The Development of a Fragment-Based *in Silico* Profiler for the Prediction of Thiol Reactivity and Toxicity

David Jonathan Ebbrell

A thesis submitted in partial fulfillment of the requirements of Liverpool John Moores University for the degree of Doctor of Philosophy

February 2018

Acknowledgements

First, I would like to thank my supervisor Dr Steve Enoch. Throughout my studies, he has provided me with a tremendous amount of support and has given a substantial amount of his time and knowledge into this thesis. I would like to give a similar thanks to Dr Judith Madden and Prof Mark Cronin who have always been willing to help and provide guidance when needed.

I am also very grateful to Prof Terry Schultz (University of Tennessee) who kindly contributed the experimental data utilized throughout this thesis.

I would also like to thank my colleagues in the Chemoinformatics group for providing an enjoyable experience and a pleasant working environment. In particular, Dr Claire Mellor for her invaluable assistance at the beginning of my studies and to my fellow PhD students: Iva Lukac (now Dr Lukac), Maria Sapounidou, Joanna Zarnecka and Julia Pletz

In addition to my work colleagues, I would like to thank my family and friends for the support they have given me throughout my studies. Finally, I would like to dedicate this thesis to my Mum.

Abstract

Regulatory toxicology in the 21st century is faced with the challenge of having to replace its use of experimental animals in chemical risk assessment with alternative methods. This is due to the introduction of the REACH legislation and the seventh amendment to the cosmetics directive. Such alternative methods include the use of *in vitro* (cell culture/tissue etc.), *in chemico* (chemical experiments e.g. determination of reactivity) and *in silico* (computational) approaches. Importantly, it is envisaged that data from all these alternative sources will be required for the prediction of the animal-based endpoints used in regulatory toxicology. One of the key computational approaches used for data gap filling is category formation and read-across. When using this approach to assess the potential toxicity of a chemical, a chemical category is best defined based on a common molecular initiating event e.g. the formation of a covalent bond with biological nucleophile via the same chemical mechanism. The structural features that define a chemical's membership of such a category can be encoded computationally as structural alerts, which in turn, can be grouped together to form an *in silico* profiler.

The work discussed in this thesis addresses the key shortcoming of traditional *in silico* profilers, this being that current *in silico* profilers provided no information about the rate of covalent bond formation for chemicals containing the same structural alert but with different substituents. The research within this thesis addresses this problem through the introduction of a fragment-based approach to *in silico* profiler development. This fragment-based approach introduces the use of calculated activation energies determined through the use of quantum mechanics calculations which enable chemical reactivity to be predicted. Chapter 3 outlines the development of the approach for α,β-unsaturated aldehydes, ketones and esters which form covalent bonds through Michael addition. Chapter 4 extends the work outlined in Chapter 3 demonstrating how the fragment-based profiler can be used to predict both chemical reactivity and skin sensitisation and toxicity to *Tetrahymena pyriformis.* Finally, Chapter 5 extends the approach to chemicals capable

of reacting with proteins via an S_N2 mechanism demonstrating the approach can be applied to any mechanistic domain for which data exist. Overall, this thesis outlines an approach for the development of novel fragment-based *in silico* profilers capable of quantitatively predicting chemical reactivity and by extension toxicity. It is envisaged that the work outlined in this thesis will be of use primarily in regulatory toxicology, within such tools as the OECD QSAR toolbox.

Table of Contents

List of abbreviations in thesis:

AOP – Adverse Outcome Pathway

B3LYP – Beck 3 parameter exchange correlation functional and the Lee-Yang-Parr correlation functional

DFT – Density Functional Theory

D_{kk} - Difference between Log k_{OW} and Log k_{GSH}

DMSO – Dimethyl Sulfoxide

DNA – Deoxyribose Nucleic Acid

DTNB – 5,5-Dithio-bis-2-nitrobenzoic Acid

Eact – Activation Energy

EC3 – Concentration required to elicit a three-fold proliferation to vehicle controls

EC_{50} – 50% Effective Concentration

ECHA – European Chemical Agency

EMA – European Medicines Agency

EPA – Environmental Protection Agency

EU – European Union

FDA – Food and Drug Administration

GGA – Generalized Gradient Approximation

GSH – Glutathione

GTO – Gaussian Type Orbitals

GPMT – Guinea Pig Maximisation Test

HF – Hartree-Fock

HOMO – Highest Occupied Molecular Orbital

IGC_{50} – 50% Growth Impairment Concentration

INT – Intermediate

k_{GSH} – Glutathione Kinetic Rate Constant

K_{OW} – Octanol : Water Partition Coefficient

LCAO – Linear Combination of Atomic Orbitals

LDA – Local Density Approximation

LLNA – Local Lymph Node Assay

LUMO – Lowest Unoccupied Molecular Orbital

MeOH – Methanol

MIE – Molecular Initiating Event

OECD – Organisation for Economic Co-operation and Development

PCM – Polarizable Continuum Model

pEC3 – Logarithmic molar EC3 value

pKa – Acid dissociation constant

QM – Quantum Mechanics

QSAR – Quantitative Structure Activity Relationship

RC_{50} – 50% Reactivity Concentration

REACH – Registration, Evaluation, Authorisation and Restriction of Chemicals

SAR – Structure Activity Relationship

SAS – Solvent Accessible Surface

SCF – Self-Consistent Field

SMARTS – Smiles Arbitrary Target Specification

SMILES – Simplified Molecular Input Line Entry System

S_N1 – Aliphatic Unimolecular Nucleophilic Substitution

S_N2 – Aliphatic Bimolecular Nucleophilic Substitution

S_NAr – Nucleophilic Aromatic Substitution

STO – Slater Type Orbitals

TG – Test Guidelines

TS – Transition State

VP – Vapour Pressure

Chapter 1. Introduction

1.1 Regulatory toxicology

Regulatory toxicology is concerned with the risk assessment and management of chemicals that humans and the environment are exposed to (1). The regulatory frameworks specifies a series of toxicological endpoints that must be assessed – these being defined as the recorded observation as a result of exposing a chemical to a biological system (e.g. an organism, the environment or a whole population). The nature and number of the required endpoints differs depending on a chemical's intended usage. For example, skin sensitisation is a key endpoint that must be assessed for chemicals used as ingredients in cosmetics, but is less significant for chemicals used in the formulation of drugs (although it could still be important depending on the intended route of administration). As might be expected, differing regulatory agencies are responsible for enforcing the relevant regulatory framework within different countries. For example in the US, the main regulatory agencies are the Environmental Protection Agency (EPA), and the Food and Drug Administration (FDA) (2). Whilst within the European Union, the main regulatory agencies are the European Chemical Agency (ECHA) for industrial chemicals and cosmetics and the European Medicines Agency (EMA) for medicines. The international nature of the cosmetics sector, and other chemical industries, requires a harmonised system of toxicological test guidelines suitable for assessing a given endpoint. This is provided by the Paris-based Organisation for Economic Co-operation and Development (OECD) which allows its 35 member countries to agree test guidelines to enable the assessment of the different endpoints required under a given regulatory environment (3). The test guidelines provided by the OECD typically contain information on the dosing of test chemical used, the species and number of test animals to use, the duration of the test and the procedure.

The majority of the endpoints defined by the OECD guidelines involve the use of laboratory animals. For example, there are four test guidelines for the determination of the skin sensitisation potential

of a chemical. These being test guidelines (TG): TG 406, TG 429, TG 442a and TG 442b (4-7). Where determining skin sensitisation is concerned, the preferred test guidelines are either the Guinea Pig Maximisation Test (GPMT, TG 406) or the Local Lymph Node Assay (LLNA, TG 429, TG 442a and TG 442b). It is worth noting that although the OECD provide a set of harmonised test guidelines for the determination of toxicological endpoints, companies do not have to follow them in order to meet the regulatory criteria for a given chemical. However, such is the nature and acceptance of the OECD test guidelines that deviations need to be fully justified in order to be accepted by the relevant regulatory authority.

Replacement of animal testing and legislation

As previously mentioned, the majority of the OECD harmonised test guidelines involve the use of animals. However, the seventh amendment to the cosmetics directive required a complete ban on the testing of cosmetic products on animals within the European Union from March 2009. Additionally, this amendment also banned the marketing of products that have been tested on animals outside the EU (8, 9). This has led to an effective worldwide ban as the EU cosmetics market is the largest in the world consisting of over 500 million people and is worth in excess of €77 billion (10). The enforcement of the ban on animal testing was staggered over a number of years to allow alternative methods to be developed for the relevant endpoints (a summary of the endpoints affected is shown in Table 1.1) (11). However, despite this staged process the number of endpoints for which suitable alternative testing methods exist is currently limited.

Table 1.1: The testing and marketing ban dates for toxicological endpoints as required by the cosmetic directive along with their respective OECD test guidelines, species and number of animals in the study (4, 12-27).

Endpoint	OCED guideline	Species	Number of animals	Testing ban date	Marketing ban date
Acute toxicity	TG 420	Rodent	5 animals of one sex per dose group	11 March 2009	11 March 2009
	TG 423	Rodent	3 animals of a single sex per step (average 2-4 steps)		
	TG 425	Rodent	5 maximum		
Skin sensitisation	TG 406	Guinea pig	10 minimum	11 March 2009	11 March 2013
	TG 429	Mice	4 minimum per dose group		
	TG 442a	Mice	4 minimum per dose group		
	TG 442b	Mice	4 minimum per dose group		
Sub-acute and sub-chronic toxicity	TG 411	Rodent / Mammal	At least 10 of each sex at each dose level	11 March 2009	11 March 2013
	TG 412	Rodent	5 of each sex		
	TG 413	Rodent	10 of each sex		
Genotoxicity and mutagenicity	TG 474	Mammalian (rodent preferred)	5 per sex	11 March 2009	11 March 2013
	TG 475	Mammalian (rodent preferred)	5 per sex		
	TG 478	Mammalian (rodent preferred)	Sufficient amount to provide statistical power		
	TG 483	Mammalian (rodent preferred)	5 males per group		
	TG 488	Rodent	5 minimum		

Endpoint	OCED guideline	Species	Number of animals	Testing ban date	Marketing ban date
Toxicokinetics and metabolism	TG 417	Same as that used in other toxicological studies performed with the test substance of interest	4 minimum of one sex for each dose group	11 March 2009	11 March 2009
Carcinogenicity	TG 451	Rodents	At least 50 animals of each sex per dose group	11 March 2009	11 March 2013
Carcinogenicity	TG 453	Rodent preferred	At least 50 animals of each sex per dose group	11 March 2009	11 March 2013
Reproductive and developmental toxicity	TG 416	Rodent preferred	20 females minimum	11 March 2009	11 March 2013
Reproductive and developmental toxicity	TG 421	Rodent	10 per sex	11 March 2009	11 March 2013

The Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) also places pressure on animal usage by ensuring all chemicals produced or imported into the EU in quantities of one tonne per annum (or more) are assessed for human and environment hazards (28). However, REACH does not ban animal testing but suggests that alternative methods should be used where possible. This means that there is a clear benefit for the application of such approaches to deal with the scale of information required by this legislation. As an example of the scale of the requirements, it has been estimated that there are up to 30,000 chemicals currently used in the EU that do not have data for a full safety assessment. Such safety assessment typically requires information from 10 *in vivo* toxicity endpoints (Table 1.2).

Table 1.2: The toxicological and ecotoxicological *in vivo* endpoints required under the REACH legislation along with their respective OECD test guidelines with information on the species, number of animals and duration of the endpoint study (4, 12-14, 18-22, 26, 27, 29-33).

Tonnes per year	Toxicological endpoint	OECD test guideline(s)	Species	Number of animals	Duration
1-10	Skin sensitisation	TG 406	Guinea pig	10 minimum	24 days
	Short-term toxicity on invertebrates	TG 202	*Daphnia* sp.	At least 20 at each test concentration	14 days
	Acute toxicity: Oral	TG 420	Rodent	5 animals of one sex per dose group	14 days
		TG 423	Rodent	3 animals of a single sex per step (average 2-4 steps)	14 days
		TG 425	Rodent	5 maximum	48 hours
10-100	Eye irritation	TG 405	Rabbit	3	8 hours
	Genotoxicity	TG 474	Mammalian (rodent preferred)	5 per sex	24-72 hours
		TG 475	Mammalian (rodent preferred)	5 per sex	39-45 hours
		TG 478	Mammalian (rodent preferred)	Sufficient to provide statistical power	until second half of pregnancy
		TG 483	Mammalian (rodent preferred)	5 males per group	1 week
		TG 488	Rodent	5 minimum	11 – 14 weeks
	Acute toxicity: Inhalation	TG 403	Mammalian (rodent preferred)	10 at each concentration	14 days

Tonnes per year	Toxicological endpoint	OECD test guideline(s)	Species	Number of animals	Duration
10-100	Short-term repeat dose toxicity	TG 407	Rodent	10 at each dose level	14 or 28 days
	Reproductive/developmental toxicity	TG 416	Rodent	20 females minimum	21-24 days (duration of pregnancy)
		TG 421	Rodent	10 per sex	63 days
	Short term toxicity on fish	TG 203	Various fish species	At least 7 at each test concentration	96 hours

In summary, the implementation of the seventh amendment to the cosmetic directive bans the use of laboratory animals for the testing and marketing of cosmetics products within the EU. In addition to this, REACH requires all chemicals produced or imported into the EU in quantities of one tonne per annum (or more) to be assessed for human and environmental hazards (28). It is clear that for differing policy reason both pieces of legislation necessitate the use of alternative testing methods rather than conventional animal-based OECD test guidelines. These methods cover the use of *in vitro* (cell culture/tissue etc.), *in chemico* (chemical experiments e.g. determination of reactivity) and *in silico* (computational) methods (34-36).
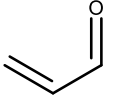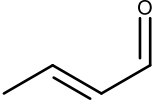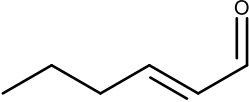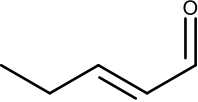
1.2 Category formation

One of the key computational alternative methods that has been suggested to assist in the reduction of animal based testing in toxicological risk assessment is category formation and read-across (35, 37). The OECD defines a category as *"Chemicals whose physical-chemical, toxicological and ecotoxicological properties that are likely to be similar or follow a regular pattern as a result of structural similarity may be considered as a group or category"* (38, 39). This is similar to the definition provided by ECHA, which defines a group as "*substances that are structurally similar with physicochemical, toxicological, eco-toxicological and/or environmental fate properties that are*

likely to be similar or to follow a regular pattern" (40, 41). The similarity of chemicals can depend on a number of factors. This may be defined by chemical shape, the presence of various functional groups or the mechanism through which the chemical initiates a toxicological response (the Molecular Initiating Event). The term read-across is used to describe a method of predicting a physicochemical property or toxicological endpoint for a chemical by using existing data from chemicals within the same group (42, 43). Read-across can be performed either qualitatively or quantitatively depending on the type of data available for the other category members. For example, the toxicity (for a hypothetical endpoint) of pent-2-enal can be predicted from the other members of the category shown in Table 1.3.   If the other category members only had categorical data associated with them (for the endpoint of interest) then it would only be possible to carry out a qualitative read-across prediction for pent-2-enal. However, if numerical data were available for the category members then it would be possible to perform a quantitative read-across prediction leading to the prediction of potency (an important consideration in risk assessment). Quantitative read-across predictions typically involve developing a local (or an MIE-based) QSAR which can be used to relate chemical properties (e.g. chemical reactivity, molecular weight, steric factors, electronic factors etc) to a given toxicological endpoint (42, 43). For example, a recent study was able to predict the skin sensitisation potency of 16 chemicals which were grouped into a category based on their ability to react with skin proteins via a Schiff base mechanism (44). This allowed a QSAR model to be developed relating Taft values (σ^* - as a measure of reactivity) and hydrophobicity (LogP - as a measure of skin permeability) to skin sensitisation potency (expressed as pEC3 values, Equation 1.1). This QSAR model was subsequently used to predict skin sensitisation potency for a number of additional chemicals also acting via a Schiff base mechanism (e.g. 2-hexenal, orange data point in Figure 1.1).

Table 1.3: Qualitative read-across for the prediction of toxicity for pent-2-enal

Chemical	prop-2-enal	but-2-enal	hex-2-enal	pent-2-enal
Structure				
Toxicity	✓	✓	✓	?
Read-across prediction	✓	✓	✓	✓
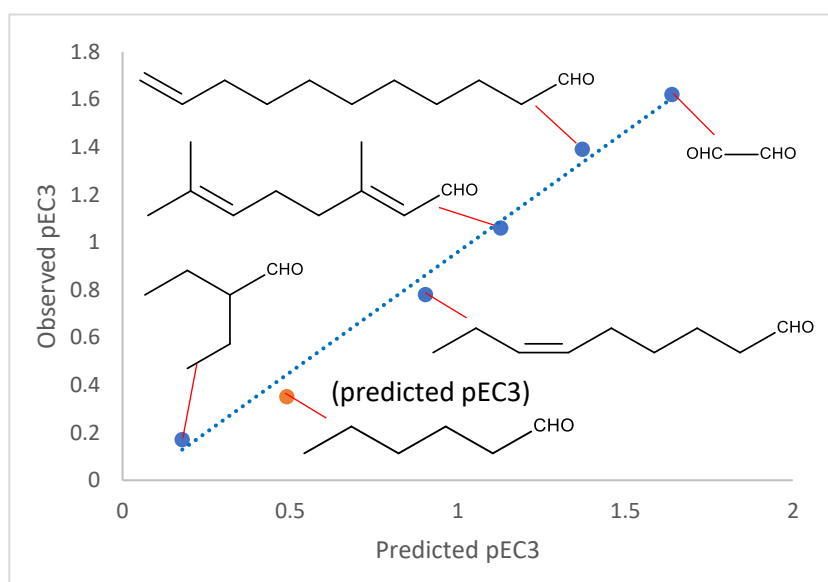
Figure 1.1: Quantitative read-across to predict the 2-hexenal (shown in orange)

$$\text{pEC3}_{obs} = 1.12 \sum \sigma^* + 0.42 \, LogP - 0.62 \qquad (1.1)$$

N = 16 R^2 = 0.95 R$^2_{adj}$ = 0.95

N = number of values, R^2 = Coefficient of determination (how well the variation in the data is modelled by the regression line), R$^2_{adj}$ = Adjusted R^2 based on the number of independent variables, R$^2_{pred}$ = how well the model predicts the response for new observations (calculated by systematically

removing each observation in the dataset) S = the standard deviation of the distance between the data values and the fitter values, these statistical measurements will be used throughout this thesis.

The key challenge with category-formation is defining the similarity that forms the basis for the group of chemicals. Common methods of defining similarity include: common physical chemical properties, common molecular fingerprints and common functional groups (45). However, the method that is used most often for defining similarity is grouping chemicals based on ability to elicit a common MIE. To recap, an MIE is defined as the initial interaction between a chemical and the biological system that leads to an adverse toxicological effect (often measured as an endpoint, as discussed previously). A number of classes of MIEs have been defined ranging from binding to key receptor sites on proteins and/or enzymes and the disruption of the electron transport chain in the mitochondria (46, 47). Additionally, the formation of a covalent bond between an electrophilic chemical (or an electrophilic metabolite) and a biological nucleophile (e.g. lysine, cysteine or DNA) has been shown to be a key MIE for a number of toxicities. Importantly, studies have demonstrated that knowledge of the chemistry that defines these covalent reactions can be used to build mechanism-based categories, within which QSAR and read-across can be utilised to predict endpoints such as skin sensitisation, respiratory sensitisation, acute aquatic toxicity and hepatotoxicity (48-53). Additionally, these investigations have also shown that the rate of these chemical reactions is key in the prediction of toxicological potency within such mechanism-based categories (43, 54-56).

1.3 Covalent reaction chemistry

As stated above, chemicals may be grouped into mechanism-based categories on their ability to form a covalent bond with a protein or DNA. Research has shown there to be six mechanistic domains relevant to toxicity through which such covalent bonds may be formed, these being: Michael addition, nucleophilic aromatic substitution (S_NAr), bimolecular nucleophilic substitution (S_N2), aliphatic unimolecular nucleophilic substitution (S_N1), acylation and Schiff base formation (57-

60). Of these six mechanisms, two are the focus of this thesis: Michael addition and S_N2. Chemicals that act via Michael addition (known as Michael acceptors) typically contain an alkene or alkyne group polarised by the presence of an electron-withdrawing group (e.g. carbonyl or a nitro group) (61). The presence of this group results in a partial positive charge on the β-carbon of the π-bond causing the chemical to be susceptible to nucleophilic attack at this position (57, 58). This nucleophilic attack results in a negative charge which is then stabilised through resonance with the electron-withdrawing group. The final step in the mechanism is protonation of the α-carbon (e.g. from the deprotonated thiol group or a deprotonated water molecule) to produce the final product (this mechanism is summarised in Figure 1.2). The key factors that influence the rate of the Michael addition reaction are:

- The nature of the electron-withdrawing group: This pulls the electron density away from the alkene or alkyne resulting in the partial positive charge on the β-carbon. As such, an increase in electron-withdrawing ability results in an increased partial charge on the β-carbon which, in turn, increases the rate of reaction (i.e. a nitro group is more reactive than a ketone).

- Steric bulk at the β–carbon: the addition of alkyl and aryl groups at the β–carbon results in less accessible surface area around the site of nucleophilic attack. This leads to a lower number of potential reactive collisions between the electrophile and nucleophile resulting in a decrease in the rate of the reaction.

- The nature of the α-group: as the negative charge of the intermediate (partially) resides on the α-carbon (resonance stabilised by the electron-withdrawing group), the presence of substituents at this position has a significant effect on the rate of reaction. The addition of an electron-withdrawing group (e.g. cyano) increases the amount of resonance stabilisation of the intermediate. This results in an increase in the rate of reaction. In contrast, the presence of an electron-donating group (e.g. an alkyl) has the opposite effect, decreasing the rate of the Michael addition reaction.
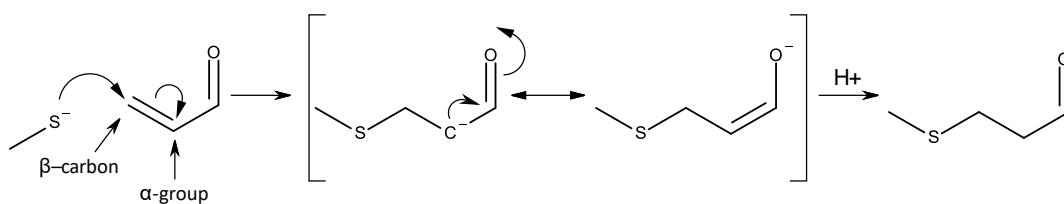
Figure 1.2. The mechanism for the Michael addition reaction between prop-2-enal and methyl thiolate

The second mechanism discussed in this thesis is bimolecular nucleophilic substitution (S_N2). This typically occurs at an aliphatic carbon, nitrogen, sulphur or halogen atom with an electronegative leaving group attached (e.g. a halogen). The S_N2 reaction involves the formation of a covalent bond between the nucleophile and the reactive centre in the electrophile. This occurs simultaneously with the breaking of the bond between the reactive centre and the leaving group. Therefore, the S_N2 reaction has a single transition state connecting the reactants and the product (mechanism is summarised in Figure 1.3). The key factors that influence the rate of the S_N2 reaction are:

- Nature of the leaving group: the S_N2 reaction cannot occur unless a suitable leaving group is attached to the reactive centre. The ability of the leaving group to leave depends on the strength of its conjugate base, where the weakest bases are considered as the best leaving groups. The strength of the conjugate base depends on electronegativity, as electronegativity decreases, the strength of the conjugate base decreases. Therefore, for the halogens, iodine is considered the best leaving group with fluorine being the poorest.

- Steric bulk around the reactive centre: Increased steric bulk makes it difficult for the nucleophile to attack the reactive centre of the electrophile resulting in a decrease in the rate of reaction (for example, $CH_3Cl > CH_2RCl > CHR_2Cl >> CR_3Cl$, in order of most to least reactive where R = alkyl or aryl groups). It is worth noting that tertiary carbons atoms (e.g. CR_3Cl) are frequently considered as reacting via an S_N1 mechanism rather than S_N2.
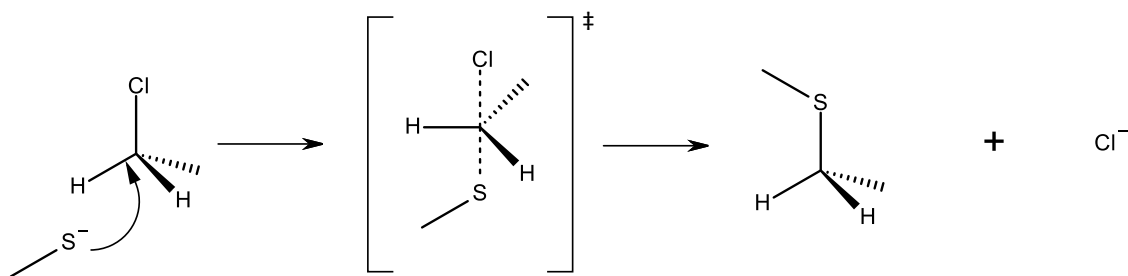
Figure 1.3. The mechanism for bimolecular nucleophilic substitution (S_N2) between chloroethane and methyl thiolate. Note that the electrophilic atom can vary (e.g. carbon, nitrogen, sulphur or halogen)

Although they are not the focus of this thesis, there are an additional four mechanisms associated with covalent bond formation relevant to toxicity (S_NAr, S_N1, acylation and Schiff base formation) (57-60). Nucleophilic aromatic substitution (S_NAr) occurs between a nucleophile such as cysteine or lysine group and an activated aromatic system. This mechanism is analogous to the S_N2 reaction apart from that the reactive carbon is part of an aromatic ring. Chemicals capable of undergoing an S_NAr reaction typically contain electron-withdrawing groups on the aromatic ring located in one, or more, of the 2-, 4- or 6- positions (relative to the electronegative leaving group). In contrast to the S_N2 mechanism, the S_NAr reaction occurs in two steps, the first being nucleophilic attack at the carbon attached to the electronegative leaving group. This results in the formation of a resonance-stabilised carbanion intermediate. The second step of the reaction is the subsequent loss of the leaving group and the reformation of the aromatic ring (the mechanism is summarised in Figure 1.4). The key factors that influence the rate of the S_NAr reaction are:

- Activating groups: the presence of activating groups at either the 2-, 4- and/or 6- positions to the electronegative leaving group. This is due to the stabilisation of the resonance-stabilised carbanion intermediate (Figure 1.4). A greater electron-withdrawing ability of the group at the 2-, 4- and/or 6- positions results in an increase in stabilisation (and therefore reactivity). Additionally, the more electron withdrawing substituents there are the greater the stabilisation is (i.e. 2- and 4- is better than 2- alone). Activating groups at the 3- and 5-

position to the electronegative leaving group have significantly less effect on resonance stability. Contrastingly, electron-donating groups (e.g. alkyl groups) have a deactivating effect on reactivity due to their ability to destabilise the resonance-stabilised intermediate. This results in a decreased rate of reaction.

- Nature of the leaving group: an increase in electronegativity results in an increase in stabilisation of the carbanion intermediate resulting in an increase in reactivity. As such, fluorine is the most reactive halide and iodine is the least reactive. It is worth nothing that this is the opposite of what is seen with S_N2 reactivity.

- Steric bulk around the reactive site: an increase in the steric bulk around the reactive site results in a lower number of potential collisions which, in turn, decreases reactivity.
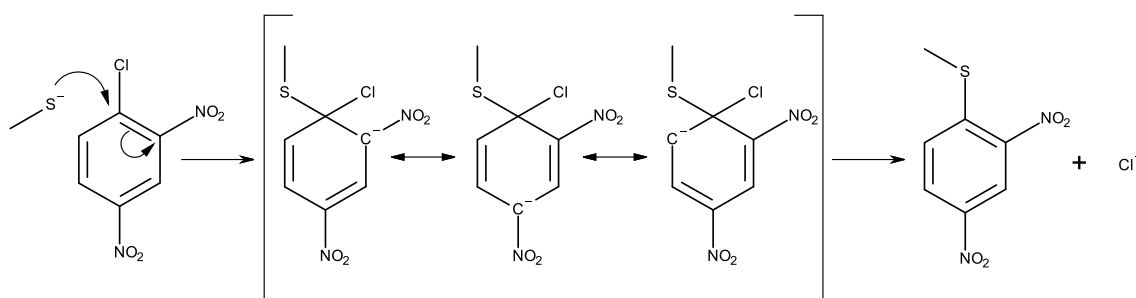
Figure 1.4. The mechanism for nucleophilic aromatic substitution (S_NAr) between 2-4-dinitro-chlorobenzene and a methyl thiolate

Unimolecular nucleophilic substitution (S_N1) is similar to the S_N2 mechanism in that it occurs at an aliphatic carbon with an electronegative leaving group attached (e.g. halogen). In contrast to the S_N2 mechanism, the S_N1 mechanism occurs in two steps with the rate limiting step being the formation of a three coordinate carbocation species. This species then undergoes nucleophilic attack producing the final product (the mechanism is summarised in Figure 1.5). The key factors that influence the rate of the S_N1 reaction are:

- Nature of the carbocation species: the more stable the carbocation intermediate, the faster S_N1 reactivity will be. The reactivity of electrophilic carbon groups is the inverse of S_N2

13

reactivity (i.e. $CR_3Cl > CHR_2Cl > CH_2RCl \gg CH_3Cl$, where R = alkyl or aryl groups). Additionally, the carbocation species can be stabilised if the carbocation is next to a double bond or aromatic system. This is as it allows for the delocalisation of the charge to produce multiple resonance structures resulting in an increase in rate of reaction.

- Nature of the leaving group: reactivity increases with increasing leaving group ability. Leaving group ability for S_N1 chemicals follows a similar trend to S_N2 chemicals with iodine being the most reactive halogen leaving group and fluorine being least reactive halogen leaving group.
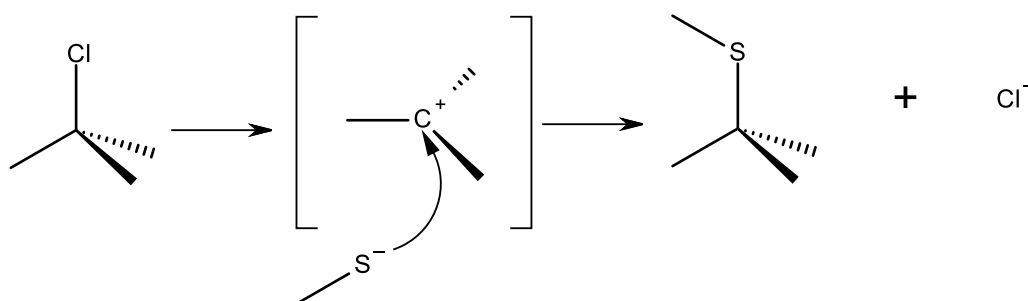
Figure 1.5. The mechanism for unimolecular nucleophilic substitution (S_N1) reaction between a 2-chloro-2-methyl-propane and methyl thiolate

Chemicals capable of reacting via an acylation mechanism have a carbonyl group (or carbonyl type moiety) adjacent to an electronegative leaving group. The mechanism is an addition-elimination reaction, with the first step involving nucleophilic attack at the carbon of the carbonyl group. This results in the formation of a tetrahedral structure featuring a negatively charged oxygen atom. The electrons from this negative charge then reform the carbonyl double bond resulting in the expulsion of the leaving group (the mechanism is summarised in Figure 1.6). The key factors that influence the rate of the acylation reaction are:

- Nature of the leaving group: the reactivity of chemicals reacting via acylation is primarily dependent on the leaving group ability, which is related to the strength of its conjugate

acid. For example, considering the halides as leaving groups results in the order F > Cl > Br > I due to the decreasing acid strengths (i.e. HI is a much weaker acid that HF).

- Steric bulk around the reactive site: an increase in the steric bulk around the reactive site results in a lower number of potential collisions which, in turn, decreases reactivity.
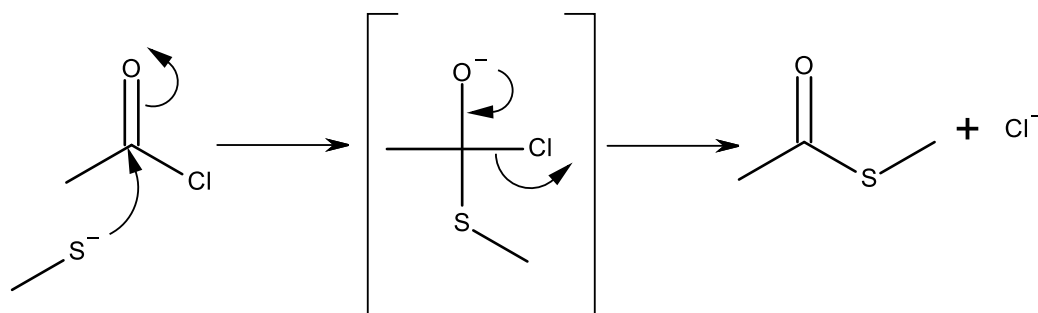

Figure 1.6. Acylation reaction between acetyl chloride and methyl thiolate

The final mechanism relevant to toxicity is Schiff base formation. This mechanism is different from the others that have been outlined as it only occurs between an aldehyde or a ketone and lysine (a nitrogen based biological nucleophile). The mechanism involves a two-step process, with the first step involving nucleophilic attack by nitrogen on the carbonyl group of either an aldehyde or a ketone. A proton transfer then occurs resulting in the formation of an enolamine. This species then undergoes a further proton transfer and the loss of water to produce the final Schiff base (Figure 1.7). The key factors in Schiff base formation reactivity are:

- The nature of the substituents attached to the aldehyde or ketone: any substituents that pull electron density away from the carbon increases reactivity. Contrastingly anything that donates electron density decreases reactivity.
- Steric bulk around the reactive site: an increase in the steric bulk around the reactive site results in a lower number of potential collisions that lead to a reaction which, in turn, decreases reactivity.
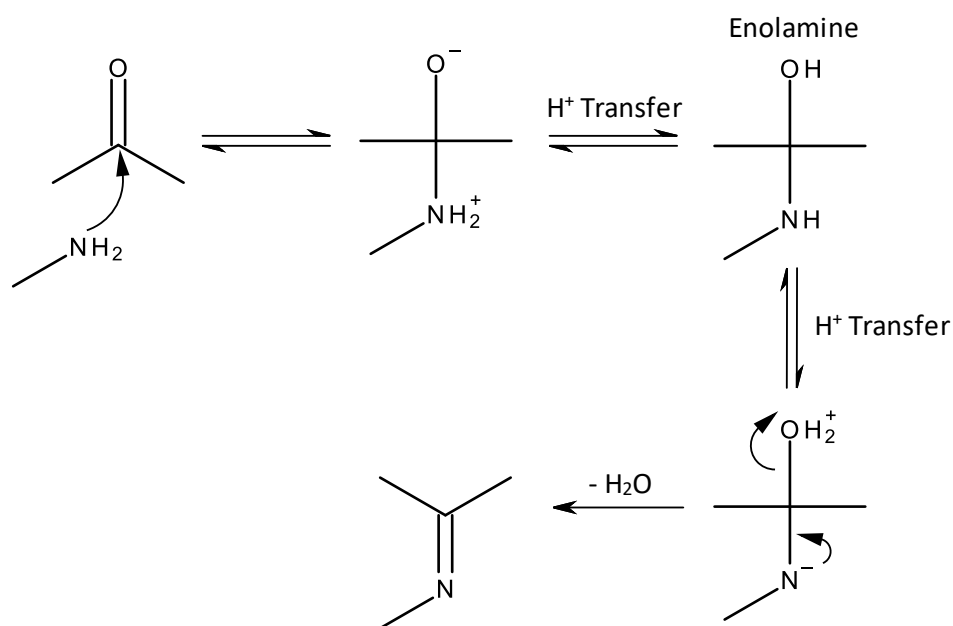
Figure 1.7. The mechanism for Schiff base formation between a propan-2-one and methylamine

1.4 Structural alerts, *in silico* profilers and category formation

As discussed previously, the key challenge in category formation is the need to define chemical similarity. If the above mechanistic information is to be used to define chemical similarity then it needs to be encoded computationally. One key *in silico* method that achieves this is in the development and use of structural alerts. Generally speaking, a structural alert is a chemical fragment or substructure associated with a particular measured property (biological or otherwise). In terms of covalent reaction chemistry, structural alerts can be used to define the structural boundaries associated with each of the mechanisms outlined i.e. used to identify chemicals that have the right structural features to undergo a particular reaction. For example, the structural alert for α,β-unsaturated aldehydes which have the potential to form covalent bonds with biological nucleophiles through a Michael addition mechanism (Figure 1.8). The presence of a common structural alert within a database of chemicals enables them to be identified and subsequently grouped into a mechanism-based category. Clearly, in order to define the wide range of potential structural features associated with the chemistry discussed above requires a large number of

complex structural alerts. It is these collections of structural alerts that have been termed *in silico* profilers and encoded into a number of tools such as the OECD QSAR Toolbox that may be used within regulatory toxicology (62). The OECD QSAR toolbox being the key software tool that is used by regulatory agencies and the chemical industry to fill data gaps. The OECD QSAR toolbox operates using a sequential workflow composed of various modules, these being in order: chemical input, profiling, endpoints, category definition, filling data gaps and report.
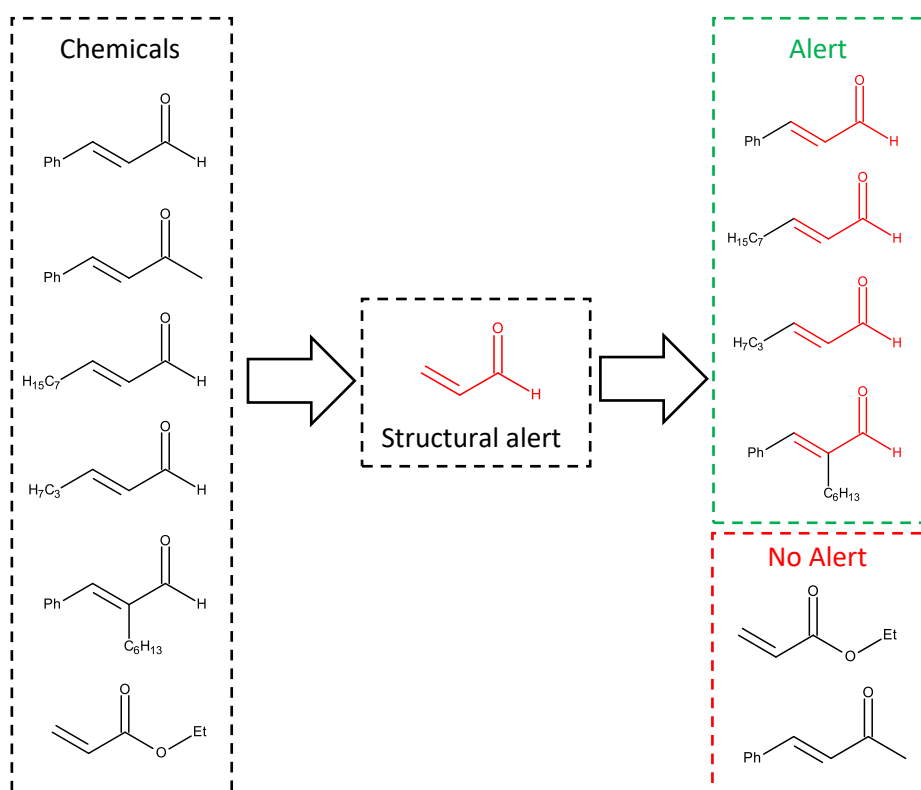
Figure 1.8 Development of a mechanistic category for α,β-unsaturated aldehydes using structural alerts

Given the relevance of covalent bond formation to toxicity, it is perhaps unsurprising that the chemistry detailed above has been incorporated into the OECD QSAR Toolbox as a number of *in silico* profilers (48, 53, 57, 58, 63-68). Additionally, a number of studies have demonstrated that experimental reactivity data (i.e. the rate at which chemicals form covalent bonds) can be used to predict toxicological potency. For example, glutathione reactivity data has been used in previous studies to predict aquatic toxicity and skin sensitisation for chemicals acting via Michael addition

(43, 56, 69, 70). Importantly, these studies highlighted that potency was highly correlated with the rate of the Michael addition reaction. Related computational studies have also shown that calculated descriptors based on either activation energies or electrophilicity are also capable of predicting potency for both of these endpoints for chemicals acting via Michael addition (9, 49, 52, 71-76). These studies are discussed in depth in Chapter 3, Section 3.1. The key limiting factor for these computational studies is the need for complex and time consuming Quantum Mechanics (QM) calculations (requiring proprietary software). This limits the inclusion of these approaches in freely available tools such as the OECD QSAR toolbox. With this in mind, the key aims of this thesis were to develop an *in silico* profiler capable of predicting reactivity and toxicity for two mechanistic domains (Michael addition and S_N2). The approach being based on a fragment method in which a database of pre-calculated energy of activation values are used within the *in silico* profiler, thus removing the need for the end-user to perform QM calculations.

1.5 Endpoints studied in this thesis

There are two toxicological endpoints that are the focus of the predictive modelling within this thesis. These being: skin sensitisation, as determined in the LLNA and aquatic toxicity, as determined in the *Tetrahymena pyriformis* growth impairment assay. The prediction of these endpoints is discussed in Chapter 4 (chemicals causing skin sensitisation and toxicity to *Tetrahymena pyriformis* via Michael addition) and Chapter 5 (chemicals causing toxicity to *Tetrahymena pyriformis* via an S_N2 mechanism).

1.5.1 Skin sensitisation and the local lymph node assay

As discussed previously, one of the key animal tests by which a chemical's ability to cause skin sensitisation can be assessed within the EU is the LLNA (49). Compared to the other available animal assays for skin sensitisation (the GPMT assay) the LLNA has the advantage of providing a quantitative estimation of potency. This estimation is based upon the measurement of proliferative responses of lymphocytes induced in the draining lymph nodes following topical exposure of a test

chemical to mice (77). Typically, these are expressed as EC3 values, this being the amount of test chemical needed to elicit a three-fold increase in proliferative activity when compared to vehicle controls (78). The greater the skin sensitisation potency, the lower the EC3 value, with chemicals that do not elicit a three-fold increase in proliferation of lymphocytes in the lymph nodes being considered as non-sensitisers. Additionally, pEC3 values are recorded that are based on molar concentrations rather than by weight (calculation through Equation 1.2) (9). When attempting to predict skin sensitisation potency, it is important to note that there is an inherent two-fold repeatability error associated with the LLNA (equivalent of 0.3 log units when using pEC3 values) (79). Chemicals can be assigned to one of five potency-based classes based on their EC3 value. These being:

LLNA EC3%	Potency Category
EC3 < 0.1	Extreme
0.1 ≤ EC3 < 1.0	Strong
1.0 ≤ EC3 < 10.0	Moderate
10.0 ≤ EC3 < 100.0	Weak
100.0 ≤ EC3	Non-sensitizer

$$pEC3 = Log\ (EC3/Molecular\ weight) \hspace{3cm} (1.2)$$

The protocol for the LLNA is as follows (80):

1. Day 1: The mice are initially treated with a solution of sodium lauryl sulphate. After one hour either the appropriate dilution of the test substance, the vehicle or a positive control is then applied to the dorsum of each ear.

2. Days 2, 3: Sodium lauryl sulphate and the test chemical is reapplied (or vehicle/positive control) to the dorsum of each ear.

3. Days 4 and 5: The animals undergo no treatment.

4. Days 6: The mice are injected with radioactive ^{3}H-thymidine (^{3}H-T). Five hours later the animals are sacrificed and the auricular lymph nodes are drained.

5. The extent of lymphocyte proliferation is determined by quantification of radioactive ^{3}H-T incorporation in the draining lymph nodes.

The mechanistic detail of the LLNA is sufficiently well understood that the key steps have been detailed in an AOP (summarised in Figure 1.9).[81] An AOP details the knowledge linking the initial chemical interaction (the MIE), through a series of biological key events to the adverse effect (in this case sensitisation of the skin) [82]. The AOP for the LLNA can be summarised as:

1. The test chemical is exposed to the skin

2. The test chemical (if not already electrophilic) is then activated (through metabolism and/or oxidation) to produce an electrophilic substance

3. This active substance then covalently binds with either a cysteine or lysine moiety (acting as biological nucleophiles). This has been denoted as the MIE in the AOP

4. This results in the activation of keratinocytes and dendritic cells which causes lymphocyte proliferation. These have been denoted as key events in the AOP

5. The culmination of these steps (the adverse effect) leads to skin sensitisation
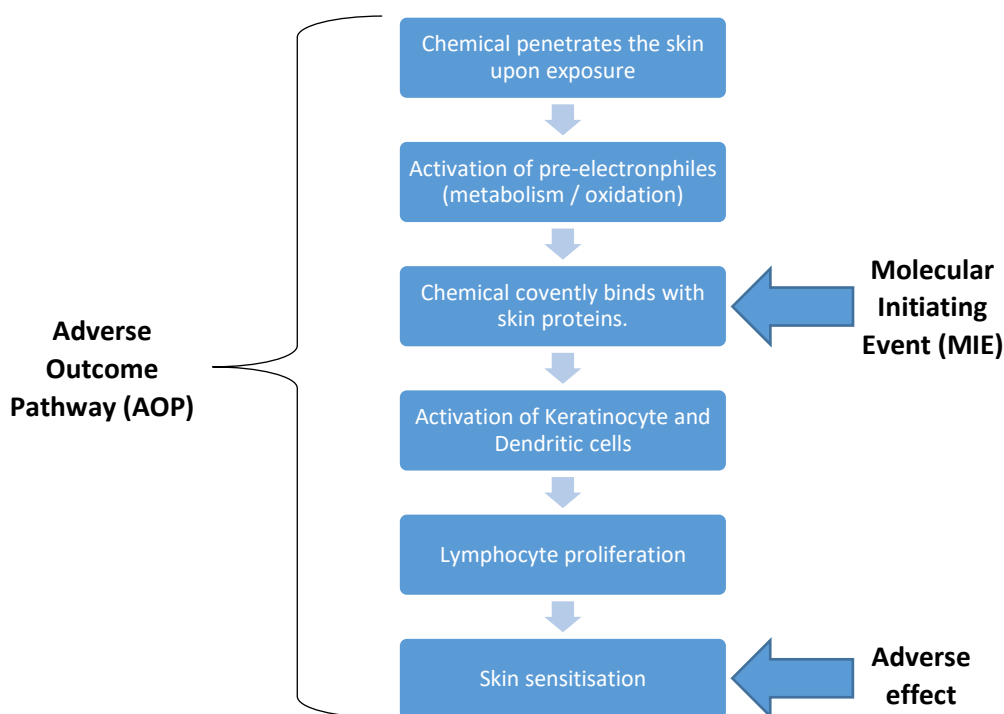
Figure 1.9: Summary of the AOP for skin sensitisation as determined in the LLNA

1.5.2 *Tetrahymena pyriformis* growth impairment assay

The assessment of environmental toxicity is an important aspect of all risk assessment legislation. One of the key endpoints associated with environmental toxicity is the ability of a chemical to cause acute aquatic toxicity. To this end, the *Tetrahymena pyriformis* growth impairment assay was developed as a quantitative estimation of a chemical's acute aquatic toxicity potency (83). The assay measures the growth impairment of *Tetrahymena pyriformis* after exposure to a test chemical where population growth impairment is the reduction in population density (i.e. cells per unit volume) under specified conditions (time, temperature etc.). Toxicity is typically quantified as the IGC_{50} (impairment growth concentration for 50% of control populations). This is achieved by exposing the ciliate to varying concentrations of the test chemical dissolved in a stock solution over a 40 hour period. Growth inhibition data is then collected either spectrophotometrically or by electronic particle counting (an instrument that counts the number of physical particles or cells). The *Tetrahymena pyriformis* growth impairment assay offers several advantages over traditional *in vivo* fish assays in that *Tetrahymena* can be cultured both easily and economically and that the

toxicity of chemicals can be quantified rapidly. This has led to the development of a database containing 2072 chemicals with toxicity to *Tetrahymena pyriformis* data. Importantly, values generated from the *Tetrahymena pyriformis* growth impairment assay have been shown to correlate well with data from the fathead minnow lethality test (which involves live fish) for a dataset of 250 chemicals covering a wide range of common mechanisms (83). The majority of chemicals cause toxicity through (polar or non-polar) narcosis which is the nonspecific disturbance of cell membranes as a result of the accumulation of toxicants in biological membranes. Because of this, baseline toxicity normally scales with the ability of a chemical to permeate cell membranes and therefore their hydrophobicity (67, 68). However, there are various reactive mechanisms associated with aquatic toxicity that are analogous to those for skin sensitisation (e.g. Michael addition). Chapters 4 and 5 of this thesis outline how calculated reactivity data can be utilised to predict toxicity to *Tetrahymena pyriformis* for chemicals acting via Michael addition and the S_N2 mechanism respectively.

1.6 *In chemico* methods

In chemico assays involve the use of chemistry-based experiments. For example, the quantification of chemical reactivity between a chemical and a biological nucleophile. The major concept behind the use of *in chemico* assays is that chemical reactivity and toxicity are linearly related for chemicals whose MIE is covalent binding to a biological macromolecule.

1.6.1 Glutathione depletion assay

The glutathione (GSH) depletion assay is a kinetics based *in chemico* assay which utilises the reactive peptide GSH which comprises glutamate, cysteine and glycine (Figure 1.10). The functional group of GSH that is responsible for the formation of covalent bond with an electrophilic chemical is the thiol of cysteine (57). This is useful as glutathione models the effects of biological nucleophiles (therefore mimics the effects of an *in vivo* system). The protocol for the glutathione depletion assay is as follows:

1. Varying concentrations of test chemicals are dissolved in a stock solution containing DMSO

2. A buffered GSH solution is then introduced into the mixture

3. After 60 minutes, the remaining GSH is quantified from its reaction with 5,5-dithio-bis-2-nitrobenzoic acid (DTNB)

4. The difference between initial GSH concentration and that quantified with DTNB represents how much GSH has reacted with the electrophile

Values are typically expressed as EC_{50} (mmol) or RC_{50} (mmol) which means the effective concentration (for EC_{50} or reactive concentration for RC_{50}) which results in 50% of GSH being depleted in 120 minutes (84). Studies have shown a direct correlation between GSH kinetic rate data (k_{GSH}) and GSH depletion data (RC_{50}). This shows that GSH reactivity data can be used as a surrogate for kinetic rate data (56). However, this study also highlights that the GSH depletion assay is a lot less sensitive to chemicals in the low and high reactivity ranges (i.e. GSH depletion data isn't able to distinguish extremely fast reacting chemicals from one another as well as is observed when using kinetic rate data). In addition to glutathione, the depletion of other reactive peptides (e.g. lysine) has also been studied (72, 85, 86). Lysine reactivity data is likely to be useful for investigating the Schiff base domain (as Schiff base formation only occurs with nitrogen based nucleophiles) and for the investigation of respiratory sensitisation (63).
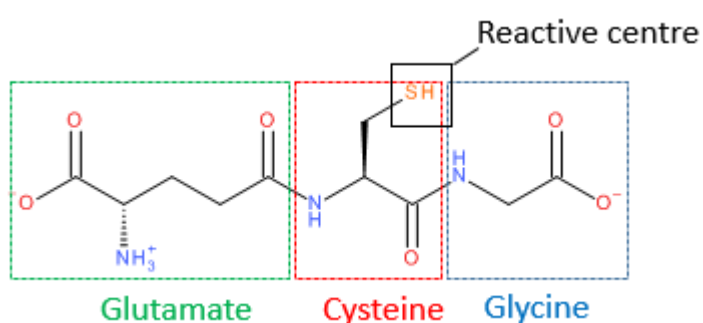
Figure 1.10: Glutathione (GSH) with the reactive centre of cysteine (thiol group) highlighted.

1.7 Concluding remarks

In this chapter, the current regulations that are the drivers of the need for non-animal test methods in regulatory toxicological have been outlined, these being the seventh amendment to the cosmetics directive and REACH. One of the key *in silico* methods that is being promoted to address these challenges is category formation and read-across. For a number of endpoints the key approach for the development of such chemical categories is grouping chemicals based on their ability to form covalent bonds with biological nucleophiles. Research has shown there to be six mechanistic domains relevant to toxicity through which such covalent bonds may be formed. This chapter highlighted how this mechanistic information can be encoded computationally as an *in silico* profiler enabling category formation and read-across predictions to be made. Finally, the biological assays which are the focus of this thesis were outlined. These being the *in vivo* LLNA, the *in vitro Tetrahymena pyriformis* growth impairment assay and the *in chemico* glutathione depletion assay. Chapter 2 outlines the theory behind the quantum mechanical methods used to computationally model chemical reactivity within this thesis. Chapters 3, 4 and 5 outline the use and development of *in silico* methods (through the use of QSAR models) for the prediction of chemical reactivity for these endpoints.

Research aims of this thesis

The overall aim of this thesis was to develop a fragment-based *in silico* profiler for chemicals acting via the Michael addition and S_N2 mechanisms and demonstrate the utility of the approach for the prediction of thiol reactivity and toxicity. The specific objectives to achieve this were:

- To develop fragments for chemicals that react with proteins via Michael addition

  - This involved the identification of suitable descriptors capable of predicting chemical reactivity (the results being outlined in Chapter 3)

- To validate the fragment-based *in silico* profiler for Michael addition

  - This involved investigating the ability of the fragment-based *in silico* profiler to predict chemical reactivity as determined in a glutathione-based depletion assay, the prediction of skin sensitisation potency as determined in the local lymph node assay and, finally, toxicity to *Tetrahymena pyriformis* (the results being outlined in Chapter 4)

- To extend the fragment-based approach to other mechanistic domains relevant to toxicology

  - This involved extending the approach to chemicals capable of reacting with proteins via an S_N2 mechanism. As for Michael addition, this profiler was validated via an investigation into its ability to predict glutathione reactivity and toxicity to *Tetrahymena pyriformis*

Chapter 2. Theoretical background of quantum mechanics calculations

In this thesis quantum mechanics methods were applied to predict the glutathione reactivity and, in turn, toxicity for chemicals acting via Michael addition and S_N2 mechanisms (Chapters 3 and 4 for Michael acceptors and Chapter 5 for chemicals acting via an S_N2 mechanism). In this chapter, the theory behind the quantum mechanics methods used will be outlined. The chapter will begin by discussing one of the fundamental equations in quantum mechanics - the Schrödinger equation (Section 2.1.1). The Born-Oppenheimer approximation, Hartree-Fock theory and the use of Linear Combinations of Atomic Orbitals (LCAO) will then be discussed as approaches towards (approximately) solving the Schrödinger equation (Sections 2.1.1 – 2.3). The limitations behind these methods will also be discussed. Finally, an alternative approach to solving the Schrödinger equation, density functional theory (the method for calculations used in Chapters 3-5) will be presented (Section 2.3 onwards). The information within this chapter was taken from several literature sources (87-90).

2.1 Molecular orbital calculations

There are various quantum theories for the treatment of molecular systems. One of the most widely used is molecular orbital theory which describes the distribution of electrons in molecules in a similar way to how atomic orbitals describe the distribution of electrons in atoms. One of the key fundamental concepts of molecular orbital calculations is that electrons within orbitals can be expressed as moving through space like a wave, this wave is represented by the wave function (Ψ). The wavefunction is used in the Schrödinger equation and contains all information that can be determined experimentally for a molecule, such as dipole moments and polarizability, although the wave function itself is not experimentally measurable. The following sections outline the Schrödinger equation and the various approximations that are required in order to solve it for multi-electron systems.

2.1.1 The Schrödinger equation

In 1926, Erwin Schrödinger proposed that if the motion of a particle (for example, an electron) behaves like a wave, it should be possible to describe it using a wave equation. This led to the development of the Schrödinger equation (Equation 2.1).

$$\left\{-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)+v\right\}\Psi(\mathbf{r},\mathrm{t})=i\hbar\frac{\partial\Psi(\mathbf{r},\mathrm{t})}{\partial t} \tag{2.1}$$

This equation refers to a single particle of mass (m) moving through space with position (**r** = position vector with coordinates x, y and z*)* over time (t) under the influence of an external field (*v*). ℏ refers to Plank's constant divided by 2π and *i* is the square root of -1. Assuming that the potential energy *v* does not change with time, this results in the time-independent Schrödinger equation:

$$\left\{-\frac{\hbar^2}{2m}\nabla^2+v\right\}\Psi(\mathbf{r})=E\Psi(\mathbf{r}) \tag{2.2}$$

where:

$$\nabla^2=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2} \tag{2.3}$$

It is usual to abbreviate the left side of the equation to *H* known as the Hamiltonian operator:

$$\mathcal{H}=-\frac{\hbar^2}{2m}\nabla^2+v \tag{2.4}$$

The concept of an operator in quantum mechanics is important. An operator is a mathematical representation of measurable parameters within a physical system. For example, quantitative values (such as energy, position or linear momentum) can be determined by using the appropriate operator. The Hamiltonian is one of the most commonly used operators which can be used to determine the energy of a system. This is determined by calculating the following integral (note that Ψ^* means the wave function may be a complex number):

$$E = \frac{\int \Psi^* \mathcal{H} \Psi dt}{\Psi^* \Psi dt} \qquad (2.5)$$

As a result, it is possible to simplify the time-independent Schrödinger equation to:

$$\mathcal{H}\Psi = E\Psi \qquad (2.6)$$

In order to solve the Schrödinger equation the values of E and function Ψ need to be determined. This equation takes on the standard eigenfunction form where an eigenfunction is a function that is "operated" on by an operator to produce a value (also called an eigenvalue) multiplied by the eigenfunction. In this case, the Hamiltonian operator acts on the Ψ (eigenfunction) operator, the equation returns the wave function multiplied by the energy (eigenvalue). The Schrödinger equation cannot be solved exactly for systems that involve three or more interacting particles (meaning that it can only be solved exactly for a hydrogen atom which has a single electron and a single proton). This is because many methods that aim to solve the Schrödinger equation do not account for the interactions between electrons (electron-electron correlation; for example, Hartree-Fock theory - see Section 2.2). Therefore, any solution for systems with two or more interacting electrons can only be an approximation to the true solution of the Schrödinger equation. In order to solve the Schrödinger equation (Equation 2.6), a number of key parameters are required, determination of these two values allows the energy of a system to be calculated using Equation 2.5. These parameters being:

- The solution to the Hamiltonian ($\mathcal{H}$ in Equation 2.6): This provides information on the kinetic energy of particles and interactions between particles (electrons and nuclei, see Section 2.2).

- The solution to the electronic wavefunction: This is a mathematical function that describes the motion of the electron through space (related to atomic and molecular orbitals, see Section 2.1.3).

2.1.2 The Born-Oppenheimer approximation and the electronic Hamiltonian

In order to solve the Schrödinger equation, one needs to define both the Hamiltonian and the wavefunction. This would enable the energy of the system to be calculated using Equation 2.5. The Hamiltonian is concerned with the interactions between electrons and nuclei. Electrons move much faster than the nuclei (as they are much lighter); therefore, wherever the nuclei are, the electrons will change their position almost instantaneously for those given nuclei. This disparity in motion leads to the Born-Oppenheimer approximation in which the nuclei are considered as fixed with only the motion of the electrons being considered in the Hamiltonian. Applying this approximation results in the electronic wavefunction which enables the Schrödinger equation to be solved (to a good approximation). Consequently, the total wavefunction of a molecule under the Born-Oppenheimer approximation can be expressed as:

$$\Psi(\text{nuclei}, \text{electrons}) = \Psi(\text{electrons})\Psi(\text{nuclei}) \qquad (2.7)$$

where the total energy of the molecule is a sum of the nuclear charge (the electrostatic repulsion between positively charged nuclei) and the electronic energy (interactions between electrons with each other and with the nucleus):

$$E_{\text{total}} = E_{\text{nuc}}(\text{nuclei} - \text{nuclei}) + E(\text{electrons} - \text{nuclei and electrons} - \text{electrons}) \quad (2.8)$$

As stated above, applying the Hamiltonian operator to the electronic wave function enables the energy of a multi-electron system to be determined. Applying the Born-Oppenheimer approximation allows for the definition of the electronic Hamiltonian operator for a molecule with nuclei A, B and electrons i and j (Equation 2.9, in atomic units).

$$\mathcal{H} = \underbrace{\sum_i \frac{\nabla_i^2}{2}}_{T_e} - \underbrace{\sum_{A,i} \frac{Z_A}{r_{Ai}}}_{V_{en}} + \underbrace{\sum_{A,i} \frac{Z_A Z_B}{R_{AB}}}_{V_{nn}} + \underbrace{\sum_{i>j} \frac{1}{r_{ij}}}_{V_{ee}} \qquad (2.9)$$

where T_e is the operator for the kinetic energy of the electrons, V_{en} is the operator for the Coulomb attraction between electrons and nuclei, V_{nn} is the repulsion between nuclei and V_{ee} is electron repulsion. R represents the distance between the respective particles and Z represents the atomic number of the nucleus. Note that there is no term representing the kinetic energy of the nuclei as these are assumed static under the Born-Oppenheimer approximation, hence this is called the electronic Hamiltonian operator. This provides the solution to the electronic Hamiltonian ($\mathcal{H}$) as part of the Schrödinger equation (Equation 2.6). However, in order to calculate the energy of the system (Equation 2.5) and thus fully solve the Schrödinger equation a solution to the wavefunction must also be derived.

2.2 The Wavefunction and Hartree-Fock theory

So far, the electron Hamiltonian under the Born-Oppenheimer approximation has been defined as part of the solution to the Schrödinger equation. The remaining portion that needs to be solved is the wavefunction, which involves determining a set of orbitals that corresponds to the minimum energy of the system. One method of determining the wavefunction for polyelectronic systems is through Hartree-Fock theory. The main difference between systems with a single electron and polyelectronic systems is the interactions between electrons. Three contributing interactions occur between particles in such a system:

- Coulomb interactions - The potential energy from the interaction between charged particles (electrons with electrons and electrons with nuclei). This is dependent upon the distance between the interacting particles.

- Exchange interactions – Exchange interactions occur through exchange (the changing of position) of indistinguishable electrons, for example electrons with the same spin. This results in a change in sign (+ / -) when expressing these electrons as part of a wavefunction (with respect to the anti-symmetric principle, see wavefunction section of Section 2.2). This is the basis of the Pauli exclusion principle which states that two or more electrons with

identical spin cannot occupy the same quantum state. This means that two electrons in the same orbitals must have opposite spins.

- Core interactions – This describes the motion of a single electron moving in a field of bare nuclei and does not take into consideration electron-electron correlation. This is dependent on the distance between the electron and nuclei.

When attempting to solve the electronic wavefunction for polyelectronic systems, all three contributions need to be taken into consideration. It is important to note that in the Hartree-Fock approach electrons in spin orbitals are assumed to be moving through an average charge distribution made up of the other electrons. As a result, the position of these electrons are assumed to not be influenced by other electrons. Therefore, Hartree-Fock methods do not take into account electron correlations, this is the key drawback with pure Hartree-Fock approaches. The solution to Hartree-Fock is obtained through the use of the Fock operator (f) which is obtained through Equation 2.10.

$$f_i(1) = \mathcal{H}^{Core}(1) + \sum_{j=1}^{N} \{ \mathcal{I}_j(1) - K_j(1) \} \tag{2.10}$$

where $\mathcal{H}^{Core}$ represents the core Hamiltonian interactions and $\mathcal{I}_j(1) - K_j(1)$ represent contributions from coulomb ($\mathcal{I}_j(1)$) and exchanged ($K_j(1)$) interactions. This can be simplified to:

$$f_i x_i = \sum_j \varepsilon_{ij} x_j(1) \tag{2.11}$$

This takes the standard eigenfunction form:

$$f_i x_i = \varepsilon_i x_i \tag{2.12}$$

The Fock operator (f) is a one-electron Hamiltonian for an electron in a polyelectronic system. The Fock operator acts on the molecular orbital x_i (the eigenfunction) and returns a series of orbitals x_i multiplied by a set of (unknown) orbital coefficients ε_i (the eigenvalue). Equation 2.12 is

obtained by defining the three terms for electrons (core, coulomb and exchange) which are present in the Hamiltonian operator. Considering an electron in orbital x_i the Hamiltonian operator can be written using the three terms for the three different contributions to the energy:

$$\left[-\frac{1}{2}\nabla_1^2 - \sum_{A=1}^{M} \frac{Z_A}{r_{iA}} \right] x_i(1) + \sum_{j \neq i} \left[\int d\tau_2 x_j(2)x_j(2)\frac{1}{r_{12}} \right] x_1(1) - \sum_{j \neq i} \left[\int d\tau_2 x_j(2)x_j(2)\frac{1}{r_{12}} \right] x_1(1)$$

$$= \sum_j \varepsilon_{ij} x_j(1) \tag{2.13}$$

Equation 2.13 can be simplified by introducing expressions for the three operators, these being:

The core Hamiltonian operator $\mathcal{H}^{Core}(1)$:

$$\mathcal{H}^{Core}(1) = -\frac{1}{2}\nabla_1^2 - \sum_{A=1}^{M} \frac{Z_A}{r_{1A}} \tag{2.14}$$

The Coulomb operator ($\mathcal{I}_j(1)$):

$$\mathcal{I}_j(1) = \int d\tau_2 x_j(2)\frac{1}{r_{12}} x_j(2) \tag{2.15}$$

The exchange operator ($K_J(1)$):

$$K_j(1)x_i(1) = \left[\int d\tau_2 x_j(2)\frac{1}{r_{12}} x_i(2) \right] x_j(1) \tag{2.16}$$

This means Equation 2.13 can be re-written as:

$$\mathcal{H}^{Core}(1)x_i(1) + \sum_{j \neq i}^{N} \mathcal{I}_j(1)x_i(1) - \sum_{j \neq i}^{n} K_j(1)x_i(1) = \sum_j \varepsilon_{ij} x_j(1) \tag{2.17}$$

This results in following equation:

$$\left[\mathcal{H}^{Core}(1) + \sum_{j=1}^{N} \{\mathcal{I}_j(1) - K_j(1)\} \right] x_i(1) = \sum_j \varepsilon_{ij} x_j(1) \tag{2.18}$$

Equation 2.18 s then simplified to derive the Fock operator (Equation 2.10). Note that the electrons are assumed to be in a fixed field composed of nuclei and other electrons. This has important implications for the solution to the Hartree-Fock equation, as electrons will naturally affect the other electrons in the system (electron-electron correlation). This lack of a term for electron-electron correlation in the Hamiltonian is the key drawback to Hartree-Fock theory. In order to solve the Hartree-Fock equation (Equation 2.12) the coefficient values (c_i) that correspond to optimal orbitals (i.e. the minimum - the ones which are lowest in energy for a given geometry) need to be obtained. Additionally, there needs to be an appropriate method of describing the orbitals in order to solve the Hartree-Fock equation (Equation 2.12). Assuming that there is a way to describe the orbitals, the solution to finding the best set of coefficients can be determined by adopting a Self-Consistent Field (SCF) approach. This is an iterative process and works as follow:

1. Orbital coefficients (c_i) are initially estimated

2. These are then used to calculate the Coulomb and exchange operators

3. The Hartree-Fock equations are then solved, giving a second set of orbital coefficients (c_i)

4. The energy of the system is calculated

5. Steps 1-4 are repeated until the energy of the system no longer decreases for consecutive sets of orbital coefficients. The wavefunction is said to be self-consistent at this point

Having defined a method for determining a set of coefficients for a given set of orbitals, the next step is to find a method of describing the orbitals themselves that are used within the SCF approach. One solution is to express the wavefunction as a product of one-electron spin orbitals. This requires the following assumptions: 1) electrons do not interact with one another and 2) the total electronic wave function describing the motion of the two electrons is equal to the product of two hydrogen atom spin orbitals. However, in order to have a valid wavefunction, the wavefunction must also satisfy the antisymmetric principle. This states that the expression of electrons as wavefunctions must be antisymmetric with respect to the exchange of any two electron co-ordinates i.e. a change

in sign (+ to − and vice versa). This is the basis of the Pauli exclusion principle which states that two or more electrons with identical spin cannot occupy the same quantum state. This means that two electrons in the same orbitals must have opposite spins. Considering the electronic wavefunction in this way results in:

$$\Psi(X_2, X_1) = x1(X_2)\, x2(X_1) - x2(X_2)\, x1(X_1) \tag{2.19}$$

where $X1$ and $X2$ are electrons containing special co-ordinates including spin of the respective electrons and $x1$ and $x2$ are spin orbitals. This can be then be expressed as a Slater determinant with the general form:

$$\Psi(X_1, X_2 \ldots X_n) = \frac{1}{\sqrt{N!}} \begin{vmatrix} x1(X_1) & x2(X_1) & \ldots & xn(X_1) \\ x1(X_2) & x2(X_2) & \cdots & xn(X_2) \\ \vdots & \vdots & \ddots & \vdots \\ x1(X_n) & x2(X_n) & \cdots & xn(X_n) \end{vmatrix} \tag{2.20}$$

where N is the number of electrons. Each column in the matrix denotes an orbital and each row denotes an electron such that each electron is associated with every orbital.

This shows that the wavefunction can be expressed as a Slater determinant which is a product of one-electron spin orbitals (where spin orbitals are a product of special and spin functions). This provides a way of describing the orbitals whilst the Hartree-Fock equations (Equations 2.10 − 2.18) provide a method of determining which coefficient values correspond to the "best" orbitals. This is useful for describing atomic orbitals of atoms; however, in polyatomic molecules the atomic orbitals overlap and form molecular orbitals. In quantum mechanics molecular orbitals are accounted for by using a Linear Combination of Atomic Orbitals (LCAO) approach. In using the LCAO approach, each molecular orbital can be written as a summation in the following form:

$$\psi_i = \sum_{\mu=1}^{K} C_{\mu i} \phi_\mu \tag{2.21}$$

where ψ_i is a spatial molecular orbital, ϕ_μ is one of K atomic orbitals and $C_{\mu i}$ is a coefficient. For example the bonding molecular orbital (resulting from constructive interaction between atomic

orbitals) of a hydrogen molecule (labelled $1\sigma_g$) can be expressed as a combination of two hydrogen 1s orbitals ($1S_A$ and $1S_B$). The antibonding molecular orbital (labelled $1\sigma_u$ as a result of destructive interference of atomic orbitals) can be expressed as the difference in the two hydrogen 1s orbitals and is higher in energy than the bonding molecular orbitals:

$$1\sigma_g = A\,(1S_A + 1S_B) \tag{2.22}$$

$$1\sigma_u = A^*\,(1S_A - 1S_B) \tag{2.23}$$

where A and A^* are normalisation factors. The one electron functions (ϕ_μ) are commonly called basis functions and typically correspond to atomic orbitals. In accordance with variation theorem the coefficient values ($C_{\mu i}$) that are required are those which give the lowest energy wavefunction (i.e. when the energy is at a minimum). As previously, the solution to these coefficients can be found using the SCF approach.

2.2.1 Basis set

As shown by Equation 2.21 a linear combination of one-electron orbitals ϕ_ν (also called basis functions) can be used to represent each spin orbital. The two commonly used basis functions in electronic structure calculations are: Slater Type Orbitals (STO) and Gaussian Type Orbitals (GTO). Slater type orbitals are more accurate but some of the integrals in STOs are difficult to solve and increase in mathematical complexity once several electrons are introduced. Fortunately, it is possible to model STOs using a linear combination of GTOs. Although a minimum of three times as many GTOs are required to reach a similar level of accuracy to a single STO. The number of GTOs included in the basis sets can vary which can have an effect on the level of accuracy of the basis set. The common ones being minimal basis set, double zeta basis set, triple zeta basis set or a split valence basis set. Additionally, there are functions that can be included to augment basis sets for example polarization and diffuse functions. These different basis sets and functions are discussed below.

2.2.1.1 Minimal basis set

A minimal basis set is constructed using one basis function (STO or GTO) for each atomic orbital in the atom. For example, a hydrogen atom would consist of a 1s orbital requiring a single basis function. A carbon atom consists of 1s, 2s, $2p_x$, $2p_y$ and $2p_z$ orbitals meaning that a total number of five basis functions would be required. Therefore, a molecule such as methane (CH_4) would use a total number of nine basis function, one for each hydrogen atom (1 x 4) and five for the carbon atom. Common minimal basis sets include STO-3G and STO-4G (generally STO-nG) where n is the number of GTOs used to represent each orbital. As stated previously, three times as many GTOs are required to properly represent each STO; therefore, STO-3G is considered the minimum sized basis set that can be used in *ab inito* molecular orbital calculations. Additionally, the computational effort increases with the number of functions in the Gaussian expression - meaning that using an STO-4G basis set requires more computational effort than using an STO-3G basis set. Typically, the choice of basis set is usually a compromise between accuracy and computational cost.

2.2.1.2 Double zeta and triple zeta basis sets

In comparison to a minimal basis set, a double zeta basis set doubles the number of all basis functions per orbital. Therefore, a hydrogen atom would use two basis functions, a carbon atom would use 10 basis functions and a methane molecule would use a total number of 18 basis functions. Similarly triple zeta triples the number of basis functions required (H = three basis functions, C = 15 basis functions, CH_4 = 27 basis functions).

2.2.1.3 Split valence basis set

In many cases, it is too computationally demanding to calculate the number of functions required for every orbital when using a double or triple zeta basis set. Split valence basis sets are a variation of double zeta sets in that the number of functions used to describe the atomic orbitals is double for the valence orbitals only (rather than for all orbitals). This approach can be rationalised in terms

of the fact that chemical bonding occurs between valence orbitals only. For example, a double zeta split valence basis set for a hydrogen atom would still require two basis functions as there is only a single 1s valence orbital. A double zeta split valence basis set for the valence orbitals of a carbon atoms would require eight basis functions – two each for the n = 2 valence orbitals (2s, $2p_x$, $2p_y$ and $2p_z$). In contrast, the 1s inner shell only requires a single basis function (as it is being described using a minimal basis set). Thus, a double zeta split valence basis set for carbon has a total of nine basis functions. A molecule of methane (CH_4) would require a total number of 17 basis functions using a double zeta split valence basis set reducing the computational effort required in comparison to a full double zeta basis set (which would require 18 basis functions).

The notation for split-valence basis sets takes the form X-YZG where X = the number of GTOs used to describe the STOs for the core orbitals, Y = the number of GTOs used for the first STO and Z = the number of GTOs used in the second STO (recalling that two STOs are required per orbital in a double zeta basis set). For example, the 3-21G basis set uses three GTOs to describe the core orbitals. The valence orbitals are described with a combination of two GTOs for the first STO and a single GTO for the second STO. Common split-valence basis sets include 3-21G, 4-31G and 6-31G. This approach can be extended to produce split valence triple zeta basis sets where the valence orbitals are described using three GTOs e.g. 6-311G.

2.2.1.4 Polarization functions

The basis functions discussed so far have been centred on atomic nuclei. However, the charge distribution about an atom in a molecule is usually perturbed in comparison to a single atom. For example, the 1s orbital in the hydrogen atom is symmetrical. As this hydrogen atom approaches another atom (such as a carbon atom) the 1s orbitals will be polarized when mixed with the carbon atoms 2p orbitals upon bonding. This polarization is not described with the use of s-functions alone; however, the addition of p-orbital functions enables this process to be correctly accounted for. Similarly, d-orbital functions can be used to describe the polarization of p-orbitals and f-orbital

functions for the polarization of d-orbitals (Figure 2.2). The use of polarization basis functions is indicated by the use of an asterisk (*). For example, a 6-31G* basis set uses a double zeta split valence basis set with polarization being taken into account for the p-orbitals. An additional asterisk (e.g. 6-31G**) indicates that polarization has also been taken into account for the s orbitals. These polarized basis sets can be additionally expressed as 6-31G(d) for 6-31G* and 6-31G(d,p) for 6-31G**.

$2p_z$ $3d_{xy}$
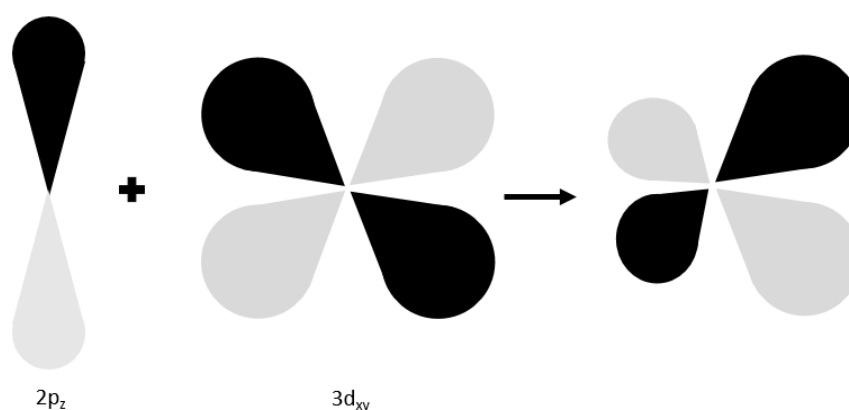
Figure 2.2: The addition of d type orbital functions (orbital $3d_{xy}$) into p orbitals ($2p_z$).

2.2.1.5 Diffuse functions

All of the basis sets discussed so far fail to deal with species such as anions and molecules containing lone pairs of electrons. These systems have a significant amount of electron density far away from the atomic nucleus. To resolve this, diffuse functions can be added to the basis set. The use of diffuse functions is indicated with a "+" for a set of s and p type GTOs e.g. 6-31+G, whilst "++" indicates that diffuse functions are included for hydrogen atom as well as heavy atoms e.g. 6-31++G.

2.2.2 Summary of Hartree-Fock theory

Hartree-Fock theory provides a way of determining which orbitals result in the "best wavefunction" (i.e. that with the lowest energy). This is achieved by using the SCF method where the solution is iteratively refined to achieve the "best" set of orbital coefficients (i.e. those which result in the

lowest energy) until further iterations result in no change (at this point it is said to be self-consistent). The use of Hartree-Fock theory assumes that electrons in spin orbitals are moving through an average charge distribution of other electrons. Although this provides a suitable means of solving the Schrödinger equation, it fails to account for electron-electron correlation. This has led to the development of alternative methods to account for electron-electron correlation, the primary example being density functional theory.

2.3 Density functional theory

So far, this chapter has discussed methods that provide an approximate solution to the Schrödinger equation, primarily Hartree-Fock theory. In Hartree-Fock theory, the electrons are assumed to be moving in an average potential of other electrons and the position of these electrons are assumed to be uninfluenced by other electrons. Importantly, Hartree-Fock methods do not take into account electron-electron correlation. Density Functional Theory (DFT) attempts to deal with these shortcomings by replacing the many-body electronic wavefunction with the electron density. In a similar way to how Hartree-Fock is constructed from a set of N single-electron wavefunctions, DFT considers single-electron functions. The idea of DFT was developed with the underlying concept that there is a relationship between the total electron energy and the overall electronic density. This concept is supported by a study by Hohenberg and Kohn in 1964 that shows that ground state energies along with other properties of a system can be uniquely defined by the electron density. This can be expressed by stating that energy is a function of electron density p(r). This is called a functional (a function of a function). In DFT, the energy functional is written as a sum of two terms:

$$E[p(r)] = \int V_{ext}(r)p(r)dr + F[p(r)] \qquad (2.24)$$

The first term ($V_{ext}(r)$) represents the interaction of the electrons with the external potential (typically Coulombic interactions with the nuclei). Whilst the last term $F[p(r)]$ is the sum of the kinetic energy of the electrons and the contribution of interelectronic interactions. Similar to

Hartree-Fock, the lower the energy value is, the closer it is to the true energy of the ground state structure. This allows a variation approach to be used to achieve the best functional i.e. the lower in energy the better. However, the key problem with this equation is that it is unknown what the function $F[p(r)]$ is. It was proposed by Kohn and Sham in 1965 that $F[p(r)]$ could be approximated as a sum of three terms.(91)

$$F[p(r)] = E_{KE}[p(r)] + E_H[p(r)] + E_{XC}[p(r)] \qquad (2.25)$$

where $E_{KE}[p(r)]$ represents the kinetic energy, $E_H[p(r)]$ is the electron-electron Coulombic energy and $E_{XC}[p(r)]$ is the contribution from electron-electron exchange and electron-electron correlation. It is important to note that the first term $E_{KE}[p(r)]$ is defined as the kinetic energy of a system of non-interacting electrons:

$$E_{KE}[p(r)] = \sum_{i=1}^{N} \int \psi_i(r)\left(-\frac{\nabla^2}{2}\right)\psi_i dr \qquad (2.26)$$

The term $E_H[p(r)]$ (also known as the Hartree electrostatic energy) arises from interaction between two charge densities resulting in:

$$E_H[p(r)] = \frac{1}{2}\iint \frac{p(r_1)p(r_2)}{|r_1 - r_2|}dr_1 dr_2 \qquad (2.27)$$

Substituting Equations 2.26 and 2.27 into Equation 2.25 and adding an additional term for the electron-nuclear interaction results in:

$$E[p(r)] = \sum_{i=1}^{N} \int \psi_i(r)\left(-\frac{\nabla^2}{2}\right)\psi_i dr + \frac{1}{2}\iint \frac{p(r_1)p(r_2)}{|r_1 - r_2|}dr_1 dr_2 + E_{XC}[p(r)]$$

$$- \sum_{A=1}^{M} \int \frac{Z_A}{|r - R_A|}p(r)dr \qquad (2.28)$$

Importantly, the above equation defines the contributions of the exchange-correlation energy function $E_{XC}[p(r)]$ and the contribution due to the difference between the true kinetic energy of the system $E_{KE}[p(r)]$.

The electron density p(r) of a system can be represented as a sum of one-electron orbitals:

$$p(r) = \sum_{i=1}^{N} |\psi_i(r)|^2 \tag{2.29}$$

Using this expression and applying the appropriate variational conditions results in the one-electron Kohn-Sham equation:

$$\left\{ -\frac{\nabla_i^2}{2} + v(r) + \int \frac{p(r_2)}{r_{12}} dr_2 + V_{XC}[r_1] \right\} \psi_i(r_1) = \varepsilon_i \psi_i(r_1) \tag{2.30}$$

where $v(r)$ is the external potential, the orbitals' energies are expressed as ε_i and V_{XC} is the exchange-correlation function which is related to the exchange correlation energy by:

$$V_{XC}[r] = \left(\frac{\delta E_{XC}[p(r)]}{\delta p(r)} \right) \tag{2.31}$$

This allows the total electronic energy to be calculated using Equation 2.28.

Similar to the solution of Hartree-Fock, the Kohn-Sham equations are solved using a self-consistent approach where an initial estimation for electron density p(r) is used to solve Equation 2.30. This derives a set of orbitals resulting in an improved value for the electron density. This value is then fed back into Equation 2.30 until the values are self-consistent (see SCF method for Hartree-Fock in Section 2.1.3).

The appeal behind DFT comes from the inclusion of an electron-electron exchange-correlation functional (Equation 2.25). In DFT, even simple approximations to the exchange-correlation functional can give favourable results. For example, studies have shown that the geometry of chemicals optimized through DFT to agree with experimentally determined geometry in

comparison to HF methods (92). Additionally, DFT has been shown to provide a better estimation of the energy change for organic reactions in comparison to HF (93). The simplest approximation to the contribution of the exchange-correlation functional is the Local Density Approximation (LDA) where the functional at a given position is computed exclusively from the value of the electron density at that position (and is therefore local). This method assumes that the electron density corresponds to that of a homogenous electron gas where the electron density is constant. The total exchange-correlation energy for a system can then be obtained by integrating over all space:

$$E_{XC}[p(r)] = \int p(r)\varepsilon_{XC}(p(r))dr \qquad (2.32)$$

where $\varepsilon_{XC}(p(r))$ is the exchange-correlation energy per electron as a function of the electron density in the homogenous electron gas. The exchange-correlation functional is obtained by differentiation of Equation 2.33:

$$V_{XC}[r] = p(r)\frac{d\varepsilon_{XC}(p(r))}{dp(r)} + \varepsilon_{XC}(p(r)) \qquad (2.33)$$

Although LDA approaches have shown to perform surprisingly well, they are inadequate for some problems, e.g. describing hydrogen bonds.

As such, extensions of this method have been developed. For example, the Generalized Gradient Approximation (GGA) (or gradient-corrected functional) which incorporates both electron density and its gradient to better describe the non-uniformity of molecular densities. These gradient corrections are typically divided into separate exchange and correlation contributions e.g. Becke gradient exchange correction and the Lee-Yang-Parr correlation functional (BLYP). An extension of this is the use of a hybrid functional which mixes GGA with other features from *ab initio* methods (e.g. Hartree-Fock methods). This has advantages as Hartree-Fock provides an exact means of treating the exchange contribution. This approach works by combining the exchange contribution from the Hartree-Fock equation with the correlation component from the LDA. For example, B3LYP

which tends to be the most commonly used method for DFT (and is the method used throughout this thesis).

2.4 Concluding remarks

This chapter has discussed various quantum theories for the treatment of molecular systems. The Schrödinger equation was introduced as the key concept that allows various properties of a particle (e.g. an electron) to be determined. This is through the use of the wavefunction (a function that describes the wave like motion of a particle) and the Hamiltonian (an operator used to derive the energy of a system by defining the kinetic energy of particles and interactions between particles). The Hartree-Fock equation was shown to be an *ab initio* method suitable for approximately solving the Schrödinger equation by using a self-consistent method (adopting variation theorem). This method assumes electrons to be in a fixed field composed of nuclei and other electrons. Therefore, Hartree-Fock theory fails to account for electron-electron correlation. Density functional theory deals with these shortcomings by replacing the many-body electronic wavefunction with the electron density. The appeal of DFT comes from the inclusion of an exchange-correlation functional to account for electron-electron correlation. Given this, all calculations used within this thesis utilised a hybrid functional form of DFT with the inclusion of additional diffusion and polarizability functions B3LYP/6-31G+(d), see Chapters 3, 4 and 5).

Chapter 3. Development of a fragment-based *in silico* profiler for Michael addition

3.1 Introduction

It is well established that various toxicological effects can occur as a result of covalent bond formation between electrophilic chemicals and nucleophilic centres in proteins. This includes a range of toxicities such as skin sensitisation, respiratory sensitisation, aquatic toxicity and liver toxicity (28, 43, 52, 53, 70, 94, 95). A key mechanism through which a covalent bond can form between an electrophile and a protein is Michael addition. Such chemicals, known as Michael acceptors, contain a π-bond adjacent to a polarising group such as an α,β-unsaturated aldehyde (Figure 3.1) (61). The presence of the polarising group results in a partial positive charge on the β-carbon of the π-bond, causing this position to become susceptible to a reaction with a biological nucleophile, typically cysteine or lysine residues (57, 58). Computational research has suggested that Michael addition can potentially occur via one of two mechanisms depending on whether the nucleophile is anionic or neutral (49, 52, 73-75). The mechanism in which the nucleophile is an anion (e.g. methyl thiolate) results in a resonance stabilised intermediate. This resonance stabilised intermediate lies in an energy minimum between two transition states along the reaction co-ordinate. The intermediate is then protonated carbon (e.g. from the deprotonated thiol group or a deprotonated water molecule) to produce the final product known as a Michael adduct (mechanism shown in Figure 3.2, energy diagram in Figure 3.3). The rate-determining step for this mechanism being determined through the formation of the first transition state. Additional research has outlined an alternative mechanism in which nucleophilic attack occurs via a neutral species (e.g. methyl thiol) resulting in the formation of a single four-membered ring transition state structure, with the formation of this species being the rate determining step (Figure 3.2, energy diagram in Figure 3.4) (52, 73, 74).
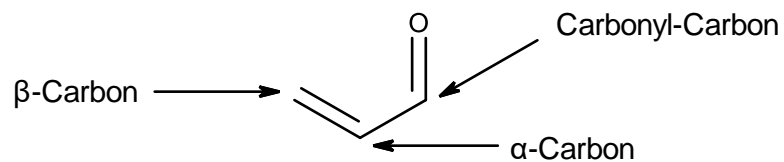
Figure 3.1: The structure of a typical Michael acceptor (prop-2-enal) showing the α-carbon, the β-carbon and the carbonyl-carbon
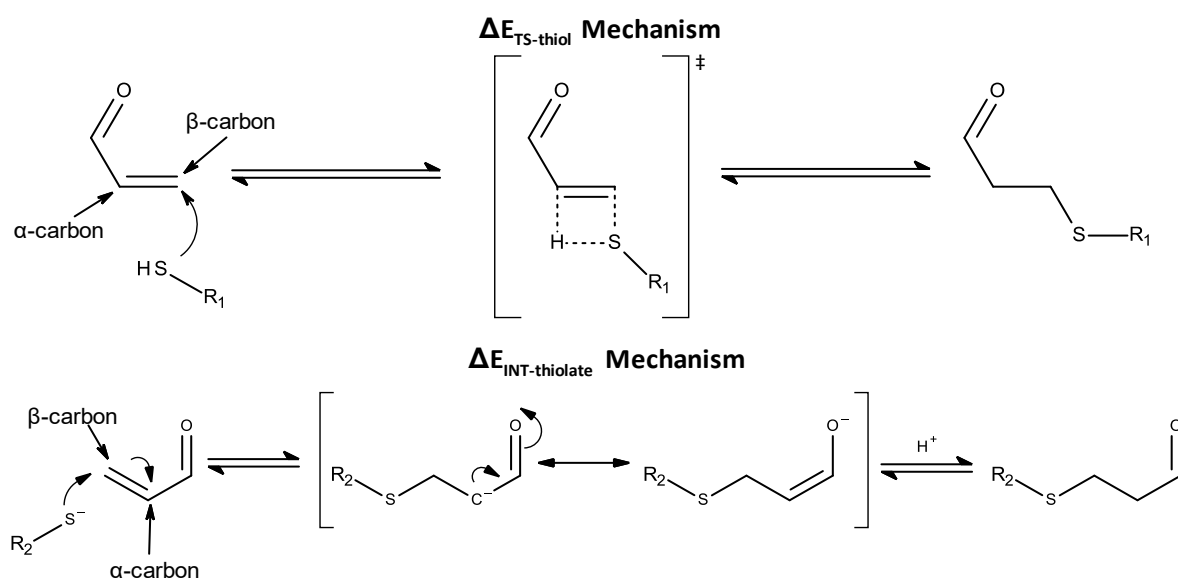
Figure 3.2: The two proposed mechanisms for Michael addition thiol reactivity resulting in the formation of a four-membered transition state structure ($\Delta E_{TS\text{-}Thiol}$) or a resonance stabilised intermediate ($\Delta E_{INT\text{-}Thiolate}$) ($R_1$ and R_2 = glutathione, alkyl)
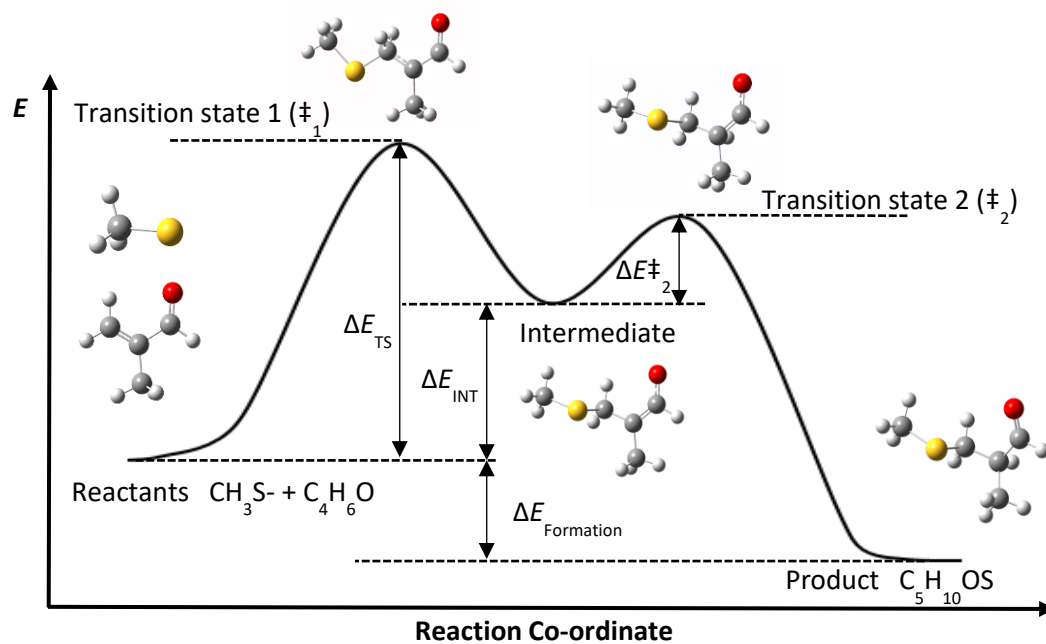
Figure 3.3: Energy diagram for the Michael addition reaction with a thiolate nucleophile through the formation of a stabilised intermediate structure (N.B. ΔE_{TS} and ΔE_{INT} correspond to $\Delta E_{TS\text{-}Thiolate}$ and $\Delta E_{INT\text{-}Thiolate}$ respectively)

Figure 3.4: Energy diagram for the Michael addition reaction with a thiol nucleophile through the formation of a four-membered ring transition state structure (N.B. ΔE_{TS} corresponds to $\Delta E_{TS\text{-}Thiol}$)

The mechanistic knowledge outlined above has allowed for the development of structural alerts enabling chemicals with the potential to react via Michael addition to be easily identified (57). As outlined in Chapter 1, structural alerts can be grouped together into *in silico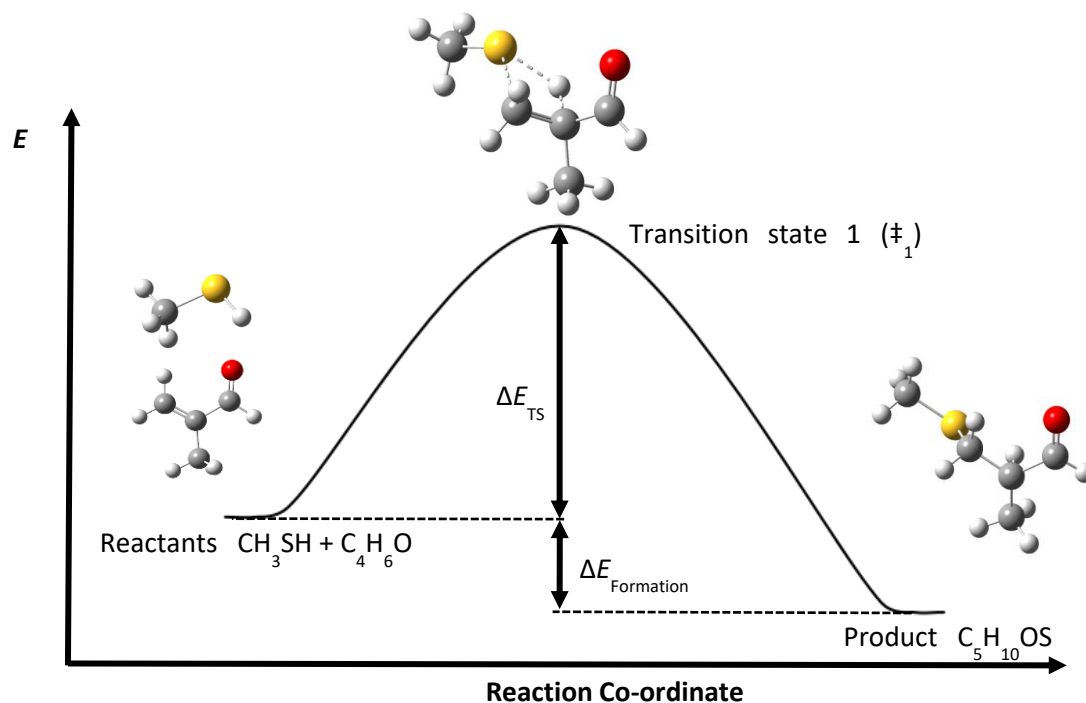* profilers which are able to identify mechanisms associated with specific toxicological outcomes, such as skin sensitisation (48). Whilst *in silico* profilers are useful for identifying features associated with potential toxicity, the information they provide is qualitative i.e. a binary yes or no for the presence of a structural feature associated with reactivity and/or toxicity. Importantly, they provide no information concerning either the rate of covalent bond formation or toxicological potency. This is information is key as when using knowledge of covalent mechanisms to predict toxicity, a primary assumption is that the rate of covalent bond formation (reactivity) is proportional to toxicological potency. 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) (42, 43, 49, 70, 95). The use of computational and *in chemico* methods for the prediction of toxicological potency is discussed in depth in Chapter 1.

Given the importance of quantifying chemical reactivity for the prediction of toxicity, it is unsurprising that there have been a number of studies that have utilised computational descriptors for the prediction of chemical reactivity and toxicity (51, 71, 95-99). These are typically derived from quantum mechanics calculations based on the structure of the reactant electrophile alone, and include descriptors such as energy values of the Highest Occupied Molecular Orbital (HOMO) and Lowest Unoccupied Molecular Orbital (LUMO) and the electrophilicity index (this index is derived from chemical potential and chemical hardness which are related to the HOMO and LUMO of a chemical) (ω) (100, 101). However, these descriptors quantify only the electronic portion of chemical reactivity and do not account for factors such as steric hindrance at the reactive site. An example of a descriptor that accounts for both electronic and steric factors is the energy of activation (Eact), this is derived by calculating the energy difference between the optimised energy

of the reactants (electrophile and nucleophile) and optimised transition state ($\Delta E_{\text{TS-Thiol}}$, Figure 3.4) or intermediate structures ($\Delta E_{\text{INT-Thiolate}}$, Figure 3.3) linking the reactants to the products (Equations 3.1 and 3.2 respectively).

$$\Delta E_{\text{TS-Thiol}} = (E_{\text{Transition state}} - E_{\text{Electrophile}} + E_{\text{Thiol}}) \tag{3.1}$$

$$\Delta E_{\text{INT-Thiolate}} = (E_{\text{Intermediate}} - E_{\text{Electrophile}} + E_{\text{thiolate}}) \tag{3.2}$$

A number of computational studies have utilised the four-coordinate transition state mechanism to predict reactivity towards glutathione (52, 73, 74). For example, the glutathione reactivity of 22 Michael acceptors (nine α,β-unsaturated aldehydes, 12 α,β-unsaturated ketones and one α,β-unsaturated ester) was successfully predicted using calculated $\Delta E_{\text{TS-Thiol}}$ (R^2 = 0.90, Model 3.1) (73). The resulting QSAR model also included descriptors relating to the rate of the reverse reaction ($\Delta E_{\text{TS-Thiol.Back}}$) and steric hindrance at the beta position (SAS(R)).

$$\text{Log } K_{\text{GSH}} = -0.03\,\Delta E_{TS-Thiol} + 1.42\,Log\,SAS(R) + 0.03\,\Delta E_{TS-Thiol-Back} - 2.14 \text{ (Model 3.1)}$$

$$N = 22, R^2 = 0.90, R^2_{\text{adj}} = 0.88, s = 0.34$$

A second study using a group of 35 Michael acceptors (10 α,β-unsaturated aldehydes, 11 α,β-unsaturated ketones and 14 α,β-unsaturated esters) also successfully used calculated $\Delta E_{\text{TS-Thiol}}$ to predict glutathione reactivity (R^2 = 0.96, Model 3.2). In contrast to the previous QSAR model outlined above, the authors of this study included an indicator variable for the presence of an α-substituent (I_a) (74). The predicted Log k_{GSH} values were subsequently used, in conjunction with a descriptor to account for hydrophobicity (Log k_{OW}), to predict toxicity to the ciliate *Tetrahymena pyriformis* for the same set of chemicals (R^2 = 0.92 for predicted Log EC_{50} against experimental Log EC_{50}) (details of the *Tetrahymena pyriformis* growth impairment assay is discussed in Chapter 1).

$$\text{Log } K_{\text{GSH}} = -0.06 \ \Delta E_{TS-Thiol} - 1.25 \ I_a + 5.08 \qquad \text{(Model 3.2)}$$

$$N = 35, R^2 = 0.96$$

In a separate study, the same authors also used the four-membered transition state approach to predict the toxicity to *Tetrahymena pyriformis* for 45 α,β-unsaturated esters (52). The authors used the model developed in the previous study (Model 3.2) to generate predicted Log k_{GSH} values. Initial modelling of the 45 chemicals in the dataset resulted in a reasonable correlation between predicted toxicity (calculated from predicted Log k_{GSH} and hydrophobicity) and experimental toxicity (R^2 of 0.68, Model 3.3). However, the authors noted that these results were inferior to the prediction of toxicity to *Tetrahymena pyriformis* for a set of 35 Michael acceptors (R^2 = 0.68 vs R^2 of 0.96 respectively for predicted Log k_{GSH} against experimental Log k_{GSH}). To improve the correlation between glutathione reactivity and toxicity to *Tetrahymena pyriformis*, the authors assigned the chemicals into one of three groups based on their reaction rate and hydrophobicity. The first group (15 highly reactive chemicals with low hydrophobicity, Model 3.4a-c for aldehydes, ketones and esters respectively) and the second group (12 moderately reactive chemicals with moderate hydrophobicity, Model 3.5) were predicted using calculated $\Delta E_{\text{TS-Thiol}}$ values and Log k_{OW} values. The authors suggested that this meant that the aquatic toxicity of these two groups of chemicals was being primarily driven by the rate of covalent bond formation (see Chapter 1 section 1.2.1.1 for aquatic toxicity mechanisms). The toxicity of the third group (18 slow reacting chemicals with high hydrophobicity, Model 3.6) was well-predicted based on Log k_{OW} alone suggesting these chemicals exert their toxicity through narcosis. The overall predictivity for the 45 chemicals (using three separate models for the three groups) significantly improved (R^2 = 0.83 versus 0.68 for predicted EC$_{50}$ against experimental Log EC$_{50}$, statistics for individual models 3.4 – 3.6 are not provided by the authors).

$$\text{Log EC}_{50} = -0.44 \; Log \; k_{OW} - 0.50 \; Log \; k_{GSH} - 2.26 \qquad \text{(Model 3.3)}$$

$$N = 45, R^2 = 0.68$$

$$\text{Log EC}_{50} = -0.45 \; Log \; k_{OW} - 0.51 \; Log \; k_{GSH} - 2.4 \qquad \text{(Model 3.4a)}$$

$$\text{Log EC}_{50} = -0.36 \; Log \; k_{OW} - 0.57 \; Log \; k_{GSH} - 2.4 \qquad \text{(Model 3.4b)}$$

$$\text{Log EC}_{50} = -0.31 \; Log \; k_{OW} - 0.68 \; Log \; k_{GSH} - 2.4 \qquad \text{(Model 3.4c)}$$

$$\text{Log EC}_{50} = -0.33 \; Log \; k_{OW} - 0.67 \; Log \; k_{GSH} - 2.5 \qquad \text{(Model 3.5)}$$

$$\text{Log EC}_{50} = -0.87 \; Log \; k_{OW} - 0.67 - 0.99 \qquad \text{(Model 3.6)}$$

Research has also shown that it is possible to use the energy difference of a resonance stabilised intermediate of the Michael addition reaction instead of the transition state structure, $\Delta E_{INT\text{-}Thiolate}$ to predict glutathione reactivity and skin sensitisation potency (49, 75). This was achieved by using methyl thiolate as a nucleophile as opposed to methyl thiol (mechanism in Figure 3.2). For example, a recent drug discovery study used calculated $\Delta E_{INT\text{-}Thiolate}$ alone to predict the reactivity towards glutathione (expressed as reaction half-lives, $t_{1/2}$) for a set of 16 acrylamides acting through Michael addition ($R^2 = 0.92$ for reactions half-lives, $t_{1/2}$ against $\Delta E_{INT\text{-}Thiolate}$, no Model provided by the authors). In contrast to the work utilising $\Delta E_{TS\text{-}Thiol}$ values discussed previously, the modelling did not require any additional descriptors for steric or electronic effects at either the α- or β-carbons. However, it is worth noting that such descriptors may not be required due to the similarity of the chemicals in the data set (15 of the 16 chemicals were unsubstituted at the α-carbon and 14 of the 16 chemicals were unsubstituted at the β-carbon) (75).

An additional study used calculated $\Delta E_{INT\text{-}Thiolate}$ as a descriptor to predict the skin sensitisation potency (as determined in the local lymph node assay, discussed in detail in Chapter 1) for a set of 25 Michael acceptors (Model 3.7). A descriptor for the solvent accessible surface area (SAS) at the β-carbon was also included as a measurement of steric bulk around the reaction site (49). However, this approach failed to predict the skin sensitisation potency of six chemicals due to them being volatile or having the ability to polymerise.

$$pEC3 = -1.60\,\Delta E_{INT-Thiolate} + 0.02\,SAS \qquad \text{(Model 3.7)}$$

$$N = 25, R^2 = 0.79, R^2_{adj} = 0.78, s = 0.13$$

Experimentally determined reactivity has also been used for the prediction of skin sensitisation potential for Michael acceptors. This was demonstrated in a recent study which determined the kinetic rate constants of 14 Michael acceptors (43). This study successfully predicted the skin sensitisation potential of 10 Michael acceptors (Model 3.7) whilst three chemicals were identified as outliers. These chemicals were *trans*-2-hexenal, ethylene glycol dimethacrylate and ethyl acrylate. The chemicals 2-hexenal and ethyl acrylate were found to be volatile which resulted in loss of the chemical from the surface of the skin. This resulted in a lower skin sensitisation potency than would be expected based on reactivity alone. It was hypothesised that the final outlier, ethylene glycol methacrylate is less potent than predicted due to free-radical polymerisation under LLNA conditions.

$$pEC3 = 0.24\,Log\,k_{GSH} + 2.11 \qquad \text{(Model 3.7)}$$

$$N = 25, R^2 = 0.79, R^2_{adj} = 0.78, s = 0.13$$

The use of calculated $\Delta E_{INT\text{-}Thiolate}$ has also been used to predict skin sensitisation potency for a group of 12 chemicals acting via a nucleophilic aromatic substitution (S_NAr) mechanism (Model 3.8, R^2 =

0.64 for $\Delta E_{INT\text{-}Thiolate}$ against pEC3). The authors also used the same descriptor to develop a classification model capable of discriminating between sensitizers and non-sensitizers (non-sensitizers had reaction barriers > 10 kcal/mol) (76). One chemical (chlorothalonil) was considered an outlier for being more potent than predicted due to its low hydrophobicity (resulting in an improved R^2 of 0.74 from 0.64 for $\Delta E_{INT\text{-}Thiolate}$ against pEC3).

$$\text{pEC3} = -0.31\,\Delta E_{INT-Thiolate} + 4.90 \qquad \text{(Model 3.8)}$$

$$N = 12, R^2 = 0.62$$

It is clear from the literature that *in silico* methods involving the calculation $\Delta E_{TS\text{-}Thiol}$ or $\Delta E_{INT\text{-}Thiolate}$ are capable of predicting both chemical reactivity and, in turn, toxicity. The published studies have demonstrated that both descriptors have a similar degree of predictivity. However, both approaches require the use of computationally demanding quantum chemical calculations requiring access to commercial software. This limits their use, and inclusion, in freely available *in silico* tools currently finding widespread use in regulatory toxicology (for example, the OECD QSAR Toolbox). Given this, the aim of this chapter was to develop a fragment-based *in silico* profiler for Michael addition. The advantages of a fragment-based *in silico* profiler are the ability to generate predictions quicker than current quantum mechanics methods and without the need to use proprietary software. The concept is similar to that of a traditional *in silico* profiler in that query chemicals (inputted as SMILES) are compared to a database of information (in this case fragments encoded as SMARTS) to generate an output. In the case of a fragment-based profiler the output is a quantitative prediction of $\Delta E_{INT\text{-}Thiolate}$ instead of the qualitative output yes/no prediction of a traditional *in silico* profiler (discussed in detail in Chapter 1). The concept of the fragment-based *in silico* profiler is summarised in Figure 3.5.
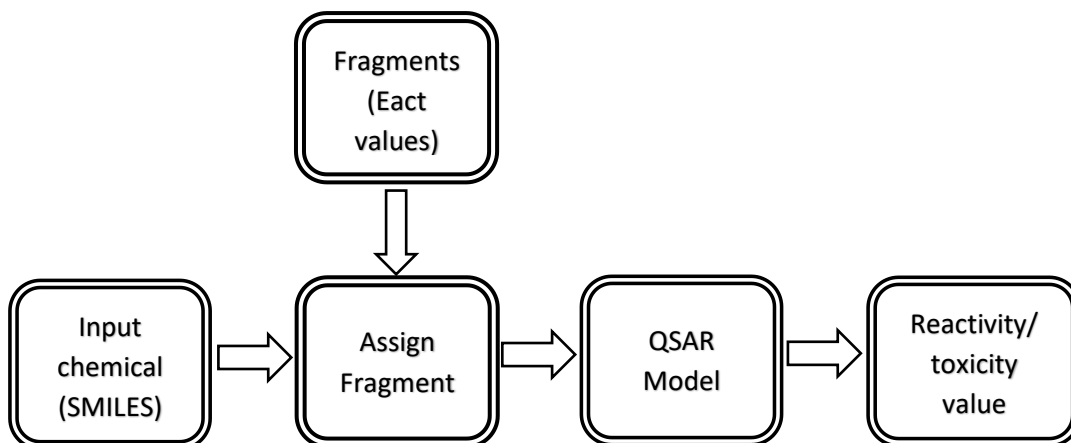
Figure 3.5. Concept of the *in silico* fragment-based profiler to predict reactivity

Given the similarity in the published predictivity when using either $\Delta E_{INT\text{-Thiolate}}$ or $\Delta E_{TS\text{-Thiol}}$ values, the method used calculated $\Delta E_{INT\text{-Thiolate}}$ as this allows values to be obtained without the need for transition state calculations. This chapter focuses on the development of the fragments required for the fragment-based profiler for Michael addition (the application of the fragment-based profiler for Michael addition in predicting thiol reactivity and related toxicity can be found in Chapter 4).

3.2 Methods

3.2.1 Dataset

Fragments were developed for linear α,β-unsaturated aldehydes, ketones and esters with varying alkyl and aryl substitutions at three positions (Figure 3.6). Profiling of a database of 212 Michael acceptors with glutathione reactivity data showed 54 chemicals were within the domain of linear α,β-unsaturated aldehydes, ketones and esters (alkyl and aryl). The modelling of the reactivity of these 54 chemicals using the fragment-based reactivity profiler is discussed in Chapter 4. An additional 17 chemicals (nine α,β-unsaturated nitros, three α,β-unsaturated nitriles and six α,β-unsaturated cyclic ketones) were also included as part of the model validation in Chapter 4. The remaining 141 chemicals were excluded for various reasons: 64 chemicals belonged to additional chemical groups (containing sulphur, phosphorous, halogen or furan functional groups), 41

chemicals had multiple sites of reactivity, 20 chemicals were alkynes, eight chemicals had di-β

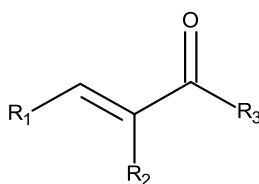substitutions, five chemicals were un-reactive and three had extended conjugation.

Figure 3.6: General structure for α,β-unsaturated aldehydes (R$_3$ = H), α,β-unsaturated ketones (R$_3$

= C) and α,β-unsaturated esters (R$_3$ = OC). R$_1$ and R$_2$ = H, varying alkyl and aryl substitutions

3.2.2 Computational methods

All calculations were carried out using the Gaussian 09 suite of software using density functional

theory at the B3LYP/6-31+G(d) level of theory using water as a solvent employing Polarizable

Cotinuum Model (PCM) as an implicit solvation model (102). Energies of activation ($\Delta E_{\text{INT-Thiolate}}$)

values were calculated using thiolate as a model nucleophile (See Equation 3.2). This was achieved

by calculating the energy of the optimised structures for the thiolate nucleophile, the Michael

acceptor and the intermediate structure (three calculations). All optimisations were obtained using

the "loose" keyword (as there could potentially be a large number of calculations to carry out in

developing a fragment-based method). The use of a thiolate (rather than a thiol) nucleophile allows

an intermediate to be isolated on the potential energy surface. This significantly simplifies the

calculations as the intermediate can be isolated using a simple energy minimisation calculation

rather than a transition state calculation. This approach is in keeping with previous research (49).

Calculated $\Delta E_{\text{INT-Thiolate}}$ values in the fragment analysis section (Sections 3.4.2 – 3.4.3) are quoted to

the nearest kcal/mol (calculated using Equation 3.1). Values for $\Delta E_{\text{TS-Thiolate}}$ were obtained using scan

calculations to determine the highest point of energy for the reaction between the electrophile and

nucleophile. Scan calculations were carried out with a bond length of 2.9 Å between the β-carbon

and the sulphur of the thiol nucleophile. A series of seven calculations were then carried out where

the bond length between the β-carbon and the sulphur of the thiol nucleophile decreased by 0.1 Å with each calculation. This allowed the bond length which resulted in the highest energy structure (i.e. the transition state) to be determined. This resulted in a total number of nine calculations required for the transition state structure (one for each reactant and seven to determine the transition state structure). Transition state structures were subsequently characterised by the presence of a single negative eigenvalue connecting the transition state to the reactants and product on the potential energy surface.

3.2.2 Statistical analysis

Minitab (version 17) was used for the linear regression analysis in Section 3.4.1 in order to identify the level of correlation (R^2 values) between $\Delta E_{TS-Thiolate}$ and $\Delta E_{INT-Thiolate}$ for a representative series of Michael acceptors. Additional linear regression analysis for the prediction of glutathione reactivity, toxicity to *Tetrahymena pyriformis* and skin sensitisation is discussed in Chapter 4.

3.3 Investigation of activation energies for Michael addition chemicals

The key aspect of the fragment-based *in silico* profiler for Michael addition is the use of a database containing fragments with pre-calculated activation energy values. As discussed above, two different descriptors as a measurement of activation energy have been successfully used to predict reactivity and toxicity ($\Delta E_{INT-Thiolate}$ and $\Delta E_{TS-Thiol}$). The calculation of $\Delta E_{INT-Thiolate}$ is computationally less demanding than $\Delta E_{TS-Thiol}$ (three calculations versus several). As such the calculations in this chapter focused on the determination of $\Delta E_{INT-Thiolate}$. However, using the energy of the intermediate structure ($\Delta E_{INT-Thiolate}$) as a substitute for the transition state depends on the assumption that the energy difference between the transition state and intermediate is consistent between all Michael acceptors. Currently there have been no studies which investigate the transition state structure which results from the reaction between the Michael acceptor and a thiolate nucleophile ($\Delta E_{TS-Thiolate}$). Additionally, there are no studies which have investigated the correlation between $\Delta E_{INT-Thiolate}$ and $\Delta E_{TS-Thiolate}$. Given this, an investigation was carried out to assess the correlation between

these two descriptors. A good correlation would show that the easier to calculate $\Delta E_{INT\text{-}Thiolate}$ could be used as a descriptor for the fragment database. Additionally, the effect on the calculated $\Delta E_{INT\text{-}Thiolate}$ values when utilising the two key conformers for electrophile structures was also investigated.

3.3.1 Analysis of transition state energy versus intermediate energy values

An analysis was carried out to investigate the level of correlation between $\Delta E_{INT\text{-}Thiolate}$ and $\Delta E_{TS\text{-}Thiolate}$ for a series of representative Michael acceptors (prop-2-enal, but-2-enal, buten-2-one, penten-2-one, methyl prop-2-enoate and methyl but-2-enoate). The calculated $\Delta E_{INT\text{-}Thiolate}$ and $\Delta E_{TS\text{-}Thiolate}$ were calculated based on the cis intermediate geometry (see section 3.3.2). The calculated $\Delta E_{TS\text{-}Thiolate}$ and $\Delta E_{INT\text{-}Thiolate}$ showed significant differences, with the intermediates being consistently lower in energy (Table 3.1). The results showed the α,β-unsaturated aldehydes to have the lowest calculated $\Delta E_{INT\text{-}Thiolate}$ and $\Delta E_{TS\text{-}Thiolate}$ values (chemicals 1 and 2 in Table 3.1), whilst the α,β-unsaturated esters were the highest in energy (chemicals 5 and 6 in Table 3.1). The calculations also accounted for the effects of steric hindrance at the reaction site with chemicals with no alkyl substituent at the β-carbon being lower in energy than their substituted equivalents (compare chemicals 1, 3 and 5 with 2, 4 and 6 respectively in Table 3.1). Importantly, the calculations also showed the energy values for transition state structures and intermediate structures to be highly correlated (R^2 = 0.96, Figure 3.8). This is unsurprising given that the geometries of the transition state and intermediate structures showed them to be extremely similar in nature, with the major difference being a shorter C-S bond length for the intermediate (Table 3.1). This analysis supports the use of the simpler method to calculate $\Delta E_{INT\text{-}Thiolate}$ (as opposed to $\Delta E_{TS\text{-}Thiolate}$) as the descriptor for reactivity for use in the fragment-based *in silico* profiler for Michael addition.
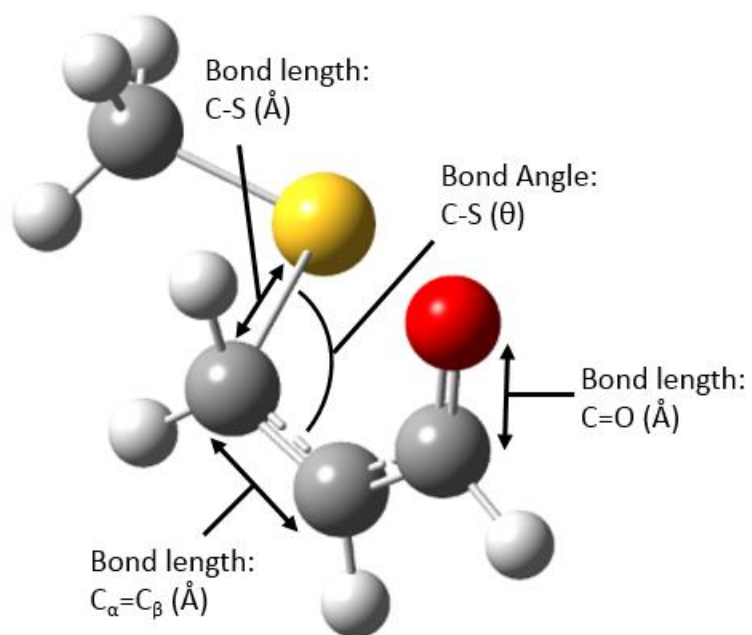
Figure 3.7. The key structural features of Michael acceptors. Bond angle and bond length values for intermediate and transition state structures shown in Table 3.1

Table 3.1: Key bond angles and bond lengths shown in Figure 3.7 for intermediate and transition state structures

ID	Chemical	C-S (θ) INT	C-S (θ) TS	C-S (Å) INT	C-S (Å) TS	$\Delta E_{INT\text{-}Thiolate}$ (kcal/mol)	$\Delta E_{TS\text{-}Thiolate}$ (kcal/mol)
1	Prop-2-enal	112.1	116.8	1.9	2.6	-5.4	2.5
2	But-2-enal	108.5	106.5	1.9	2.6	-1.5	5.1
3	Buten-2-one	112.4	104.6	1.9	2.7	-0.8	5.6
4	Penten-2-one	108.6	110.7	1.9	2.5	4.5	9.7
5	Methyl prop-2-enoate	113.2	117.4	1.9	2.5	4.9	7.7
6	Methyl but-2-enoate	109.4	110.8	1.9	2.3	10.2	12.4

Figure 3.8: The correlation between $\Delta E_{TS\text{-Thiolate}}$ (kcal/mol) and $\Delta E_{INT\text{-Thiolate}}$ (kcal/mol) for a representative set of Michael acceptors (Table 3.1)

3.3.2 The effect of *cis* and *trans* intermediates on calculated $\Delta E_{INT\text{-Thiolate}}$

The previous analysis established that calculated $\Delta E_{INT\text{-Thiolate}}$ can be used as opposed to $\Delta E_{TS\text{-Thiolate}}$ as a descriptor for reactivity for use in t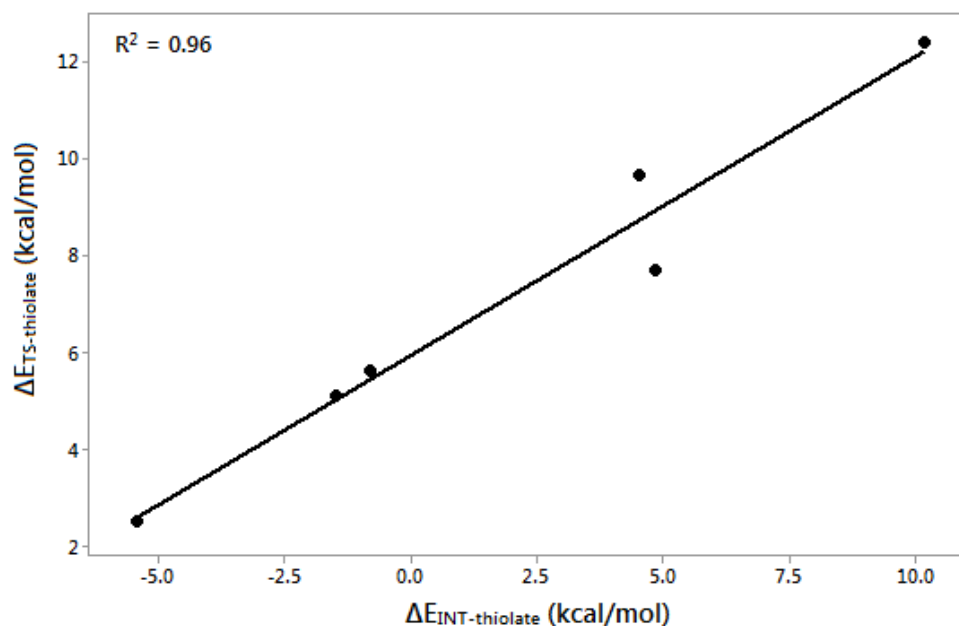he fragment-based *in silico* profiler for Michael addition. In addition to this, an investigation into the effect of the potential for $\Delta E_{INT\text{-Thiolate}}$ to vary due to rotation around the carbon-carbon double bond was also undertaken. Analysis of the reactant structure showed that there are two potential starting geometries for Michael acceptors. These being the *cis*-conformer where the carbon-carbon double bond is in the *cis*-position to the carbonyl, and the *trans*-conformer where the carbon-carbon double bond is in the *trans*-position to the carbonyl. These two conformers are related to each other via a 180° rotation around the α-carbon and the carbonyl-carbon. For example, there are two possible conformers for prop-2-enal (Figure 3.9).

α-carbon | α-carbon
cis-conformer | *trans*-conformer
$\Delta E_{\text{TS-thiolate}} = -5.4$ | $\Delta E_{\text{TS-thiolate}} = -2.6$
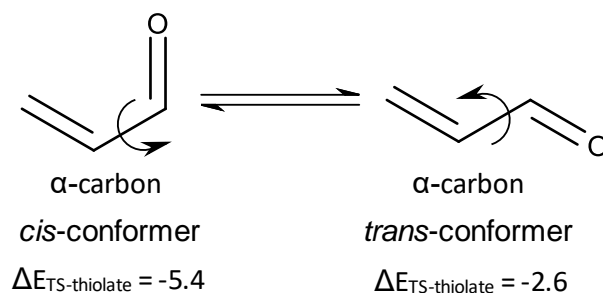
Figure 3.9. The two possible conformers of prop-2-enal highlighting the α-carbon

Upon nucleophilic attack by a thiolate ion electron density in the π-bond is shifted resulting in a resonance-stabilised intermediate. This results in the formation of a partial double bond between the carbonyl-carbon and the α-carbon resulting in restricted rotation around this bond. This partial double bond results in the formation of two potential geometrical isomers for the intermediate, denoted here as a *cis*-intermediate and a *trans*-intermediate (mechanism showing the formation of each isomer is given in Figure 3.10).
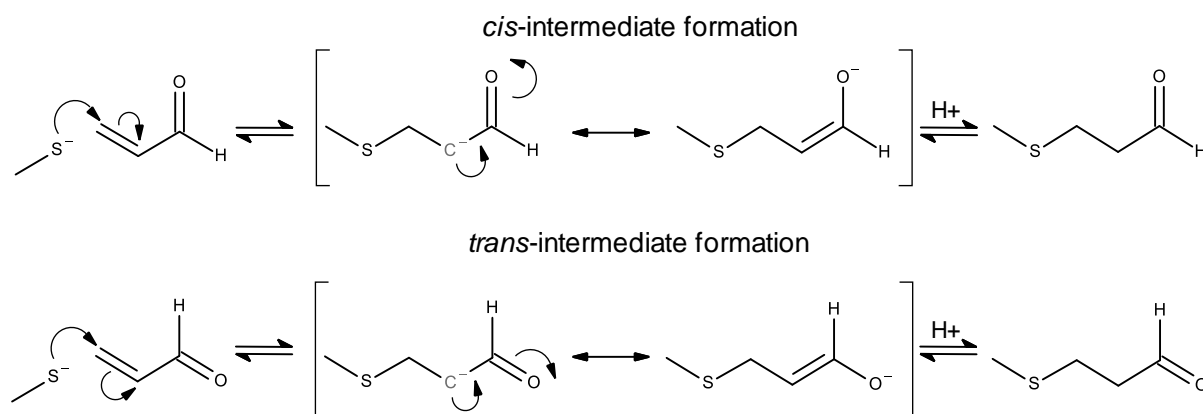
Figure 3.10: The different intermediate geometries of Michael acceptors reacting with a thiolate nucleophile

Calculating $\Delta E_{\text{INT-Thiolate}}$ for the chemicals in the previous analysis showed that the formation of the *cis*-intermediate was lower in energy than the formation of the *trans*-intermediate for all six chemicals (Figure 3.11, individual values shown in Table 3.2). This effect was greatest for α,β-unsaturated aldehydes and ketones (chemicals 1 – 4 in Table 3.3). Whilst the difference in energy

was significantly lower for the α,β-unsaturated esters (chemicals 5 and 6 in Table 3.2). This analysis highlighted the importance of having a consistent approach towards the choice of reactant conformations from which to calculate $\Delta E_{\text{INT-Thiolate}}$ (more so for α,β-unsaturated aldehydes and ketones than the α,β-unsaturated esters). The formation of the *cis*-intermediate was calculated to be lower in energy for all chemicals investigated, and was thus utilised for all subsequent fragment calculations in the development of the fragment-based *in silico* profiler for Michael addition.
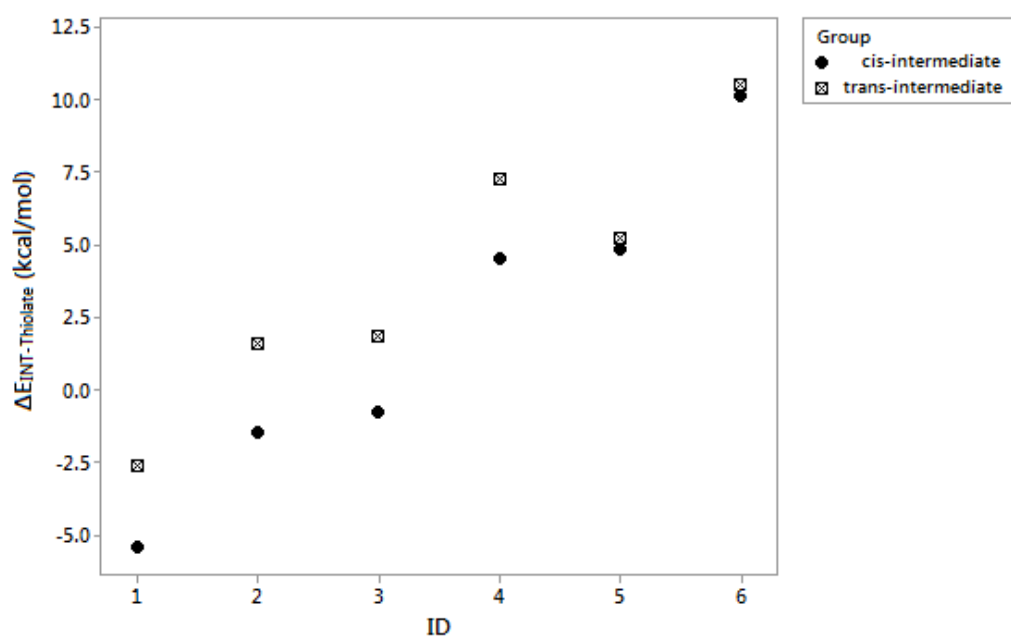

Figure 3.11: The calculated $\Delta E_{\text{INT-Thiolate}}$ difference between *trans* (⊠) and *cis* (●) intermediates. Structures are: (1) prop-2-enal, (2) but-2-enal, (3) buten-2-one, (4) penten-2-one, (5) methyl-prop-2-enoate and (6) methyl-but-2-enoate

Table 3.2: Calculated $\Delta E_{INT\text{-}Thiolate}$ (kcal/mol) values for *cis* and *trans* structures (Figures 3.9)

ID	Chemical	*Cis*-intermediate $\Delta E_{INT\text{-}Thiolate}$ (kcal/mol)	*Trans*-intermediate $\Delta E_{INT\text{-}Thiolate}$ (kcal/mol)	*Trans*-intermediate $\Delta E_{INT\text{-}Thiolate}$ − *Cis*-intermediate $\Delta E_{INT\text{-}Thiolate}$ (kcal/mol)
1	Prop-2-enal	-5.4	-2.6	2.8
2	But-2-enal	-1.5	1.6	3.1
3	Buten-2-one	-0.8	1.9	2.7
4	Penten-2-one	4.5	7.2	2.7
5	Methyl prop-2-enoate	4.9	5.2	0.3
6	Methyl but-2-enoate	10.2	10.5	0.3

3.4 Development of fragments for Michael addition

The above analysis showed that $\Delta E_{INT\text{-}Thiolate}$ was highly correlated with the true transition state for thiolate addition ($\Delta E_{TS\text{-}Thiolate}$) meaning $\Delta E_{INT\text{-}Thiolate}$ could be used as the descriptor associated with the fragments. Additionally, the analysis showed that the $\Delta E_{INT\text{-}Thiolate}$ for chemicals in the *cis*-intermediate were lower in energy than the equivalent values for the *trans*-intermediate As such $\Delta E_{INT\text{-}Thiolate}$ for chemicals in the *cis*-conformation was used as a descriptor for all fragments. Having established this, the next step was to develop the fragments required to cover the applicability domain for α,β-unsaturated aldehydes, ketones and esters (chemical groups excluded from this domain are stated in the methods, Section 3.2.1). This is covered in Section 3.4.1 – 3.5. Sections 3.4.2 and 3.4.3 outline the development of fragments for α,β-unsaturated aldehydes and for α,β-unsaturated ketones and esters respectively. This next section (Section 3.4.1) will deal with the protocol for developing fragments. This protocol was adopted in order to enable a systematic approach to the SAR analysis. The SAR analysis aimed to investigate the variability of $\Delta E_{INT\text{-}Thiolate}$ upon changing the substitutions at the various R-groups for the Michael acceptors. This was to establish the point at which $\Delta E_{INT\text{-}Thiolate}$ no longer varied with a change in substitution at the respective R-group.

3.4.1 Rules for fragment development

In order to enable a systematic approach to the fragment development, the following protocol was utilised:

1. All $\Delta E_{INT\text{-}Thiolate}$ values were calculated for fragments from the *cis*-intermediate.

2. The $\Delta E_{INT\text{-}thiolate}$ values for straight chains at each R-position were compared with the $\Delta E_{INT\text{-}Thiolate}$ values of a chain containing one carbon less (or in the case of methyl with hydrogen). For example, ethyl was compared to methyl and propyl compared to ethyl.

3. Branched chains $\Delta E_{INT\text{-}Thiolate}$ values were compared to the $\Delta E_{INT\text{-}Thiolate}$ value of their straight chain equivalent. For example, *i*-propyl and *t*-butyl were compared with ethyl.

4. Aryl substitutions were compared to the equivalent alkyl chain (or in the case of benzene with hydrogen). For example, methyl benzene with methyl, ethyl benzene with ethyl.

5. Only one R group was investigated at a time whilst the other R groups remained constant. (e.g. R_1 and R_2 remained as hydrogen whilst the effect of substituents at the R_3 position was investigated).

6. Individual calculated $\Delta E_{INT\text{-}Thiolate}$ were rounded to the nearest integer before comparing values rather than rounding the difference in $\Delta E_{INT\text{-}Thiolate}$ between the two values (see Figure 3.12).

7. A cut off value of > 1.0 kcal/mol was used to assess if there was a significant difference in $\Delta E_{INT\text{-}Thiolate}$ following a change in substituents (see Figure 3.12).

8. Unrounded $\Delta E_{INT\text{-}Thiolate}$ values for fragments (to 1 decimal place) were used in the modelling of reactivity and toxicity in Chapters 4 and 5 (unlike rounded values which are used in the analysis).

To explain rules 6, 7 and 8 in more detail, consider *i*-propyl and *t*-butyl substitution at the R_1 position for α-β-unsaturated aldehydes. The $\Delta E_{INT\text{-}Thiolate}$ is calculated to be 3.7 kcal/mol (rounded to 4.0 kcal/mol, see rule 6) and 5.6 kcal/mol (rounded to 6.0 kcal/mol, see rule 6) for α,β-unsaturated

aldehyde substituted by these fragments at the R_1 position respectively. Both these fragments were compared to an ethyl substituent at the R_1 position (ethyl substituent $\Delta E_{INT\text{-}Thiolate}$ value being 2.7 kcal/mol, 3.0 kcal/mol when rounded). The results showed that there was a difference in $\Delta E_{INT\text{-}Thiolate}$ of 1.0 kcal/mol between ethyl and *i*-propyl, whilst there was a difference of 3.0 kcal/mol between ethyl and *t*-butyl (summarized in Figure 3.12). Therefore, an ethyl group could be used to predict the $\Delta E_{INT\text{-}Thiolate}$ of an *i*-propyl group (within 1.0 kcal/mol difference, see rule 7), but not *t*-butyl (as there is a >1.0 kcal/mol difference). The results of this analysis meaning that to cover branched chains for α,β-unsaturated aldehydes substituted at the R_1 position ethyl (to cover *i*-propyl substation) and *t*-butyl were included in the fragment database. Note that although the integer rounded values were used in the fragment analysis, the calculated $\Delta E_{INT\text{-}Thiolate}$ values assigned to fragments in the database were rounded to 1 decimal place (i.e. 2.7 kcal/mol and 5.6 kcal/mol would be used for ethyl and *t*-butyl respectively at R_1, see rule 8 and Figure 3.12).

$\Delta E_{INT\text{-}Thiolate}$ = 3.7 kcal/mol
(Rounded to 4.0 kcal/mol)

$\Delta E_{INT\text{-}Thiolate}$ = 2.7 kcal/mol
(Rounded to 3.0 kcal/mol)

$\Delta E_{INT\text{-}Thiolate}$ = 5.6 kcal/mol
(Rounded to 6.0 kcal/mol)

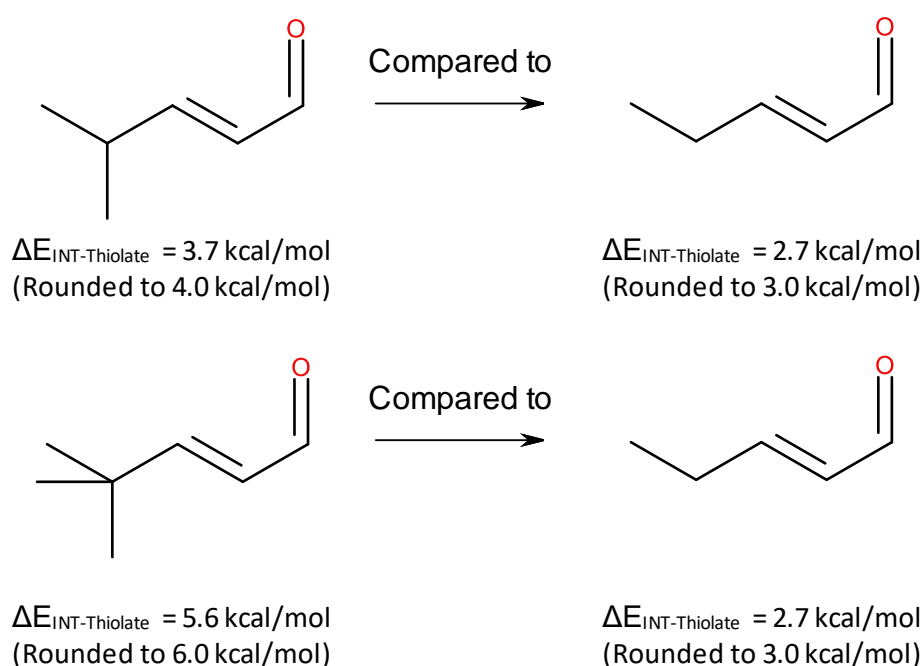$\Delta E_{INT\text{-}Thiolate}$ = 2.7 kcal/mol
(Rounded to 3.0 kcal/mol)

Figure 3.12: The analysis of fragments for α,β-unsaturated aldehydes containing *i*-propyl and *t*-butyl substitution at the R_1 positions

3.4.2 Development of fragments for α,β-unsaturated aldehydes

The eight rules stated above allowed the SAR in terms of $\Delta E_{INT\text{-}Thiolate}$ to be investigated for positions R_1 and R_2 for α,β-unsaturated aldehydes. Where R_1 and R_2 were hydrogen, linear alkyl chains, methyl, ethyl and propyl), branched alkyl chains (*i*-propyl or *t*-butyl), or aryl (phenyl, benzyl, phenethyl and 3-phenylpropane) substituents. All $\Delta E_{INT\text{-}Thiolate}$ values in the following fragment analysis are rounded to the nearest integer (see rule 6).

Calculated $\Delta E_{INT\text{-}Thiolate}$ SAR at position R_1 for α,β-unsaturated aldehydes

Initially the SAR for the variability of $\Delta E_{INT\text{-}Thiolate}$ when extending the chain length at R_1 position for α,β-unsaturated aldehydes was investigated. The calculated $\Delta E_{INT\text{-}Thiolate}$ value increased by 3 kcal/mol when extending the chain length from hydrogen to methyl substitution (compare chemical 2 with 1 in Table 3.3). A larger increase of 5 kcal/mol was calculated when the alkyl substituent at R_1 was extended from methyl to ethyl (compare chemical 3 with 2 in Table 3.3). In contrast, calculated $\Delta E_{INT\text{-}Thiolate}$ remained relatively constant when extending the alkyl chain further from ethyl to propyl (compare chemical 3 with 4 in Table 3.3). This was also seen in calculated $\Delta E_{INT\text{-}Thiolate}$ values when the alkyl chain was extended further (up to, and including, pentyl - Figure 3.13). This showed that $\Delta E_{INT\text{-}Thiolate}$ for an ethyl substituent at the R_1 can be used for the prediction of a propyl group or larger.
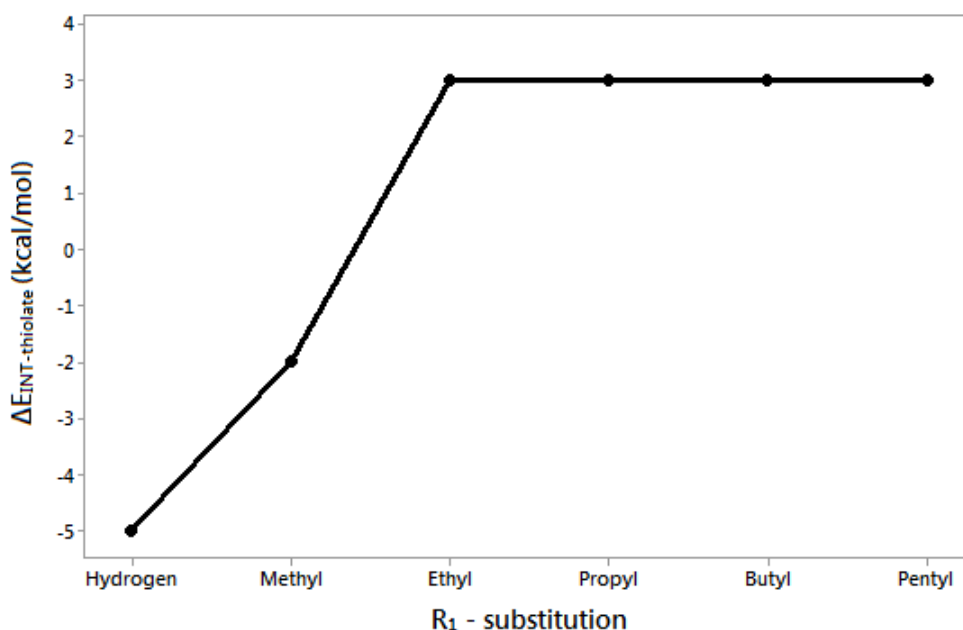
Figure 3.13: The change of calculated $\Delta E_{INT\text{-Thiolate}}$ (kcal/mol) for α,β-unsaturated aldehydes with increasing alkyl chain length at the R_1 position. All $\Delta E_{INT\text{-Thiolate}}$ are rounded to the nearest integer

In addition to linear alkyl chains, both *i*-propyl and *t*-butyl groups were also included in the analysis. These substituents were compared to their linear chain equivalents i.e. both *i*-propyl and *t*-butyl were compared to ethyl (Figure 3.12). An energy difference of 1 kcal/mol was calculated between *i*-propyl and ethyl (compare chemical 5 with 3 in Table 3.3), whilst an energy difference of 3 kcal/mol was calculated between *t*-butyl and ethyl substitution (compare chemical 6 with 3 in Table 3.3). This analysis showed that an ethyl chain was a suitable fragment for the prediction of calculated $\Delta E_{INT\text{-thiolate}}$ for *i*-propyl. In contrast, the calculations showed that a fragment for a *t*-butyl group needed to be included at the R_1 position. This is presumably due to its increased size (compared to its parent linear alkyl chain, ethyl) making nucleophilic attack by the thiolate ion more difficult due to steric hindrance. Figure 3.14 summarises the fragments that are required to predict the calculated $\Delta E_{INT\text{-Thiolate}}$ for linear and branched α,β-unsaturated aldehydes (denoted as the parent chemical in Figure 3.14). The effect of aryl groups on the calculated $\Delta E_{INT\text{-Thiolate}}$ for the α,β-unsaturated aldehydes was also investigated at the R_1 position. The results showed that the calculated $\Delta E_{INT\text{-Thiolate}}$ increases significantly by 8 kcal/mol on going from hydrogen to phenyl

(compare chemical 7 with 1 in Table 3.3). As expected, the results also showed that calculated $\Delta E_{INT\text{-}Thiolate}$ decreased by 2 kcal/mol upon the addition of a CH_2 group between the β-carbon of the alkene and the benzene ring (compare chemical 8 with 7 in Table 3.3). In terms of defining fragments for the effect of a benzene ring it is necessary to compare the aryl substituent with the corresponding alkyl substituent. For example, an energy difference of 3 kcal/mol was calculated when comparing the $\Delta E_{INT\text{-}Thiolate}$ value for phenyl to methyl (compare chemical 8 with 2 in Table 3.3). This energy difference is significantly in excess of the 1 kcal/mol cut-off. In contrast, an energy difference of 1 kcal/mol was calculated when comparing the $\Delta E_{INT\text{-}Thiolate}$ values of ethyl and phenethyl (compare chemicals 9 and 3 in Table 3.3). This means that three fragments were required to define the effect of a benzene ring at the R_1 position: phenyl, benzyl and an ethyl chain. The ethyl fragment being used to predict chemicals with a benzene ring two or more carbons away from the β-carbon of the alkene.

Table 3.3: Calculated $\Delta E_{INT\text{-}Thiolate}$ (kcal/mol) values for α,β-unsaturated aldehydes substituted at the R_1 position (R groups as defined in Figure 3.6). The fragment substituent column indicates the substituent used to predict the $\Delta E_{INT\text{-}Thiolate}$ of the respective R-group following the above analysis. The resulting fragment calculated $\Delta E_{INT\text{-}Thiolate}$ from using that substitution is also included

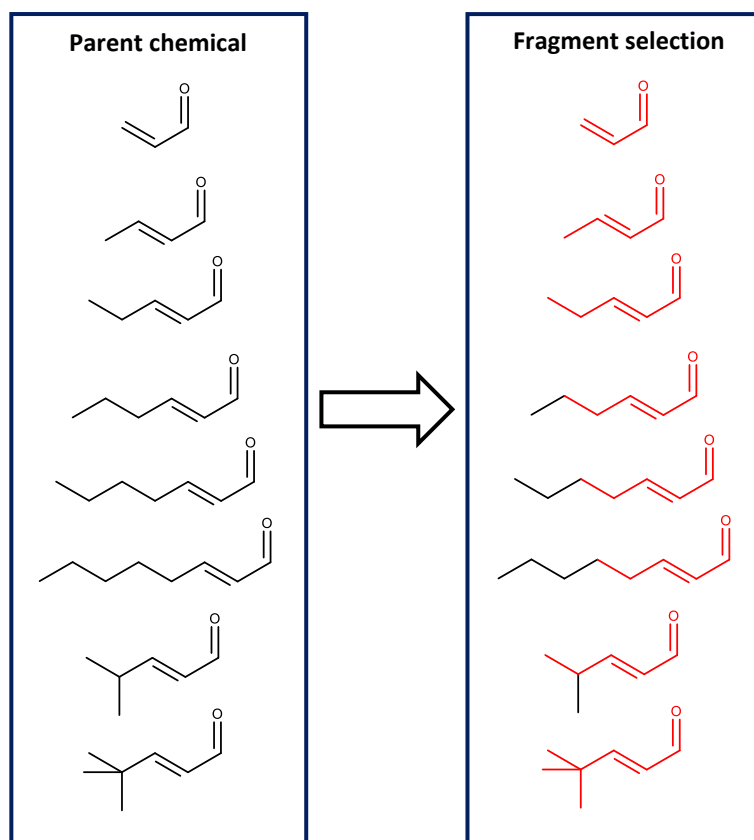| ID	Substituent name – R_1	$\Delta E_{INT\text{-}Thiolate}$ (kcal/mol)	Fragment substituent	Fragment $\Delta E_{INT\text{-}Thiolate}$ (kcal/mol)
1	Hydrogen	-5	H	-5
2	Methyl	-2	CH_3	-2
3	Ethyl	3	CH_2CH_3	3
4	Propyl	3	CH_2CH_3	3
5	*i*-Propyl	3	CH_2CH_3	3
6	*t*-Butyl	6	*t*-butyl	6
7	Phenyl	3	C_6H_5	3
8	Benzyl	1	$CH_2C_6H_5$	1
9	Phenethyl	2	CH_2CH_3	3
10	3-Phenylpropane	2	CH_2CH_3	3

Figure 3.14: The fragment selection for a series of α,β-unsaturated aldehydes substituted at the R_1 position ($\Delta E_{INT\text{-}Thiolate}$ values in Table 3.3). The rationale behind the selected fragments for the parent chemical is discussed in Section 3.4.2

Calculated $\Delta E_{INT\text{-}Thiolate}$ SAR at position R_2 for α,β-unsaturated aldehydes

An analogous analysis into the effect of alkyl chain length on the α,β-unsaturated aldehydes substituted at position R_2 showed only a small change in calculated $\Delta E_{INT\text{-}Thiolate}$ values (within 1 kcal /mol) when going from methyl to ethyl to propyl (compare chemicals 4 and 3 with 2 in Table 3.4). This enabled methyl to be used as a fragment for all linear chains at the R_2 position. However, a difference in calculated $\Delta E_{INT\text{-}Thiolate}$ of 2 kcal/mol and 4 kcal/mol was calculated when comparing both *i*-propyl and *t*-butyl with ethyl (compare chemicals 5 and 6 with 3 in Table 3.4). These calculated results showed the need to include fragments for both *i*-propyl and *t*-butyl at the R_2 position. Similarly, an analysis into the effect of aryl groups on the α,β-unsaturated aldehydes

substituted at position R_2 showed the calculated $\Delta E_{INT\text{-}Thiolate}$ value to decrease by 4 kcal/mol when going from hydrogen to phenyl (compare chemical 7 with 1 in Table 3.4). This decrease in energy can be rationalised in terms of the ability of the aromatic ring system to stabilise the negative charge present on the α-carbon in the intermediate structure (Figure 3.2). Only a small difference in $\Delta E_{INT\text{-}Thiolate}$ was calculated (within 1 kcal/mol when rounded) when comparing benzene groups that were one, two or three carbons away from the α-carbon (compare chemicals 8 with 2, 9 with 3 and 10 with 4 in Table 3.4) with a methyl group (Table 3.4). The results of this analysis showed that only phenyl had a significant effect on calculated $\Delta E_{INT\text{-}Thiolate}$ values at the R_2 position. The calculations showed that a methyl fragment could be used to predict the calculated $\Delta E_{INT\text{-}Thiolate}$ of benzene groups one or more carbons away from the α-carbon. This resulted in two fragments being required for aryl substituents at the R_2 position (methyl and phenyl).

Table 3.4: Calculated $\Delta E_{INT\text{-}Thiolate}$ (kcal/mol) values for α,β-unsaturated aldehydes substituted at the R_2 position (R groups as defined in Figure 3.6). The fragment substituent column indicates the substituent used to predict the $\Delta E_{INT\text{-}Thiolate}$ of the respective R-group following the above analysis. The resulting fragment calculated $\Delta E_{INT\text{-}Thiolate}$ from using that substitution is also included

ID	Substituent name – R_2	$\Delta E_{INT\text{-}Thiolate}$ (kcal/mol)	Fragment substituent	Fragment $\Delta E_{INT\text{-}Thiolate}$ (kcal/mol)
1	Hydrogen	-5	H	-5
2	Methyl	-2	CH_3	-2
3	Ethyl	-2	CH_3	-2
4	Propyl	-2	CH_3	-2
5	*i*-Propyl	0	*i*-Propyl	0
6	*t*-Butyl	2	*t*-Butyl	2
7	Phenyl	-9	C_6H_5	-9
8	Benzyl	-2	CH_3	-2
9	Phenethyl	-3	CH_3	-2
10	3-Phenylpropane	-3	CH_3	-2

3.4.3 Development of fragments for α,β-unsaturated ketones and esters

The approach was extended to α,β-unsaturated ketones and esters using the same substituents as were used for the α,β-unsaturated aldehydes. The α,β-unsaturated ketones and esters required an additional R group to be taken into consideration (R_3 in Figure 3.6).

Calculated $\Delta E_{INT\text{-}Thiolate}$ SAR at position R_1 for ketones and esters

For the α,β-unsaturated ketones six fragments were required at the R_1 position (individual values shown in Table 3.5) while for the α,β-unsaturated esters a total number of five fragments were required (individual values shown in Table 3.6). The same set of fragments as discussed for the α,β-unsaturated aldehydes substituted at the R_1 position were required (Section 3.4.2, Table 3.3) with the following exceptions:

- An *i*-propyl fragment was required for the α,β-unsaturated ketones as *i*-propyl was not within 1 kcal/mol of ethyl when rounded (compare chemical 5 with 3 in Table 3.5)
- A benzyl fragment was not required for either of the α,β-unsaturated ketones or esters as methyl was within 1 kcal/mol when rounded (compare chemical 8 and 1 in Table 3.5 and Table 3.6 for ketones and esters respectively)

Table 3.5: Calculated $\Delta E_{INT\text{-}Thiolate}$ values for α,β-unsaturated ketones substituted at the R_1 position (R groups as defined in Figure 3.6). The fragment substituent column indicates the substituent used to predict the $\Delta E_{INT\text{-}Thiolate}$ of the respective R-group following the above analysis. The resulting fragment $\Delta E_{INT\text{-}Thiolate}$ calculated from using that substitution is also included

ID	Substituent SMILES - R_1	$\Delta E_{INT\text{-}Thiolate}$ (kcal/mol)	Fragment substituent	Fragment $\Delta E_{INT\text{-}Thiolate}$ (kcal/mol)
1	Hydrogen	-1	H	-1
2	Methyl	5	CH_3	5
3	Ethyl	7	CH_2CH_3	7
4	Propyl	7	CH_2CH_3	7
5	*i*-Propyl	9	*i*-propyl	9
6	*t*-Butyl	10	*t*-butyl	9
7	Phenyl	7	C_6H_5	7
8	Benzyl	5	CH_3	5
9	Phenethyl	7	CH_2CH_3	7

Table 3.6: Calculated $\Delta E_{INT\text{-}Thiolate}$ values for polarised esters substituted at the R_1 position (R groups as defined in Figure 3.6). The fragment substituent column indicates the substituent used to predict the $\Delta E_{INT\text{-}Thiolate}$ of the respective R-group following the above analysis. The resulting fragment $\Delta E_{INT\text{-}Thiolate}$ calculated from using that substitution is also included

ID	Substituent name – R_1	$\Delta E_{INT\text{-}Thiolate}$ (kcal/mol)	Fragment substituent	Fragment $\Delta E_{INT\text{-}Thiolate}$ (kcal/mol)
1	Hydrogen	5	H	5
2	Methyl	10	CH_3	10
3	Ethyl	13	CH_2CH_3	13
4	Propyl	13	CH_2CH_3	13
5	*i*-Propyl	14	CH_2CH_3	13
6	*t*-Butyl	16	*t*-Butyl	16
7	Phenyl	13	C_6H_5	13
8	Benzyl	11	CH_3	10
9	Phenethyl	12	CH_2CH_3	13

Calculated ΔE$_{INT\text{-}Thiolate}$ SAR at position R2 for α,β-unsaturated ketones and esters

The analysis at the R$_2$ position for the α,β-unsaturated ketones required a total of six fragments (individual values shown in Table 3.7) whilst the analysis for α,β-unsaturated esters substituted at the R$_2$ position required a total of six fragments (individual values shown in Table 3.8). The same set of fragments as discussed for the α,β-unsaturated aldehydes substituted at the R$_2$ position were required (Section 3.4.2, Table 3.4) with the following exceptions:

- An ethyl fragment was required for both the α,β-unsaturated ketones and esters to predict aryl substituents that were two or more carbons away (compare chemicals 9 and 10 with 3 in Table 3.7 and 3.8 for ketones and esters respectively)

- An *i*-propyl fragment was not required for α,β-unsaturated ketones and esters as *i*-propyl was within 1 kcal/mol of ethyl when rounded (compare chemical 5 with 3 in Table 3.7 and Table 3.8 for ketones and esters respectively)

- A benzyl fragment was required for both the α,β-unsaturated ketones and esters as methyl was not within 1 kcal/mol of benzyl rounded (compare chemical 8 with 2 in Table 3.7 and Table 3.8 for ketones and esters respectively)

Table 3.7: Calculated $\Delta E_{INT\text{-}Thiolate}$ values for α,β-unsaturated ketones substituted at the R_2 positions

(R groups as defined in Figure 3.6). The fragment substituent column indicates the substituent used

to predict the $\Delta E_{INT\text{-}Thiolate}$ of the respective R-group following the above analysis. The resulting

fragment $\Delta E_{INT\text{-}Thiolate}$ calculated from using that substitution is also included

ID	Substituent name – R_2	$\Delta E_{INT\text{-}Thiolate}$ (kcal/mol)	Fragment substituent	Fragment $\Delta E_{INT\text{-}Thiolate}$ (kcal/mol)
1	Hydrogen	-1	H	-1
2	Methyl	5	CH_3	5
3	Ethyl	4	CH_3	5
4	Propyl	4	CH_3	5
5	*i*-Propyl	5	CH_2CH_3	5
6	*t*-Butyl	7	*t*-butyl	7
7	Phenyl	-2	C_6H_5	-2
8	Benzyl	2	$CH_2C_6H_5$	2
9	Phenethyl	3	CH_2CH_3	4
10	3-Phenylpropane	3	CH_2CH_3	4

Table 3.8: Calculated $\Delta E_{INT\text{-}Thiolate}$ values for α,β-unsaturated esters substituted at the R_2 position (R groups as defined in Figure 3.6). The fragment substituent column indicates the substituent used to predict the $\Delta E_{INT\text{-}Thiolate}$ of the respective R-group following the above analysis. The resulting fragment $\Delta E_{INT\text{-}Thiolate}$ calculated from using that substitution is also included

ID	Substituent name – R_2	$\Delta E_{INT\text{-}Thiolate}$ (kcal/mol)	Fragment substituent	Fragment $\Delta E_{INT\text{-}Thiolate}$ (kcal/mol)
1	Hydrogen	5	H	5
2	Methyl	10	CH_3	10
3	Ethyl	9	CH_3	10
4	Propyl	10	CH_3	10
5	*i*-Propyl	8	CH_3CH_2	8
6	*t*-Butyl	11	*t*-butyl	11
7	Phenyl	1	C_6H_5	1
8	Benzyl	7	$CH_2C_6H_5$	7
9	Phenethyl	8	CH_2CH_3	9
10	3-Phenylpropane	8	CH_2CH_3	9

Calculated $\Delta E_{INT\text{-}Thiolate}$ SAR at position R_3 for α,β-unsaturated ketones and esters

Unlike the α,β-unsaturated aldehydes, α,β-unsaturated ketones and esters required an additional R group to be taken into consideration (R_3 in Figure 3.6 - all values for α,β-unsaturated ketones and esters shown in Table 3.9 and Table 3.10 respectively). The analysis of linear and branched chains at the R_3 position for α,β-unsaturated ketones showed the calculated $\Delta E_{INT\text{-}Thiolate}$ values increased by 1 kcal/mol when going from methyl to ethyl (compare chemical 2 with 1 in Table 3.9). This meant that a methyl group could be used to predict the calculated $\Delta E_{INT\text{-}thiolate}$ for linear chains two or more carbons in length at the R_3 position. For the analysis of branched chains, calculated $\Delta E_{INT\text{-}Thiolate}$ values also increased by only 1 kcal/mol when going from ethyl to *i*-propyl substituents (compare chemical 4 with 2 in Table 3.9). In contrast, the calculated $\Delta E_{INT\text{-}Thiolate}$ value increased by 3 kcal/mol when going from ethyl to *t*-butyl (compare chemical 5 with 2 in Table 3.9). Therefore, a fragment

for *t*-butyl was required, but a fragment for *i*-propyl was not required as the results showed that

ethyl could be used as a fragment for *i*-propyl. Finally, the analysis for aryl groups at the R_3 position

showed a decrease in the calculated $\Delta E_{\text{INT-Thiolate}}$ values of 2 kcal/mol when going from a methyl

substituent to a phenyl substituent (compare chemical 6 with 1 in Table 3.9). This is likely to be due

to the increased electron withdrawing ability of the aryl ring which results in additional stabilisation

of the negative charge present in the structure of the intermediate (Figure 3.2). This electron

withdrawing effect is decreased by the addition of an extra alkyl carbon upon going from phenyl to

benzyl. This resulted in an increase in the calculated $\Delta E_{\text{INT-Thiolate}}$ value of only 1 kcal/mol (compare

chemical 7 with 6 in Table 3.9). The results of these calculations showed four fragments to be

required at the R_3 position for α,β-unsaturated ketones: methyl, phenyl, ethyl and *t*-butyl. The

calculated trends in the $\Delta E_{\text{INT-Thiolate}}$ SAR for the R_3 position for the α,β-unsaturated esters showed

the same results, with an analogous set of fragments being required (Table 3.10).

Table 3.9: Calculated $\Delta E_{\text{INT-Thiolate}}$ values for α,β-unsaturated ketones substituted at the R_3 position

(R groups as defined in Figure 3.6). The fragment substituent column indicates the substituent used

to predict the $\Delta E_{\text{INT-Thiolate}}$ of the respective R-group following the above analysis. The resulting

fragment $\Delta E_{\text{INT-Thiolate}}$ calculated from using that substitution is also included

ID	Substituent name – R_3	$\Delta E_{\text{INT-Thiolate}}$ (kcal/mol)	Fragment substituent	Fragment $\Delta E_{\text{INT-Thiolate}}$ (kcal/mol)
1	Methyl	-1	CH_3	-1
2	Ethyl	0	CH_3	-1
3	Propyl	0	CH_3	-1
4	*i*-Propyl	1	CH_2CH_3	0
5	*t*-Butyl	3	*t*-butyl	3
6	Phenyl	-3	C_6H_5	-3
7	Benzyl	-2	CH_3	-1
8	Phenethyl	-1	CH_3	-1

Table 3.10: Calculated $\Delta E_{INT\text{-}Thiolate}$ values for α,β-unsaturated esters substituted at the R_3 position (R groups as defined in Figure 3.6). The fragment substituent column indicates the substituent used to predict the $\Delta E_{INT\text{-}Thiolate}$ of the respective R-group following the above analysis. The resulting fragment $\Delta E_{INT\text{-}Thiolate}$ calculated from using that substitution is also included

ID	Substituent name – R_3	$\Delta E_{INT\text{-}Thiolate}$ (kcal/mol)	Fragment substituent	Fragment $\Delta E_{INT\text{-}Thiolate}$ (kcal/mol)
1	Methyl	5	H	5
2	Ethyl	5	OCH_3	5
3	Propyl	5	OCH_3	5
4	*i*-Propyl	6	OCH_2CH_3	5
5	*t*-Butyl	7	Ot-butyl	7
6	Phenyl	-0	OC_6H_5	-0
7	Benzyl	5	OCH_3	5
8	Phenethyl	5	OCH_3	5

3.5 Overall domain of the fragments

The analysis outlined above for α,β-unsaturated aldehydes, α,β-unsaturated ketones and esters resulted in the definition of 294 fragments which are summarised in Table 3.11. The table shows which substitutions were required at the respective R-groups for each chemical class. All aldehydes, ketones and esters required no greater than an ethyl substitution at all R-positions when considering linear alkyl chains. Additionally, steric factors were shown to have a significant effect on the calculated $\Delta E_{INT\text{-}Thiolate}$ for all R-groups for aldehydes ketones and esters, with *t*-butyl having a greater effect on calculated $\Delta E_{INT\text{-}Thiolate}$ than *i*-propyl. The electron withdrawing effect of the benzene ring was also calculated to have an effect on calculated $\Delta E_{INT\text{-}Thiolate}$ (when compared to linear alkyl chains) when linked by either one or two carbons at positions R_1, R_2 or (where relevant) R_3.

Table 3.11: Summary of the fragments defined for α,β-unsaturated aldehydes, α,β-unsaturated ketones and esters (R groups as defined in Figure 3.6)

Chemical class	R_1	R_2	R_3
α,β-unsaturated aldehydes	Alkyl: H, CH_3, CH_2CH_3, t-butyl Aryl: C_6H_5, $CH_2C_6H_5$, CH_2CH_3 [for $(CH_2)_nC_6H_5$, n>=2]	Alkyl: H, CH_3, i-propyl, t-butyl Aryl: C_6H_5, CH_3 [for $(CH_2)_nC_6H_5$, n>=1]	H
α,β-unsaturated ketones	Alkyl: H, CH_3, CH_2CH_3, i-propyl, t-butyl Aryl: C_6H_5, CH_3 [for $(CH_2)_nC_6H_5$, n>=1]	Alkyl: H, CH_3, CH_2CH_3, t-butyl Aryl: C_6H_5, $CH_2C_6H_5$, CH_2CH_3 [for $(CH_2)_nC_6H_5$, n>=2]	Alkyl: CH_3, CH_2CH_3, t-butyl Aryl: C_6H_5, CH_3 [for $(CH_2)_nC_6H_5$, n>=1]
α,β-unsaturated esters	Alkyl: H, CH_3, CH_2CH_3, t-butyl Aryl: C_6H_5, CH_3 [for $(CH_2)_nC_6H_5$, n>=1]	Alkyl: H, CH_3, CH_2CH_2, t-butyl Aryl: C_6H_5, $CH_2C_6H_5$, CH_2CH_2 [for $(CH_2)_nC_6H_5$, n>=2]	Alkyl: OCH_3, OCH_2CH_3, O-t-butyl Aryl: OC_6H_5, OCH_3 [for $(CH_2)_nC_6H_5$, n>=1]

3.6 Concluding remarks

This chapter has focused on the development of fragments for a fragment-based *in silico* profiler for Michael addition. The analysis showed that $\Delta E_{INT\text{-}Thiolate}$ was highly correlated with $\Delta E_{TS\text{-}Thiolate}$ showing that $\Delta E_{INT\text{-}Thiolate}$ could be used as the descriptor for reactivity for use in the fragment-based *in silico* profiler for Michael addition. This significantly reduced the number of calculations required per chemical (three calculations versus a minimum of nine). Additionally, calculations also showed the geometry of the reactant Michael acceptor to have an effect on the calculated $\Delta E_{INT\text{-}Thiolate}$ values, with the *cis*-intermediate being lower in energy. The $\Delta E_{INT\text{-}Thiolate}$ SAR analysis showed that a total number of 294 fragments were required (30 α,β-unsaturated aldehydes, 144 α,β-unsaturated ketones and 120 α,β-unsaturated esters) to cover the domain of linear α,β-unsaturated aldehydes,

ketones and esters. This covers all α,β-unsaturated aldehydes, ketones and esters with alkyl or aryl substitutions at the 3 R-positions (effectively covering a potentially vast number of chemicals). The definition of these fragments is the key step in the development of a fragment-based profiler. The ability of the developed fragments to predict thiol reactivity, skin sensitisation and *Tetrahymena pyriformis* toxicity will be discussed in Chapter 4.

Chapter 4. Application of the fragment-based *in* silico profiler for Michael Addition in predicting thiol reactivity, skin sensitisation and toxicity to *Tetrahymena pyriformis*

4.1 Introduction

Chapter 3 focused on the development of a fragment-based *in silico* profiler for Michael addition. The approach involved defining the fragments, and associated $\Delta E_{INT-Thiolate}$ values, required to cover the domain of linear α,β-unsaturated aldehydes, ketones and esters. This analysis resulted in the definition of 294 unique fragments. The aim of this chapter is to validate the ability of the *in silico* profiler, based on these fragments, and $\Delta E_{INT-Thiolate}$ values to predict glutathione reactivity, skin sensitisation potency and toxicity to *Tetrahymena pyriformis* for chemicals within the domain of the profiler.

4.2 Methods

4.2.1 Computational Methods

The fragment-based *in silico* profiler for Michael addition was developed based on a set of polarised aldehydes, ketones, and esters with varying alkyl and aryl substitutions (Figure 4.1) (103).

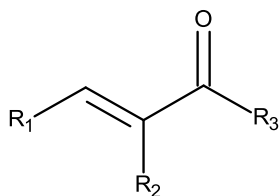Figure 4.1: Domain covered by fragment method for Michael acceptors. (R_1 = Hydrogen, alkyl, aryl) (R_2 = Hydrogen, alkyl, aryl) (R_3 = H) for polarised aldehydes, (R_3 = CH, C-alkyl, C-aryl) for polarised ketones, (R_3 = OCH, OC-alkyl, OC-aryl) for polarised esters.

The approach utilises a database of fragments with pre-calculated activation energy values ($\Delta E_{INT-Thiolate}$) calculated using Density Functional Theory (DFT) at the B3LYP/6-31G+(d) level of theory using water as a solvent employing Polarizable Cotinuum Model (PCM) as an implicit solvation model (calculations performed using Gaussian09 with water as solvent) (102). The approach is summarised in Figure 4.2. 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 (104). The Solvent Accessible Surface area (SAS) at the α-position was calculated for each chemical using the Chimera software (105).

4.2.2 Dataset for glutathione reactivity

The –Log RC_{50} values for various Michael acceptors were determined using a previously published spectrophotometric peptide depletion assay (106) (this data was gathered by Prof T. W. Schultz and the research group at the Department of Compariative Medicine research group at the University of Tennessee). Average -Log RC_{50} values were calculated for chemicals that had multiple experimental values. The – Log RC_{50} values for poorly soluble chemicals were determined by the addition of 50% MeOH. Profiling of a database of 212 Michael acceptors with glutathione reactivity data showed 54 chemicals were within the domain of the fragment-based *in silico* profiler for Michael addition (linear α,β-unsaturated aldehydes, ketones and esters). An additional 18 chemicals (nine nitros, three nitriles and six cyclic ketones) were also included as part of the model validation which required the development of additional fragments. This resulted in a subsequent dataset of 72 chemicals comprising 13 aldehydes, 17 ketones, 24 esters, nine nitro compounds, three nitrile containing compounds and six cyclic ketones (Table 4.1). Standard deviation values indicate an average experimental error of 0.13 log units for the reactivity measurements.

4.2.3 Data Set for *Tetrahymena pyriformis* and skin sensitization

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 *in silico* profiler (defined in Figure 4.1). This data was compiled in a study by Ruusmann et al based on the available data relating to the *Tetrahymena pyriformis* growth impairment assay and the work carried out by Prof T. W. Schultz (83, 107). 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 –Log EC_{50} values) (54). A similar analysis of skin sensitization data gathered from the mouse LLNA resulted in a dataset of 38 Michael acceptors within the applicability domain of the fragment-based *in silico* profiler for Michael addition (data gathered from studies by Gerberick et al, Kern et al and Natsch et al) (108-110). 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 (as shown in 1.2). 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 a final dataset of 27 skin sensitising chemicals (111).

4.2.4 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.

4.3 Prediction of glutathione reactivity using the fragment-based *in silico* profiler

4.3.1 Predictions for α,β-unsaturated aldehydes, ketones and esters

The ability of the fragment-based *in silico* profiler for Michael addition to predict glutathione reactivity was investigated for 54 chemicals (chemicals 1-54 in Table 4.1: 13 α,β-unsaturated aldehydes, 17 α,β-unsaturated ketones and 24 α,β-unsaturated esters).

Table 4.1: Michael acceptors with corresponding –Log RC_{50} values investigated in the current study. Where RC_{50} is the concentration of reactive chemical required to deplete GSH by 50 % in 120 minutes. Fragment names are based on their IUPAC name

ID	Chemical	-Log RC_{50} (mM)	Fragment	Predicted –Log RC_{50} (mM)			
				Model 4.1	Model 4.2	Model 4.3	Model 4.4
Aldehyde							
1	*trans*-Pent-2-enal	0.48	*trans*-Pent-2-enal	0.02	0.81	0.57	0.55
2	*trans*-Oct-2-enal	0.56	*trans*-Pent-2-enal	0.02	0.81	0.57	0.55
3	*trans*-Non-2-enal	0.39	*trans*-Pent-2-enal	0.02	0.81	0.57	0.55
4	*trans*-Hex-2-enal	0.37	*trans*-Pent-2-enal	0.02	0.81	0.57	0.55
5	*trans*-Prop-2-enal	1.14	*trans*-Prop-2-enal	0.89	1.74	1.38	1.34
6	*trans*-2-Methylbut-2-enal	-1.07	*trans*-2-Methylbut-2-enal	-0.43	-0.68	-0.87	-0.96
7	2-Methyl-pent-2-enal	-1.32	2-Methyl-pent-2-enal	-0.66	-0.82	-0.97	-1.05
8	4-Methyl-pent-2-enal	-0.06	4-Methyl-pent-2-enal	-0.48	0.30	0.12	0.12
9	*trans*-But-2-enal	0.67	*trans*-But-2-enal	0.31	0.98	0.70	0.66
10	E-Dec-2-enal	0.67	*trans*-Pent-2-enal	0.02	0.81	0.57	0.55
11	*trans*-Dec-2-enal	0.77	*trans*-Pent-2-enal	0.02	0.81	0.57	0.55
12	*trans*-Cinnamaldehyde	0.02	*trans*-Cinnamaldehyde	-0.43	0.24	0.06	0.05
13	α-Methyl-*trans*-cinnamaldehyde *	-1.33	α-Methyl-*trans*-cinnamaldehyde	-0.72	-1.04	-1.18	-1.27
Ketone							
14	Methyl vinyl ketone	1.23	Methyl vinyl ketone	0.20	1.19	0.93	0.92
15	Hex-1-en-3-one	1.23	Methyl vinyl ketone	0.20	1.19	0.93	0.92

ID	Chemical	-Log RC$_{50}$ (mM)	Fragment	Predicted –Log RC$_{50}$ (mM)			
				Model 4.1	Model 4.2	Model 4.3	Model 4.4
16	Pent-1-en-3-one	1.29	Methyl vinyl ketone	0.20	1.19	0.93	0.92
17	Pent-3-en-2-one	0.83	Pent-3-en-2-one	-0.60	0.30	0.14	0.15
18	Hept-3-en-2-one	0.17	Hex-3-en-2-one	-0.96	0.08	-0.03	0.00
19	Oct-3-en-2-one	0.24	Hex-3-en-2-one	-0.96	0.08	-0.03	0.00
20	Non-3-en-2-one	0.27	Hex-3-en-2-one	-0.96	0.08	-0.03	0.00
21	Dec-3-en-2-one	0.24	Hex-3-en-2-one	-0.96	0.08	-0.03	0.00
22	Hex-4-en-3-one	0.46	Hex-4-en-3-one	-0.72	0.23	0.08	0.10
23	Oct-1-en-3-one	1.78	Methyl vinyl ketone	0.20	1.19	0.93	0.92
24	3-Methyl-pent-3-en-2-one	-0.99	3-Methyl-pent-3-en-2-one	-1.33	-1.16	-1.23	-1.25
25	5-Methyl-hept-2-en-4-one	0.44	5-Methyl-hept-2-en-4-one	-1.00	0.05	-0.05	-0.01
26	*trans*-Non-3-en-2-one	0.22	Hex-3-en-2-one	-0.96	0.08	-0.03	0.00
27	4-Phenyl-but-3-en-2-one	-0.55	4-Phenyl-but-3-en-2-one	-1.00	-0.44	-0.55	-0.55
28	*trans*-Chalcone *	0.40	*trans*-Chalcone	-0.75	-0.29	-0.43	-
29	2-Hydroxychalcone	0.83	*trans*-Chalcone	-0.75	-0.29	-0.43	-
30	4-Hydroxychalcone	0.39	*trans*-Chalcone	-0.75	-0.29	-0.43	-
	Esters						
31	Isobutyl acrylate	0.32	Methyl acrylate	-0.66	0.61	0.46	0.50
32	n-Hexylacrylate	0.09	Methyl acrylate	-0.66	0.61	0.46	0.50
33	Butyl acrylate	0.11	Methyl acrylate	-0.66	0.61	0.46	0.50

ID	Chemical	-Log RC$_{50}$ (mM)	Fragment	Predicted –Log RC$_{50}$ (mM)			
				Model 4.1	Model 4.2	Model 4.3	Model 4.4
34	Methyl crotonate	-1.33	Methyl crotonate	-1.45	-0.22	-	-
35	Ethyl acrylate	0.29	Methyl acrylate	-0.66	0.61	0.46	0.50
36	Methyl acrylate	0.31	Methyl acrylate	-0.66	0.61	0.46	0.50
37	Methyl methacrylate	-1.84	Methyl methacrylate	-1.38	-1.31	-1.37	-1.40
38	*t*-butyl acrylate	-0.11	*t*-butyl acrylate	-0.97	0.42	0.31	0.38
39	Propyl acrylate	0.07	Methyl acrylate	-0.66	0.61	0.46	0.50
40	2-Hydroxy ethyl acrylate	0.57	Methyl acrylate	-0.66	0.61	0.46	0.50
41	2-Hydroxyethyl methacrylate	-1.52	Methyl methacrylate	-1.38	-1.31	-1.37	-1.40
42	2-Hydroxypropyl methacrylate	-1.33	Methyl methacrylate	-1.38	-1.31	-1.37	-1.40
43	Phenyl acrylate	1.64	Phenyl acrylate	0.10	1.06	0.81	-
44	Isoamyl acrylate	0.17	Methyl acrylate	-0.66	0.61	0.46	0.50
45	N-pentylacrylate	0.09	Methyl acrylate	-0.66	0.61	0.46	0.50
46	Ethyl crotonate	-1.25	Methyl crotonate	-1.45	-0.22	-	-
47	Methyl *trans*-pent-2-enoate	-0.70	Methyl *trans*-pent-2-enoate	-1.53	-0.26	-	-
48	Ethyl *trans*-hex-2-enoate	0.12	Methyl *trans*-pent-2-enoate	-1.53	-0.26	-0.30	-0.22
49	Methyl-hex-2-enoate	-0.39	Methyl *trans*-pent-2-enoate	-1.53	-0.26	-0.30	-0.22
50	Methyl-4-methyl-pent-2-enoate	-0.11	Methyl-4-methyl-pent-2-enoate	-2.08	-0.59	-0.56	-0.45
51	Ethyl tiglate	-1.15	Methyl tiglate	-2.08	-1.70	-1.66	-1.64
52	Ethyl methacrylate *	-1.53	Methyl methacrylate	-1.38	-1.31	-1.37	-1.40

ID	Chemical	-Log RC$_{50}$ (mM)	Fragment	Predicted –Log RC$_{50}$ (mM)			
				Model 4.1	Model 4.2	Model 4.3	Model 4.4
53	Butyl methacrylate *	-1.63	Methyl methacrylate	-1.38	-1.31	-1.37	-1.40
54	2-Ethylhexyl acrylate *	0.36	Methyl acrylate	-0.66	0.61	0.46	0.50
Nitro							
55	1-Nitro-1-cyclohexene	1.56	2-Nitrobut-2-ene	-	-	-	0.96
56	4-Methyl-β-nitrostyrene (mixture of *cis* and *trans*)	0.94	2-Nitroethylbezene	-	-	-	0.85
57	*trans*-β-Nitrostyrene	1.21	2-Nitroethylbezene	-	-	-	0.85
58	*trans*-4-Methyl-β-nitrostyrene	1.09	2-Nitroethylbezene	-	-	-	0.85
59	*trans*-4-Chloro-β-nitrostyrene	1.14	2-Nitroethylbezene	-	-	-	0.85
60	*trans*-4-Bromo-β-nitrostyrene	1.18	2-Nitroethylbezene	-	-	-	0.85
61	4-Fluoro-β-nitrostyrene	1.29	2-Nitroethylbezene	-	-	-	0.85
62	*trans*-4-Methoxy-β-nitrostyrene	1.36	2-Nitroethylbezene	-	-	-	0.85
63	*trans*-β-Methyl-β-nitrostyrene	1.19	2-Nitroprop-1-en-1-yl-benzene	-	-	-	0.82
Nitrile							
64	2-Methyleneglutaronitrile	-1.36	2-Methylprop-2-ene-nitrile	-	-	-	-1.28
65	Cyclohexene-1-carbonitrile (1-cyanocyclohexene)	-1.45	2-Methylbut-2-ene-nitrile	-	-	-	-1.57
66	1-Cyclopentene-1-carbonitrile	-1.31	2-Methylbut-2-ene-nitrile	-	-	-	-1.57
Cyclic ketones							

ID	Chemical	-Log RC$_{50}$ (mM)	Fragment	Predicted –Log RC$_{50}$ (mM)			
				Model 4.1	Model 4.2	Model 4.3	Model 4.4
67	2-Cyclohexen-1-one	0.50	Hex-3-en-2-one	-	-	-	0.00
68	2-Cyclopenten-1-one	0.18	Hex-3-en-2-one	-	-	-	0.00
69	2-Methyl-2-cyclopenten-1-one	-1.00	3-Methyl-hex-3-en-2-one	-	-	-	-1.35
70	4,4-Dimethyl-2-cyclohexen-1-one	-0.01	5-Methyl-hex-3-en-2-one	-	-	-	-0.10
71	1-Acetyl-1-cyclohexene	-0.31	3-Methyl-hex-3-en-2-one	-	-	-	-1.35
72	1-Acetyl-1-cyclopentene	-0.59	3-Methyl-hex-3-en-2-one	-	-	-	-1.35

-Log RC$_{50}$ were provided by T. W Schultz using a previously published spectrophotometric peptide depletion assay.[18] * indicates - chemicals that were unreactive in the standard 120 minute GSH assay with DMSO, for these chemicals values were obtained using 50% MeOH as solvent.

Initial modelling, using only the calculated $\Delta E_{\text{INT-Thiolate}}$, failed to produce a statistically significant model due to chemicals with an α-substituent being consistently over-predicted (Model 4.1 in Table 4.2 and Figure 4.2, chemicals with an α-substituent shown as ⊠). Inclusion of a solvent accessible surface area (SAS) descriptor for the α-position resulted in a significantly improved model (Model 4.2 in Table 4.2 and Figure 4.2). The mechanistic relevance of this descriptor likely stems from the nature of the intermediate in the Michael reaction which involves the formation of a resonance stabilised negative charge on the α-carbon atom. The solvation of this charge plays a key role in the stability of the transition state and thus overall reactivity. This solvation effect can be modelled by the inclusion of the steric SAS parameter, with the less solvent accessible α-substituted chemicals being less stabilised, due to solvent molecules being sterically hindered from solvating the charge by the presence of the substituent, compared to chemicals without an α-substituent. Given this, values for SAS at the α-position were introduced to all fragments in the fragment-based *in silico* profiler for Michael addition (in addition to $\Delta E_{\text{INT-Thiolate}}$).

Table 4.2: Summary statistics for Models 4.1-4.4 as shown in Figure 3. Model 1 has no SAS at the α-position value as it uses $\Delta E_{\text{INT-Thiolate}}$ as a single descriptor.

Model	N	a	B	c	R^2	R^2-adj	R^2-Pred	Average Error
				-Log RC_{50} = a + b. $\Delta E_{\text{INT-Thiolate}}$ + c.SAS α				
4.1	54	0.80	-0.15	X	0.52	0.51	0.48	0.60
4.2	54	-1.05	-0.09	-0.11	0.77	0.76	0.74	0.41
4.3	52	-1.30	-0.07	0.12	0.81	0.80	0.78	0.37
4.4	48	-1.48	-0.06	0.13	0.87	0.86	0.85	0.29

Model 4.2 successfully improves the prediction for the majority of the chemicals in the dataset. However, closer inspection of the data shows methyl and ethyl crotonate (shown as ⊠ in Model 4.2 Figure 4.2) to be significant outliers with errors of 1.07 and 0.99 log units respectively. Both methyl and ethyl crotonate have high predicted Log VP values (Table 4.3). As the experimental assay is carried out in scintillation vials, loss of the compound during the reaction may cause the results

to be unreliable (84). It is possible that this is not apparent for the unsubstituted volatile esters as they react sufficiently quickly. This is likely to be due to less steric hindrance; therefore, the reaction occurs before significant loss due to evaporation. Therefore, the effect on the slower reacting β-substituted esters is more easily detected. Given this, these chemicals were considered to be outside the domain of the model and were removed from the analysis.

Table 4.3. Predicted –Log RC_{50} values of β-substituted esters with corresponding error and Log Vapour Pressure (VP) values.

Compound	-Log RC_{50} (mM)	Predicted -Log RC_{50} (mM)	Error	Log VP
Methyl crotonate	-1.33	-0.32	1.01	1.26
Ethyl crotonate	-1.25	-0.32	0.93	0.91
Methyl *trans*-pent-2-enoate	-0.70	-0.37	0.33	0.98
Ethyl *trans*-hex-2-enoate	0.12	-0.37	-0.49	0.14
Methyl-hex-2-enoate	-0.39	-0.37	0.02	0.54
Methyl-4-methyl-pent-2-enoate	-0.11	-0.69	-0.58	0.80
Ethyl tiglate	-1.16	-1.72	-0.56	0.52

An additional set of chemicals were also poorly predicted by Model 4.3 (Figure 4.3, chemicals highlighted as ⊠), these being chemicals in which a phenyl ring conjugated to the carbonyl or ester moiety acts as the polarising group (Table 4.4). The reactivity of these chemicals was consistently under-predicted with error values ranging from 0.76 – 0.92 log units. Interestingly, the analogous chemical 4-phenyl-but-3-en-2-one in which the polarising group is a simple alkyl ketone is well predicted by Model 4.3 with an error of -0.04 log units. This suggests that the total electron-withdrawing effect of a conjugated phenyl group at position R_3 is not fully captured in the calculations (it is important to note that additional chemicals where R_3 is alkyl or hydrogen and the β-position is substituted with an aromatic ring are well predicted by the model). Removing these

four chemicals from Model 4.3 resulted in Model 4.4 (Table 4.2 and Figure 4.2) with an average error of 0.28 Log units.
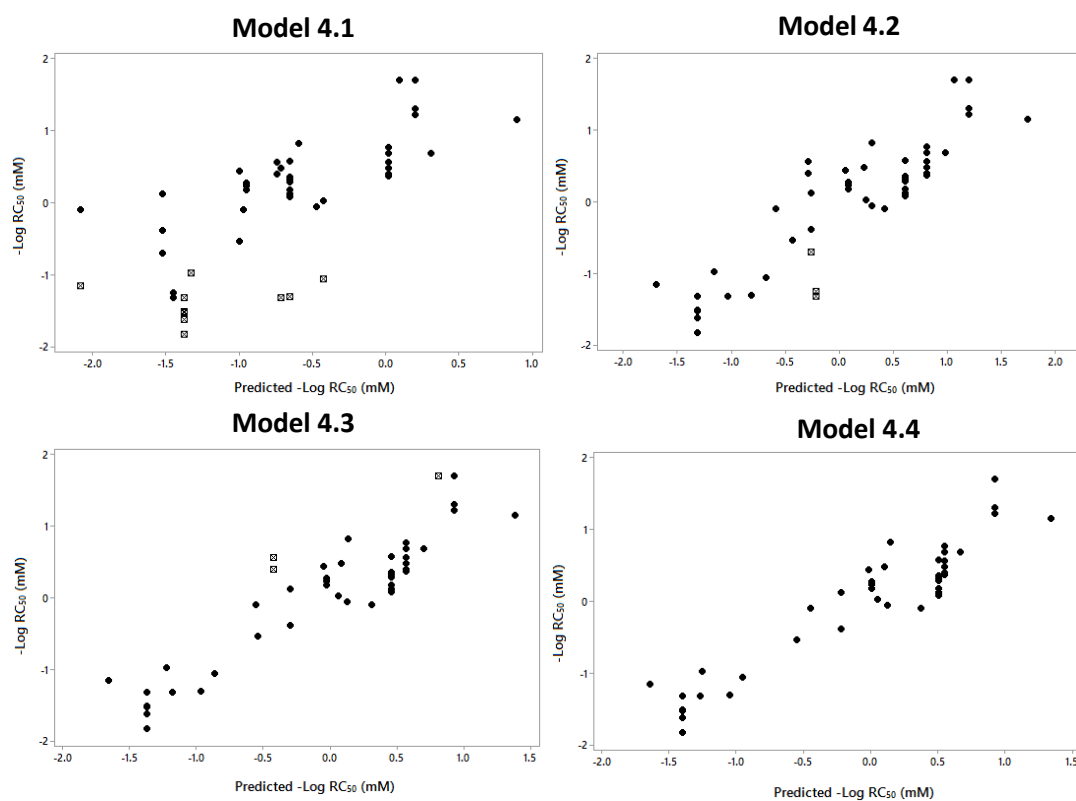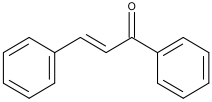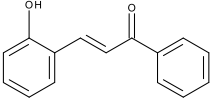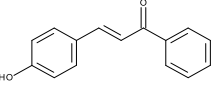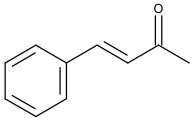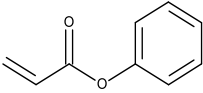

Figure 4.2: Predicted versus experimental values for –Log RC$_{50}$ for all models in the current study. Model 4.1: $\Delta E_{INT\text{-Thiolate}}$ only; Model 4.2: $\Delta E_{INT\text{-Thiolate}}$ with SAS at the α-position included. Model 4.3: $\Delta E_{INT\text{-Thiolate}}$ with SAS at the α-position excluding three volatile β-esters, Model 4.4: $\Delta E_{INT\text{-Thiolate}}$ with SAS at the α-position excluding three volatile β-esters and four compounds with a phenyl electron withdrawing group.

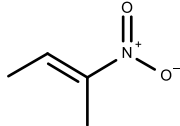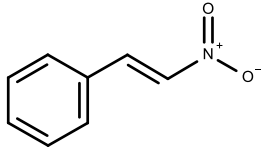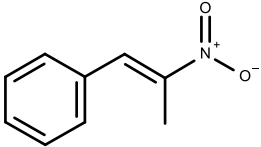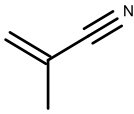Table 4.4: Predicted −log RC_{50} values for chemicals with a conjugated phenyl polarising group.
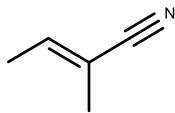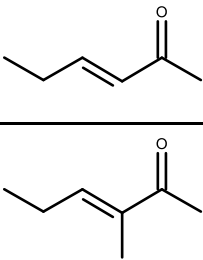
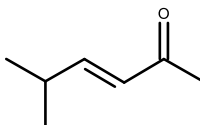| Chemical	Structure	-Log RC_{50}	Predicted -Log RC_{50}	Error
Chalcone		0.40	-0.37	0.77
2-Hydroxy-chalcone		0.56	-0.37	0.92
4-Hydroxy-chalcone		0.39	-0.37	0.76
4-Phenyl-but-3-en-2-one		-0.55	-0.51	-0.04
Phenyl-acrylate		1.64	0.94	0.76

4.3.2 Prediction of additional chemical classes

To demonstrate how the fragment-based *in silico* profiler may be expanded to cover additional chemical classes, a second dataset of 18 chemicals (compounds 55 – 72 in Table 4.1) with reactivity data was investigated. The chemicals within this dataset required $\Delta E_{\text{INT-Thiolate}}$ values for an additional five fragments to be calculated, along with three fragments previously defined (Table 4.5). These $\Delta E_{\text{INT-Thiolate}}$ values were used in conjunction with Model 4.4 to predict −Log RC_{50} values for these 18 chemicals with an average error of 0.62 log units (Figure 4.3a shows the predicted values for these 18 chemicals, shown as ⊠, in comparison to the chemicals used in the derivation of Model 4.4, shown as ●). The results suggest that, for the polarised nitro, substituents at the α-position have significantly less effect on reactivity than for chemicals polarised by an aldehyde, ketone or ester moiety. This can be rationalised in terms of the resonance stabilisation of the

intermediate for the polarised nitros for which two possible resonance forms exist (Figure 4.4). It is possible that the nitro group is sufficiently polarising that the negative charge is localised mainly on the oxygen rather than the α-carbon, resulting in solvation at this position becoming less important. Excluding the SAS parameter for the polarised nitros (in effect assuming that these chemicals have an SAS value equivalent to hydrogen) results in a significant improvement in the predicted –Log RC_{50} values for these chemicals (Figure 4.3b), with an average error of 0.44 log units. Interestingly, among the polarised nitros three of the compounds contain halogenated phenyl groups at the β-position, these are predicted well (see chemicals 59-61 in Table 4.1). This suggests that using phenyl alone was sufficient and that the applicability domain of this method may extend further to alkyl and phenyl groups with varying substitutions.

Table 4.5: Fragments required to predict the reactivity of polarised nitros, polarised nitriles and cyclic ketones.

Chemical	Fragment used	New or existing fragment
Polarised nitros		
1-Nitro-1-cyclohexene		New
4-Methyl-β-nitrostyrene *Trans*-β-nitrostyrene *Trans*-4-methyl-β-nitrostyrene *Trans*-4-chloro-β-nitrostyrene *Trans*-4-bromo-β-nitrostyrene *Trans*-4-fluoro-β-nitrostyrene *Trans*-4-methoxy-β-nitrostyrene		New
Trans-β-methyl-β-nitrostyrene		New
Polarised nitriles		
2-Methyleneglutaronitrile		New
Cyclohexene-1-carbonitrile 1-Cyclopentene-1-carbonitrile		New
Polarised cyclic ketones		
Cyclohex-2-en-1-one Cyclopent-2-en-1-one		Existing
2-Methyl-cyclopent-2-en-1-one 1-Acetyl-cyclohex-1-ene 1-Acetyl-cyclopent-1-ene		Existing

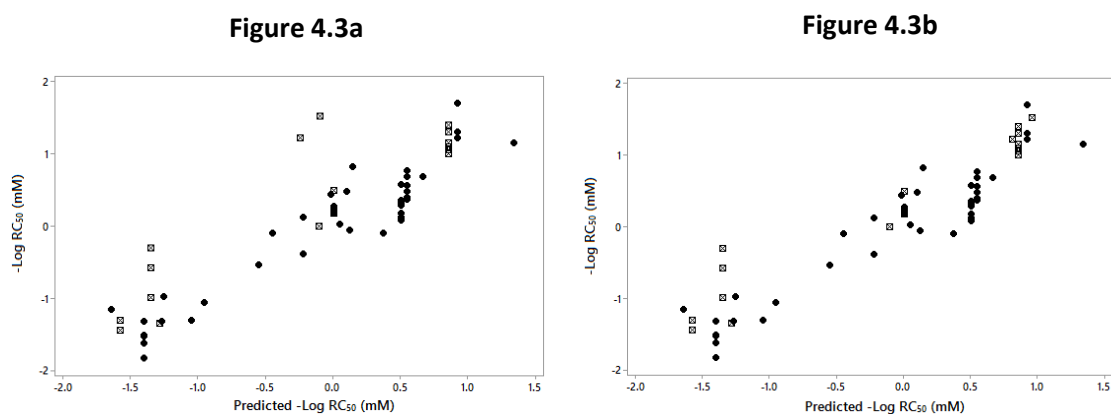| Chemical	Fragment used	New or existing fragment
4,4-Dimethyl-cyclohex-2-en-1-one		Existing

Figure 4.3a   **Figure 4.3b**

Figure 4.3: Predicted versus experimental –Log RC_{50} values for polarised nitros, polarised nitriles and polarised cyclic ketones (⊠) using Model 4.4 in comparison to the polarised aldehydes, ketones and esters in the initial dataset (●). Figure 4.3a shows polarised nitros with the inclusion of the SAS descriptor. Figure 4.3b shows polarised nitros with the SAS descriptor value set to hydrogen for all chemicals.
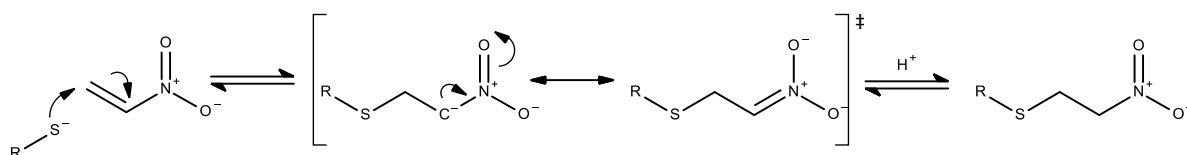
Figure 4.4: Michael addition mechanism for the reaction between thiol nucleophile and nitroethene (R = alkyl, GSH).

In summary, the fragment-based *in silico* profiler was successful in predicting the thiol reactivity of Michael acceptors. This required the use of an additional descriptor to model the solvent accessible surface area at the alpha position (SAS α). The analysis showed that the fragment-based *in silico* profiler could be extended to additional Michael acceptors by calculating the necessary fragments.

The process through which the thiol reactivity of Michael acceptors was predicted using the fragment-based *in silico* profiler is summarized in Figure 4.5.
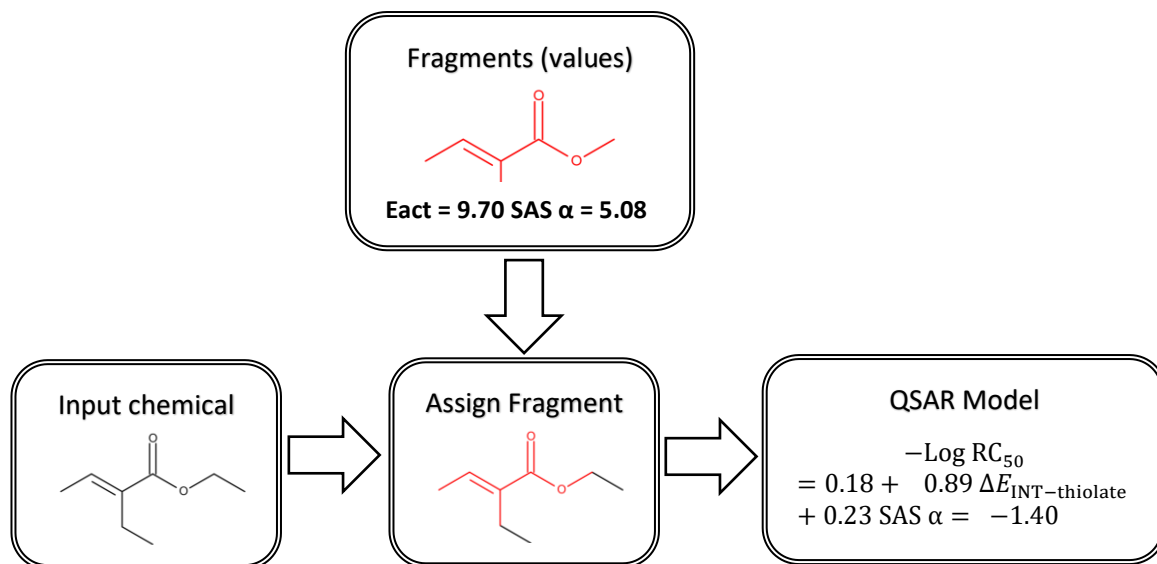
Figure 4.5: A summary of the fragment-based *in silico* profiler for predicting thiol reactivity (-Log RC$_{50}$).

4.4 Prediction of toxicity using the predicted reactivity values

Having shown that the $\Delta E_{\text{INT-Thiolate}}$ values from the *in silico* profiler for Michael addition were successful in predicting glutathione reactivity (from QSAR Model 4.4 in Table 4.2). The next step was to assess the ability of the predicted reactivity values (-Log RC$_{50}$) in predicting toxicity (toxicity to *Tetrahymena pyriformis* and skin sensitisation.

4.4.1 Prediction of toxicity to *Tetrahymena pyriformis*

Analysis of the *Tetrahymena pyriformis* data within the applicability domain of the fragment-based *in silico* profiler resulted in a dataset of 62 chemicals (14 aldehydes, 12 ketones and 36 esters) with corresponding EC$_{50}$ values (Table 4.6).

Table 4.6 The 62 chemicals used in the assessment of the fragment method for predicting *Tetrahymena pyriformis* toxicity. Chemical names, SMILES, experimental –Log EC_{50} mmol/l, –Log RC_{50}(pred), D_{kk} and predicted -Log EC_{50} mmol/l for the respective models are shown. N.B -Log EC_{50} values were calculated with Model 4.6a for fast reacting chemicals (1-43) and Model 4.6b for slower reacting chemicals (44-62).

ID	Chemical	SMILES	-Log EC_{50} (mmol/l)	– Log RC_{50}(pred)	Log k_{ow}	D_{kk}	Predicted -Log EC_{50}	
							Model 4.5	Model 4.6a/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

ID	Chemical	SMILES	-Log EC$_{50}$ (mmol/l)	– Log RC$_{50}$(pred)	Log k_{ow}	D$_{kk}$	Predicted -Log EC$_{50}$ Model 4.5	Predicted -Log EC$_{50}$ Model 4.6a/b
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	CCCCC(=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

ID	Chemical	SMILES	-Log EC$_{50}$ (mmol/l)	– Log RC$_{50}$(pred)	Log k_{ow}	D$_{kk}$	Predicted -Log EC$_{50}$	
							Model 4.5	Model 4.6a/b
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

97

ID	Chemical	SMILES	-Log EC$_{50}$ (mmol/l)	– Log RC$_{50}$(pred)	Log k_{ow}	D$_{kk}$	Predicted -Log EC$_{50}$	
							Model 4.5	Model 4.6a/b
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

Initial modelling using the −Log RC_{50}(pred) values alone showed a clear trend (R^2 = 0.45) between reactivity and toxicity to *Tetrahymena pyriformis* (Model 4.5 in Figure 4.6). 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) (70). However, in comparison with the current study (using −Log RC_{50}(pred) as a measure of reactivity) the previously published 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 their slower reaction with proteins (70).

$$-\text{Log EC}_{50} = 0.63 + 0.61(-Log\ RC_{50}(pred)) \qquad \text{(Model 4.5)}$$

$$N = 62, R^2 = 0.45, R^2_{adj} = 0.44, s = 0.46$$

Consistent with this hypothesis, a related study showed that splitting the data into fast reacting and slow reacting classes resulted in significantly improved modelling results (52). 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 4.1 denoted D_{kk} (the difference between Log k_{OW} and Log k_{GSH}), where chemicals with a D_{kk} < 3 are fast reacting and those with D_{kk} > 3 are slow reacting. Although the authors used kinetic glutathione data (-Log K_{GSH}) these values have been shown to be highly correlated with glutathione depletion data (-Log RC_{50}) (56). Applying these criteria to the current dataset, using −Log RC_{50}(pred) as a measure of reactivity resulted in Models 4.6a and 4.6b (fast and slow reacting chemicals respectively). Forty-three chemicals were assigned to the fast reacting class (chemicals 1 – 43 in Table 4.6), whilst 23 chemicals were assigned to the slow reacting class (chemicals 44-62 in Table 4.6). 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$_{50}$(pred) (Model 4.6a), whilst the chemicals in the slow reacting class required both –Log RC$_{50}$(pred) and log K_{ow} (Model 4.6b). Figure 4.7 shows the correlation plots for Models 4.6a and 4.6b.

$$D_{kk} = Log\ (k_{ow}/- RC_{50}(pred)) = Log\ k_{ow} - (-Log\ RC_{50}(pred)) \tag{4.1}$$

$$-Log\ EC_{50} = 0.41 + 0.94 - (Log\ RC_{50}(pred)) \qquad \text{(Model 4.6a)}$$

$$N = 43, R^2 = 0.78, R^2_{adj} = 0.77, s = 0.30$$

$$-Log\ EC_{50} = -1.82 + 0.35(-Log\ RC_{50}(pred)) + Log\ k_{OW} \qquad \text{(Model 4.6b)}$$

$$N = 19, R^2 = 0.85, R^2_{adj} = 0.83, s = 0.31$$

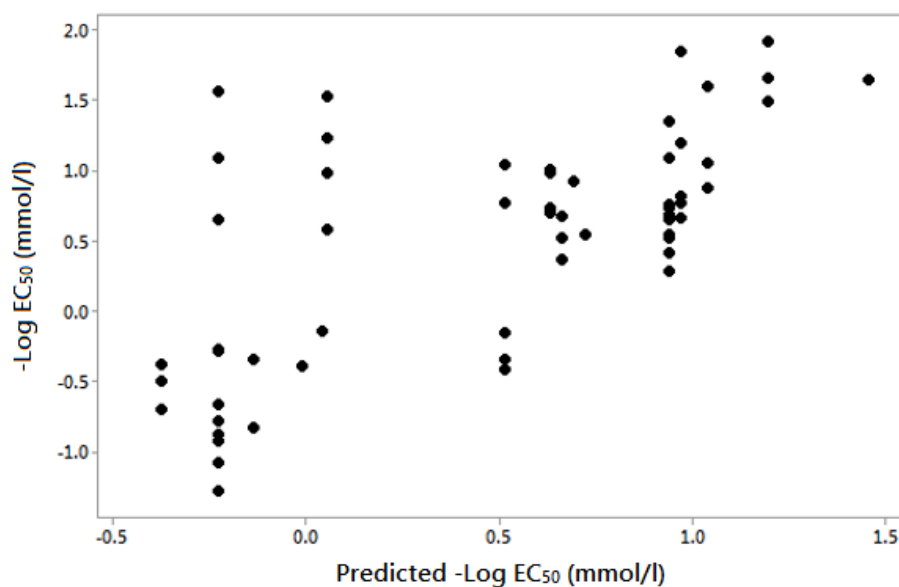Figure 4.6. The predicted -Log EC$_{50}$ values against experimental –Log EC$_{50}$ values for all 62 Michael acceptors using –Log RC$_{50}$(pred) alone (Model 4.5).
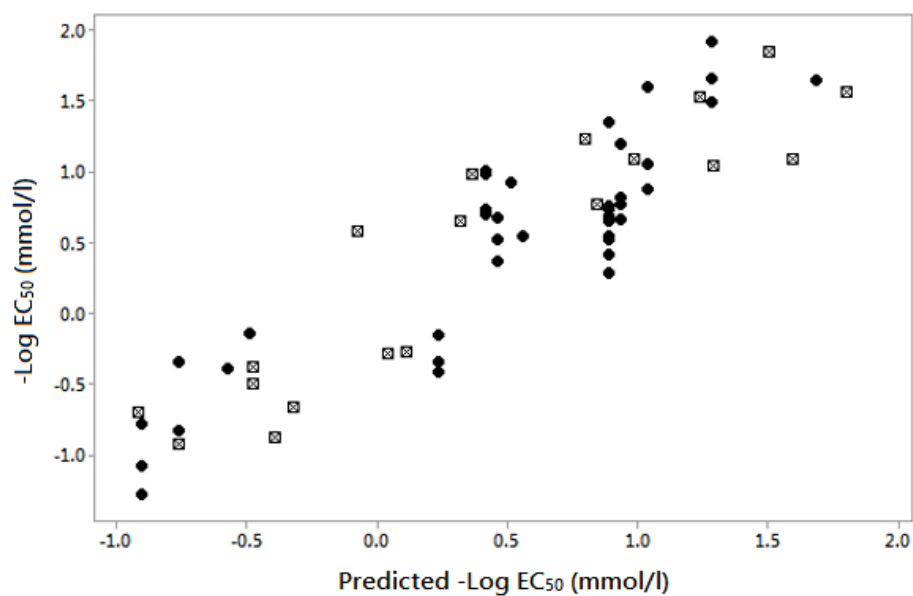
Figure 4.7: The predicted -Log EC_{50} (mmol/l) against experimental -Log EC_{50} (mmol/l) of all 43 fast reacting chemicals (●) (Model 4.6a, chemicals 1-43 in Table 4.6) and 19 slower reacting chemicals (⊠) (Model 4.6b, chemicals 44-62 in Table 1) requiring hydrophobicity to be taken into account

4.4.2 Prediction of skin sensitisation potency as measured 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 mouse LLNA using both experimental and computational measures of reactivity (43, 49, 94, 95). In keeping with these studies, the fragment-based *in silico* profiler was used to predict pEC3 values for the 26 Michael acceptors within the previously defined applicability domain. These chemicals are shown in Table 4.7.

Table 4.7. The 26 chemicals used in the assessment of the ability of the predicted reactivity values to predict skin sensitization potency (pEC3). Chemical names, SMILES, experimental pEC3 with error values, $-Log RC_{50}$(pred), Log VP and predicted pEC3 values for all models are shown.

ID	Chemical	SMILES	pEC3	$-Log RC_{50}$(pred)	Log VP	Predicted pEC3	
						Model 4.7	Model 4.8
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	CC1=CC[C@H](CC1=O)C(=C)C	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	α-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-Hex-2-enal	CCC\C=C\C=O	1.25	0.41	0.71	1.38(0.13)	-
11	α-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	α-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)

ID	Chemical	SMILES	pEC3	-Log RC$_{50}$(pred)	Log VP	Predicted pEC3	
						Model 4.7	Model 4.8
18	5-Methyl-2-phenyl-hex-2-enal	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*-Dec-2-enal	CCCCCCC/C=C/C=O	1.79	0.41	-1.08	1.38(-0.41)*	1.95(0.16)
21	Galbanone	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)

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

‡ Chemicals with an additional –Log RC$_{50}$(pred) value use this value for Model 4.8 for reasons discussed in the text

An initial analysis of the correlation between pEC3 and –Log RC$_{50}$(pred) resulted in a model with in extremely poor statistics (Model 4.7 in Figure 4.8). Despite this, 13 of the chemicals were predicted within a two-fold 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 two-fold error of the LLNA (79). Any chemicals outside of the two-fold error of the experimental assay were considered to be outliers (labelled in Table 4.7) and were analysed in an attempt to rationalise the error in their predicted values.

$$Predicted\ pEC_3 = 1.35 - 0.05 - (Log\ RC_{50}(pred)) \qquad \text{(Model 4.7)}$$
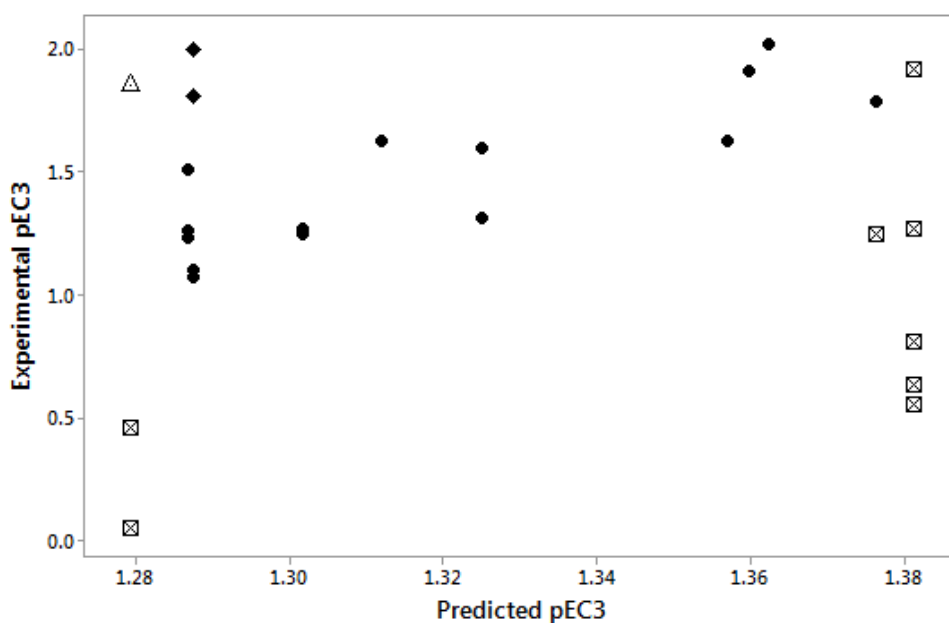
$$N = 26, R^2 = 0.00, R^2_{adj} = 0.00, s = 0.30$$

Figure 4.8. Predicted pEC3 versus experimental pEC3 for all 26 Michael acceptors shown in Table 2. ⊠ = volatile chemicals; ♦ = galbanone and spirogalbanone; △ = 5,5-dimethyl-3-methylene-dihydro-2(3H)-furone.

Most of the compounds with the largest errors are chemicals that are volatile, with the majority of these being acrylates and methacrylates (chemicals 1-5 in Table 4.7). 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 (43). In addition, research has also suggested that the acrylate and methacrylates are susceptible to polymerisation driven by free radical chemistry in the skin (112, 113). Interestingly, the toxicity of a large number of similar chemicals towards *Tetrahymena pyriformis* were well predicted (chemicals 23-62 in Table 4.6). 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 in Table 4.7), and two additional acrylates (chemicals 13 and 24 in Table 4.7). Three of these chemicals were removed despite being relatively well-predicted (chemicals 10, 13 and 24 in Table 4.7) 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 predicted reactivity values (chemicals 21 and 25 in Table 4.7). The skin sensitisation potency of these two chemicals was predicted using 3-methyl-pent-3-en-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) (103). 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 4.9) (114). 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 4.9). 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 cyclic ring twists out of the plane of the carbonyl group upon

nucleophilic attack at the alternative site (data not shown). This will result in decrease in conjugation of the β-carbon within the double ring and the carbonyl group making a reaction less likely to happen. 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 *in silico* profiler in that it enables the investigation of alternative sites of chemical reactivity through the use of alternate fragments.
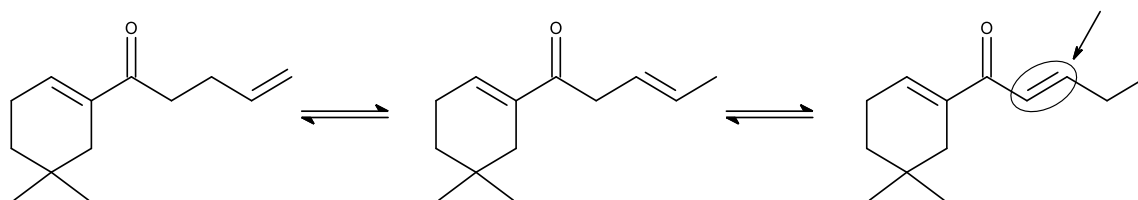

Figure 4.9. Isomerisation of galbanone to produce an extended conjugated chemical highlighting a possible additional site of reactivity

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 *in silico* profiler 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 (103). In keeping with this analysis, the analogous chemicals in the skin sensitisation data were well predicted (chemicals 6, 7 and 14 in Table 4.7). Inspection of the data used to develop the fragment-based *in silico* profiler 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 *in silico* profiler 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 nine chemicals resulting in a final model based on 17 chemicals

with an R^2 = 0.77 (Figure 4.10, Model 4.8). 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 (43, 94). However, the use of –Log RC_{50}(pred) 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).

$$\text{Predicted pEC}_3 = 1.77 + 0.43 - (Log\ RC_{50}(pred)) \qquad \text{(Model 4.8)}$$

$$N = 17, R^2 = 0.76, R^2_{adj} = 0.76, s = 0.12$$
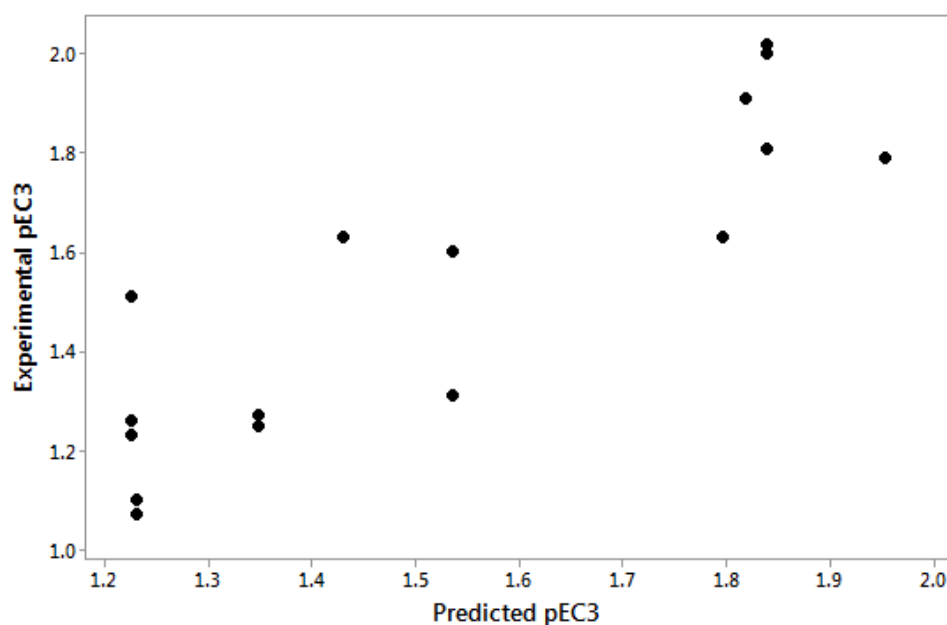

Figure 4.10. The predicted pEC3 against experimental pEC3 for Model 4.8 (predicted values shown in Table 4.7)

4.5 Concluding remarks

The first aim of this chapter was to predict thiol reactivity for Michael acceptors using the $\Delta E_{INT\text{-}Thiolate}$ values from the *in silico* profiler for Michael addition. The second aim was to validate the ability of the predicted –Log RC_{50} values (generated from the use of the fragment based *in silico* profiler) to predict toxicity to *Tetrahymena* pyriformis and skin sensitisation potency (as determined in the LLNA). The results showed that a QSAR combined from pre-calculated $\Delta E_{INT\text{-}Thiolate}$

values, coupled with a descriptor for the solvent accessible surface area at the α-carbon, was able to accurately predict chemical reactivity as measured in a glutathione depletion assay. Two sets of chemicals were poorly predicted by the approach, these being: volatile esters with an extended substituent at the β-carbon and chemicals containing a conjugated benzene ring as part of the polarising group. The study also demonstrated the ease with which the approach can be extended to other chemical classes by the calculation of additional fragments and their associated $\Delta E_{INT\text{-}Thiolate}$ and SAS values. Additionally, the results of this study showed the predicted reactivity values (-Log RC_{50}(pred)) were able to predict both endpoints within well-defined, end-point specific applicability domains. The results also showed the importance of considering slow versus fast reacting Michael acceptors when modelling toxicity to *Tetrahymena pyriformis* and that polymerisation and volatility are important considerations in successfully predicting skin sensitisation potency. These results were in keeping with previously published studies that have 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 *in silico* profiler 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 to undertake complex quantum mechanics calculations. The next chapter adopts the method of the *in silico* profiler to predict the reactivity and toxicity of chemicals that form covalent bonds through an alternative mechanism (bimolecular nucleophilic substitution - S_N2).

Chapter 5. The development of a fragment-based *in silico* profiler for the S$_N$2 mechanism and its application in predicting thiol reactivity and toxicity to *Tetrahymena pyriformis*

5.1 Introduction

Chapter 3 outlined the development of an *in silico* profiler for Michael addition applicable to α,β-unsaturated aldehydes, ketones and esters. The profiler involved defining a set of 294 fragments developed from a SAR analysis of the effect of varying a range of substituents on calculated $\Delta E_{INT\text{-}Thiolate}$ values. Chapter 4 demonstrated the ability of this fragment-based profiler to predict glutathione reactivity, skin sensitisation potency (as defined in the mouse LLNA) and toxicity to *Tetrahymena pyriformis* for chemicals within its applicability domain. This chapter outlines the extension of this approach to chemicals capable of reacting with biological nucleophiles via a bimolecular nucleophilic substitution (S$_N$2) mechanism. These reactions most commonly occur at sp^3 carbon atoms bound to an electronegative leaving group such as a halogen (it is worth noting that toxicologically relevant S$_N$2 reactions have also been shown to occur at nitrogen, sulphur and chlorine) (57). Unlike Michael addition, the S$_N$2 reaction has no stable intermediate as the attack by the nucleophile and loss of the leaving group happen simultaneously (Figure 5.1). As such, the energy profile for the S$_N$2 reaction consists of a single transition state linking the reactants and products on the potential energy surface (Figure 5.2). The rate limiting step for this reaction is the simultaneous formation of the carbon-nucleophile (or an equivalent nucleophilic atom e.g. nitrogen or sulphur) bond and breaking of the carbon-leaving group bond. This type of reaction has the potential to be reversible, depending on the similarity in pKa values between the nucleophile and the leaving group (similar pKa values indicating reversibility). However, this is unlikely to be the case in the reaction between chlorinated/brominated alkanes and a thiolate nucleophile (the focus of this chapter) due to the large difference in pKa values (-2.2, -4.7 and 10.5 for chloride, bromide and

thiolate respectively). In contrast to Michael addition, there have been no computational attempts to predict reactivity and toxicity for chemicals acting via an S_N2 mechanism. As such, the aim of this chapter was to develop a fragment-based *in silico* profiler for the S_N2 mechanism and to validate the ability of the *in silico* profiler and $\Delta E_{TS\text{-}Thiolate}$ values to predict glutathione reactivity and toxicity to *Tetrahymena pyriformis*.
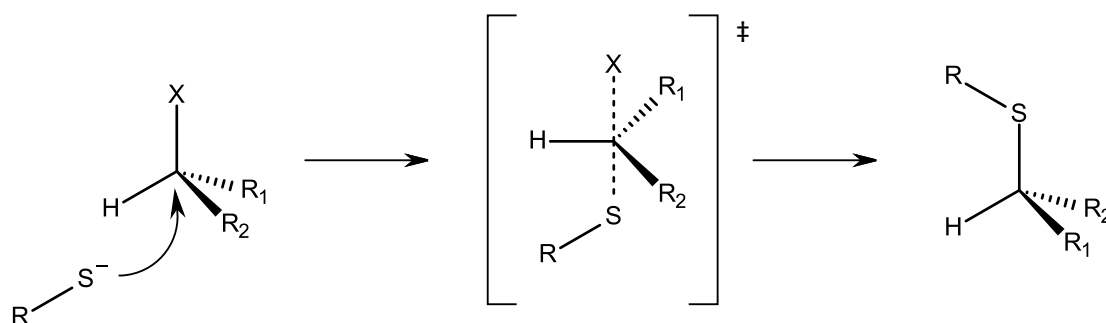

Figure 5.1. The formation of a covalent bond between an electrophilic chemical and cysteine via an S_N2 mechanism, (X = halogen, R = alkyl, R_1 = R_2 = hydrogen or carbon)
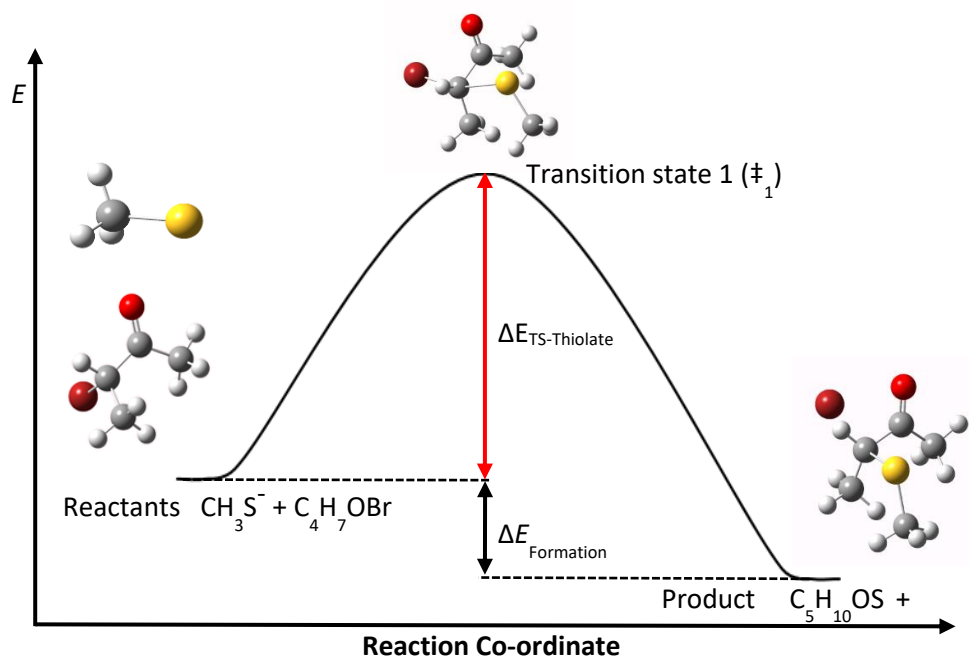

Figure 5.2. Energy diagram for the S_N2 reaction between 2-bromobutan-2-one and a thiolate nucleophile resulting in the formation of a transition state structure (energy of activation shown as $\Delta E_{TS\text{-}Thiolate}$). The energy difference between the reactants and the products is shown as $\Delta E_{Formation}$

5.2 Methods

5.2.1 Data set

Thirty-one chemicals were identified as acting via an S_N2 mechanism from a study by D. W. Roberts and T. W. Schultz (chemicals shown in Table 5.7) (42). Three chemicals were excluded from this dataset, these being: 3-bromo-acetyl-coumarin, ethyl iodoacetate and 2-iodoacetamide. 3-Bromo-acetyl-coumarin was excluded from the analysis due to it having multiple sites of electrophilic reactivity. The other two chemicals contained iodine as the leaving group. It was not possible to perform calculations on these chemicals due to the chosen basis set only being applicable to elements in the first three rows of the periodic table (see Chapter 2 for information on basis sets). This resulted in a dataset of 28 activated S_N2 chemicals. All chemicals in the dataset had associated glutathione reactivity data (-Log RC_{50}) and *Tetrahymena pyriformis* toxicity data (-Log IGC_{50}) (information on respective assays found in Chapter 1) (42).

5.2.2 Computational methods

All calculations were carried out using the Gaussian 09 suite of software using density functional theory at the B3LYP/6-31G+(d) level of theory using water as a solvent employing Polarizable Cotinuum Model (PCM) as an implicit solvation model (102). The $\Delta E_{TS-Thiolate}$ values were obtained using scan calculations to determine the highest point of energy on the potential energy surface for the reaction between the electrophile and thiolate nucleophile. All scan calculations were performed using an initial bond length of 2.9 Å between the halogenated carbon atom and the sulphur of the nucleophile. In keeping with Chapter 3, all calculations used methyl thiolate as a model nucleophile. A series of seven calculations were then carried out in which the bond length between the halogenated carbon and the sulphur of the thiolate nucleophile was decreased by 0.1 Å with each calculation. This mapped the reaction coordinate enabling the highest energy point corresponding to the transition state structure to be identified. All transition state structures were optimised and subjected to frequency analysis in order to identify a single negative eigenvalue

connecting the transition state to the reactants and products on the potential energy surface. All calculations were carried out using the "opt=loose" keyword. Calculated $\Delta E_{TS\text{-}Thiolate}$ values in the fragment analysis section (Sections 5.4.2 – 5.4.3) are quoted to the nearest kcal/mol. The approach is analogous to the method adopted in Chapter 3 in which a database of fragments with pre-calculated activation energy values ($\Delta E_{INT\text{-}Thiolate}$) is used as the basis for the fragment-based *in silico* profiler for Michael addition (see Figure 3.5, Chapter 3).

5.2.3 Statistical analysis

Linear regression analysis was used to develop quantitative structure-activity relationship models to obtain correlations between $\Delta E_{TS\text{-}Thiolate}$ values, –Log RC_{50}(pred) and toxicity to *Tetrahymena pyriformis* (-Log IGC_{50}) values using the Minitab (version 17) statistical software.

5.3 Investigation of S_N2 transition state geometries

Initial modelling efforts were focussed on the characterisation of the transition state structures for five brominated S_N2 chemicals with differing electron-withdrawing groups (Table 5.3). All five chemicals had similar transition state structures with S-C bond lengths of 2.7 Å and C-Br bond lengths between 2.2-2.3 Å (Figure 5.3 and Table 5.1). The five chemicals had calculated $\Delta E_{TS\text{-}Thiolate}$ values in the following order (least reactive to most reactive): 2-bromo acetamide > methyl-2-bromoacetate > 2-bromoacetic acid > 1-bromopropan-2-one > 2-bromo-1-phenylethan-1-one.
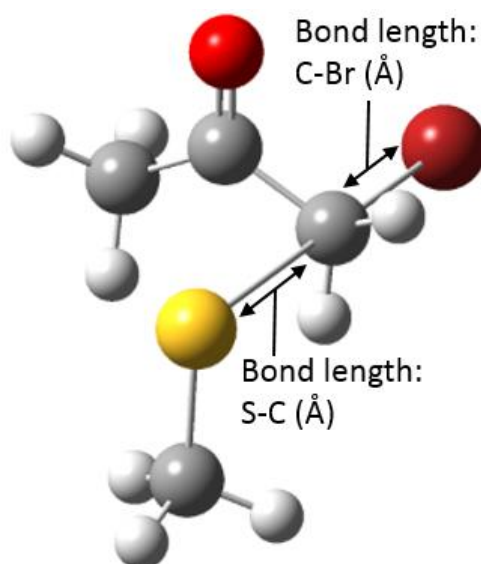
Figure 5.3. The calculated transition state geometry of 1-bromopropan-2-one upon reaction with methyl thiolate

Table 5.1. The calculated bond lengths (Å) and $\Delta E_{TS\text{-}Thiolate}$ (kcal/mol) values for a representative series of brominated S_N2 chemicals activated by an electron-withdrawing group

ID	Chemical	S-C (Å)	C-Br (Å)	$\Delta E_{TS\text{-}Thiolate}$ (kcal/mol)	$\Delta E_{TS\text{-}Thiolate}$ + sodium (kcal/mol)
1	1-Bromopropan-2-one	2.7	2.2	-0.7	5.8
2	2-Bromo-1-phenylethan-1-one	2.7	2.2	-1.7	4.9
3	Methyl-2-bromoacetate	2.7	2.3	0.4	6.9
4	2-Bromoacetic acid	2.7	2.3	-0.4	6.1
5	2-Bromo acetamide	2.7	2.3	2.2	8.7

It is worth noting some of the $\Delta E_{TS\text{-}Thiolate}$ values in Table 5.1 are negative (i.e. the transition state is lower in energy than the reactants). These values are unusual at first glance, as the $\Delta E_{TS\text{-}Thiolate}$ values were obtained from the highest point of energy along the reaction co-ordinate and have negative frequencies associated with them (indicating them to be true transition states). Given that the pKa of a thiolate ion is reported to be roughly 12.0 it is unlikely that this species will exist in a significant amounts at neutral pH. In reality, it is likely that the thiolate is stabilised by a positively charge

counter ion such as sodium. With this in mind, a sodium ion was included in the calculation as part of the reactants and transition state structure. This resulted in a large increase in the $\Delta E_{\text{TS-Thiolate}}$ (kcal/mol) for all chemicals (see Table 5.1). This was something not immediately obvious in the analysis with the Michael acceptors and the analysis was carried out using the resonance-stabilised intermediate of the Michael addition reaction as opposed to the use of a transition state structure. However, in order to have a rational comparison with the Michael addition analysis (see Chapter 4), $\Delta E_{\text{TS-Thiolate}}$ without the addition of sodium was used for the S_N2 analysis. Although this has an effect on the $\Delta E_{\text{TS-Thiolate}}$ values, the values will be effected equally and therefore the predictive outcome of the values will not change.
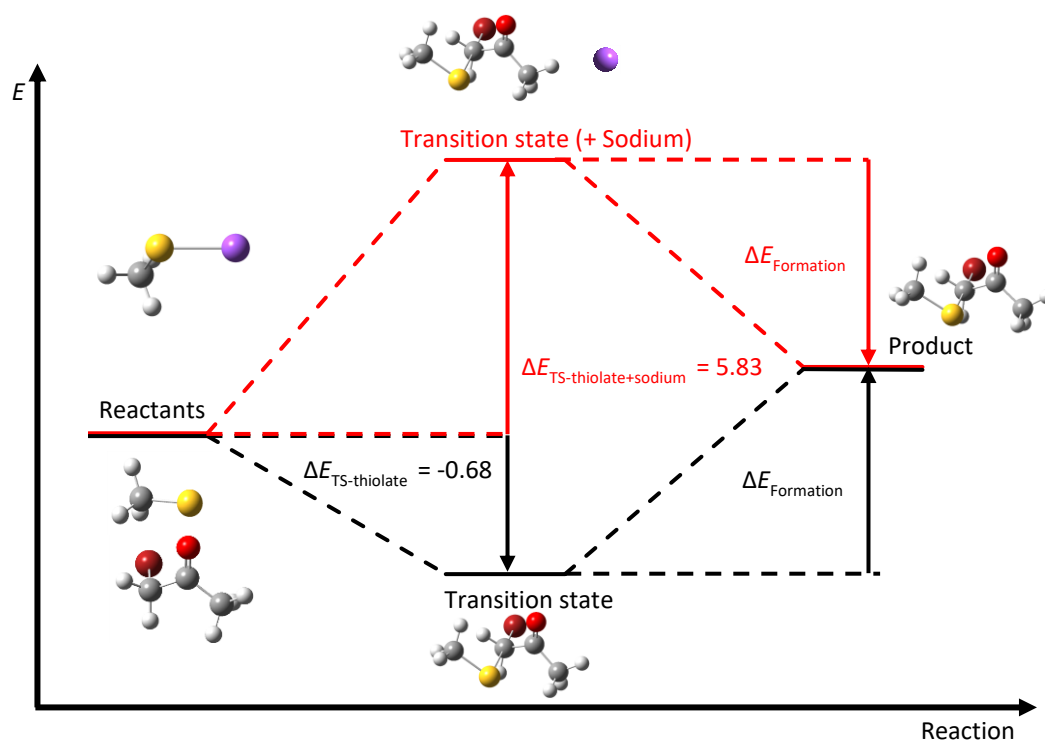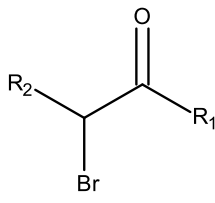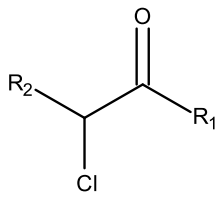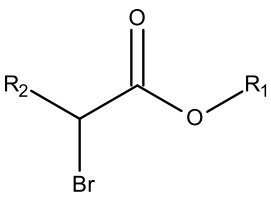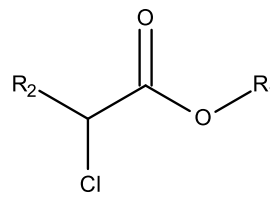

(Figure 5.4). Comparison of $\Delta E_{\text{TS-Thiolate}}$ values for 1-bromoproan-2-one with and without the addition of sodium (see Table 5.1 for values)
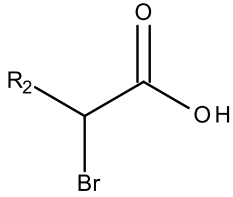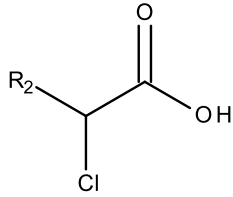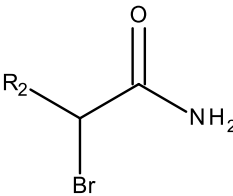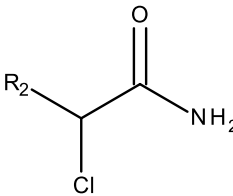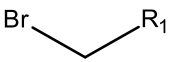
5.4 Development of fragments for S_N2 chemicals

The above analysis showed that the negative $\Delta E_{TS\text{-}Thiolate}$ values associated with some of the S_N2 chemicals could be resolved by including a counter ion (e.g. sodium) in the calculations. However, in order to remain consistent with the Michael addition analysis (see Chapter 4), $\Delta E_{TS\text{-}Thiolate}$ without the addition of sodium was chosen as the key descriptor in the development of a fragment-based profiler for the S_N2 mechanism. The utility of this descriptor for modelling glutathione reactivity and toxicity is the focus of this chapter.

As outlined in Chapter 3 the key process in the development of a fragment-based *in silico* profiler is the investigation of the SAR relationship between a reactivity descriptor (in this case $\Delta E_{TS\text{-}Thiolate}$) and the possible R-group substituents. Inspection of the chemicals in the dataset showed that there were three factors that varied for the five types of electron-withdrawing groups present (ketones, esters, acids, amides and aromatic groups), these being; the halogen leaving group (bromine or chlorine) and varying substituents at the R_1 and R_2 positions (R-groups as defined in Table 5.2). Therefore, the analysis focused on the effects of these substituents on the calculated $\Delta E_{TS\text{-}Thiolate}$ values within the applicability domain of the experimental assay. Groups where there was no variation in substituents were not investigated (for example, a fragment analysis into R_2 substituents for ketones was not carried out as there is only a single substituent at R_2 for ketones in the dataset, R_2 = H). This resulted in three groups for SAR analysis: position R_1 for brominated ketones and esters, position R_2 for brominated esters, acids and amides and position R_1 for chlorinated esters. Although no analysis was carried out for groups containing only one chemical, a fragment was still included for these chemicals to investigate the use of the $\Delta E_{TS\text{-}Thiolate}$ values for the prediction of glutathione reactivity and toxicity to *Tetrahymena pyriformis.* Section 5.4.2 outlines the fragment development for brominated chemicals with substituents at the R_1 (ketone and esters) and R_2 positions (esters, acids and amides). Section 5.4.3 outlines the fragment development of the chlorinated chemicals at the R_1 position (esters).

Table 5.2. The domain of chemicals in the dataset (all chemicals shown in Table 5.3). The table shows which substituents are present for the various R groups (R_1 and R_2) for each electron withdrawing group class. Additionally, the number of chemicals in each chemical class (N) is shown

Electron withdrawing group	Brominated chemicals	Chlorinated chemicals
Ketones		
	R_1 = CH_2CH_3, *t*-butyl, phenyl, naphthalene, pyrene, thiophene R_2 = H	R_1 = *t*-butyl R_2 = H
	N = 7	
Esters		
	R_1 = CH_3, CH_2CH_3, $(CH_2)_2CH_3$, *t*-butyl, phenyl R_2 = H, CH_3, CH_2CH_3, $(CH_2)_2CH_3$	R_1 = CH_3, CH_2CH_3, $(CH_2)_2CH_3$, *t*-butyl R_2 = H
	N = 15	

Electron withdrawing group	Brominated chemicals	Chlorinated chemicals
Acids		
	R_1 = N/A	R_1 = N/A
	R_2 = CH_2CH_3, $(CH_2)_2CH_3$	R_2 = CH_2CH_3
	N = 3	
Amides		
	R_1 = N/A	R_1 = N/A
	R_2 = H, CH_3	R_2 = H
	N = 3	
Aromatic group		None
	R_1 = Nitrobenzene	
	N = 1	

5.4.1 Rules for fragment development for chemicals acting via an S$_N$2 mechanism

In Chapter 3, a set of rules were established for the development of fragments for Michael addition enabling a systematic approach to be undertaken for their development. This included key information about the structure and geometries of the fragments. In addition, the process of how fragments were compared to one another to establish how many fragments were needed in the final profiler was also detailed. The aim of these rules being to enable the domain of the profiler to be easily extended in the future (as more data become available). Similarly, fragments for the profiler for chemicals acting via an S$_N$2 mechanism were developed using an analogous set of rules:

1. All fragments were developed using the transition state structure upon reaction with a thiolate nucleophile using $\Delta E_{TS\text{-}Thiolate}$ values as the key reactivity descriptor.

2. The $\Delta E_{TS\text{-}Thiolate}$ values for straight chains at each R-position were compared with the $\Delta E_{TS\text{-}Thiolate}$ values of a chain containing one carbon less (or in the case of methyl with hydrogen where applicable); for example, ethyl was compared to methyl and propyl compared to ethyl.

3. Branched chains $\Delta E_{TS\text{-}Thiolate}$ values were compared to the $\Delta E_{TS\text{-}Thiolate}$ value of their straight chain equivalent; for example, *t*-butyl was compared with ethyl.

4. Only ketones and esters contained aromatic substituents (at R$_1$). In all cases these were compared to a methyl group; for example, phenyl, naphthalene, pyrene and thiophene were compared to methyl.

5. Only one R group was investigated at a time whilst the other R group remained constant. For example, R$_1$ remained as hydrogen whilst the effect of substituents at the R$_2$ position was investigated.

6. Individual calculated $\Delta E_{TS\text{-}Thiolate}$ values were rounded to the nearest integer before comparing values rather than rounding the difference in $\Delta E_{TS\text{-}Thiolate}$ between the two values.

7. A cut off value of 1.0 kcal/mol was used to assess if there was a significant difference between two substituents (to determine the need for the inclusion of a fragment in the profiler).

8. Unrounded $\Delta E_{TS\text{-Thiolate}}$ values for fragments (to one decimal place) were used in the modelling of reactivity and toxicity (unlike rounded values, which were used in the fragment analysis).

5.4.2 Development of fragments for brominated chemicals

The eight rules stated above allowed the SAR in terms of $\Delta E_{TS\text{-Thiolate}}$ values to be investigated for positions R_1 and R_2 for chemicals acting via an S_N2 mechanism (where the R-groups investigated are shown in Table 5.2). All $\Delta E_{TS\text{-Thiolate}}$ values in the following fragment analysis will be rounded to the nearest integer (see rule 6). Of the 28 chemicals in the dataset (shown in Table 5.7), 21 were brominated and seven were chlorinated. For the brominated chemicals this covered, 10 brominated esters, six brominated ketones, two brominated acids, two brominated amides and 1-bromomethyl-4-nitrobenzene. Analysis of the dataset revealed three groups that were large enough to allow a SAR analysis to be undertaken (11 chemicals with substituents at position R_1 for brominated ketones and esters, eight chemicals with substituents at position R_2 for brominated esters, acids and amides and four chemicals with substituents at position R_1 for chlorinated esters).

Calculated $\Delta E_{TS\text{-Thiolate}}$ values SAR at position R_1 for brominated ketones and esters

Initially the SAR for $\Delta E_{TS\text{-Thiolate}}$ values when extending the chain length at the R_1 position for brominated ketones and esters was investigated (Figure 5.5, substituent values in Table 5.3). This group contained seven brominated ketones and 10 brominated esters covering seven and five substituents at the R_1 position respectively. Calculated $\Delta E_{TS\text{-Thiolate}}$ values increased by 1 kcal/mol when extending the chain length from a methyl to an ethyl substituent at the R_1 position for both electron-withdrawing groups (compare chemicals 2 with 1 and 9 with 8 in Table 5.3). Similarly, the change in calculated $\Delta E_{TS\text{-Thiolate}}$ values was less than 1 kcal/mol when comparing ethyl to a *t*-butyl

group for both electron-withdrawing groups (compare chemicals 3 with 2 and 11 with 9 in Table 5.3). Additionally, methyl could be used for aromatic substituents phenyl, naphthalene, pyrene and thiophene for ketones (compare chemicals 4-7 with 1 in Table 5.3). The result of this analysis showed that a methyl group could be used to predict the $\Delta E_{TS\text{-}Thiolate}$ values of all alkyl groups (and aryl groups for ketones) at the R_1 position. The exception being the need for an ethyl group for the prediction of chemicals where R_1 = t-butyl (chemicals 11 and 3 in Table 5.3). This resulted in four fragments being required to cover the domain of ketones and esters with substituents at the R_1 position (these being R_1 = methyl and ethyl for both ketones and esters).
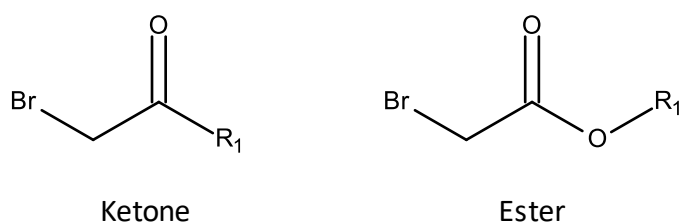

Figure 5.5. The general structure of primary brominated ketones and esters investigated in this study, R_1 groups as defined in Table 5.3

Table 5.3. Calculated $\Delta E_{TS\text{-}Thiolate}$ (kcal/mol) values for brominated ketones and esters substituted at the R_1 positions

ID	Substituent name R_1	$\Delta E_{TS\text{-}Thiolate}$ (kcal/mol)	Fragment substituent	Fragment $\Delta E_{TS\text{-}Thiolate}$ (kcal/mol)
Ketones				
1	Methyl	-1	CH_3	-1
2	Ethyl	0	CH_3	-1
3	*t*-Butyl	0	CH_2CH_3	0
4	Phenyl	-2	CH_3	-1
5	Naphthalene	-2	CH_3	-1
6	Pyrene	0	CH_3	-1
7	Thiophene	-1	CH_3	-1
Esters				
8	Methyl	0	CH_3	0
9	Ethyl	1	CH_3	0
10	Propyl	1	CH_3	0
11	*t*-butyl	1	CH_2CH_3	1
12	Phenyl	0	CH_3	0

Calculated $\Delta E_{TS\text{-}Thiolate}$ values SAR at position R_2 for brominated esters, acids and amides.

The analysis for varying substituents at the R_2 position was applicable to three chemicals groups (esters, acids and amides, Figure 5.6). There were 11 brominated esters, two brominated acids and two brominated amides in the dataset covering four substituents for esters and two for both acids and amides at the R_2 position (based on Table 5.4). The results showed that for all three electron-withdrawing groups, the calculated $\Delta E_{TS\text{-}Thiolate}$ values differed significantly when going from hydrogen to methyl substituents (compare chemicals 2 with 1, 6 with 5 and 10 with 9 in Table 5.4). This increase in the calculated $\Delta E_{TS\text{-}Thiolate}$ value is expected due to increased steric bulk around the reactive site. However, increasing the chain length further from methyl to ethyl resulted in the

calculated $\Delta E_{TS-Thiolate}$ values being within 1 kcal/mol (compare chemicals 3 with 2 and 7 with 6 in

Table 5.4). This consistency in calculated $\Delta E_{TS-Thiolate}$ values was also seen when extending the chain

length from ethyl to propyl (compare chemicals 4 with 3 and 8 with 7 in Table 5.4). This showed

that only the addition of the methyl group (going from primary to secondary halide) has an effect

on calculated $\Delta E_{TS-Thiolate}$ values, (increasing the chain length further resulted in no change in the

calculated $\Delta E_{TS-Thiolate}$ values). This resulted in two fragments being used to cover the brominated

electron-withdrawing groups for R_2 substituents (this being R_2 = Hydrogen and methyl). Although

the groups for brominated acids and amides are small, it can be assumed that their applicability

extends to acids and amides with larger substituents at the R_2 position. This assumption is based

on consistency in calculated $\Delta E_{TS-Thiolate}$ values when extending the chain length at the R_2 position

for other chemical groups (e.g. brominated esters substituted at the R_2 position – see chemicals 1-
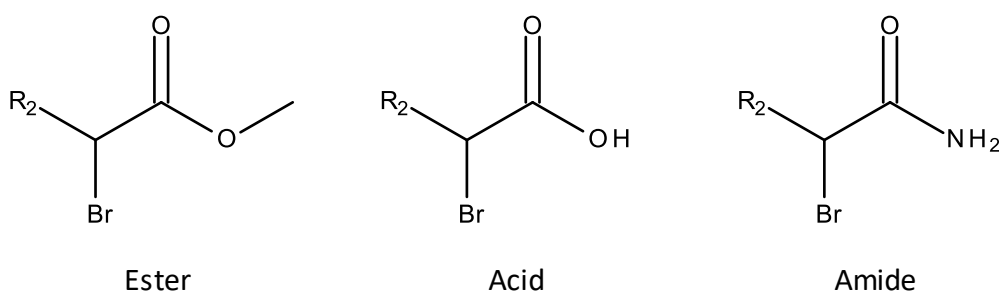
4 in Table 5.4).

Figure 5.6. The general structure of secondary brominated esters, acid and amides investigated in

this study (R_2 groups as defined in Table 5.4)

Table 5.4. Calculated $\Delta E_{\text{TS-Thiolate}}$ (kcal/mol) values for brominated esters, acids and amides at the R_2 positions

ID	Substituent name R	$\Delta E_{\text{TS-Thiolate}}$ (kcal/mol)	Fragment Substituent	Fragment $\Delta E_{\text{TS-Thiolate}}$ (kcal/mol)
Esters				
1	Hydrogen	0	H	0
2	Methyl	5	CH_3	5
3	Ethyl	5	CH_3	5
4	Propyl	5	CH_3	5
Acids				
5	Hydrogen	0	H	0
6	Methyl	5	CH_3	5
7	Ethyl	4	CH_3	5
8	Propyl	4	CH_3	5
Amides				
9	Hydrogen	2	H	2
10	Methyl	7	CH_3	7

Brominated chemicals for which no SAR analysis was possible

Of the 21 chemicals there was only a single chemical activated by nitrobenzene group, this prevented a SAR analysis from being carried out for this class. Given this, a fragment was included for 1-bromomethyl-4-nitrobezene to assess the ability of the $\Delta E_{\text{TS-Thiolate}}$ values to predict glutathione reactivity and toxicity to *Tetrahymena pyriformis* for this chemical.

5.4.3 Development of fragments for chlorinated chemicals

The dataset contained seven chlorinated chemicals - four esters, one ketone, one acid and an amide. Given this, the only group for which a SAR analysis could be carried out was the chlorinated esters substituted at the R_1 position (R groups as defined in Table 5.2). This analysis resulted in the

same outcome as was seen for the brominated esters, where no change in the $\Delta E_{TS-Thiolate}$ values were calculated beyond a methyl substituent at the R_1 position (chemicals 1-4 in Table 5.5). This resulted in two fragments being required to cover the four chlorinated esters in the dataset (these being R_1 = methyl and ethyl).
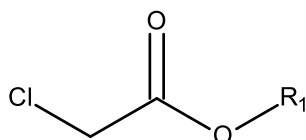

Figure 5.7. The general structure of secondary brominated esters, acid and amides investigated in this study (R_1 groups as defined in Table 5.5)

Table 5.5. Calculated $\Delta E_{TS-Thiolate}$ (kcal/mol) values for chlorinated esters substituted at the R_1 positions

ID	Substituent name R_1	$\Delta E_{TS-Thiolate}$ (kcal/mol)	Fragment Substituent	Fragment $\Delta E_{TS-Thiolate}$ (kcal/mol)
1	Methyl	3	CH_3	3
2	Ethyl	4	CH_3	3
3	Propyl	3	CH_3	3
4	*t*-Butyl	4	CH_2CH_3	3

Chlorinated chemicals for which no SAR analysis was possible
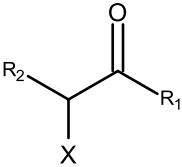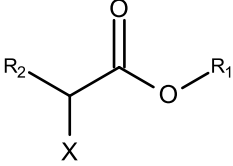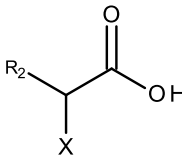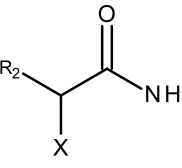
Of the seven chlorinated chemicals in the dataset, there were three chemicals for which no SAR analysis could be applied due to there being no other structurally related chemicals (1-chloropinacalone, 2-chloroacetamide and 2-chlorobutyric acid). The fragments used to define these chemicals are discussed in the next section.
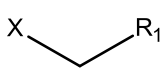
5.5 Applicability domain of the fragment-based *in silico* profiler for the S$_N$2 mechanism

The above analysis resulted in the definition of ten fragments for the brominated chemicals and five fragments for the chlorinated chemicals (summarised in Table 5.6). All chemical classes for

brominated chemicals showed that a methyl substituent was capable of predicting the $\Delta E_{TS\text{-}Thiolate}$

values for alkyl and aryl substituents at both R-positions. The exception being the need for an ethyl

group for *t*-butyl substituents for brominated ketones and esters. The SAR analysis for chlorinated

esters substituted at the R_1 position resulted in an identical outcome to that calculated for the

brominated equivalents. Given this, an assumption was made that an analogous set of fragments

to those defined for the brominated chemicals could be applied to extend the applicability domain

of the profiler to cover an equivalent set of chlorinated chemicals. As such, the same set of R_1 and

R_2 substituents were used to cover both brominated and chlorinated chemicals. This resulted in a

total of 20 fragments to cover the expanded domain (fragments defined in Table 5.6).

Table 5.6. The fragments required to cover the domain of chemicals in Table 5.2. Substituents used

in the expanded domain for chlorinated chemicals are shown in italics.

Chemical group	Structure	X = Br	X = Cl
Ketones		$R_1 = CH_3, CH_2CH_3$ $R_2 = H$	$R_1 = CH_3, (CH_2CH_3)$ $R_2 = H$
Esters		$R_1 = CH_3, CH_2CH_3$ $R_2 = H, CH_3$	$R_1 = CH_3, CH_2CH_3$ $R_2 = H, (CH_3)$
Acids		$R_1 = N/A$ $R_2 = H, CH_3$	$R_1 = N/A$ $R_2 = (H), CH_3$
Amides		$R_1 = N/A$ $R_2 = H, CH_3$	$R_1 = N/A$ $R_2 = H, (CH_3)$

Chemical group	Structure	X = Br	X = Cl
Aromatic group	X⌄R₁	R₁ = Nitrobenzene	R₁ = *(Nitrobenzene)*
Total Number of Fragments		N = 10	N = 10

5.6 Prediction of glutathione reactivity and *Tetrahymena pyriformis* toxicity using the fragment-based *in silico* profiler for SN2

The above analysis identified the need for 15 fragments to cover the structural domain of the 28 chemicals within the dataset with an additional five fragments being required to cover the expanded domain of chlorinated chemicals. This rationale is similar to the results from Chapter 4 where 24 fragments were required for the prediction of reactivity for 54 Michael acceptors. The ability of the fragments and corresponding $\Delta E_{TS\text{-Thiolate}}$ were then used to predict glutathione reactivity (-Log RC_{50}) and *Tetrahymena pyriformis* toxicity (-Log IGC_{50}). Briefly, RC_{50} corresponds to reactive concentration, which results in 50% of GSH being depleted in 120 minutes (information on glutathione depletion assay found in Chapter 1, Section 1.6.1). Additionally, IGC_{50} corresponds to impairment growth concentration for 50% of control populations when exposed to the test chemical over a 40 hour period. Using the $\Delta E_{TS\text{-Thiolate}}$ values as the independent variables in a linear regression analysis showed them to be capable of predicting glutathione reactivity (Model 5.1, Figure 5.8, experimental and predicted values in Table 5.7) (103).

$$-\mathrm{Log}\,RC_{50}(\mathrm{pred}) = 1.03 - 0.34\,\Delta E_{TS-Thiolate} \qquad \text{(Model 5.1)}$$

N = 28, R^2 = 0.85, R^2-adj = 0.84, R^2-pred = 0.82, Average error = 0.31
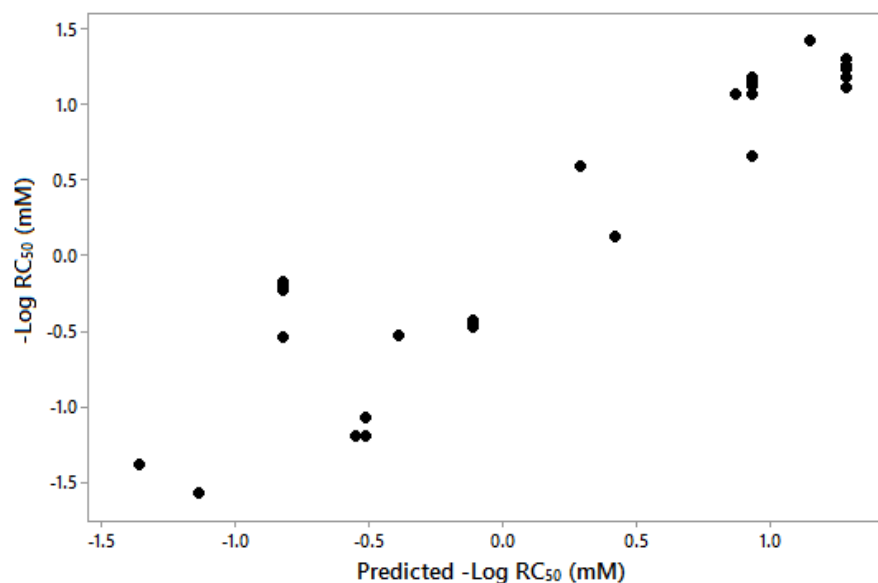
Figure 5.8. The predicted –Log RC_{50} against experimental –Log RC_{50} for all 28 S_N2 chemicals

The predicted glutathione reactivity data values obtained from Model 5.1 (–Log RC_{50}(pred)) were subsequently utilised as the independent variable for the prediction of toxicity to *Tetrahymena pyriformis* (-Log IGC_{50}). The resulting regression analysis is shown in Model 5.2, with the statistics of this model being in keeping with previous research in which experimental glutathione reactivity data were used to predict toxicity to *Tetrahymena pyriformis* (Figure 5.9) (42).
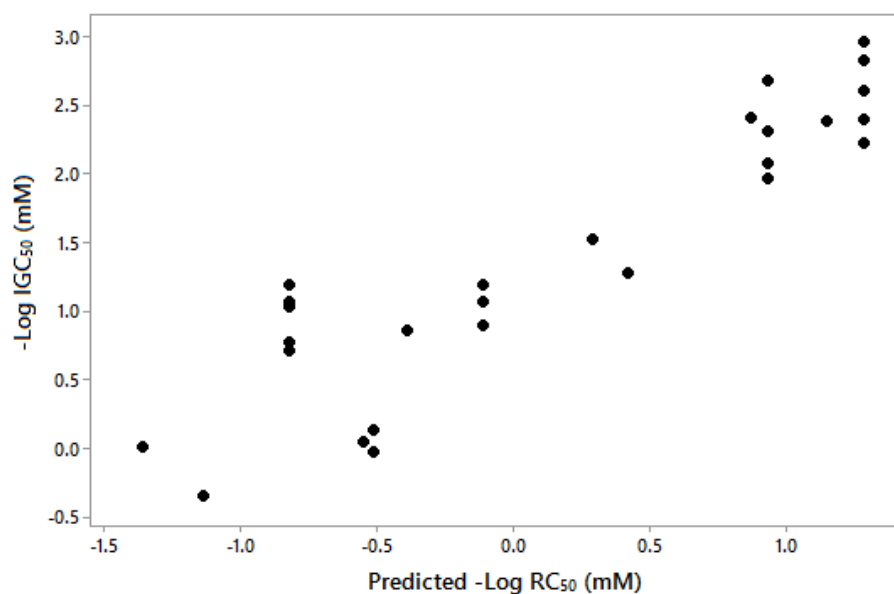
Figure 5.9. The predicted –Log RC_{50} against experimental –Log IGC_{50} values for all 28 S_N2 chemicals

$$-\text{Log IGC}_{50} = 1.27 + 1.00 - Log\ RC_{50}\ (pred) \qquad \text{(Model 5.2)}$$

N = 28, R^2 = 0.84, R^2-adj = 0.83, R^2-pred = 0.81, Average error = 0.31

N = 28, R^2 = 0.84, R^2-adj = 0.83, R^2-pred = 0.81, Average error = 0.31

Table 5.7. Chemicals acting via an S_N2 mechanism with corresponding –Log RC_{50}, -Log IGC_{50} (toxicity to *Tetrahymena pyriformis*), predicted –Log RC_{50} and Predicted –Log IGC_{50} values investigated in the current study. Chemical names taken directly from reference (2). Fragment names are based on their IUPAC name

ID	Chemical	-Log RC_{50} (mM)	-Log IGC_{50} (mM)	Fragment	Predicted –Log RC_{50} (mM) (Model 5.1)	Predicted –Log IGC_{50} (mM) (Model 5.2)
1	1-Bromomethyl-4-nitrobenzene	0.66	2.30	1-Bromomethyl-4-nitrobenzene	0.93	2.20
2	2-(2-Bromoacetyl) thiophene	1.11	2.22	1-Bromo-2-propanone	1.28	2.55
3	1-Bromo-2-butanone	1.30	2.60	1-Bromo-2-propanone	1.28	2.55
4	1-Bromopinacolone	1.42	2.38	1-Bromo-2-butanone	1.14	2.41
5	1-Chloropinacolone	0.12	1.27	1-Chloro-2-butanone	0.42	1.69
6	2-Bromoacetophenone	1.26	2.82	1-Bromo-2-propanone	1.28	2.55
7	2-(2-Bromoacetyl) naphthalene	1.18	2.96	1-Bromo-2-propanone	1.28	2.55
8	1-(Bromoacetyl) pyrene	1.24	2.39	1-Bromo-2-propanone	1.28	2.55
9	Ethyl bromoacetate	1.07	2.68	Methyl bromoacetate	0.92	2.19
10	Ethyl chloroacetate	-0.48	1.06	Methyl chloroacetate	-0.12	1.15
11	Methyl bromoacetate	1.18	2.96	Methyl bromoacetate	0.92	2.19
12	Methyl chloroacetate	-0.45	0.89	Methyl chloroacetate	-0.12	1.15
13	Propyl bromoacetate	1.14	2.08	Methyl bromoacetate	0.92	2.19
14	Propyl chloroacetate	-0.43	1.18	Methyl chloroacetate	-0.12	1.15
15	*t*-Butyl bromoacetate	1.07	2.68	Ethyl bromoacetate	0.86	2.13

ID	Chemical	-Log RC$_{50}$ (mM)	-Log IGC$_{50}$ (mM)	Fragment	Predicted –Log RC$_{50}$ (mM) (Model 5.1)	Predicted –Log IGC$_{50}$ (mM) (Model 5.2)
16	*t*-Butyl chloroacetate	-0.53	0.85	Ethyl chloroacetate	-0.40	0.87
17	Phenyl bromoacetate	1.14	2.08	Methyl bromoacetate	0.92	2.19
18	2-Bromoacetamide	0.59	1.52	2-Bromoacetamide	0.28	1.55
19	2-Chloroacetamide	-1.20	0.04	2-Chloroacetamide	-0.56	0.71
20	Methyl-2-bromopropionate	-0.18	1.18	Methyl-2-bromopropionate	-0.83	0.44
21	Methyl-2-bromobutyrate	-0.54	1.02	Methyl-2-bromopropionate	-0.83	0.44
22	2-Bromopropionamide	-1.40	0.00	2-Bromopropanamide	-1.36	-0.09
23	2-Bromobutyric acid	-1.20	0.12	2-Bromopropanoic acid	-0.52	0.75
24	2-Chlorobutyric acid	-1.58	-0.35	2-Chloropropanoic acid	-1.14	0.13
25	2-Bromovaleric acid	-1.08	-0.04	2-Bromopropanic acid	-0.52	0.75
26	Ethyl-2-bromovalerate	-0.23	0.70	Methyl-2-bromopropionate	-0.83	0.44
27	Ethyl-2-bromobutyrate	-0.20	0.77	Methyl-2-bromopropionate	-0.83	0.44
28	Ethyl-2-bromopropionate	-0.23	1.06	Methyl-2-bromopropionate	-0.83	0.44

5.7 Concluding remarks

The aim of this chapter was to develop a fragment-based profiler for the S_N2 mechanism by adopting the method that was successfully applied to the Michael addition domain in Chapters 3 and 4. The results showed that the fragment-based profiler for the S_N2 mechanism was able to predict both glutathione reactivity and toxicity to *Tetrahymena pyriformis* for a series of activated S_N2 chemicals. The predicted toxicity values towards *Tetrahymena pyriformis* were in keeping with a previous study in which experimental reactivity data were used to predict toxicity. The results of this chapter highlight the benefits of developing fragment-based reactivity profilers of this type, and their application in the prediction of toxicological endpoints for which the formation of a covalent bond is the key driver of potency within well-defined mechanistic domains.

Chapter 6: Discussion

This final chapter will present an overall summary of the findings of and the conclusions drawn from the work in this thesis. The first section will focus on the background to the project and the rationale for the research undertaken. This will be followed by an overview of the key findings from the research presented in Chapters 3, 4 and 5. The final part of the chapter will focus on potential future work and applications of the research with regards to the potential for extending the *in silico* fragment-based reactivity profiler to other mechanistic domains and nucleophiles relevant to toxicity. The work will be discussed in the context of developing alternatives to traditional toxicity testing.

6.1 Summary of work

The work in this thesis has been focused on the development of alternative methods for toxicological risk assessment, more specifically the development of fragment-based *in silico* profilers for thiol reactivity. The motivation for the development of alternative methods arises from the seventh amendment to the cosmetic directive and the REACH legislation, which demand greater speed of chemical assessment with minimal animal usage. One of the key *in silico* methods used to satisfy the non-animal testing approaches envisaged by such legislation is category formation and read-across. The importance of considering the MIE when developing a chemical category was outlined in Chapter 1. This chapter also highlighted the importance of the mechanistic organic chemistry associated with the formation of a covalent bond between an electrophile and a biological macromolecule when developing such chemical categories. A number of studies have shown that such mechanistic chemistry can be encoded as structural alert-based *in silico* profilers which have been encoded into a number of freely available *in silico* tools, such as the OECD QSAR Toolbox. Finally, Chapter 1 outlined that research has shown that, when available, both experimental and computational measurements of reactivity can be used to build QSAR models within a mechanism-based category, enabling toxicological potency to be predicted. From a

computational point of view the limiting factor for the inclusion of these approaches in tools such as the OECD QSAR Toolbox is the need for complex and time-consuming QM calculations. With this in mind, the key aim of this thesis was to develop fragment-based *in silico* profilers capable of predicting reactivity (and, by extension, toxicity) for chemicals acting via Michael addition and S_N2 mechanisms.

Chapter 2 outlined the theory behind the QM methods used throughout this thesis. One of the fundamental equations used in QM was introduced – the Schrödinger equation. The Schrödinger equation allows a number of important properties of a particle to be determined (such as energy). The Schrödinger equation is composed of two key parts, the Hamiltonian (provides information on kinetic energy and interaction between particles) and the wavefunction (a mathematical function that describes the motion of the particle through space). Additional concepts were defined in order to solve the Schrödinger equation, these being the Born-Oppenheimer approximation and HF theory. Density functional theory (DFT) was outlined as an alternative to solving the Schrödinger equation by replacing the many-body electronic wavefunction used in HF with the electron density. The appeal of DFT comes from the inclusion of an exchange-correlation functional to account for electron-electron correlation. The QM methods used in this thesis utilised a hybrid functional form of DFT with the inclusion of additional diffusion and polarizability functions.

Chapter 3 focused on the development of fragments which formed the basis of the novel fragment-based *in silico* profiler for Michael addition. The development of the fragment-based *in silico* profiler required the use of an appropriate descriptor which could successfully be used to relate chemical structure to chemical reactivity. Two descriptors for the Michael addition reaction were investigated, both being calculated using DFT. The first of these being the calculated activation energy derived from utilising the energy of the transition state structure linking the reactants and products for chemicals acting via Michael addition ($\Delta E_{TS\text{-}Thiolate}$). The second descriptor that was considered related to the activation energy calculated using the energy of the ionised intermediate

structure that exists on the Michael addition potential energy surface ($\Delta E_{\text{Int-Thiolate}}$). Analysis within Chapter 3 showed these two descriptors to be highly correlated with one another. Given this, it was decided that all fragments for chemicals acting via Michael addition would be calculated using $\Delta E_{\text{INT-Thiolate}}$ due to it being significantly easier to compute. This enabled a SAR analysis to be carried out to investigate the variability of $\Delta E_{\text{INT-Thiolate}}$ upon changing the substitutions (aliphatic and aromatic) at the various R-groups for the Michael acceptors. This was to establish the point at which $\Delta E_{\text{INT-Thiolate}}$ no longer varied with a change in substitution at the respective R-group. The calculated $\Delta E_{\text{INT-Thiolate}}$ stopped varying after 2 aliphatic carbons at all R groups for α,β-unsaturated aldehydes, ketones and esters. The exception being the need to include branched chains (*i*-propyl and *t*-butyl) at various R positions for the three groups. This resulted in the definition of 294 fragments to cover the domain of α,β-unsaturated aldehydes, ketones and esters (defined fragments are summarised in Table 3.11). This covers all α,β-unsaturated aldehydes, ketones and esters with alkyl or aryl substitutions at the various R-positions (potentially covering an vast number of chemicals). The development of such fragments is a crucial step in the development of the fragment-based *in silico* profiler for Michael addition thiol reactivity.

Chapter 4 demonstrated the ability of the fragment-based *in silico* profiler for Michael addition developed in Chapter 3 to predict thiol reactivity as determined in a glutathione depletion assay. The results showed that fragment-based calculations for $\Delta E_{\text{INT-Thiolate}}$ values, in conjunction with an additional descriptor related to the solvent accessible surface area at the α-position, were capable of predicting the glutathione reactivity for the majority of the chemicals in the dataset. Two sets of chemicals were poorly predicted by the approach; volatile esters with an extended substituent at the β-carbon and chemicals containing a conjugated benzene ring as part of the polarising group. Chapter 4 also demonstrated that the predicted glutathione reactivity values generated by the fragment-based *in silico* profiler for Michael addition could also be used to predict skin sensitisation potency and toxicity to *Tetrahymena pyriformis* within well-defined, endpoint specific applicability domains. The novel research output from this chapter being the development of an *in silico* profiler

for Michael addition and its ability to generate rapid predictions of glutathione reactivity. The utility of these values in the prediction of toxicity was demonstrated by predicting skin sensitisation potency and toxicity to *Tetrahymena pyriformis* within well-defined, end-point specific applicability domains.

The expansion of the fragment-based *in silico* profiler to chemicals which form covalent bonds through an S_N2 mechanism was investigated in Chapter 5. The research enabled a fragment-based *in* silico profiler for the S_N2 mechanism to be developed based on a total of 19 fragments covering chlorinated and brominated chemicals activated by an electron-withdrawing group. This Chapter also outlined the use of this profiler for the prediction of toxicity to *Tetrahymena pyriformis*. The results highlighted the ability of the fragment-based *in silico* profiler to predict reactivity and toxicity of chemicals which form covalent bonds via mechanisms other than Michael addition something not extensively studied in the literature to date. This study expands on the novelty of the fragment-based *in silico* profiler developed in Chapter 4 by including chemicals from additional mechanistic domains.

6.2 Prospects for future work - expansion of the fragment-based *in silico* profiler to additional mechanistic domains and endpoints

The work carried out in this thesis highlights the development and use of two novel fragment-based *in silico* profilers for the Michael addition and the S_N2 mechanistic domains. These profilers were shown to be capable of predicting both glutathione-based chemical reactivity and toxicity (skin sensitisation and aquatic toxicity). Given this, it is likely that the approach will be applicable to the other mechanistic domains relevant to toxicology (S_NAr, S_N1, acylation and Schiff base formation). The expansion of the fragment-based *in silico* profiler to additional domains is dependent on the availability of reactivity data. This is likely to be derived primarily from glutathione and other reactivity data sources (a database of 3089 chemicals with reactivity data exists covering all six mechanisms discussed). It may also be possible that additional fragments could be developed from

alternative data sources giving the opportunity to cover a wider range of chemistry within each of the mechanistic domains. Additionally given the high degree of correlation between glutathione reactivity and toxicity to *Tetrahymena pyriformis* it is possible that the latter could be used as a surrogate for reactivity data. The benefit of such an analysis becomes clear when considering that a database of 2072 chemicals with toxicity to *Tetrahymena pyriformis* values is publically available (83). Alternatively, an investigation could be carried out into the development of fragments with nitrogen-based nucleophiles such as lysine or butylamine (72, 85, 86). Lysine reactivity is likely to be of use for the development of fragments for the Schiff base domain as chemicals acting via this mechanism do so only with nitrogen based nucleophiles. As an example of the importance of this to toxicology, one only has to consider that lysine reactivity has been implicated as being key in endpoints such as respiratory sensitisation (where the MIE is covalent bond formation covering similar chemistry to skin sensitisation).

In addition to the expansion to additional mechanistic domains relevant to toxicological endpoints, there are other areas of research where knowledge of the rate of covalent mechanisms may be of use. For example, in the development of covalently binding drugs. The significance of reactivity for covalent acting drugs was investigated in a study in which calculated $\Delta E_{INT\text{-}Thiolate}$ values were used to predict the reactivity towards glutathione for a set of Michael acceptors (this study is discussed in detail in Chapter 3 Section 3.1) (75). These are classified as "warhead" type drugs which have found use in the treatment of serious disease such as cancer. Consider the chemicals ibrutinib and acalabrutinib (Figure 6.2) - both of these chemicals bind with biological proteins through Michael addition due to the presence of the alkene / alkyne moiety adjacent to an electron-withdrawing carbonyl group. Acalabrutinib is designed to be more potent and selective than ibrutinib. The increased potency of acalabrutinib is likely due to the difference in reactivity between an alkyne and alkene Michael acceptor with the alkyne being significantly more reactive. However, design of such drugs has been considered relatively controversial due to their affinity to form covalent bonds with off-target proteins. As such, the development of covalently binding drugs involves a process

of carefully tuning the reactivity and specificity which is complementary to the target (115). Therefore, a method that allows the reactivity of chemical to be predicted rapidly would clearly be beneficial in the design process of such molecules.
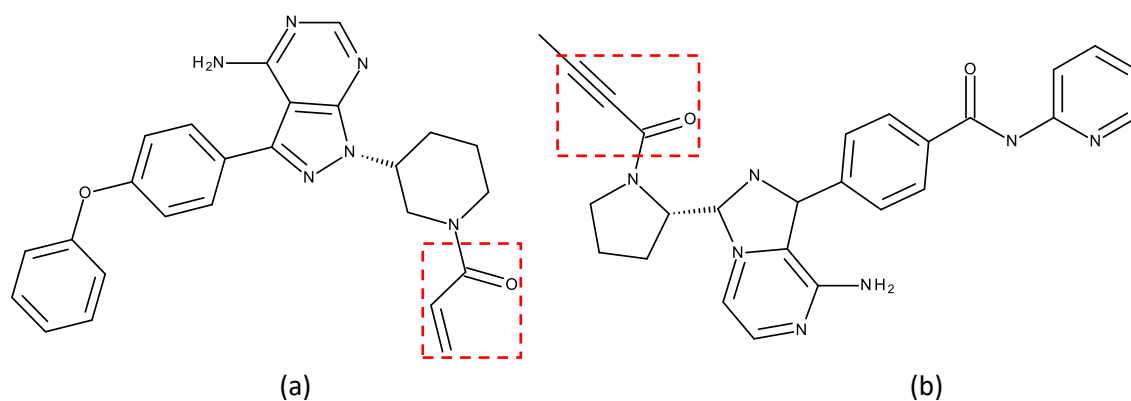

Figure 6.2. Ibrutinib (A) and Acalabrutinib (B) highlighting the reactivity centres in red

6.3 Concluding remarks

The work in this thesis has focused on the development of novel fragment-based *in silico* profilers for the prediction of thiol reactivity and toxicity for chemicals acting via Michael addition and S_N2 mechanistic domains. The approach enables chemical reactivity to be predicted without the use of time-consuming quantum mechanics calculations which involve the use of proprietary software. The work in this thesis has also shown that it is possible to expand the domain of the fragment-based *in silico* profilers through the calculation of additional fragments, enabling additional chemicals and/or mechanistic domains to be readily investigated. This method is likely to be of primary use in regulatory toxicology with the potential to be implemented in such freely available tools as the OECD QSAR toolbox. Such implementation would enable the development of mechanism-based QSARs for reactivity to be utilised within the chemical category approach to data gap filling. Additionally, this approach is likely to be of use in other areas of research such as the development of covalent binding drugs such as anti-cancer agents.

References

1. Gelbke HP. Regulatory toxicology: objectives and tasks defined by the working group of the German society of experimental and clinical pharmacology and toxicology. *Toxicol. Lett*. 2002; 126(3):159-60.

2. Bendele RA. Safety Assessment Programs for United-States Regulatory Agencies - a Perspective of Requirements and Compliance. *Toxicol Pathol*. 1994; 22(2):95-104.

3. OECD. About the OECD Available online at http://www.oecd.org/about/ (accessed September 06, 2017). 1961.

4. OECD. Test No. 406: Skin Sensitisation. Available online at http://www.oecd-ilibrary.org/environment/test-no-406-skin-sensitisation_9789264070660-en;jsessionid=37lm9ggskc3fj.x-oecd-live-03 (accessed October 17, 2017). 1992.

5. OECD. Test No. 429: Skin Sensitisation Local Lymph Node Assay. Available online at http://www.oecd-ilibrary.org/environment/test-no-429-skin-sensitisation_9789264071100-en;jsessionid=23rr48t9ve6bs.x-oecd-live-03 (accessed November 13, 2017). 2010.

6. OECD. Test No. 442A: Skin Sensitisation Local Lymph Node Assay: DA. Available online at http://www.oecd-ilibrary.org/environment/test-no-442a-skin-sensitization_9789264090972-en;jsessionid=23rr48t9ve6bs.x-oecd-live-03 (accessed November 13, 2017). 2010.

7. OECD. Test No. 442B: Skin Sensitisation Local Lymph Node Assay: BrdU-ELISA. Available online at http://www.oecd-ilibrary.org/environment/test-no-442b-skin-sensitization_9789264090996-en;jsessionid=23rr48t9ve6bs.x-oecd-live-03 (accessed November 13, 2017). 2010.

8. EC. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Off. *J. Eur. Union*, L396/1 of 30.12.2006. Commission of the European Communities. 2006.

9. Enoch SJ, Cronin MT, Schultz TW, Madden JC. Quantitative and mechanistic read across for predicting the skin sensitization potential of alkenes acting via Michael addition. *Chem Res Toxicol.* 2008; 21(2):513-20.

10. Cosmetics Europe. cosmetics and personal care industry overview 2015. Available online at: https://www.cosmeticseurope.eu/cosmetics-industry (accessed June 19th 2017). 2015

11. European Commission. Timetables for the phasing-out of animal testing in the framework of the 7th Amendment to the Cosmetics Directive (Council Directive 76/768/EEC) Available online at http://ec.europa.eu/DocsRoom/documents/13124/attachments/1/translations (accessed September 05, 2017). 2004.

12. OECD. Test No. 420: Acute Oral Toxicity - Fixed Dose Procedure. Available online at http://www.oecd-ilibrary.org/environment/test-no-420-acute-oral-toxicity-fixed-dose-procedure_9789264070943-en;jsessionid=37lm9ggskc3fj.x-oecd-live-03 (accessed October 17, 2017). 2002.

13. OECD. Test No. 423: Acute Oral Toxicity - Acute Toxic Class Method. Available online at http://www.oecd-ilibrary.org/environment/test-no-423-acute-oral-toxicity-acute-toxic-class-method_9789264071001-en;jsessionid=37lm9ggskc3fj.x-oecd-live-03 (accessed October 17, 2017). 2002.

14. OECD. Test No. 425: Acute Oral Toxicity: Up-and-Down Procedure. Available online at http://www.oecd-ilibrary.org/environment/test-no-425-acute-oral-toxicity-up-and-down-procedure_9789264071049-en;jsessionid=37lm9ggskc3fj.x-oecd-live-03 (accessed October 17, 2017). 2008.

15.   OECD. Test No. 411: Subchronic Dermal Toxicity: 90-day Study. Available online at http://www.oecd-ilibrary.org/environment/test-no-411-subchronic-dermal-toxicity-90-day-study_9789264070769-en;jsessionid=37lm9ggskc3fj.x-oecd-live-03 (accessed October 17, 2017). 1981.

16.   OECD. Test No. 412: Subacute Inhalation Toxicity: 28-Day Study. Available online at http://www.oecd-ilibrary.org/environment/test-no-412-subacute-inhalation-toxicity-28-day-study_9789264070783-en;jsessionid=37lm9ggskc3fj.x-oecd-live-03 (accessed October 17, 2017). 2017.

17.   OECD. Test No. 413: Subchronic Inhalation Toxicity: 90-day Study. Available online at http://www.oecd-ilibrary.org/environment/test-no-413-subchronic-inhalation-toxicity-90-day-study_9789264070806-en;jsessionid=37lm9ggskc3fj.x-oecd-live-03 (accessed October 17, 2017). 2017.

18.   OECD. Test No. 474: Mammalian Erythrocyte Micronucleus Test. Available online at http://www.oecd-ilibrary.org/environment/test-no-474-mammalian-erythrocyte-micronucleus-test_9789264264762-en;jsessionid=37lm9ggskc3fj.x-oecd-live-03 (accessed October 17, 2017). 2016.

19.   OECD. Test No. 475: Mammalian Bone Marrow Chromosomal Aberration Test. Available online   at   http://www.oecd-ilibrary.org/environment/test-no-475-mammalian-bone-marrow-chromosomal-aberration-test_9789264264786-en;jsessionid=37lm9ggskc3fj.x-oecd-live-03 (accessed October 17, 2017). 2016.

20.   OECD. Test No. 478: Rodent Dominate Lethal Test. Available online at http://www.oecd-ilibrary.org/environment/test-no-478-rodent-dominant-lethal-test_9789264264823-en;jsessionid=37lm9ggskc3fj.x-oecd-live-03 (accessed October 17, 2017). 2016.

21.   OECD. Test No. 483: Mammalian Spermatogonial Chromosomal Aberration Test. Available online   at   http://www.oecd-ilibrary.org/environment/test-no-483-mammalian-spermatogonial-chromosomal-aberration-test_9789264264847-en;jsessionid=37lm9ggskc3fj.x-oecd-live-03 (accessed October 17, 2017). 2016.

22.   OECD. Test No. 488: Transgenic Rodent Somatic and Germ Cell Gene Mutation Assays. Available online at http://www.oecd-ilibrary.org/environment/test-no-488-transgenic-rodent-somatic-and-germ-cell-gene-mutation-assays_9789264203907-en;jsessionid=37lm9ggskc3fj.x-oecd-live-03 (accessed October 17, 2017). 2013.

23.   OECD.   Test   No.   417:   Toxicokinetic.   Available   online   at   http://www.oecd-ilibrary.org/environment/test-no-417-toxicokinetics_9789264070882-en;jsessionid=37lm9ggskc3fj.x-oecd-live-03 (accessed October 17, 2017). 2010.

24.   OECD. Test No. 451: Carcinogenicity Stuidies. Available online at http://www.oecd-ilibrary.org/environment/test-no-451-carcinogenicity-studies_9789264071186-en;jsessionid=37lm9ggskc3fj.x-oecd-live-03 (accessed October 17, 2017). 2009.

25.   OECD. Test No. 453: Combined Chronic Toxicity/Carcinogenicity Studies. Available online at http://www.oecd-ilibrary.org/environment/test-no-453-combined-chronic-toxicity-carcinogenicity-studies_9789264071223-en;jsessionid=37lm9ggskc3fj.x-oecd-live-03   (accessed October 17, 2017). 2009.

26.   OECD. Test No. 416: Two-Generation Reproduction Toxicity. Available online at http://www.oecd-ilibrary.org/environment/test-no-416-two-generation-reproduction-toxicity_9789264070868-en;jsessionid=37lm9ggskc3fj.x-oecd-live-03 (accessed October 17, 2017). 2001.

27.   OECD. Test No. 421: Reproduction/Developmental Toxicity Screening Test. Available online at   http://www.oecd-ilibrary.org/environment/test-no-421-reproduction-developmental-toxicity-screening-test_9789264264380-en;jsessionid=37lm9ggskc3fj.x-oecd-live-03 (accessed October 17, 2017). 2016.

28.   Aptula AO, Roberts DW. Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: General principles and application to reactive toxicity. *Chem Res Toxicol*. 2006; 19(8):1097-105.

29.   OECD. Test No. 202: Daphnia sp. Acute Immobilisation Test. Available online at http://www.oecd-ilibrary.org/environment/test-no-202-daphnia-sp-acute-immobilisation-test_9789264069947-en (accessed October 17, 2017). 2004.

30.   OECD. Test No. 405: Acute Eye Irritation/Corrosion. Available online at http://www.oecd-ilibrary.org/environment/test-no-405-acute-eye-irritation-corrosion_9789264185333-en;jsessionid=37lm9ggskc3fj.x-oecd-live-03 (accessed October 17, 2017). 2017.

31.   OECD. Test No. 403: Acute Inhalation Toxicity. Available online at http://www.oecd-ilibrary.org/environment/test-no-403-acute-inhalation-toxicity_9789264070608-en;jsessionid=37lm9ggskc3fj.x-oecd-live-03 (accessed October 17, 2017). 2009.

32.   OECD. Test No. 407: Repeated Dose 28-day Oral Toxicity in Rodents. Available online at http://www.oecd-ilibrary.org/environment/test-no-407-repeated-dose-28-day-oral-toxicity-study-in-rodents_9789264070684-en;jsessionid=37lm9ggskc3fj.x-oecd-live-03 (accessed October 17, 2017). 2008.

33.   OECD. Test No. 203: Fish, Acute Toxicity Test. Available online at http://www.oecd-ilibrary.org/environment/test-no-203-fish-acute-toxicity-test_9789264069961-en (accessed October 17, 2017). 1992.

34.   Combes R, Barratt M, Balls M. An overall strategy for the testing of chemicals for human hazard and risk assessment under the EU REACH system. *Atla-Altern Lab Anim*. 2003;31(1):7-19.

35.   Schultz TW, Rogers K, Aptula AO. Read-across to rank skin sensitization potential: subcategories for the Michael acceptor domain. *Contact Dermatitis*. 2009; 60(1):21-31.

36.   Doke SK, Dhawale SC. Alternatives to animal testing: A review. *Saudi Pharm J*. 2015;23(3):223-9.

37.   Cronin MTD, Enoch SJ, Hewitt M, Madden JC. Formation of Mechanistic Categories and Local Models to Facilitate the Prediction of Toxicity. *Altex-Altern Anim Ex*. 2011; 28(1):45-9.

38.   Jaworska J, Nikolova-Jeliazkova N. How can structural similarity analysis help in category formation? *SAR QSAR Environ Res*. 2007; 18(3-4):195-207.

39.   OECD. Guidance on grouping of chemicals, second edition. Available online at http://www.oecd.org/chemicalsafety/risk-assessment/chapter3dataevaluation.htm (accessed September 05, 2017). 2014.

40.   ECHA. Grouping of substances and read-across appraoch Part 1: Introductory note Available online at https://echa.europa.eu/support/registration/how-to-avoid-unnecessary-testing-on-animals/grouping-of-substances-and-read-across (accessed September 06, 2017). 2013.

41.   Koleva YK, Madden JC, Cronin MT. Formation of categories from structure-activity relationships to allow read-across for risk assessment: toxicity of alpha,beta-unsaturated carbonyl compounds. *Chem Res Toxicol*. 2008; 21(12):2300-12.

42.   Roberts DW, Schultz TW, Wolf EM, Aptula AO. Experimental Reactivity Parameters for Toxicity Modeling: Application to the Acute Aquatic Toxicity Of S(N)2 Electrophiles to *Tetrahymena pyriformis*. *Chem Res Toxicol*. 2010; 23(1):228-34.

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

44.   Roberts DW, Aptula AO, Patlewicz G. 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*. 2006; 19(9):1228-33.

45.   van Leeuwen K, Schultz TW, Henry T, Diderich B, Veith GD. Using chemical categories to fill data gaps in hazard assessment (vol 20, pg 207, 2009). *SAR QSAR Environ Res*. 2009; 20(5-6):591-.

46.   Nelms MD, Ates G, Madden JC, Vinken M, Cronin MTD, Rogiers V, et al. Proposal of an in silico profiler for categorisation of repeat dose toxicity data of hair dyes. *Arch Toxicol*. 2015; 89(5):733-41.

47.   Mellor CL, Steinmetz FP, Cronin MTD. Using Molecular Initiating Events to Develop a Structural Alert Based Screening Workflow for Nuclear Receptor Ligands Associated with Hepatic Steatosis. *Chem Res Toxicol*. 2016; 29(2):203-12.

48.   Enoch SJ, Madden JC, Cronin MTD. Identification of mechanisms of toxic action for skin sensitisation using a SMARTS pattern based approach. *SAR QSAR Environ Res*. 2008; 19(5-6):555-78.

49.   Enoch SJ, Roberts DW. Predicting Skin Sensitization Potency for Michael Acceptors in the LLNA Using Quantum Mechanics Calculations. *Chem Res Toxicol*. 2013; 26(5):767-74.

50.   Enoch SJ, Seed MJ, Roberts DW, Cronin MTD, Stocks SJ, Agius RM. Development of Mechanism-Based Structural Alerts for Respiratory Sensitization Hazard Identification. *Chem Res Toxicol*. 2012; 25(11):2490-8.

51.   Enoch SJ, Roberts DW, Cronin MTD. Mechanistic Category Formation for the Prediction of Respiratory Sensitization. *Chem Res Toxicol*. 2010; 23(10):1547-55.

52.   Mulliner D, Schuurmann G. Model Suite for Predicting the Aquatic Toxicity of alpha,beta-Unsaturated Esters Triggered by Their Chemoavailability. *Mol Inform*. 2013; 32(1):98-107.

53.   Hewitt M, Enoch SJ, Madden JC, Przybylak KR, Cronin MTD. Hepatotoxicity: A scheme for generating chemical categories for read-across, structural alerts and insights into mechanism(s) of action. *Crit Rev Toxicol*. 2013; 43(7):537-58.

54.   Schultz TW, Yarbrough JW, Pilkington TB. Aquatic toxicity and abiotic thiol reactivity of aliphatic isothiocyanates: Effects of alkyl-size and -shape. *Environ Toxicol Phar*. 2007; 23(1):10-7.

55.   Bajot F, Cronin MTD, Roberts DW, Schultz TW. Reactivity and aquatic toxicity of aromatic compounds transformable to quinone-type Michael acceptors. *SAR QSAR Environ Res*. 2011; 22(1-2):51-65.

56.   Bohme A, Thaens D, Paschke A, Schuurmann G. Kinetic glutathione chemoassay to quantify thiol reactivity of organic electrophiles--application to alpha,beta-unsaturated ketones, acrylates, and propiolates. *Chem Res Toxicol*. 2009; 22(4):742-50.

57.   Enoch SJ, Ellison CM, Schultz TW, Cronin MTD. A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity. *Crit Rev Toxicol*. 2011; 41(9):783-802.

58.   Enoch SJ, Cronin MTD. A review of the electrophilic reaction chemistry involved in covalent DNA binding. *Crit Rev Toxicol*. 2010; 40(8):728-48.

59.   Roberts DW, Aptula AO, Cronin MTD, Hulzebos E, Patlewicz G. Global (Q)SARs for skin sensitisation - assessment against OECD principles. *SAR QSAR Environ Res*. 2007; 18(3-4):343-65.

60.   Roberts DW, Patlewicz G, Kern PS, Gerberick F, Kimber I, Dearman RJ, et al. Mechanistic applicability domain classification of a local lymph node assay dataset for skin sensitization. *Chem Res Toxicol*. 2007; 20(7):1019-30.

61.   Schultz TW, Yarbrough JW, Hunter RS, Aptula AO. Verification of the structural alerts for Michael acceptors. *Chem Res Toxicol*. 2007; 20(9):1359-63.

62.   OECD. The OECD QSAR Toolbox for Grouping Chemicals into Gategories. Available online at https://www.qsartoolbox.org/ (accessed Novemeber 06, 2017). 2012.

63.   Enoch SJ, Roberts DW, Cronin MT. Mechanistic category formation for the prediction of respiratory sensitization. *Chem Res Toxicol*. 2010; 23(10):1547-55.

64.   Enoch SJ, Hewitt M, Cronin MTD, Azam S, Madden JC. Classification of chemicals according to mechanism of aquatic toxicity: An evaluation of the implementation of the Verhaar scheme in Toxtree. *Chemosphere*. 2008; 73(3):243-8.

65.   Roberts DW, Aptula AO, Patlewicz G. Electrophilic chemistry related to skin sensitization. Reaction mechanistic applicability domain classification for a published data set of 106 chemicals tested in the mouse local lymph node assay. *Chem Res Toxicol*. 2007; 20(1):44-60.

66.   Enoch SJ, Cronin MTD. Development of new structural alerts suitable for chemical category formation for assigning covalent and non-covalent mechanisms relevant to DNA binding. *Mutat Res-Gen Tox En*. 2012; 743(1-2):10-9.

67.   Verhaar HJM, Vanleeuwen CJ, Hermens JLM. Classifying Environmental-Pollutants .1. Structure-Activity-Relationships for Prediction of Aquatic Toxicity. *Chemosphere*. 1992; 25(4):471-91.

68.   Verhaar HJM, Ramos EU, Hermens JLM. Classifying environmental pollutants .2. Separation of class 1 (baseline toxicity) and class 2 ('polar narcosis') type compounds based on chemical descriptors. *J Chemometr*. 1996; 10(2):149-62.

69.   Bohme A, Thaens D, Schramm F, Paschke A, Schuurmann G. Thiol Reactivity and Its Impact on the Ciliate Toxicity of alpha,beta-Unsaturated Aldehydes, Ketones, and Esters. *Chem Res Toxicol*. 2010; 23(12):1905-12.

70.   Yarbrough JW, Schultz TW. Abiotic sulfhydryl reactivity: A predictor of aquatic toxicity for carbonyl-containing alpha,beta-unsaturated compounds. *Chem Res Toxicol*. 2007; 20(3):558-62.

71.   Cronin MTD, Aptula AO, Duffy JC, Netzeva TI, Rowe PH, Valkova IV, et al. Comparative assessment of methods to develop QSARs for the prediction of the toxicity of phenols to Tetrahymena pyriformis. *Chemosphere*. 2002; 49(10):1201-21.

72.   Chan K, Poon R, O'Brien PJ. Application of structure-activity relationships to investigate the molecular mechanisms of hepatocyte toxicity and electrophilic reactivity of alpha,beta-unsaturated aldehydes. *J Appl Toxicol*. 2008; 28(8):1027-39.

73.   Schöwbel JAH, Madden JC, Cronin MTD. Examination of Michael addition reactivity towards glutathione by transition-state calculations. *SAR QSAR Environ Res*. 2010; 21(7-8):693-710.

74.   Mulliner D, Wondrousch D, Schuurmann G. Predicting Michael-acceptor reactivity and toxicity through quantum chemical transition-state calculations. *Org Biomol Chem*. 2011; 9(24):8400-12.

75.   Flanagan ME, Abramite JA, Anderson DP, Aulabaugh A, Dahal UP, Gilbert AM, et al. Chemical and Computational Methods for the Characterization of Covalent Reactive Groups for the Prospective Design of Irreversible Inhibitors. *J Med Chem*. 2014; 57(23):10072-9.

76.   Promkatkaew M, Gleeson D, Hannongbua S, Gleeson MP. Skin Sensitization Prediction Using Quantum Chemical Calculations: A Theoretical Model for the SNAr Domain. *Chem Res Toxicol*. 2014; 27(1):51-60.

77.   Basketter DA, Evans P, Fielder RJ, Gerberick GF, Dearman RJ, Kimber I. Local lymph node assay - validation, conduct and use in practice. *Food Chem Toxicol*. 2002; 40(5):593-8.

78.   Basketter DA, Blaikie L, Dearman RJ, Kimber I, Ryan CA, Gerberick GF, et al. Use of the local lymph node assay for the estimation of relative contact allergenic potency. *Contact Dermatitis*. 2000; 42(6):344-8.

79.   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.

80.   Anderson SE, Siegel PD, Meade BJ. The LLNA: A Brief Review of Recent Advances and Limitations. *J Allergy (Cairo)*. 2011; 2011:424203.

81.   OECD, 2012. The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. Part 2: Use of the AOP to Develop Chemical Categories and Integrated Assessment and Testing Approaches. Series on Testing and Assessmentp.No. 168. ENV/JM/MONO(2012)/PART2.

82.   Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, et al. Adverse Outcome Pathways: A Conceptual Framework to Support Ecotoxicology Research and Risk Assessment. *Environ Toxicol Chem*. 2010; 29(3):730-41.

83.   Schultz TW. Tetratox: *Tetrahymena pyriformis* population growth impairment endpoint - A surrogate for fish lethality. *Toxicology Methods*. 1997; 7(4):289-309.

84.   Schultz TW, Yarbrough JW, Johnson EL. Structure-activity relationships for reactivity of carbonyl-containing compounds with glutathione. *SAR QSAR Environ Res*. 2005; 16(4):313-22.

85.   Dahal UP, Gilbert AM, Obach RS, Flanagan ME, Chen JM, Garcia-Irizarry C, et al. Intrinsic reactivity profile of electrophilic moieties to guide covalent drug design: N-alpha-acetyl-L-lysine as an amine nucleophile. *Medchemcomm*. 2016; 7(5):864-72..

86.   Gerberick GF, Vassallo JD, Bailey RE, Chaney JG, Morrall SW, Lepoittevin JP. Development of a peptide reactivity assay for screening contact allergens. *J. Toxicol Sci*. 2004;81(2):332-43.

87.   Orio M, Pantazis DA, Neese F. Density functional theory. *Photosynth Res*. 2009; 102(2-3):443-53.

88.   Cramer CJ. Essentials of Computational Chemistry Theories and Models. Second ed. West Sussex: Wiley; 2004.

89.   Jensen F. Introduction to Computational Chemistry. Second ed. West Sussex: Wiley; 2007.

90.   Leach AR. Molecular Modelling: Principles and Applications. Second ed. Essex: Longman; 2001.

91.   Kohn W, Sham LJ. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review*. 1965; 140:A 1133 - A 8.

92.   ElAzhary AA, Suter HU. Comparison between optimized geometries and vibrational frequencies calculated by the DFT methods. *J Phys Chem-Us*. 1996; 100(37):15056-63.

93.   Simon L, Goodman JM. How reliable are DFT transition structures? Comparison of GGA, hybrid-meta-GGA and meta-GGA functionals. *Org Biomol Chem*. 2011; 9(3):689-700.

94.   Natsch A, Emter R, Gfeller H, Haupt T, Ellis G. Predicted Skin Sensitizer Potency Based on *In Vitro* Data from KeratinoSens and Kinetic Peptide Binding: Global Versus Domain-Based Assessment. *J. Toxicol Sci*. 2015; 143(2):319-32.

95.   Enoch SJ, Cronin MTD, Schultz TW, Madden JC. Quantitative and mechanistic read across for predicting the skin sensitization potential of alkenes acting via Michael addition. *Chem Res Toxicol*. 2008; 21(2):513-20.

96.   Schwöbel JAH, Wondrousch D, Koleva YK, Madden JC, Cronin MTD, Schuurmann G. Prediction of Michael-Type Acceptor Reactivity toward Glutathione. *Chem Res Toxicol*. 2010; 23(10):1576-85.

97.   Enoch SJ, Cronin MTD, Schultz TW, Madden JC. An evaluation of global QSAR models for the prediction of the toxicity of phenols to Tetrahymena pyriformis. *Chemosphere*. 2008; 71(7):1225-32.

98.   Duchowicz PR, Mercader AG, Fernandez FM, Castro EA. Prediction of aqueous toxicity for heterogeneous phenol derivatives by QSAR. *Chemometr Intell Lab*. 2008; 90(2):97-107.

99.   Miller MD, Yourtee DM, Glaros AG, Chappelow CC, Eick JD, Holder AJ. Quantum mechanical structure-activity relationship analyses for skin sensitization. *J Chem Inf Model*. 2005; 45(4):924-9.

100.   Parr RG, Von Szentpaly L, Liu SB. Electrophilicity index. *J Am Chem Soc*. 1999; 121(9):1922-4.

101.   Wondrousch D, Bohme A, Thaens D, Ost N, Schuurmann G. Local Electrophilicity Predicts the Toxicity-Relevant Reactivity of Michael Acceptors. *J. Phys Chem Lett*. 2010; 1(10):1605-10.

102.   Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, et al. Gaussian 09, revision A.1. Wallingford, CT.; 2009.

103.   Ebbrell DJ, Madden JC, Cronin MT, Schultz TW, Enoch SJ. Development of a Fragment-Based *in Silico* Profiler for Michael Addition Thiol Reactivity. *Chem Res Toxicol*. 2016; 29(6):1073-81.

104.   US EPA. 2015. Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.11. United States Environmental Protection Agency, Washington, DC, USA.

105.   Molecular graphics and analyses were performed with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311). .

106.   Nelms MD, Cronin MTD, Schultz TW, Enoch SJ. Experimental verification, and domain definition, of structural alerts for protein binding: epoxides, lactones, nitroso, nitros, aldehydes and ketones. *SAR QSAR Environ Res*. 2013; 24(9):695-709.

107.   Ruusmann V, Maran U. 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*. 2013; 27(7):583-603.

108.   Gerberick GF, Ryan CA, Kern PS, Schlatter H, Dearman RJ, Kimber I, et al. Compilation of historical local lymph node data for evaluation of skin sensitization alternative methods. *Dermatitis*. 2005;16(4):157-202.

109.   Kern PS, Gerberick GF, Ryan CA, Kimber I, Aptula A, Basketter DA. Local Lymph Node Data for the Evaluation of Skin Sensitization Alternatives: A Second Compilation. *Dermatitis*. 2010; 21(1):8-32.

110.   Natsch A, Ryan CA, Foertsch L, Emter R, Jaworska J, Gerberick F, et al. A dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing prevalidation. *J Appl Toxicol.*. 2013; 33(11):1337-52.

111.   Dumont C, Barroso J, Matys I, Worth A, Casati S. 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*. 2016; 34:220-8.

112.   Barner-Kowollik C. Acrylate Free Radical Polymerization: From Mechanism to Polymer Design. *Macromol Rapid Commun*. 2009; 30(23):1961-3.

113.   Pham PD, Monge S, Lapinte V, Raoul Y, Robin JJ. Various radical polymerizations of glycerol-based monomers. *Eur J Lipid Sci Technol.*. 2013;115(1):28-40.

114.   Hubert AJ, Reimlinger H. The Isomerization of Olefins Part I. Based-Catalysed Isomerization of Olefins. *Synthesis*. 1969:97-112.

115.   Singh J, Petter RC, Baillie TA, Whitty A. The resurgence of covalent drugs. *Nat Rev Drug Discov*. 2011; 10(4):307-17.

Appendices

Appendix I. Tables containing chemical names and SMILES for all Tables in the thesis

Appendix II. Copies of published work

Appendix I. Tables containing chemical names and SMILES for all Tables in the thesis

Table 3.1a Chemical names and SMILES for chemicals in Table 3.1 and Table 3.2

ID	Chemical	SMILES
1	Prop-2-enal	C=CC=O
2	But-2-enal	CC=CC=O
3	Buten-2-one	C=CC(=O)C
4	Penten-2-one	CC=CC(=O)C
5	Methyl prop-2-enoate	C=CC(=O)OC
6	Methyl but-2-enoate	CC=CC(=O)OC

Table 4.1a Chemical Names and SMILES for chemicals in Table 4.1

ID	Chemical	SMILES
		Aldehyde
1	*trans*-Pent-2-enal	O=C/C=C/CC
2	*trans*-Oct-2-enal	O=C/C=C/CCCCC
3	*trans*-Non-2-enal	O=C/C=C/CCCCCC
4	*trans*-Hex-2-enal	O=C/C=C/CCC
5	*trans*-Prop-2-enal	O=CC=C

ID	Chemical	SMILES
6	*trans*-2-Methylbut-2-enal	O=C/C(C)=C/C
7	2-Methyl-pent-2-enal	O=CC(=CCC)C
8	4-Methyl-pent-2-enal	O=CC=CC(C)C
9	*trans*-But-2-enal	O=C/C=C/C
10	E-Dec-2-enal	O=C/C=C\CCCCCCC
11	*trans*-Dec-2-enal	O=C/C=C/CCCCCCC
12	*trans*-Cinnamaldehyde	O=C\C=C\c1ccccc1
13	α-Methyl-*trans*-cinnamaldehyde	C\C(C=O)=C/c1ccccc1
Ketone		
14	Methyl vinyl ketone	O=C(C=C)C
15	Hex-1-en-3-one	O=C(C=C)CC
16	Pent-1-en-3-one	O=C(C=C)CC
17	Pent-3-en-2-one	O=C(C=CC)C
18	Hept-3-en-2-one	O=C(C=CCC)C
19	Oct-3-en-2-one	O=C(C=CCCC)C
20	Non-3-en-2-one	O=C(C=CCCCC)C
21	Dec-3-en-2-one	O=C(C=CCCCCC)C
22	Hex-4-en-3-one	O=C(C=CC)CC
23	Oct-1-en-3-one	O=C(C=C)CCCCC

ID	Chemical	SMILES
24	3-Methyl-pent-3-en-2-one	O=C(C(=CC)C)C
25	5-Methyl-hept-2-en-4-one	CC(C)C(=O)C=CC
26	*trans*-Non-3-en-2-one	CC(=O)/C=C/CCCCC
27	4-Phenyl-but-3-en-2-one	CC(=O)\C=C\C1=CC=CC=C1
28	*trans*-Chalcone	O=C(\C=C\c1ccccc1)c1ccccc1
29	2-Hydroxychalcone	Oc1ccccc1\C=C\C(=O)c1ccccc1
30	4-Hydroxychalcone	Oc1ccc(\C=C\C(=O)c2ccccc2)cc1
	Esters	
31	Methyl crotonate	CC(C)COC(=O)C=C
32	Ethyl acrylate	CCCCCCOC(=O)C=C
33	Methyl acrylate	CCCCOC(=O)C=C
34	Methyl methacrylate	COC(=O)\C=C\C
35	*t*-butyl acrylate	CCOC(=O)C=C
36	Propyl acrylate	COC(=O)C=C
37	2-Hydroxy ethyl acrylate	COC(=O)C(C)=C
38	2-Hydroxyethyl methacrylate	CC(C)(C)OC(=O)C=C
39	2-Hydroxypropyl methacrylate	CCCOC(=O)C=C
40	Phenyl acrylate	OCCOC(=O)C=C
41	Isoamyl acrylate	CC(=C)C(=O)OCCO

ID	Chemical	SMILES
42	N-pentylacrylate	CC(O)COC(=O)C(C)=C
43	Ethyl crotonate	CC(=C)C(=O)Oc1ccccc1
44	Methyl *trans*-pent-2-enoate	CC(C)CCOC(=O)C=C
45	Ethyl *trans*-hex-2-enoate	CCCCCOC(=O)C=C
46	Methyl-hex-2-enoate	CCOC(=O)\C=C\C
47	Methyl-4-methyl-pent-2-enoate	CC\C=C\C(=O)OC
48	Ethyl tiglate	CCC\C=C\C(=O)OCC
49	Ethyl methacrylate	CCC\C=C\C(=O)OC
50	Butyl methacrylate	COC(=O)\C=C\C(C)C
51	2-Ethylhexyl acrylate	CCOC(=O)C(\C)=C\C
52	Methyl crotonate	CCOC(=O)C(C)=C
53	Ethyl acrylate	CCCCOC(=O)C(C)=C
54	Methyl acrylate	CCCCC(CC)COC(=O)C=C
Nitro		
55	1-Nitro-1-cyclohexene	C1CCCC=C1N(=O)(=O)
56	4-Methyl-β-nitrostyrene (mixture of *cis* and *trans*)	N(=O)(=O)C=Cc1ccc(C)cc1
57	*trans*-β-Nitrostyrene	c1ccccc1/C=C/N(=O)=O
58	*trans*-4-Methyl-β-nitrostyrene	O=N(=O)/C=C/c1ccc(C)cc1

ID	Chemical	SMILES
59	*trans*-4-Chloro-β-nitrostyrene	c1cc(Cl)ccc1/C=C/N(=O)(=O)
60	*trans*-4-Bromo-β-nitrostyrene	O=N(=O)/C=C/c1ccc(Br)cc1
61	4-Fluoro-β-nitrostyrene	O=N(=O)C=Cc1ccc(F)cc1
62	*trans*-4-Methoxy-β-nitrostyrene	O=N(=O)/C=C/c1ccc(OC)cc1
63	*trans*-β-Methyl-β-nitrostyrene	c1ccccc1/C=C/(C)N(=O)(=O)
Nitrile		
64	2-Methyleneglutaronitrile	N#CC(=C)CCC#N
65	Cyclohexene-1-carbonitrile (1-cyanocyclohexene)	C1(C#N)=CCCCC1
66	1-Cyclopentene-1-carbonitrile	N#CC1=CCCC1
Cyclic Ketones		
67	2-Cyclohexen-1-one	O=C1CCCC=C1
68	2-Cyclopenten-1-one	O=C1CCC=C1
69	2-Methyl-2-cyclopenten-1-one	CC1=CCCC1=O
70	4,4-Dimethyl-2-cyclohexen-1-one	CC1(C)CCC(=O)C=C1
71	1-Acetyl-1-cyclohexene	CC(=O)C1=CCCCC1
72	1-Acetyl-1-cyclopentene	CC(=O)C1=CCCC1

Table 4.3a Chemical names and SMILES for chemicals in Table 4.3

ID	Chemical	SMILES
1	Methyl crotonate	C/C=C/C(=O)OC
2	Ethyl crotonate	C/C=C/C(=O)OCC
3	Methyl *trans*-pent-2-enoate	CC/C=C/C(=O)OC
4	Ethyl *trans*-hex-2-enoate	CCC/C=C/C(=O)OCC
5	Methyl-hex-2-enoate	CCC\C=C\C(=O)OCC
6	Methyl-4-methyl-pent-2-enoate	COC(=O)\C=C\C(C)C
7	Ethyl tiglate	CCOC(=O)C(\C)=C\C

Table 4.4a Chemical names and SMILES for chemicals in Table 4.4

ID	Chemical	SMILES
1	Chalcone	O=C(\C=C\c1ccccc1)c1ccccc1
2	2-Hydroxy-chalcone	Oc1ccccc1\C=C\C(=O)c1ccccc1
3	4-Hydroxy-chalcone	Oc1ccc(\C=C\C(=O)c2ccccc2)cc1
4	4-Phenyl-but-3-en-2-one	CC(=O)\C=C\C1=CC=CC=C1
5	Phenyl-acrylate	CC(=C)C(=O)Oc1ccccc1

Table 4.5a Chemical names and SMILES for chemicals in Table 4.5

ID	Chemical	SMILES
1	1-Nitro-1-cyclohexene	C1CCCC=C1N(=O)(=O)
2	4-Methyl-β-nitrostyrene	N(=O)(=O)C=Cc1ccc(C)cc1
3	*Trans*-4-methyl-β-nitrostyrene	O=N(=O)/C=C/c1ccc(C)cc1
4	*Trans*-4-chloro-β-nitrostyrene	c1cc(Cl)ccc1/C=C/N(=O)(=O)
5	*Trans*-4-bromo-β-nitrostyrene	O=N(=O)/C=C/c1ccc(Br)cc1
6	*Trans*-4-fluoro-β-nitrostyrene	O=N(=O)C=Cc1ccc(F)cc1
7	*Trans*-4-methoxy-β-nitrostyrene	O=N(=O)/C=C/c1ccc(OC)cc1
8	*Trans*-β-methyl-β-nitrostyrene	c1ccccc1/C=C/(C)N(=O)(=O)
9	2-Methyleneglutaronitrile	N#CC(=C)CCC#N
10	Cyclohexene-1-carbonitrile	C1(C#N)=CCCCC1
11	1-Cyclopentene-1-carbonitrile	N#CC1=CCCC1
12	Cyclohex-2-en-1-one	O=C1CCCC=C1
13	Cyclopent-2-en-1-one	O=C1CCC=C1
14	2-Methyl-cyclopent-2-en-1-one	CC1=CCCC1=O
15	1-Acetyl-cyclohex-1-ene	CC(=O)C1=CCCCC1
16	1-Acetyl-cyclopent-1-ene	CC(=O)C1=CCCC1
17	4,4-Dimethyl-cyclohex-2-en-1-one	CC1(C)CCC(=O)C=C1

Table 4.6a Chemical names and SMILES for chemicals in Table 4.6

ID	Chemical	SMILES
1	Prop-2-enal	C=CC=O
2	(2E)-But-2-enal	C\C=C\C=O
3	(2E)-3-(Furan-2-yl)prop-2-enal	O=C\C=C\c1ccco1
4	(2E)-Pent-2-enal	CC\C=C\C=O
5	4-Methylpent-2-enal	CC(C)\C=C\C=O
6	Hex-2-enal	CCC\C=C\C=O
7	(2E)-3-Phenylprop-2-enal	O=C\C=C\c1ccccc1
8	(2E)-3-[4-(Dimethylamino)phenyl]prop-2-enal	CN(C)c1ccc(\C=C\C=O)cc1
9	Hept-2-enal	CCCC\C=C\C=O
10	(2E)-Oct-2-enal	CCCCC\C=C\C=O
11	(2E)-2-Methylbut-2-enal	C\C=C(/C)C=O
12	Non-2-enal	CCCCCC\C=C\C=O
13	2-Methylpent-2-enal	CC\C=C(/C)C=O
14	But-3-en-2-one	CC(=O)C=C
15	Pent-1-en-3-one	CCC(=O)C=C
16	Hex-1-en-3-one	CCCC(=O)C=C

ID	Chemical	SMILES
17	Pent-3-en-2-one	C\C=C\C(C)=O
18	Hex-4-en-3-one	CCC(=O)\C=C\C
19	Oct-1-en-3-one	CCCCCC(=O)C=C
20	Hept-3-en-2-one	CCC\C=C\C(C)=O
21	Oct-3-en-2-one	CCCC\C=C\C(C)=O
22	Oct-2-en-4-one	CCCCC(=O)\C=C\C
23	2-Methylcyclopent-2-en-1-one	CC1=CCCC1=O
24	3-Methylpent-3-en-2-one	C\C=C(/C)C(C)=O
25	Non-3-en-2-one	CCCCC\C=C\C(C)=O
26	2-Hydroxyethyl prop-2-enoate	OCCOC(=O)C=C
27	2-Hydroxypropyl prop-2-enoate	CC(O)COC(=O)C=C
28	Methyl prop-2-enoate	COC(=O)C=C
29	Ethyl prop-2-enoate	CCOC(=O)C=C
30	Propyl prop-2-enoate	CCCOC(=O)C=C
31	2-Methylpropyl prop-2-enoate	CC(C)COC(=O)C=C
32	2-Hydroxyethyl 2-methylprop-2-enoate	CC(=C)C(=O)OCCO
33	Butyl prop-2-enoate	CCCCOC(=O)C=C

ID	Chemical	SMILES
34	Benzyl prop-2-enoate	C=CC(=O)OCc1ccccc1
35	3-Methylbutyl prop-2-enoate	CC(C)CCOC(=O)C=C
36	Pentyl prop-2-enoate	CCCCCOC(=O)C=C
37	Cyclohexyl prop-2-enoate	C=CC(=O)OC1CCCCC1
38	Methyl 2-methylprop-2-enoate	COC(=O)C(C)=C
39	Hexyl prop-2-enoate	CCCCCCOC(=O)C=C
40	2-Methylpropyl (2E)-but-2-enoate	C\C=C\C(=O)OCC(C)C
41	Butan-2-yl (2E)-but-2-enoate	CCC(C)OC(=O)\C=C\C
42	Butyl (2E)-but-2-enoate	CCCCOC(=O)\C=C\C
43	2-Ethoxyethyl 2-methylprop-2-enoate	CCOCCOC(=O)C(C)=C
44	(2E)-Dec-2-enal	CCCCC\C=C\C=O
45	Heptyl prop-2-enoate	CCCCCCCOC(=O)C=C
46	Ethyl 2-methylprop-2-enoate	CCOC(=O)C(C)=C
47	Methyl (2E)-oct-2-enoate	CCCCC\C=C\C(=O)OC
48	Methyl (2E)-3-phenylprop-2-enoate	COC(=O)\C=C\c1ccccc1
49	Methyl (2E)-2-methylbut-2-enoate	COC(=O)C(\C)=C\C
50	Propan-2-yl 2-methylprop-2-enoate	CC(C)OC(=O)C(C)=C

ID	Chemical	SMILES
51	Propyl 2-methylprop-2-enoate	CCCOC(=O)C(C)=C
52	Methyl non-2-enoate	CCCCCC\C=C\C(=O)OC
53	Ethyl (2E)-3-phenylprop-2-enoate	CCOC(=O)\C=C\c1ccccc1
54	Ethyl (2E)-2-methylbut-2-enoate	CCOC(=O)C(\C)=C\C
55	Methyl (2E)-2-methylpent-2-enoate	CC\C=C(/C)C(=O)OC
56	2-Methylpropyl 2-methylprop-2-enoate	CC(C)COC(=O)C(C)=C
57	Butyl 2-methylprop-2-enoate	CCCCOC(=O)C=C
58	Propyl (2E)-3-phenylprop-2-enoate	CCCOC(=O)\C=C\c1ccccc1
59	Benzyl 2-methylprop-2-enoate	CC(=C)C(=O)OCc1ccccc1
60	Butyl (2E)-3-phenylprop-2-enoate	CCCCOC(=O)\C=C\c1ccccc1
61	Hexyl 2-methylprop-2-enoate	CCCCCCOC(=O)C(C)=C
62	2-Ethylhexyl 2-methylprop-2-enoate	CCCCC(CC)COC(=O)C(C)=C

Table 4.7a Chemical names and SMILES for chemicals in Table 4.7

ID	Chemical	SMILES
1	Methyl methacrylate	CC(=C)C(=O)OC

ID	Chemical	SMILES
2	2-Hydroxypropyl methacrylate	CC(COC(=O)C(=C)C)O
3	Ethyl acrylate	CCOC(=O)C=C
4	Methyl acrylate	COC(=O)C=C
5	Butyl acrylate	CCCCOC(=O)C=C
6	r-Carvone	CC(=C)C1CC=C(C)C(=O)C1
7	L-Carvone	CC1=CC[C@H](CC1=O)C(=C)C
8	α-Butyl cinnamic aldehyde	CCCC\C(C=O)=C/c1ccccc1
9	Linalool aldehyde	C\C(C=O)=C/CCC(C)(O)C=C
10	*trans*-Hex-2-enal	CCC\C=C\C=O
11	α-Amyl cinnamic aldehyde	CCCCC/C(=C\c1ccccc1)/C=O
12	α-Hexylcinnamaldehyde	CCCCCC\C(C=O)=C/c1ccccc1
13	2-Ethylhexyl-acrylate	CCCCC(CC)COC(=O)C=C
14	Perillaldehyde	CC(=C)C1CCC(C=O)=CC1
15	1-(p-Methoxyphenyl)-1-penten-3-one	CCC(=O)\C=C\c1ccc(OC)cc1
16	α-Methyl-cinnamic aldehyde	C\C(C=O)=C/c1ccccc1
17	Benzylidene acetone	CC(=O)\C=C\c1ccccc1
18	5-Methyl-2-phenyl-hex-2-enal	CCCC\C=C(\C=O)c1ccccc1
19	Cinnamic aldehyde	O=C\C=C\c1ccccc1

ID	Chemical	SMILES
20	*trans*-Dec-2-enal	CCCCCCC/C=C/C=O
21	Galbanone	CC1(C)CCC=C(C1)C(=O)CCC=C
22	5,5-Dimethyl-3-methylene-dihydro-2(3H)-furone	CC1(C)CC(=C)C(=O)O1
23	Diethyl maleate	CCOC(=O)/C=C/C(=O)OCC
24	2-Hydroxyethyl acrylate	C=CC(=O)OCCO
25	Spirogalbanone	C=CCCC(=O)C1=CCCC2(CCCC2)C1
26	Pomarose	C\C=C\C(=O)C(\C)=C(/C)C(C)C

Table 5.1a Chemical names and SMILES for chemicals in Table 5.1

ID	Chemical	SMILES
1	1-Bromopropan-2-one	CC(=O)CBr
2	2-Bromo-1-phenylethan-1-one	BrCC(=O)c1ccccc1
3	Methyl-2-bromoacetate	COC(=O)CBr
4	2-Bromoacetic acid	OC(=O)CBr
5	2-Bromo acetamide	NC(=O)CBr

Table 5.7a Chemical names and SMILES for chemicals in Table 5.7

ID	Chemical	SMILES
1	1-Bromomethyl-4-nitrobenzene	[O-][N+](=O)c1ccc(CBr)cc1
2	2-(2-Bromoacetyl) thiophene	BrCC(=O)c1cccs1
3	1-Bromo-2-butanone	CCC(=O)CBr
4	1-Bromopinacolone	CC(C)(C)C(=O)CBr
5	1-Chloropinacolone	CC(C)(C)C(=O)CCl
6	2-Bromoacetophenone	BrCC(=O)c1ccccc1
7	2-(2-Bromoacetyl) naphthalene	BrCC(=O)c1ccc2ccccc2c1
8	1-(Bromoacetyl) pyrene	BrCC(=O)c1ccc2ccc3cccc4ccc1c2c34
9	Ethyl bromoacetate	CCOC(=O)CBr
10	Ethyl chloroacetate	CCOC(=O)CCl
11	Methyl bromoacetate	COC(=O)CBr
12	Methyl chloroacetate	COC(=O)CCl
13	Propyl bromoacetate	CCCOC(=O)CBr
14	Propyl chloroacetate	CCCOC(=O)CCl
15	*t*-Butyl bromoacetate	CC(C)(C)OC(=O)CBr
16	*t*-Butyl chloroacetate	CC(C)(C)OC(=O)CCl
17	Phenyl bromoacetate	BrCC(=O)Oc1ccccc1

ID	Chemical	SMILES
18	2-Bromoacetamide	NC(=O)CBr
19	2-Chloroacetamide	NC(=O)CCl
20	Methyl-2-bromopropionate	COC(=O)C(C)Br
21	Methyl-2-bromobutyrate	CCC(Br)C(=O)OC
22	2-Bromopropionamide	CC(Br)C(N)=O
23	2-Bromobutyric acid	CCC(Br)C(O)=O
24	2-Chlorobutyric acid	CCC(Cl)C(O)=O
25	2-Bromovaleric acid	CCCC(Br)C(O)=O
26	Ethyl-2-bromovalerate	CCCC(Br)C(=O)OCC
27	Ethyl-2-bromobutyrate	CCOC(=O)C(Br)CC
28	Ethyl-2-bromopropionate	CCOC(=O)C(C)Br

Appendix II. Copies of published work

Contributions of authors to papers:

Development of a fragment-based *in silico* profiler for Michael addition thiol reactivity

D. J. EBBRELL , J. C. MADDEN , M. T. D. CRONIN , T. W. SCHULTZ AND S. J ENOCH

D. J. Ebbrell carried out the bulk of the analysis and generated the first draft of the paper. T. W. Schultz kindly donated experimental reactivity data used in the paper. S. J. Enoch provided guidance on the contents of the paper and corrections for multiple drafts. J. C. Madden and M. T. D. Cronin provided additional corrections for the final draft prior to submission.

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

D. J. EBBRELL , J. C. MADDEN , M. T. D. CRONIN , T. W. SCHULTZ AND S. J ENOCH

D. J. Ebbrell carried out the bulk of the analysis and generated the first draft of the paper. T. W. Schultz provided his knowledge regarding the experimental data used. S. J. Enoch provided guidance on the contents of the paper and corrections for multiple drafts. J. C. Madden and M. T. D. Cronin provided additional corrections for the final draft prior to submission.

Development of a fragment-based *in silico* profiler for S_N2 thiol reactivity and its application in predicting toxicity of chemicals towards *Tetrahymena pyriformis*

*Yet to be published

D. J. Ebbrell , J. C. Madden , M. T. D. Cronin , T. W. Schultz and S. J Enoch

D. J. Ebbrell carried out the bulk of the analysis and generated the first draft of the paper. T. W. Schultz provided his knowledge regarding the experimental data used. S. J. Enoch provided guidance on the contents of the paper and corrections for multiple drafts. J. C. Madden and M. T. D. Cronin provided additional corrections for the final draft prior to submission.

Development of a Fragment-Based *in Silico* Profiler for Michael Addition Thiol Reactivity

David J. Ebbrell,[†] Judith C. Madden,[†] Mark T. D. Cronin,[†] Terry W. Schultz,[‡] and Steven 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 37996, United States

S *Supporting Information*
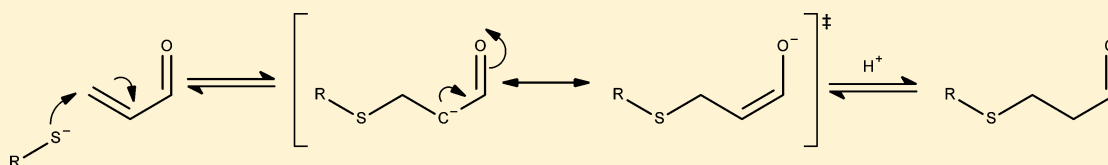
ABSTRACT: The Adverse Outcome Pathway (AOP) paradigm details the existing knowledge that links the initial interaction between a chemical and a biological system, termed the molecular initiating event (MIE), through a series of intermediate events, to an adverse effect. An important example of a well-defined MIE is the formation of a covalent bond between a biological nucleophile and an electrophilic compound. This particular MIE has been associated with various toxicological end points such as acute aquatic toxicity, skin sensitization, and respiratory sensitization. This study has investigated the calculated parameters that are required to predict the rate of chemical bond formation (reactivity) of a dataset of Michael acceptors. Reactivity of these compounds toward glutathione was predicted using a combination of a calculated activation energy value (E_{act}, calculated using density functional theory (DFT) calculation at the B3YLP/6-31G+(d) level of theory, and solvent-accessible surface area values (SAS) at the α carbon. To further develop the method, a fragment-based algorithm was developed enabling the reactivity to be predicted for Michael acceptors without the need to perform the time-consuming DFT calculations. Results showed the developed fragment method was successful in predicting the reactivity of the Michael acceptors excluding two sets of chemicals: volatile esters with an extended substituent at the β-carbon and chemicals containing a conjugated benzene ring as part of the polarizing group. Additionally the study also demonstrated the ease with which the approach can be extended to other chemical classes by the calculation of additional fragments and their associated E_{act} and SAS values. The resulting method is likely to be of use in regulatory toxicology tools where an understanding of covalent bond formation as a potential MIE is important within the AOP paradigm.

■ INTRODUCTION

The Adverse Outcome Pathway (AOP) paradigm has been promoted as a key approach that may enable the demands of the seventh amendment to the cosmetic directive and REACH to be met.[1] An AOP details the existing knowledge that links the initial interaction between a chemical and a biological system, through a series of intermediate events, to an adverse effect.[2] Clearly, biological pathways, the perturbation of which, can lead to an adverse effect, are diverse and complex. Thus, the AOP concept is concerned with defining only the key, testable events in a given pathway. Consequently, there are significant efforts to develop *in silico*, *in chemico*, and *in vitro* methods that enable such key events to be predicted and/or tested. The ultimate aim is that a series of alternative tests (developed from the knowledge of an AOP) will enable an animal test for a regulatory end point to be replaced. For example, the recently defined AOP for skin sensitization has led to the development of a number of non-animal testing methods which may be used (in combination) to replace *in vivo* studies.[3] Within the AOP approach, *in silico* methods are typically used to define the chemistry associated with the initial chemical interaction between a chemical and the biological system, termed the molecular initiating event (MIE).

An important example of a well-defined MIE is the formation of a covalent bond between a biological nucleophile, such as the thiol group of cysteine or the amine group of lysine, and an electrophilic chemical such as acrolein.[4] This particular MIE has been associated with various adverse outcomes such as skin sensitization, respiratory sensitization, acute aquatic toxicity, liver toxicity, chromosomal aberration, and a wide range of idiosyncratic drug toxicities.[5−10] Given the importance of covalent bond formation as an MIE, various *in chemico* assays have been used to investigate the potential correlation between rate of covalent bond formation (reactivity) of chemicals and their ability to elicit a toxicological effect.[11] There are a number of reactive mechanisms by which an electrophilic chemical may react with a biological nucleophile. An important and well-studied mechanism is Michael addition. For a chemical to act via Michael addition it must have an electron-withdrawing group adjacent to a carbon–carbon double bond; this results in an electron-deficient carbon at the β-position. This allows for nucleophilic attack such as a thiolate nucleophile at the electron-deficient β-position
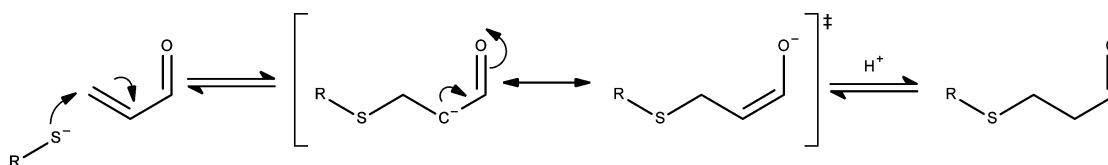
Figure 1. Proposed mechanism and transition state of acrolein (an electrophile) and a thiol nucleophile (R = glutathione, alkyl).

resulting in the formation of a resonance-stabilized carbanion at the α-position, the carbanion is then protonated to produce the final product, a Michael adduct (Figure 1).[12] When considering Michael addition thiol reactivity, there are three important factors: the impact of the electron-withdrawing group, substitution at the α-position (where the inductive effect of the substituent can stabilize/destabilize the negative charge at this position), and substitution at the β-position of the carbon−carbon double bond.

There have been many attempts to relate predict the reactivity and toxicity of chemicals known to act via Michael addition both experimentally (*in chemico*) and computationally (*in silico*).[13−24] *In chemico* approaches involve either the determination of the kinetic rate constant or, more typically, spectrophotometric methods that involve determination of the concentration of the electrophile required to deplete a model nucleophile such as glutathione.[18] In contrast, *in silico* methods, such as the works of Mulliner et al. and Schwobel et al., use quantum mechanical methods to calculate the energy of activation for these types of electrophilic reactions, enabling the experimental rate values to be predicted using simple quantitative structure activity models.[19,22] Furthermore, such *in silico* methods have been applied for the prediction of toxicity data where covalent protein binding is the MIE.

It is clear from the literature that *in silico* methods involving the calculation of the activation energy are capable of predicting both chemical reactivity and, in turn, toxicity. However, these approaches require the use of time-consuming quantum chemical calculations which require proprietary software. This limits their use and inclusion in freely available *in silico* tools currently finding widespread use in regulatory toxicology (for example, the OECD QSAR Toolbox). Therefore, the aim of this study is to develop an *in silico* profiler capable of predicting chemical reactivity for Michael acceptors. This approach is based on a fragment method in which a database of pre-calculated energy of activation values are used within the *in silico* profiler, thus removing the need for the end-user to perform such calculations.

METHODS

Dataset. The RC_{50} values for various Michael acceptors were determined using a previously published spectrophotometric peptide depletion assay.[25] Where RC_{50} is the concentration of electrophile required to deplete the concentration of glutathione (GSH) by 50% in 120 minutes. Average RC_{50} values were calculated for chemicals which had multiple experimental values. RC_{50} values for poorly soluble chemicals were determined by the addition of 50% MeOH. A structurally diverse set of experimental data was profiled using previously published alerts for polarized aldehydes, ketones, esters, nitros, nitriles, and cyclic ketones.[12] This resulted in a subsequent dataset of 72 chemicals covering 13 aldehydes, 17 ketones, 24 esters, 9 nitro compounds, 3 nitrile-containing compounds, and 6 cyclic ketones (Table 1). Additionally, individual standard deviation values are stated; these values result in an average experimental error of 0.13 log unit.

Computational Methods. All calculations were carried out using the Gaussian 09 suite of software using density functional theory (DFT) utilizing the B3LYP/6-31G+(d) level of theory.[26] Energies of activation (E_{act}) values for transition-state structures were calculated using thiolate as a model nucleophile. The use of a thiolate (rather than a thiol) nucleophile allows an intermediate to be isolated on the potential energy

surface. This significantly simplifies the calculations, as the intermediate can be isolated using a simple energy minimization calculation rather than a transition-state calculation. The solvent-accessible surface area (SAS) at the α-position was calculated for each chemical using the Chimera software.[27] The *in silico* profiler was encoded as a workflow using the open source KNIME environment. All experimental and calculated data are available in the Supporting Information, including the fragment which is used for each chemical in Table 1, calculated E_{act} (kcal/mol), SAS values, and predicted $-\log RC_{50}$ values for each model.

Statistical Analysis. Linear regression analysis was used to develop quantitative structure−activity relationship (QSAR) models to obtain correlations between $-\log RC_{50}$ values and the calculated descriptors (E_{act} and SAS values) using the Minitab (version 17) statistical software.

RESULTS AND DISCUSSION

The initial aim of this study was to develop a fragment-based *in silico* profiler capable of predicting chemical reactivity for polarized alkenes (aldehydes, ketones, and esters, chemicals 1−54 in Table 1). This was achieved by systematically varying a series of alkyl and aryl substituents at each of the R groups (as shown in Figure 2) in order to establish the point at which increasing the alkyl chain size failed to increase the activation energy by more than 1 kcal/mol (all analysis carried out by rounding the energy difference to the nearest kcal/mol). For example, examining how the calculated activation energy changes when varying the substituents at position R_1 for a series of aldehydes ($R_2 = R_3 =$ hydrogen) shows that on going from methyl to ethyl the activation energy increases by 4.2 kcal/mol. In contrast, extending the alkyl chain further from ethyl to propyl, decreases the activation energy by 0.2 kcal/mol (Table 2). This change is significantly less than the cutoff value of 1 kcal/mol (or less), meaning that all alkyl chains of two carbons or more can be reasonably predicted using the calculated activation energy value of the ethyl group. This analysis enables two fragments to be defined that can be used to calculate the activation energy of chemicals with simple alkyl chains at this position ($R =$ Me and Et). The analysis also showed the need to include isopropyl and *tert*-butyl groups due to their increased steric hindrance. An analogous analysis was carried out into the effect of alkyl chain length on the polarized aldehydes at position R_2 (Table 2).

The effect of a benzene ring on the calculated activation energy for the polarized aldehydes was also investigated at positions R_1 and R_2. Taking the effect at R_1 as an example, the results showed that the activation energy increases significantly on going from $R_1 =$ Me to Ph (-1.5 to 3.4 kcal/mol). As expected, the results also showed that increasing the number of CH_2 groups between the alkene and the benzene ring caused a decrease in the associated activation energy (compare $R_1 = C_6H_5$ to $CH_2C_6H_5$). In terms of defining fragments for the effect of a benzene ring at this position, it is useful to compare the aryl substituent with the corresponding alkyl substituent. For example, comparing the activation energy values of $R_1 = CH_2C_6H_5$ to $R_1 = CH_3$ shows there to be an energy difference of 2.7 kcal/mol, which when rounded to the nearest kcal/mol is significantly in excess of the 1 kcal/mol (or less) cutoff. In contrast, comparing $R_1 = CH_2CH_2C_6H_5$ to $R_1 = CH_2CH_3$ shows there to be an energy difference of 1.1 kcal/mol

Table 1. Michael Acceptors with Corresponding −Log RC$_{50}$ Values Investigated in the Current Study[a]

ID	chemical	SMILES	RC$_{50}$ average	−log RC$_{50}$ average (mM)
		Aldehydes		
1	*trans*-2-pentenal	O=C/C=C/CC	0.33 ± 0.02	0.48
2	*trans*-2-octenal	O=C/C=C/CCCCC	0.28 ± 0.02	0.56
3	*trans*-2-nonenal	O=C/C=C/CCCCCC	0.41 ± 0.05	0.39
4	*trans*-2-hexenal	O=C/C=C/CCC	0.43 ± 0.11	0.37
5	acrolein	O=CC=C	0.07 ± 0.03	1.14
6	*trans*-2-methyl-2-butenal	O=C/C(C)=C/C	11.71 ± 1.88	−1.07
7	2-methyl-2-pentenal	O=CC(=CCC)C	20.74 ± 1.21	−1.32
8	4-methyl-2-pentenal	O=CC=CC(C)C	1.15 ± 0.15	−0.06
9	*trans*-2-butenal	O=C/C=C/C	0.21 ± 0.02	0.67
10	(*E*)-2-decen-1-al	O=C/C=C\CCCCCC	0.21 ± 0.05	0.67
11	*trans*-2-decenal	O=C/C=C/CCCCCC	0.17 ± 0.02	0.77
12	*trans*-cinnamaldehyde	O=C\C=C\C1=CC=CC=C1	0.96 ± 0.24	0.02
13	α-methyl-*trans*-cinnamaldehyde[b]	C\C(C=O)=C/C1=CC=CC=C1	21.60 ± 7.04	−1.33
		Ketones		
14	methyl vinyl ketone	O=C(C=C)C	0.06 ± 0.03	1.23
15	1-hexen-3-one	O=C(C=C)CCC	0.06 ± 0.00	1.23
16	1-penten-3-one	O=C(C=C)CC	0.05 ± 0.00	1.29
17	3-penten-2-one	O==C(C=CC)C	0.15 ± 0.09	0.83
18	3-hepten-2-one	O=C(C = CCCC)C	0.67 ± 0.11	0.17
19	3-octen-2-one	O=C(C=CCCCC)C	0.57 ± 0.16	0.24
20	3-nonen-2-one	O=C(C=CCCCCC)C	0.54 ± 0.11	0.27
21	3-decen-2-one	O=C(C=CCCCCCC)C	0.58 ± 0.16	0.24
22	4-hexen-3-one	O=C(C=CC)CC	0.34 ± 0.05	0.46
23	1-octen-3-one	O=C(C=C)CCCCC	0.02 ± 0.01	1.78
24	3-methyl-3-penten-2-one	O=C(C(=CC)C)C	9.77 ± 1.23	−0.99
25	5-methyl-2-hepten-4-one	CC(C)C(=O)C=CC	0.37 ± 0.02	0.44
26	*trans*-3-nonen-2-one	CC(=O)/C=C/CCCC	0.60 ± 0.03	0.22
27	4-phenyl-3-buten-2-one	CC(=O)\C=C\C1=CC=CC=C1	3.53 ± 0.07	−0.55
28	*trans*-chalcone[b]	O=C(\C=C\C1=CC=CC=C1)C1=CC=CC=C1	0.40 ± 0.08	0.40
29	2-hydroxychalcone	OC1=CC=CC=C1\C=C\C(=O)C1=CC=CC=C1	0.28 ± 0.08	0.83
30	4-hydroxychalcone	OC1=CC=C(\C=C\C(=O)C2=CC=CC=C2)C=C1	0.41 ± 0.29	0.39
		Esters		
31	isobutyl acrylate	CC(C)COC(=O)C=C	0.48 ± 0.06	0.32
32	*n*-hexyl acrylate	CCCCCCOC(=O)C=C	0.82 ± 0.08	0.09
33	butyl acrylate	CCCCOC(=O)C=C	0.77 ± 0.02	0.11
34	methyl crotonate	COC(=O)\C=C\C	21.25 ± 4.95	−1.33
35	ethyl acrylate	CCOC(=O)C=C	0.52 ± 0.05	0.29
36	methyl acrylate	COC(=O)C=C	0.49 ± 0.10	0.31
37	methyl methacrylate	COC(=O)C(C)=C	69.19 ± 7.12	−1.84
38	*tert*-butyl acrylate	CC(C)(C)OC(=O)C=C	1.28 ± 0.030	−0.11
39	propyl acrylate	CCCOC(=O)C=C	0.85 ± 0.08	0.07
40	2-hydroxyethyl acrylate	OCCOC(=O)C=C	0.27 ± 0.03	0.57
41	2-hydroxyethyl methacrylate	CC(=C)C(=O)OCCO	33.40 ± 1.33	−1.52
42	2-hydroxypropyl methacrylate	CC(O)COC(=O)C(C)=C	21.15 ± 9.20	−1.33
43	phenyl acrylate	CC(=C)C(=O)OC1=CC=CC=C1	0.02 ± 0.01	1.64
44	isoamyl acrylate	CC(C)CCOC(=O)C=C	0.68 ± 0.22	0.17
45	*n*-pentyl acrylate	CCCCCOC(=O)C=C	0.81 ± 0.02	0.09
46	ethyl crotonate	CCOC(=O)\C=C\C	17.95 ± 0.78	−1.25
47	methyl *trans*-2-pentenoate	CC\C=C\C(=O)OC	5.05 ± 0.42	−0.70
48	ethyl *trans*-2-hexenoate	CCC\C=C\C(=O)OCC	0.76 ± 0.10	0.12
49	methyl 2-hexenoate	CCC\C=C\C(=O)OC	2.46 ± 1.37	−0.39
50	methyl 4-methyl-2-pentenoate	COC(=O)\C=C\C(C)C	1.28 ± 0.25	−0.11
51	ethyl tiglate	CCOC(=O)C(\C)=C\C	14.34 ± 3.32	−1.15
52	ethyl methacrylate[b]	CCOC(=O)C(C)=C	33.75	−1.53
53	butyl methacrylate[b]	CCCCOC(=O)C(C)=C	43.27	−1.63
54	2-ethylhexyl acrylate[b]	CCCCC(CC)COC(=O)C=C	0.44 ± 0.03	0.36
		Nitros		
55	1-nitro-1-cyclohexene	C1CCCC=C1N(=O)(=O)	0.03 ± 0.01	1.56
56	4-methyl-β-nitrostyrene (mixture of *cis* and *trans*)	N(=O)(=O)C=Cc1ccc(C)cc1	0.10 ± 0.03	0.94

Table 1. continued

ID	chemical	SMILES	RC$_{50}$ average	−log RC$_{50}$ average (mM)
		Nitros		
57	*trans*-β-nitrostyrene	c1ccccc1/C=C/N(=O)=O	0.09 ± 0.02	1.21
58	*trans*-4-methyl-β-nitrostyrene	O=N(=O)/C=C/c1ccc(C)cc1	0.08 ± 0.01	1.09
59	*trans*-4-chloro-β-nitrostyrene	c1cc(Cl)ccc1/C=C/N(=O)(=O)	0.07 ± 0.03	1.14
60	*trans*-4-bromo-β-nitrostyrene	O=N(=O)/C=C/c1ccc(Br)cc1	0.07 ± 0.00	1.18
61	4-fluoro-β-nitrostyrene	O=N(=O)C=Cc1ccc(F)cc1	0.05 ± 0.01	1.29
62	*trans*-4-methoxy-β-nitrostyrene	O=N(=O)/C=C/c1ccc(OC)cc1	0.04 ± 0.02	1.36
63	*trans*-β-methyl-β-nitrostyrene	c1ccccc1/C=C/(C)N(=O)(=O)	0.06 ± 0.01	1.19
		Nitriles		
64	2-methyleneglutaronitrile	N#CC(=C)CCC#N	22.92 ± 3.45	−1.36
65	cyclohexene-1-carbonitrile (1-cyanocyclohexene)	C1(C#N)=CCCCC1	28.16 ± 40.45	−1.45
66	1-cyclopentene-1-carbonitrile	N#CC1=CCCC1	20.51 ± 0.95	−1.31
		Cyclic Ketones		
67	2-cyclohexen-1-one	O=C1CCCC=C1	0.32 ± 0.13	0.50
68	2-cyclopenten-1-one	O=C1CCC=C1	0.67 ± 0.17	0.18
69	2-methyl-2-cyclopenten-1-one	CC1=CCCC1=O	9.92 ± 1.24	−1.00
70	4,4-dimethyl-2-cyclohexen-1-one	CC1(C)CCC(=O)C=C1	1.01 ± 0.11	−0.01
71	1-acetyl-1-cyclohexene	CC(=O)C1=CCCCC1	2.06 ± 0.54	−0.31
72	1-acetyl-1-cyclopentene	CC(=O)C1=CCCC1	3.90 ± 4.19	−0.59

[a]RC$_{50}$ is defined as the concentration of reactive chemical required to deplete GSH by 50% in 120 min. Average RC$_{50}$ values are given for chemicals with multiple measurements. RC$_{50}$ values were provided by T. W Schultz, obtained using a previously published spectrophotometric peptide depletion assay.[18] [b]These chemicals were unreactive in the standard 120 min GSH assay with DMSO; RC$_{50}$ for these chemicals values were obtained using 50% MeOH as solvent.
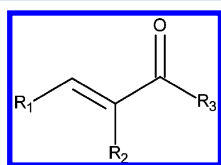
Figure 2. General structure for polarized aldehydes (R$_3$ = H), polarized ketones (R$_3$ = C), and polarized esters (R$_3$ = OC).

Table 2. Calculated Activation Energy Values for Polarized Aldehydes[a]

R$_1$	R$_2$	R$_3$	E$_{act}$ (kcal/mol)
CH$_3$	H	H	−1.5
CH$_2$CH$_3$	H	H	2.7
CH$_2$CH$_2$CH$_3$	H	H	2.5
isopropyl	H	H	3.7
tert-butyl	H	H	5.6
C$_6$H$_5$	H	H	3.4
CH$_2$C$_6$H$_5$	H	H	1.2
CH$_2$CH$_2$C$_6$H$_5$	H	H	1.8
CH$_2$CH$_2$CH$_2$C$_6$H$_5$	H	H	2.2
H	H	H	−5.4
H	CH$_3$	H	−1.7
H	CH$_2$CH$_3$	H	−1.9
H	CH$_2$CH$_2$CH$_3$	H	−2.0
H	isopropyl	H	−0.3
H	*tert*-butyl	H	2.4
H	C$_6$H$_5$	H	−8.7
H	CH$_2$C$_6$H$_5$	H	−2.2
H	CH$_2$CH$_2$C$_6$H$_5$	H	−2.5
H	CH$_2$CH$_2$CH$_2$C$_6$H$_5$	H	−2.5

[a]R groups as defined in Figure 2. Analogous data for polarized ketones and esters are available in the Supporting Information.

(sufficiently close to the 1 kcal/mol cutoff). This means that two fragments are required to define the effect of a benzene ring at the

β-position (R$_1$), with R$_1$ = CH$_2$CH$_3$ being used to predict chemicals with a benzene ring three or more carbons away from the β-carbon of the alkene. As previously, an analogous analysis was carried out for the polarized aldehydes at the α-position (R$_2$) (Table 2).

The structure−activity analysis into the effect of alkyl and aryl substituents on the calculated activation energy was repeated for the polarized ketones and esters in the dataset (varying groups at positions R$_1$, R$_2$, and R$_3$; data shown in the Supporting Information), resulting in the definition of 407 fragments, which are summarized in Table 3. These fragments cover both singly substituted chemicals and all possible combinations of the fragments shown in Table 3.

Predicting Glutathione Reactivity Using Fragment-Based *in Silico* Profiler. The ability of the fragment-based *in silico* profiler to predict GSH reactivity was investigated for a total of 54 chemicals (13 polarized aldehydes, 17 polarized ketones, and 24 polarized esters). Initial modeling using only the calculated activation energy value (E$_{act}$) failed to produce a statistically significant model due to chemicals with an α-substituent being consistently over-predicted (model 1 in Table 4 and Figure 3, where chemicals with an α-substituent are shown as filled squares). Inclusion of a SAS descriptor for the α-position resulted in a significantly improved model (model 2 in Table 4 and Figure 3). The mechanistic relevance of this descriptor likely stems from the nature of the intermediate in the Michael reaction, which involves the formation of a resonance-stabilized negative charge on the α-carbon atom. The solvation of this charge plays a key role in the stability of the transition state and thus overall reactivity. This solvation effect can be modeled by the inclusion of the steric SAS parameter, with the less solvent-accessible α-substituted chemicals being less stabilized due to solvent molecules being sterically hindered from solvating the charge by the presence of the sub-stituent compared to chemicals without an α-substituent.

Model 2 successfully improves the prediction for the majority of the chemicals in the dataset. However, closer inspection of the

Table 3. Summary of the Fragments Defined for Polarized Aldehydes, Ketones, and Esters[a]

chemical class	R_1	R_2	R_3
polarized aldehydes	alkyl: H, CH_3, CH_2CH_3, isopropyl, *tert*-butyl	alkyl: H, CH_3, isopropyl, *tert*-butyl	H
	aryl: C_6H_5, $CH_2C_6H_5$, CH_2CH_3 [for $(CH_2)_nC_6H_5$, $n \geq 2$]	aryl: C_6H_5, CH_3 [for $(CH_2)_nC_6H_5$, $n \geq 1$]	
polarized ketones	alkyl: H, CH_3, CH_2CH_3, isopropyl, *tert*-butyl	alkyl: H, CH_3, isopropyl, *tert*-butyl	alkyl: CH_3, isopropyl, *tert*-butyl
	aryl: C_6H_5, CH_3 [for $(CH_2)_nC_6H_5$, $n \geq 1$]	aryl: C_6H_5, $CH_2C_6H_5$, CH_2CH_3 [for $(CH_2)_nC_6H_5$, $n \geq 2$]	aryl: C_6H_5, $CH_2C_6H_5$, CH_2CH_3 [for $(CH_2)_nC_6H_5$, $n \geq 2$]
polarized esters	alkyl: H, CH_3, CH_2CH_3, isopropyl, *tert*-butyl	alkyl: H, CH_3, isopropyl, *tert*-butyl	alkyl: OCH_3, O-isopropyl, O-*tert*-butyl
	aryl: C_6H_5, CH_3 [for $(CH_2)_nC_6H_5$, $n \geq 1$]	aryl: C_6H_5, $CH_2C_6H_5$, $CH_2CH_2C_6H_5$, $CH_2CH_2CH_3$ [for $(CH_2)_nC_6H_5$, $n \geq 3$]	aryl: OC_6H_5, OCH_3 [for $(CH_2)_nC_6H_5$, $n \geq 1$]

[a] R groups as defined in Figure 2.

Table 4. Summary Statistics for Models 1−4 As Shown in Figure 3

model	N	a	b	c	R^2	R^2-adj	R^2-pred	average error
				$-\log RC_{50} = a + bE_{act} + cSAS\alpha$				
1	54	0.80	−0.15	—[a]	0.52	0.51	0.48	0.60
2	54	−1.05	−0.09	−0.11	0.77	0.76	0.74	0.41
3	51	−1.30	−0.07	0.12	0.81	0.80	0.78	0.37
4	47	−1.48	−0.06	0.13	0.87	0.86	0.85	0.29

[a] Model 1 has no SAS value, as it uses E_{act} as a single descriptor.
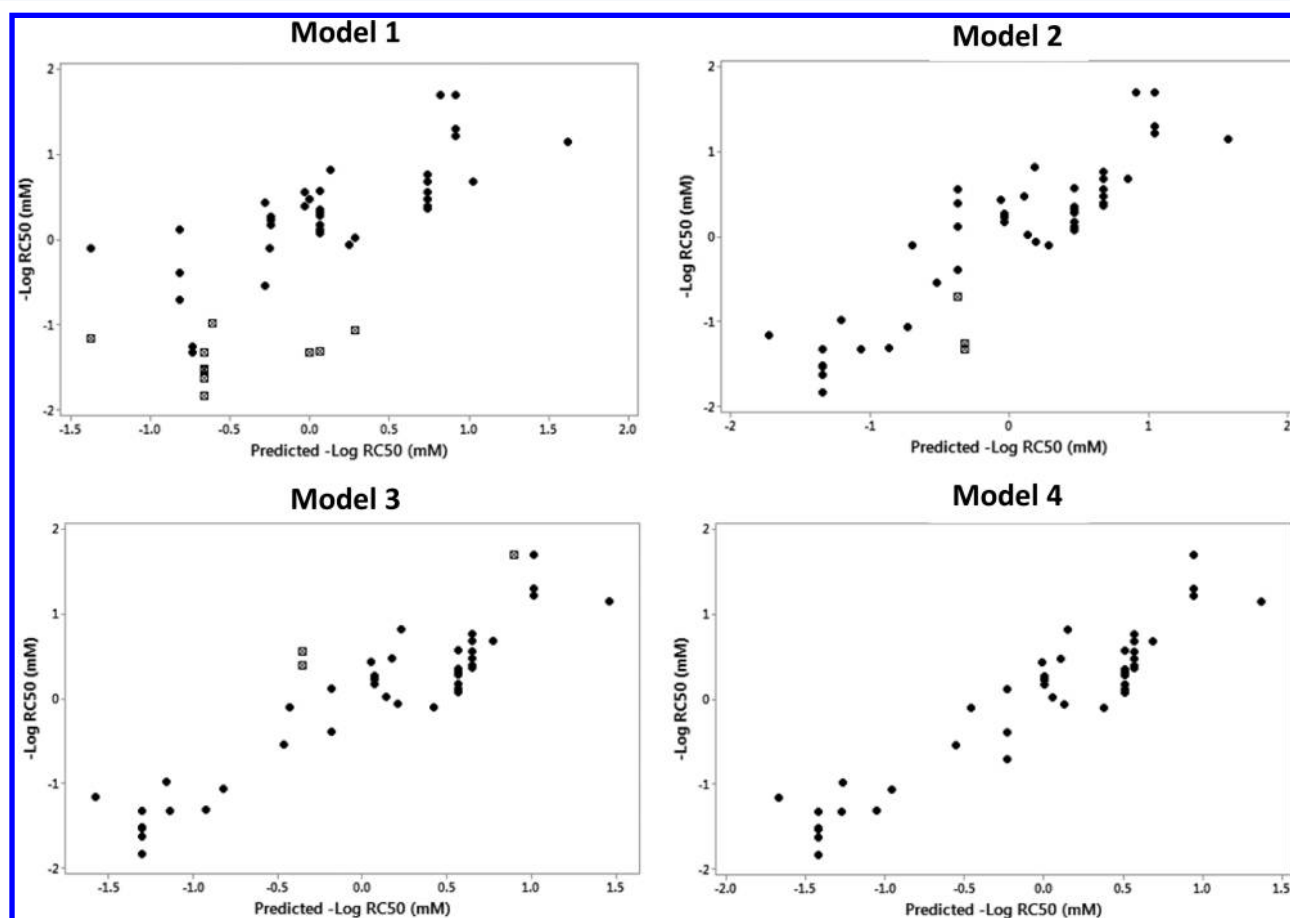
Figure 3. Predicted versus experimental values for $-\log RC_{50}$ for all models in the current study: model 1, E_{act} only; model 2, E_{act} with SAS at the α-position included; model 3, E_{act} with SAS at the α-position excluding three volatile β-esters; model 4, E_{act} with SAS at the α-position, excluding three volatile β-esters and four compounds with a phenyl electron-withdrawing group.

data shows methyl and ethyl crotonate to be significant outliers, with errors of 1.07 and 0.99 log units, respectively (Figure 3). Both methyl and ethyl crotonate have high predicted log VP

values (Table 5), as the experimental assay is carried out in scintillation vials loss of the compound during the reaction may cause an issue.[21] It may be possible that this is not being shown
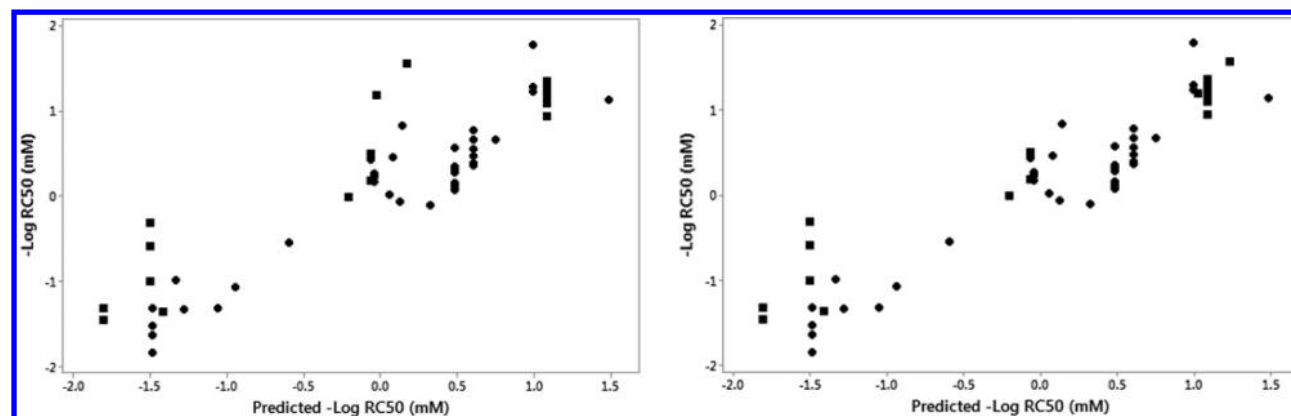
Figure 4. Predicted versus experimental −log RC$_{50}$ values for polarized nitros, polarized nitriles, and polarized cyclic ketones (shown as filled in squares) using model 4 in comparison to the polarized aldehydes, ketones, and esters in the initial dataset (shown as filled circles). Left: polarized nitros with the inclusion of the SAS descriptor. Right: polarized nitros with the SAS descriptor value set to hydrogen for all chemicals.

Table 5. Predicted Error Values for Predicted −Log RC$_{50}$ of β-Substituted Esters with Corresponding Log Vapor Pressure (VP) Values

compound	−log RC$_{50}$ (mM)	predicted −log RC$_{50}$ (mM)	error	log VP
methyl crotonate	−1.33	−0.32	1.01	1.26
ethyl crotonate	−1.25	−0.32	0.93	0.91
methyl *trans*-2-pentenoate	−0.70	−0.37	0.33	0.98
ethyl *trans*-2-hexenoate	0.12	−0.37	−0.49	0.14
methyl 2-hexenoate	−0.39	−0.37	0.02	0.54
methyl 4-methyl-2-pentenoate	−0.11	−0.69	−0.58	0.80
ethyl tiglate	−1.16	−1.72	−0.56	0.52

with the unsubstituted esters, as they are reacting sufficiently fast for the reaction to occur before the loss of reactive compound. This is therefore having a greater effect on the slower reacting β-substituted esters. With this in mind it could be suggested that

β-substituted esters with log VP values of 0.9 or greater are out of the predictive domain of this model.

An additional set of chemicals were also poorly predicted by model 3 (Figure 3, chemicals highlighted as filled squares), these being chemicals in which a phenyl ring conjugated to the carbonyl or ester moiety acts as the polarizing group (Table 6). The reactivity of these chemicals was consistently underpredicted, with error values ranging from 0.76 to 0.92 log unit. Interestingly, the analogous chemical 4-phenyl-3-buten-2-one in which the polarizing group is a simple alkyl ketone is well predicted by model 3, with an error of −0.04 log unit. This suggests that the full electron-withdrawing effect of a conjugated phenyl group at position R$_3$ is not fully captured in the calculations. (It is important to note that additional chemicals, where R$_3$ = alkyl or hydrogen and the β-position is substituted with an aromatic ring, are well predicted by the model; see Supporting Information, Table S1, chemicals 12, 13, 27, 57−61.) Removing these four chemicals from model 3 resulted in model 4 (Table 4 and Figure 3), with an average error of 0.28 log unit.

Table 6. Predicted −Log RC$_{50}$ Values for Chemicals with a Conjugated Phenyl Polarizing Group

Chemical	Structure	-Log RC$_{50}$	Predicted -Log RC$_{50}$	Error
Chalcone		0.40	-0.37	0.77
2-Hydroxy-chalcone		0.56	-0.37	0.92
4-Hydroxy-chalcone		0.39	-0.37	0.76
4-Phenyl-3-buten-2-one		-0.55	-0.51	-0.04
Phenyl-acrylate		1.64	0.94	0.76

Table 7. Fragments Required To Predict Reactivity of Polarized Nitros, Polarized Nitriles, and Cyclic Ketones

Chemical	Fragment used	New or existing fragment
Polarised Nitros		
1-Nitro-1-cyclohexene		New
4-Methyl-b-nitrostyrene *Trans*-b-nitrostyrene *Trans*-4-methyl-b-nitrostyrene *Trans*-4-chloro-b-nitrostyrene *Trans*-4-bromo-β-nitrostyrene 4-Fluoro-b-nitrostyrene *Trans*-4-methoxy-b-nitrostyrene		New
Trans-b-methyl-b-nitrostyrene		New
Polarised nitriles		
2-Methyleneglutaronitrile		New
Cyclohexene-1-carbonitrile 1-Cyclopentene-1-carbonitrile		New
Polarised cyclic ketones		
2-Cyclohexen-1-one 2-Cyclopenten-1-one		Existing
2-Methyl-2-cyclopenten-1-one 1-Acetyl-1-cyclohexene 1-Acetyl-1-cyclopentene		Existing
4,4-Dimethyl-2-cyclohexen-1-one		Existing
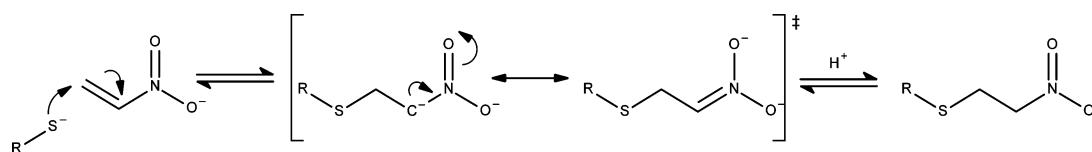
Figure 5. Michael addition mechanism for the reaction between thiol nucleophile and nitroethene (R = alkyl, GSH).

Prediction of Other Chemical Classes Using the Fragment-Based *in Silico* Profiler. To demonstrate how the fragment-based *in silico* profiler may be expended to cover additional chemical classes, a second dataset of 18 chemicals (compounds 55−72 in Table 1) with reactivity data was investigated. The chemicals within this dataset required E_{act} values for an additional five fragments to be calculated, along with three fragments previously defined (Table 7). These E_{act} values were used in conjunction with model 4 to predict $-\log \mathrm{RC}_{50}$ values for these 18 chemicals, with an average error of 0.62 log unit (Figure 4, left-hand plot shows the predicted values for these 18 chemicals as square data points in comparison to the chemicals used in the derivation of model 4). The results suggest that for the polarized nitros that substituents at the α-position have significantly less effect on reactivity than for chemicals polarized by either an aldehyde, ketone, or ester moiety. This can be rationalized in terms of the resonance stabilization of the intermediate for the polarized nitros for which two possible resonance forms exist (Figure 5). It is possible that the nitro group is sufficiently polarizing that the negative charge is localized mainly on the oxygen rather than the α-carbon, resulting in solvation at this position becoming less important. Excluding the SAS parameter for the polarized nitros (in effect assuming that these chemicals have an SAS value equivalent to hydrogen) results in a significant improvement in the predicted $-\log \mathrm{RC}_{50}$ values for these chemicals (Figure 4, right-hand plot), with an average error of 0.44 log unit. Interestingly, among the polarized nitros three of the compounds contain halogenated phenyl groups at the β- position, these are predicted well (see Supporting Information, Table S1, chemicals 59−61). This suggests that using phenyl alone was sufficient enough of a prediction and that the applicability domain of this method may extend further to alkyl and phenyl with varying substitutions.

■ CONCLUSIONS

The aim of this work was to develop an *in silico* profiler capable of predicting reactivity for polarized aldehydes, ketones, and esters acting via Michael addition. The results showed that a

combination of pre-calculated E_{act} values coupled with a descriptor for the solvent-accessible surface area at the α-carbon was able to accurately predict chemical reactivity as measured in a glutathione depletion assay. Two sets of chemicals were poorly predicted by the approach: volatile esters with an extended substituent at the β-carbon, and chemicals containing a conjugated benzene ring as part of the polarizing group. The study also demonstrated the ease with which the approach can be extended to other chemical classes by the calculation of additional fragments and their associated E_{act} and SAS values. This approach, along with the associated *in silico* profiler, enables chemical reactivity to be predicted without the use of time-consuming quantum mechanics calculations and is likely to be of use in regulatory toxicology tools where an understanding of covalent bond formation as a potential molecular initiating event is important within the adverse outcome pathway paradigm.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.chemrestox.6b00099.

> All data for 72 chemicals (13 aldehydes, 17 ketones, 24 esters, 9 nitro compounds, 3 nitrile-containing compounds, and 6 cyclic ketones), including chemical names, SMILES, experimental values (RC$_{50}$ average, RC$_{50}$, and $-\log$ RC$_{50}$), calculated values (fragment E_{act} and fragment SAS), the fragment used, and predicted $-\log$ RC$_{50}$ values for models 1–4 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Tel: + 44 151 231 2164. Fax: + 44 151 231 2170. E-mail: s.j.enoch@ljmu.ac.uk.

Notes

The authors declare no competing financial interest.

■ ABBREVIATIONS

AOP, Adverse Outcome Pathway; DFT, density functional theory; E_{act}, activation energy; MIE, molecular initiating event; OECD, Organisation for Economic Co-operation and Development; QSAR, quantitative structure–activity relationship; REACH, Registration, Evaluation, Authorisation, and Restriction of Chemicals; SAS, solvent-accessible surface area; VP, vapor pressure

■ REFERENCES

(1) EC Regulation No. 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Off. J. Eur. Union, L396/1 of 30.12.2006. Commission of the European Communities, 2006.

(2) Ankley, G. T., Bennett, R. S., Erickson, R. J., Hoff, D. J., Hornung, M. W., Johnson, R. D., Mount, D. R., Nichols, J. W., Russom, C. L., Schmieder, P. K., Serrrano, J. A., Tietge, J. E., and Villeneuve, D. L. (2010) Adverse Outcome Pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environ. Toxicol. Chem.* 29, 730–741.

(3) Przybylak, K. R., and Schultz, T. W. (2013) Informing Chemical Categories through the Development of Adverse Outcome Pathways, in *Chemical Toxicity Prediction: Category Formation and Read-across*, pp 44–67, Royal Society of Chemistry.

(4) 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.

(5) Aptula, A. O., Enoch, S. J., and Roberts, D. W. (2009) Chemical mechanisms for skin sensitization by aromatic compounds with hydroxy and amino groups. *Chem. Res. Toxicol.* 22, 1541–1547.

(6) Enoch, S. J., Roberts, D. W., and Cronin, M. T. D. (2009) Electrophilic Reaction Chemistry of Low Molecular Weight Respiratory Sensitizers. *Chem. Res. Toxicol.* 22, 1447–1453.

(7) Verhaar, H. J. M., van Leeuwen, C. J., and Hermens, J. L. M. (1992) Classifying environmental-pollutants 0.1. Structure-activity-relationships for prediction of aquatic toxicity. *Chemosphere* 25, 471–491.

(8) Obach, R. S., Kalgutkar, A. S., Soglia, J. R., and Zhao, S. X. (2008) Can in vitro metabolism-dependent covalent binding data in liver microsomes distinguish hepatotoxic from nonhepatotoxic drugs? An analysis of 18 drugs with consideration of intrinsic clearance and daily dose. *Chem. Res. Toxicol.* 21, 1814–1822.

(9) Mekenyan, O., Todorov, M., Serafimova, R., Stoeva, S., Aptula, A., Finking, R., and Jacob, E. (2007) Identifying the structural requirements for chromosomal aberration by incorporating molecular flexibility and metabolic activation of chemicals. *Chem. Res. Toxicol.* 20, 1927–1941.

(10) Kalgutkar, A. S., and Didiuk, M. T. (2009) Structural alerts, reactive metabolites, and protein covalent binding: how reliable are these attributes as predictors of drug toxicity? *Chem. Biodiversity* 6, 2115–2137.

(11) Schwobel, J. A. H., Madden, J. C., and Cronin, M. T. D. (2011) Application of a computational model for Michael addition reactivity in the prediction of toxicity to Tetrahymena pyriformis. *Chemosphere* 85, 1066–1074.

(12) 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. *Crit. Rev. Toxicol.* 41, 783–802.

(13) 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.

(14) 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.

(15) 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.

(16) Rodriguez-Sanchez, N., Schultz, T. W., Cronin, M. T. D., and Enoch, S. J. (2013) Experimental verification of structural alerts for the protein binding of cyclic compounds acting as Michael acceptors. *SAR QSAR Environ. Res.* 24, 963–977.

(17) Bohme, A., Thaens, D., Schramm, F., Paschke, A., and Schuurmann, G. (2010) Thiol reactivity and its impact on the ciliate toxicity of α,β-unsaturated aldehydes, ketones, and esters. *Chem. Res. Toxicol.* 23, 1905–1912.

(18) 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.

(19) Mulliner, D., Wondrousch, D., and Schuurmann, G. (2011) Predicting Michael-acceptor reactivity and toxicity through quantum chemical transition-state calculations. *Org. Biomol. Chem.* 9, 8400–8412.

(20) Schwobel, J. A. H., Wondrousch, D., Koleva, Y. K., Madden, J. C., Cronin, M. T. D., and Schuurmann, G. (2010) Prediction of Michael-type acceptor reactivity toward glutathione. *Chem. Res. Toxicol.* 23, 1576–1585.

(21) 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.

(22) Schwobel, J. A. H., Madden, J. C., and Cronin, M. T. D. (2010) Examination of Michael addition reactivity towards glutathione by transition-state calculations. *SAR QSAR Environ. Res. 21*, 693−710.

(23) Cee, V. J., Volak, L. P., Chen, Y., Bartberger, M. D., Tegley, C., Arvedson, T., McCarter, J., Tasker, A. S., and Fotsch, C. (2015) Systematic study of the glutathione (GSH) reactivity of N-Arylacrylamides: 1. Effects of aryl substitution. *J. Med. Chem. 58*, 9171−9178.

(24) Dahal, U. P., Gilbert, A. M., Obach, R. S., Flanagan, M. E., Chen, J. M., Garcia-Irizarry, C., Starr, J. T., Schuff, B., Uccello, D. P., and Young, J. A. (2016) Intrinsic reactivity profile of electrophilic moieties to guide covalent drug design: N-[α]-acetyl-l-lysine as an amine nucleophile. *MedChemComm*, DOI: 10.1039/C6MD00017G.

(25) Nelms, M. D., Cronin, M. T. D., Schultz, T. W., and Enoch, S. J. (2013) Experimental verification, and domain definition, of structural alerts for protein binding: epoxides, lactones, nitroso, nitros, aldehydes and ketones. *SAR QSAR Environ. Res. 24*, 695−709.

(26) 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., Jr., 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, Gaussian Inc., Wallingford, CT.

(27) Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem. 25*, 1605−1612.
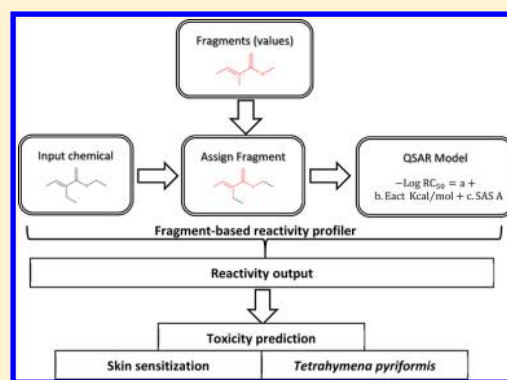
Validation of a Fragment-Based Profiler for Thiol Reactivity for the Prediction of Toxicity: Skin Sensitization and *Tetrahymena pyriformis*

David J. Ebbrell,[†][ORCID] Judith C. Madden,[†] Mark T. D. Cronin,[†] Terry W. Schultz,[‡] and Steven J. Enoch*,[†]

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

[‡]Department of Comparative Medicine, College of Veterinary Medicine, The University of Tennessee, Knoxville, Tennessee 37996, United States

ABSTRACT: This study outlines the use of a recently developed fragment-based thiol reactivity profiler for Michael acceptors to predict toxicity toward *Tetrahymena pyriformis* and skin sensitization potency as determined in the Local Lymph Node Assay (LLNA). The results showed that the calculated reactivity parameter from the profiler, $-\log RC_{50}(\text{calc})$, was capable of predicting toxicity for both end points with excellent statistics. However, the study highlighted the importance of a well-defined applicability domain for each end point. 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 $-\log RC_{50}(\text{calc})$ as a descriptor, while the second required the addition of a descriptor for hydrophobicity. Modeling of the LLNA required only a single descriptor, $-\log RC_{50}(\text{calc})$, enabling potency to be predicted. The applicability domain excluded chemicals capable of undergoing polymerization and those that were predicted to be volatile. The modeling results for both end points, using the $-\log RC_{50}(\text{calc})$ value from the profiler, were in keeping with previously published studies that have utilized experimentally determined measurements of reactivity. These results demonstrate that the output from the fragment-based thiol reactivity profiler can be used to develop quantitative structure–activity relationship models where reactivity toward 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 sensitization 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 polarizing 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 stabilized 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 sensitization.[10] While *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
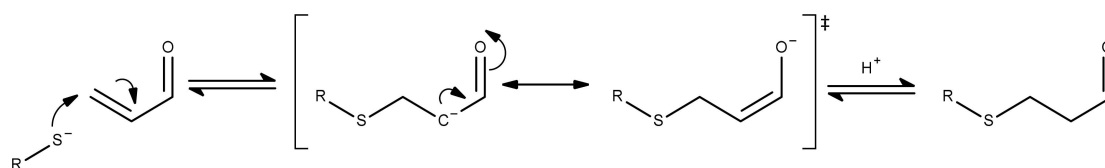
Figure 1. Michael addition reaction between acrolein and a thiol nucleophile (R = glutathione, alkyl).
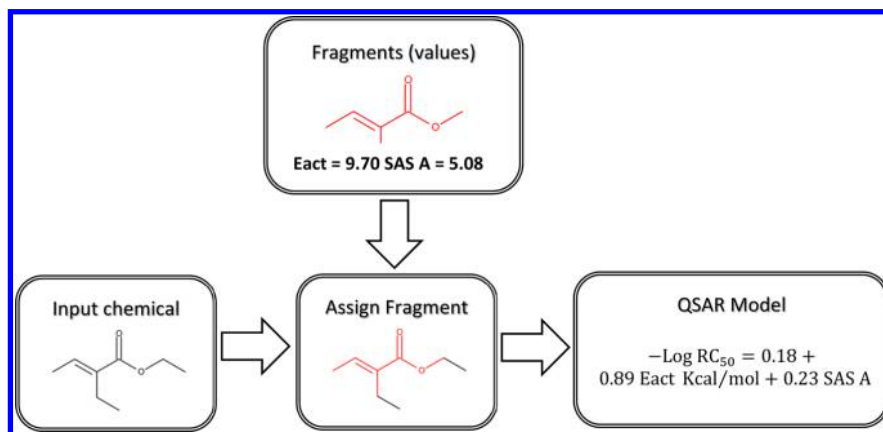

Figure 2. Summary of the workflow used to predict reactivity ($-\log RC_{50}$).

sensitization potency.[3,4] Previous studies have also utilized Hammett and Taft descriptors to model chemical reactivity for the prediction of skin sensitization.[17] These descriptors were derived from extensive studies into the effect of substituents upon the acidic dissociation constant (pK_a) 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 don't 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 sensitization.[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 is capable of predicting potency for aquatic toxicity and skin sensitization. However, the derivation of Eact 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 utilizing 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 toward glutathione (expressed as $-\log RC_{50}$) through the use of fragments with precalculated 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 precalculated activation energy values. The methodology was encoded as a KNIME workflow through which chemicals of interest can be 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 $-\log RC_{50}$ values) based on a previously developed QSAR model; this process is summarized in Figure 2. Therefore, given the availability of a fragment-based profiler, the aim of this study was to validate the calculated $-\log RC_{50}$ values generated from the fragment based profiler for thiol reactivity in predicting toxicity to *Tetrahymena pyriformis* and skin sensitization potency (as determined in the LLNA) for Michael acceptors.

■ METHODS

Computational Methods. The previously published fragment-based profiler for thiol reactivity was utilized in the current study to predict reactivity toward 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 2 h time period.[22] The fragment-based reactivity profiler was trained on a set of polarized 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 sensitization data sets were generated using a previously developed KNIME workflow encoding the fragment-based
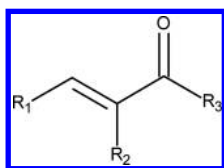
Figure 3. Domain covered by the fragment method for Michael acceptors. (R_1 = hydrogen, alkyl, aryl) (R_2 = hydrogen, alkyl, aryl) (R_3 = H) for polarized aldehydes, (R_3 = CH, C-alkyl, C-aryl) for polarized ketones, and (R_3 = OCH, OC-alkyl, OC-aryl) for polarized esters.

reactivity profiler for thiol reactivity (this workflow, including calculated fragments is available from the authors on request).[21] The workflow utilizes a database of fragments with precalculated activation energy values (Eact) calculated using density functional theory (DFT) at the B3LYP/6-31G+(d) level of theory (calculations performed using Gaussian 09 and with water as a solvent).[23] The workflow is summarized in Figure 2. Descriptors for hydrophobicity (log K_{ow}) and vapor pressure (log VP) were calculated using the KOWWIN (V1.68) and MPBPWIN (V1.43) modules of the EPI suite.[24]

Data Sets for *Tetrahymena pyriformis* and Skin Sensitization. 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 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-h exposure period to the test chemical (also recorded as EC_{50} values).[26] A similar analysis of skin sensitization data gathered from the Local Lymph Node Assay (LLNA) resulted in a data set 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 3-fold stimulation in the lymph nodes; this is reported as an EC3 value (% weight) for the chemical. If the chemical does not produce a 3-fold stimulation, it is not considered a sensitizer. All EC3 values were converted to pEC3 values (eq 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 a final data set of 26 skin sensitizing chemicals.[30]

$$pEC3 = \log(EC3/\text{molecular weight}) \tag{1}$$

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 standardized residuals as identified by Minitab. Chemicals for which a mechanistic rationale enabling outlying behavior 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 the toxicity of Michael acceptors toward *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 data set of 62 chemicals (14 aldehydes, 12 ketones, and 36 esters) with corresponding EC_{50} values (Table 1). Initial modeling 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 4). Interestingly, this value is lower than that published on a data set 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]

$$\log(1/EC_{50}) = 0.63 + 0.61 - \log RC_{50}(\text{calc})$$

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

$$\text{(model 1)}$$

Consistent with this hypothesis, a related study showed that splitting the data into fast reacting and slow reacting classes resulted in significantly improved modeling results.[5] Importantly, the toxicity to *Tetrahymena pyriformis* for the fast reacting chemicals could be predicted from experimental reactivity alone, while those in the slow reacting class required both hydrophobicity and reactivity. The authors suggested a reactivity cutoff to distinguish the two classes based on model 1, where chemicals with a $D_{kk} < 3$ were fast reacting, and those with $D_{KK} > 3$ were slow reacting. Applying these criteria to the current data set, using −log RC_{50}(calc) as a measure of reactivity resulted in models 2a and 2b (fast and slow reacting chemicals, respectively). Forty-three chemicals were assigned to the fast reacting class (chemicals 1−43 in Table 1), while 19 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), while 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 2a and 2b.

$$D_{KK} = \log(K_{ow}/-\log RC_{50}(\text{calculated}))$$

$$= \log K_{ow} - -\log RC_{50}(\text{calculated}) \tag{model 2}$$

$$\log(1/EC_{50}) = 0.41 + 0.94 - \log RC_{50}(\text{calc})$$

$$N = 43,\ R^2 = 0.78,\ R^2 - \text{adj} = 0.77,\ s = 0.30$$

$$\text{(model 2a)}$$

$$\log(1/EC_{50}) = -1.82 + 0.35 - \log RC_{50}(\text{calculated})$$

$$+ 0.89\log K_{ow}$$

$$N = 19,\ R^2 = 0.85,\ R^2 - \text{adj} = 0.83,\ s = 0.31$$

$$\text{(model 2b)}$$

PREDICTION OF SKIN SENSITIZATION POTENCY AS DEFINED IN THE LLNA

The rate of covalent bond formation has also been shown to be important for the prediction of skin sensitization 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 −log RC_{50}(calc) resulted in extremely poor statistics (model 3 in Figure 6). Despite this, 13 of the chemicals were predicted within a 2-fold error of the corresponding

Table 1. 62 Chemicals Used in the Assessment of the Fragment Method for Predicting *Tetrahymena pyriformis* Toxicity (log 1/ EC$_{50}$ mmol/L)^a

ID	chemical	SMILES	log (1/EC$_{50}$) (mmol/L)	−log RC$_{50}$(calc)	logK$_{ow}$	D$_{kk}$	predicted log (1/ EC$_{50}$) (mmol/L) 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	CCCC\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

D

Table 1. continued

ID	chemical	SMILES	log $(1/EC_{50})$ (mmol/L)	$-$log RC_{50}(calc)	logK_{ow}	D_{kk}	predicted log $(1/EC_{50})$ (mmol/L) model 1	model 2a/b
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	CCCC(CC)COC(=O)C(C)=C	1.57	-1.40	4.64	6.04	-0.23	1.80

[a]Chemical names, SMILES, experimental log $(1/EC_{50})$ (mmol/L) with $-$log RC_{50}(calc), D_{kk}, and predicted log $(1/EC_{50})$ (mmol/L) for the respective models are shown. log $(1/EC_{50})$ (mmol/L) values were calculated with model 2a for fast reacting chemicals (1−43) and model 2b for slower reacting chemicals (44−62).
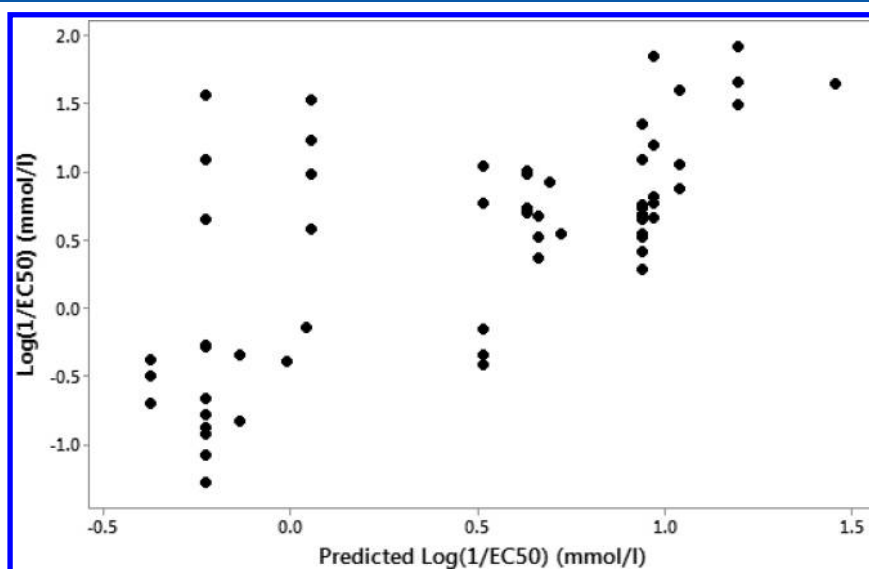

Figure 4. Predicted log $(1/EC_{50})$ (mmol/L) values against experimental log $(1/EC_{50})$ (mmol/L) values for all 62 Michael acceptors using $-$log RC_{50}(calc) alone (model 1).
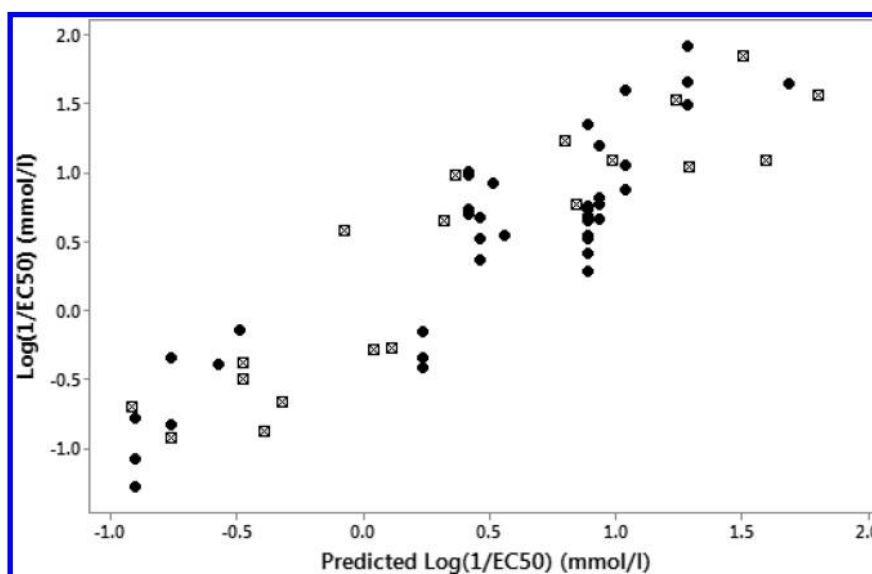
Figure 5. Predicted log $(1/EC_{50})$ (mmol/L) against experimental log $(1/EC_{50})$ (mmol/L) of all 43 fast reacting chemicals (bold circles) (model 2a, chemicals 1−43 in Table 1)) and 19 slower reacting chemicals (squares) (model 2b, chemicals 44−62 in Table 1) requiring hydrophobicity to be taken into account.

experimental value (chemicals with a predicted value within 0.3 log units of the experimental value). These predictions are within the experimental 2-fold error of the LLNA.[31] Any chemicals outside of the 2-fold error of the experimental assay were considered as outliers (labeled in Table 2) and were analyzed to rationalize the error in their predictions.

Table 2. 26 Chemicals Used in the Assessment of the Fragment-Based Reactivity Algorithm's Ability to Predict Skin Sensitization Potency (pEC3)[a]

| ID	chemical	SMILES	pEC3	−log RC$_{50}$(calc)	log VP	predicted pEC3	
						model 3[b]	model 4
1	methyl methacrylate	CC(=C)C(=O)OC	0.05	−1.40	1.59	1.28(1.23)[b]	
2	2-hydroxypropyl methacrylate	CC(COC(=O)C(=C)C)O	0.46	−1.40	−1.10	1.28(0.82)[b]	
3	ethyl acrylate	CCOC(=O)C=C	0.55	0.50	1.61	1.38(0.83)[b]	
4	methyl acrylate	COC(=O)C=C	0.63	0.50	1.95	1.38(0.75)[b]	
5	butyl acrylate	CCCCOC(=O)C=C	0.81	0.50	0.74	1.38(0.57)[b]	
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	CCCC(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)[b]	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)[b]	1.95(0.16)
21	galbone	CC1(C)CCC=C(C1)C(=O)CCC=C	1.81	−1.25 (0.15)[c]	−1.72	1.29(−0.52)[b]	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)[b]	
23	diethyl maleate	CCOC(=O)/C=C/C(=O)OCC	1.91	0.10	−0.72	1.36(−0.55)[b]	1.82(−0.09)
24	2-hydroxyethyl acrylate	C=CC(=O)OCCO	1.92	0.50	−0.85	1.38(−0.54)[b]	
25	spirogalbanone	C=CCCC(=O)C1=CCCC2(CCCC2)C1	2.00	−1.25 (0.15)[c]	−3.02	1.29(−0.71)[b]	1.84(−0.16)
26	pomarose	C\C=C\C(=O)C(=O)C(C)=C(/C)C(C)C	2.02	0.15	−0.55	1.36(−0.66)[b]	1.84(−0.18)

[a]Chemicals, SMILES, experimental pEC3 with error values, −log RC$_{50}$(calc), log VP, and predicted pEC3 values for all models are shown. [b]Notes chemicals with predictions outside of the experimental error for that model. Error values for predicted pEC3 values for all chemicals are shown in parentheses. [c]Chemicals with additional −log RC$_{50}$(calc) values use this value for model 3 for the reasons discussed in the text.

F
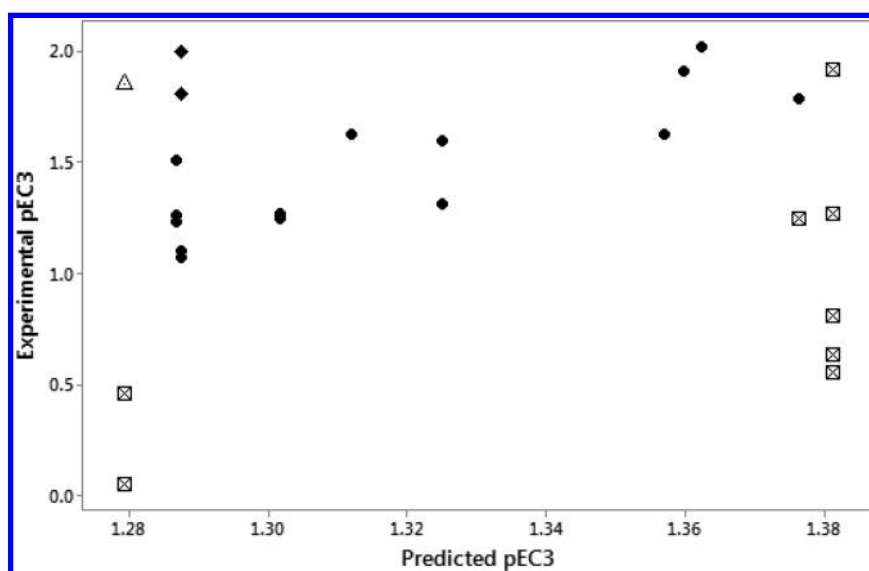
Figure 6. Predicted pEC3 versus experimental pEC3 for all 26 Michael acceptors shown in Table 2. □ = volatile chemicals; ♦ = galbanone and spirogalbanone; △ = 5,5-dimethyl-3-methylene-dihydro-2(3*H*)-furone.
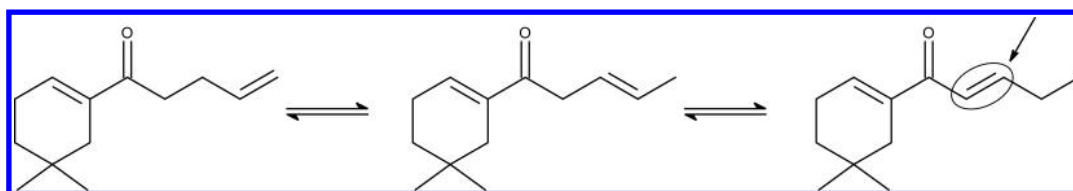
Figure 7. Isomerization of galbanone to produce extended conjugated chemicals highlighting a possible additional site of reactivity.

$$\text{predicted pEC}_3 = 1.35 + -0.05 - \log \text{RC}_{50}(\text{calc})$$

$$N = 26, \; R^2 = 0.00, \; R^2 - \text{adj} = 0.00, \; s = 0.3$$

(model 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 sensitization 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 polymerization driven by free radical chemistry in the skin.[32,33] Interestingly, the toxicity of a large number of similar chemicals toward *Tetrahymena pyriformis* was well predicted (chemicals 26−62 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 is a case of applying a cautionary applicability domain to the model for these types of chemicals.

In contrast to the overprediction 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 sensitization potency of these two chemicals was predicted using 3-methly-3-penten-2-one as the reference fragment to take into 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 favored when the alkene group is unsubstituted (CH₂=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 sensitizing ability of these chemicals. The more reactive alternative site for Michael addition was utilized 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-(3*H*)-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

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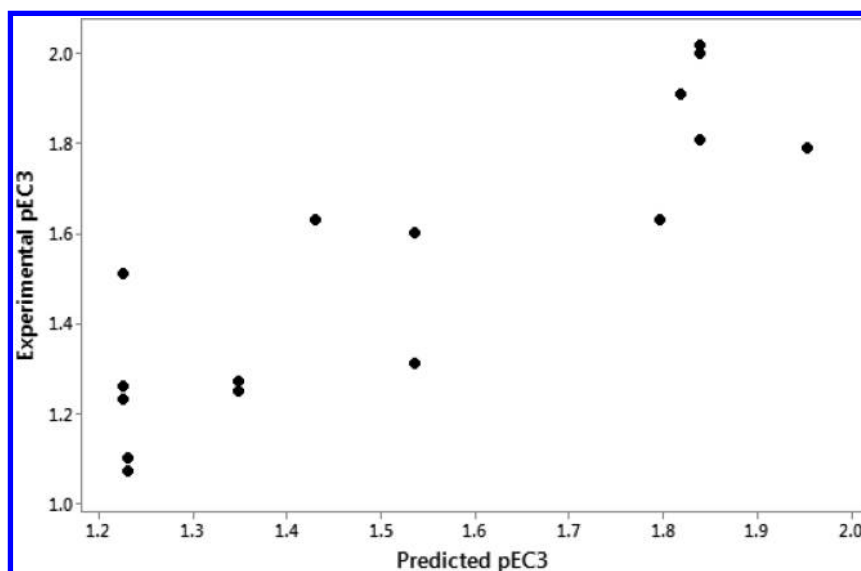
Figure 8. Predicted pEC3 against experimental pEC3 for model 4 (predicted values are shown in Table 2).

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 sensitization 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* data set analyzed 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 an applicability domain similar to that published using experimentally determined kinetic rate constants, in that volatile chemicals and those that can polymerize are excluded.[3,4] However, the use of $-\log RC_{50}(\text{calc})$ in the current study enabled a greater number of chemicals to be predicted (17 versus 10) while maintaining a similar level of statistical accuracy ($R^2 = 0.76$ versus 0.84).

$$pEC3 = 1.77 + 0.43 - \log RC_{50}(\text{calculated})$$

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

(model 4)

CONCLUSIONS

The aim of this work was to validate the fragment-based reactivity profiler for thiol reactivity for the prediction of toxicity to *Tetrahymena pyriformis* and skin sensitization potency for Michael acceptors. The results of this study showed that the predicted reactivity values ($-\log RC_{50}(\text{calc})$) was able to predict both end points within well-defined, end-point specific applicability domains. The results showed the importance of considering slow versus fast reacting Michael acceptors when modeling toxicity to *Tetrahymena pyriformis* and polymerization and volatility to be important in successfully predicting skin sensitation potency. These results were in keeping with previously published studies that have utilized experimentally determined measurements of chemical reactivity to model the same end points. The statistical quality of resulting QSAR models demonstrated that the predicted reactivity values generated by the fragment-based profiler for thiol reactivity are on 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 toward 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 approach outlined for the development of the fragment-based *in silico* profiler could be extended to other end points, for example, genotoxicity where nitrogen acts as the nucleophile. Such developments would be dependent on the availability of reactivity data for nitrogen-based nucleophiles.

AUTHOR INFORMATION

Corresponding Author

*Tel: + 44 151 231 2164. Fax: + 44 151 231 2170. E-mail: s.j.enoch@ljmu.ac.uk.

ORCID

David J. Ebbrell: 0000-0001-5329-7539

Notes

The authors declare no competing financial interest.

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

■ 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. *Toxicol. Sci. 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. *Mol. Inf. 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. *Crit. Rev. Toxicol. 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. *Crit. Rev. Toxicol. 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. *J. Phys. Chem. Lett. 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. *Org. Biomol. Chem. 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, Gaussian, Inc., Wallingford, CT.

(24) US EPA (2015) *Estimation Programs Interface Suite for Microsoft Windows*, v 4.11, United States Environmental Protection Agency, Washington, DC.

(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. *Environ. Toxicol. Pharmacol. 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. *J. Appl. Toxicol. 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*, National Institute of Environmental Health Sciences, Research Triangle Park, NC.

(32) Barner-Kowollik, C. (2009) Acrylate Free Radical Polymerization: From Mechanism to Polymer Design. *Macromol. Rapid Commun. 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. *Eur. J. Lipid Sci. Technol. 115*, 28−40.

(34) Hubert, A. J., and Reimlinger, H. (1969) The Isomerization of Olefins Part I. Based-Catalysed Isomerization of Olefins. *Synthesis* *1969*, 97−112.

Development of a fragment-based *in silico* profiler for S$_N$2 thiol reactivity and its application in predicting toxicity of chemicals towards *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
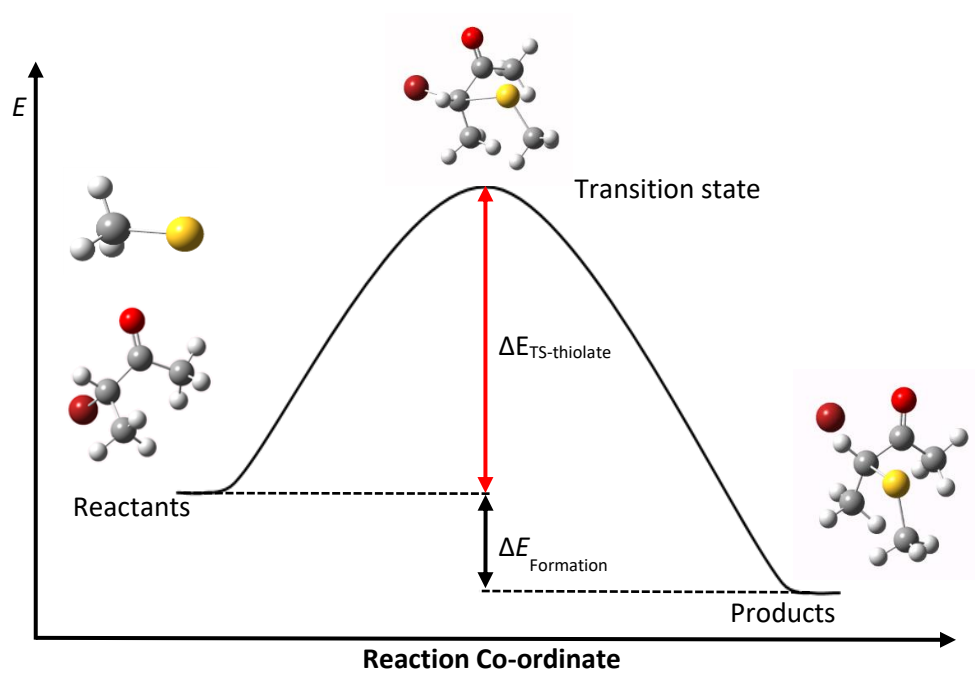
*Corresponding author
Tel: + 44 151 231 2164
Fax: + 44 151 231 2170
Email: s.j.enoch@ljmu.ac.uk

Graphical abstract

Abstract

This study outlines the development of a fragment-based *in in silico* profiler for S_N2 thiol reactivity. The profiler was developed based on the chemical space covered by a previously published dataset of glutathione reactivity data covering S_N2 chemicals activated by the presence of an adjacent π-system. The approach is in keeping with a recently developed fragment-based *in silico* profiler for Michael addition thiol reactivity and involved developing a database of structural alert-type fragments with associated activation energy values ($\Delta E_{INT\text{-}thiolate}$). These energy values were calculated using density functional theory with a B3YLP functional coupled to a 6-31G+(d) basis set. The fragment-based *in silico* profiler for S_N2 thiol reactivity was utilized to predict glutathione reactivity and toxicity towards *Tetrahymena pyriformis* of a series of S_N2 chemicals. The results showed the fragment-based *in silico* profiler was able to successfully predict both glutathione reactivity and toxicity to *Tetrahymena pyriformis*. Overall the results of this study extend the previous fragment-based profiler development to the S_N2 domain and further validate the approach. The study also highlights the ability of the fragment-based *in silico* profilers to predict toxicological potency where the formation of a covalent bond is the key Molecular Initiating Event.

Introduction

Human health risk assessment faces significant challenges as there is a demand for greater speed of chemical assessment and the requirement to minimise or avoid animal usage. The Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) regulation states that all chemicals produced or imported into the European Union in quantities of one ton per annum (or more) need to be assessed for human and environmental hazards.[1] To test all of these chemicals using traditional animal methods (*in vivo*) would be costly, time consuming and raises ethical concerns in terms of animal usage.[2] The seventh amendment to the cosmetics directive bans the use of animal testing for cosmetic products, hence there is clearly a need to develop robust alternative methods. The Adverse Outcome Pathway (AOP) paradigm is seen as the key approach that will enable the demands of seventh amendment to the cosmetic directive to be met in regulatory toxicology. An AOP details the existing knowledge that links the initial interaction between a chemical and a biological system, termed the Molecular Initiating Event (MIE), through a series of intermediate key events, to an adverse outcome.[3] Within the AOP approach, computational (*in silico*) methods are typically used to define the chemistry associated with the MIE

The formation of a covalent bond between electrophilic chemicals and biological nucleophiles (such as cysteine and lysine groups of proteins) is an example of a well-defined MIE. Bimolecular nucleophilic substitution (S_N2) is an example of a mechanism through which covalent bonds may form between electrophiles and nucleophiles. This typically occurs at an aliphatic carbon, nitrogen, sulphur or halogen atom bound to an electronegative leaving group (Figure 1).[4] Unlike Michael addition, the S_N2 reaction has no stable intermediate as the attack by the nucleophile and loss of the leaving group is assumed to happen simultaneously.
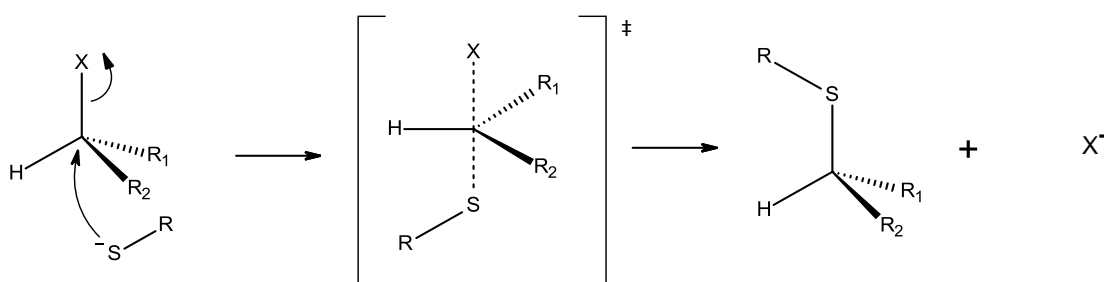

Figure 1: The formation of a covalent bond between an electrophilic chemical and cysteine via an S_N2 mechanism (X = halogen, R = alkyl, R_1 = R_2 = hydrogen or carbon)

Given the importance of covalent bond formation as an MIE in various toxicities for example skin sensitisation and aquatic toxicity, several studies have attempted to predict toxicological potency using *in chemico* and *in silico* methods. [5-14] In terms of toxicity relating to covalent bond formation, one of the primary assumptions is that toxicological potency and rate of covalent bond formation (or reactivity)

are proportionally related. Consequently, *in chemico* analysis is often used to quantify the rate of reactivity between electrophilic chemicals and nucleophilic peptides, such as glutathione, representing the biological system. For example, a study by Roberts et al predicted toxicity towards *Tetrahymena pyriformis* for a set of 60 chemicals potentially able to react via an S_N2 mechanism using glutathione depletion data (expressed as RC_{50} values).[15] The study found that it was possible to predict toxicity to *Tetrahymena pyriformis* by assigning chemicals to groups based on their reaction mechanism characteristics. This resulted in the definition of four groups: eight non-activated primary halides (of which all were unreactive), nine chemicals activated by an unsaturated hydrocarbon, 22 chemicals activated by an unsaturated activating group and 21 chemicals whose nature and/or mechanism of reaction could not be assigned solely as acting via an S_N2 mechanism. Glutathione reactivity data were successfully used to predict the toxicity of the largest group of 22 chemicals identified as acting via an S_N2 mechanism (chemicals activated by an unsaturated activating group). In addition, the toxicity of a further nine chemicals whose mechanism could not initially be assigned as definitely acting via an S_N2 mechanism were also well predicted using glutathione data, suggesting that these chemicals most likely act via S_N2 (with the potential competing reactions playing little or no role in determining toxicity). This resulted in a final model of 31 chemicals whose toxicity to *Tetrahymena pyriformis* could be successfully predicted using glutathione reactivity data alone. It was suggested by Roberts et al that the additional chemicals may be acting through alternative mechanisms such as via competing S_N1 reactions or after elimination reactions leading to the production of Michael acceptors.[15]

Previous research has shown that data from the *in vitro Tetrahymena pyriformis* growth impairment assay correlates well with experimental reactivity data for other mechanistic domains.[7, 9, 15-17] Although data from *in chemico* reactivity assays have been used successfully to relate reactivity to toxicity (such as to *Tetrahymena pyriformis*) recent efforts have focussed on obviating the need to conduct these laboratory experiments entirely through the development of *in silico* alternatives. Such efforts involve the calculation of quantum mechanical descriptors, to enable prediction of toxicity or reactivity directly from structure. [11-13, 18] For example, descriptors such as the energy values of reactants and transition state and/or key intermediate structures (typically using a model nucleophile e.g. methyl thiol). In doing this it is possible to calculate the energy difference between the reactants and transition states and/or intermediate structures computationally, this is termed the activation energy (Eact). Previous studies have shown calculated Eact to correlate well with reaction rate for Michael acceptors.[11, 12, 14, 19]

A recent study by the current authors showed it was possible to predict experimental reactivity (expressed as -log RC_{50} values) for a set of Michael acceptors using a fragment-based *in silico* profiler where fragments are stored in a database with their respective, pre-calculated activated energy (Eact) values.[19] Fragments were developed for Michael acceptors by defining the length of alkyl chain beyond which further increases in chain length failed to significantly increase Eact. Query chemicals (input as SMILES) are compared to the database of fragments with pre-calculated Eact values along with an

additional descriptor that models the solvent accessible surface area (SAS) at the α-position. Once the query chemical has been assigned a reference fragment, its corresponding Eact and SAS-α values are used to calculated –log RC_{50} using a defined QSAR model. Furthermore, the potential of this fragment-based profiler, for Michael acceptors, to predict toxicities associated with covalent bond formation was assessed using *Tetrahymena pyriformis* toxicity data and skin sensitisation potency as measured in the Local Lymph Node Assay (LLNA).[20] These studies demonstrated that it was possible to predict toxicity to *Tetrahymena pyriformis* using two different models that differentiated between fast and slow reacting chemicals; the model for slow reacting chemicals incorporated an additional descriptor for lipophilicity (log K_{OW}). Additionally skin sensitisation potency as measured in the LLNA was successfully modelled within a well-defined applicability domain where volatile chemicals and those with the potential to polymerise were excluded. These findings were in keeping with previously published studies for both toxicity to *Tetrahymena pyriformis* and skin sensitisation potency. [21, 22]

Given that experimental reactivity towards glutathione, and toxicity that is associated with such reactivity, were successfully predicted for linear Michael acceptors in previous studies, it plausible that an analogous method could be successfully applied to other mechanistic domains. As such, the aim of this study was to develop a fragment based *in silico* profiler for the prediction of S_N2 thiol reactivity. The fragments developed for this are based on the chemicals investigated in a previous study by Roberts et al.[15] This will be achieved by adopting a similar method as applied previously to Michael acceptors.[19] Additionally, the ability of fragment-based *in silico* profiler for S_N2 thiol reactivity to predict both glutathione reactivity and toxicity towards *Tetrahymena pyriformis* was also investigated.

Methods

Data set

Thirty-one chemicals were identified as acting via an S_N2 mechanism from reference (chemicals shown in Table 6).[10] Three chemicals were excluded from this dataset, these being: 3-bromo-acetyl-coumarin, ethyl iodoacetate and 2-iodoacetamide. 3-Bromo-acetyl-coumarin was excluded from the analysis due to it having multiple sites of electrophilic reactivity. The other two chemicals contained iodine as the leaving group. It was not possible to perform calculations on these chemicals due to the chosen basis set only being applicable to elements in the first three rows of the periodic table. This resulted in a dataset of 28 activated S_N2 chemicals. All chemicals in the dataset had associated glutathione reactivity data (-log RC_{50}) and *Tetrahymena pyriformis* toxicity data (-log IGC_{50}).[10]

Computational methods – calculation of $\Delta E_{TS\text{-}thiolate}$

All calculations were carried out using the Gaussian 09 suite of software using density functional theory at the B3LYP/6-31G+(d) level of theory with water as a solvent.[23] The $\Delta E_{TS\text{-}thiolate}$ values were obtained using scan calculations to determine the highest point of energy on the potential energy surface for the reaction between the electrophile and thiolate nucleophile. All scan calculations were performed using

an initial bond length of 2.9 Å between the halogenated carbon atom and the sulphur of the nucleophile. All calculations used methyl thiolate as a model nucleophile. A series of seven calculations were then carried out in which the bond length between the halogenated carbon and the sulphur of the thiolate nucleophile was decreased by 0.1 Å with each calculation. This mapped the reaction coordinate enabling the highest energy point corresponding to the transition state structure to be identified. All transition state structures were subjected to frequency analysis in order to identify a single negative eigenvalue connecting the transition state to the reactants and products on the potential energy surface. All calculations were carried out using the "opt=loose" keyword.

It is worth nothing some of the $\Delta E_{TS\text{-thiolate}}$ values obtained for some of the chemicals were negative (i.e. the transition state is lower in energy than the reactants). These values are unusual at first glance, as the $\Delta E_{TS\text{-Thiolate}}$ values were obtained from the highest point of energy along the reaction co-ordinate and have negative frequencies associated with them (indicating them to be true transition states). Given that the pKa of a thiolate ion is reported to be roughly 12.0 it is unlikely that this species will exist in a significant amounts at neutral pH. In reality, it is likely that the thiolate is stabilised by a positively charge counter ion such as sodium. With this in mind, a sodium ion was included in the calculation as part of the reactants and transition state structure. This resulted in a large increase in the $\Delta E_{TS\text{-Thiolate}}$ (kcal/mol) for all chemicals. This was something not immediately obvious in the analysis with the Michael acceptors and the analysis was carried out using the resonance-stabilised intermediate of the Michael addition reaction as opposed to the use of a transition state structure. However, in order to have a rational comparison with the Michael addition analysis, $\Delta E_{TS\text{-Thiolate}}$ without the addition of sodium was used for the S_N2 analysis. Although this has an effect on the $\Delta E_{TS\text{-Thiolate}}$ values, the values will be effected equally and therefore the predictive outcome of the values will not change.

Fragment analysis

All fragment development utilised the following set of rules (analogous to those utilised during the development of the previously published fragment-based *in silico* profiler for Michael addition thiol reactivity[19]:

1. All fragments were developed using the transition state structure upon reaction with a thiolate nucleophile using $\Delta E_{TS\text{-thiolate}}$ values as the key reactivity descriptor.
2. The $\Delta E_{TS\text{-thiolate}}$ values for straight chains at each R-position were compared with the $\Delta E_{TS\text{-thiolate}}$ values of a chain containing one carbon less (or in the case of methyl with hydrogen where applicable); for example, ethyl was compared to methyl and propyl compared to ethyl.
3. Branched chains $\Delta E_{TS\text{-thiolate}}$ values were compared to the $\Delta E_{TS\text{-thiolate}}$ value of their straight chain equivalent; for example, *t*-butyl was compared with ethyl.

4. Only ketones and esters contained aromatic substituents (at R_1). In all cases these were compared to a methyl group; for example, benzene, naphthalene, pyrene and thiophene were compared to methyl.

5. Only one R group was investigated at a time whilst the other R group remained constant. For example, R_1 remained as hydrogen whilst the effect of substituents at the R_2 position was investigated.

6. Individual calculated $\Delta E_{TS\text{-}thiolate}$ values were rounded to the nearest integer before comparing values rather than rounding the difference in $\Delta E_{TS\text{-}thiolate}$ between the two values.

7. A cut off value of 1.0 kcal/mol was used to assess if there was a significant difference between two substituents (to determine the need for the inclusion of a fragment in the profiler).

8. Unrounded $\Delta E_{TS\text{-}thiolate}$ values for fragments (to one decimal place) were used in the modelling of reactivity and toxicity.
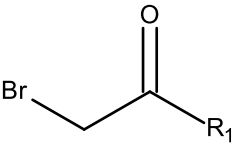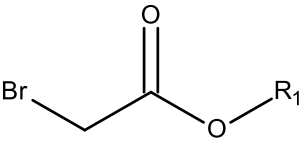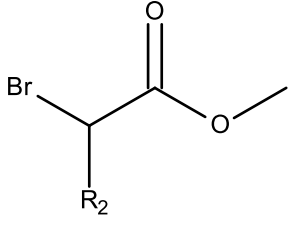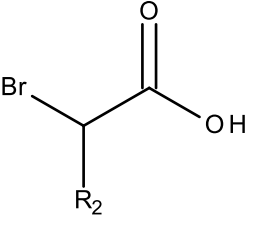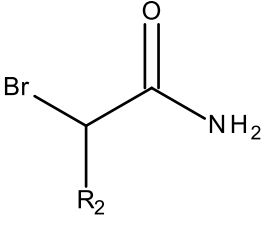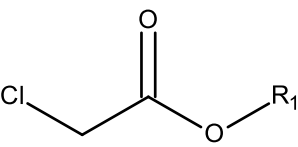
Statistical analysis

Linear regression analysis was used to develop quantitative structure-activity relationship models to obtain correlations between $\Delta E_{TS\text{-}Thiolate}$ values, predicted $-\log RC_{50}$ and toxicity to *Tetrahymena pyriformis* ($-\log IGC_{50}$) values using the Minitab (version 17) statistical software.

Results and Discussion

Inspection of the chemicals in the dataset showed that there were three factors that varied for the five types of electron-withdrawing groups present (ketones, esters, acids, amides and aromatic groups), these being; the halogen leaving group (bromine or chlorine) and varying substituents at the R_1 and R_2 positions (R-groups as defined in Table 1). Therefore, the analysis focused on the development of fragments capable of predicting the effects of these substituents on the calculated $\Delta E_{TS\text{-}thiolate}$ values within the domain of the experimental assay.

Table 1: Structures of the chemicals utilised in the SAR fragment analysis in the current study

R_1 = CH_2CH_3, *t*-butyl, phenyl, naphthalene, pyrene, thiophene	R_1 = CH_3, CH_2CH_3, $(CH_2)_2CH_3$, *t*-butyl, phenyl	
Brominated ketones and esters (N = 11)		

R_2 = H, CH_3, CH_2CH_3, $(CH_2)_2CH_3$	R_2 = CH_2CH_3, $(CH_2)_2CH_3$	R_2 = H, CH_3
Brominated esters, acids and amides (N = 8)		
R_1 = CH_3, CH_2CH_3, $(CH_2)_2CH_3$, *t*-butyl		
Chlorinated esters (N = 4)		

Development of fragments for brominated chemicals

Of the 28 chemicals in the dataset (shown in Table 6), 21 were brominated and seven were chlorinated. For the brominated chemicals this covered, 10 brominated esters, six brominated ketones, two brominated acids, two brominated amides and 1-bromomethyl-4-nitrobenzene. Analysis of the dataset revealed three groups that were large enough to allow a SAR analysis to be undertaken (11 chemicals with substituents at position R_1 for brominated ketones and esters, eight chemicals with substituents at position R_2 for brominated esters, acids and amides and four chemicals with substituents at position R_1 for chlorinated esters).

Calculated $\Delta E_{TS\text{-}thiolate}$ values SAR at position R_1 for brominated ketones and esters

Initially the SAR for $\Delta E_{TS\text{-}thiolate}$ values when extending the chain length at the R_1 position for brominated ketones and esters was investigated (chemicals as shown in Table 1). This group contained seven brominated ketones and 10 brominated esters covering seven and five substituents at the R_1 position respectively. Calculated $\Delta E_{TS\text{-}thiolate}$ values increased by 1.0 kcal/mol when extending the chain length from a methyl to an ethyl substituent at the R_1 position for both electron-withdrawing groups (compare chemical 2 with 1, and 9 with 8 in Table 2). Similarly, the change in calculated $\Delta E_{TS\text{-}thiolate}$ values was less than 1.0 kcal/mol when comparing ethyl to a *t*-butyl group for both electron-withdrawing groups (compare chemical 3 with 2, and 11 with 9 in Table 2). Additionally, methyl could be used for aromatic substituents benzene, naphthalene, pyrene and thiophene for ketones (compare chemicals 4-7 with 1 in

Table 2). The result of this analysis showed that a methyl group could be used to predict the $\Delta E_{TS\text{-thiolate}}$ values of all alkyl groups (and aryl groups for ketones) at the R_1 position. The exception being the need for an ethyl group for the prediction of chemicals where R_1 = t-butyl (compare chemical 11 with 3 in Table 2). This resulted in four fragments being required to cover the domain of ketones and esters with substituents at the R_1 position (these being R_1 = methyl and ethyl for both ketones and esters).

Table 2: Calculated $\Delta E_{TS\text{-thiolate}}$ (kcal/mol) values for brominated ketones and esters at the R_1 position

ID	Substituent name R_1	$\Delta E_{TS\text{-thiolate}}$ (kcal/mol)	Fragment substituent	Fragment $\Delta E_{TS\text{-thiolate}}$ (kcal/mol)
Ketones				
1	Methyl	-1.0	CH_3	-1.0
2	Ethyl	0.0	CH_3	-1.0
3	t-Butyl	0.0	CH_2CH_3	0.0
4	Benzene	-2.0	CH_3	-1.0
5	Naphthalene	-2.0	CH_3	-1.0
6	Pyrene	0.0	CH_3	-1.0
7	Thiophene	-1.0	CH_3	-1.0
Esters				
8	Methyl	0.0	CH_3	0.0
9	Ethyl	1.0	CH_3	0.0
10	Propyl	1.0	CH_3	0.0
11	t-Butyl	1.0	CH_2CH_3	1.0
12	Benzene	0.0	CH_3	0.0

Calculated $\Delta E_{TS\text{-thiolate}}$ values SAR at position R_2 for brominated esters, acids and amides.

The analysis for varying substituents at the R_2 position was applicable to three chemicals groups (chemicals as shown in Table 1). There were 11 brominated esters, two brominated acids and two brominated amides in the dataset covering four substituents for esters and two for both acids and amides at the R_2 position. The results showed that for all three electron-withdrawing groups, the calculated $\Delta E_{TS\text{-thiolate}}$ values differed significantly when going from hydrogen to methyl substituents (compare chemical 2 with 1, 6 with 5, and 10 with 9 in Table 3). This increase in the calculated $\Delta E_{TS\text{-thiolate}}$ value is expected due to increased steric bulk around the reactive site. However, increasing the chain length further from methyl to ethyl resulted in the calculated $\Delta E_{TS\text{-thiolate}}$ values being within 1.0 kcal/mol (compare chemical 3 with 2, and 7 with 6 in Table 3). This consistency in calculated $\Delta E_{TS\text{-thiolate}}$ values

was also seen when extending the chain length from ethyl to propyl (compare chemical 4 with 3, and 8 with 7 in Table 3). This showed that only the addition of the methyl group (going from primary to secondary halide) has an effect on calculated $\Delta E_{TS\text{-thiolate}}$ values, (increasing the chain length further resulted in no change in the calculated $\Delta E_{TS\text{-thiolate}}$ values). This resulted in two fragments being used to cover the brominated electron-withdrawing groups for R_2 substituents (this being R_2 = methyl and hydrogen). Although the groups for brominated acids and amides are small, it can be assumed that their applicability extends to acids and amides with larger substituents at the R_2 position. This assumption is based on consistency in calculated $\Delta E_{TS\text{-thiolate}}$ values when extending the chain length at the R_2 position for other chemical groups (e.g. brominated esters at the R_2 position – see chemicals 1-4 in Table 3).

Table 3: Calculated $\Delta E_{TS\text{-thiolate}}$ (kcal/mol) values for brominated esters, acids and amides at the R_2 position

ID	Substituent name R	$\Delta E_{TS\text{-thiolate}}$ (kcal/mol)	Fragment Substituent	Fragment $\Delta E_{TS\text{-thiolate}}$ (kcal/mol)
Esters				
1	Hydrogen	0.0	H	0.0
2	Methyl	5.0	CH$_3$	5.0
3	Ethyl	5.0	CH$_3$	5.0
4	Propyl	5.0	CH$_3$	5.0
Acids				
5	Hydrogen	0.0	H	0.0
6	Methyl	5.0	CH$_3$	5.0
7	Ethyl	4.0	CH$_3$	5.0
8	Propyl	4.0	CH$_3$	5.0
Amides				
9	Hydrogen	2.0	H	2.0
10	Methyl	7.0	CH$_3$	7.0

Brominated chemicals for which no SAR analysis was possible

Of the 21 chemicals there was only a single chemical activated by nitrobenzene group, this prevented a SAR analysis from being carried out for this class. Given this, a fragment was included for 1-bromomethyl-4-nitrobezene to assess the ability of the $\Delta E_{TS\text{-thiolate}}$ values for the prediction of glutathione reactivity and toxicity to *Tetrahymena pyriformis* for this chemical.

Calculated $\Delta E_{TS\text{-thiolate}}$ values SAR at position R_1 for chlorinated esters

The dataset contained seven chlorinated chemicals - four esters, one ketone, one acid and an amide. Given this, the only group for which a SAR analysis could be carried out for was the chlorinated esters at the R_1 position (R groups as defined in Table 1). This analysis resulted in the same outcome as was seen for the brominated esters, where no change in the $\Delta E_{TS\text{-thiolate}}$ values were calculated beyond a methyl substituent at the R_1 position (chemicals 1-4 in Table 4). This resulted in two fragments being required to cover the four chlorinated esters in the dataset (these being R_1 = methyl and ethyl).

Table 4: Calculated $\Delta E_{TS\text{-thiolate}}$ (kcal/mol) values for chlorinated esters at the R_1 position

ID	Substituent name R_1	$\Delta E_{TS\text{-thiolate}}$ (kcal/mol)	Fragment Substituent	Fragment $\Delta E_{TS\text{-thiolate}}$ (kcal/mol)
1	Methyl	3.0	CH_3	3.0
2	Ethyl	4.0	CH_3	3.0
3	Propyl	3.0	CH_3	3.0
4	*t*-Butyl	4.0	CH_2CH_3	3.0

Chlorinated chemicals for which no SAR analysis was possible

Of the seven chlorinated chemicals in the dataset, there were three chemicals for which no SAR could be carried out due to there being no other structurally related chemicals (1-chloropinacalone, 2-chloroacetamide and 2-chlorobutyric acid). The fragments used to define these chemicals are discussed in the next section.

Applicability domain of the fragment-based *in silico* profiler for S_N2 thiol reactivity

The above analysis resulted in the definition of ten fragments for the brominated chemicals and five fragments for the chlorinated chemicals (summarised in Table 5). All chemical classes for brominated chemicals showed that a methyl substituent was capable of predicting the $\Delta E_{TS\text{-thiolate}}$ values for alkyl and aryl substituents at both R-positions. The exception being the need for an ethyl group for *t*-butyl substituents for brominated ketones and esters. The SAR analysis for chlorinated esters at the R_1 position resulted in an identical outcome to that calculated for the brominated equivalents. Given this, an assumption was made that an analogous set of fragments to those defined for the brominated chemicals could be applied to extend the applicability domain of the profiler to cover an equivalent set of chlorinated chemicals. As such, the same set of R_1 and R_2 substituents were used to cover both brominated and chlorinated chemicals. This resulted in a total of 20 fragments to cover the expanded domain (fragments shown in italics in Table 5).

Table 5: The fragments required to cover the domain of chemicals in the dataset. Substituents used in the expanded domain for chlorinated chemicals are shown in italics.

Chemical group	Structure	X = Br	X = Cl
Ketones		R_1 = CH$_3$, CH$_2$CH$_3$ R_2 = H	R_1 = CH$_3$, *(CH$_2$CH$_3$)* R_2 = H
Esters		R_1 = CH$_3$, CH$_2$CH$_3$ R_2 = H, CH$_3$	R_1 = CH$_3$, CH$_2$CH$_3$ R_2 = H, *(CH$_3$)*
Acids		R_1 = N/A R_2 = H, CH$_3$	R_1 = N/A R_2 = *(H)*, CH$_3$
Amides		R_1 = N/A R_2 = H, CH$_3$	R_1 = N/A R_2 = H, *(CH$_3$)*
Aromatic group		R_1 = Nitrobenzene	R_1 = *(Nitrobenzene)*
Total Number of Fragments		N = 10	N = 10

Validation of the fragment-based *in silico* profiler for S$_N$2 thiol reactivity

The above analysis identified the need for 15 fragments to cover the structural domain of the 28 chemicals within the dataset. An additional five fragments were defined enabling the domain of the

chlorinated chemicals to be expanded to cover the equivalent chemical space as defined for the brominated chemicals. Using the $\Delta E_{TS\text{-thiolate}}$ values associated with each fragment as the independent variable in a linear regression analysis showed them to be capable of predicting glutathione reactivity (Model 1 and Figure 2, experimental and predicted values in Table 6).

$$-\text{predicted} \log RC_{50} = 1.03 - 0.34\,\Delta E_{TS-\text{Thiolate}} \qquad \text{(Model 1)}$$

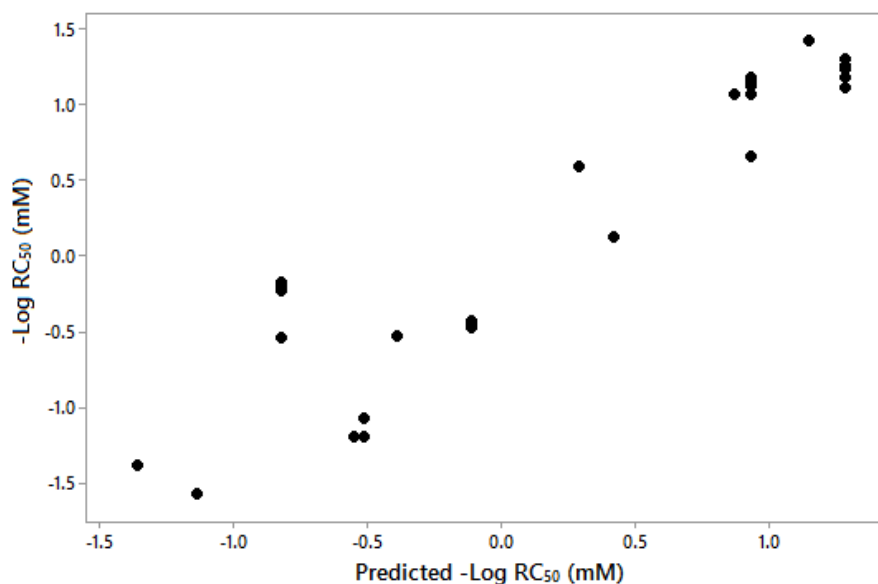$$N = 28,\ R^2 = 0.85,\ R^2\text{-adj} = 0.84,\ R^2\text{-pred} = 0.82,\ \text{Average error} = 0.31$$

Figure 2: Correlation between predicted and experimental –Log RC_{50} and for all 28 S_N2 chemicals

The predicted glutathione reactivity data values obtained from Model 1 (predicted –log RC_{50}) were subsequently utilised as the independent variable for the prediction of toxicity to *Tetrahymena pyriformis* (-log IGC_{50}). The resulting regression analysis is shown in Model 2 and Figure 3, with the statistics of this model being in keeping with previous research in which experimental glutathione reactivity data were used to predict toxicity to *Tetrahymena pyriformis*.[16]

$$-\log IGC_{50} = 1.27 + 1.00 - \text{predicted} \log RC_{50} \qquad \text{(Model 2)}$$

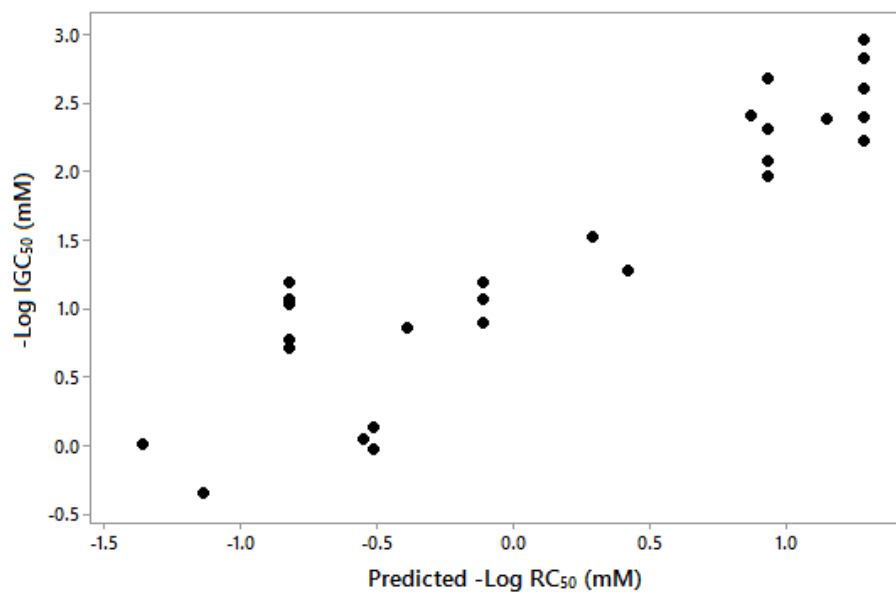$$N = 28,\ R^2 = 0.84,\ R^2\text{-adj} = 0.83,\ R^2\text{-pred} = 0.81,\ \text{Average error} = 0.31$$

Figure 3: Correlation between predicted –log RC$_{50}$ values from Model 1 and experimental –log IGC$_{50}$ values for the 28 chemicals in the dataset

Table 6: Chemicals acting via an S_N2 mechanism with corresponding –Log RC_{50}, -Log IGC_{50} (toxicity to *Tetrahymena pyriformis*), predicted –Log RC_{50} and –Log IGC_{50} values investigated in the current study. Chemicals names taken directly from reference (2). Fragment names are based on their IUPAC name

ID	Chemical	-Log RC_{50} (mM)	-Log IGC_{50} (mM)	Fragment	Predicted –Log RC_{50} (mM) (Model 1)	Predicted –Log IGC_{50} (mM) (Model 2)
1	1-Bromomethyl-4-nitrobenzene	0.66	2.30	1-Bromomethyl-4-nitrobenzene	0.93	2.20
2	2-(2-Bromoacetyl) thiophene	1.11	2.22	1-Bromo-2-propanone	1.28	2.55
3	1-Bromo-2-butanone	1.30	2.60	1-Bromo-2-propanone	1.28	2.55
4	1-Bromopinacolone	1.42	2.38	1-Bromo-2-butanone	1.14	2.41
5	1-Chloropinacolone	0.12	1.27	1-Chloro-2-butanone	0.42	1.69
6	2-Bromoacetophenone	1.26	2.82	1-Bromo-2-propanone	1.28	2.55
7	2-(2-Bromoacetyl) naphthalene	1.18	2.96	1-Bromo-2-propanone	1.28	2.55
8	1-(Bromoacetyl) pyrene	1.24	2.39	1-Bromo-2-propanone	1.28	2.55
9	Ethyl bromoacetate	1.07	2.68	Methyl bromoacetate	0.92	2.19
10	Ethyl chloroacetate	-0.48	1.06	Methyl chloroacetate	-0.12	1.15
11	Methyl bromoacetate	1.18	2.96	Methyl bromoacetate	0.92	2.19
12	Methyl chloroacetate	-0.45	0.89	Methyl chloroacetate	-0.12	1.15
13	Propyl bromoacetate	1.14	2.08	Methyl bromoacetate	0.92	2.19
14	Propyl chloroacetate	-0.43	1.18	Methyl chloroacetate	-0.12	1.15
15	*t*-Butyl bromoacetate	1.07	2.68	Ethyl bromoacetate	0.86	2.13
16	*t*-Butyl chloroacetate	-0.53	0.85	Ethyl chloroacetate	-0.40	0.87
17	Phenyl bromoacetate	1.14	2.08	Methyl bromoacetate	0.92	2.19
18	2-Bromoacetamide	0.59	1.52	2-Bromoacetamide	0.28	1.55
19	2-Chloroacetamide	-1.20	0.04	2-Chloroacetamide	-0.56	0.71

20	Methyl-2-bromopropionate	-0.18	1.18	Methyl-2-bromopropionate	-0.83	0.44
21	Methyl-2-bromobutyrate	-0.54	1.02	Methyl-2-bromopropionate	-0.83	0.44
22	2-Bromopropionamide	-1.40	0.00	2-Bromopropanamide	-1.36	-0.09
23	2-Bromobutyric acid	-1.20	0.12	2-Bromopropanoic acid	-0.52	0.75
24	2-Chlorobutyric acid	-1.58	-0.35	2-Chloropropanoic acid	-1.14	0.13
25	2-Bromovaleric acid	-1.08	-0.04	2-Bromopropanic acid	-0.52	0.75
26	Ethyl-2-bromovalerate	-0.23	0.70	Methyl-2-bromopropionate	-0.83	0.44
27	Ethyl-2-bromobutyrate	-0.20	0.77	Methyl-2-bromopropionate	-0.83	0.44
28	Ethyl-2-bromopropionate	-0.23	1.06	Methyl-2-bromopropionate	-0.83	0.44

Conclusions

The aim of this work was to develop a fragment-based profiler for S_N2 thiol reactivity by adopting a similar method that was previously successfully applied to the Michael addition domain. The results showed that the fragment-based *in silico* profiler was able to predict glutathione reactivity for a series of activated S_N2 chemicals. Additionally, the *in silico* reactivity predictions were shown to predict toxicity to *Tetrahymena pyriformis* for chemicals identified as acting via an S_N2 mechanism. The ability of the reactivity values derived from the fragment-based *in silico* profiler to predict toxicity to *Tetrahymena pyriformis* were in-keeping with previous research that demonstrated the use of experimental reactivity data to predict the same endpoint. The results of this study highlights the benefits of developing fragment-based reactivity profilers of this type, and their application in the prediction of toxicological endpoints for which the formation of a covalent bond is the key driver of potency.

Funding

Abbreviations

AOP – Adverse Outcome Pathway

DFT – Density Functional Theory

Eact – Energies of activation

LLNA – Local Lymph Node Assay

MIE – Molecular Initiating Event

QSAR – Quantitative Structure Activity Relationship

REACH – Registration, Evaluation, Authorisation, and restriction of CHemicals

SAS – Solvent Accessible Surface area

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) Przybylak, K., and Schultz, T. W. (2013) Informing chemical categories throguh the delopment of adverse outomce pathways, In *Chemical Toxicity Prediction: Category Formation and Read-across* pp 44-67, Royal Society of Chemistry.

(3) Ankley, G. T., Bennett, R. S., Erickson, R. J., Hoff, D. J., Hornung, M. W., Johnson, R. D., Mount, D. R., Nichols, J. W., Russom, C. L., Schmieder, P. K., Serrrano, J. A., Tietge, J. E., and Villeneuve, D. L. (2010) Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environ. Toxicol. Chem. 29*, 730-741.

(4) 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. *Crit. Rev. Toxicol. 41*, 783-802.

(5)   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.

(6)   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.

(7)   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.

(8)   Rodriguez-Sanchez, N., Schultz, T. W., Cronin, M. T. D., and Enoch, S. J. (2013) Experimental verification of structural alerts for the protein binding of cyclic compounds acting as Michael acceptors. *SAR QSAR Environ. Res. 24*, 963-977.

(9)   Bohme, A., Thaens, D., Schramm, F., Paschke, A., and Schuurmann, G. (2010) Thiol reactivity and its impact on the ciliate toxicity of α,β-unsaturated aldehydes, ketones, and esters. *Chem. Res. Toxicol. 23*, 1905-1912.

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

(11)   Mulliner, D., Wondrousch, D., and Schuurmann, G. (2011) Predicting Michael-acceptor reactivity and toxicity through quantum chemical transition-state calculations. *Org. Biomol. Chem. 9*, 8400-8412.

(12)   Schwobel, J. A. H., Wondrousch, D., Koleva, Y. K., Madden, J. C., Cronin, M. T. D., and Schuurmann, G. (2010) Prediction of Michael-type acceptor reactivity toward glutathione. *Chem. Res. Toxicol. 23*, 1576-1585.

(13)   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.

(14)   Schwobel, J. A. H., Madden, J. C., and Cronin, M. T. D. (2010) Examination of Michael addition reactivity towards glutathione by transition-state calculations. *SAR QSAR Environ. Res. 21*, 693-710.

(15)   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_N2 electrophiles to *Tetrahymena pyriformis*. *Chem. Res. Toxicol. 23*, 228-234.

(16)   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.

(17)   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.

(18)   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.

(19)   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.

(20)   Ebbrell, D. J., Madden, J. C., Cronin, M. T., Schultz, T. W., and Enoch, S. J. (2017) Validation of a Fragment-Based Profiler for Thiol Reactivity for the Prediction of Toxicity: Skin Sensitization and Tetrahymena pyriformis. *Chem. Res. Toxicol.* 30, 604-613

(21)   Mulliner, D., and Schuurmann, G. (2013) Model suite for predicting the aquatic toxicity of α,β-unsaturated esters triggered by their chemoavailability. *Mol. Inf. 32*, 98-107.

(22)   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.

(23)   Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, et al. (2009) Gaussian 09, revision A.1. Wallingford, CT