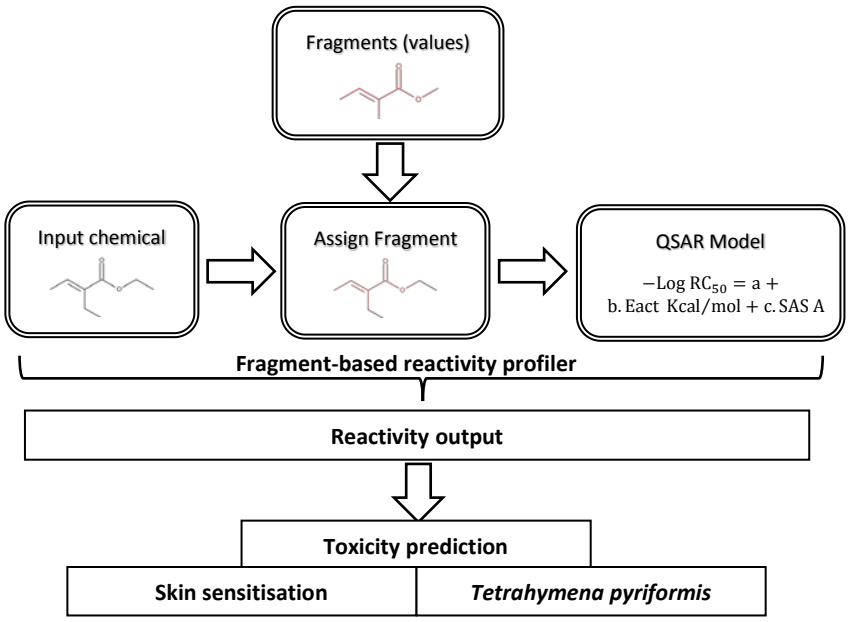
Validation of a Fragment-Based Profiler for Thiol Reactivity for the Prediction of Toxicity: Skin Sensitisation and *Tetrahymena pyriformis*

D. J. Ebbrell [†], J. C. Madden [†], M. T. D. Cronin [†], T. W. Schultz [‡] and S. J Enoch [†*]

[†] School of Pharmacy and Bimolecular Sciences, Liverpool John Moores University, Byrom Street, Liverpool, L3 3AF, England.

[‡] Department of Comparative Medicine, College of Veterinary Medicine, The University of Tennessee, Knoxville, Tennessee, USA

*Corresponding author
Tel: + 44 151 231 2164
Fax: + 44 151 231 2170
Email: s.j.enoch@ljmu.ac.uk


Fragments (values)
Input chemical
Assign Fragment
QSAR Model
$-\text{Log RC}_{50} = \text{a} + \text{b. Eact Kcal/mol} + \text{c. SAS A}$
Fragment-based reactivity profiler
Reactivity output
Toxicity prediction
Skin sensitisation
Tetrahymena pyriformis

Abstract

This study outlines the use of a recently developed fragment-based thiol reactivity profiler for Michael acceptors to predict toxicity towards *Tetrahymena pyriformis* and skin sensitisation potency as determined in the Local Lymph Node Assay (LLNA). The results showed that the calculated reactivity parameter from the profiler, -log RC_{50}(calc), was capable of predicting toxicity for both endpoints with excellent statistics. However, the study highlighted the importance of a well-defined applicability domain for each endpoint. In terms of *Tetrahymena pyriformis* this domain was defined in terms of how fast or slowly a given Michael acceptor reacts with thiol leading to two separate quantitative structure-activity models. The first, for fast reacting chemicals required only –Log RC_{50}(calc) as a descriptor, whilst the second required the addition of a descriptor for hydrophobicity. Modelling of the LLNA required only a single descriptor, -log RC_{50}(calc), enabling potency to be predicted. The applicability domain excluded chemicals capable of undergoing polymerisation and those that were predicted to be volatile. The modelling results for both endpoints, using the –log RC_{50}(calc) value from the profiler, were in keeping with previously published studies that have utilised experimentally determined measurements of reactivity. This results demonstrate the output from the fragment-based thiol reactivity profiler can be used to develop quantitative structure-activity relationship models where reactivity towards thiol is a driver of toxicity.

Introduction

It is well established that various toxicological effects can occur as a result of covalent bond formation between electrophilic chemicals and biological nucleophiles such as lysine and cysteine groups of proteins. This includes toxicological effects in both humans and environmental species, for example skin sensitisation or aquatic toxicity.[1-6] One mechanism resulting in covalent bond formation is Michael addition. Chemicals that act via Michael addition (known as Michael acceptors) are typically organic chemicals that contain a π-bond adjacent to a polarising group, such as a carbonyl.[7] This results in a partial positive charge on the β-carbon of the π-bond, causing the electrophilic chemical to become susceptible to a reaction with a biological nucleophile with either a negative charge or a lone pair of electrons. [8, 9] This nucleophilic attack at the β-carbon of the Michael acceptor results in a resonance stabilised carbanion intermediate, with a negative charge residing on the α-carbon. This α-carbon is then protonated to produce the final product (known as a Michael adduct) (Figure 1).

Knowledge of this mechanism has allowed for the development of structural alerts to identify chemicals that may act via Michael addition, and consequently have the potential to cause toxicological effects. [8, 9] Structural alerts can be grouped together to form the basis of an '*in silico* profiler' for mechanisms associated with specific toxicological outcomes, such as the structural alerts developed to identify the potential mechanism of action for skin sensitisation.[10] Whilst *in silico* profilers are useful for identifying features associated with potential toxicity the information they provide is qualitative (i.e. a binary yes or no for the presence of a structural feature); they provide no information concerning toxicological potency. When using knowledge of covalent mechanisms to predict toxicological potency, a primary assumption is that the rate of covalent bond formation (reactivity) is proportional to toxicity.[11] As a result of this assumption, there has been an increase in the number of studies focused on predicting potency using computational methods and/or *in chemico* reactivity measurements (i.e. experimental reactivity measurements that do not require the use of laboratory animals). A common experimental approach is the measurement of depletion of reactive peptides (such as glutathione) upon exposure to the test chemical over a fixed time period.[2] There have been many experimental studies which have successfully linked reactivity, as measured in an *in chemico* assay, to toxicity e.g. to *Tetrahymena pyriformis* measured in the *in vitro Tetrahymena pyriformis* growth impairment assay.[12-16] Similarly, results of kinetic peptide depletion assays have also been used in the prediction of skin sensitisation potency.[3, 4] Previous studies have also utilised Hammett and Taft descriptors to model chemical reactivity for the prediction of skin sensitisation.[17] These descriptors were derived from extensive studies into the effect of substituents upon the acidic dissociation constant (pKa) in model acid systems. These efforts further demonstrate the possibility that potency can be predicted for reactive chemicals within well-defined mechanistic domains.

A number of approaches have been published that make use of chemical descriptors derived from computational (*in silico*) approaches aimed at quantifying chemical reactivity. These are typically derived from quantum mechanics calculations, and include descriptors, such as energy values of the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) and the electrophilic index (ω).[18] These descriptors are then used to relate the electronic properties of the test chemical to their reactivity or to their toxicity, directly. However, these descriptors quantify only the electronic portion of chemical reactivity and account for factors such as steric hindrance at the reactive site.[6] Another common descriptor is the energy of activation (Eact) in which the energy difference between a test chemical and a model nucleophile, (with its respective transition state structure) is used. This has been performed successfully for the prediction of both aquatic toxicity and skin sensitisation.[19, 20] Importantly, this type of descriptor offers the advantage that it accounts for both electronic and steric factors involved in chemical reactivity, with studies showing that this approach capable of predicting potency for aquatic toxicity and skin sensitisation. However, the derivation of Eact requires is reliant on the quantum mechanics calculations capable of 'mapping' out the reaction pathway including the identification of key intermediates and/or transition state structures. This can be a time-consuming process, requiring significant expertise in the application of such methods.

Given the challenges of utilising quantum mechanics calculations to derive Eact values for use in predictive toxicology a recent study by the current authors showed that it is possible to predict experimentally derived reactivity towards glutathione (expressed as $-\log RC_{50}$) through the use of fragments with pre-calculated Eact values for Michael acceptors.[21] This approach involved defining the length of alkyl chain of the Michael acceptor beyond which further increases failed to significantly increase the activation energy. This enabled appropriate fragments to be generated which could be stored in a database along with pre-calculated activation energy values. The methodology was encoded as a KNIME workflow through which chemicals of interest can inputted using SMILES strings and are then compared to the fragments encoded as SMARTS patterns. The fragments are associated with their corresponding Eact values and an additional parameter that models the solvent accessible surface (SAS) at the α-position of the Michael acceptor. Once the query chemical has been assigned a fragment, its corresponding Eact and SAS values are used to predict its reactivity (expressed as $-\text{Log } RC_{50}$ values) based on a previously developed QSAR model; this process is summarised in Figure 2. Therefore given the availability of a fragment-based profiler, the aim of this study was to validate the calculated $-\text{Log } RC_{50}$ values generated from the fragment based reactivity profiler for thiol reactivity in predicting toxicity to *Tetrahymena* pyriformis and skin sensitisation potency (as determined in the LLNA) for Michael acceptors.

Methods

Computational methods

The previously published fragment-based reactivity profiler for thiol reactivity was utilised in the current study to predict reactivity towards a thiol nucleophile (defined a –log RC_{50}(calc)).[21] Briefly, this profiler was developed from a set of linear Michael acceptors with experimentally determined RC_{50} values, where the RC_{50} is the concentration of the electrophile required to deplete the concentration of glutathione by 50% over a fixed two hour time period.[22] The fragment-based reactivity profiler was trained on a set of polarised aldehydes, ketones, and esters with varying alkyl and aryl substitutions (Figure 3).[21]

The –Log RC_{50}(calc) values for chemicals in the *Tetrahymena pyriformis* and skin sensitisation datasets were generated using a previously developed KNIME workflow encoding the fragment-based reactivity profiler for thiol reactivity (this workflow, including calculated fragments is available from the authors on request).[21] The workflow utilises a database of fragments with pre-calculated activation energy values (Eact) calculated using Density Functional Theory (DFT) at the B3LYP/6-31G+(d) level of theory (calculations performed using Gaussian09 and with water as a solvent).[23] The workflow is summarised in Figure 2. Descriptors for hydrophobicity (Log K_{ow}) and vapour pressure (Log VP) were calculated using the KOWWIN (V1.68) and MPBPWIN (V1.43)modules of EPI suite .[24]

Datasets for *Tetrahymena pyriformis* and skin sensitisation

A set of 62 Michael acceptors from a database of 2072 chemicals with experimental toxicity values to *Tetrahymena pyriformis* were identified as being within the applicability domain of the of the fragment-based thiol reactivity profiler (defined in Figure 3).[25] These toxicity data were obtained using an *in vitro* assay, which quantifies 50% growth inhibition of the ciliate *Tetrahymena pyriformis* over a 40-hour exposure period to the test chemical (also recorded as EC_{50} values).[26] A similar analysis of skin sensitisation data gathered from the Local Lymph Node Assay (LLNA) resulted in a dataset of 38 Michael acceptors within the applicability of the fragment-based thiol reactivity profiler.[27-29] The LLNA is an *in vivo* based assay in which the stimulation of the lymph nodes of mice is measured upon exposure to a test chemical. The recorded value is the concentration required to elicit a three-fold stimulation in the lymph nodes, this is reported as an EC3 value (% weight) for the chemical. If the chemical does not produce a threefold stimulation it is not considered a sensitiser. All EC3 values were converted to pEC3 values (Equation 1). As the test vehicle is known to influence pEC3 values, only chemicals for which the vehicle was recorded to be Acetone: olive oil, AOO 4:1) were included in the analysis, this resulted in final dataset of 27 skin sensitising chemicals. [30]

$$\mathrm{pEC3} = \mathrm{Log}\,(\mathrm{EC3}/\mathrm{Molecular\ weight}) \qquad (1)$$

Statistical analysis

Linear regression analysis was used to develop quantitative structure-activity relationship models to obtain correlations between calculated –Log RC_{50} values and toxicity values using the Minitab (version 17) statistical software. Outliers were identified following linear regression analysis as chemicals with large standardised residuals as identified by Minitab. Chemicals for which a mechanistic rationale enabling outlying behaviour to be explained were subsequently removed from the analysis.

Results and Discussion

The aim of this study was to investigate the ability of a recently published fragment-based thiol reactivity profiler to predict toxicity of Michael acceptors towards *Tetrahymena pyriformis* and the LLNA.[25, 27-29] Analysis of the *Tetrahymena pyriformis* data within the applicability domain of the fragment-based thiol reactivity profiler resulted in a dataset of 62 chemicals (14 Aldehydes, 12 Ketones and 36 esters) with corresponding EC_{50} values (Table1). Initial modelling using the –Log RC_{50}(calc) values alone showed a clear trend (R^2 = 0.45) between reactivity and toxicity to *Tetrahymena pyriformis* (Model 1 in Figure 5 using equation 3). Interestingly, this value is lower than that published on a dataset of 41 Michael acceptors using experimentally determined glutathione depletion data (R^2 = 0.85).[2] However, in comparison with the current study (using –log RC_{50}(calc) as a measure of reactivity) this study using experimental reactivity data also failed to predict the toxicity to *Tetrahymena pyriformis* of slow reacting chemicals such as methacrylate esters. It was suggested that for these chemicals toxicity is driven by both hydrophobicity and reactivity due to them reacting slowly with proteins.[2]

$$\text{Log } (1/EC_{50}) = 0.63 + 0.61 \text{ –Log } RC_{50}\text{(calc)} \qquad \text{(Model 1)}$$

$$N = 62,\ R^2 = 0.45,\ R^2\text{-adj} = 0.44,\ s = 0.46$$

Consistent with this hypothesis a related study showed splitting the data into fast reacting and slow reacting classes resulted in significantly improved modelling results.[5] Importantly, the toxicity to *Tetrahymena pyriformis* for the fast reacting chemicals could be predicted from experimental reactivity alone, whilst those in the slow reacting class required both hydrophobicity and reactivity. The authors suggested a reactivity cut-off to distinguish the two classes based on equation 2, where chemicals with a $D_{kk} < 3$ being fast reacting and those with $D_{KK} > 3$ being slow reacting. Applying these criteria to the current dataset, using –Log RC_{50}(calc) as a measure of reactivity resulted in models 2 and 3 (fast and slow reacting chemicals respectively). Forty three chemicals were assigned to the fast reacting class (chemicals 1 – 43 in Table 1), whilst 23 chemicals were assigned to the slow reacting class (chemicals 44-62 in Table 1). In keeping with the previously published work using experimentally determined reactivity data, toxicity to *Tetrahymena pyriformis* for the chemicals in the fast reacting class required

only –log RC_{50}(calc) (model 2a), whilst the chemicals in the slow reacting class required both –log RC_{50}(calc) and log K_{ow} (model 2b). Figure 5 shows the correlation plots for models 2 and 3.

$$D_{KK} = \text{Log}\,(K_{ow}/-\text{Log}\,RC_{50}(\text{calculated})) = LogK_{ow} - -\text{Log}\,RC_{50}(\text{calculated}) \qquad \textbf{(2)}$$

$$\text{Log}\,(1/EC_{50}) = 0.41 + 0.94 - \text{Log}\,RC_{50}(\text{calc}) \qquad \text{(Model 2a)}$$
$$N = 43,\ R^2 = 0.78,\ R^2\text{-adj} = 0.77,\ s = 0.30$$

$$\text{Log}\,(1/EC_{50}) = -1.82 + 0.35 - \text{Log}\,RC_{50}\,(\text{calculated}) + 0.89\,\text{Log}K_{ow} \qquad \text{(Model 2b)}$$
$$N = 19,\ R^2 = 0.85,\ R^2\text{-adj} = 0.83,\ s = 0.31$$

Prediction of skin sensitisation potency as defined in the LLNA

The rate of covalent bond formation has also been shown to be important for the prediction of skin sensitisation potency as determined in the LLNA using both experimental and computational measures of reactivity.[3, 4, 6, 19] In keeping with these studies, the fragment-based reactivity algorithm was used to predict pEC3 values for the 26 Michael acceptors within the previously defined applicability domain. These chemicals are shown in Table 2. An initial analysis of the correlation between pEC3 and –LogRC_{50}(calc) resulted in extremely poor statistics (equation 7). Despite this, 13 of the chemicals were predicted within a twofold error of the corresponding experimental value (chemicals with a predicted value within 0.3 log units of the experimental value). These predictions are within the experimental twofold error of the LLNA.[31] Any chemicals outside of the two-fold error of the experimental assay were consider as outliers (labelled in Table 2) and were analysed to rationalise the error in their predictions.

$$\text{Predicted pEC}_3 = 1.35 + -0.05 - \text{Log}\,RC_{50}(\text{calc}) \qquad \text{(model 3)}$$
$$N = 26,\ R^2 = 0.00,\ R^2\text{-adj} = 0.00,\ s = 0.3$$

The majority of compounds with the largest errors are chemicals that are volatile, with the majority of these being acrylates and methacrylates (chemicals 1-5 in Table 2). Previous research has shown that the skin sensitisation potency of these volatile chemicals is less than might be expected based on their experimentally determined chemical reactivity.[3] In addition, research has also suggested that the acrylate and methacrylates chemicals are susceptible to polymerisation driven by free radical chemistry in the skin.[32, 33] Interestingly, the toxicity of a large number of similar chemicals towards *Tetrahymena pyriformis* were well predicted (chemicals 23-62 in Table 1). This highlights the importance of defining the applicability domain of any predictive model (experimental or computational) based on a detailed

understanding of the mechanistic chemistry of the assay. This mechanistic rationale resulted in the removal of a total of six volatile chemicals (chemicals 1-5 and 10), and two additional acrylates (chemicals 13 and 24). Three of these chemicals were removed despite being relatively well-predicted (chemicals 10, 13 and 24) as no mechanistic rationale could be offered as to why they were correctly predicted compared to the other chemicals identified. This being a case of applying a cautionary applicability domain to the model for these types of chemicals.

In contrast to the over-prediction of the majority of volatile chemicals, galbanone and spirogalbanone were significantly under predicted using the fragment-based reactivity algorithm (chemicals 21 and 25 in Table 2). The skin sensitisation potency of these two chemicals was predicted using 3-methly-3-penten-2-one as the reference fragment to take account of the effect of an alkyl group at the α-position (which causes a decrease in the rate of the Michael addition reaction).[21] However, it is possible that a second site of Michael addition reactivity exists for these chemicals due to their reported ability to undergo double bond migration (highlighted part of the structure shown in Figure 7).[34] This type of migration is particularly favoured when the alkene group is unsubstituted (CH_2=CR) as is the case with galbanone and spirogalbanone (Figure 7). Predicting the glutathione reactivity of spirogalbanone and galbanone with the reference fragment 3-penten-2-one (to reflect the second potential site of reactivity) resulted in an improved pEC3 prediction of 1.84 (versus 1.36) for both galbanone (pEC3 =1.81) and spirogalbanone (pEC3 = 2.00). Importantly, it is likely that only one of these two possible sites of reactivity can undergo Michael addition at any one time as calculations show that the steric bulk of the cyclic ring enables only one of the alkene moieties to be conjugated with the carbonyl group at a time (data not shown). The predicted values suggest that the more reactive migrated site is primarily responsible for the skin sensitising ability of these chemicals. The more reactive alternative site for Michael addition was utilised for these chemicals enabling them to remain within the applicability domain of the model. This analysis demonstrates one of the strengths of the fragment-based thiol reactivity profiler in that it enables the investigation of alternative sites of chemical reactivity through the use of alternate fragments.

The final chemical that was poorly predicted was 5,5-dimethyl-3-methylene-dihydro-2-(3H)-furone, This chemical is a cyclic Michael acceptor in which only the α-carbon of the alkene is part of the ring system. The development of the fragment-based reactivity algorithm showed that the glutathione reactivity of cyclic Michael acceptors in which both the α- and β-carbons of the alkene were part of the ring could be successfully predicted using linear reference fragments.[21] In keeping with this analysis, the analogous chemicals in the skin sensitisation data were well predicted (chemicals 6, 7 and 14 in Table 2). Inspection of the data used to develop the fragment-based reactivity algorithm shows that it does not contain chemicals in which only the α-carbon of the double bond is part of the ring. In addition, these types of chemicals are also not present in the *Tetrahymena pyriformis* dataset analysed in the

current study. Therefore, it is impossible to ascertain as to whether the fragment-based reactivity algorithm is under-predicting the glutathione reactivity of these chemicals or if these chemicals are more potent in the LLNA than is predicted from reactivity alone.

The analysis outlined enabled the removal of 11 chemicals resulting in a final model based on 17 chemicals with an $R^2 = 0.77$ (Figure 8, model 4). Importantly, this model has a similar applicability domain to that published using experimentally determined kinetic rate constants, in that volatile chemicals and those that can polymerise are excluded.[3, 4] However, the use of $-\text{Log } RC_{50}$(calc) in the current study enabled a greater number of chemicals to be predicted (17 versus 10), whilst maintaining a similar level of statistical accuracy ($R^2 = 0.77$ versus 0.84).

$$\text{pEC3} = 1.77 + 0.43 - \text{Log } RC_{50} \text{ (calculated)} \qquad \text{(model 4)}$$

$$N = 17,\ R^2 = 0.76,\ R^2\text{-adj} = 0.76,\ s = 0.12$$

Conclusions

The aim of this work was to validate the fragment-based reactivity profiler for thiol reactivity for prediction of toxicity to *Tetrahymena pyriformis* and skin sensitisation potency for Michael acceptors. The results of this study showed the predicted reactivity values (-Log RC_{50}(calc)) was able to predict both endpoints within well-defined, end-point specific applicability domains. The results showed the importance of considering slow versus fast reacting Michael acceptors when modelling toxicity to *Tetrahymena pyriformis* and polymerisation and volatility to be important in successfully predicting skin sensitation potency. These results were in keeping with previously published studies that has utilised experimentally determined measurements of chemical reactivity to model the same endpoints. The statistical quality of resulting QSAR models demonstrated that the predicted reactivity values generated by the fragment-based profiler for thiol reactivity are on a par with using experimentally determined values. However, the use of an *in silico* approach offers clear benefits in terms of the ability to predict reactivity towards thiol for Michael acceptors in an efficient manner, without the need to perform either time-consuming and expensive experimental assays or undertake complex quantum mechanics calculations. The fragment-based *in silico* profiler could be developed further for additional endpoints such as genotoxicity for where lysine is the nucleophile. Such developments are dependent on the availability of reactivity data.

Abbreviations

Eact – Energies of Activation

HOMO – Highest Occupied Molecular Orbital

LUMO – Lowest Unoccupied Molecular Orbitals

SAS – Solvent Accessible Surface area

SMARTS – Smiles Arbitrary Target Specification

SMILES – Simplified Molecular Input Line Entry System

Funding

The research described in this article was funded in part by the 2013 LUSH prize for cosmetics and a research bursary to D.J.E from Liverpool John Moores University.

References

(1) Aptula, A. O., and Roberts, D. W. (2006) Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: General principles and application to reactive toxicity. *Chem Res Toxicol 19*, 1097-1105

(2) Yarbrough, J. W., and Schultz, T. W. (2007) Abiotic sulfhydryl reactivity: A predictor of aquatic toxicity for carbonyl-containing alpha,beta-unsaturated compounds. *Chem Res Toxicol 20*, 558-562.

(3) Roberts, D. W., and Natsch, A. (2009) High throughput kinetic profiling approach for covalent binding to peptides: application to skin sensitization potency of Michael acceptor electrophiles. *Chem Res Toxicol 22*, 592-603.

(4) Natsch, A., Emter, R., Gfeller, H., Haupt, T., and Ellis, G. (2015) Predicted Skin Sensitizer Potency Based on *In* Vitro Data from KeratinoSens and Kinetic Peptide Binding: Global Versus Domain-Based Assessment. *Toxicological Sciences 143*, 319-332.

(5) Mulliner, D., and Schuurmann, G. (2013) Model Suite for Predicting the Aquatic Toxicity of alpha,beta-Unsaturated Esters Triggered by Their Chemoavailability. *Molecular Informatics 32*, 98-107.

(6) Enoch, S. J., Cronin, M. T. D., Schultz, T. W., and Madden, J. C. (2008) Quantitative and mechanistic read across for predicting the skin sensitization potential of alkenes acting via Michael addition. *Chem Res Toxicol 21*, 513-520.

(7) Schultz, T. W., Yarbrough, J. W., Hunter, R. S., and Aptula, A. O. (2007) Verification of the structural alerts for Michael acceptors. *Chem Res Toxicol 20*, 1359-1363.

(8) Enoch, S. J., and Cronin, M. T. D. (2010) A review of the electrophilic reaction chemistry involved in covalent DNA binding. *Critical Reviews in Toxicology 40*, 728-748.

(9) Enoch, S. J., Ellison, C. M., Schultz, T. W., and Cronin, M. T. D. (2011) A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity. *Critical Reviews in Toxicology 41*, 783-802.

(10) Enoch, S. J., Madden, J. C., and Cronin, M. T. D. (2008) Identification of mechanisms of toxic action for skin sensitisation using a SMARTS pattern based approach. *Sar Qsar Environ Res 19*, 555-578.

(11) Bohme, A., Thaens, D., Paschke, A., and Schuurmann, G. (2009) Kinetic Glutathione Chemoassay To Quantify Thiol Reactivity of Organic Electrophiles-Application to alpha,beta-Unsaturated Ketones, Acrylates, and Propiolates. *Chem Res Toxicol 22*, 742-750.

(12) Bajot, F., Cronin, M. T. D., Roberts, D. W., and Schultz, T. W. (2011) Reactivity and aquatic toxicity of aromatic compounds transformable to quinone-type Michael acceptors. *Sar Qsar Environ Res 22*, 51-65.

(13) Bohme, A., Thaens, D., Schramm, F., Paschke, A., and Schuurmann, G. (2010) Thiol Reactivity and Its Impact on the Ciliate Toxicity of alpha,beta-Unsaturated Aldehydes, Ketones, and Esters. *Chem Res Toxicol 23*, 1905-1912.

(14) Roberts, D. W., Schultz, T. W., Wolf, E. M., and Aptula, A. O. (2010) Experimental Reactivity Parameters for Toxicity Modeling: Application to the Acute Aquatic Toxicity Of S(N)2 Electrophiles to Tetrahymena pyriformis. *Chem Res Toxicol 23*, 228-234.

(15) Schultz, T. W., Ralston, K. E., Roberts, D. W., Veith, G. D., and Aptula, A. O. (2007) Structure-activity relationships for abiotic thiol reactivity and aquatic toxicity of halo-substituted carbonyl compounds. *Sar Qsar Environ Res 18*, 21-29.

(16) Schultz, T. W., Sparfkin, C. L., and Aptula, A. O. (2010) Reactivity-based toxicity modelling of five-membered heterocyclic compounds: Application to Tetrahymena pyriformis. *Sar Qsar Environ Res 21*, 681-691.
(17) Roberts, D. W., Aptula, A. O., and Patlewicz, G. (2006) Mechanistic applicability domains for non-animal based prediction of toxicological endpoints. QSAR analysis of the Schiff base applicability domain for skin sensitization. *Chem Res Toxicol 19*, 1228-1233.
(18) Wondrousch, D., Bohme, A., Thaens, D., Ost, N., and Schuurmann, G. (2010) Local Electrophilicity Predicts the Toxicity-Relevant Reactivity of Michael Acceptors. *Journal of Physical Chemistry Letters 1*, 1605-1610.
(19) Enoch, S. J., and Roberts, D. W. (2013) Predicting Skin Sensitization Potency for Michael Acceptors in the LLNA Using Quantum Mechanics Calculations. *Chem Res Toxicol 26*, 767-774.
(20) Mulliner, D., Wondrousch, D., and Schuurmann, G. (2011) Predicting Michael-acceptor reactivity and toxicity through quantum chemical transition-state calculations. *Organic & Biomolecular Chemistry 9*, 8400-8412.
(21) Ebbrell, D. J., Madden, J. C., Cronin, M. T., Schultz, T. W., and Enoch, S. J. (2016) Development of a Fragment-Based in Silico Profiler for Michael Addition Thiol Reactivity. *Chem Res Toxicol 29*, 1073-1081.
(22) Schultz, T. W., Yarbrough, J. W., and Johnson, E. L. (2005) Structure-activity relationships for reactivity of carbonyl-containing compounds with glutathione. *Sar Qsar Environ Res 16*, 313-322.
(23) Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuii, H., Carcicato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakaiima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Mongomery, J. A., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kabayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Toamsi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Faraks, O., Foresmann, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. (2009) Gaussian 09, revision A.1, Wallingford, CT.
(24) US EPA. 2015. Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.11. United States Environmental Protection Agency, Washington, DC, USA.
(25) Ruusmann, V., and Maran, U. (2013) From data point timelines to a well curated data set, data mining of experimental data and chemical structure data from scientific articles, problems and possible solutions. *J Comput Aided Mol Des 27*, 583-603.
(26) Schultz, T. W., Yarbrough, J. W., and Pilkington, T. B. (2007) Aquatic toxicity and abiotic thiol reactivity of aliphatic isothiocyanates: Effects of alkyl-size and -shape. *Environmental Toxicology and Pharmacology 23*, 10-17.
(27) Gerberick, G. F., Ryan, C. A., Kern, P. S., Schlatter, H., Dearman, R. J., Kimber, I., Patlewicz, G. Y., and Basketter, D. A. (2005) Compilation of historical local lymph node data for evaluation of skin sensitization alternative methods. *Dermatitis 16*, 157-202.
(28) Kern, P. S., Gerberick, G. F., Ryan, C. A., Kimber, I., Aptula, A., and Basketter, D. A. (2010) Local Lymph Node Data for the Evaluation of Skin Sensitization Alternatives: A Second Compilation. *Dermatitis 21*, 8-32.
(29) Natsch, A., Ryan, C. A., Foertsch, L., Emter, R., Jaworska, J., Gerberick, F., and Kern, P. (2013) A dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing prevalidation. *Journal of Applied Toxicology 33*, 1337-1352.
(30) Dumont, C., Barroso, J., Matys, I., Worth, A., and Casati, S. (2016) Analysis of the Local Lymph Node Assay (LLNA) variability for assessing the prediction of skin sensitisation potential and potency of chemicals with non-animal approaches. *Toxicol In Vitro 34*, 220-228.

(31) ICCVAM. 2009. Recommended Performance Standards: Murine Local Lymph Node Assay. NIH Publication Number 09-7357. Research Triangle Park, NC: National Institute of Environmental Health Sciences.
(32) Barner-Kowollik, C. (2009) Acrylate Free Radical Polymerization: From Mechanism to Polymer Design. *Macromolecular Rapid Communications 30*, 1961-1963.
(33) Pham, P. D., Monge, S., Lapinte, V., Raoul, Y., and Robin, J. J. (2013) Various radical polymerizations of glycerol-based monomers. *European Journal of Lipid Science and Technology 115*, 28-40.
(34) Hubert, A. J., and Reimlinger, H. (1969) *Synthesis*, 97-112.

Table 1. The 62 chemicals used in the assessment of the fragment method for predicting *Tetrahymena pyriformis* toxicity (Log $1/EC_{50}$ mmol/l). Chemical names, SMILES, experimental Log $1/EC_{50}$ mmol/l with–LogRC_{50}(calc), D_{kk} and predicted Log $1/EC_{50}$ mmol/l for the respective models are shown. N.B log $1/EC_{50}$ values were calculated with model 2a for fast reacting chemicals (1-43) and model 2b for slower reacting chemicals (44-62).

ID	Chemical	SMILES	Log($1/EC_{50}$) (mmol/l)	–Log RC_{50}(calc)	LogK_{ow}	D_{kk}	Predicted Log ($1/EC_{50}$)	
							Model 1	Model 2a/b
1	prop-2-enal	C=CC=O	1.65	1.34	0.19	-1.15	1.45	1.66
2	(2E)-but-2-enal	C\C=C\C=O	0.88	0.66	0.60	-0.06	1.04	1.04
3	(2E)-3-(furan-2-yl)prop-2-enal	O=C\C=C\c1ccco1	0.37	0.05	1.19	1.14	0.66	0.46
4	(2E)-pent-2-enal	CC\C=C\C=O	0.66	0.55	1.09	0.54	0.97	0.94
5	4-methylpent-2-enal	CC(C)\C=C\C=O	0.82	0.55	1.51	0.96	0.97	0.94
6	hex-2-enal	CCC\C=C\C=O	0.77	0.55	1.58	1.03	0.97	0.94
7	(2E)-3-phenylprop-2-enal	O=C\C=C\c1ccccc1	0.68	0.05	1.82	1.77	0.66	0.46
8	(2E)-3-[4-(dimethylamino)phenyl]prop-2-enal	CN(C)c1ccc(\C=C\C=O)cc1	0.52	0.05	2.00	1.95	0.66	0.46
9	hept-2-enal	CCCC\C=C\C=O	1.05	0.66	2.07	1.41	1.04	1.04
10	(2E)-oct-2-enal	CCCCC\C=C\C=O	1.20	0.55	2.57	2.02	0.97	0.94
11	(2E)-2-methylbut-2-enal	C\C=C(/C)C=O	-0.14	-0.96	1.15	2.11	0.04	-0.49
12	non-2-enal	CCCCCC\C=C\C=O	1.60	0.66	3.06	2.40	1.04	1.04
13	2-methylpent-2-enal	CC\C=C(/C)C=O	-0.39	-1.05	1.64	2.69	-0.01	-0.58
14	but-3-en-2-one	CC(=O)C=C	1.50	0.92	0.41	-0.51	1.20	1.27
15	pent-1-en-3-one	CCC(=O)C=C	1.49	0.92	0.90	-0.02	1.20	1.29
16	hex-1-en-3-one	CCCC(=O)C=C	1.66	0.92	1.39	0.47	1.20	1.29
17	pent-3-en-2-one	C\C=C\C(C)=O	0.54	0.15	0.82	0.67	0.72	0.56
18	hex-4-en-3-one	CCC(=O)\C=C\C	0.93	0.10	1.31	1.21	0.69	0.51
19	oct-1-en-3-one	CCCCCC(=O)C=C	1.92	0.92	2.37	1.45	1.20	1.29
20	hept-3-en-2-one	CCC\C=C\C(C)=O	0.70	0.00	1.80	1.80	0.63	0.42
21	oct-3-en-2-one	CCCC\C=C\C(C)=O	0.74	0.00	2.29	2.29	0.63	0.42
22	oct-2-en-4-one	CCCCC(=O)\C=C\C	1.01	0.00	2.29	2.29	0.63	0.42
23	2-methylcyclopent-2-en-1-one	CC1=CCCC1=O	-0.83	-1.25	1.26	2.51	-0.14	-0.77
24	3-methylpent-3-en-2-one	C\C=C(/C)C(C)=O	-0.34	-1.25	1.37	2.62	-0.14	-0.77
25	non-3-en-2-one	CCCCC\C=C\C(C)=O	0.98	0.00	2.79	2.79	0.63	0.42
26	2-hydroxyethyl prop-2-enoate	OCCOC(=O)C=C	0.69	0.50	-0.25	-0.75	0.94	0.88
27	2-hydroxypropyl prop-2-enoate	CC(O)COC(=O)C=C	0.65	0.50	0.17	-0.33	0.94	0.89
28	methyl prop-2-enoate	COC(=O)C=C	0.55	0.50	0.73	0.23	0.94	0.89
29	ethyl prop-2-enoate	CCOC(=O)C=C	0.52	0.50	1.22	0.72	0.94	0.89
30	propyl prop-2-enoate	CCCOC(=O)C=C	0.53	0.50	1.71	1.21	0.94	0.89

31	2-methylpropyl prop-2-enoate	CC(C)COC(=O)C=C	0.29	0.50	2.13	1.63	0.94	0.89
32	2-hydroxyethyl 2-methylprop-2-enoate	CC(=C)C(=O)OCCO	-1.08	-1.40	0.30	1.70	-0.23	-0.91
33	butyl prop-2-enoate	CCCCOC(=O)C=C	0.52	0.50	2.20	1.70	0.94	0.89
34	benzyl prop-2-enoate	C=CC(=O)OCc1ccccc1	1.35	0.50	2.44	1.94	0.94	0.89
35	3-methylbutyl prop-2-enoate	CC(C)CCOC(=O)C=C	0.41	0.50	2.62	2.12	0.94	0.89
36	pentyl prop-2-enoate	CCCCCOC(=O)C=C	0.54	0.50	2.69	2.19	0.94	0.89
37	cyclohexyl prop-2-enoate	C=CC(=O)OC1CCCCC1	0.76	0.50	3.00	2.50	0.94	0.89
38	methyl 2-methylprop-2-enoate	COC(=O)C(C)=C	-1.28	-1.40	1.28	2.68	-0.23	-0.91
39	hexyl prop-2-enoate	CCCCCCOC(=O)C=C	0.73	0.50	3.18	2.68	0.94	0.89
40	2-methylpropyl (2E)-but-2-enoate	C\C=C\C(=O)OCC(C)C	-0.34	-0.19	2.54	2.73	0.51	0.24
41	butan-2-yl (2E)-but-2-enoate	CCC(C)OC(=O)\C=C\C	-0.42	-0.19	2.54	2.73	0.51	0.24
42	butyl (2E)-but-2-enoate	CCCCOC(=O)\C=C\C	-0.16	-0.19	2.61	2.80	0.51	0.24
43	2-ethoxyethyl 2-methylprop-2-enoate	CCOCCOC(=O)C(C)=C	-0.78	-1.40	1.49	2.89	-0.23	-0.91
44	(2E)-dec-2-enal	CCCCC\C=C\C=O	1.85	0.55	3.55	3.00	0.97	1.50
45	heptyl prop-2-enoate	CCCCCCCOC(=O)C=C	1.09	0.50	3.67	3.17	0.94	1.59
46	ethyl 2-methylprop-2-enoate	CCOC(=O)C(C)=C	-0.93	-1.40	1.77	3.17	-0.23	-0.76
47	methyl (2E)-oct-2-enoate	CCCCC\C=C\C(=O)OC	0.77	-0.19	3.10	3.29	0.51	0.84
48	methyl (2E)-3-phenylprop-2-enoate	COC(=O)\C=C\c1ccccc1	0.58	-0.94	2.36	3.30	0.05	-0.08
49	methyl (2E)-2-methylbut-2-enoate	COC(=O)C(\C)=C\C	-0.70	-1.64	1.69	3.33	-0.38	-0.92
50	propan-2-yl 2-methylprop-2-enoate	CC(C)OC(=O)C(C)=C	-0.88	-1.40	2.18	3.58	-0.23	-0.40
51	propyl 2-methylprop-2-enoate	CCCOC(=O)C(C)=C	-0.66	-1.40	2.26	3.66	-0.23	-0.33
52	methyl non-2-enoate	CCCCCC\C=C\C(=O)OC	1.04	-0.19	3.60	3.79	0.51	1.29
53	ethyl (2E)-3-phenylprop-2-enoate	CCOC(=O)\C=C\c1ccccc1	0.99	-0.94	2.85	3.79	0.05	0.36
54	ethyl (2E)-2-methylbut-2-enoate	CCOC(=O)C(\C)=C\C	-0.50	-1.64	2.18	3.82	-0.38	-0.48
55	methyl (2E)-2-methylpent-2-enoate	CC\C=C(/C)C(=O)OC	-0.38	-1.64	2.18	3.82	-0.38	-0.48
56	2-methylpropyl 2-methylprop-2-enoate	CC(C)COC(=O)C(C)=C	-0.28	-1.40	2.67	4.07	-0.23	0.04
57	butyl 2-methylprop-2-enoate	CCCCOC(=O)C=C	-0.27	-1.40	2.75	4.15	-0.23	0.11
58	propyl (2E)-3-phenylprop-2-enoate	CCCOC(=O)\C=C\c1ccccc1	1.23	-0.94	3.34	4.28	0.05	0.80
59	benzyl 2-methylprop-2-enoate	CC(=C)C(=O)OCc1ccccc1	0.65	-1.40	2.98	4.38	-0.23	0.32
60	butyl (2E)-3-phenylprop-2-enoate	CCCCOC(=O)\C=C\c1ccccc1	1.53	-0.94	3.83	4.77	0.05	1.24
61	hexyl 2-methylprop-2-enoate	CCCCCCOC(=O)C(C)=C	1.09	-1.40	3.73	5.13	-0.23	0.99
62	2-ethylhexyl 2-methylprop-2-enoate	CCCCC(CC)COC(=O)C(C)=C	1.57	-1.40	4.64	6.04	-0.23	1.80

Table 2. The 27 chemicals used in the assessment of the fragment-based reactivity algorithm's ability to predict skin sensitisation potency (pEC3). Chemicals SMILES, experimental pEC3 with error values,–LogRC_{50}(calc), Log VP and predicted pEC3 values for all models are shown.

ID	Chemical	SMILES	pEC3	-Log RC_{50}(calc)	Log VP	Predicted pEC3	
						Model 3	Model 4
1	Methyl methacrylate	CC(=C)C(=O)OC	0.05	-1.40	1.59	1.28(1.23)	-
2	2-Hydroxypropyl methacrylate	CC(COC(=O)C(=C)C)O	0.46	-1.40	-1.10	1.28(0.82)	-
3	Ethyl acrylate	CCOC(=O)C=C	0.55	0.50	1.61	1.38(0.83)	-
4	Methyl acrylate	COC(=O)C=C	0.63	0.50	1.95	1.38(0.75)	-
5	Butyl acrylate	CCCCOC(=O)C=C	0.81	0.50	0.74	1.38(0.57)	-
6	r-Carvone	CC(=C)C1CC=C(C)C(=O)C1	1.07	-1.25	-0.86	1.29(0.22)	1.23(0.16)
7	L-Carvone	CC1=CC[C@H](CC1=O)C(=C)C	1.10	-1.25	-0.86	1.29(0.19)	1.23(0.16)
8	a-Butyl cinnamic aldehyde	CCCC\C(C=O)=C/c1ccccc1	1.23	-1.26	-2.55	1.29(0.06)	1.23(0.00)
9	Linalool aldehyde	C\C(C=O)=C/CCC(C)(O)C=C	1.25	-0.98	-2.51	1.30(0.05)	1.35(0.10)
10	trans-2-Hexenal	CCC\C=C\C=O	1.25	0.41	0.71	1.38(0.13)	-
11	a-Amyl cinnamic aldehyde	CCCCC/C(=C\c1ccccc1)/C=O	1.26	-1.26	-3.47	1.29(0.03)	1.23(-0.03)
12	α-Hexylcinnamaldehyde	CCCCCC\C(C=O)=C/c1ccccc1	1.26	-1.26	-3.45	1.29(0.03)	1.23(-0.03)
13	2-Ethylhexyl-acrylate	CCCCC(CC)COC(=O)C=C	1.27	0.50	-0.71	1.38(0.11)	-
14	Perillaldehyde	CC(=C)C1CCC(C=O)=CC1	1.27	-0.98	-1.32	1.30(0.03)	1.35(0.08)
15	1-(p-Methoxyphenyl)-1-penten-3-one	CCC(=O)\C=C\c1ccc(OC)cc1	1.31	-0.55	-2.73	1.33(0.02)	1.54(0.23)
16	a-Methyl-cinnamic aldehyde	C\C(C=O)=C/c1ccccc1	1.51	-1.26	-1.59	1.29(-0.22)	1.23(-0.28)
17	Benzylidene acetone	CC(=O)\C=C\c1ccccc1	1.60	-0.55	-2.00	1.33(-0.27)	1.54(-0.06)
18	5-Methyl-2-phenyl-2-hexenal	CCCC\C=C(\C=O)c1ccccc1	1.63	-0.79	-2.55	1.31(-0.32)	1.43(-0.20)
19	Cinnamic aldehyde	O=C\C=C\c1ccccc1	1.63	0.05	-1.46	1.36(-0.27)	1.80(0.17)
20	trans-2-Decenal	CCCCCCC/C=C/C=O	1.79	0.41	-1.08	1.38(-0.41)	1.95(0.16)
21	Galbone	CC1(C)CCC=C(C1)C(=O)CCC=C	1.81	-1.25 (0.15)‡	-1.72	1.29(-0.52)	1.84(0.03)
22	5,5-Dimethyl-3-methylene-dihydro-2(3H)-furone	CC1(C)CC(=C)C(=O)O1	1.85	-1.40	-0.76	1.28(-0.57)	-
23	Diethyl maleate	CCOC(=O)/C=C/C(=O)OCC	1.91	0.10	-0.72	1.36(-0.55)	1.82(-0.09)
24	2-Hydroxyethyl acrylate	C=CC(=O)OCCO	1.92	0.50	-0.85	1.38(-0.54)	-
25	Spirogalbanone	C=CCCC(=O)C1=CCCC2(CCCC2)C1	2.00	-1.25 (0.15) ‡	-3.02	1.29(-0.71)	1.84(-0.16)
26	Pomarose	C\C=C\C(=O)C(\C)=C(/C)C(C)C	2.02	0.15	-0.55	1.36(-0.66)	1.84(-0.18)

*Notes chemicals with predictions outside of the experimental error for that model. Error values for predicted pEC3 values for all chemicals are shown in brackets.

‡ Chemicals with an additional –LogRC_{50}(calc) values use this value for Model 3 for reasons discussed in the text

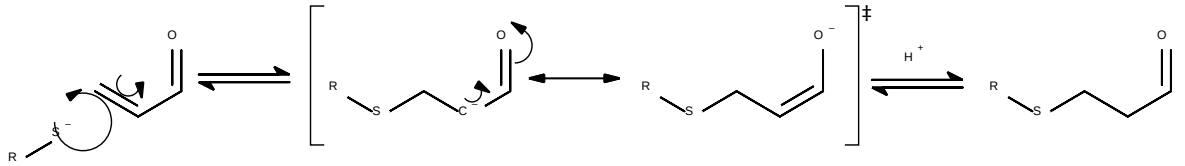


Figure 1. Michael addition reaction between acrolein and a thiol nucleophile (R = glutathione, alkyl).

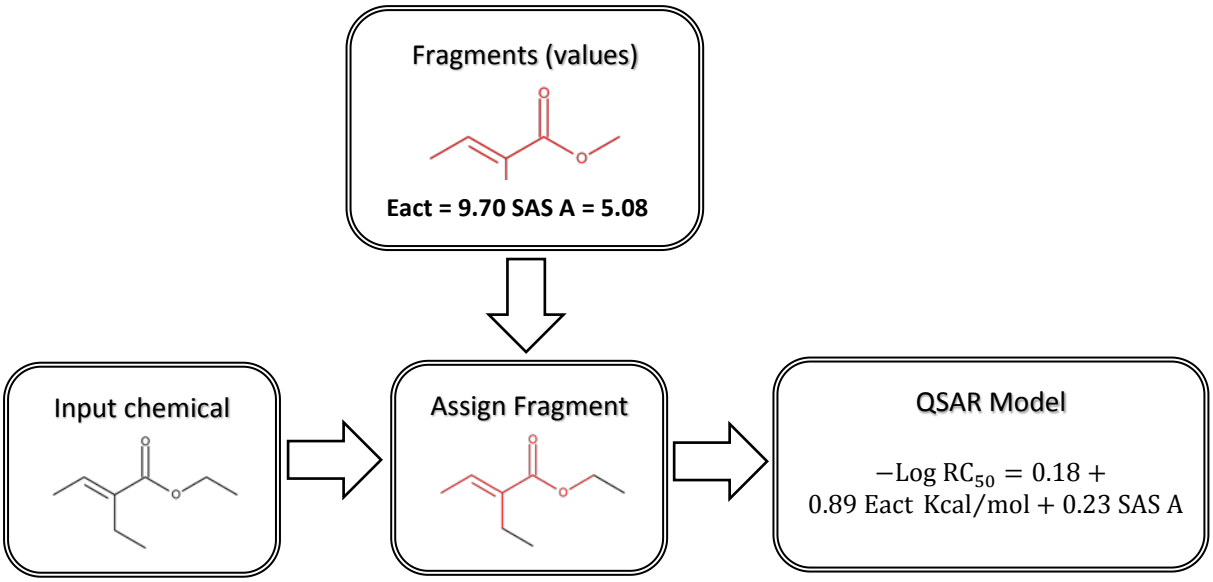


Figure 2. A summary of the workflow used to predict reactivity (-Log RC_{50}).

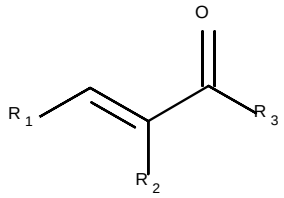


Figure 3: Domain covered by fragment method for Michael acceptors. (R_1 = Hydrogen, alkyl, aryl) (R_2 = Hydrogen, alkyl, aryl) (R_3 = H) for polarised aldehydes, (R_3 = CH, C-alkyl, C-aryl) for polarised ketones. (R_3 = OCH, OC-alkyl, OC-aryl) for polarised esters.

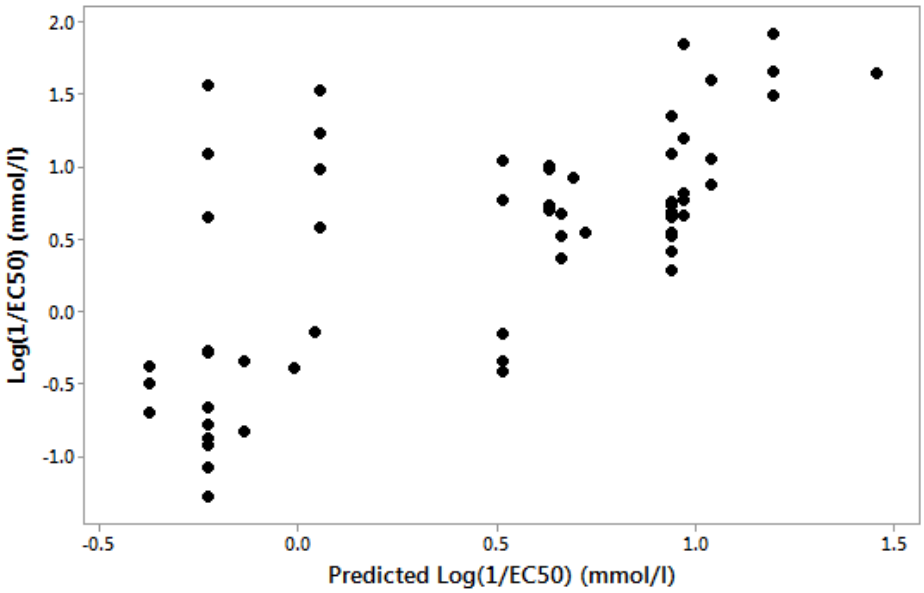


Figure 4. The predicted Log $1/EC_{50}$ values against experimental $1/EC_{50}$ values for all 62 Michael acceptors using –Log RC_{50}(calc) alone (model 1).

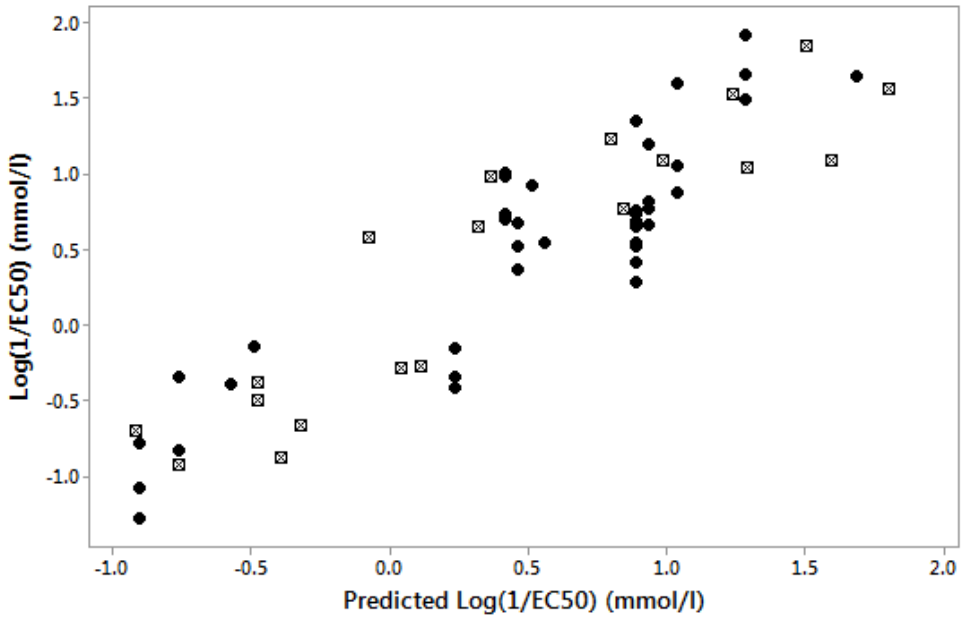


Figure 5: The predicted Log ($1/EC_{50}$) (mmol/l) against experimental Log ($1/EC_{50}$) (mmol/l) of all 43 fast reacting chemicals (bold circles) (model 2a, chemicals 1-43 in Table 1)) and 19 slower reacting chemicals (squares) (model 2b, chemicals 44-62 in Table 1) requiring hydrophobicity to be taken into account

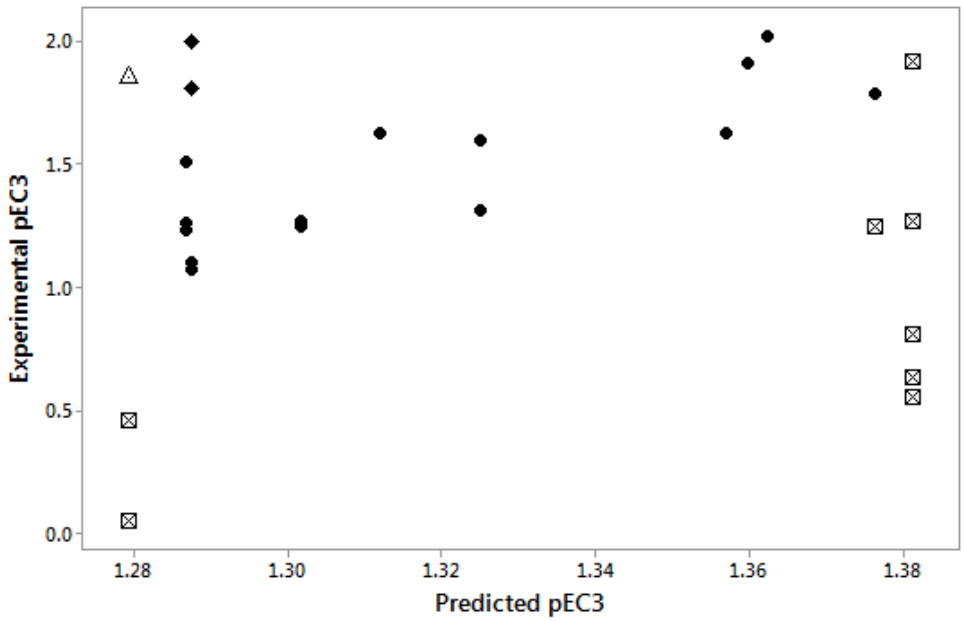


Figure 6. Predicted pEC3 versus experimental pEC3 for all 26 Michael acceptors shown in Table 2. Unfilled squares = volatile chemicals; filled diamonds = galbanone and spirogalbanone; unfilled triangle = 5,5-dimethyl-3-methylene-dihydro-2(3H)-furone

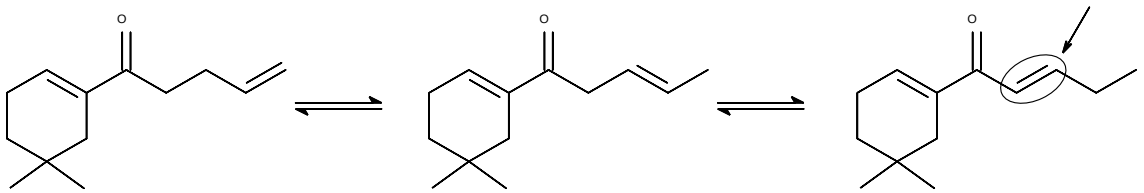


Figure 7. Isomerisation of galbanone to produce extended conjugated chemical highlighting a possible additional site of reactivity.

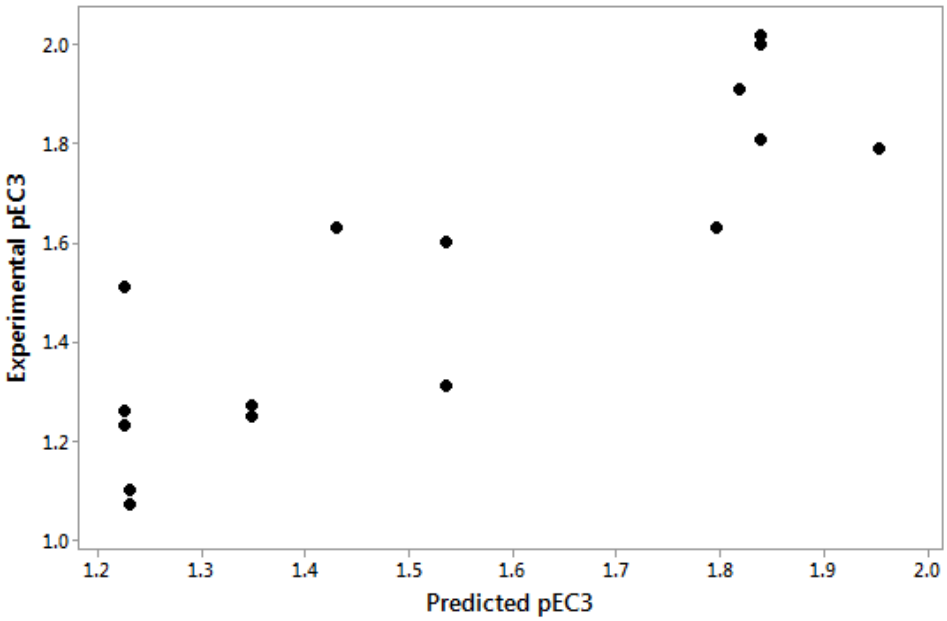


Figure 8. The predicted pEC3 against experimental pEC3 for model 4 (predicted values shown in Table 2).