Development of an *In Silico* Profiler for Respiratory Sensitisation

Steven J. Enoch, David W. Roberts, Judith C. Madden and Mark T.D. Cronin

*QSAR and Molecular Modelling Group, School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Liverpool, UK*

**Summary** — In this article, we outline work that led the QSAR and Molecular Modelling Group at Liverpool John Moores University to be jointly awarded the 2013 Lush Science Prize. Our research focuses around the development of *in silico* profilers for category formation within the Adverse Outcome Pathway paradigm. The development of a well-defined chemical category allows toxicity to be predicted via read-across. This is the central approach used by the OECD QSAR Toolbox. The specific work for which we were awarded the Lush Prize was for the development of such an *in silico* profiler for respiratory sensitisation. The profiler was developed by an analysis of the mechanistic chemistry associated with covalent bond formation in the lung. The data analysed were collated from clinical reports of occupational asthma in humans. The impact of the development of *in silico* profilers on the Three Rs is also discussed.

**Key words:** adverse outcome pathway, *in silico*, QSAR, read-across, respiratory sensitisation.

**Address for correspondence:** Steve Enoch, School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Liverpool L3 3AF, UK. E-mail: s.j.enoch@ljmu.ac.uk

**Introduction**

The advent of the seventh amendment to the Cosmetics Directive in the European Union (EU) has led to the banning of animal testing for new cosmetics ingredients (1). This legislation is extremely important in terms of placing the EU at the forefront of animal welfare for chemical safety assessment within this industry. Such legislation is also important, given the widespread use of Low Molecular Weight (LMW) chemicals (defined as chemicals with a molecular mass of less than 1,000g/mol) in the cosmetics industry, and the subsequent burden that animal testing would have imposed (in the absence of the Cosmetics Directive). However, it is also important that this legislation does not come at too high an economic cost, as the cosmetics sector in the EU is worth over €60 billion and employs over 150,000 people (with a further 350,000 in related employment; 2).

There are a number of regulatory endpoints that are of importance to the cosmetics industry, for which non-animal test methods will need to be developed in order to comply with the Cosmetics Directive. The need to develop alternative, non-animal, test methods for chemical safety assessment has led to the emergence of the Adverse Outcome Pathway (AOP) concept (3–8). An AOP relates a series of key events that link the initial interaction between a chemical and a biological system and an adverse effect at the organ level (which in turn can be linked to the biological system or even the population). The aim of an AOP is not to outline every minute detail of the biological pathway that is perturbed, leading to toxicity, but rather to outline the key processes that can be tested by using either *in silico*, *in chemico* or *in vitro* methods. The importance of this approach was recognised by the award of the 2012 Lush Science Prize to Brigitte Landesmann and co-workers for their work on hepatotoxicity (4).

Within the AOP paradigm, *in silico* approaches have focused on defining the chemistry associated with the initial interaction between a chemical and the biological system, the so-called Molecular Initiating Event (MIE; 9–11). The chemistry associated with the MIEs can be compiled into *'in silico'* profilers, which enable chemicals to be grouped into mechanism-based categories, allowing for predictions of toxicity to be made by using read-across (12–15). In addition, such profilers enable chemical inventories to be prioritised for further *in vitro* and/or *in chemico* investigations (rather than testing every chemical in an unstructured manner).

Perhaps unsurprisingly, the ability of a chemical to cause either skin and/or respiratory sensitisation is of particular interest to scientists developing new cosmetics products. In brief, an individual can become sensitised to a LMW chemical via an initial (dermal or inhalation) exposure (the induction phase), with subsequent re-exposure resulting...
in the observed toxicity (the elicitation phase), i.e. contact dermatitis in the skin and asthma-like symptoms in the lung (16). The ability to predict the likelihood of skin and respiratory sensitisation is especially important, given that once individuals’ immune systems are sensitised to a chemical, they may potentially remain sensitised to it for the remainder of their lives. The situation can be further complicated by the ability of chemicals that are activated to a common metabolite to cause cross-sensitisation (the metabolite of both chemicals being ultimately responsible for induction and elicitation). Finally, in terms of chemicals capable of causing respiratory sensitisation, there is evidence that the induction phase can take place in either the skin or the lung, with the elicitation phase taking place on re-exposure in the lung (17). These multiple factors combine to present a seemingly complex mechanistic picture that underpins the likelihood that a LMW chemical will result in either skin and/or respiratory sensitisation.

The availability of historical data from tests, including the Local Lymph Node Assay (LLNA) and (to a lesser extent) the Guinea-Pig Maximisation Test (GPMT), has led to a significant mechanistic understanding of skin sensitisation (18–20). This mechanistic knowledge has been recently reviewed and published as an AOP (21, 22). The AOP outlines the evidence that the formation of a covalent bond between a LMW chemical and a protein is the MIE for skin sensitisation (23, 24). In terms of chemistry, this means that, in order for a chemical to cause skin sensitisation, it must either be directly electrophilic or be activated (metabolised or oxidised) to an electrophile. Mechanistic knowledge, combined with the availability of toxicological data, has allowed several in silico profilers to be developed (24–28). In contrast, the situation for respiratory sensitisation is, in terms of mechanistic understanding, significantly less well-defined, in the main due to the lack of a predictive animal assay. The lack of (publicly available) historical animal data has prevented the type of analysis being undertaken that is required in order to develop an in silico profiler for respiratory sensitisation.

**Structural Alerts, In Silico Profilers and Toxicological Data**

As we have stated, the availability of toxicological data is the key factor in the development of in silico profilers. These data are important, as they allow us to develop a hypothesis with regard to what the MIE for a given endpoint is likely to be. Subsequent analysis allows us to develop a series of structural alerts (molecular fragments related to the MIE) that together form a profiler. A key aspect, which might be important for regulatory acceptance, is that each structural alert is supported by peer-reviewed literature that relates the chemistry to the MIE. In terms of respiratory sensitisation data, a number of publications have attempted to develop statistically-based Quantitative Structure–Activity Relationship models (29–32). However, these models (and statistical models in general) have met with very limited regulatory acceptance, due to their lack of transparency in terms of the underlying statistical algorithm and in relation to the biochemical mechanism leading to toxicity. Despite this, these publications gave us access to 40 chemicals associated with reports of respiratory sensitisation in humans. These data were typically drawn from clinical reports of individuals being sensitised to LMW chemicals in the workplace (also known as occupational asthma). In addition to these published data, we were also able to gain access to an additional unpublished data set of over 60 chemicals from colleagues at the University of Manchester, UK, giving us a final data set of 104 LMW chemicals associated with respiratory sensitisation in humans.

**Development of an In Silico Profiler for Respiratory Sensitisation**

The major part of the work for which we were jointly awarded the 2013 Lush Science Prize, related to the development of an in silico profiler for respiratory sensitisation. The profiler, which has been recently accepted for inclusion in the next version of the OECD QSAR Toolbox, was developed from a mechanistic chemistry analysis of the data set of 104 LMW chemicals outlined above. Our initial interest in this area of research stemmed from a study where we showed that, for a small number of respiratory sensitisers, the most likely MIE was the formation of a covalent bond in the lung (33). We then outlined how such mechanistic information could be used to predict respiratory sensitisation by read-across for a second, slightly larger, data set of chemicals (34). In both studies we outlined the importance of the underlying mechanistic chemistry as the guiding principle in the process of grouping chemicals. Our research suggested that two key factors drive the MIE for respiratory sensitisation: chemical reactivity (electrophilicity) and the ability to cross-link proteins (Figure 1). Additionally, we suggested that a highly electrophilic chemical can cause sensitisation without the need for protein cross-linking (for example, cyanocrylates). In contrast, a chemical that is less electrophilic requires multiple reactive centres resulting in protein cross-linking (for example, diisocyanates). Our mechanistic rationale for the MIE offered a significant improvement on the previous hypothesis that all chemicals that cause respiratory sensitisation must have multiple reactive centres (17, 29, 30, 32).

The availability of the larger data set of respiratory sensitisation data enabled us to extend our
analysis, allowing us to outline the detailed mechanistic chemistry associated with the MIE for LMW chemicals (35). The analysis enabled us to identify and publish a set of 52 structural alerts that defined the chemistry associated with covalent protein binding in the lung. An important aspect is the availability of the associated metadata for each structural alert, which documents the reaction mechanism and supporting peer-reviewed literature. This information is of central importance for profilers, if they are going to be used to group chemicals together in regulatory toxicology (for example, when using the OECD QSAR Toolbox).

The type of mechanistic chemistry analysis that we used for the development of our *in silico* profiler, also enabled us to explore areas of closely related chemical space for which no toxicological data are available. In practice, this allows us to either extend a structural alert (which has data associated with it) to cover closely related chemicals, or to define multiple structural alerts covering both the parent and metabolite structures. This is the reason why, on first inspection, it often seems as if too many structural alerts are defined from a given set of data; for example, we recently published a set of 52 structural alerts in the respiratory sensitisation profiler from a data set of only 104 chemicals (35). As an example of our approach to this area of profiler development, let us consider the chemical hexamine, which is reported in the data set as being a respiratory sensitiser in humans. Our analysis of the mechanistic chemistry for this chemical showed that it readily releases formaldehyde, which is capable of cross-linking proteins (Figure 2).

The mechanistic information that we gained from the chemistry of hexamine toxicity allowed us to define other chemicals also likely to cause respiratory sensitisation due to the release of formaldehyde. It is worth remembering that these chemicals are not associated with toxicological data; instead, they are related through a common MIE. Importantly, we make it clear to the user of the profiler that there are no toxicity data in the associated metadata for the alerts defined from these chemicals. In our profiler,

---

**Figure 1:** Examples of chemicals able to cause respiratory sensitisation due to electrophilicity alone (cyanoacrylates) or requiring multiple centres to be capable of protein cross-linking (di-isocyanates)

**Figure 2:** Bio-activation of hexamine to formaldehyde and subsequent Schiff base reactions resulting in protein cross-linking and potential respiratory sensitisation

\[ R = \text{carbon.} \]
we were able to define an additional six structural alerts (including hexamine) for chemicals capable of releasing formaldehyde (Figure 3).

**Profiling levels: Mechanistic domains, mechanistic alerts and structural alerts**

In keeping with the previous profilers we have developed (9, 10, 13), we have recently assigned each of the 52 published structural alerts (35) identified from the mechanistic analysis of the respiratory sensitisation data set, to two additional tiers: mechanistic domains and mechanistic alerts (13). The three profiling tiers were defined as follows:

— **Mechanistic domain**: One of the six general reaction mechanisms, as defined by Aptula and Roberts (36), these being: acylation, Michael addition, Schiff base formation, \( S_N 1 \), \( S_N 2 \), and \( S_NAr \).

— **Mechanistic alert**: One or more structural alerts grouped together, based on the presence of a common reaction site. One or more structural alerts grouped together, based on each of the structural alerts being activated to a common electrophile responsible for reactivity.

— **Structural alert**: A fragment in a molecule related to covalent protein binding.

Thus, within each mechanistic domain, mechanistic alerts were created on the basis of the presence of a common reactive centre (the site of attack by a biological nucleophile), as defined by a group of structural alerts. For example, alkenes acting as Michael acceptors, due to the influence of a polarising moiety, were grouped into a mechanistic alert based on the presence of the \( sp^2 \) carbon atom as the target of nucleophilic attack. In addition, chemicals that had been shown to be activated to a common electrophile, were also grouped into a single mechanistic alert — for example, a mechanistic alert consisting of the six structural alerts, shown in Figure 3, that release formaldehyde (formaldehyde being the common electrophile responsible for toxicity). This new analysis has enabled us to improve the definition of a number of the published structural alerts in terms of ensuring a single alert identifies a single chemical class. This resulted in an updated set of 66 structural alerts. The breakdown of mechanistic domains, mechanistic alerts and structural alerts for this updated version of the respiratory sensitisation profiler, is detailed in the supplementary information (SMARTS patterns are also included).

The importance of being able to profile chemicals at either the mechanistic alert or structural alert level is best illustrated by using two example chemicals for which no respiratory sensitisation data are available. Imagine a situation where we are trying to assess the respiratory sensitisation potential of propyl cyanoacrylate and \( p \)-tolylmethoxymethanol. Our first step is to profile these target chemicals for information about their poten-

---

**Figure 3:** Structural alerts defined for chemicals capable of releasing formaldehyde
tial ability to react covalently in the lung (Table 1). The results of this profiling show that propyl cyanoacrylate is capable of undergoing a Michael addition reaction due to the polarised alkene moiety (Figure 4). Subsequent profiling of the respiratory sensitisation data set at the structural alert level identifies two analogues, methyl cyanoacrylate and ethyl cyanoacrylate, which have both been reported as being sensitisers. The presence of the common cyanoacrylate alert allows us to make a read-across prediction that propyl acrylate is also likely to be a respiratory sensitisier. In this example, there is no need to go to the next, less strict, mechanistic alert level of profiling, as we have been able to identify analogues at the structural alert level.

Inspection of the profiling results for the second target chemical, \( p \)-tolylmethoxymethanol, show it to be capable of reacting with proteins via a Schiff base mechanism, due to its ability to release formaldehyde (mechanisms as shown in Figures 2 and 3). However, for this chemical there are no analogues in the data set that contain the same benzylhemiformal structural alert. It is in cases such as this that the additional mechanistic alert level of profiling becomes useful, as it allows us to identify chemicals that do not share a common structural alert, but that instead share very closely related chemistry. For \( p \)-tolylmethoxymethanol, profiling at the mechanistic alert level allows us to identify a single analogue, hexamine, that has been shown to cause respiratory sensitisation in humans. This allows us to predict that \( p \)-tolylmethoxymethanol is likely to be a sensitisier based on a one-to-one read-across. It is important to state that within the AOP paradigm both read-across predictions would, in an ideal scenario, be supported with additional data drawn from \textit{in chemico} and/or \textit{in vitro} assays (Table 1).

Our general approach to using \textit{in silico} profilers is summarised in the flowchart in Figure 5. The flowchart includes an additional step, outlining a method for dealing with chemicals that do not contain a structural alert related to the formation of a covalent bond in the lung. For chemicals such as these, our suggestion is to profile them for simple

![Figure 4: Michael addition mechanism for chemicals containing a cyanoacrylate structural alert](image)

Table 1: Chemical structures and \textit{in silico} profiling results for propyl cyanoacrylate and \( p \)-tolylmethoxymethanol

<table>
<thead>
<tr>
<th>Name</th>
<th>Mechanistic domain</th>
<th>Mechanistic alert</th>
<th>Structural alert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Propyl cyanoacrylate</td>
<td>Michael addition</td>
<td>Polarised alkenes 5</td>
<td>Cyanoacrylates 2</td>
</tr>
<tr>
<td>( p )-Tolylmethoxymethanol</td>
<td>Schiff base</td>
<td>( p )-Formaldehyde 1</td>
<td>Benzyloxymethanols 0</td>
</tr>
</tbody>
</table>

The number of analogues, excluding the target chemical, identified in the data set at each profiling level are shown in bold.
organic functional groups (such profilers are available in tools such as the OECD QSAR Toolbox). Structural analogues can then be identified, based on the common functional groups. This final profiling tier also offers the opportunity to identify chemicals that do not cause respiratory sensitisation, by using negative skin sensitisation data. If such a group of chemicals were all negative in the LLNA, then this could be used as evidence for the absence of protein reactivity. In turn, this could be used as an indication that they are also unlikely to result in sensitisation in the lung via covalent bond formation. Again, such predictions should be used as part of a weight-of-evidence approach to chemical risk assessment. However, it is noted that there is no formal relationship between the absence of skin sensitisation and of respiratory sensitisation; it is at this point in a weight-of-evidence approach that further information from the tests suggested by the AOP could be valuable. A recent article builds on our work in this area, outlining how mechanistic information, including in chemico data relating to protein reactivity, can be used to devise an AOP-driven testing strategy for respiratory sensitisation (37).

**Replacement of Animals in Regulatory Toxicology**

Our work in developing in silico profilers, and specifically, a profiler for respiratory sensitisation, offers tools that can be used as part of the AOP approach to chemical risk assessment. As we have outlined, in silico profilers encode the mechanistic information associated with the MIE for organ toxicity (9, 10, 13, 26, 35). This information can then be used to group chemicals together, and to make predictions via read-across, a process that has been supported at the OECD within the development of the OECD QSAR Toolbox. Currently, the LLNA assay is sometimes used to assess the potential of a chemical to cause respiratory sensitisation. This approach, at worst, is using a different organ (skin versus lung) in a different species (mouse versus human) to assess a chemical’s ability to sensitise the human lung. It is worth noting that variations of the assay, involving exposure via inhalation, are also possible. Taking a broader view, the development of AOPs will lead to more-detailed mechanistic knowledge and to the development of better, more-relevant, non-animal

---

**Figure 5: Flowchart illustrating how the respiratory sensitisation profiler can be used to form chemical categories and make predictions by using read-across**

1. Is a known covalent mechanism present in the target chemical?  
2. Are there any known respiratory sensitisers that contain the same structural alert as the target chemical?  
3. Are there any known respiratory sensitisers that contain the same mechanistic alert as the target chemical?  
4. Can the presence of a common functional (e.g., a simple alcohol) group be used to identify suitable analogues?

Use these chemicals to form a category and predict respiratory sensitisation potential via read-across

Category formation not possible for the target chemical
assays. To this end, we are currently involved in an international working group that is developing an AOP for respiratory sensitisation. However, despite the progress that the AOP paradigm will surely bring to regulatory risk assessment, at the time of writing, within well-defined limitations, chemistry-driven \textit{in silico} profilers offer one of the key solutions to the problem of making predictions of organ toxicity, when no other data or information are available.

**Conclusions, Future Outlook and Perspectives**

In this article, we have outlined the development of an \textit{in silico} profiler for respiratory sensitisation, for which we were jointly awarded the 2013 Lush Science Prize. Our approach is based around a detailed understanding of the MIE leading to organ-level toxicity, allowing for chemical category formation and read-across. The approach is in-line with the OECD QSAR Toolbox (for which we have developed a number of the profilers), and is currently one of the key methods to make predictions of toxicity without the use of animals. However, significant challenges remain, especially when we look to developing alternative methods for toxicity endpoints associated with repeated-dose exposure. Specifically, in terms of \textit{in silico} approaches, there is a requirement for more-detailed profilers that encode the mechanistic information leading to toxicity in differing organs, for example, the liver, kidney and the heart. Therefore, we have used the Lush Prize money to co-fund, with Liverpool John Moores University, a PhD studentship for work on developing new and improved \textit{in silico} profilers for organ-level toxicity. In the wider predictive toxicology field, this work needs to be matched with the development of new \textit{in chemico} and \textit{in vitro} assays that enable other key events in AOPs to be investigated. This will allow risk assessment to be carried out by using a weight-of-evidence approach that uses a variety of non-animal test data.

**Acknowledgements**

Professor Raymond Agius, Dr Martin Seed and Dr Jill Stocks, from the Centre for Occupational and Environmental Health at the University of Manchester, UK, are gratefully acknowledged for their efforts in collating the original data set. The research outlined in this article was funded, in part, by the following: the European Community’s Seventh Framework Programme (FP7/2007-2013) COSMOS Project under Grant Agreement Number 266835 and from Cosmetics Europe, European Chemicals Agency Service Contract No. ECHA/2008/20/ECA.203, and a framework contract (Agreement Number 300000801) with the Organisation for Economic Co-operation and Development (OECD).

**References**


