feel it is a joint interest for both industry and regulatory bodies to develop a system for reporting alternative methods, such that they can be easily interpreted and evaluated in the risk assessment. This should also be considered when designing a testing strategy.

Considerations on the goal of formats

The goal of the formats is to streamline how alternative methods are reported, and in no way tries to limit or fix which methods are used. The underlying (database of) methods that are described in some detail (see section on levels) can be easily expanded once new methods have been developed.

For industry, it is vital that they can report findings of alternative methods in an accepted format and that the underlying models or methods used are described and stored somewhere to avoid duplication of effort.

For regulators, it is vital that they can see how a certain result is achieved, that results of different methods are reported and that the underlying methods or models can be traced and scrutinized. This is needed to evaluate whether the alternative methods provided in the risk assessment are adequate for the test endpoint in question and provide sufficient certainty for regulatory decision making.

Levels of reporting formats

The current reporting formats have three levels.

- *Level one* – Reporting of end conclusion of using alternative methods, based on the summaries for each method or model, for a specific substance and endpoint (e.g., Substance Y for bioaccumulation).
- *Level two* – Reporting of the prediction and conclusion for a specific substance and endpoint, for a single method or model.
- *Level three* – Description of a specific method or model, based on the OECD criteria.

Level one is the top level reporting format that provides essential information and the conclusions for a specific substance and endpoint. The conclusions from each underlying method or model are repeated so the reasoning and weight of evidence is transparent. As part of this level, a summary of essential substance characteristics (as input to the models or methods) can be given.

Some of the information is dependent on the regulatory framework in question. It can also addresses cut-off criteria, screening criteria, thresholds, classification and labeling issues.

Level two is the reporting level for an individual model or method, for a specific substance and endpoint. The format states the basic

Examples

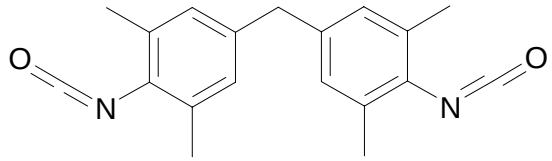
Unfortunately, examples of reporting formats for the endpoint bioaccumulation are not yet available. For illustration purposes, we have provided examples of reporting formats for the endpoint of Skin Irritation.

The example consists of separate parts that are electronically linked (but collated for this example)

- Substance identity (Cas nr. 101657-77-6)
- Level 1 report for skin irritation, purpose of classification and labeling
- Level 2 report for the Gerner model (specific)
- Level 2 report for the DerekfW model (specific)
- Level 3 report for the Gerner model (generic)
- Level 3 report for the DerekfW model (generic)

Hopefully, these examples will stimulate the discussion on how to use the results from alternative methods in risk assessment and allow others to evaluate the results.

Identity

Chemical Name (English)	4,4'-methylenebis(2,6-dimethylphenyl cyanate)
CAS RN	101657-77-6
EINECS/ELINCS-nr.	CAS RN not found in ESIS
SMILES	O=C=Nc1c(C)cc(cc1C)Cc2cc(C)c(c(C)c2)N=C=O
Structure (2D):	
Molecular Weight	306.36 g/mol
Bruto Formula	$C_{19}H_{18}N_2O_2$

Physico-Chemical parameters

Parameter	**Value**	**Unit**	**Source**
Melting point	135 107	°C	(estimate) confidential test
Water Solubility	5.3 6.5	mg/l	(estimate) confidential test
Log Kow	7.4 7.6		(estimate) confidential test
Surface tension	37.8	mN/m	est. Chemsketch 8
Lipid solubility	3.87	??	Confidential test
Hydrolysis	Unknown		
pH in water solubility test	Unknown		

LEVEL 1 EU Classification & Labelling – Skin Irritation

Substance

ITS for substance:	4,4'-methylenebis(2,6-dimethylphenyl cyanate), Identity – Example 2.doc

Endpoint

Regulatory endpoint:	EU Classification and Labelling for dangerous substances and preparations: http://ecb.jrc.it/Legislation/1967L0548EC.htm

Data – QSARs, category approach, in-vivo & in vitro test data

Does the intended use of the chemical give any indication for corrosive properties?	Result	Yes, reactive chemicals – skin corrosion or irritation is likely
	Reliability	2
	Reasoning	No data is available on the use of this substance but isocyanates are known to spontaneously react with water, forming a primary amine (known alert for skin irritancy) and carbondioxide.
Is the pH of the substance indicative of corrosive properties (2>pH>11.5)?	Result	No data available. Skin corrosion not likely
	Reliability	2
	Reasoning	No strongly acidic or basic functionality is present, also not after reaction with water.
Is the substance an organic hydroperoxide?	Result	No – Not corrosive to skin (not R34)
	Reliability	1
	Reasoning	Substance is not an organic hydroperoxide
Is the substance an organic peroxide?	Result	No – Not irritant to skin (not R38)
	Reliability	1
	Reasoning	Substance is not an organic peroxide
Does the substance contain impurities (> 0.1%) that are known skin irritants or corrosives?	Result	No – No classification needed for impurities
	Reliability	1
	Reasoning	
Results of the Gerner exclusion rules for skin irritation:	Level 2:	L2 - Gerner - Example 2.doc
	Result	Not a skin irritant (NOT R38), and not a skin corrosive (NOT R34/35)
	Reliability	1
	Reasoning	The combination of four applicable rules is thought to be give sufficient evidence of the absence of skin irritation potential.
Results of the DEREKfW 8.0 prediction for skin irritation:	Level 2:	L2 - DEREKfW - Example 2.doc
	Result	Skin irritant (mammalian)
	Reliability	1-2
	Reasoning	The isocyanide alert (2X) indicates potential skin irritation. The evaluation of the potential for skin penetration is invalidated by a suspect log K_{ow} estimation. When the experimental value is used, the evaluation would be that skin penetration of the substance is NOT favorable.
Available *in-vitro* data	Result	No data available

	Reliability	
	Reasoning	
Available *in-vivo* data	Result	No data available
	Reliability	
	Reasoning	

Conclusion

Weighted summary of the presented data	Result	Not a skin irritant, NOT R38 or R34/35
	Reliability	1
	Reasoning	pH, chemical class and purity of the substance do not require classification. Physico-chemical properties of the substance indicate absence of skin irritation potential (Gerner rules). The presence of a structural alert (isocyanide, DEREKfW) indicates potential for skin irritation, but this potential is diminished by the phys.chem. properties. DEREKfW also indicates the importance of physico-chemical properties favouring or hindering skin penetration in the interpretation of the validity of the alert. Overall the substance is evaluated as not requiring C&L for skin irritation or skin corrosion.
Need for further testing? > Physico-chemical or related to model input > *In vitro* testing > *In vivo* testing	>. >.	

LEVEL 2 Gerner skin irritation model

MODEL

Model Name	Gerner physico-chemical exclusion rules for skin irritation
Level 3 Description	L3 - GERNER SKIN IRRITATION.doc
Endpoint description (dependent variable)	NOT Classifying for EU C&L as R38 (irritant to skin) and/or R34/R35 (corrosive to skin)
Model Descriptors (independent variables)	Physico-chemical parameters, see Identity – Example 2.doc

DOMAIN

Prediction for substance	4,4'-methylenebis(2,6-dimethylphenyl cyanate), Identity – Example 2.doc	
Model Domain	Chemical:	EU New Substances, no organometallic compounds Purity of the substance should be >95%
	Descriptor:	See Level 3 Description, L3 - GERNER SKIN IRRITATION.doc

PREDICTION

Applicable classes	Class All – organic substances, not salts or metal containing Class CN – compounds only containing C,H,O and N atoms			
Algorithm (rules that apply to this substance)	General algorithm of the exclusion rules: **IF (rule) THEN substance is NOT R38 and/or R34/45**			
	Class	**Rule**	**Result**	**Goodness of fit**
	CN	mol.weight > 290 g/mol	NOT R34/35	338/338
	CN	log K_{ow} > 4.5	NOT R34/35	119/119
	CN	aqueous solubility < 0.1 mg/l	NOT R38	104/104
	CN	log K_{ow} > 5.5	NOT R38	85/85
Remarks				
Structural analogues from training set	Not given – no means available to search the training set for structural analogues.			

CONCLUSION

Result	NOT R38 (irritant to skin) or R34/35 (corrosive to skin)
Reliability (Klimitsch)	1
Reasoning	The aqueous solubility rule for the CN class gave one false negative in the external validation set (borderline substance). However in combination with the three other applicable rules the quality of the prediction is thought to be sufficient. The rules based on molecular weight and log K_{ow} don't have exceptions in the training set, and did not give any false negatives in the external validation set.

LEVEL 2 DEREKfW skin irritation model

MODEL

Model Name	DEREKfW8.0
Level 3 Description	L3 - DEREKfW SKIN IRRITATION.doc
Endpoint description (dependent variable)	Skin Irritation (mammalian). Not necessarily strong enough to lead to classification (alert dependent)

DOMAIN

Prediction for substance	4,4'-methylenebis(2,6-dimethylphenyl cyanate), Identity – Example 2.doc	
Domain	Chemical:	Organic substances that contain at least one alert. The substance is a diisocyanate and thus contains the isocyanate structural alert for skin irritation. The examples show that the alert (isocyanate) can be a substituent of benzylic ring systems. Therefore the chemical is clearly within the domain of the structural alert.

PREDICTION

Algorithm	There is no algorithm, only a qualitative evaluation of structural alerts (leading to skin irritation) and parameters for skin penetration (favouring or hindering the potential skin irritation caused by the structural alert.	**Result**
	Alert identified: R1-N=C=O, R1= carbon atom (2X)	**Irritant to skin (mammals)**
	Parameters calculated for skin uptake evaluation ***Log Kp:*** *-2.036 Calc. by the Potts & Guy equation.* ***Log P:*** *3.596 Calc. by the Moriguchi estimation* ***MW****:. 306.37 g/mol* *Skin penetration is favoured by relatively lipophilic molecules (Log K_{ow} = 1-4) of low molecular weight (<500).*	**Skin penetration favorable for skin irritation**
Remarks	The presence of two isocyanate alerts in one structure strengthens the prediction of skin irritation potential. The estimation of log P (=log K_{ow}) differs strongly from the experimental value and other estimations (ClogP and KOWWIN QSARs).	
Structural analogues	The structural alert is illustrated with 5 analogues. These are however smaller than the submitted chemical. See Annex 1 (DEREKfW result): *Known irritants which fire the alert include:* *Methyl isocyanate* *Ethyl isocyanate* *Phenyl isocyanate* *Toluene diisocyanate*	

CONCLUSION

Result	Skin irritant
Reliability (Klimitsch)	1-2
Reasoning	The presence of an alert for skin irritation (2X) indicates potential skin irritation. The alert is thought to be valid, the substance is well within the structural domain of the alert. The evaluation of the potential for skin penetration is hampered by a suspect log K_{ow} estimation. When the experimental value is used, the evaluation would be that skin penetration of the substance is NOT favorable. The interpretation of the combination of the effect of the structural alert and the influence of skin penetration is left completely to the end user, no definite prediction is given by the algorithm. The quality of the overall prediction is therefore thought to be 1-2 (structural alert 1, skin penetration evaluation 2).

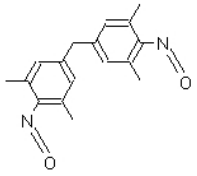
Annex I ***DEREK for Windows report***

Version: *8.0.1*

Species: *human*
mammal
SuperEndpoints: *Irritation*

Compound Name:
Log Kp: *-2.036 Calculated by the Potts & Guy equation*
Log P: *3.596 Calculated by the Moriguchi estimation*
Molecular Weight: *306.365 Calculated by LPS*

Submitted Compound:



List of alerts found:

211 Isocyanate. Irritation (of the skin, eye and respiratory tract). Number of matches = 2

Alert overview: 211 Isocyanate

R1—N=C=O

R1 = C

Known irritants which fire the alert include:
Methyl isocyanate
Ethyl isocyanate
Phenyl isocyanate
Toluene diisocyanate

Isocyanates are highly reactive substances and generally irritating to the skin, eyes and respiratory tract. Hydrolysis and reaction with biologically important molecules, including proteins, occurs rapidly. Irritation to the respiratory tract may occur at low concentrations. E.g. exposure of humans to 2ppm methyl isocyanate for 1-5 minutes produced tears and irritation to the nose and throat. Diisocyanates are generally stronger irritants than monoisocyanates. A polymeric isocyanate, polymethylene polyphenyl isocyanate, has been classified as irritating to the skin, eyes and respiratory tract.

N.B. A structural alert for irritancy indicates some potential for this effect. Additionally, except for highly reactive corrosive substances, the skin and eye irritation potential of a chemical is very dependent on physicochemical properties which influences the concentrations at and exposure to component tissues. Skin penetration is favoured by relatively lipophilic molecules (Log P(octanol/water)= 1-4) of low molecular weight (<500). For many classes of chemicals (e.g. aliphatic amines) eye irritation is greatest for the more water soluble compounds which readily dissolve in the aqueous tear film on the cornea and conjunctiva. Liquid substances (cf.solids) have good tissue contact and are more likely to be irritating, particularly to the skin. Highly reactive corrosive chemicals may penetrate tissue as a result of corrosive damage with a lower dependence on solubility characteristics.

References:

Title: *The Dictionary of Substances and their Effects on CD-ROM.*
Author: *Anonymous.*
Source: *The Dictionary of Substances and their Effects on CD-ROM, SilverPlatter Information, Boston, 1996.*

Title: *Toxicology of the Eye.*
Author: *Grant WM.*
Source: *Toxicology of the Eye, Grant WM, Charles C Thomas, Springfield, 1962.*

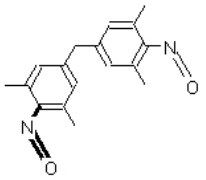
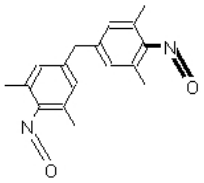
Title: *Cyanides and nitriles.*
Author: *Hartung R.*
Source: *Patty's Industrial Hygiene and Toxicology, 4th edition, volume 2D, Clayton GD and Clayton FE (editors), John Wiley, New York, 1994, 3119-3172.*

Title: *Respiratory effects of inhaled isocyanates.*
Author: *Karol MH.*
Source: *Critical Reviews in Toxicology, 1986, 16, 349-379.*

Title: *Mechanisms of activation of the sensory irritant receptor by airborne chemicals.*
Author: *Nielsen GD.*
Source: *Critical Reviews in Toxicology, 1991, 21, 183-208.*

Title: *Industrial hygiene.*
Author: *Schrenk HH.*
Source: *Industrial and Engineering Chemistry, 1955, 47, 107A-108A.*

Locations:

Examples: (211 Isocyanate)
(No examples)

Custom Examples: (211 Isocyanate)
(No examples)

LEVEL 3 QSAR model of Gerner et al.,

1. QSAR identifier

Literature model (2004/5) and software package DSS (2000), the latter is not evaluated

2. Source

The empirical rulebase model uses physical-chemical cut off values for specific empirical classes, that predicts the absence of skin corrosion or irritation. The model is developed by Gerner and co-workers at BfR in Berlin, Germany and was first reported in 2000 and updated in 2004 (Gerner et al. and Zinke et al.). More information and its potential use in testing strategies are described in (Walker et al, 2005).

2.1 Reference(s) to scientific papers and/or software packages

Gerner, I., Graetschel, G., Kahl, J., Schlede, E. Development of a Decision Support System for the Introduction of Alternative Methods into Local Irritation/Corrosion Testing Strategies: Development of a Relational Data Base. *ATLA* **2000**, 28, 11-28.

Gerner, I., Zinke, S., Graetschel, G., Schlede, E. Development of a Decision Support System for the Introduction of Alternative Methods into Local Irritancy/Corrosivity Testing Strategies. Creation of Fundamental Rules for a Decision Support System. *ATLA* **2000**, 28, 665-698.

Zinke, S., Gerner, I., Graetschel, G., Schlede, E. Local irritation/corrosion testing strategies: Development of a decision support system for the introduction of alternative methods. *ATLA* **1999**, 28, 29-40.

Zinke, S. and Gerner, I. Local irritation/corrosion testing strategies: Extending a decision support system by applying self-learning classifiers. *ATLA* **2000**, 28, 651-663.

Gerner, I., Walker J.D., Hulzebos, E., Schlegel, K., Use of physicochemical property limits to develop rules for identifying chemical substances with no skin irritation or corrosion potential, QSAR Comb. Sci., **23**, 726-733 (2004).

Walker, J.D., Gerner I., Hulzebos, E., Schlegel, K. (Q)SARs for predicting skin irritation and corrosion: Mechanisms, transparency and applicability of predictions, QSAR Comb. Sci., **23**, 721-725 (2004).

Walker, J.D., Gerner, I., Hulzebos, E.T., Schlegel, K. The skin irritation corrosion rules estimation tool (SICRET), QSAR Comb. Sci., **24**, 378-384 **(**2005).

2.2 Date of publication

A number of publications are given though key dates are notably 1999/2000.

2.3 Identification of the model developer(s)/authors

Dr. I. Gerner and co-workers at BfR. Matthias Herzler is the (Q)SAR contact point.

Dr. Matthias Herzler
Bundesinstitut für Risikobewertung (BfR)
Sicherheit von Stoffen und Zubereitungen
Toxikologie der Pestizide
Thielallee 88-92

14195 Berlin
Fon 01888 412 4402
Fax 01888 412 3260
Mail m.herzler@bfr.bund.de

2.4 Contact details of the model developer(s)/authors

The model can be derived from literature data.

2.5 Indication of whether the model is proprietary of non-proprietery

The model that predicts the chemicals is not proprietary, the details of the training set are.

3. Type of model

3.1 ■ 1-D (Q)SARs Empirical formulas

3.2 □ 2-D (Q)SARs

3.3 □ 3-D (Q)SARs

3.4 □ Battery of models

Overall prediction depends on applicability of multiple models/rules

3.5 □ Expert system

Overall prediction depends on application of multiple models/rules and use of data in knowledge base

3.6 ■ Empirical system

3.6 □ Neural network

3.7 □ Other

4. Definition of the model

The rabbit skin irritation test is the bases for the model (OECD404). The outcome of the test into a regulatory application is a two step process. The application of the chemical on the skin can result in erythema and oedema. The severeness and persistency of the effects is reflected in Draize irritation scores, that need to be reported in prescribed time intervals 1h, 24, 48 and 72h up to 21 days when effects are persisting. In the second step the scores are categorised using certain cut offs of the Draize scores, including persistency, for regulatory decision making in EU. The three categories are non-irritant, irritant, or corrosive. The classification and labelling of chemicals is used for risk reduction measures for workers and consumers that are exposed to these chemicals.

The endpoint that the model predicts is not the outcome of the skin irritation test, the effects reported as Draize scores, but it predicts the categorisation of the chemical. The model can therefore be directly used for regulatory EU classification and labelling purposes.

4.1 Defined endpoint

4.1.1. Species:

The relevant test guideline determines the species being modelled though is typically a rabbit.

4.1.2 Endpoint:

The endpoint is EU classification and labelling for skin irritation.

4.1.3 Units of measurement:

The unit of measurement has to be interpreted as the chemical is corrosive, irritant and non-irritant.

4.1.4 Reference to a specific protocol:

The reference to the experimental protocol is OECD 404

4.2 Number of descriptors used as independent variables

Six, see below

4.3 Identification of descriptors (names, symbols)

Molecular weight (g/Mol)
Log Kow
Aqueous solution (a.s in g/l.)
Surface tension (s.t. in mN/m)
Lipid solubility (l.s. in g/kg)
Vapour pressure (v.p. in Pa)

4.4 Explicit algorithm for generating prediction from descriptors

4.5

The algorithm of the model is described as physical chemical cut-off values for specific empirical chemical classes above or below, which the absence of corrosive or irritation classification is predicted. Empirical classes are described as C or Chal, meaning that chemicals only contain C, H and O atoms, or only C, H, O and halogen atoms. For example, a physical chemical cut-off value is that C chemicals with a log Kow of < -3.1 will not be irritants or corrosives.

The model can be used to predict the absence of skin irritation classification of organic chemicals without any statistical methodology. All chemicals in the database that are classified for skin irritation are excluded from the rules.

Three prerequisites are stated. The pH of the aqueous solution of the chemical should be outside the corrosive boundaries meaning that the pH of the chemicals should not be above 11 or below 1.5, which already implies classification as a corrosive (OECD, 404). The chemical predicted should have at least a purity of 95%, as irritant or corrosive impurities might cause false negative predictions. When there are other reasons to assume high reactivity the rules (e.g. oxidisers) might give false negatives.

Though the physical limit values are empirically derived, the mechanism underlining these limit values is that most organic chemicals first have to penetrate the skin before being reactive is discussed in Walker et al. (2004).

Table 1. Chemical groups, physicochemical properties, number of chemicals in each group that were used to develop rules to identify chemicals with no skin irritation or skin corrosion potential (From Walker et al., 2005)

Chemical Group	Physicochemical property	# chem passed/ # chem tested	No Skin Irritation (I) or Corrosion (C)
All chemicals	melting point > 200°C	291/297*	No I or C
All chemicals	log P_{ow} or log K_{ow} < -3.1	56/56	No I or C
All chemicals	lipid solubility < 0.01 g/kg	60/60	No C
Group C ($C_xH_yO_z$)	melting point > 55°C	128/130*	No I or C
Group C ($C_xH_yO_z$)	molecul.weight > 350 g/Mol	93/93	No C
Group C ($C_xH_yO_z$)	surface tension > 62 mN/m	94/95**	No C
Group C ($C_xH_yO_z$)	vapour pressure < 0.0001 Pa	73/73	No I
Group CN ($C_xH_yO_zN_a$)	lipid solubility < 0.4 g/kg	56/56	No I or C
Group CN ($C_xH_yO_zN_a$)	molecul.weight > 290 g/Mol	338/338	No C
Group CN ($C_xH_yO_zN_a$)	aqueous solubility < 0.1 g/l	280/280	No C
Group CN ($C_xH_yO_zN_a$)	log P_{ow} or log K_{ow} > 4.5	119/119	No C
Group CN ($C_xH_yO_zN_a$)	vapour pressure < 0.001 Pa	273/273	No C
Group CN ($C_xH_yO_zN_a$)	molecul.weight > 540 g/Mol	86/86	No I
Group CN ($C_xH_yO_zN_a$)	melting point > 180°C	153/153	No I
Group CN ($C_xH_yO_zN_a$)	aqueous solubil. < 0.0001 g/l	104/104	No I
Group CN ($C_xH_yO_zN_a$)	log P_{ow} or log K_{ow} > 5.5	85/85	No I
Group CNHal ($C_xH_yO_zN_aF$,Cl,Br or I)	log P_{ow} or log K_{ow} > 3.8	70/70	No I or C
Group CNHal ($C_xH_yO_zN_aF$,Cl,Br or I)	aqueous solubility < 0.1 g/l	135/135	No C
Group CNHal ($C_xH_yO_zN_aF$,Cl,Br or I)	molecul.weight > 370 g/Mol	109/109	No C
Group CNHal ($C_xH_yO_zN_aF$,Cl,Br or I)	lipid solubil. < 400 g/kg	76/76	No C
Group CNHal ($C_xH_yO_zN_aF$,Cl,Br or I)	molecul.weight > 380 g/Mol	99/99	No I
Group CNHal ($C_xH_yO_zN_aF$,Cl,Br or I)	lipid solubil. < 4 g/kg	29/29	No I
Group CNHal ($C_xH_yO_zN_aF$,Cl,Br or I)	aqueous solubil. < 0.001 g/l	78/78	No I
Group CNS ($C_xH_yO_zN_aS_b$)	molecul.weight > 620 g/Mol	53/53	No C
Group CNS ($C_xH_yO_zN_aS_b$)	melting point > 50°C	179/180*	No C
Group CNS ($C_xH_yO_zN_aS_b$)	surface tension > 62 mN/m	92/92	No C
Group CNS ($C_xH_yO_zN_aS_b$)	melting point > 120°C	137/137	No I
Group CNS ($C_xH_yO_zN_aS_b$)	log P_{ow} or log K_{ow} < 0.5	96/96	No I
Group CHal ($C_xH_yO_zF$,Cl,Br or I)	molecul.weight > 370 g/Mol	24/24	No I or C
Group CHal ($C_xH_yO_zF$,Cl,Br or I)	molecul.weight > 280 g/Mol	59/59	No C

*chemicals that did not pass were organic salts which release strong inorganic acids or bases when in contact with aqueous substrates/organic media

**chemical that did not pass was a skin de-fatting ether with high vapour pressure at 20°C

Table 2: Additional rules for skin irritation/corrosion
(By Ingrid Gerner and Matthias Herzler not mentioned in the Gerner et al., 2004 and Walker et al., 2005 publications)

GROUP	IF				THEN NOT	REF.
	Parameter	Qualifier	Value	Unit		
All	log P_{OW} or log K_{OW}	>	9		R34, R35	[2]
C	a.s.	<	0.0001	g/mol	R34 or R35	[2]
CHal	log P_{OW} or log K_{OW}	>	4.5		R34 or R35	[2]
CHal	m.p.	>	65	°C	R34 or R35	[2]

[2] Gerner I, Herzler M. (2004) submitted to ECVAM on July 12, 2004

4.6 Goodness-of-fit statistics

The third column in table 1 shows the goodness-of-fit.

4.6 Information on the applicability domain of the model

4.6.1 Are full details of the training set given, including details of chemical names, structural formulae, CAS numbers (if available), and data for all descriptor and response variables.

Chemical names, structural formulae and CAS numbers are only available to the Competent Authorities of the EU member states. The German BfR has put these data in a database and data are confidential. However, the excel file containing the empirical formulas and outcome of the tests are not confidential and could be made available.

The information on descriptor and response values is available in an excel file and could be made publicly available as no confidential data are included. However this excel file is not yet made publicly available.

4.6.2 If the data used to develop the model were based on the processing of raw data (e.g., the averaging of replicate values)

For each chemical(one notification) one test was performed. No averaging of replicate values has been done.

4.6.3 Is there an adequate description of the data processing?

The data processing is adequately described.

4.6.4 Are the raw data provided?

The raw data are available and provided to the evaluator for the purpose of external validation by Rorije and Hulzebos (2005).

4.6.5 Does application of the appropriate statistical method(s) to the training set result in the same (Q)SAR model?

The results of the validation (Rorije and Hulzebos, 2005) show that application of the same method results in the same model.

The following remarks should be included:

Melting point and Vapour Pressure have their cut-off values set non-conservative. All rules based on **melting point** or vapour pressure have exceptions; sometimes a substantial part (44%) of the irritant/corrosive substances is not covered by the chosen cut-off value. It is suggested that the melting point rules are either removed, or that the cut-off values are set at more conservative values e.g., the values covering 100 percent or 100 percent. The rules using **vapour pressure** cannot be redefined using conservative cut-off values since these would in effect make the rules non-applicable to any substance (a cut-off value of 0 Pa would be needed). It is suggested that vapour pressure will be dismissed as a parameter to base exclusion rules for skin irritancy on.

Surface Tension The two exclusion rules based on Surface Tension have not been evaluated because of the limited applicability of the exclusion rules. These rules

apply only to 10 / 201 substances in the validation set, and only 2 of these 10 substances were not covered by any other rules.

4.6.6 Is there a specification of the statistical method(s)used to develop the QSAR (including details of any software packages used)?

There is a specification of the method used. This is however not a statistical method. It is a visual/graphical method that shows at which descriptor value no classification is noticed. No algorithm to determine the cut-off values for specific parameters has been used.

4.7 External validation/Predictivity

4.7.1 An indication whether the model has been validated by using a test set that is independent of the training set?

This has been done twice. First time the external validation is described in the publication of Zinke et al. [Zinke 1999] on the set of rules described in the same paper. The second external validation is presented in the present report by (Rorije and Hulzebos, 2005)

4.7.2.1 If an external validation has been performed, is the following information available Zinke et al:

In the first validation exercise the rule base, including the use of structural alerts was tested with 331 substances not used for the training the model, which contained 1000 chemicals (Zinke et al., 1999, tables VII and VIII). For skin corrosion a validation was carried out. For skin irritation no such validation was presented.

a) number of test structures;

   282 (already excluded the skin irritants (16) and the chemicals for which no experimental data was available (33)

b) the identity of the test structures;

c) the specific identity of the chemicals is not publicly available. More details are known to the CA's of the EU member states;

d) the approach for selecting the test structures;

   The next 331 chemicals submitted after deriving the rules were used;

e) the statistical analysis of the predictive performance of the model? (e.g., including sensitivity, specificity, and positive and negative predictivities for classification models);

   As the model only predicts the absence of effects, the prediction performance can only be given as specificity and false negatives. The specificity is expressed as the number chemicals that are correctly predicted as not classified divided by the number of chemicals that are negative based on the experimental test. False negative is the fraction of chemicals that showed to be irritating/corrosive, while the absence of skin irritant effects was predicted by the model.

f) the results of the prediction?
   The results were that the specificity was 63.2% (163/258), was correctly predicted not corrosive. The percentage false negatives was 4.2 % (1/24).

4.7.2.1 In the second validation exercise reported in the present report the following information was available:

a) **number** of test structures was available: 201.

b) **the identity**: this was only known as empirical formulas in the excel datasheet. More details are known to the CA's of the EU member states;

c) the approach for selecting the test structures;

the approach is known: the next 201 chemicals submitted after deriving the rules were used;

Definition of the applicability domain. The distribution of the test set among the empirical classes and descriptor values was compared with the training set. It was concluded that the test set was very similar to the training set and the test set can be considered a real external validation set.

e) **T**he statistical analysis of the predictive performance of the model? (e.g. including sensitivity, specificity, and positive and negative predictivities for classification models);

See above, only specificity and false negatives can be derived

f) a comparison of the predictive of the model against previously-defined quantitative performance criteria?
Rorije and Hulzebos, 2005 shows that
If a corrosive or irritant potential based on pH would be applied before applying the physico chemical exclusion (a prerquisite); if the recommended newly calculated cut-off values for melting point would be applied; and if the recommendation to remove the K_{ow} rule for CNS compounds is followed, the statistics for the performance of the exclusion rules on the external validation set improve as shown below:

Incorrect prediction of NOT R34/35	0	0%
Incorrect prediction of NOT R34/35/38	1	0.5%
Correct predictions of NOT R34/35	58	29.1%
Correct predictions of NOT R34/35/38	85	42.7%
No prediction – test result NOT R34/R35/R38	34	17.1%
No prediction – test result R34/R35 or R38	21	10.6%
total	199	100.0%

5 Mechanistic Interpretation, if possible

5.1 In the case of a SAR, is there a description of the molecular events that underlie the reactivity of the molecule (e.g. description of how substructural features could act as nucleophiles or electrophiles, or form part or all of a receptor-binding region)? See 5.2.

5.2 In the case of a QSAR, do the descriptors have a physicochemical interpretation that is consistent with a known mechanism (of biological action)?

The very reactive chemicals are excluded from the model (but included in the testing strategy according to OECD 404., because the model is empirically modelling skin absorption.

5.3 Are any literature references cited in support of the proposed mechanistic basis of the (Q)SAR?

In other related publications literature references are supporting the empirical/mechanistic bases (e.g. Walker et al. 2004) and Hulzebos et al. 2005).

6. Applications of the model

Suggestions for possible applications for the model.

Skin irritation is predicted in terms of EU classification, chemicals are predicted as non-corrosives, non-irritants. The model can be applied to organic chemicals including the prerequisites on high reactivity, pH and purity for accepting negative predictions. Those chemicals that are predicted non-irritants are neither corrosives and need not be classified for skin irritation. The potential mechanism is often reactivity. Example chemicals are provided, which can possibly be used as analogues or categories, including EU classification if known.

7. Miscellaneous information

No additional information

9. References

OECD (Organisation for Economic Cooperation and Development). Guideline for Testing Chemicals No 404, Skin irritation, Paris, (2002). (http://www.oecd.org/dataoecd/45/25/2741642.doc).

Anon, 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, *Official Journal of the European Communities 196*, 16.8.1967, 1-98 (1967).

EC. Commission Directive 2001/59/EC of 6 August 2001 adapting to technical progress for the 28th time Council Directive 67/548/EEC on the approximation of the laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances *Official Journal L 225 , 21/08/2001 P. 0001 – 0333*,Office for Official Publications of the European Communities, Luxembourg, **2001**.

OECD, Harmonised integrated classification system for human health and environmental hazards of chemical substances and mixtures, *http://www.oecd.org/ehs/* **2001**.

Rorije E. and Hulzebos, E. (2005), Evaluation of (Q)SARs for the prediction of skin irritation/corrosion potential. SEC report, publicatie in prep. Available at the ECB website: http://ecb.jrc.it/QSAR/ Documents/Evaluation of skin irritation QSARs

LEVEL 3: DEREK FOR WINDOWS MODEL FOR SKIN IRRITATION

(copied from ECB proposal for sensitization , developed in consultation with LHASA Ltd)

Content

1. QSAR identifier

Derek for Windows skin irritation rulebase. Version No 8.

2. Source

2.1 Reference(s) to scientific papers and/or software package:

- Greene, N., Judson, N.P., Langowski, J.J., Marchant, C.A. (1999). Knowledge-based expert systems for toxicity and metabolism prediction: DEREKfW, StAR and METEOR. *SAR and QSAR in Environmental Research* 10, 299-314.
- Sanderson, D.M., Earnshaw, C.G. (1991). Computer prediction of possible toxic action from chemical structure; The DEREK system. *Human & Experimental Toxicology* 10, 261-273.
- Zinke, S., Gerner, I., Schlede, E. (2002). 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 DEREKFW Expert System. *ATLA* 30, 285-298.
- Greene, N. (2002). Computer systems for the prediction of toxicity: an update. *Advanced Drug Delivery Reviews* 54, 417-431.

2.2 Date of publication:

A number of publications though key dates are notably 1986 when the first DEREK system was created at Schering Agrochemicals in the UK and 1989 when LHASA Ltd adopted the DEREK system and began coordinating the main development of the structure-toxicity knowledge base.

2.3 Identification of the model developer(s)/authors:

Lhasa Limited
LHASA is the acronym for Logic and Heuristics Applied to Synthetic Analysis)

2.4 Contact details of the model developer(s)/authors:

22-23 Blenheim Terrace,
Woodhouse Lane,
Leeds LS2 9HD, UK
Tel: +44 (0)113 394 6020
Fax: +44 (0)113 394 6099
Email: info@lhasalimited.org
Web: www.lhasalimited.org

2.5 Indication of whether the model is proprietary or non-proprietary: Proprietary

3. Type of model

3.1	☐	2D SAR
3.2	☐	3D SAR (e.g. pharmacophore)
3.3	☐	Regression-based QSAR
3.4	☐	3D QSAR
3.3	☐	Battery of (Q)SARs (overall prediction depends on application of multiple models/rules)
3.4	☑	Expert system (overall prediction depends on application of multiple models/rules and use of data in a knowledge base)
3.5	☐	Neural network
3.6		Other

4. Definition of the model

4.1 Dependent variable being modeled:

4.1.1 Species

The relevant test guideline (OECD404) determines the species being modeled though is typically rabbit

4.1.2 Endpoint (including exposure time)

The endpoint is defined as reactivity (acid, bases, oxidisers, reductors, surfactants) and for similar chemicals EU classification for irritation and corrosion is added. The model can be used as an indicator for reactivity and/or a supplier of analogues.

4.1.3 Units of measurement

Qualitative predictions are made which do not incorporate any specific unit of measurement.

4.1.4 Reference to specific experimental protocol(s):

The skin irritation knowledge encoded within Derek includes both public and proprietary data. Information about the experimental conditions is only given in the references associated with a given alert. Since only a subset of these are fully referenced, the quality of the data used in the derivation of an alert cannot be fully verified.

4.2 Number of descriptors used as independent variables:

Not applicable

4.3 Identification of descriptors (names, symbols):

Not applicable

4.4 Explicit algorithm for generating predictions from the descriptors:

DerekfW8.0 provides an explicit description of the substructure and substituents. When a query structure is processed, the alerts that match are displayed in a hierarchy called the prediction tree and are highlighted in bold in the query structure. The prediction tree includes the endpoint, and reasoning outcome, the number and name of the alert, and the example from the knowledge base if it exactly matches the query structure. The alert description provides a description depicting the structural requirement for the toxicophore detected and a reference to show the bibliographic references used. Some rules are extremely general with substructures only taking into account the immediate environment of a functional group. In other cases, the descriptions are much more specific. This means that remote fragments that may modulate skin irritation are not always taken into consideration in the assessment.

DEREKfW contains 25 structural alerts for skin irritation:

These alerts include some examples and the algorithms for the SAR are described including possible attachments.

4.5 Goodness-of-fit statistics

DEREKfW does not provide the full details of the training data used to develop an alert. Only a subset of the references and example chemicals used to develop the alert are provided for illustrative purposes.

4.6 Information on the applicability domain of the model

DEREKfW includes some inclusion/exclusion rules associated with an alert. These are documented in the alert description as particular substituents. For some skin irritation rules there are very clear descriptions of what is covered by a specific substructure. In other cases the rules are extremely general. Physical properties (Log P and MW) are used to limit the domain for skin irritation, by accounting for skin permeability (where dermal absorption is relevant). DEREKfW has limited means of flagging which chemistries are covered in the rulebase and which are not. The program is not suitable for polymers.

4.7 Information on the mechanistic basis/interpretation of the model

All the rules in Derek are based on either hypotheses relating to mechanisms of action of a chemical class or observed empirical relationships, the ideas for which come from a variety of sources, including published data or suggestions from the DEREK collaborative group.

This group consists of toxicologists who represent Lhasa Ltd and members who meet at regular intervals to give advice and guidance on the rule development work and predictions made by the program. The hypotheses underpinning each alert are documented in the alert descriptions as comments. These comments often include descriptions of features acting as electrophiles or nucleophiles. However, the detail depends on the specific alert. Some alerts contain no comments, aside from the modulating factors of skin penetration.

5. Development of the model

5.1 Explanation of the method (approach) used to generate each descriptor

Any information would be found in the comments section of the alert but this is not systemically provided.

5.2 Selection of descriptors

5.2.1 Indication of initial number of descriptors screened

Not applicable

5.2.2 Explanation of the method (approach) used to select the descriptors and develop the model from them

Not applicable

5.2.3 Indication of final number of descriptors included in the model:
Not applicable

5.3 Information on experimental design for data splitting into training and validation sets.

Not applicable

5.4 Availability of the training set

5.5.1	☐	Chemical names (common names and/or IUPAC names)
5.5.2	☐	CAS numbers
5.5.3	☐	1D representation of chemical structure (e.g. SMILES)
5.5.4	☐	2D representation of chemical structure (e.g. ISIS sketch file)
5.5.5	☐	3D representation of chemical structure (e.g. MOL file)
5.5.6	☐	Data for each descriptor variable
5.5.7	☐	Data for the dependent variable

DEREK rules describe generalised structure-activity relationships and do not record internally the specific chemical structures on which they are based. Derek is a knowledge base as opposed to a database. This does mean it is possible to use data from confidential sources as a basis for new rules without revealing exact chemicals to end-users. This provides a means by which proprietary data can be used without revealing potentially sensitive information.

This is a clear advantage for the purposes of securing business confidentially, but reduces the transparency of the system. The training set information available is limited to a few key example compounds to illustrate the scope of the alert.

6. Validation of the model

6.1 Statistics obtained by leave-one-out cross-validation
None

6.2 Statistics obtained by leave-many-out cross-validation
None

6.3 Statistics obtained by Y-scrambling
None

6.4 Statistics obtained by external validation
None

6.5 Definition of the applicability domain of the model

Evaluation exercise was performed by Hulzebos and Posthumus (2005) for DEREKfW 5.0, however the evaluation set of circa 50 chemicals were not detected, as the two alerts for skin irritation of that DEREKfW version were not present in the chemicals

6.6 Availability of the external validation set

6.6.1	☐	Chemical names (common names and/or IUPAC names)
6.6.2	☐	CAS numbers
6.6.3	☐	1D representation of chemical structure (e.g. SMILES)
6.6.4	☐	2D representation of chemical structure (e.g. ISIS sketch file)
6.6.5	☐	3D representation of chemical structure (e.g. MOL file)
6.6.6	☐	Data for each descriptor variable
6.6.7	☐	Data for the dependent variable

Not applicable for DEREKfW 8.0

7. Applications of the model

Suggestions for possible applications for the model.
Skin irritation is predicted as a potential hazard. The potential mechanism is often reactivity. Example chemicals are provided, which can possibly be used as analogues or categories, including EU classification if known.

8. Miscellaneous information

Needed?

- DerekfW is essentially a knowledge archive of structure-toxicity relationships.
- DerekfW is limited in that it identifies only 'activating' fragments, meaning the negative prediction is based solely on the lack of structural alerts. Only qualitative outcomes are provided, no measure of potency is provided. Training sets of chemicals containing these structural alerts are not provided. DerekfW does not provide a comprehensive list of references used in the development of each alert. Insufficient information is provided about the quality of the data used in the development of each alert.
- No clear explanation of the domain of applicability is provided that would alert the user as to when a query structure was within or outside the chemical domain of Derek.

- Some of the alerts within DerekfW are very general, explaining the high number of false positives in the external validation studies.
- DerekfW covers a small subset of chemical space, a huge number of rules would need to be developed in order to account for each chemical class. Development of DerekfW is incremental, focusing on each chemical class in turn. DerekfW would improve from adding more information about the modulating factors in the environment of an alert such as remote groups or by calculation of other physiochemical descriptors.

9. References

OECD (Organisation for Economic Cooperation and Development). Guideline for Testing Chemicals No 404, Skin irritation, Paris, (2002). http://www.oecd.org/dataoecd/45/25/2741642.doc).

Anon, 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, *Official Journal of the European Communities 196*, 16.8.1967, 1-98 (1967).

EC. Commission Directive 2001/59/EC of 6 August 2001 adapting to technical progress for the 28th time Council Directive 67/548/EEC on the approximation of the laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances *Official Journal L 225 , 21/08/2001 P. 0001 – 0333*,Office for Official Publications of the European Communities, Luxembourg, **2001**.

OECD, Harmonised integrated classification system for human health and environmental hazards of chemical substances and mixtures, *http://www.oecd.org/ehs/* **2001**.