LJMU Research Online

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

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

http://researchonline.ljmu.ac.uk/id/eprint/3606/

Article

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. Shultz [‡] 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
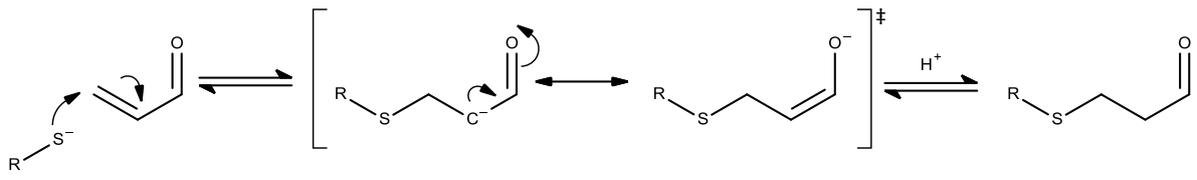
Abstract

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 electrophilic compound. This particular MIE has been associated with various toxicological endpoints such as acute aquatic toxicity, skin sensitisation and respiratory sensitisation. 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 towards glutathione was predicted using a combination of a calculated activation energy value (Eact, 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 alpha 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 polarising 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 Eact 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 molecular initiating event is important within the adverse outcome pathway paradigm.

Introduction

The Adverse Outcome Pathway (AOP) paradigm has been promoted as a key approach that may enable the demands of 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 endpoint to be replaced. For example, the recently defined AOP for skin sensitisation 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 sensitisation, respiratory sensitisation, 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 resulting in the formation of a resonance stabilised 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 stabilise/destabilise 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 (GSH). [18] In contrast *in silico* methods, such as the work 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 molecular initiating event.

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. 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 such calculations.

Methods

Data set

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%. 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 polarised aldehydes, ketones, esters, nitros, nitriles and cyclic ketones. [11] This resulted in a subsequent dataset of 72 chemicals covering 13 aldehydes, 17 ketones, 24 esters, nine nitro compounds, three nitrile containing compounds and six cyclic ketones (Table 1). Additionally individual standard

deviation values are stated, these values result in an average experimental error of 0.13 log units.

[Table 1 here]

<u>Computational methods</u>

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 (Eact) values for transition state structures were calculated use 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 minimisation 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 supplementary information. This includes the fragment which is used for each chemical in table 1, calculated Eact (Kcal/mol), SAS values and predicted –Log RC$_{50}$ values for each model.

<u>Statistical analysis</u>

Linear regression analysis was used to develop quantitative structure-activity relationship models to obtain correlations between -log RC$_{50}$ values and the calculated descriptors (Eact and SAS values) using the Minitab (version 17) statistical software.

<u>Results and Discussion</u>

The initial aim of this study was to develop a fragment-based *in silico* profiler capable of predicting chemical reactivity for polarised 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 cut-off 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 *i*-propyl and *t*-butyl groups due to their increased steric hindrance. An analogous analysis was carried out into the effect of alkyl chain length on the polarised aldehydes at position R_2 (Table 2).

The effect of a benzene ring on the calculated activation energy for the polarised 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) cut-off. 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, (sufficiently close to the1 kcal/mol s) cut-off. This means that two fragments are required to define the effect of a benzene ring at the β-position (R_1), with $R_1 = CH_2CH_3$ 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 polarised aldehydes at the α-position (R_2) (Table 2).

[Table 2 here]

[Table 3 here]

The structure-activity analysis into the effect of alkyl and aryl substituents on the calculated activation energy was repeated for the polarised ketones and esters in the dataset (varying groups at positions R1, R2 and R3, data shown in the supplementary information) resulting in the definition of 407 fragments, these are summarised 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 glutathione reactivity was investigated for a total of 54 chemicals (13 polarised aldehydes, 17 polarised ketones and 24 polarised esters). Initial modelling using only the calculated activation energy value (Eact) 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, Chemicals with an alpha-substituent shown as filled squares). Inclusion of a solvent accessible surface area (SAS) descriptor for the α-position resulted in the 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 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.

 [Table 4 here]

Model 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 to be significant outliers with errors of 1.07 and 0.99 log units respectively (Figure 4). 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 with the unsubstituted esters as they are reacting sufficiently fast enough 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.

[Table 5 here]

An additional set of chemicals were also poorly predicted by model 3 (Figure 4, chemicals highlighted as filled squares), these being chemicals in which a phenyl ring conjugated to the carbonyl or ester moiety acts as the polarising group (Table 6). 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-3-buten-2-one in which the polarising group is a simple alkyl ketone is well predicted by model 3 with an error of -0.04 log units. 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 is alkyl or hydrogen and the β-position is substituted with aromatic ring are well predicted by the model – see supplementary 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 units.

[Table 6 here]

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 Eact values for an additional five fragments to be calculated, along with three fragments previously defined (Table 7). These Eact values were used in conjunction with model 4 to predict $-logRC_{50}$ values for these 18 chemicals with an average error of 0.62 log units (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 polarised nitros that substituents at the α-position have significantly less effect on reactivity than for chemicals polarised by either 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 5). 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 $-logRC_{50}$ values for these chemicals (Figure 4, right hand lot), 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 supplementary 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.

[Table 7 here]

Conclusions

The aim of this work was to develop an *in silico* profiler capable of predicting reactivity for polarised aldehydes, ketones and esters acting via Michael addition. The results showed that a combination of pre-calculated Eact 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 Eact and SAS values. The approach, and 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.

Abbreviations

AOP – Adverse Outcome Pathway

DFT – Density Functional Theory

Eact – Energies of activation

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

Supporting information description

SAS – Solvent Accessible Surface area

VP – Vapour Pressure

The supplementary information contains all data for 72 chemicals (13 aldehydes, 17 ketones, 24 esters, nine nitro compounds, three nitrile containing compounds and six cyclic ketones). This includes chemical names, SMILES, experimental values (RC_{50} average, RC_{50} –$\log RC_{50}$), calculated values (fragment Eact and fragment SAS A), the fragment used and predicted –$\log$ RC_{50} values for models 1-4. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1)   EC. (2006) 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.

(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 andrisk 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 .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. Biodivers. 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. Re.s 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-[small alpha]-acetyl-l-lysine as an amine nucleophile. *Medchemcomm*.

(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., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kabayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Toamsi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Faraks, O., Foresmann, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. (2009) Gaussian 09, revision A.1, Wallingford, CT.

(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.

Table 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.

ID	Chemicals	SMILES	RC50 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\CCCCCCC	0.21 ± 0.05	0.67
11	*Trans*-2-Decenal	O=C/C=C/CCCCCCC	0.17 ± 0.02	0.77
12	*Trans*-cinalamdehyde	O=C\C=C\C1=CC=CC=C1	0.96 ± 0.24	0.02
13	α-Methyl-*trans*-cinnamaldehyde *	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=CCC)C	0.67 ± 0.11	0.17
19	3-Octen-2-one	O=C(C=CCCC)C	0.57 ± 0.16	0.24
20	3-Nonen-2-one	O=C(C=CCCCC)C	0.54 ± 0.11	0.27
21	3-Decen-2-one	O=C(C=CCCCCC)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 *	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-hydroxy chalcone	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-Hexylacrylate	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

14

40	2-Hydroxy ethyl 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-pentylacrylate	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-pentnoate	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-pentnoate	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 *	CCOC(=O)C(C)=C	33.75	-1.53
53	Butyl methacrylate *	CCCCOC(=O)C(C)=C	43.27	-1.63
54	2-Ethylhexyl acrylate *	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
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

Average RC_{50} values are given for chemicals with multiple measurements, RC_{50} values 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.

Table 2: Calculated activation energy values for polarised aldehydes (R groups as defined in Figure 2). Analogous data for polarised ketones and esters is available in the supplementary information.

R_1	R_2	R_3	Eact (kcal/mol)
CH_3	H	H	-1.5
CH_2CH_3	H	H	2.7
$CH_2CH_2CH_3$	H	H	2.5
i-propyl	H	H	3.7
t-butyl	H	H	5.6

C_6H_5	H	H	3.4
$CH_2C_6H_5$	H	H	1.2
$CH_2CH_2C_6H_5$	H	H	1.8
$CH_2CH_2CH_2C_6H_5$	H	H	2.2
H	H	H	-5.4
H	CH_3	H	-1.7
H	CH_2CH_3	H	-1.9
H	$CH_2CH_2CH_3$	H	-2.0
H	i-propyl	H	-0.3
H	t-butyl	H	2.4
H	C_6H_5	H	-8.7
H	$CH_2C_6H_5$	H	-2.2
H	$CH_2CH_2C_6H_5$	H	-2.5
H	$CH_2CH_2CH_2C_6H_5$	H	-2.5

Table 3: Summary of the fragments defined for polarised aldehydes, ketones and esters (R groups as defined in Figure 2)

Chemical class	R_1	R_2	R_3
Polarised aldehydes	Alkyl: H, CH_3, CH_2CH_3, i-propyl, t-butyl Aryl: C_6H_5, $CH_2C_6H_5$, CH_2CH_3	Alkyl: H, CH_3, i-propyl, t-butyl Aryl: C_6H_5, CH_3 [for $(CH_2)_nC_6H_5$, n>=1]	H

	[for $(CH_2)_nC_6H_5$, n>=2]		
Polarised 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, *i*-propyl, *t*-butyl Aryl: C_6H_5, $CH_2C_6H_5$, CH_2CH_3 [for $(CH_2)_nC_6H_5$, n>=2]	Alkyl: CH_3, *i*-propyl, *t*-butyl Aryl: C_6H_5, $CH_2C_6H_5$, CH_2CH_3 [for $(CH_2)_nC_6H_5$, n>=2]
Polarised esters	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, *i*-propyl, *t*-butyl Aryl: C_6H_5, $CH_2C_6H_5$, $CH_2CH_2C_6H_5$, $CH_2CH_2CH_3$ [for $(CH_2)_nC_6H_5$, n>=3]	Alkyl: OCH_3, O-*i*-propyl, O-*t*-butyl Aryl: OC_6H_5, OCH_3 [for $(CH_2)_nC_6H_5$, n>=1]

Table 4: Summary statistics for models 1-4 as shown in Figure 3. Model 1 has no SAS value as it uses Eact as a single descriptor.

Model	N	a	b	c	R^2	R^2-adj	R^2-Pred	Average Error
			-Log RC_{50} = a + b.Eact + c.SAS Alpha					
1	54	0.80	-0.15	X	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 | 52 | -1.30 | -0.07 | 0.12 | 0.81 | 0.80 | 0.78 | 0.37 |
| 4 | 48 | -1.48 | -0.06 | 0.13 | 0.87 | 0.86 | 0.85 | 0.29 |

Table 5. The predicted error values for predicted –Log RC_{50} of β-substituted esters with corresponding Log Vapour Pressure (VP) values.

Compound	-Log RC50 (mM)	Predicted -Log RC50 (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

Chemical				
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

Table 6: Predicted –log RC_{50} values chemicals with a conjugated phenyl polarising group.

Chemical	Structure	-Log RC_{50}	Predicted -Log RC_{50}	Error
Chalcone		0.40	-0.37	0.77
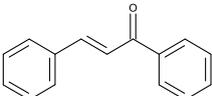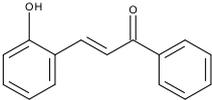
Chemical	Structure			
2-Hydroxy-chalcone		0.56	-0.37	0.92
4-Hydroxy-chalcone		0.39	-0.37	0.76
4-Phenyl-3-buten-2-one		-0.55	-0.51	-0.04
Phenyl-acrylate		1.64	0.94	0.76
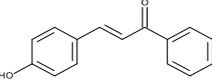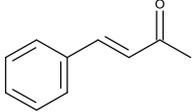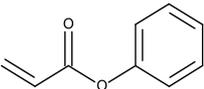
Table 7: Fragments required to predict reactivity of polarised nitros, polarised nitriles and cyclic ketones.

Chemical	Fragment used	New or existing fragment
	Polarised Nitros	
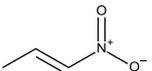
1-Nitro-1-cyclohexene		New
4-Methyl-b-nitrostyrene *Trans*-b-nitrostyrene *Trans*-4-methyl-b-nitrostyrene *Trans*-4-chloro-b-nitrostyrene *Trans*-4-bromo-β-nitrostyrene 4-Fluoro-b-nitrostyrene *Trans*-4-methoxy-b-nitrostyrene		New
Trans-b-methyl-b-nitrostyrene		New
Polarised nitriles		
2-Methyleneglutaronitrile		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
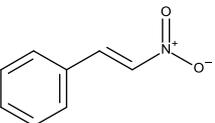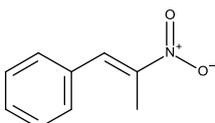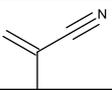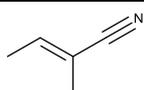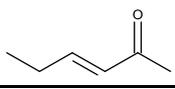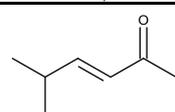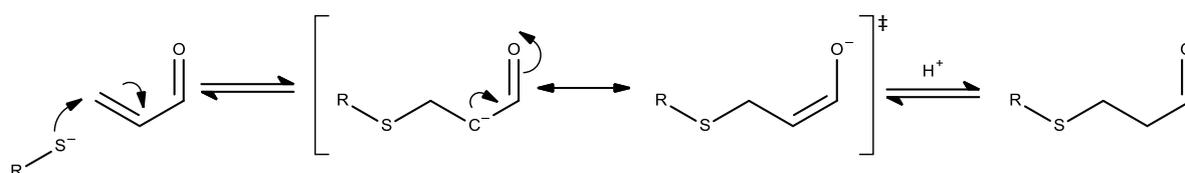
Figure 1. The proposed mechanism and transition state of acrolein (an electrophile) and a thiol
nucleophile (R = glutathione, alkyl).
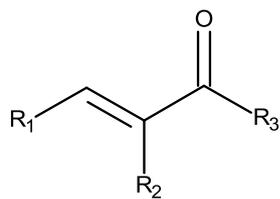
Figure 2: General structure for polarised aldehydes (R_3 = H), polarised ketones (R_3 = C) and polarised esters (R_3 = OC)
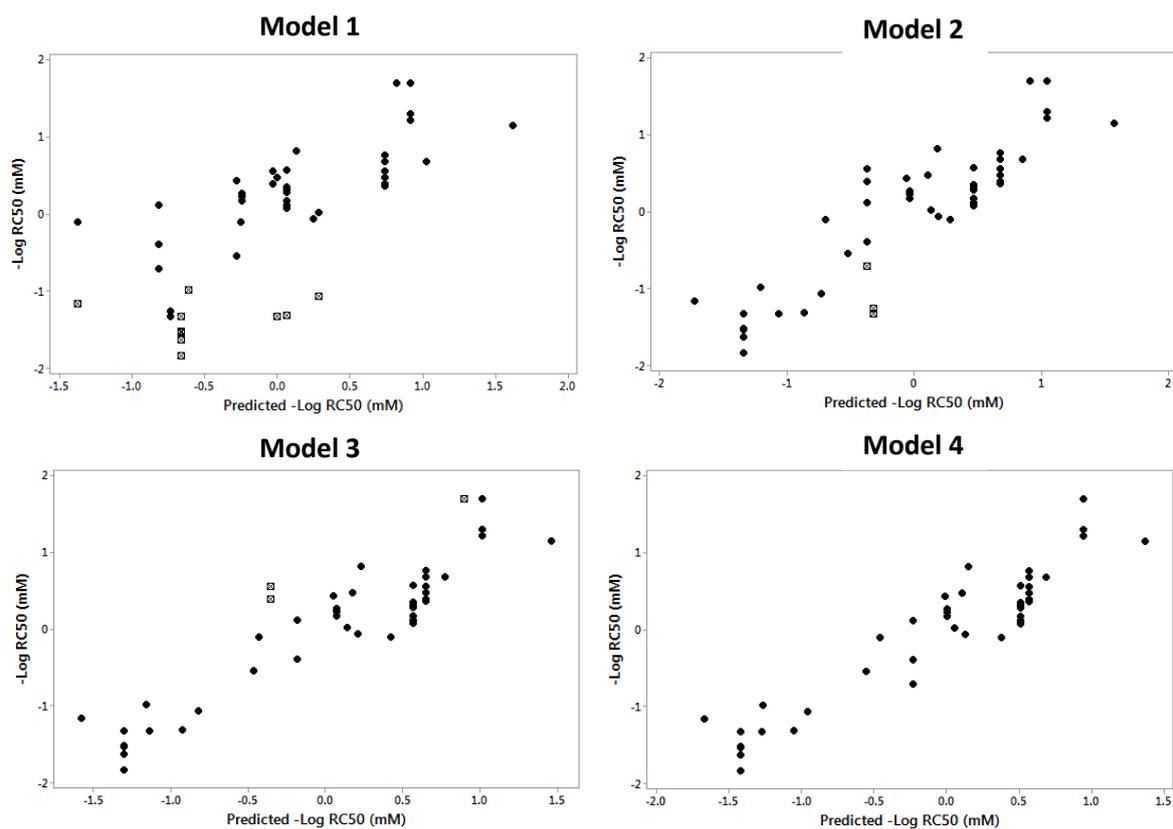
Figure 3: Predicted versus experimental values for -LogRC$_{50}$ for all models in the current study. Model 1: Eact only; model 2: Eact with SAS at the α-position included model 3: Eact with SAS at the α-position excluding three volatile β-esters 4: Eact with SAS at the α-position excluding three volatile β-esters and four compound with phenyl electron withdrawing group
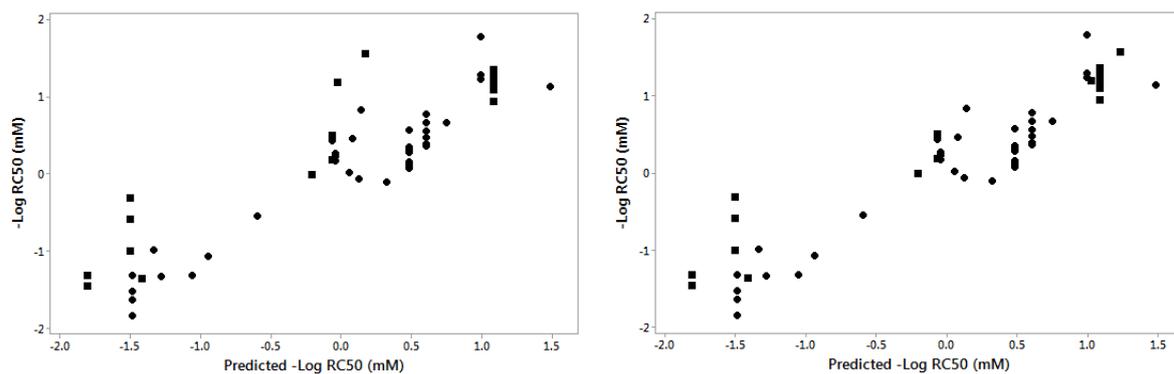
Figure 4: Predicted versus experimental –Log RC$_{50}$ values for polarised nitros, polarised nitriles and polarised cyclic ketones (shown as filled in squares) using model 4 in comparison to the polarised aldehydes, ketones and esters in the initial dataset (shown as filled circles). Left hand plot shows polarised nitros with the inclusion of the SAS descriptor. Right hand plot shows polarised nitros with the SAS descriptor value set to hydrogen for all chemicals.
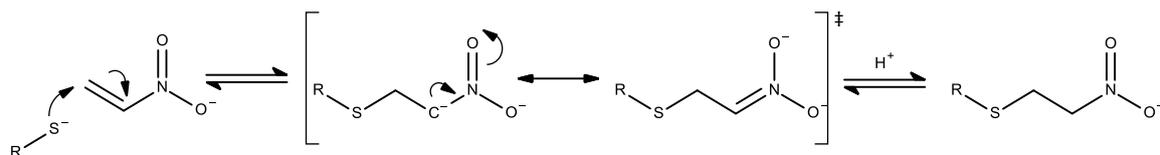
Figure 5: Michael addition mechanism for the reaction between thiol nucleophile and nitroethene (R = alkyl, GSH).