

Ebbrell, DJ, Madden, JC, Cronin, MTD, Schultz, T and Enoch, SJ

Development of a fragment-based in silico profiler for SN2 thiol reactivity and its application in predicting toxicity of chemicals towards *Tetrahymena pyriformis*

<http://researchonline.ljmu.ac.uk/id/eprint/11931/>

Article

Citation (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

**Ebbrell, DJ, Madden, JC, Cronin, MTD, Schultz, T and Enoch, SJ (2019)
Development of a fragment-based in silico profiler for SN2 thiol reactivity
and its application in predicting toxicity of chemicals towards *Tetrahymena
pyriformis*. *Computational Toxicology*. 13. ISSN 2468-1113**

LJMU has developed [LJMU Research Online](#) for users to access the research output of the University more effectively. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LJMU Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain.

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact researchonline@ljmu.ac.uk

<http://researchonline.ljmu.ac.uk/>

Development of a fragment-based *in silico* profiler for SN2 thiol reactivity and its application in predicting toxicity of chemicals towards *Tetrahymena pyriformis*

David Ebbrell, Judith Madden, Mark Cronin, Terry Schultz, Steve Enoch

Keywords

Bimolecular nucleophilic substitution

Thiol reactivity

Glutathione reactivity

Tetrahymena pyriformis

Density functional theory

1.1 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 the 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 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 here]

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 linearly 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. methane 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, 20].

A recent study by the current authors showed it was possible to predict experimental reactivity (expressed as pRC₅₀ 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 [20]. 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 calculate -log RC₅₀ 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). (21) 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 [11, 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's 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 largest group of chemicals investigated in a previous study by Roberts et al (S_N2 compounds activated by a carbonyl electron-withdrawing group) [15]. This was achieved by adopting a similar method as applied previously to Michael acceptors [20]. 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. 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.

2.0 Methods

2.1 Data set

Thirty-two chemicals were identified as being within the defined S_N2 domain (S_N2 compounds activated by a carbonyl electron-withdrawing group) from the literature [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 29 activated S_N2 chemicals. All chemicals in the dataset had associated glutathione reactivity data (pRC_{50}) and *Tetrahymena pyriformis* toxicity data ($pIGC_{50}$) [10]. The full experimental dataset, including SMILES strings, can be found in the supplementary information.

2.2 Computational methods – analysis of S_N2 reaction profile and reactivity descriptor selection

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 $\Delta E_{React-TS}$ 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. This is in keeping with previous research into the development of a fragment-based method for Michael addition reactivity [20]. 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 methane 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. The “loose” keyword maximises the step sized required to reach convergence for optimization typically reducing the amount of time required to run the calculation. Values for $\Delta E_{\text{React-TS}}$ for secondary halides were obtained using their S-isomer geometries. This isomer was selected as a set of test calculations showed there to be less than 1.0 kcal/mol difference between the two isomers (data not shown). Given this, a single geometrical isomer was selected in order to simplify the development of the fragment-based method. All calculated values can be found in the supplementary information.

2.3 Statistical analysis

Linear regression analysis was used to develop quantitative structure-activity relationship models to obtain correlations between $\Delta E_{\text{React-TS}}$ values, predicted $-\log \text{RC}_{50}$ and toxicity to *Tetrahymena pyriformis* (pIGC₅₀) values using the Minitab (version 17) statistical software. This was also carried out to obtain appropriate statistical values for models (R^2 , R^2 -adj and LOO validation / R^2 -pred values).

3.0 Results and Discussion

The aim of this study was to extend an existing fragment-based *in silico* profiler capable of predicting chemical reactivity for Michael acceptors to a set of S_N2 compounds activated by a carbonyl electron-withdrawing group. The key advantage of this approach being the development of a fragment-based library of computational reactivity values that remove the need to perform time-consuming DFT calculations (that require specialist software).

3.1 Analysis of S_N2 transition states

The key aspect of the fragment-based *in silico* profiler for S_N2 is the use of a database containing fragments with pre-calculated $\Delta E_{\text{React-TS}}$ values. Upon inspection of the S_N2 transition states, a number of the $\Delta E_{\text{React-TS}}$ 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 $\Delta E_{\text{React-TS}}$ values were obtained from the highest energy point along the reaction co-ordinate and have negative frequency associated with them (indicating them to be true transition states). These negative $\Delta E_{\text{React-TS}}$ values can be explained by the existence of a pre-reaction complex on the potential energy surface that was identified as being lower in energy than the transition state. For example, the smallest brominated

fragment in the dataset (1-bromopropan-2-one) has a pre-reaction complex that is 1.39 kcal/mol lower in energy than the reactants ($\Delta E_{\text{React-PreRC}}$ see Table 1 and Figure 2). This makes the energy difference between the pre-reaction complex and the transition state ($\Delta E_{\text{PreRC-TS}}$) 0.70 kcal/mol. This resulted in a calculated $\Delta E_{\text{React-TS}}$ of -0.69 kcal/mol (see Table 1 and Figure 2). This analysis is supported by studies on the S_N2 energy profile showing it to feature a single transition state and two minima, corresponding to pre- and post-reaction complexes [24-26]. Although $\Delta E_{\text{PreRC-TS}}$ is a closer representation of the activation energy, $\Delta E_{\text{React-TS}}$ was selected as the descriptor with which to develop the fragment-based profiler for S_N2 thiol reactivity. This was due to the fact that the calculation of $\Delta E_{\text{PreRC-TS}}$ required significantly more computational effort compared to the calculation of $\Delta E_{\text{React-TS}}$ – an important consideration when developing the fragment database.

[Table 1 here]

[Figure 2 here]

3.2 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 [20]:

1. All fragments were developed using the transition state structure upon reaction with a thiolate nucleophile using $\Delta E_{\text{React-TS}}$ values as the key reactivity descriptor.
2. The $\Delta E_{\text{React-TS}}$ values for straight chains at each R-position were compared with the $\Delta E_{\text{React-TS}}$ value for methyl (or in the case of methyl with hydrogen where applicable)
3. Branched chains $\Delta E_{\text{React-TS}}$ values were compared to the $\Delta E_{\text{React-TS}}$ value for methyl to investigate the importance of steric bulk.
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. 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). This cut-off value was identified as being the variability in the computationally determined reactivity values during the development of a fragment-based profiler for Michael addition [20].
7. Activation energy values for fragments to two decimal places were used in the modelling of reactivity and toxicity. These were defined as $\Delta E_{\text{React-TS-fragment}}$.

Inspection of the chemicals in the dataset showed that there were three factors that varied for the four types of carbonyl electron-withdrawing groups present (ketones, esters, acids and amides), these being; the halogen leaving group (bromine or chlorine) and varying substituents at the R₁ and/or R₂ positions (R-groups as defined in Table 2). Therefore, the analysis focused on the development of fragments capable of predicting the effects of these substituents on the calculated $\Delta E_{\text{React-TS}}$ values within the domain of the experimental assay.

[Table 2 here]

3.3 Development of fragments for brominated chemicals

Of the 29 chemicals in the dataset, 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 R₁ for brominated ketones and esters, eight chemicals with substituents at position R₂ for brominated esters, acids and amides and four chemicals with substituents at position R₁ for chlorinated esters).

3.4 Calculated $\Delta E_{\text{React-TS}}$ values SAR at position R₁ for brominated ketones and esters

Initially the SAR for $\Delta E_{\text{React-TS}}$ values when extending the chain length at the R₁ position for brominated ketones and esters was investigated (chemicals as shown in Table 2). This group contained seven brominated ketones and 10 brominated esters covering seven and five substituents at the R₁ position respectively. Calculated $\Delta E_{\text{React-TS}}$ values increased by 1.0 kcal/mol when extending the chain length from a methyl to an ethyl substituent at the R₁ position for both carbonyl electron-withdrawing groups (compare chemical 2 with 1, and 9 with 8 in Table 3). A methyl group was also calculated to be a suitable fragment for a *t*-butyl group when the electron-withdrawing group was an ester (compare chemicals 8 and 11 in Table 3). However, this was not the case when the electron-withdrawing group was a ketone (compare chemicals 1 and 3 in Table 3). Inspection of the transition state structures for these chemicals showed the presence of the oxygen linker in the ester group significantly reduced the steric hindrance around the reactive centre (compared to the ketone). Additionally, methyl could be used for aromatic substituents benzene and thiophene for ketones (compare chemicals 4 and 7 with 1 in Table 3). This was not the case for naphthalene or pyrene, leading to the need to define a fragment for this functional group (compare chemicals 5, 6 and 1 in Table 3). The result of this analysis showed that five fragments were required to cover the domain of ketones and esters with substituents at the R₁ position (these being R₁ = methyl, *t*-butyl, naphthalene and pyrene for ketones and R₁ = methyl for esters).

[Table 3 here]

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

The analysis for varying substituents at the R_2 position was applicable to four chemical groups (brominated ketones, esters, acids and amides, chemicals as shown in Table 2). There were seven brominated ketones, 11 brominated esters, two brominated acids and two brominated amides in the dataset covering four substituents for esters and two for ketones acids and amides at the R_2 position. The results showed that for all four carbonyl electron-withdrawing groups, the calculated $\Delta E_{\text{React-TS}}$ values differed significantly when going from hydrogen to methyl substituents (compare chemical 2 with 1, 3 with 4, 7 with 9 and 11 with 12 in Table 4). This increase in the calculated $\Delta E_{\text{React-TS}}$ 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 $\Delta E_{\text{React-TS}}$ values being within 1.0 kcal/mol (compare chemical 5 with 4, and 9 with 8 in Table 4). This consistency in calculated $\Delta E_{\text{React-TS}}$ values was also seen when extending the chain length from ethyl to propyl (compare chemical 6 with 5, and 10 with 9 in Table 4). This showed that only the addition of the methyl group (going from primary to secondary halide) has an effect on calculated $\Delta E_{\text{React-TS}}$ values, (increasing the chain length further resulted in no change in the calculated $\Delta E_{\text{React-TS}}$ values). This resulted in two fragments being used to cover the brominated carbonyl electron-withdrawing groups for R_2 substituents (this being R_2 = methyl and hydrogen). Although the groups for brominated ketones, 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 $\Delta E_{\text{React-TS}}$ 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 3-6 in Table 4).

[Table 4 here]

3.6 Calculated $\Delta E_{\text{React-TS}}$ values SAR at position R_1 for chlorinated esters

The dataset contained seven chlorinated chemicals - four esters, two ketones, 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 2). This analysis resulted in the same outcome as was seen for the brominated esters, where no change in the $\Delta E_{\text{React-TS}}$ values were noted beyond a methyl substituent at the R_1 position (chemicals 1-4 in Table 5). This resulted in a single fragment being required to cover the four chlorinated esters in the dataset (R_1 = methyl).

[Table 5 here]

Chlorinated chemicals for which no SAR analysis was possible

Of the eight chlorinated chemicals in the dataset, there were four chemicals for which no SAR could be carried out due to there being no other structurally related chemicals (3-chloro-2-butanone, 1-chloropinacalone, 2-chloroacetamide and 2-chlorobutyric acid). The fragments used to define these chemicals are discussed in the next section.

3.7 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 (For a total of 15 fragments, shown as non-italicised substituents in Table 6). Note that a fragment for brominated acids with a hydrogen substituted at the R_2 position was also included to expand the applicability domain of the method, although this fragment was not required to model any of the chemicals in the dataset. All chemical classes for brominated chemicals showed that a methyl substituent was capable of predicting the $\Delta E_{\text{React-TS}}$ values for alkyl and aryl substituents at both R -positions. 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 26 fragments to cover the expanded domain (fragments shown in italics in Table 6).

[Table 6 here]

3.8 Validation of the fragment-based in silico profiler for S_N2 thiol reactivity

The above analysis identified the need for 18 fragments to cover the structural domain of the 29 chemicals within the dataset. An additional eight 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. In order for the fragment library of calculated energy values ($\Delta E_{\text{React-TS-fragment}}$) to be useful in computational toxicology they must be at least equally predictive of reactivity and toxicity when compared to utilising the non-fragment energy values ($\Delta E_{\text{React-TS}}$). This analysis showed that both $\Delta E_{\text{React-TS}}$ and $\Delta E_{\text{React-TS-fragment}}$ values were capable of predicting glutathione reactivity and toxicity to *Tetrahymena pyriformis* for the 29 chemicals within the dataset (correlations shown in Figures 3 and 4, summary statistics shown in Table 7. Importantly, the fragment-based approach offers a clear benefit in terms of the ability to predict reactivity and toxicity to *Tetrahymena pyriformis* without the need for time consuming quantum mechanical calculations. Predicted values for all chemicals can be found in the supplementary information). These results are also in keeping with the published fragment-based profiler for the prediction of Michael acceptor reactivity [20]. Finally, it is worth noting that the $\Delta E_{\text{React-TS}}$ values for the three chemicals featuring an acid moiety as the electron-withdrawing

group (2-bromobutyric acid, 2-bromovaleric acid and 2-chlorobutyric acid) were calculated in their protonated form. However, these chemicals are likely to be mostly ionised under experimental conditions (pH = 7.4). Interestingly, utilising the ionised form of these chemicals resulted in a large increase in $\Delta E_{\text{React-TS}}$ which made the prediction of reactivity for these chemicals significantly worse (for example, the $\Delta E_{\text{React-TS}}$ value for 2-chlorobutyric acid increased from 6.38 kcal/mol to 22.24 kcal/mol). This is likely due to the use of an implicit solvation model to model the effect of water, which fails to capture the solvation of the carboxylate anion accurately (especially when compared to non-ionised chemicals). Given this, the fragment-based profiler utilised protonated fragments for acid-based electron withdrawing groups.

[Figure 3 here]

[Figure 4 here]

[Table 7 here]

4.0 Conclusions

The aim of this work was to develop a fragment-based profiler for S_N2 thiol reactivity for compounds activated by a carbonyl electron-withdrawing group 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 and toxicity to *Tetrahymena pyriformis* for a series of S_N2 chemicals activated by a carbonyl electron-withdrawing group. Importantly, this study shows that such calculated values are as predictive of reactivity and toxicity as the equivalent energy values derived from complete chemical structures. The availability of such a fragment-based library removes the need to perform time-consuming DFT calculations in computational toxicology. The fragment library also enables reactivity and toxicity of novel S_N2 compounds (i.e. those not covered included in the dataset) to be predicted using the QSAR models outlined in this study. It is anticipated that the approach outlined in this study (and previous studies) will lead to the fragment-based approach for predicting chemical reactivity being incorporated into key computational tools utilized in regulatory toxicology. The availability of a fragment-library will enable quantitative structure-activity models to be constructed for the prediction of reactivity and toxicity where covalent bond formation is the key driver.

Acknowledgments

The research in this manuscript was funded in part by the 2013 LUSH prize for cosmetics and a research bursary to David Ebbrell from Liverpool John Moores University.

Abbreviations

AOP – Adverse Outcome Pathway

DFT – Density Functional Theory

Eact – Energy of activation

$\Delta E_{\text{React-Pre-reaction complex}}$ – Energy difference between the reactants and pre-reaction complex

$\Delta E_{\text{Pre-reaction complex – TS}}$ – Energy difference between pre-reaction complex and transition state structure

$\Delta E_{\text{TS – Post-reaction complex}}$ – Energy difference between transition state structure and post-reaction complex

$\Delta E_{\text{Post-reaction complex – Products}}$ – Energy difference between the post-reaction complex and products

$\Delta E_{\text{React-TS}}$ – Energy difference between reactants and transition state structure calculated for full chemical structure

$\Delta E_{\text{React-TS-fragment}}$ – Energy difference between reactants and transition state structure for the fragment structures

GSH - Glutathione

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

S_N1 – Unimolecular nucleophilic substitution

S_N2 – Bimolecular nucleophilic substitution

TS – Transition state

References

1. 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. doi:10.1021/Tx0601004
2. Przybylak K, Schultz TW. Informing Chemical Categories through the Development of Adverse Outcome Pathways. *Chemical Toxicity Prediction: Category Formation and Read-across*. Issues in Toxicology: Royal Society of Chemistry; 2013. p. 44-67. doi:10.1039/9781849734400-00044
3. 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. doi: 10.1002/Etc.34
4. Enoch SJ, Ellison CM, Schultz TW, Cronin MTD. A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity. *Critical Reviews in Toxicology.* 2011;41(9):783-802. doi: 10.3109/10408444.2011.598141
5. 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. doi: 10.1080/10659360500204152
6. Schultz TW, Yarbrough JW, Hunter RS, Aptula AO. Verification of the structural alerts for Michael acceptors. *Chem Res Toxicol.* 2007;20(9):1359-63. doi:10.1021/Tx700212u
7. Bajot F, Cronin MT, 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. doi:10.1080/1062936X.2010.528449
8. Rodriguez-Sanchez N, Schultz TW, Cronin MTD, Enoch SJ. Experimental verification of structural alerts for the protein binding of cyclic compounds acting as Michael acceptors. *Sar Qsar Environ Res.* 2013;24(11):963-77. doi:10.1080/1062936x.2013.820793

9. Bohme A, Thaens D, Schramm F, Paschke A, Schuurmann G. Thiol Reactivity and Its Impact on the Ciliate Toxicity of α,β -Unsaturated Aldehydes, Ketones, and Esters. *Chem Res Toxicol*. 2010;23(12):1905-12. doi:10.1021/Tx100226n
10. Yarbrough JW, Schultz TW. Abiotic sulfhydryl reactivity: A predictor of aquatic toxicity for carbonyl-containing α,β -unsaturated compounds. *Chem Res Toxicol*. 2007;20(3):558-62. doi:10.1021/Tx600344a
11. Mulliner D, Schuurmann G. Model Suite for Predicting the Aquatic Toxicity of α,β -Unsaturated Esters Triggered by Their Chemoavailability. *Mol Inform*. 2013;32(1):98-107. doi:10.1002/minf.201200101
12. Schwobel 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. doi:10.1021/Tx100172x
13. 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. doi:10.1021/Tx4000655
14. Schwobel 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. doi:10.1080/1062936x.2010.528943
15. 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. doi:10.1021/Tx90003648
16. Schultz TW, Ralston KE, Roberts DW, Veith GD, Aptula AO. Structure-activity relationships for abiotic thiol reactivity and aquatic toxicity of halo-substituted carbonyl compounds. *Sar Qsar Environ Res*. 2007;18(1-2):21-9. doi:10.1080/10629360601033424
17. Schultz TW, Sparfkin CL, Aptula AO. Reactivity-based toxicity modelling of five-membered heterocyclic compounds: Application to *Tetrahymena pyriformis*. *Sar Qsar Environ Res*. 2010;21(7-8):681-91. doi:10.1080/1062936x.2010.52893
18. 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. doi:10.1021/tx700322g
19. 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. doi:10.1039/C1ob06065a
20. 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. doi:10.1021/acs.chemrestox.6b00099
21. Ebbrell DJ, Madden JC, Cronin MT, Schultz TW, Enoch SJ. Validation of a Fragment-Based Profiler for Thiol Reactivity for the Prediction of Toxicity: Skin Sensitization and *Tetrahymena pyriformis*. *Chem Res Toxicol*. 2017. doi:10.1021/acs.chemrestox.6b00361
22. 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. doi:10.1021/tx800431x
23. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, et al. Gaussian 09, revision A.1. Wallingford, CT.; 2009.
24. Liu X, Zhang JX, Yang L, Hase WL. How a Solvent Molecule Affects Competing Elimination and Substitution Dynamics. Insight into Mechanism Evolution with Increased Solvation. *J Am Chem Soc*. 2018;140(35):10995-1005. doi:10.1021/jacs.8b04529
25. Manikandan P, Zhang JX, Hase WL. Chemical Dynamics Simulations of $X^- + CH_3Y \rightarrow XCH_3 + Y^-$ Gas-Phase $S(N)2$ Nucleophilic Substitution Reactions. Nonstatistical Dynamics and Nontraditional Reaction Mechanisms. *J Phys Chem A*. 2012;116(12):3061-80. doi:10.1021/jp211387c
26. Yang SY, Fleurat-Lessard P, Hristov I, Ziegler T. Free energy profiles for the identity $S(N)2$ reactions $Cl^- + CH_3Cl$ and $NH_3 + H_3BNH_3$: A constraint ab initio molecular dynamics study. *J Phys Chem A*. 2004;108(43):9461-8. doi:10.1021/jp046954j

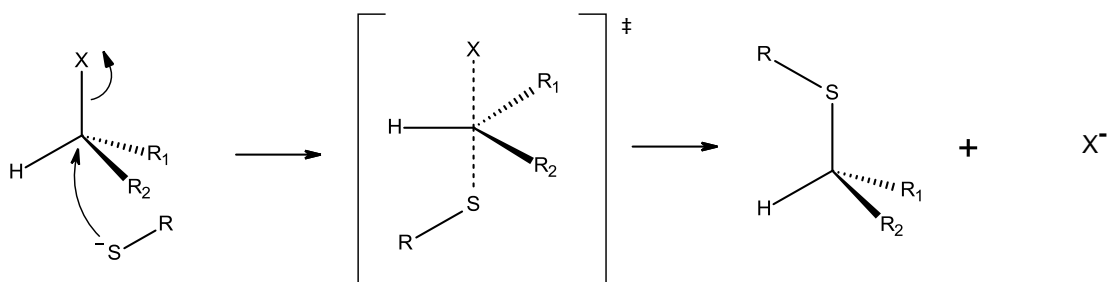


Figure 1: The formation of a covalent bond between an electrophilic chemical and cysteine via an S_N2 mechanism (X = halogen, R = alkyl, R₁ = R₂ = hydrogen or carbon)

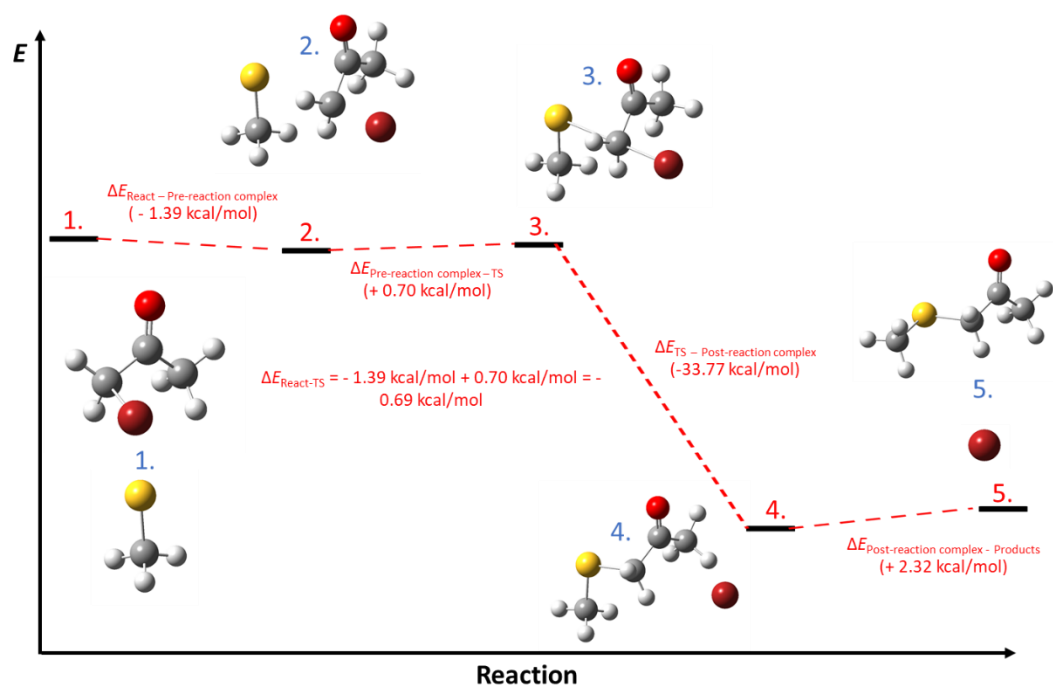


Figure 2. Energy profile for the reaction between methane thiolate and 1-bromopropan-2-one, see Table 1 for energy values and bond lengths

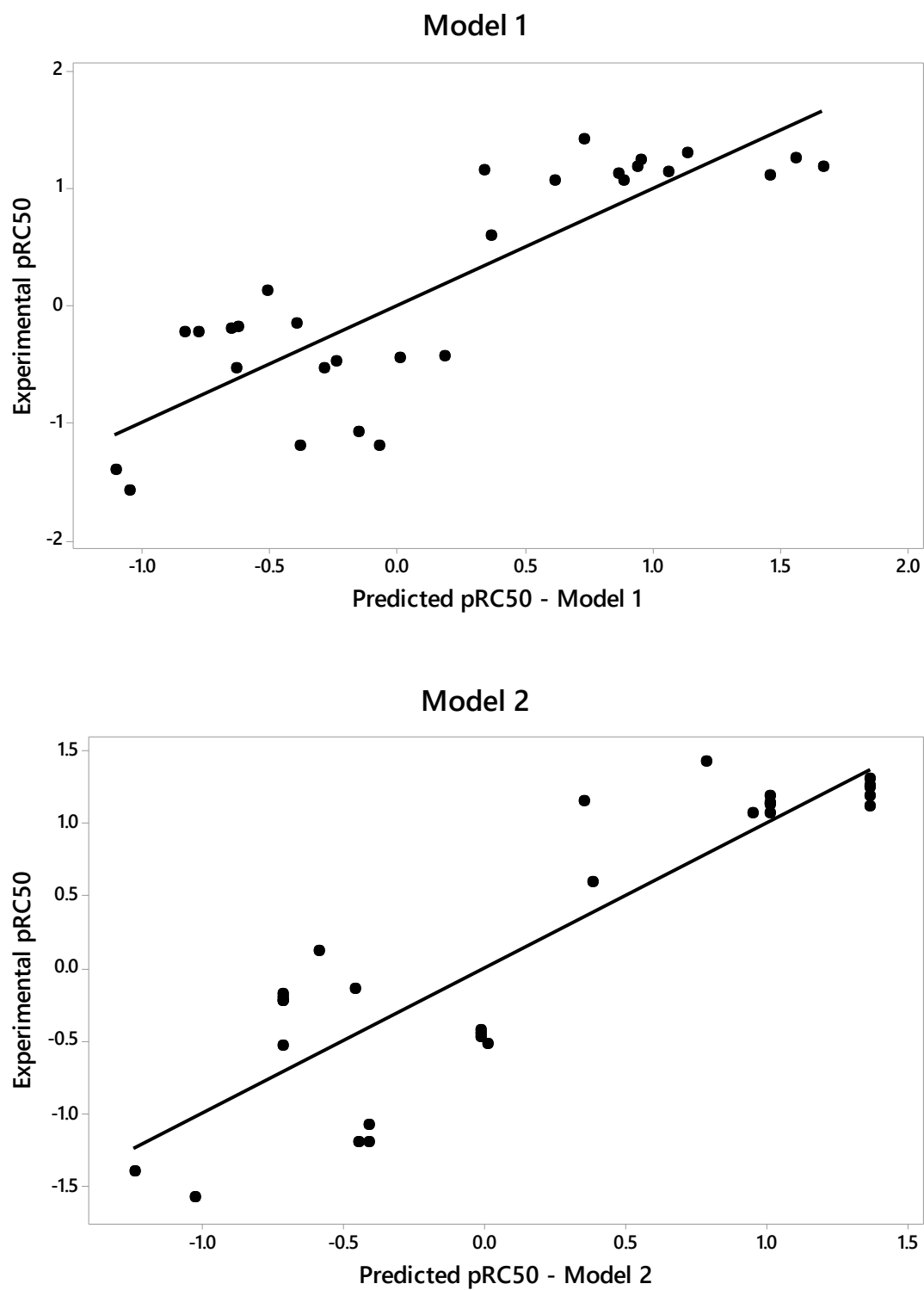


Figure 3: Correlations between predicted pRC_{50} values calculated using $\Delta E_{\text{React-TS}}$ and experimental values (model 1) and predicted pRC_{50} values calculated using $\Delta E_{\text{React-TS-fragment}}$ and experimental values (model 2) for all 29 carbonyl electron-withdrawing $\text{S}_{\text{N}}2$ chemicals.

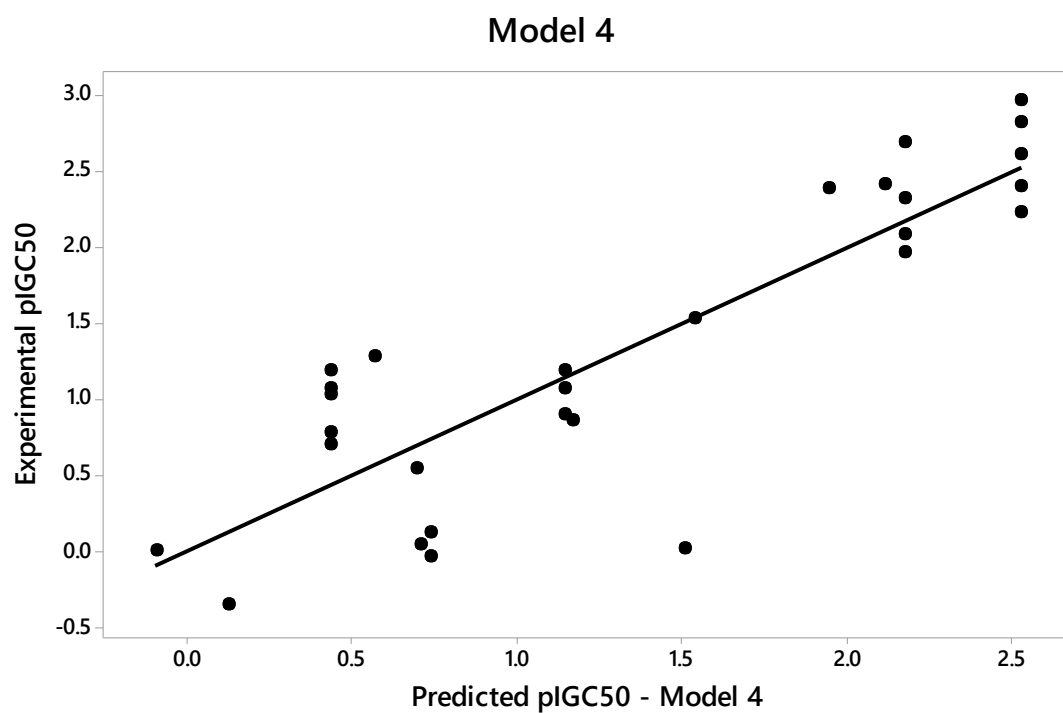
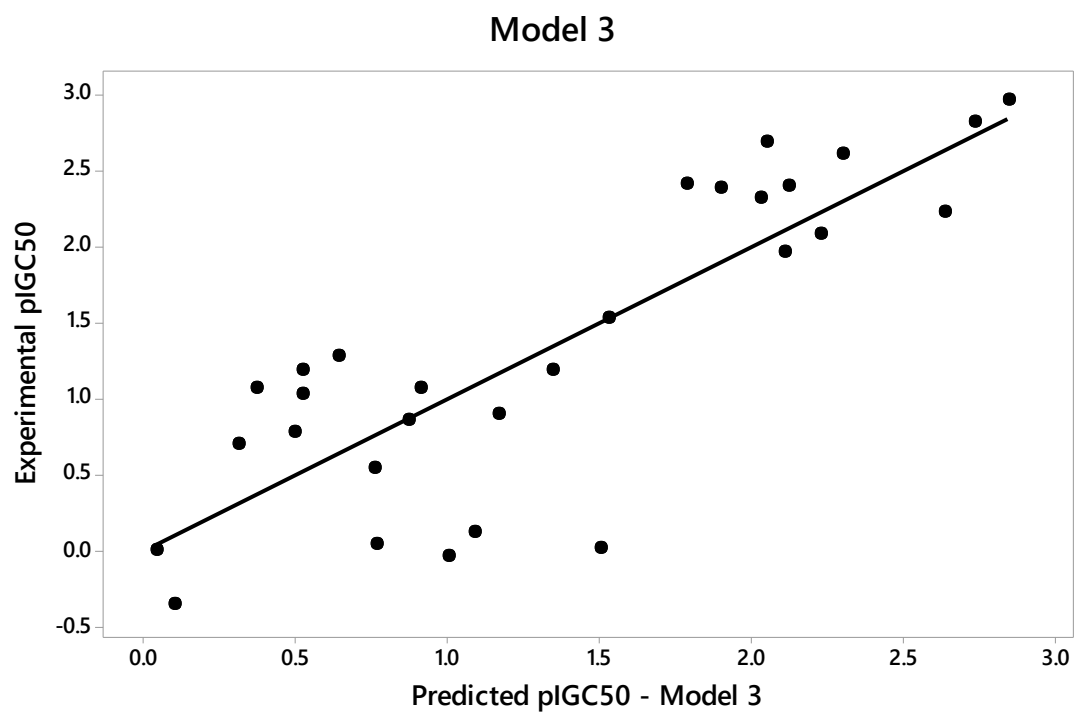
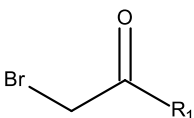
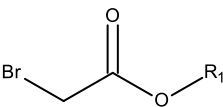
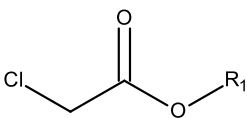
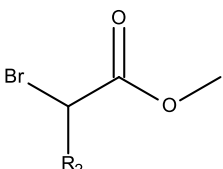
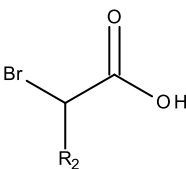
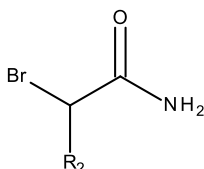
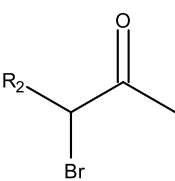
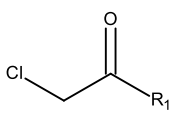


Figure 4: Correlations between predicted pIGC₅₀ values calculated using $\Delta E_{\text{React-TS}}$ and experimental values (model 3) and predicted pIGC₅₀ values calculated using $\Delta E_{\text{React-TS-fragment}}$ and experimental values (model 4) for all 29 carbonyl electron-withdrawing S_N2 chemicals.

Table 1: Energy values and bond lengths of the species for the reaction between methane thiolate and 1-bromopropan-2-one (* ΔE refers to the change in energy from the previous species / step in the reaction sequence (i.e. 1 \rightarrow 2, 2 \rightarrow 3, etc.). Refer to Figure 2 for energy diagram and optimised geometries)

Species number	Species name	C-Br (Å)	C-S (Å)	ΔE^* (kcal/mol)
1	Reactants (React)	2.0	N/A	N/A
2	Pre-reaction complex	2.0	3.2	-1.39
3	Transition State (TS)	2.2	2.7	0.70
4	Post-reaction complex	3.8	1.9	-33.77
5	Products	N/A	1.8	2.32

Table 2: Structures of the chemicals utilised in the SAR fragment analysis in the current study (*no SAR analysis was possible for these groups as they only contained one compound)

Groups where SAR analysis was possible		
 <p>$R_1 = \text{CH}_2\text{CH}_3, t\text{-butyl, phenyl, naphthalene, pyrene, thiophene}$</p> <p>$R_1$ – Brominated Ketones (N = 6)</p>	 <p>$R_1 = \text{CH}_3, \text{CH}_2\text{CH}_3, (\text{CH}_2)_2\text{CH}_3, t\text{-butyl, phenyl}$</p> <p>$R_1$ – Brominated Esters (N = 5)</p>	 <p>$R_1 = \text{CH}_3, \text{CH}_2\text{CH}_3, (\text{CH}_2)_2\text{CH}_3, t\text{-butyl}$</p> <p>$R_1$ – Chlorinated Esters (N = 4)</p>
 <p>$R_2 = \text{H, CH}_3, \text{CH}_2\text{CH}_3, (\text{CH}_2)_2\text{CH}_3$</p> <p>$R_2$ – Brominated Esters (N = 4)</p>	 <p>$R_2 = \text{CH}_2\text{CH}_3, (\text{CH}_2)_2\text{CH}_3$</p> <p>$R_2$ – Brominated Acids (N = 2)</p>	 <p>$R_2 = \text{H, CH}_3$</p> <p>R_2 Brominated Amides (N = 2)</p>
 <p>$R_2 = \text{H, CH}_3$</p> <p>R_2 – Brominated Ketones (N = 2)</p>		
Groups where SAR analysis was not possible*		
		

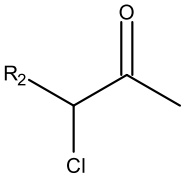
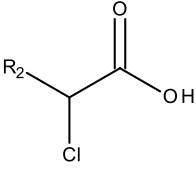
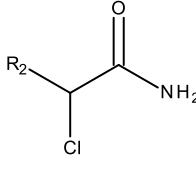
$R_1 = t\text{-butyl}$		
R_1 Chlorinated Ketones (N = 1)		
 $R_2 = \text{CH}_3$	 $R_2 = \text{CH}_3$	 $R_2 = \text{H}$
R_2 Chlorinated Ketones (N = 1)	R_2 – Chlorinated Acids (N = 1)	R_2 – Chlorinated Amides (N = 1)

Table 3: Calculated $\Delta E_{\text{React-TS}}$ (kcal/mol) values for brominated ketones and esters at the R_1 position, where fragment substituent represents the substitution at the respective R-position. Where appropriate, $\Delta\Delta E_{\text{React-TS}}$ values show the energy difference between the fragment value and the full substituent using the rounded integer data.

ID	R_1 Substituent	$\Delta E_{\text{React-TS}}$ (kcal/mol)	Fragment substituent	Fragment $\Delta\Delta E_{\text{React-TS}}$ (kcal/mol)
Ketones				
1	Methyl	-0.7	CH_3	-
2	Ethyl	-0.3	CH_3	+0.4
3	<i>t</i> -Butyl	1.0	<i>t</i> -Butyl	+1.3
4	Benzene	-1.7	CH_3	-1.0
5	Naphthalene	-2.0	Naphthalene	0.0
6	Pyrene	1.3	Pyrene	+2.0
7	Thiophene	-1.3	CH_3	+0.5
Esters				
8	Methyl	0.4	CH_3	-
9	Ethyl	0.5	CH_3	+0.1
10	Propyl	0.6	CH_3	+0.2
11	<i>t</i> -Butyl	1.4	CH_3	+1.0
12	Benzene	0.0	CH_3	+0.4

Table 4: Calculated $\Delta E_{\text{React-TS}}$ (kcal/mol) values for brominated esters, acids and amides at the R_2 position, where fragment substituent represents the substitution at the respective R-position. Where appropriate, $\Delta\Delta E_{\text{React-TS}}$ values show the energy difference between the fragment value and the full substituent using the rounded integer data. Chemicals denoted with a star are not in the experimental dataset and are included for comparison to allow the fragment library to be defined.

ID	R_2 substituent	$\Delta E_{\text{React-TS}}$ (kcal/mol)	Fragment Substituent	Fragment $\Delta\Delta E_{\text{React-TS}}$ (kcal/mol)
Ketones				
1	Hydrogen	-0.7	H	-
2	Methyl	2.3	CH ₃	-
Esters				
3	Hydrogen	0.4	H	-
4	Methyl	5.5	CH ₃	-
5	Ethyl	5.3	CH ₃	-0.2
6	Propyl	5.5	CH ₃	0.0
Acids				
7	Hydrogen*	-0.4	H	-
8	Methyl*	4.6	CH ₃	-
9	Ethyl	4.0	CH ₃	-0.6
10	Propyl	3.7	CH ₃	-0.9
Amides				
11	Hydrogen	2.2	H	-
12	Methyl	7.0	CH ₃	-

Table 5: Calculated $\Delta E_{\text{React-TS}}$ (kcal/mol) values for chlorinated esters at the R_1 position, where fragment substituent represents the substitution at the respective R-position. Where appropriate, $\Delta\Delta E_{\text{React-TS}}$ values show the energy difference between the fragment value and the full substituent using the rounded integer data.

ID	R_1 substituent	$\Delta E_{\text{React-TS}}$ (kcal/mol)	Fragment Substituent	Fragment $\Delta\Delta E_{\text{React-TS}}$ (kcal/mol)
1	Methyl	3.3	CH ₃	-
2	Ethyl	4.2	CH ₃	+0.9
3	Propyl	2.8	CH ₃	+0.5
4	<i>t</i> -Butyl	4.3	CH ₃	+1.0

Table 6: 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.

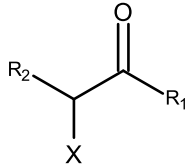
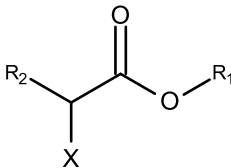
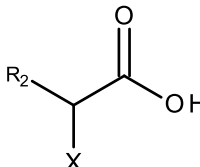
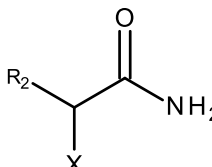
Chemical group	Structure	X = Br	X = Cl
Ketones		R ₁ = CH ₃ , <i>t</i> -Butyl, naphthalene, pyrene R ₂ = H, CH ₃	R ₁ = (<i>CH</i> ₃), <i>t</i> -Butyl, <i>naphthalene, pyrene</i> R ₂ = H, (<i>CH</i> ₃)
Esters		R ₁ = CH ₃ R ₂ = H, CH ₃	R ₁ = CH ₃ R ₂ = H, (<i>CH</i> ₃)
Acids		R ₁ = N/A R ₂ = (<i>H</i>), CH ₃	R ₁ = N/A R ₂ = (<i>H</i>), CH ₃
Amides		R ₁ = N/A R ₂ = H, CH ₃	R ₁ = N/A R ₂ = H, (<i>CH</i> ₃)
Total Number of Fragments		N = 13	N = 13

Table 7: Summary of the QSAR models shown in Figures 3 and 4. R^2_{pred} is leave-one-out cross-validation. All models utilised the full 29 chemicals in the dataset.

Model	Equation	R^2 (%)	R^2_{adj} (%)	R^2_{pred} (%)	S
1	$\text{pRC50} = 1.05 - 0.31 \Delta E_{\text{React-TS}}$	71.49	70.44	67.77	0.53
2	$\text{pRC50} = 1.13 - 0.34 \Delta E_{\text{React-TS-fragment}}$	78.13	77.32	75.37	0.46
3	$\text{pIGC50} = 2.22 - 0.31 \Delta E_{\text{React-TS}}$	69.74	68.62	66.25	0.56
4	$\text{pIGC50} = 2.29 - 0.34 \Delta E_{\text{React-TS-fragment}}$	74.28	73.32	71.16	0.52