Evaluation of (Q)SARs for the prediction of Eye Irritation/Corrosion Potential

Physicochemical exclusion rules

Ivanka Tsakovska, Tatiana Netzeva and Andrew Worth

2005 EUR 21897 EN
Evaluation of (Q)SARs for the prediction of Eye Irritation/Corrosion Potential

*Physicochemical exclusion rules*

Ivanka Tsakovska, Tatiana Netzeva and Andrew Worth
LEGAL NOTICE

Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of the following information.

A great deal of additional information on the European Union is available on the Internet. It can be accessed through the Europa server (http://europa.eu.int)
Abstract

The traditional test method for assessing the potential of chemicals to cause eye irritation and corrosion is the Draize rabbit eye irritation test. However, for scientific and animal welfare reasons, there is an impetus to use alternative (non-animal) methods for the determination of local irritation or corrosion, to replace or reduce the need for animal testing. This is reflected in the proposed REACH (Registration, Evaluation and Authorisation of Chemicals) regulation which calls for the development and application of valid quantitative structure–activity relationship (Q)SAR models for the assessment of chemicals.

A set of rules (rulebase) has been developed by the German Bundesinstitut für Risikobewertung (BfR) and incorporated into a Decision Support System (DSS), which is used within the BfR to support regulatory decisions. The rulebase is based on the combined use of two predictive approaches: a) physicochemical exclusion rules to identify chemicals with no skin irritation/corrosion or eye irritation/corrosion potential; and b) structural inclusion rules to identify chemicals with skin irritation/corrosion or eye irritation/corrosion potential.

In this study, an evaluation was performed of the physicochemical rulebase (comprising 31 physicochemical exclusion rules) for predicting the absence of eye irritation/corrosion. In particular, the following aspects were addressed: a) an assessment of the derivation of the rules by using the rulebase training set of 1358 substances (343 irritants/corrosives and 1015 non-irritants and non-corrosives); b) an external validation by using a test set of 199 chemicals (45 irritants/corrosives and 154 non-irritants and non-corrosives); c) an assessment of the suitability of the test set used for validation; d) an assessment of the compliance of the rule-base with the OECD principles for the validation of (Q)SARs.

According to the results of this study: a) the physicochemical exclusion rules for eye irritation/corrosion comply well with the OECD validation principles; b) predictions of no adverse effect (NOT R34/R35/R36/R41) can be made for 20 out of the 199 chemicals in the test set; c) 3 of the 45 irritants/corrosives are falsely predicted as non-irritant or non-corrosive; d) the probability of a negative prediction being correct (Negative Predictive Value) is 0.87; and e) approximately 10% of Draize rabbit eye tests could be avoided by relying on the predictions of no adverse effect.

On the basis of the results, some recommendations are made regarding the improvement of the physicochemical exclusion rules for eye irritation/corrosion.
Contents

Abstract .............................................................................................................................. 1
Abbreviations .................................................................................................................... 3
Introduction .................................................................................................................... 4
Method ............................................................................................................................ 5

Results ............................................................................................................................ 7
  Evaluation of the physicochemical cut-off values by using the training set .............. 8
    Molecular weight ...................................................................................................... 8
    Octanol-water partition coefficient ......................................................................... 9
    Lipid solubility ......................................................................................................... 10
    Aqueous solubility ................................................................................................... 11
    Melting Point .......................................................................................................... 12
  External validation using a test set of chemicals ....................................................... 13
  Investigation of the suitability of the test set for validation ..................................... 14
    Comparison of the distributions of the training and test set over the physicochemical domain ................................................................. 14
    Molecular weight .................................................................................................. 15
    Octanol-Water partition coefficient ...................................................................... 15
    Lipid solubility ........................................................................................................ 16
    Aqueous solubility ................................................................................................. 17
    Melting point ......................................................................................................... 18
  Investigation of the structural similarity of the training and test set ....................... 18
  External validation of the physicochemical rules using a test set of 199 compounds .............................................................................................. 20
    Molecular weight .................................................................................................. 20
    Octanol–water partition coefficient ...................................................................... 20
    Lipid solubility ........................................................................................................ 21
    Aqueous solubility ................................................................................................. 21
    Melting point ......................................................................................................... 22
    Combined application of the 31 exclusion rules ................................................. 23
  Compliance of the exclusion rules with the OECD principles of (Q)SAR validation ......................................................................................... 25

Discussion and Conclusions ......................................................................................... 29
  Molecular weight ...................................................................................................... 29
  Octanol-water partition coefficient .......................................................................... 30
  Lipid solubility ......................................................................................................... 31
  Aqueous solubility ................................................................................................... 31
  Melting point .......................................................................................................... 32
  Overall evaluation of the 31 exclusion rules ......................................................... 33

Acknowledgements ...................................................................................................... 34

References ..................................................................................................................... 35

Appendix 1: Box plots illustrating the distributions of physicochemical properties used in the definition of the BfR exclusion rules ......................................................... 37
Abbreviations

BfR: Bundesinstitut für Risikobewertung
DSS: Decision Support System
ECB: European Chemical Bureau
EU: European Union
NCD: New Chemicals Database
NPV: Negative Predictive Value
OECD: Organisation for Economic Cooperation and Development
QSAR: Quantitative Structure-Activity Relationship
RIVM: Dutch Institute for Public Health and the Environment
R36: EU risk phrase. Irritating to eyes (moderate eye irritation reversible within ≤ 21 days)
R41: EU risk phrase. Risk of serious damage to eyes (moderate but persistent eye lesions, eye corrosion)
R34: EU risk phrase. Causes burns (skin corrosion caused by a 4-hours skin contact)
R35: EU risk phrase. Causes severe burns (skin corrosion caused by a 3-minutes skin contact)
Introduction

The traditional test method for assessing the potential of chemicals to cause eye irritation and corrosion is the Draize rabbit eye irritation test, as incorporated into EU Test Method B.5 and OECD Test Guideline 405 (1). However, for scientific and animal welfare reasons, there is an impetus to use alternative (non-animal) methods for the determination of local irritation or corrosion, to replace or reduce the need for animal testing.

This is reflected in the proposed REACH (Registration, Evaluation and Authorisation of Chemicals) regulation (2) which calls for the development and application of valid quantitative structure–activity relationship (Q)SAR models for the assessment of chemicals. One application could be the screening of hazardous chemicals prior to in vitro and/or in vivo testing.

Chemicals are classified on the basis of their eye irritation/corrosion potential according to European Union (EU) or United Nations (UN) criteria within the Globally Harmonized Classification System (3).

A Decision Support System (DSS) has been developed by the German Bundesinstitut für Risikobewertung (BfR) that applies physicochemical rules and structural alerts to identify potential skin and eye irritant/corrosive chemicals (4,5,6,7). The system was developed in order to support regulatory decision making for hazard assessment of chemicals, and is based on the combined use of two predictive approaches:

1. physicochemical property limits (physicochemical exclusion rules) to identify chemicals with no skin irritation/corrosion or eye irritation/corrosion potential.
2. structural alerts (inclusion rules) to identify chemicals with skin irritation/corrosion or eye irritation/corrosion potential.

The current (2005) version of the DSS is based on a training set of 1358 chemicals with experimental data for skin and eye irritation and corrosion. Only pure substances were considered (95% purity). The training set was compiled from confidential data submitted under the EU New Chemicals Notification procedure, and contained within the New Chemicals Database (NCD). The NCD is maintained by the European Chemical Bureau (ECB; 8).

Recently, an evaluation of the DSS physicochemical exclusion rules for skin irritation and corrosion was performed by the Dutch National Institute for Public Health and the Environment (RIVM), with financial support from the ECB (9). The aim of the present project was to perform a similar evaluation of the DSS physicochemical exclusion rules for eye irritation/corrosion. The current project also included an external validation of the physicochemical rules by using a test set of 199 chemicals not used in the derivation of the rules.
Method

In the BfR DSS system five physiochemical parameters are used to define the exclusion rules for eye irritation/corrosion (Table 1). These parameters are experimentally measured properties generated by protocols used by industry for the notification of New Chemicals in the EU (10). The physicochemical cut-off values in the BfR rulebase aim to predict the absence of eye irritation/corrosion according to the EU classification and labelling system. They are defined in terms of physicochemical limit values identified in groups of irritants/corrosives included in the BfR database.

Table 1. Physicochemical properties used in the definition of the exclusion rules

<table>
<thead>
<tr>
<th>Physicochemical property</th>
<th>Unit</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular Weight</td>
<td>g/mol</td>
<td>MW</td>
</tr>
<tr>
<td>Octanol-Water partition coefficient</td>
<td>none</td>
<td>log P</td>
</tr>
<tr>
<td>Lipid solubility</td>
<td>g/kg</td>
<td>l.s.</td>
</tr>
<tr>
<td>Aqueous solubility</td>
<td>g/l</td>
<td>a.s.</td>
</tr>
<tr>
<td>Melting Point</td>
<td>°C</td>
<td>m.p.</td>
</tr>
</tbody>
</table>

The eye irritation/corrosion caused by a chemical is characterized using the following EU risk phrases:

R36: Irritating to eyes (moderate eye irritation reversible within ≤ 21 days)
R41: Risk of serious damage to eyes (moderate but persistent eye lesions, eye corrosion)

According to international risk assessment guidelines, skin corrosion potential excludes further considerations on a similar hazardous potential to eyes, since the chemical has already proved to have corrosive properties. Thus the following effects induced by local contact to skin are also assumed to be predictive of eye damage:

R34: Causes burns (skin corrosion caused by a 4-hour skin contact)
R35: Causes severe burns (skin corrosion caused by a 3-minute skin contact)

For the project implementation, the latest set of physicochemical rules in the BfR database were used (Table 2), as communicated by the BfR (Dr Matthias Herzler). The rules are either applicable to all substances (Group “All” in Table 2) or to particular chemical classes (classes C, CN, CNHal, CNS, and CHal in Table 2). The description of the chemical classes is given in Table 3. The substances that do not fall into one of the chemical classes are predicted by the physicochemical exclusion rules as defined for all substances.
Table 2. Physicochemical exclusion rules for eye irritation/corrosion

<table>
<thead>
<tr>
<th>ID</th>
<th>Group</th>
<th>IF parameter</th>
<th>Qualifier</th>
<th>Value</th>
<th>Unit</th>
<th>Then NOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All</td>
<td>m.p.</td>
<td>&gt;</td>
<td>200</td>
<td>°C</td>
<td>R34 or R35</td>
</tr>
<tr>
<td>2</td>
<td>All</td>
<td>logP</td>
<td>&gt;</td>
<td>9</td>
<td></td>
<td>R34, R35, R36 or R41</td>
</tr>
<tr>
<td>3</td>
<td>All</td>
<td>logP</td>
<td>&lt;</td>
<td>-3.1</td>
<td></td>
<td>R34 or R35</td>
</tr>
<tr>
<td>4</td>
<td>All</td>
<td>l.s.</td>
<td>&lt;</td>
<td>0.01</td>
<td>g/kg</td>
<td>R34 or R35</td>
</tr>
<tr>
<td>5</td>
<td>All</td>
<td>a.s.</td>
<td>&lt;</td>
<td>0.00002</td>
<td>g/l</td>
<td>R41</td>
</tr>
<tr>
<td>6</td>
<td>All</td>
<td>a.s.</td>
<td>&lt;</td>
<td>0.000005</td>
<td>g/l</td>
<td>R36</td>
</tr>
<tr>
<td>7</td>
<td>All</td>
<td>m.w.</td>
<td>&gt;</td>
<td>650</td>
<td>g/mol</td>
<td>R36 a)</td>
</tr>
<tr>
<td>8</td>
<td>C</td>
<td>m.p.</td>
<td>&gt;</td>
<td>55</td>
<td>°C</td>
<td>R34 or R35</td>
</tr>
<tr>
<td>9</td>
<td>C</td>
<td>m.w.</td>
<td>&gt;</td>
<td>380</td>
<td>g/mol</td>
<td>R34, R35, R36 or R41</td>
</tr>
<tr>
<td>10</td>
<td>C</td>
<td>a.s.</td>
<td>&lt;</td>
<td>0.0001</td>
<td>g/l</td>
<td>R34, R35, R36 or R41</td>
</tr>
<tr>
<td>11</td>
<td>C</td>
<td>a.s.</td>
<td>&lt;</td>
<td>0.0005</td>
<td>g/l</td>
<td>R36 or R41</td>
</tr>
<tr>
<td>12</td>
<td>CN</td>
<td>l.s.</td>
<td>&lt;</td>
<td>0.4</td>
<td>g/kg</td>
<td>R34 or R35</td>
</tr>
<tr>
<td>13</td>
<td>CN</td>
<td>m.w.</td>
<td>&gt;</td>
<td>290</td>
<td>g/mol</td>
<td>R34 or R35</td>
</tr>
<tr>
<td>14</td>
<td>CN</td>
<td>a.s.</td>
<td>&lt;</td>
<td>0.1</td>
<td>g/l</td>
<td>R34 or R35</td>
</tr>
<tr>
<td>15</td>
<td>CN</td>
<td>logP</td>
<td>&gt;</td>
<td>4.5</td>
<td></td>
<td>R34 or R35</td>
</tr>
<tr>
<td>16</td>
<td>CNHal</td>
<td>logP</td>
<td>&gt;</td>
<td>3.8</td>
<td></td>
<td>R34, R35 or R41</td>
</tr>
<tr>
<td>17</td>
<td>CNHal</td>
<td>a.s.</td>
<td>&lt;</td>
<td>0.1</td>
<td>g/l</td>
<td>R34 or R35</td>
</tr>
<tr>
<td>18</td>
<td>CNHal</td>
<td>m.w.</td>
<td>&gt;</td>
<td>370</td>
<td>g/mol</td>
<td>R34 or R35</td>
</tr>
<tr>
<td>19</td>
<td>CNHal</td>
<td>l.s.</td>
<td>&lt;</td>
<td>400</td>
<td>g/kg</td>
<td>R34 or R35</td>
</tr>
<tr>
<td>20</td>
<td>CNHal</td>
<td>a.s.</td>
<td>&lt;</td>
<td>0.004</td>
<td>g/l</td>
<td>R41</td>
</tr>
<tr>
<td>21</td>
<td>CNS</td>
<td>m.w.</td>
<td>&gt;</td>
<td>620</td>
<td>g/mol</td>
<td>R34, R35 or R36</td>
</tr>
<tr>
<td>22</td>
<td>CNS</td>
<td>m.p.</td>
<td>&gt;</td>
<td>50</td>
<td>°C</td>
<td>R34 or R35</td>
</tr>
<tr>
<td>23</td>
<td>CNS</td>
<td>logP</td>
<td>&lt;</td>
<td>-2</td>
<td></td>
<td>R34 or R35</td>
</tr>
<tr>
<td>24</td>
<td>CNS</td>
<td>logP</td>
<td>&gt;</td>
<td>1.5</td>
<td></td>
<td>R36</td>
</tr>
<tr>
<td>25</td>
<td>CNS</td>
<td>logP</td>
<td>&gt;</td>
<td>3.6</td>
<td></td>
<td>R41</td>
</tr>
<tr>
<td>26</td>
<td>CNS</td>
<td>a.s.</td>
<td>&lt;</td>
<td>0.006</td>
<td>g/l</td>
<td>R36 or R41</td>
</tr>
<tr>
<td>27</td>
<td>CHal</td>
<td>m.w.</td>
<td>&gt;</td>
<td>280</td>
<td>g/mol</td>
<td>R34 or R35</td>
</tr>
<tr>
<td>28</td>
<td>CHal</td>
<td>m.w.</td>
<td>&gt;</td>
<td>370</td>
<td>g/mol</td>
<td>R34, R35 or R36</td>
</tr>
<tr>
<td>29</td>
<td>CHal</td>
<td>m.p.</td>
<td>&gt;</td>
<td>65</td>
<td>°C</td>
<td>R34 or R35</td>
</tr>
<tr>
<td>30</td>
<td>CHal</td>
<td>logP</td>
<td>&gt;</td>
<td>4.5</td>
<td></td>
<td>R34, R35 or R41</td>
</tr>
</tbody>
</table>

Set of rules confirmed by the BfR (Dr Matthias Herzler)

Table 3. The chemical classes defined in the BfR database

<table>
<thead>
<tr>
<th>Class (designation)</th>
<th>Chemical elements included in the structure</th>
<th>Empirical Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>C, H and O atoms</td>
<td>CnH2Oz</td>
</tr>
<tr>
<td>CN</td>
<td>C, H, O and N atoms</td>
<td>CnH2O2Na</td>
</tr>
<tr>
<td>CNHal</td>
<td>C, H, O, N and Halogen atoms</td>
<td>CnH2O2NaHalb</td>
</tr>
<tr>
<td>CNS</td>
<td>C, H, O, N and S atoms</td>
<td>CnH2O2NSb</td>
</tr>
<tr>
<td>CHal</td>
<td>C, H, O and Halogen atoms</td>
<td>CnH2O2Halb</td>
</tr>
</tbody>
</table>
The project was performed in three phases:

1. Evaluation of the physicochemical cut-off values as defined on the basis of the training set of 1358 substances.

2. External validation using a test set of 199 chemicals from the NCD.

2.1. Investigation of the suitability of the test set for validation.
   a. Comparison of the distributions of the training and test set over the physicochemical domain.
   b. Investigation of the structural similarity between the training and the test set.

2.2. Evaluation of the predictivity of the BfR DSS by comparing the predictions with the actual classifications of the substances proposed by the EU Working Group on Classification and Labelling.

3. Evaluation of the compliance of the BfR rule-base for eye irritation/corrosion with the OECD principles for validation of (Q)SARs (11).

**Results**

The training set of 1358 substances contains 1015 non-irritants and non-corrosives (74.74%), and 343 eye irritants/corrosives (25.26%), of which:

- 57 substances (4.20%) are labelled R36
- 77 substances (5.67%) are labelled R34 or R35
- 209 substances (15.39%) are labelled R41

The distribution of substances into the chemical classes is given in Table 4.

**Table 4. Distribution of substances according to chemical class and EU risk phrase**

<table>
<thead>
<tr>
<th>Class</th>
<th>No of substances</th>
<th>Not labelled</th>
<th>R36</th>
<th>R41</th>
<th>R34 / R35</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>261</td>
<td>207</td>
<td>14</td>
<td>31</td>
<td>9</td>
</tr>
<tr>
<td>CHal</td>
<td>79</td>
<td>46</td>
<td>4</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td>CN</td>
<td>494</td>
<td>372</td>
<td>25</td>
<td>75</td>
<td>22</td>
</tr>
<tr>
<td>CNHal</td>
<td>177</td>
<td>143</td>
<td>3</td>
<td>23</td>
<td>8</td>
</tr>
<tr>
<td>CNS</td>
<td>186</td>
<td>140</td>
<td>6</td>
<td>35</td>
<td>5</td>
</tr>
</tbody>
</table>
**Evaluation of the physicochemical cut-off values by using the training set**

For the purposes of the evaluation the various exclusion rules (Table 2) are grouped according to the physicochemical parameter used. The proximity of the cut-off values to the highest/lowest value of the given physicochemical parameter in the training set of 1358 substances as a whole and in the separate classes is discussed for each rule separately. The so called “safety margin” is considered as a measure for the “safe” definition of the rule. It is equal to the difference between the cut-off value defined in the DSS for each exclusion rule and the actual highest/lowest physicochemical parameter value in the group of the eye irritants/corrosives.

**Molecular weight**

**Class All:**  \[ \text{IF m.w.} > 650 \text{ g/mol} \text{ THEN NOT R36} \]

Molecular weights of the substances in the training set according to the labelling are given in Table 5.

**Table 5. Ranges of the molecular weight according to the classification of the substances**

<table>
<thead>
<tr>
<th>Labelling</th>
<th>Number of substances</th>
<th>Lowest MW</th>
<th>Highest MW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not labelled</td>
<td>1015</td>
<td>88</td>
<td>79000</td>
</tr>
<tr>
<td>R36</td>
<td>57</td>
<td>114</td>
<td>647</td>
</tr>
<tr>
<td>R41</td>
<td>209</td>
<td>66</td>
<td>998</td>
</tr>
<tr>
<td>R34, R35</td>
<td>77</td>
<td>66</td>
<td>1046</td>
</tr>
</tbody>
</table>

The number of non-polymeric substances is 1350 and the highest non-polymeric molecular weight is 1506 g/mol. There are 57 eye irritants in the training set, with molecular weights in the range 114-647 g/mol. Therefore, the cut-off values cover all eye irritants and have a small safety margin (3 g/mol) between the highest molecular weight and the cut-off value.

**Class C:**  \[ \text{IF m.w.} > 380 \text{ g/mol} \text{ THEN NOT R34, R35, R36 or R41} \]

From the 261 substances in the class C, 54 are irritants/corrosives (R36, R41, R34, R35). They have molecular weights in the range 66-370 g/mol. The cut-off values cover all eye irritants and corrosives and have a safety margin of 10 g/mol.

**Class CN:**  \[ \text{IF m.w.} > 290 \text{ g/mol} \text{ THEN NOT R34 or R35} \]

There are 22 corrosive compounds (R34, R35) in this class with a range of the molecular weight 89-281 g/mol. The cut-off values cover all eye corrosives (R34, R35) and have a safety margin of 9 g/mol.

**Class CNHal:**  \[ \text{IF m.w.} > 370 \text{ g/mol} \text{ THEN NOT R34 or R35} \]
From the 177 compounds in the class, 8 are corrosives (R34, R35) and the range of their molecular weight is 127-370 g/mol. Therefore the cut-off value coincides with the highest molecular weight in the group of the corrosives.

**Class CNS:** \[\text{IF } \text{m.w.} > 620 \text{ g/mol} \text{ THEN NOT } \text{R34, R35 or R36}\]
The number of irritants/corrosives (R36, R34, R35) is 11 and the range of the molecular weight is 145-508 g/mol. The cut-off value is set with a relative large safety margin of 112 g/mol.

**Class CHal:** \[\text{IF } \text{m.w.} > 280 \text{ g/mol} \text{ THEN NOT } \text{R34 or R35} \text{ IF } \text{m.w.} > 370 \text{ g/mol} \text{ THEN NOT } \text{R34, R35 or R36}\]
The number of compounds with R34, R35 (first exclusion rule for the class) is 13 and they have molecular weight ranging from 146 g/mol to 276 g/mol. The cut-off value is defined adequately and covers all of them with a safety margin of 4 g/mol.

The number of compounds labelled R36, R34, R35 (second exclusion rule for the class) is 17 and the highest value of the molecular weight in the group is 368 g/mol. The cut-off value for compounds labelled R36, R34, R35 covers all of them and has a narrow safety margin of 2 g/mol.

**Octanol-water partition coefficient**

For all 1358 compounds in the training set measured log P values range from -7 to 25. In order to evaluate the cut-off values defined, the ranges of the log P for the non-irritants, irritants and corrosives are analysed (Table 6).

**Table 6. Ranges of log P according to the classification of the substances**

<table>
<thead>
<tr>
<th>Labelling</th>
<th>Number of compounds</th>
<th>Lowest log P value</th>
<th>Highest log P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not labelled</td>
<td>1015</td>
<td>-5.1</td>
<td>25</td>
</tr>
<tr>
<td>R36</td>
<td>57</td>
<td>-6</td>
<td>2.4</td>
</tr>
<tr>
<td>R41</td>
<td>209</td>
<td>-7</td>
<td>9</td>
</tr>
<tr>
<td>R34, R35</td>
<td>77</td>
<td>-3.1</td>
<td>8.7</td>
</tr>
</tbody>
</table>

**Class ALL:** \[\text{IF } \text{log P} > 9 \text{ THEN NOT } \text{R34, R35, R36 or R41} \text{ IF } \text{log P} < -3.1 \text{ THEN NOT } \text{R34 or R35}\]

Analysing Table 6 it is obvious that the cut-off values (both exclusion rules) are equal to the highest log P value in the group of compounds labelled R34, R35, R36, R41 and respectively the lowest log P value in the group of R34, R35 labelled compounds. The rules therefore are defined in order to cover all irritants and corrosives/skin corrosives but without any safety margin.

**Class CN:** \[\text{IF } \text{log P} > 4.5 \text{ THEN NOT } \text{R34 or R35}\]
In this class the range of log P is from -5 to 25. There are 22 compounds with R34, R35 and the highest value in this group is 4.2. The margin between the cut-off value and the highest log P in the group is 0.3.

**Class CNHal:** \[ \text{IF log P > 3.8 THEN NOT R34, R35 or R41} \]
The range of log P values among the 177 compounds in the class is -2.4–13.6. There are 31 corrosive substances (R34, R35, R41) with log P values ranging from -2.3 to 7. Obviously the cut-off value is not properly defined - after the compound (R34) with log P = 3.7 (in the margins of cut-off) there is a compound (R41) with log P = 7.

**Class CNS:** \[ \text{IF log P < \(-2\) THEN NOT R34 or R35} \]
\[ \text{IF log P > 1.5 THEN NOT R36} \]
\[ \text{IF log P > 3.6 THEN NOT R41} \]
In this class there are 46 compounds with R34, R35, R36, R41 and log P values from -6 to 8.7. The number of skin corrosives (R34, R35) is 5 with log P ranging from -1.3 to 8.7. Thus the first rule for this class is properly defined with a safety margin of 0.7. The irritants (R36) are 6 with log P values from -6 to 1.4. The second rule is on the safe side but the difference between the cut-off value and the highest log P value in the group is only 0.1. The number of substances labeled R41 is 35 and highest log P in the group is 3.5. The cut-off is properly defined but the safety margin is very narrow (0.1).

**Class CHal:** \[ \text{IF log P > 4.5 THEN NOT R34, R35 or R41} \]
The range of log P in the class is -2.6-8.3. There are 29 corrosive compounds (R34, R35, R41) with log P values from 2.6 to 3.7. Therefore the rule covers all corrosives with a large safety margin of 0.8.

**Lipid solubility**

The lipid solubility is measured for 554 compounds in the training set. It varies from 0.00002 to 2500 g/kg. Table 7 represents the ranges of the lipid solubility according to the classification of substances.

**Table 7. Ranges of lipid solubility according to the classification of the substances**

<table>
<thead>
<tr>
<th>Labelling</th>
<th>Number of compounds</th>
<th>Lowest L.s. value</th>
<th>Highest L.s. value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not labelled</td>
<td>431</td>
<td>2.10^{-9}</td>
<td>2500</td>
</tr>
<tr>
<td>R36</td>
<td>22</td>
<td>0.0018</td>
<td>1000</td>
</tr>
<tr>
<td>R41</td>
<td>68</td>
<td>2.5.10^{-5}</td>
<td>1000</td>
</tr>
<tr>
<td>R34, R35</td>
<td>33</td>
<td>0.01</td>
<td>1000</td>
</tr>
</tbody>
</table>

**Class All:** \[ \text{IF l.s. < 0.01 g/kg THEN NOT R34 or R35} \]
The corrosive substance with the lowest lipid solubility (Table 7) has a lipid solubility of 0.01 g/kg. The next lowest lipid solubility of a corrosive substance is 0.145 g/kg. The cut of value of the lipid solubility exclusion rule for all substances therefore has no safety margin.
Class CN:  \[ \text{IF l.s. < 0.4 g/kg THEN NOT R34 or R35} \]
In the CN class there are 22 corrosive substances. For 15 of them the l.s. is measured and only 3 have lipid solubility other than 1000 g/kg: 0.55 g/kg, 10 g/kg and 63 g/kg. Obviously the cut-off value of 0.4 g/kg is set on the safe side of these two values. However, establishing a cut-off value based only on four different values of l.s. concentrations gives a large uncertainty in the application of the rule.

Class CNHal:  \[ \text{IF l.s. < 400 g/kg THEN NOT R34 or R35} \]
The CNHal class contains only 3 corrosive substances, two of which have a l.s. of 1000 g/kg and the third has l.s. of 442 g/kg. The rule is therefore properly defined but with very high uncertainty.

Aqueous solubility

Experimental measurements of aqueous solubility were available for 1272 substances in the training set, with the ranges given in the Table 8.

<table>
<thead>
<tr>
<th>Labelling</th>
<th>Number of compounds</th>
<th>Lowest a.s. value</th>
<th>Highest a.s. value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not labelled</td>
<td>972</td>
<td>0.000001</td>
<td>1000</td>
</tr>
<tr>
<td>R36</td>
<td>55</td>
<td>0.000005</td>
<td>1000</td>
</tr>
<tr>
<td>R41</td>
<td>191</td>
<td>0.000005</td>
<td>1894</td>
</tr>
<tr>
<td>R34, R35</td>
<td>54</td>
<td>0.000207</td>
<td>1000</td>
</tr>
</tbody>
</table>

The 86 substances lacking a.s. data show high presence of corrosiveness/irritancy (43/86 = 50%) compared to the overall appearance of corrosiveness/irritancy in the training set data (343/1358 = 25.25%). Two main reasons why aqueous solubility is not or cannot be measured are: a) complete insolubility in water, and b) reactivity with water. This reactivity might be an important indicator for possible corrosivity or irritancy. Thus, more information about these substances related to their a.s. could be very useful in defining the a.s. exclusion rules.

Class All:  \[ \text{IF a.s. < 0.00002 g/l THEN NOT R41} \]
\[ \text{IF a.s. < 0.000005 g/l THEN NOT R36} \]
There are 55 eye irritants (R36) with a.s. from 0.000005 to 1000 g/l (Table 8). Obviously the cut-off value overlaps the lowest a.s. in the range and therefore there is no safety margin. 191 compounds are with R41 in range 0.000005-1894 g/l. The lowest values of a.s. are 0.0000242 g/l and 0.000005 g/l. Therefore the rule “If a.s.< 0.00002 g/l then not R41” is not properly defined.

Class C:  \[ \text{IF a.s. < 0.0001 g/l THEN NOT R34, R35, R36 or R41} \]
\[ \text{IF a.s. < 0.00005 g/l THEN NOT R36 or R41} \]
Out of 246 compounds in the class with measured a.s. value, 47 compounds are labelled R34, R35, R36, R41 and their a.s. values ranges from 0.000262 to 1000 g/l. Therefore the
rule “If a.s. < 0.0001 then not R34,35,36,41” is set on the safe side. A total of 42 irritants (R36, R41) have a.s. in the range 0.00972-1000 g/l. Thus, the rule “if a.s. < 0.0005 g/l then not R36, R41” is set with a safety margin.

Class CN: \[\text{IF a.s.} < 0.1 \text{ g/l THEN NOT R34 or R35}\]

There are 19 skin corrosives with experimental a.s. data in the range 0.1-1000 g/l. Therefore the cut-off value is set at the lowest a.s. value in the group.

Class CNHal: \[\text{IF a.s.} < 0.1 \text{ g/l THEN NOT R34 or R35}\]
\[\text{IF a.s.} < 0.004 \text{ g/l THEN NOT R41}\]

Measured a.s. data for 173 compounds are in the range 0.000001-1000 g/l. Only 7 are the corrosives with data for a.s. from 0.13 g/l to 135 g/l. The cut-off value is set with a safety margin of 0.03 g/l.

There are 23 R41 compounds. The lowest a.s. values in the group are 0.0048 g/l and 0.000005 g/l. The presence of a compound with a.s. = 0.000005 g/l breaks the rule “if a.s. < 0.004 g/l then not R41”.

Class CNS: \[\text{IF a.s.} < 0.006 \text{ g/l THEN NOT R36 or R41}\]

The number of eye irritants (R36, R41) is 38 and the lowest measured a.s. in the group is 0.00644 g/l. The cut-off value is set with a safety margin.

Melting Point

The ranges for the melting points of the substances in the training set are given in Table 9.

Table 9. Ranges of melting point according to the classification of the substances

<table>
<thead>
<tr>
<th>Labelling</th>
<th>Number of compounds</th>
<th>Lowest m.p.</th>
<th>Highest m.p.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not labelled</td>
<td>1015</td>
<td>-170</td>
<td>500</td>
</tr>
<tr>
<td>R36</td>
<td>57</td>
<td>-94</td>
<td>346</td>
</tr>
<tr>
<td>R41</td>
<td>209</td>
<td>-100</td>
<td>430</td>
</tr>
<tr>
<td>R34,R35</td>
<td>77</td>
<td>-76</td>
<td>358</td>
</tr>
</tbody>
</table>

All compounds in the training set have a value for the melting point. The range is from -170°C to 500°C. For 904 compounds a single value for melting point is reported, and for 454 compounds a range values is given. In this case the lower limit of the melting range is taken as input for the exclusion rule.

Class All: \[\text{IF m.p.} > 200^\circ\text{C THEN NOT R34 or R35}\]

The number of skin corrosive substances is 77 and they have melting points ranging from -76°C to 358°C. Five substances (6.5%) have a melting points > 200°C (245-360°C) but are still considered to be corrosive. The substance with the highest melting point below 200°C that still shows corrosivity has a melting point of 195°C. The cut-off value of
200°C was not chosen on the safe side, since corrosive substances with melting points of up to 358°C are present in the training set.

**Class C:** \( \text{IF m.p. > 55°C THEN NOT R34 or R35} \)

The nine skin corrosive substances in the class have melting point values from -45 to 285°C. There is one corrosive substance that breaks the above rule - a lithium salt assigned R35 which releases LiOH when in contact with aqueous substrates/organic media. It has m.p. = 285°C. Therefore the cut-off value misses one skin corrosive substance.

**Class CNS:** \( \text{IF m.p. > 50°C THEN NOT R34 or R35} \)
\( \text{IF m.p. > 200°C THEN NOT R36} \)

The CNS class contains 186 substances with melting points ranging from -39 to 430°C. There are 5 corrosive substances (R34) in total, with melting points -17, 0, 31, 120 and 360°C. Thus there are 2 substances with melting points > 50°C. The cut off value of 50°C for corrosive substances in the CNS class seems not to have been set on the safe side. There are only 6 eye irritants (R36) with melting points 50, 95, 96, 125, 126, 177°C. The rule “if m.p. > 200°C then not R36” covers all of them and has a safety margin.

**Class CHal:** \( \text{IF m.p. > 65°C THEN NOT R34 or R35} \)

The class comprises 79 substances with melting points ranging from -160 to 295°C. Thirteen substances show corrosivity, with the highest melting point being 61°C. This rule therefore covers all potential corrosive substances.

**External validation using a test set of chemicals**

The test set comprises 199 substances, of which 154 (77.39%) are not labelled and 45 (22.6%) are labelled as eye irritants/corrosives, as follows: a) 10 substances (5.03%) labelled R36; b) 28 substances (14.07%) labelled R41; and c) 7 substances (3.52%) labelled R34, R35.

The comparison between the training and the test set shows a similar distribution of eye irritants/corrosives (Figure 1). The highest difference in the number of substances is observed in the group of skin corrosives (R34, R35).
Investigation of the suitability of the test set for validation

Before evaluating the predictivity of the BfR physicochemical rules with the external test set it was necessary to investigate if the test substances are in the applicability domain of the model, i.e. if the model is applicable to the test substances. The (Q)SAR applicability domain is defined as the response and chemical structure space in which the model makes predictions with a given reliability (12). There are different methods for its description. In this study, the range-based method was used. First the distributions of the training and the test set over the physicochemical parameters were investigated and compared, and then the structural similarity between the two sets was examined in a qualitative manner.

Comparison of the distributions of the training and test set over the physicochemical domain

In order to obtain the distributions of the substances in the training and test set over the physicochemical domain the cumulative observed frequency was calculated for each physicochemical property. It includes the following steps (13):

1. Determining the number of classes \( K \) (with a given range of the physicochemical parameter values) equal to a square root of the number of substances \( N \): \( K = (N)^{1/2} \).
2. Determining the interval \( I \) for each class by dividing the range of the physicochemical parameter \( R \) by the number of classes: \( I = R / K \).
3. Counting the number of compounds (observed frequency, \( N_i \)) in each class.
4. Calculating the relative observed frequency (ROF) for each class by dividing the observed frequency (number of the compounds in the class) by the total number of compounds: \( \text{ROF}_i = N_i / N \).
5. Obtaining the cumulative observed frequency (COF) on the basis of the observed frequencies for the defined classes: \( \text{COF}_i = \text{COF}_{i-1} + \text{ROF}_i \).

In the following section the distributions for the training and the test set for each physicochemical parameter are presented and compared in order to qualify the similarity in both sets.

**Molecular weight**

The distributions of the training and test set along the molecular weight domain are given in the Figure 2. In the training set substances with m.w. values < 1500 g/mol are included (1348 substances). The distributions are cut off on the lower side because of the specificity of the variable – the lowest molecular weight in the training set is 66 g/mol. The substances in the test set have molecular weights in the range 102-810 g/mol. The cumulative observed frequencies for the test set in this range are higher. On the whole, the distributions of the molecular weight in the training and test set are similar.

**Octanol-Water partition coefficient**

The distributions of the training and test set along the log P domain are given in the Figure 3. For the training set log P values are in range -7-25, and for the test set they are in range -9.9-16.5. The figure shows that both sets have similar distributions.
Lipid solubility

Out of 1358 substances in the training set, 554 have measured lipid solubility values, which is a sufficient amount to establish a physicochemical exclusion rule. However the practical value of such a rule for new substances is questionable. Only the substances for which it is very difficult to determine an octanol-water partition coefficient this property is measured (9). Furthermore, the external validation of this particular BfR rule cannot be performed because only three substances in the test set have measured lipid solubility. However the distribution of the training set along the log(lipid solubility) domain was plotted in order to investigate the applicability of this rule (Figure 4). It is not a normal distribution. The values for l.s. are not uniformly distributed in the range: a significant proportion of substances, 134 out of 554 (24%), have l.s. = 1000 g/kg. Forty three out of 123 irritants/corrosives (35%) have l.s. = 1000 g/kg. Therefore the data basis for the exclusion rules based on lipid solubility is not representative.
Aqueous solubility

The distributions of the training and the test set along the aqueous solubility domain are given in Figure 5. Obviously they are not normal distributions. In the training set there are a number of substances with a.s. = 1000 g/l and one substance with a.s. = 1894 g/l that is the reason for the skewed distribution on the side of highly soluble substances. In the test set, 90% of the chemicals have a.s. < 200 g/l, which makes the distribution very steep on the lower side. Further selection in the training and test set is recommended to increase the quality of the defined rules based on aqueous solubility.
Melting point

The distributions of the training and test set over the melting point domain are given in the Figure 6. The range of the melting points in the training set is from -170 to 500°C and in the test set – from -170 to 400°C. The similarity in the ranges and the distributions indicates that the test set is suitable for the external validation of the physicochemical rules based on melting point.

Figure 6. Distribution of the training set (1358 substances) and the test set (196 substances) along the melting point domain

The distributions of the physicochemical properties (molecular weight, log P, aqueous solubility and melting point) within the training and test sets are also presented in the form of box plots (see Annex I). The distribution of lipid solubility is not presented in Annex I because of the limited number of experimental values in the test set.

For each distribution, the box plots provide the following information: a) the central tendency (median), i.e. the value for which one-half (50%) of the observations (when ranked) lie above that value and one-half below; b) range or variation statistics (the 25th and 75th percentiles); c) outliers, i.e. extreme values that are “far” from the median of the distribution (14).

Investigation of the structural similarity of the training and test set

Overall the training and the test set have approximately the same presence of substances in all chemical classes (Figure 7). This is an indication that the training and test set are sufficiently similar in their distribution of substances over the various chemical classes. The structural similarity could be an indication that the compounds in the test set are in the applicability domain of the physicochemical rules defined.
Figure 7. Distribution of the chemical classes in the training set (a) and the test set (b)

The distributions of the substances over the five classes for which separate physicochemical rules have been determined are given in Figure 8. These classes (C, CHal, CN, CNHal and CNS) constitute 86% of the substances in the training set (1170/1358) and 85% of the substances in the test set (170/199).

The most notable difference between the training and the test set is the larger number of the CHal compounds in the test set (15.4%) versus the training set (5.8%). However, the absolute number of CHal compounds in the training set (79) is still larger than the number of CHal compounds in the validation set (31), so as to allow valid predictions for the external validation data (16).

Figure 8. Distribution of the substances in the training and the test set over the five chemical classes for which separate exclusion rules are defined
External validation of the physicochemical rules using a test set of 199 compounds

To perform an external validation, the predictions generated by each rule were compared with the actual classifications of the test substances as agreed by the EU Working Group on Classification and Labelling. The results in terms of correct/incorrect predictions for each physicochemical rule are given in the following sections.

Molecular weight

The results from the external validation are given in the Table 10.

Table 10. External validation of the rules based on molecular weight

<table>
<thead>
<tr>
<th>Physicochemical rule</th>
<th>Prediction in the test set</th>
<th>Number of false negatives</th>
<th>Correct negative predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Class</td>
<td>IF m.w.&gt;650g/mol</td>
<td>THEN NOT R36</td>
</tr>
<tr>
<td>C (29)</td>
<td>m.w.&gt;380g/mol</td>
<td>R34, R35, R36</td>
<td>0</td>
</tr>
<tr>
<td>CN (67)</td>
<td>m.w.&gt;290g/mol</td>
<td>R34 or R35</td>
<td>0</td>
</tr>
<tr>
<td>CNHal (28)</td>
<td>m.w.&gt;370g/mol</td>
<td>R34 or R35</td>
<td>0</td>
</tr>
<tr>
<td>CNS (13)</td>
<td>m.w.&gt;620g/mol</td>
<td>R34, R35 or R36</td>
<td>0</td>
</tr>
<tr>
<td>CHal (31)</td>
<td>m.w.&gt;280g/mol</td>
<td>R34 or R35</td>
<td>0</td>
</tr>
<tr>
<td>CHal (31)</td>
<td>m.w.&gt;370g/mol</td>
<td>R34, R35, R36</td>
<td>0</td>
</tr>
</tbody>
</table>

The number of compounds in each particular class is given in parentheses

* The percentage is calculated on the basis of the total number of substances in each particular class.

In Table 10, it can be seen that there are no false negative predictions in the test set based on “molecular weight” rules. If the rule for all substances is considered in isolation, only 7% of the substances can be predicted (as not R36), and nothing can be concluded regarding the remaining 93% of the substances in the test set. The division of substances into chemical classes and the application of the general and class-specific rules leads to predictions for a higher number of compounds (76 out of 199, i.e. 38% of the test set chemicals).

Octanol–water partition coefficient

The results from the external validation are given in the Table 11.
Table 11. External validation of the rules based on log P

<table>
<thead>
<tr>
<th>Physicochemical rule</th>
<th>Prediction in the test set</th>
<th>Number of false negatives</th>
<th>Correct negative predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IF</td>
<td>THEN NOT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R34, R35, R36 or R41</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>CN (66) log P &gt; 4.5</td>
<td>R34 or R35</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>CNS (27) log P &gt; 3.8</td>
<td>R34, R35 or R41</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>CNS (13) log P &gt; 1.5</td>
<td>R36</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>CNS (13) log P &gt; 3.6</td>
<td>R41</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CHal (30) log P &gt; 4.5</td>
<td>R34, R35 or R41</td>
<td>0</td>
<td>12</td>
</tr>
</tbody>
</table>

The number of compounds in each particular class is given in parentheses.

* The percentage is calculated on the basis of the total number of substances in each particular class.

There are no false negative predictions based on log P, i.e. no substance in the test set is incorrectly predicted as an eye irritant or corrosive. As a whole the number of predictions based on the general and class-specific log P exclusion rules is lower compared with molecular weight (47 out of 199 chemicals, i.e. 24% of the test set).

It is noted that some of the rules are based on extremely high or low log P values (e.g. log P > 9 or log P < -2). Such extreme values of log P are very difficult to measure, and even when measured they may not be reliable enough to give useful information about the distribution of the chemicals.

**Lipid solubility**

Only 3 substances in the test set have experimental values of lipid solubility. Therefore, this particular test set is not adequate to carry out a meaningful external validation.

**Aqueous solubility**

The results from the external validation are given in the Table 12.
Table 12. External validation of the rules based on aqueous solubility

<table>
<thead>
<tr>
<th>Physicochemical rule</th>
<th>Prediction in the test set</th>
<th>Number of false negatives</th>
<th>Correct negative predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IF</td>
<td>THEN NOT</td>
<td>Number</td>
</tr>
<tr>
<td>All (186)</td>
<td>a.s.&lt;0.000002 g/l</td>
<td>R41</td>
<td>1</td>
</tr>
<tr>
<td>All (186)</td>
<td>a.s.&lt;0.000005 g/l</td>
<td>R36</td>
<td>1</td>
</tr>
<tr>
<td>C (28)</td>
<td>a.s.&lt;0.0001 g/l</td>
<td>R34, R35, R36 or R41</td>
<td>0</td>
</tr>
<tr>
<td>C (28)</td>
<td>a.s.&lt;0.0005 g/l</td>
<td>R36 or R41</td>
<td>0</td>
</tr>
<tr>
<td>CN (67)</td>
<td>a.s.&lt;0.1 g/l</td>
<td>R34 or R35</td>
<td>0</td>
</tr>
<tr>
<td>CNHHal (25)</td>
<td>a.s.&lt;0.1 g/l</td>
<td>R34 or R35</td>
<td>1</td>
</tr>
<tr>
<td>CNHHal (25)</td>
<td>a.s.&lt;0.004 g/l</td>
<td>R41</td>
<td>0</td>
</tr>
<tr>
<td>CNS (13)</td>
<td>a.s.&lt;0.006 g/l</td>
<td>R36 or R41</td>
<td>0</td>
</tr>
</tbody>
</table>

The number of compounds in each particular class is given in parentheses

* The percentage is calculated on the basis of the total number of substances in each particular class

According to Table 12, there are 3 incorrect predictions (false negatives):

- One substance has a.s. = 0 and is predicted as NOT R41, in contrast to the experimental classification (R41). Actually, this chemical was not tested for eye irritation but was labelled R41 on the basis of effects observed in the skin.
- Another substance has a.s. = 0 and is predicted as non-irritant (NOT R36), even though it is labelled R36 on the basis of experimental data. For this substance water solubility is not assignable because of immediate hydrolysis. It is therefore likely that the irritating effect is due to one of the products of hydrolysis.
- Another substance has a.s. = 0.0254 g/l and is predicted as not skin corrosive. This substance was not actually tested for skin irritation/corrosion but was labelled R35 due to its “intrinsic corrosive property”. This classification is probably due to the fact that the substance is a complex salt with hydrogen chloride (a strong acid).

The application of all rules based on aqueous solubility leads to predictions for 86 out of 186 test chemicals (i.e. 46% of the test set).

**Melting point**

The results from the external validation are given in the Table 13.
Table 13. External validation of the rules based on melting point

<table>
<thead>
<tr>
<th>Physicochemical rule</th>
<th>Prediction in the test set</th>
<th>Number of false negatives</th>
<th>Correct negative predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>IF</td>
<td>THEN NOT</td>
<td>Number</td>
</tr>
<tr>
<td>All (196)</td>
<td>m.p.&gt;200°C</td>
<td>R34 or R35</td>
<td>0</td>
</tr>
<tr>
<td>C (29)</td>
<td>m.p.&gt;55°C</td>
<td>R34 or R35</td>
<td>0</td>
</tr>
<tr>
<td>CNS (13)</td>
<td>m.p.&gt;50°C</td>
<td>R34 or R35</td>
<td>0</td>
</tr>
<tr>
<td>CNS (13)</td>
<td>m.p.&gt;200°C</td>
<td>R36</td>
<td>0</td>
</tr>
<tr>
<td>CHal (30)</td>
<td>m.p.&gt;65°C</td>
<td>R34 or R35</td>
<td>0</td>
</tr>
</tbody>
</table>

The number of compounds in the certain class is given in parentheses
* The percentage is calculated on the basis of the total number of substances in each particular class

There are no incorrect predictions based on the use of the melting point exclusion rules. In the class CNS, all compounds are covered by the specific rule because they have m.p. > 50°C. For comparison, 150 out of 186 CNS substances in the training set have m.p. > 50°C. However, this rule is not properly defined (see below), so a revision of the rule is recommended before it can be properly assessed by external validation.

**Combined application of the 31 exclusion rules**

In total, there were 3 incorrect classifications made using the set of 31 physicochemical exclusion rules on the 199 substances in the validation set. The overall results of the external validation in terms of correct and incorrect predictions are given in the Table 14.

Due to the specificity of the eye organ, it cannot be concluded that a substance is not corrosive, just because it is not irritant. For instance, according to the defined rules in DSS, chemicals with molecular weight > 650g/mol are predicted as non-eye irritants (NOT R36), but they may still elicit severe tissue damage resulting in local corrosion (R41) (7). Therefore, a chemical can be considered “safe” only if the assessment is based on a suitable combination of rules to exclude all risk phrases (R36, R34, R35, R41).

For this reason, in Table 14 only chemicals that predicted as not R36, R34, R35 or R41 by one or more rules and that are not labelled on the basis of experimental data are counted as correctly predicted. The percentage of these correct predictions gives an indication of the number of animal tests that could be replaced, based on the assumption that a single animal test is performed for each substance., and that a prediction of NOT R34/35/36/41 is a sufficient reason to omit an eye irritancy test. From the results of the external validation in this study, it is concluded that 10.05% of all Draize eye irritation tests could be avoided by using the BfR physicochemical exclusion rules. It is also conceivable that additional eye irritation testing could be avoided by using the inclusion rules based on structural alerts (not investigated in this study).
Table 14. External validation of the BfR exclusion rules using a test set of 199 substances

<table>
<thead>
<tr>
<th>Experimental classification</th>
<th>Predicted classification</th>
<th>No of substances</th>
<th>Percentage of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skin corrosive (R34, R35)</td>
<td>NOT R34 or R35 (incorrect prediction)</td>
<td>1</td>
<td>0.50%</td>
</tr>
<tr>
<td>Eye irritant (R36)</td>
<td>NOT R36 (incorrect prediction)</td>
<td>1</td>
<td>0.50%</td>
</tr>
<tr>
<td>Eye irritant (R41)</td>
<td>NOT R41 (incorrect prediction)</td>
<td>1</td>
<td>0.50%</td>
</tr>
<tr>
<td>Irritant/corrosive (R36, R41, R34, R35)</td>
<td>No prediction</td>
<td>42</td>
<td>21.11%</td>
</tr>
<tr>
<td>Not labelled for eye irritation/corrosion</td>
<td>NOT R34, R35, R36 or R41 (correct prediction)</td>
<td>20</td>
<td>10.05</td>
</tr>
<tr>
<td>Not labelled</td>
<td>No prediction</td>
<td>134</td>
<td>67.34</td>
</tr>
</tbody>
</table>

The exclusion rules were assessed on the basis that no prediction can be made for any substance to which no rule applies. Thus, on the basis of the exclusion rules alone, it is not possible to predict whether such substances are eye irritants or corrosives. Such predictions would require the additional consideration of the BfR inclusion rules based on structural alerts.

The specificity is the proportion of known negatives that are correctly predicted as negative. On the basis of the external validation results in Table 14, the total number of negatives = 134 + 20 = 154, and therefore specificity = 20/154 = 0.13. The fact that specificity is so low does not mean that the rules are not useful – it is simply a reflection of the large proportion of true negatives (substances that are not labelled) that are not covered by the set of rules.

A more meaningful statistic is the Negative Predictive Value (NPV). The NPV is the number of true negatives as a proportion of the number of negative predictions. It is a statistic that answers the question “If a substance is predicted as negative, what is the probability that it is actually negative?” On the basis of the external validation results in Table 14, NPV = 20/23 = 0.87.

Other statistics that are often reported for classification models (sensitivity, concordance and positive predictive value) are not meaningful in the context of this exercise (16).
Compliance of the exclusion rules with the OECD principles of (Q)SAR validation

Principles for (Q)SAR validation were proposed in 2002 at the international workshop on the “Regulatory Acceptance of QSARs for Human Health and Environment Endpoints” organised by the International Council of Chemical Associations (ICCA) and the European Chemical Industry Council (CEFIC) (16,17,18,19) and were further developed by the OECD QSAR Working Group (11). These principles were established to provide a structured basis for checking the validity of (Q)SAR models. In the present report the (Q)SAR validation principles were used to check the validity of the BfR rulebase for predicting the absence of eye irritation/corrosion. In the following section an evaluation for the each principle will be presented:

1 Defined endpoint

1.1 Does the model have a clearly defined scientific purpose (i.e. does it make predictions of a clearly defined physicochemical, biological or environmental effect)?

The BfR exclusion rules predict the absence of eye irritation/corrosion for industrial chemicals based on physical limit values.

Fulfilled.

1.2 Does the model have the potential to address (or partially address) a clearly defined regulatory need?

The defined physicochemical cut-off values predict the absence of eye irritation/corrosion, which is a clearly defined regulatory need.

Fulfilled.

1.3 Is information given about important experimental conditions that affect the measurements and therefore the prediction (e.g. sex, species, temperature, exposure period)?

The BfR DSS is based on experimental results obtained with a standardised test protocol in which the test substance is applied to the rabbit eye.

Fulfilled.

1.4 Are the units of measurements given?

The units of measurements of the eye irritation test are the Draize scores. Based on these scores chemicals are classified and labelled for eye irritation. The exclusion rules predict the absence of eye irritation/corrosion, without predicting the exact Draize scores on which classifications are based.

Not applicable.
2 Defined algorithm

2.1 In the case of a SAR is there an explicit description of the substructure, including an explicit identification of its substituents?

The SAR classes are described by empirical formulas, which are straightforward: e.g. class C, containing chemicals that only consist of C, H and/or O atoms; CHal class contains substances with C, H and/or O and Halogen atoms, etc. **Fulfilled.**

2.2 In the case of a QSAR, is the equation explicitly defined including definitions of all descriptors used?

The model is a set of decision rules based on cut-off values along quantitative descriptors. It gives the quantitative limit values outside which no classifiable eye irritation is expected. **Fulfilled.**

3 Domain of applicability

3.1 In the case of a SAR, is the substructure associated with any inclusion and/or exclusion rules on its applicability to groups of chemicals?

The exclusion rules are developed for all substances in the training set and also in the chemical classes, thus each class is associated with specific physicochemical rules. **Fulfilled.**

3.2 In the case of a SAR is the substructure associated with rules regarding the modulating effects of the substructure’s molecular environment.

The classes of chemicals for which separate rules are defined are empirically derived and have no (immediate) link with modulating effects on an active functionality in the structures. **Not applicable**, see also 3.1.

3.3 In the case of a QSAR, are the descriptor and response variables associated with inclusion and/or exclusion rules that define the variable ranges for which the QSAR is applicable?

The model predicts yes/no classification using four R phrases for irritation (R34/R35 for skin corrosive substances, R36 for eye irritant chemicals, and R41 for eye corrosive substances). These define the response values. The descriptor values are continuous values. They are formulated by cut-off values above or below which no classification is expected. **Fulfilled.**

3.4 A (graphical) expression of how the descriptor values of the chemicals in the training set are distributed in relation to the endpoints predicted by the model?

The present report gives such a graphical expression (Figures 2-6).
4a Internal performance

4a.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 confidential and are only available to the ECB and to the Competent Authorities of the EU member states.

Not yet fulfilled.

4a.2 If the data used to develop the model were based on the processing of raw data:
   a Is there an adequate description of the data processing (e.g. the averaging of replicate values)?

For each chemical one test has been performed. No averaging of replicate values has been carried out.

Not applicable.

   b Are the raw data provided?

The raw data are available.

Fulfilled.

4a.3 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. The DSS applies an empirical approach that shows at which descriptor value no classification is noticed. No algorithm to determine the cut-off values for specific parameters has been used.

Fulfilled.

4a.4 Is the QSAR associated with basic statistics for its goodness-of-fit to the training set?

True and false negative rates and negative predictivity are the only relevant statistics for this model (determined for the complete set of 31 rules).

Fulfilled.

4a.5 Is the QSAR associated with any statistics based on cross-validation or sampling?

Since the model is not based on the statistical fit to training set data, but depends solely on the choice of the physicochemical cut-off values, the calculation of cross-validation or sampling statistics does not seem very valuable.

Not applicable
4b External validation/Predictivity

4b.1 Does application of the appropriate statistical method(s) to the training set result in the same (Q)SAR model?

The results of the present report show that application of the same method results in the same model.

**Fulfilled.**

4b.2 An indication whether the model has been validated by using a test set that is independent of the training set?

There are two independent external validation of the BfR DSS. The first one is performed by Zinke *et al.* (5). The second external validation is performed in the frame of the presented project (see the Results) on the updated set of exclusion rules.

**Fulfilled.**

4b.3 If an external validation has been performed, is the following information available:

a) number of test structures;
b) the identity of the test structures;
c) the approach for selecting the test structures;
d) the statistical analysis of the predictive performance of the model? (e.g. including sensitivity, specificity, and positive and negative predictivities for classification models);
e) a comparison of the predictivity of the model against previously-defined quantitative performance criteria?

In the validation exercise included in the presented report the following information is available:

a) The number of test structures was available: 199.

**Fulfilled.**

b) The identity of the test structures was known.

**Fulfilled.**

c) The test structures were selected among the chemicals submitted after deriving the rules.

**Fulfilled.**

d) The present report shows the correct and incorrect predictions. (Sensitivity and specificity are of limited value because only the absence of effects is predicted).

**Fulfilled.**

e) The model was designed to not give any false negatives as the limit values were (mostly) put on the safe side from a precautionary point of view.

**Fulfilled.**
5  Mechanistic Interpretation

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)?
Not applicable.

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)?

Physicochemical limit values characterise the absence of the capacity of a chemical to penetrate eye tissues and/or the absence of biochemical reactivity considered necessary for relevant damaging potential to specific tissues in and around the eyeball.
Fulfilled.

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

This is addressed in references 6 and 7.
Fulfilled.

Discussion and Conclusions

In comparison with skin corrosion, the development of alternative tests and QSARs to replace the use of the Draize eye irritation test is more difficult. The reason is that the eye is a very complex organ and “eye irritation” and “eye damage” are very complex adverse effects based not only on local lesions to the conjunctiva, iris and cornea, but also on very specific lesions to the eye which may lead to impairment of sight. Several validation studies have been performed, in order to find alternative test methods appropriate for replacing the Draize eye test, but no single test or model has been found to be adequately predictive (20). The development of the BfR DSS is therefore expected to contribute to integrated testing strategies for skin and eye irritation/corrosion, which are based on in the combined use of different approaches.

The evaluation of the physicochemical exclusion rules derived from the training set of 1358 substances is summarised and discussed according to the physicochemical parameter used for the formulation of the rules. The rules are evaluated on the basis of two criteria: correct/incorrect definition and presence/absence of safety margin.

Molecular weight

A summary of the evaluation of the rules based on this physicochemical parameter is given in Table 15.
Table 15. Evaluation of the physicochemical rules based on molecular weight

<table>
<thead>
<tr>
<th>Physiochemical rule</th>
<th>Class a</th>
<th>IF</th>
<th>THEN NOT</th>
<th>Correctly defined?</th>
<th>Safety margin? b</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All (141)</td>
<td>m.w.&gt;650g/mol</td>
<td>R36</td>
<td>y</td>
<td>y (3g/mol)</td>
<td></td>
</tr>
<tr>
<td>C (53)</td>
<td>m.w.&gt;380g/mol</td>
<td>R34, R35, R36 or R41</td>
<td>y</td>
<td>y (10g/mol)</td>
<td></td>
</tr>
<tr>
<td>CN (256)</td>
<td>m.w.&gt;290g/mol</td>
<td>R34 or R35</td>
<td>y</td>
<td>y (10g/mol)</td>
<td></td>
</tr>
<tr>
<td>CNHal (80)</td>
<td>m.w.&gt;370g/mol</td>
<td>R34 or R35</td>
<td>y</td>
<td>n</td>
<td></td>
</tr>
<tr>
<td>CNS (42)</td>
<td>m.w.&gt;620g/mol</td>
<td>R34, R35 or R36</td>
<td>y</td>
<td>y (112g/mol)</td>
<td></td>
</tr>
<tr>
<td>CHal (23)</td>
<td>m.w.&gt;280g/mol</td>
<td>R34 or R35</td>
<td>y</td>
<td>y (4g/mol)</td>
<td></td>
</tr>
<tr>
<td>CHal (13)</td>
<td>m.w.&gt;370g/mol</td>
<td>R34, R35, R36</td>
<td>y</td>
<td>y (2g/mol)</td>
<td></td>
</tr>
</tbody>
</table>

Correctly/incorrectly defined cut-off values are designated by y/n, the presence/absence of safety margin is designated by y/n.

a The number of substances correctly classified by the rule is given in parentheses.
b The value of the safety margin is given in parentheses.

It can be concluded that the rules based on molecular weight are correctly defined with the cut-off values set with a safety margin. The only exception is the specific rule for the CNHal class which has no safety margin. Bearing in mind the small number of corrosives (8 substances) in this set, the representativeness of this rule is questionable.

Octanol-water partition coefficient

The summary of the evaluation of the rules based on this physicochemical parameter is given in Table 16.

Table 16. Evaluation of the physicochemical rules based on the octanol-water partition coefficient

<table>
<thead>
<tr>
<th>Physiochemical rule</th>
<th>Class a</th>
<th>IF</th>
<th>THEN NOT</th>
<th>Correctly defined?</th>
<th>Safety margin? b</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All (29)</td>
<td>log P &gt; 9</td>
<td>R34, R35, R36 or R41</td>
<td>y</td>
<td>n</td>
<td></td>
</tr>
<tr>
<td>(47)</td>
<td>log P &lt; -3.1</td>
<td>R34 or R35</td>
<td>y</td>
<td>n</td>
<td></td>
</tr>
<tr>
<td>CN(101)</td>
<td>log P &gt; 4.5</td>
<td>R34 or R35</td>
<td>y</td>
<td>y (0.3)</td>
<td></td>
</tr>
<tr>
<td>CNHal(56)</td>
<td>log P &gt; 3.8</td>
<td>R34, R35 or R41</td>
<td>n</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNS(38)</td>
<td>log P &lt; -2</td>
<td>R34 or R35</td>
<td>y</td>
<td>y (0.7)</td>
<td></td>
</tr>
<tr>
<td>CNS (71)</td>
<td>log P &gt; 1.5</td>
<td>R36</td>
<td>y</td>
<td>y (0.1)</td>
<td></td>
</tr>
<tr>
<td>CNS (37)</td>
<td>log P &gt; 3.6</td>
<td>R41</td>
<td>y</td>
<td>y (0.1)</td>
<td></td>
</tr>
<tr>
<td>CHal (9)</td>
<td>log P &gt; 4.5</td>
<td>R34, R35 or R41</td>
<td>y</td>
<td>y (0.8)</td>
<td></td>
</tr>
</tbody>
</table>

Correctly/incorrectly defined cut-off values are designated by y/n, the presence/absence of safety margin is designated by y/n.

a The number of substances correctly classified by the rule is given in parentheses.
b The value of the safety margin is given in parentheses.
The only rule that is not correctly defined is the specific rule for the class CNHal. One substance is not properly classified – it is labelled R41 and has log P = 7. Most of the rules have no or a narrow safety margin. For this reason, a redefinition of the cut-off values would improve the predictive ability of the rules.

**Lipid solubility**

The summary of the evaluation is given in the Table 17.

**Table 17. Evaluation of the physicochemical rules based on lipid solubility**

<table>
<thead>
<tr>
<th>Physiochemical rule</th>
<th>Cut-off value</th>
<th>Correctly defined?</th>
<th>Safety margin?</th>
</tr>
</thead>
<tbody>
<tr>
<td>All (59)</td>
<td>l.s. &lt; 0.01g/kg</td>
<td>R34 or R35</td>
<td>y</td>
</tr>
<tr>
<td>CN (56)</td>
<td>l.s. &lt; 0.4 g/kg</td>
<td>R34 or R35</td>
<td>y</td>
</tr>
<tr>
<td>CNHal (75)</td>
<td>l.s. &lt; 400 g/kg</td>
<td>R34 or R35</td>
<td>y</td>
</tr>
</tbody>
</table>

Correctly/incorrectly defined cut-off values are designated by y/n, the presence/absence of safety margin is designated by y/n

*a The number of substances correctly classified by the rule is given in parentheses

*b The value of the safety margin is given in parentheses

* High uncertainty in the determination of the rule

The rules based on lipid solubility have limited value because a poor dataset was used to develop the rule: in the CN class, only three corrosive substances have l.s. values different from 1000 g/kg, whereas in the CNHal class, there are only 3 corrosive substances and only one of them has a l.s. different from 1000 g/kg. Further this rule cannot be meaningfully validated by using the external test set because of limited number of substances with l.s. values (only 3 substances).

**Aqueous solubility**

The summary of the evaluation is given in the Table 18.
Table 18. Evaluation of the physicochemical rules based on aqueous solubility

<table>
<thead>
<tr>
<th>Physicochemical rule</th>
<th>Cut-off value</th>
<th>Correctly defined?</th>
<th>Safety margin?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class a</td>
<td>IF</td>
<td>THEN NOT</td>
<td></td>
</tr>
<tr>
<td>All (103)</td>
<td>a.s.&lt;0.00002 g/l</td>
<td>R41</td>
<td>n</td>
</tr>
<tr>
<td>All (38)</td>
<td>a.s.&lt;0.000005 g/l</td>
<td>R36</td>
<td>y, n</td>
</tr>
<tr>
<td>C (38)</td>
<td>a.s.&lt;0.0001 g/l</td>
<td>R34, R35, R36 or R41</td>
<td>y, y (0.000612 g/l)</td>
</tr>
<tr>
<td>C (72)</td>
<td>a.s.&lt;0.0005 g/l</td>
<td>R36 or R41</td>
<td>y, y (0.00922 g/l)</td>
</tr>
<tr>
<td>CN (231)</td>
<td>a.s.&lt;0.1 g/l</td>
<td>R34 or R35</td>
<td>y, n</td>
</tr>
<tr>
<td>CNHal (110)</td>
<td>a.s.&lt;0.1 g/l</td>
<td>R34 or R35</td>
<td>y, y (0.003 g/l)</td>
</tr>
<tr>
<td>CNHal (78)</td>
<td>a.s.&lt;0.004 g/l</td>
<td>R41</td>
<td>n</td>
</tr>
<tr>
<td>CNS (45)</td>
<td>a.s.&lt;0.006 g/l</td>
<td>R36 or R41</td>
<td>y, y (0.00044 g/l)</td>
</tr>
</tbody>
</table>

Correctly/incorrectly defined cut-off values are designated by y/n, the presence/absence of safety margin is designated by y/n.

a The number of substances correctly classified by the rule is given in parentheses.
b The value of the safety margin is given in parentheses.

One compound in class CNHal breaks two of the rules: it has a.s. = 0.000005 g/l and is labelled R41. The same compound also breaks one of the log P based rules (Table 16). This is not surprising bearing in mind the relationship between both aqueous solubility and log P. However the high number of correctly predicted substances (Tables 16 and 18) and also the satisfactory number of corrosives/irritants used in the definition of the cut-off values indicate that this compound should be considered as an outlier, and thus not taken into account in the rule definition.

Melting point

The summary of the evaluation, according to the criteria correct/incorrect definition and presence/absence of safety margin, is given in the Table 19.

Table 19. Evaluation of the physicochemical rules based on melting point

<table>
<thead>
<tr>
<th>Physicochemical rule</th>
<th>Cut-off value</th>
<th>Correctly defined?</th>
<th>Safety margin?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class a</td>
<td>IF</td>
<td>THEN NOT</td>
<td></td>
</tr>
<tr>
<td>All (234)</td>
<td>m.p.&gt;200°C</td>
<td>R34 or R35</td>
<td>n</td>
</tr>
<tr>
<td>C (92)</td>
<td>m.p.&gt;55°C</td>
<td>R34 or R35</td>
<td>n</td>
</tr>
<tr>
<td>CNS (148)</td>
<td>m.p.&gt;50°C</td>
<td>R34 or R35</td>
<td>n</td>
</tr>
<tr>
<td>CNS (57)</td>
<td>m.p.&gt;200°C</td>
<td>R36</td>
<td>y, y (23°C)</td>
</tr>
<tr>
<td>CHal (21)</td>
<td>m.p.&gt;65°C</td>
<td>R34 or R35</td>
<td>y, y (4°C)</td>
</tr>
</tbody>
</table>

Correctly/incorrectly defined cut-off values are designated by y/n, the presence/absence of safety margin is designated by y/n.

a The number of substances correctly classified by the rule is given in parentheses.
b The value of the safety margin is given in parentheses.
The exclusion rules based on melting point have cut-off values that do not cover all the eye irritants/corrosives in the training set: three out of five rules are not correctly defined. It is therefore recommended that the exclusion rules based on melting point are redefined.

**Overall evaluation of the 31 exclusion rules**

As a whole, the application of the 31 physicochemical exclusion rules to the training set (containing 343 irritant or corrosive substances) produces 7 misclassifications, which corresponds to a False Negative Rate of 2% (i.e. 7/343). The highest number of misclassifications (6 substances) is related to the physicochemical rules based on melting point.

The external validation of the physicochemical exclusion rules by using a test set of 199 substances (containing 45 irritant or corrosive substances) produced only 3 false negative predictions, which corresponds to a False Negative Rate of 7%. All three false negatives result from the use of the aqueous solubility parameter. Two of these chemicals were not actually classified on the basis of Draize eye irritant test but due to an “intrinsic corrosive property” or observed effects in the skin. For the third compound, the aqueous solubility cannot be assigned due to the immediate hydrolysis.

On the basis of the external validation, the Negative Predictive Value is 87%. In other words, if the 31 exclusion rules predict that a chemical is not R36/R41/R34/R35, then there is a probability of 0.87 that the chemical is not an irritant or corrosive.

On the basis of the evaluation performed several conclusions and recommendations can be made.

The rules based on molecular weight and octanol-water partition coefficient have a very important role in the DSS. The very similar distributions in training and test set are criteria for the reliability of the external validation. However, it is recommended to redefine the logP cut-off values to ensure higher safety margins for some of the log P rules.

The rules based on lipid solubility have limited value because of the poor dataset for determining of the rule. Moreover, these rules cannot be validated in the external test set because only 3 substances in the test set have measured lipid solubility values. It is recommended to exclude the rules based on lipid solubility from the DSS due to their limited applicability.

Some of the rules based aqueous solubility and melting point are either incorrectly defined or they lack a safety margin. It is suggested to re-define the cut-off values to cover all irritants/corrosives and to ensure an adequate safety margin.

There are some cut-off values which coincide with the highest or lowest physicochemical parameter value of the irritants/corrosives, which means a lack of safety margin. For other rules, safety margins are foreseen but they seem arbitrarily defined. The adoption of a consistent way of setting the cut-off values, including a safety margin is recommended. The calculation procedure proposed by Rorije & Hulzebos (9) could be applied.
In order to increase the practical usefulness of the DSS it is recommended to further develop the rulebase to specify the rules excluding all eye irritancy (NOT R36, R34, R35 and R41).

The limited number of compounds in some chemical classes could be a reason for the lower statistical significance of the specific rules obtained. In such cases, the integration of structurally related classes would increase the statistical quality of the defined rules.

A further validation exercise would increase the reliability of the obtained results. This could be done by expanding the test set with new substances notified in NCD. The selection should be directed to increasing the representativeness of the chemical classes, to ensure wider coverage of the structural space as well as the ranges of physicochemical values in the chemical classes.

Overall the DSS developed by BfR shows high internal and external performance. Using the set of the rules it seems possible to avoid about 10% of all Draize eye irritation tests performed for new substances.

**Acknowledgements**

The authors thank Dr Matthias Herzler (Bundesinstitut für Risikobewertung [BfR], Berlin, Germany) for his contributions. Ivanka Tsakovska acknowledges receipt of a post-doctoral grant from the European Commission’s Joint Research Centre (JRC contract 195025-2005-06 P1B30 ISP IT).
References


8. ECB website: http://ecb.jrc.it/


Appendix I: Box plots illustrating the distributions of physicochemical properties used in the definition of the BfR exclusion rules

Figure 1. Distribution of molecular weight values. In the training set substances with m.w. < 1500 g/mol are given

Figure 2. Distribution of octanol-water partition coefficient values
Figure 3. Distribution of log (aqueous solubility) values. Substances with a.s. = 0 (7 substances in the test set) are not included.

Figure 4. Distribution of melting point values.