Automated Computational Screening of the Thiol Reactivity of Substituted Alkenes

Jennifer M. Smith and Christopher N. Rowley

Department of Chemistry, Memorial University of Newfoundland, St. John's, NL, A1B 3X7 Canada

Corresponding Author:

Christopher N. Rowley,

e-mail: crowley@mun.ca

tel: +7098642229

The final publication is available at Springer via http://dx.doi.org/10.1007/s10822-015-9857-0

The total number of words of the manuscript, including entire text from title page to figure legends: **6120**

The number of words of the abstract: **215**

The number of figures: **3**

The number of tables: 6

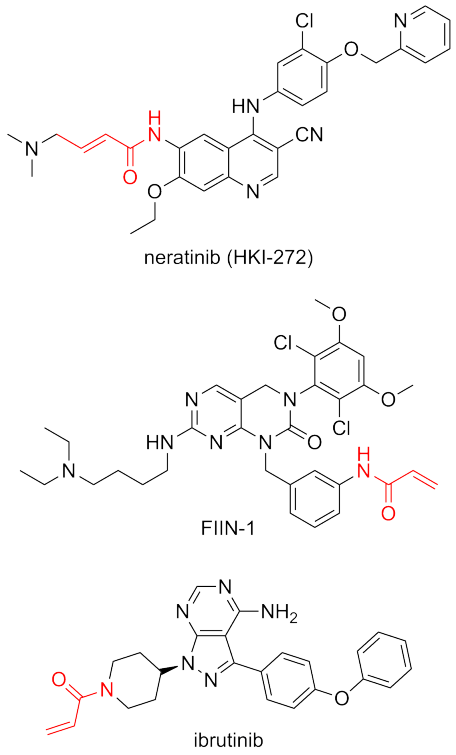
Electrophilic olefins can react with the S–H moiety of cysteine side chains. The formation of a covalent adduct through this mechanism can result in the inhibition of an enzyme. The reactivity of an olefin towards cysteine depends on its functional groups. In this study, 325 reactions of thiol-Michael-type additions to olefins were modeled using density functional theory (DFT). All combinations of ethenes with hydrogen, methyl ester, amide, and cyano substituents were included. An automated workflow was developed to perform the construction, conformation search, minimization, and calculation of molecular properties for the reactant, carbanion intermediate, and thioether products for a model reaction of the addition of methanethiol to the electrophile. Known cysteine-reactive electrophiles present in the database were predicted to react exergonically with methanethiol through a carbanion with a stability in the 30–40 kcal mol^{-1} range. 13 other compounds in our database that are also present in the PubChem database have similar properties. Natural Bond Orbital (NBO) parameters were computed and regression analysis was used to determine the relationship between properties of the olefin electronic structure and the product and intermediate stability. The stability of the intermediates is very sensitive to electronic effects on the carbon where the anionic charge is centered. The stability of the products is more sensitive to steric factors.

Introduction

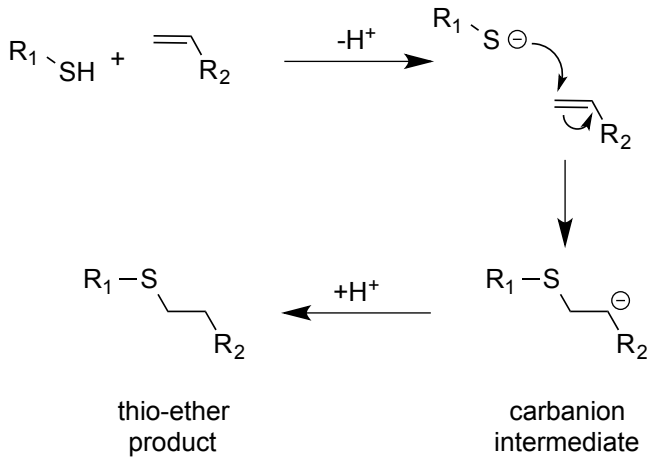
Enzyme inhibition through the reaction of a cysteine residue with an electrophile is a significant mechanism in toxicology and pharmacology [1-3]. A successful line of anti-cancer drugs has targeted active site cysteine residues of kinases [4]. These drugs can also modify cysteines in unrelated proteins, so off-target modification is a significant concern. A better understanding of properties needed for a compound to covalently modify a protein will facilitate the design of potent but selective irreversible inhibitors.

A major class of cysteine-targeting covalent modifiers are electrophilic olefins. Established irreversible inhibitors like neratinib [5], FIIN-1 [6], and ibrutinib [7] feature acrylamide warheads, although acrylate [8] and acrylonitrile [9] moieties have also shown cysteine-modifying activity (Scheme 1). In the development of selective covalent modifier drugs, medicinal chemists have frequently started from the scaffold of a drug that binds non-covalently and then added an electrophilic olefin substituent, referred to as the "warhead."[2]

The canonical mechanism for covalent modification of an electrophilic olefin begins with non-covalent binding to the target, such that the electrophile is positioned in the region of the deprotonated thiol moiety of the target cysteine. The second step is a Michael-type addition reaction, where the nucleophilic thiolate attacks the more electrophilic carbon of the olefin (Scheme 2) [10]. Conjugated addition then proceeds through formation of a carbanion at the intermediate α-carbon. This type of thiol-Michael addition to an electrophilic olefin is also of significance in toxicology [1], organic synthesis [11], and materials chemistry [12,13].



Scheme 1. Cysteine targeting covalent-modifier drug molecules with an electrophilic olefin warhead (red).



Scheme 2. General reaction mechanism for thiol Michael-type addition to an alkene.

The reactivity of the warhead can be understood from the stabilities of two key species: the carbanion intermediate and the thio-ether product [14]. The stability of the carbanion intermediate is proportional to the rate at which the addition occurs. The stability of the thio-ether product determines if the reaction will occur spontaneously and the degree to which the addition is reversible. The stabilities of

the carbanion and products are quantified by ΔG_{int} and ΔG_{prod}, respectively (Scheme 3).

The reactivity of an olefin through this mechanism is very sensitive to the substituents of the alkene moiety, so these substituents must be chosen to confer the necessary reactivity towards the target without creating a promiscuous inhibitor. The degree of electrophilicity must be carefully adjusted; warheads should be reactive enough to achieve reaction with the target cysteine, but not so reactive that the inhibitor will make off-target modifications. A recent line of development by Taunton and coworkers optimized an electrophile reacting with its target reversibly in order to limit off-target modification [14,15].

Several studies have correlated calculated molecular descriptors of electrophiles to toxicological or pharmacological properties [1,10,16]. Descriptors calculated from Natural Bond Orbital (NBO) theory, such as electrophilicity, steric accessibility, and energy levels of olefin have been successfully correlated to the rates of thiol additions of electrophiles [16]. Although these studies are informative, their scope is hindered by the relatively small number of compounds for which there is experimental data. By using quantum chemistry to study the reactants, intermediates, and products involved in the thiol-Michael addition reaction profiles, a more comprehensive survey of substituent effects can be performed.

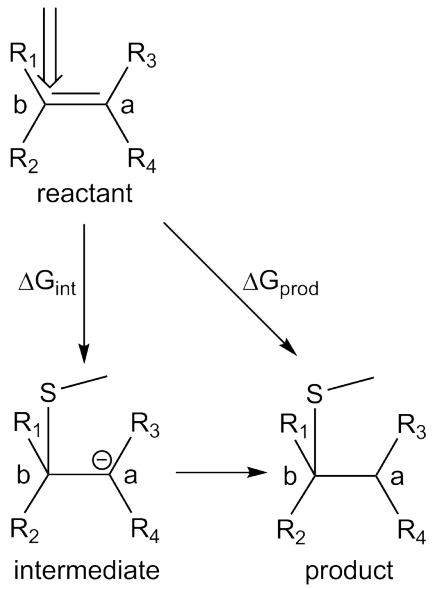
Until recently, computational research into thiol-Michael additions has been limited. Recent studies have found valuable insights by modeling small numbers of addition reactions, but there are many thousands of other possible electrophiles [17,18]. Computational workflows have been developed to calculate if a molecule containing an electrophilic warhead has a geometry that can fit in the active site of a protein target when the cysteine adduct has formed, but these methods are based on mechanical models and cannot describe the relative reactivity of the warhead [15,19-21]. A recent study by our group determined that many common DFT methods fail to correctly describe the carbanion intermediate and underestimate the stability of the products [22].This failure has limited the utility of standard DFT functionals. A newly developed class of range-separated DFT functionals correct this failure and creates the potential to use computational tools to predict the properties of large numbers of electrophilic warheads.

In this paper, we develop a computational workflow to compute the full set of olefins with substituents commonly used in pharmacology. We estimate the stabilities of the intermediate and products for the reaction of the electrophiles with a model thiol. Finally, electronic structure descriptors of the olefins are correlated to the stabilities of the intermediates and products to provide a predictive model for these energies and quantitatively determine how the properties of the reactant olefins affect them.

A large dataset of electrophilic olefins was generated to provide estimates on their ability to function as irreversible inhibitors. The compounds examined vary in substituent groups at the ‘a’ and ‘b’ carbons (Scheme 3), which provides insight into the effects of these substituents on the thermodynamics of thiol addition reactions.

We examined the effect of substituents on the four positions around a C=C double bond. The substituents include representatives of the major functional groups present in pharmaceutical electrophiles: hydrogen, methyl, methyl ester, N-methyl amide, and cyano. All combinations of these substituents were generated and analyzed by an automated procedure (see Computational Details). Except for symmetric olefins, the addition of methyl thiolate can occur at two distinct carbons, yielding two possible carbanion intermediates. The two possible thioether products of these intermediates were also computed. There were a total of 325 unique reactions after excluding reactions within this that are equivalent by symmetry.

For each reaction, the reactant olefin was constructed and its properties were calculated using DFT. The structures, properties, and energies of the complexes computed in this study are tabulated in the supplementary information, referenced according to their PubChem ID where available. In the cases where there are two possible stereoisomers, the entries are cross-referenced and the isomer with the more stable intermediate and products are indicated. The trends in these computed properties are discussed in the following sections.



Scheme 3. The reactant, intermediate, and product species of the model thiol-addition reactions. Groups $R_{1\text{-}4}$ = H, CN, CH_3, $C(=O)OCH_3$, $C(=O)NHCH_3$

General Properties of the Reaction Set

Within the dataset, 194 of the 235 reactions were predicted to be exergonic, with net Gibbs energies of reaction ranging between –11.6 kcal mol^{-1} and 12.0 kcal mol^{-1}. These moderate reaction Gibbs energies generally reflect the loss of entropy that occurs as the result of the addition and the modest strength of the C–S bond formed in the reaction. The carbanion intermediates had a broader range of stabilities. Minimum energy carbanion intermediates were identified for 254 of the 325 reactions. The stabilities of these intermediates had a broad range, from –6.1 to 74.9 kcal mol^{-1}.

Structure and Properties of Carbanions

The carbanions generally had consistent structures. The length of the C_b–S bond varied in the 1.85–2.00 Å range, with the bond length tending to be longer in less stable intermediates. The C_a center, where the carbanion is localized, is planar. Bonds between this center and carboxylate, amide, or cyano substituents were typically in the 1.38–1.42 Å range, indicating delocalization of the anionic charge to these functional groups. This is apparent in a π Natural Bond Orbital (NBO) between C_a and the C of the carboxylate group of the acrylate carbanion (Figure 1).

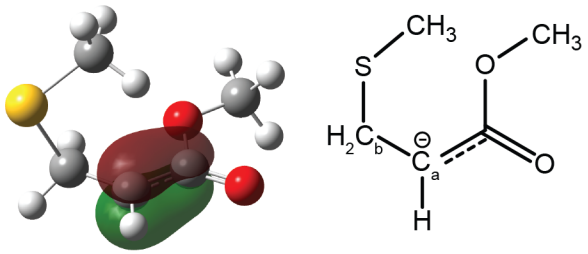
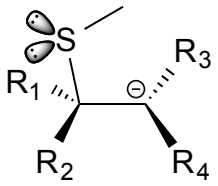


Figure 1. The NBO corresponding to π delocalization of the carbanion to the electron withdrawing methyl ester group of the carbanion intermediate of the methylacrylate addition.

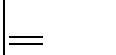
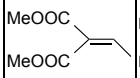
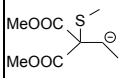
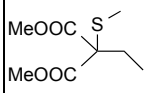
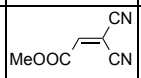
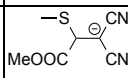
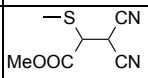
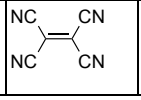
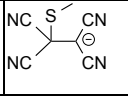
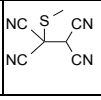
The intermediate favors conformations where the S–CH_3 bond eclipses the C_a–C_b bond because this orientation minimizes the electrostatic repulsion between the lone pair orbitals of the sulfur atom and the p-orbital on the 'a'-carbon where the anionic charge is localized [23],



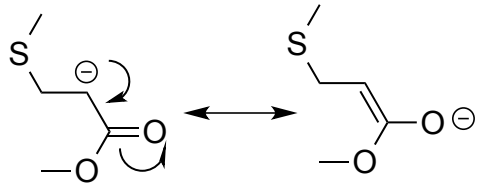
This orientation effectively blocks one face of the intermediate from protonation, causing a kinetic preference for the anti stereoisomer of the product. As a result, we only considered the anti product in our calculations.

The stability of the intermediate (ΔG_{int}) ranged from –6.1 to 74.9 kcal mol^{-1}, indicating that it is highly sensitive to the substituents. No potential energy minimum corresponding to the carbanion could be identified in 71 of the 325 reactions. These species are presumed to be incapable of undergoing a Michael-type addition with a thiol and would need to react through another mechanism, like a 1,2 σ-addition. Most of these reactions were additions to olefins that lacked an electron-withdrawing group at the 'a'-position, such as ethene or propene (e.g. Table 1, entry 1).

Table 1. The stabilities of the methanethiol addition intermediate and products for selected olefins.

Entry	Olefin	Intermediate	Product	ΔG_{int} (kcal mol^{-1})	ΔG_{prod} (kcal mol^{-1})
1		N/A		—	–11.6
102				74.9	2.5
159				0.9	–3.1
325				–6.1	0.6

The reactions with the most stable carbanion intermediates were classical Michael-acceptors, where the carbanion can be delocalized through resonance with a conjugated functional group bound to the ‘a’-carbon. An example of this is the thio-methylacrylate carbanion, which is resonance stabilized by its enolate form,



The methyl ester, amide, and cyano groups are all capable of this type of resonance, so carbanions with these substituents at the ‘a’-position tend to be stable intermediates.

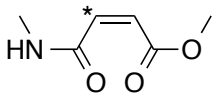
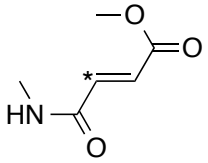
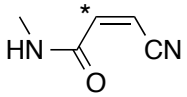
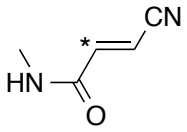
In general, substitution at the ‘b’ position gave inferior intermediate stabilities when compared with simple hydrogen substitution. However, in most cases where substitution at the ‘b’-position was combined with methyl or hydrogen groups at the ‘a’-position, intermediates reverted to a non-covalent reactant complex or were highly unstable. Olefins with electron withdrawing groups at the ‘b’ position tended to have very unstable carbanions (e.g. Table 1, entry 102), unless there were additional electron withdrawing groups at the ‘a’-position (e.g. Table 1, entry 159). Substitution with 3 or 4 electron-withdrawing groups yield very stable carbanion intermediates (e.g. Table 1, entry 325).

Trends in the Thio-Ether Products

The net Gibbs reaction energies, ΔG_{prod}, determined from the reaction of the electrophilic olefin with methanethiol to form the neutral thioether product, ranged from –11.6 to 12.0 kcal mol^{-1}. In comparison to the wide variation found for the intermediates, the stability of the thioether product was less sensitive to the substituents. The most stable product was that of entry **1**, where the reactant olefin was ethene. Broadly, the presence of electron withdrawing groups resulted in a decrease in stability of the thio-ether product. Reactions with compounds with multiple electron withdrawing substituents were generally endergonic or weakly exergonic.

The cis and trans isomers of 1,2-disubstituted olefins show an interesting trend in the stability of the thio-ether products. When bulky substituents, such as amides or esters, are cis to another group, the reaction to form the thio-ether product was significantly more exergonic than for the trans isomer (Table 2). This can be attributed to steric repulsion between bulky substituents that are cis to each other. The addition of the thiol to the C=C bond allows these groups to rotate out of plane, reducing the repulsive interaction between them. This effect is greatest when both substituents are bulky, such as amides or esters. A smaller effect is apparent when one group is bulky and the second is a smaller group, like a cyano.

Table 2. Comparison of cis vs trans isomers on intermediate and product stabilities. The site of addition is indicated by an asterisk.

Entry	Olefin	ΔG_{int} (kcal mol^{-1})	ΔG_{prod} (kcal mol^{-1})
51 cis		26.2	–11.1
45 trans		32.1	–3.3
172 cis		30.8	–6.7
184 trans		34.7	–2.9

No Correlation Between Carbanion and Product Stability

The Gibbs energies of the intermediates and products (ΔG_{int} vs. ΔG_{prod}) for each reaction where an intermediate could be identified are plotted in Figure 2. The coefficient of determination (R^2) is 0.03, indicating that there is essentially no correlation between these properties. These diverse pairings of kinetic reactivity and thermodynamic stabilities provide considerable opportunity for medicinal chemists to design an electrophilic warhead with an optimal kinetic and thermodynamic reactivity via its substituents. It also indicates that the substituent effects that yield stable intermediates are different than those that yield stable products.

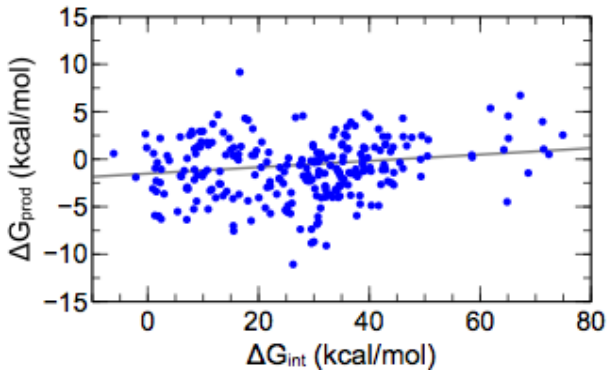


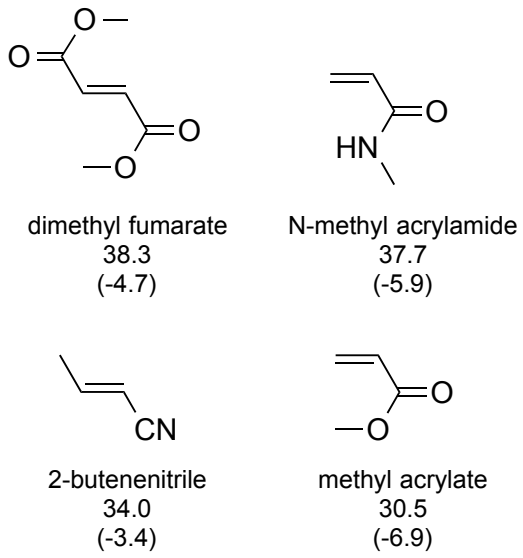
Figure 2. Correlation between intermediate and product stabilities (R^2=0.03).

Connection to Chemical Stability

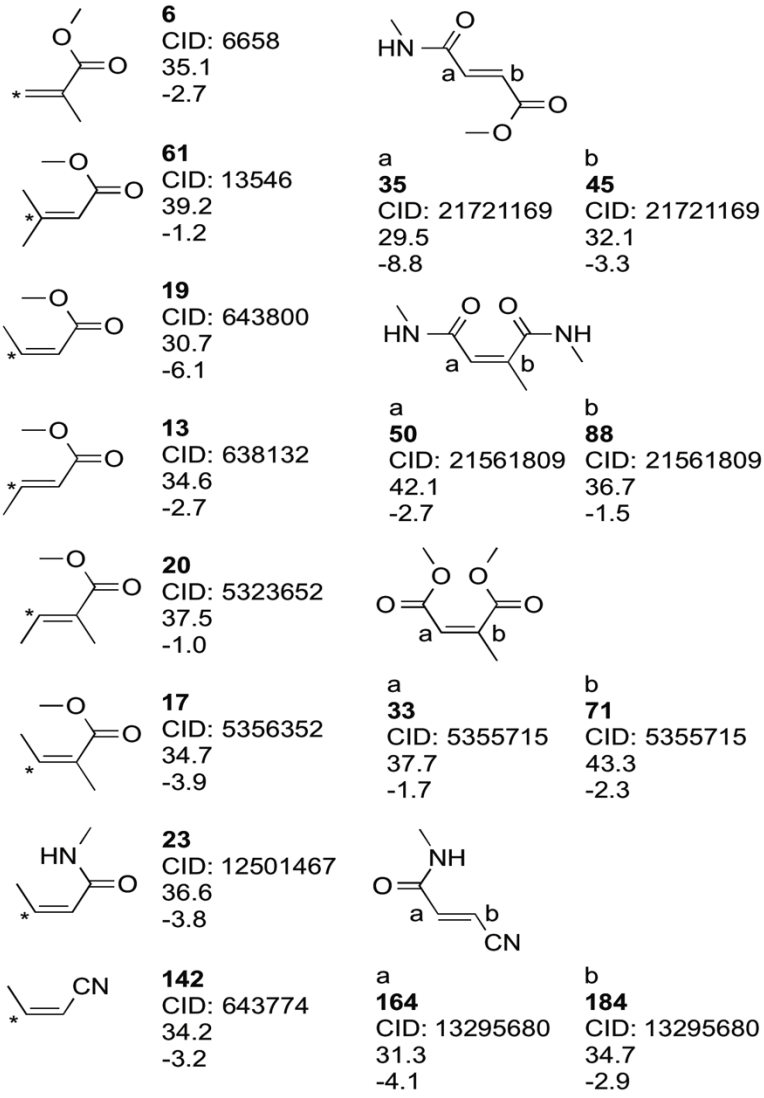
Many of the olefins in our database contain multiple electron withdrawing groups and are therefore highly electrophilic. This high electrophilicity means that it may not be possible to synthesize some of these compounds and those that could be synthesized are likely to be insufficiently stable for use in a biological environment. We identified those that have been synthesized and characterized by checking if they have been entered in the PubChem database [24]. This analysis showed that only 81 of the olefins in this test set are currently in the PubChem database.

Olefins that lacked a PubChem ID tended to have very stable carbanions; of the 92 reactions that are predicted to undergo a spontaneous addition with methanethiol and have intermediate stabilities less than 20 kcal mol^{-1}, only 12 correspond to olefins that have PubChem IDs. Although these 12 compounds have been characterized, it is apparent that they are generally too electrophilic to be relevant in a biological environment; this list of compounds includes the powerful redox reagent tetracyanoethylene and methyl cyanoacrylate, which polymerizes on exposure to moisture. Based on these trends, we can define a rough rule that an electrophile with a carbanion intermediate stability of 20 kcal mol^{-1} or lower is probably too reactive to be used as the warhead of a selective therapeutic inhibitor. Our database also contains some known cysteine-reactive electrophiles. Dimethyl fumarate, N-methyl acrylamide, methyl acrylate, 2-butenenitrile, and methyl acrylonitrile have been shown experimentally to react with glutathione, a prototypical biological thiol [25-27]. Acrylamide warheads have also been used in covalent modifiers like FIIN-1 and ibrutinib. These compounds have intermediate stabilities

that range from 29.5 to 38.3 kcal mol^{-1} (Scheme 4). These examples suggest that electrophiles with intermediate stabilities in the 30–40 kcal mol^{-1} range are still reactive towards biological thiols but are not so reactive as to be immediately deactivated by other mechanisms. The stability of these thioether products is in the moderately-exergonic range ($-3.4 \leq \Delta G_{prod} \leq -6.9$ kcal mol^{-1}).



Scheme 4. Michael acceptors that have been shown experimentally to react with cysteine thiols. The calculated carbanion intermediate stability (ΔG_{int}) and Gibbs energy of reaction (ΔG_{prod}, in parenthesis) are given below the structures. All energies are reported in kcal mol^{-1}.



Scheme 5. Database entries with PubChem IDs that exergonically react with methanethiol ($\Delta G_{prod} < -1$ kcal mol^{-1}) through a moderately stable carbanion intermediate (30 kcal mol^{-1} < ΔG_{int} < 40 kcal mol^{-1}). The site of thiol addition is indicated by an asterisk when there is one stable addition route. The site of addition is indicated by an 'a' or 'b' when there are two possible sites for addition.

Based on these criteria, we identified 13 compounds in our database that are predicted to react exergonically with methanethiol ($\Delta G_{prod} < -1$ kcal mol^{-1}) and form intermediates with moderate stability (30 kcal mol^{-1} ≤ ΔG_{int} ≤ 40 kcal mol^{-1}). The structures and predicted energies of these compounds are presented in Scheme 5. Most of these compounds contain one electron-withdrawing group and one or more methyl groups, indicating that methyl-substitution of the olefin does not necessarily destabilize the intermediate or products to an excessive degree. Four others have

electron withdrawing groups on both the 'a' and 'b' carbons. These will typically form stable carbanions and products through addition at both sites.

Selecting a warhead with a carbanion stability >40 kcal mol^{-1} would have the advantage of reducing the risk of promiscuous modification, but potentially at the cost of a slow reaction rate between the warhead and the target protein. One strategy to address this would be to design a drug with a less reactive warhead that binds to the target through strong non-covalent interactions. The non-covalent complex would eventually be converted to a covalent adduct, reducing the off-rate of the drug. Strong non-covalent interactions combined with a moderate rate of covalent modification were noted as key features of existing irreversible inhibitors of EGFR [28]. Alternatively, if the drug were designed to target active-site cysteine residues with low pKa's, less acidic cysteines elsewhere in the organism would be less susceptible to modification.

Linear Free Energy Relationships Between Reactant Descriptors and Intermediate/Product Stability

To make more quantitative predictions about thiol-reactivity based on the electronic structure of the reactant olefin, five NBO properties were calculated for each olefin: the Natural Population Analysis (NPA) atomic charges of the two olefinic carbons ($q(C_a)$ and $q(C_b)$), and three orbital energies of the olefin double bonds (E_σ, E_π, and E_{π^*}). We also considered the HOMO-LUMO gap, $\Delta E_{\pi \rightarrow \pi^*}$, and electrophilicity index, ω, as reactivity descriptors. The electrophilicity index is calculated from the NBO π and π^* orbital energies using Eqn. 1 [29].

$$\omega = \frac{E_\pi^2 + 2E_\pi E_{\pi^*} + E_{\pi^*}^2}{4\left(E_{\pi^*} - E_\pi\right)} \tag{1}$$

Steric repulsion between the thiol and the substituents and the substituents with each other is another important factor in thiol-Michael additions. Molar refractivities are commonly used steric descriptors in QSAR studies that provide a measure of the volume and polarizability of a particular atom or group of atoms [30]. The steric interaction energies at the 'a' and 'b' carbon sites were quantified by using a sum of the defined molar refractivities for each individual substituent at the specific

site (MR_a and MR_b, respectively). We also calculated the total steric repulsion between substituents cis to each other in the olefin reactant (E_{steric}). The E_{steric} term was calculated from the relative energy of the cis and trans isomers of the 1,2-substituted olefins for each pair of substituents (Table 3). The total steric repulsion is calculated as the sum of these energies for pairs of the cis-oriented substituents in the olefin reactant.

Table 3. Calculated steric repulsive energy in olefins with cis-substituents by substituent pairs. All values are in kcal mol^{-1}.

E_{steric}	$-CH_3$	$-C(=O)OCH_3$	$-C(=O)NHCH_3$	-CN
$-CH_3$	0.92	2.14	2.25	−0.32
$-C(=O)OCH_3$	2.14	4.88	6.90	2.87
$-C(=O)NHCH_3$	2.25	6.90	3.60	1.48
-CN	−0.32	2.87	1.48	0.88

Simple Linear Regression

In our preliminary analysis, we performed simple linear regression to identify relationships between the product and intermediate stabilities and descriptors of the reactants. Regression analysis revealed limited correlation of both intermediate and product stabilities with any single parameter, with the best correlation yielding R^2 values of 0.52 and 0.21 for the intermediates and products, respectively (Table 4).

Prediction of the carbanion stability using this type of simple linear correlation was moderately successful; E_σ, E_π, $\Delta E_{\pi\rightarrow\pi^*}$, and ω all have coefficients of determination (R^2) values in the 0.48–0.52 range. The E_σ descriptor showed the strongest correlation, with an R^2 of 0.52. The common feature of these parameters is that they are sensitive to the stability of the C_a–C_b bonding orbitals; olefins where these orbitals are destabilized by electron withdrawing groups tended to yield more stable carbanion intermediates.

All parameters showed fairly weak correlation with product stabilities. The highest correlation was seen for $q(C_b)$, but strength of the correlation was modest (R^2=0.21). This correlation generally reflects that the C_b–S bond will be stronger if C_b is electron deficient. The orbital energies of the olefinic double bond, $q(C_a)$, and the electrophilicity index (ω) are not strongly correlated to the stability of the products.

This is consistent with the lower sensitivity of the products to electronic effects. The stability of the products had a greater correlation to the steric effects of the substituents on the 'b'-carbon (MR_b) because there is a steric interaction between these groups and the thiol group in the products.

Table 4. Coefficients of determination (R^2) for linear regression analysis between the reactant descriptors and intermediate and product stabilities.

Descriptors	Coefficients of Determination (R^2) for Predicted Property	
	ΔG_{int}	ΔG_{prod}
$q(C_a)$	0.05	0.06
$q(C_b)$	0.11	0.21
E_σ	0.52	0.00
E_π	0.50	0.01
$\Delta E_{\pi\rightarrow\pi^*}$	0.48	0.01
E_{π^*}	0.12	0.03
ω	0.48	0.01
E_{steric}	0.01	0.00
MR_b	0.03	0.13
MR_a	0.17	0.00

Multiple Linear Regression

In an attempt to find a model with greater correlation, multiple linear regression was performed on combinations of 2–10 of the descriptors. The intermediate and product Gibbs energies were expressed as a linear combination of the descriptors with coefficients as per Eqn. 2. The regression coefficients and coefficients of determination with the highest correlation are collected in Tables 5 (intermediates) and 6 (products).

$$\Delta G_{\text{int/prod}} = aC_a + bC_b + cE_\sigma + dE_\pi + eE_{\pi^*} + fE_{\pi\rightarrow\pi^*} + g\omega + hE_{\text{steric}} + iMR_{a/b} + j \quad (2)$$

The stability of the intermediate was predicted with reasonable accuracy using the C_a charge, ω, and the sum of substituent molar refractivities at the 'a'-carbon (MR_a), yielding an R^2=0.83. This represents a significant improvement over the correlation with the C_a charge alone. Although the coefficient of determination is 0.83, the stability of each intermediate was predicted with an error of less than 3 kcal mol^{-1} using this model. This suggests that the stability of the intermediate is most sensitive to the electrophilicity of the C_a=C_b bond and the steric and electrostatic properties of C_a. In comparison, the steric terms and the charge on the 'b'-carbon were not as significant.

The prediction of the stability of the thioether products using multiple linear regression proved to be more challenging. The best fit with 4 parameters (R^2=0.63) included the atomic charges at both the 'a' and 'b'-carbon as well as steric influence from both the cis isomer effect, E_{steric}, and 'b'-carbon substituents, MR_b (Table 4). Including additional terms made only small improvement on the quality of the fit (R^2=0.66 when ω and E_{π^*} were also included). Generally, sterics are a more significant factor in the stability of the products than of the intermediates.

The highest R^2 value for the stability of products is still relatively poor compared to the performance of the model for the intermediate stability with fewer parameters. There were a greater number of products than intermediates in our data set because some intermediates could not be optimized, so we performed the correlation analysis again with only the products where an intermediate could be identified to provide a more direct comparison. The quality of this correlation remained relatively poor, indicating that the accurate prediction of the stability of thioether addition products would require the selection of different descriptors or the development of a more complex model.

Table 5. Multiple linear regression coefficients and R^2 values for calculated intermediate stability based on computed reactant properties.

ΔG_{int}	a	b	c	d	e	f	g	h	i	j
(R^2)	($q(C_a)$)	($q(C_b)$)	(E_σ)	(E_π)	(E_{π^*})	($\Delta E_{\pi\rightarrow\pi^*}$)	(ω)	(E_{steric})	(MR_a)	
0.84	60.51	12.65	−328.58	331.31	468.00	0	49.75	−0.63	−1.67	−144.46
0.83	60.42	11.23					−644.11	−0.29	−1.66	93.77
0.83	51.19						−643.79		−1.68	90.14

0.70							−659.55		−1.54	81.65
0.48							−626.18			58.99

Table 6. Multiple linear regression coefficients and R^2 values for calculated product stability based on computed reactant properties.

ΔG_{prod} (R^2)	a ($q(C_a)$)	b ($q(C_b)$)	c (E_σ)	d (E_π)	e (E_{π^*})	f ($\Delta E_{\pi \rightarrow \pi^*}$)	g (ω)	h (E_{steric})	i (MR_b)	j
0.66	23.18	17.19	152.38	796.25	−1115.81	883.25	−242.06	−0.20	0.18	−158.06
0.66	22.05	18.84			−103.64		−144.59	−0.41	0.17	80.17
0.63	21.39	18.25						−0.46	0.18	76.54
0.57	23.50	19.68						−0.28		68.05
0.21	12.15									45.39

Methods

An automated workflow was developed to generate, model, and analyze the reactants, intermediates, and products of the full dataset (Figure 3). The dataset consisted of olefins with all combinations of hydrogen, methyl, methyl ester, N-methyl amido, and cyano substituents. Methanethiol was used as the model thiol reactant. Within this system, the thiolate undergoes addition on a fixed "first" olefin carbon to avoid duplicate stereoisomers. In each reaction, the thiol site of attack was denoted as C_b, and the adjacent carbon forming a double bond with it is designated C_a. A Python script was used to generate the SMILES strings [31] corresponding to the reactants, intermediates, and products. The Open Babel program was used to generate the three-dimensional structure for each molecule [32].

Molecular topologies were generated using the Antechamber program [33]. Parameters for the reactants and products were assigned from the Generalized Amber Force Field (GAFF) [34]. A replica exchange molecular dynamics simulation (REMD)[35] was performed to find the lowest energy conformation using CHARMM version c36b2 [36]. The CHARMM program was used to run parallel tempering with the GAFF force field using a temperature range of 10–1000 K for 20 different replicas of each of the 325 reactant and product structures [37]. The GAFF force field barriers for rotation around the X–C=C–X and C(=O)–NH were increased

to 100 kcal mol^{-1} to prevent cis-trans isomerization of these bonds in the high temperature replicas. The C_b–S bond length in the products was restrained to 1.9 Å to avoid dissociation at high temperatures. The parallel tempering simulations were performed using Langevin dynamics with a friction coefficient of 50 ps^{-1}. The simulations were 1 ns in length with a 2 fs time step. Exchanges were attempted every 100 steps. The structure of each replica was energy minimized at the end of the parallel tempering simulation and the lowest energy conformation was used as the initial structure for the DFT calculations.

The REMD simulations of the intermediates required a different protocol; the electronic effects and atypical chemical bonding present in the intermediates necessitates the use of a quantum mechanical method, so the AM1 semi-empirical method was used [38]. The C_b–S bond lengths were restrained to 1.9 Å, which is a typical length for a carbanion intermediate [22]. This restraint prevents the C_b–S bond from dissociating in the high temperature replicas. The parallel tempering simulations for the intermediates were 0.2 ns in length with a 2 fs time step. Exchanges were attempted every 100 steps.

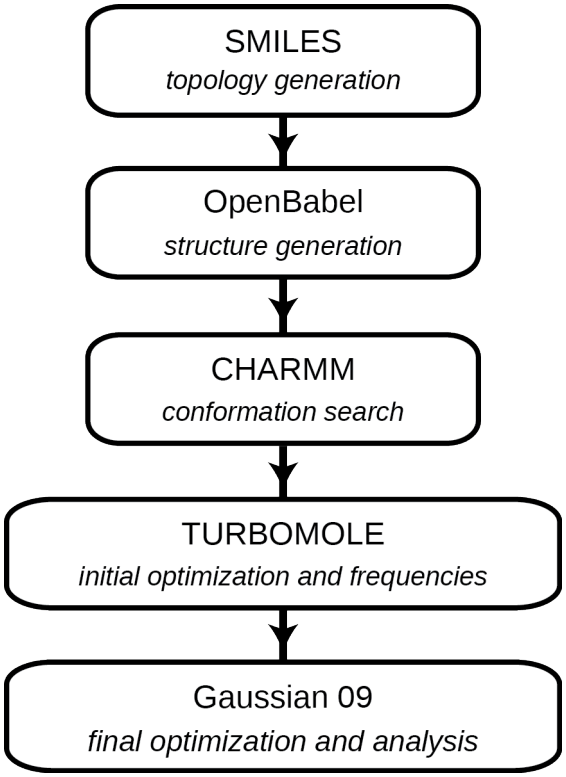


Figure 3. Computational workflow for the calculation of energies and properties for the reactants, intermediates, and products.

For each species, the lowest-energy conformation was optimized again using DFT methods. A rapid initial optimization was performed using TURBOMOLE 6.6 [39] using the PBE0 [40] exchange-correlation functional and the TZVP basis set [41]. The C_b–S bond length of the intermediates was constrained to 1.9 Å in this optimization to prevent the thiolate from dissociating. The resulting structures were optimized again using Gaussian 09 with the ωB97X-D[42] functional and TZVP basis set. This functional has been found to be very accurate for calculations involving thiol-Michael additions [22]. Frequency analysis was performed at the optimized structures to confirm that they were true minima and to compute thermal corrections.

This final optimization step in the intermediate does not constrain the C_b–S bond length like in the conformation search. As a result, many of the intermediate structures reverted to a non-covalent complex formed with the thiolate instead of a carbanion intermediate in this optimization. These molecules were reconstructed using Spartan '10 software [43] and a conformational search was performed with the C_b–S bond length constrained to 1.9 Å. A subsequent optimization was performed using the same level of theory with the constraint removed to allow the structure to relax to a true minimum. This procedure allows us to find some unstable intermediate structures that could not be located by the automatic procedure, but in most of these cases, no stable intermediate could be found by any method.

Calculation of the stability of the carbanions with respect to the thiol reactant requires the absolute solvation Gibbs energy of the proton, which is a property that is difficult to compute and has been the subject of extensive debate [44]. To avoid this issue, we estimate the deprotonation Gibbs energy of methanethiol from its experimentally-determined pK_a of 10.4 [45]. This corresponds to a Gibbs energy of 13.6 kcal mol^{-1} at 25 °C, which was added to the calculated Gibbs energy of the carbanion relative to the olefin and thiolate. This energy is reported as ΔG_{int}.

Single-point energy calculations were performed based on these optimized geometries employing the aug-cc-pVTZ basis set [46,47], denoted as ωB97X-D/aug-cc-pVTZ//ωB97X-D/TZVP. Single-point energies were also obtained using the polarizable continuum model (PCM) [48] to estimate the solvation free energy of

each species in water. Natural Bond Orbital (NBO) analysis was performed for all reactants using NBO version 3.1 [49,50].

The calculated total energies were defined by:

$$\Delta G_{total} = \Delta E_{aug} + \Delta G_{corr} + \Delta G_{solv} \quad 2)$$

where ΔE_{aug} represents the energy calculated from ωB97X-D/aug-cc-pVTZ//ωB97X-D/TZVP. ΔG_{corr} is the term for thermal correction to the Gibbs energy. ΔG_{solv} is the calculated hydration energy at the ωB97X-D/TZVP calculated using the Polarizable Continuum Model (PCM) [51].

Conclusions

A large set of model electrophilic substituted alkenes was generated using a automated computational procedure. For these compounds, thiol-ene addition reactions were modeled based on a Michael addition of methanethiol, with intermediate and product stabilities obtained for all reactions.

In general, intermediate stabilities showed strong dependence on stabilization of the anionic charge at the ‘a’-carbon by electron withdrawing ‘a’-substituents and also slight steric influence at the ‘a’-carbon. The stability of the products showed a weaker dependence on electronic effects in that some reactions with no carbanion intermediate formed highly stable products. Steric effects, however, involving ‘b’-substitution and cis versus trans isomers were deemed to have more influence on the product stability. In general, electrophiles with bulky substituents in a cis configuration will yield more stable thiol addition products than those in a trans configuration.

Ideally, a structure-activity relationship could be used to predict the reactivity of an olefin based on its chemical properties. The stabilities of the intermediates and products of these reactions were analyzed and correlated with NBO orbital energies, NPA charges, the electrophilicity index and steric parameters. Regression analysis displayed single parameters alone as having a poor predictive performance for both intermediate and product stabilities. Further analysis with multi-linear regression using different combinations of these parameters showed considerably good statistics

for predicting the intermediate stability. A model using the $q(C_a)$, ω, and MR_a, descriptors yielded an R^2 value of 0.83.

Product stability did not show as strong of a correlation with simple reactant descriptors where only R^2=0.63 could be obtained based on four parameters ($q(C_a)$, $q(C_b)$, E_{steric}, and MR_b). Nevertheless, reasonable estimates of product stability can be made based on calculated reactant descriptors. While these regression models based on reactant descriptors may provide some guidance in warhead design, it may be more sensible to directly calculate the stability of the carbanion intermediates and thioether products, as these calculations are not overly computationally demanding. This approach could be useful in screening compounds that cannot easily be synthesized or where it is difficult to determine their reactivity experimentally.

The full set of moieties that could be used as an electrophilic warhead in a covalent-modifying inhibitor is vast and this database encompasses only a small selection of possible electrophilic olefin warheads. Expanding this database to include other common functional groups, such as aryl substituents, is an ongoing effort. The accumulation of experimental data on irreversible inhibition will allow the utility of these computed properties to be evaluated quantitatively.

Acknowledgements

We thank NSERC of Canada for funding through a Discovery Grant (Application No. 418505-2012) and an Undergraduate Student Research Award for JMS. JMS thanks the Dean of Science of Memorial University for a travel grant. Computational resources were provided by the Compute Canada Consortium (CCI: djk-615-ac). We thank Archita Adluri for proofreading the manuscript. We thank the reviewers for meticulous examination of the manuscript.

References

1. Enoch SJ, Ellison CM, Schultz TW, Cronin MTD (2011) Critical Reviews in Toxicology 41(9):783
2. Potashman MH, Duggan ME (2009) Journal of Medicinal Chemistry 52(5):1231
3. Gersch M, Kreuzer J, Sieber SA (2012) Natural Product Reports 29(6):659

4. Miller RM, Taunton J. Chapter Four - Targeting Protein Kinases with Selective and Semipromiscuous Covalent Inhibitors. In: Kevan MS (ed). Methods in Enzymology. Volume Volume 548: Academic Press, 2014:93
5. Rabindran SK, Discafani CM, Rosfjord EC, Baxter M, Floyd MB, Golas J, Hallett WA, Johnson BD, Nilakantan R, Overbeek E, Reich MF, Shen R, Shi X, Tsou H-R, Wang Y-F, Wissner A (2004) Cancer Research 64(11):3958
6. Zhou W, Hur W, McDermott U, Dutt A, Xian W, Ficarro SB, Zhang J, Sharma SV, Brugge J, Meyerson M, Settleman J, Gray NS Chemistry & Biology 17(3):285
7. Pan Z, Scheerens H, Li S-J, Schultz BE, Sprengeler PA, Burrill LC, Mendonca RV, Sweeney MD, Scott KCK, Grothaus PG, Jeffery DA, Spoerke JM, Honigberg LA, Young PR, Dalrymple SA, Palmer JT (2007) ChemMedChem 2(1):58
8. Ettari R, Micale N, Schirmeister T, Gelhaus C, Leippe M, Nizi E, Di Francesco ME, Grasso S, Zappalà M (2009) Journal of Medicinal Chemistry 52(7):2157
9. Serafimova IM, Pufall MA, Krishnan S, Duda K, Cohen MS, Maglathlin RL, McFarland JM, Miller RM, Frödin M, Taunton J (2012) Nat Chem Biol 8(5):471
10. Mulliner D, Wondrousch D, Schuurmann G (2011) Organic & Biomolecular Chemistry 9(24):8400
11. Hoyle CE, Bowman CN (2010) Angewandte Chemie International Edition 49(9):1540
12. Lutolf MP, Tirelli N, Cerritelli S, Cavalli L, Hubbell JA (2001) Bioconjugate Chemistry 12(6):1051
13. Nair DP, Podgórski M, Chatani S, Gong T, Xi W, Fenoli CR, Bowman CN (2014) Chemistry of Materials 26(1):724
14. Krishnan S, Miller RM, Tian B, Mullins RD, Jacobson MP, Taunton J (2014) Journal of the American Chemical Society 136(36):12624
15. London N, Miller RM, Krishnan S, Uchida K, Irwin JJ, Eidam O, Gibold L, Cimermančič P, Bonnet R, Shoichet BK, Taunton J (2014) Nat Chem Biol 10(12):1066
16. Schwöbel JAH, Wondrousch D, Koleva YK, Madden JC, Cronin MTD, Schüürmann G (2010) Chemical Research in Toxicology 23(10):1576
17. Krenske EH, Petter RC, Zhu Z, Houk KN (2011) The Journal of Organic Chemistry 76(12):5074
18. Capoferri L, Lodola A, Rivara S, Mor M (2015) Journal of Chemical Information and Modeling 55(3):589
19. Irwin JJ, Shoichet BK, Mysinger MM, Huang N, Colizzi F, Wassam P, Cao Y (2009) Journal of Medicinal Chemistry 52(18):5712
20. Del Rio A, Sgobba M, Parenti M, Degliesposti G, Forestiero R, Percivalle C, Conte P, Freccero M, Rastelli G (2010) Journal of Computer-Aided Molecular Design 24(3):183
21. Ouyang X, Zhou S, Su CTT, Ge Z, Li R, Kwoh CK (2013) Journal of Computational Chemistry 34(4):326
22. Smith JM, Jami Alahmadi Y, Rowley CN (2013) Journal of Chemical Theory and Computation 9(11):4860

23. Hori K, Higuchi S, Kamimura A (1990) The Journal of Organic Chemistry 55(23):5900
24. Bolton EE, Wang Y, Thiessen PA, Bryant SH. Chapter 12 - PubChem: Integrated Platform of Small Molecules and Biological Activities. In: Ralph AW, David CS (eds). Annual Reports in Computational Chemistry. Volume Volume 4: Elsevier, 2008:217
25. Schmidt TJ, Ak M, Mrowietz U (2007) Bioorganic & Medicinal Chemistry 15(1):333
26. Edwards PM (1975) British Journal of Industrial Medicine 32(1):31
27. Guengerich FP, Geiger LE, Hogy LL, Wright PL (1981) Cancer Research 41(12 Part 1):4925
28. Schwartz PA, Kuzmic P, Solowiej J, Bergqvist S, Bolanos B, Almaden C, Nagata A, Ryan K, Feng J, Dalvie D, Kath JC, Xu M, Wani R, Murray BW (2014) Proceedings of the National Academy of Sciences 111(1):173
29. Parr RG, Szentpály Lv, Liu S (1999) Journal of the American Chemical Society 121(9):1922
30. J.B. Taylor PDK (1981) **Introductory Medicinal Chemistry**. Ellis Horwood Limited, Chichester
31. Weininger D (1988) Journal of Chemical Information and Computer Sciences 28(1):31
32. O'Boyle N, Banck M, James C, Morley C, Vandermeersch T, Hutchison G (2011) Journal of Cheminformatics 3(1):33
33. Wang J, Wang W, Kollman PA, Case DA (2006) Journal of Molecular Graphics and Modelling 25(2):247
34. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Journal of Computational Chemistry 25(9):1157
35. Earl DJ, Deem MW (2005) Physical Chemistry Chemical Physics 7(23):3910
36. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M (2009) Journal of Computational Chemistry 30(10):1545
37. Rowley, C.N. CHARMM Conformation Search; https://github.com/RowleyGroup/charmm-conformation (accessed May 15, 2015).
38. Dewar MJS, Thiel W (1977) Journal of the American Chemical Society 99(15):4899
39. TURBOMOLE V6.6 2014, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com.
40. Adamo C, Barone V (1999) The Journal of Chemical Physics 110(13):6158
41. Schäfer A, Huber C, Ahlrichs R (1994) The Journal of Chemical Physics 100(8):5829
42. Chai J-D, Head-Gordon M (2008) The Journal of Chemical Physics 128(8):084106
43. Spartan '10. Irvine, CA: Wavefunction Inc., 2010

44. Lamoureux G, Roux B (2006) The Journal of Physical Chemistry B 110(7):3308
45. Kreevoy MM, Eichinger BE, Stary FE, Katz EA, Sellstedt JH (1964) J Org Chem 29(6):1641
46. Dunning TH (1989) The Journal of Chemical Physics 90(2):1007
47. Woon DE, Dunning TH (1993) The Journal of Chemical Physics 98(2):1358
48. Tomasi J, Mennucci B, Cammi R (2005) Chemical Reviews 105(8):2999
49. Reed AE, Weinstock RB, Weinhold F (1985) The Journal of Chemical Physics 83(2):735
50. NBO Version 3.1, E. D. Glendening, A. E. Reed, J. E. Carpenter, and F. Weinhold.
51. Mennucci B (2012) Wiley Interdisciplinary Reviews: Computational Molecular Science 2(3):386