University of Cape Town

Department of Chemistry

Development of a group contribution method to predict the enthalpy of formation of energetic azoles

Megan Coetzee

Supervisor: Dr. Gerhard A. Venter

November 2025

Declaration:

1. I know that plagiarism is wrong. Plagiarism is using another's work and pretending it is one's own.
2. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.
3. This project is my own work.
4. I have included internet article, book, or other material references used for this project.

Signed: Megan Coetzee

Date: 2025-11-29

Development of a group contribution method to predict the enthalpy of formation of energetic azoles

Megan Coetzee

Abstract

Energetic materials (EMs) are a class of high-nitrogen content material that stores a large amount of chemical energy released upon external stimuli (e.g., friction, thermal shock, or electrostatic discharge). Azoles, which are five-membered heterocyclic aromatic compounds containing a nitrogen atom and at least one other heteroatom, are an ideal source of EMs. The enthalpy of formation ($\Delta_f H$) is a critical thermodynamic quantity that is required to estimate the detonation performance of EMs (e.g., detonation pressure and velocity). Because physical measurements of $\Delta_f H$ require time and resources, group contribution methods (GCMs) provide an alternative that is quick and cost-effective.

A GCM is built on the principle that a property of a molecule can be estimated using the sum of individual contributions associated with its smaller structural units, or groups. The enthalpy of formation of a cyclic compound can be obtained by summing the corresponding group additive values (GAV) determined from the acyclic reference molecules and adding a correction accounting for the ring strain (RS). This ring-strain correction (RSC) is generally positive and is calculated from the difference between the known value of $\Delta_f H$ of the cyclic compound and the sum of the acyclic GAVs. However, if $\Delta_f H$ of an aromatic compound is calculated in the same way, the resulting correction must account for both the RS *and* the aromatic stabilisation energy (ASE), where this sum is typically negative. Alternatively, an aromatic compound may also be built directly as a sum over GAVs that have been determined using aromatic reference molecules. The former model

has been applied to pyrrole derivatives only, and to the best of our knowledge, although the latter model has been applied to azines, it has not been attempted for azoles.

In this work, GCMs based on both approaches were developed using *ab initio* quantum mechanical (QM) calculations due to the limited availability of physical measurements of $\Delta_f H$ of azoles. The uncertainty arising primarily from the systematic error or bias associated with the G4(MP2)-6X composite method was first quantified using 47 neutral CHN-containing molecules. Reference $\Delta_f H$ values for these molecules were obtained from Active Thermochemical Tables (ATcT). A correction of $+1.11\,\mathrm{kJ\,mol^{-1}}$ resulted, while the standard uncertainty associated with this correction is $3.04\,\mathrm{kJ\,mol^{-1}}$, leading to a 95 % confidence interval of $6.08\,\mathrm{kJ\,mol^{-1}}$ for the calculated values. Subsequently, a database of 72 acyclic molecules and multiple linear regression (MLR) was used to determine 42 Benson-type GAVs, which were then applied to estimate $\Delta_f H$ of ten unsubstituted azoles, including pyrrole, two diazoles, four triazoles, two tetrazoles, and pentazole. The sum of RS and ASE was determined as the difference between the calculated $\Delta_f H$ values and the sum of the acyclic GAVs, providing a GCM based on the strain-centred approach. To extend the work to energetic azoles, the aromatic group approach was then applied and a set of 33 GAVs was obtained using MLR of a database of the calculated $\Delta_f H$ values of 60 energetic azoles. In addition to unsubstituted azoles, these azoles were also functionalised with a methyl group and explosophoric functional groups such as amine, azide, nitro, and nitramide. The resulting GCM showed a mean absolute error (MAE) of $3.22\,\mathrm{kJ\,mol^{-1}}$ and maximum error of $10.95\,\mathrm{kJ\,mol^{-1}}$ for the enthalpy of formation across all functionalised and unfunctionalised azoles.

Acknowledgements

I would first like to sincerely thank my supervisor, Dr. Gerhard Venter, for the invaluable guidance and continuous support throughout this research.

Acknowledgement is given to the Centre for High-Performance Computing (CHPC), South Africa, for providing computational resources to this research project.

Acknowledgement is given to the UCT Centre for High-Performance Computing (UCT CHPC) for providing computational resources to this research project.

I am truly grateful to my partner, Brandon, for your unwavering support, encouragement, and understanding through this journey, as your belief in me has been a constant source of motivation every step of the way.

I would like to acknowledge my parents, Jerry and Tania, for their presence and support throughout this journey, even in the smallest ways.

To my brother, Shea, a heartfelt thank you for everything you did; your presence was truly invaluable throughout this journey.

This work is based on the research supported wholly by the National Research Foundation of South Afica (Grant Number 140976).

Contents

List of Abbreviations

Abbreviation	Meaning
EM	Energetic Material
NG	Nitroglycerine
NC	Nitrocellulose
PA	Picric Acid
TNT	2,4,6-Trinitrotoluene
RDX	Royal Demolition Explosive
PETN	Pentaerythritol Tetranitrate
HMX	High Melting Explosive
HEDMs	High-energy Density Materials
CL-20	Hexanitrohexaazaisowurtzitane
MTNI	1-Methyl-2,4,5-trinitroimidazole
FPD	Feller-Petersen-Dixon
ZPE	Zero-point Energy
DFT	Density-functional Theory
HLC	Higher-level Corrections
ccCA	Correlation-Consistent Composite Approaches
QSPR	Quantitative Structure-Property Relationship
GCM	Group Contribution Method
JR	Joback and Reid
RSC	Ring Strain Correction
NIST	National Institute of Standards and Technology
HBI	Hydrogen Bond Increments
MAD	Mean Absolute Deviation
RE	Resonance Energy

ASE	Aromatic Stabilisation Energy
NICS	Nucleus Independent Chemical Shift
GAVs	Group Additive Values
ISO	International Organisation of Standardisation
GUM	Guide to the Experession of Uncertainty in Measurement
ATcT	Active Thermochemical Tables
ABC	Artificial Bee Colony
MLR	Multiple Linear Regression
MAE	Mean Absolute Error
GOAT	Global Optimization ALgorithm
MAPE	Mean Absolute Percentage Error
AE	Absolute Error
APE	Absolute Percentage Error
IG	Isogyric Reactions
ID	Isodesmic Reactions
HD	Homodesmotic Reactions
TN	Thermochemical Network
MP	Møller-Plesset
CC	Coupled-Cluster
HF	Hartree-Fock
MOs	Molecular Orbitals
CI	Configuration Interaction
MBPT	Many-Body Perturbation Theory
KS	Kohn-Sham
LDA	Local Density Approximation
LSDA	Local Spin Density Approximation
GGA	Generalised Gradient Approximation

DHDFT	Double Hybrid Density Functional Theory
HLC	High-level Correction
FC	Frozen Core

1 Introduction

This chapter starts with a brief overview of the most notable energetic materials, outlining their development and impact, followed by a discussion on their applications and fundamental properties. Then, a description of azoles is given, which highlights their structural characteristics and relevance to energetic materials. The chapter then explores methods for predicting the enthalpy of formation of energetic salts, focusing on both the theoretical approach and group contribution methods. Additionally, a brief discussion on the determination of the aromatic stabilisation energy is included. Finally, the chapter concludes with the aims and objectives.

1.1 Energetic Materials

Energetic material (EM) is a class of high-nitrogen content compounds that store a large amount of chemical energy, which is released upon its decomposition through friction, thermal shock, and electrostatic discharge.[1] Propellants, explosives, and pyrotechnics are examples of EMs that are used in a range of applications from military functions to civilian

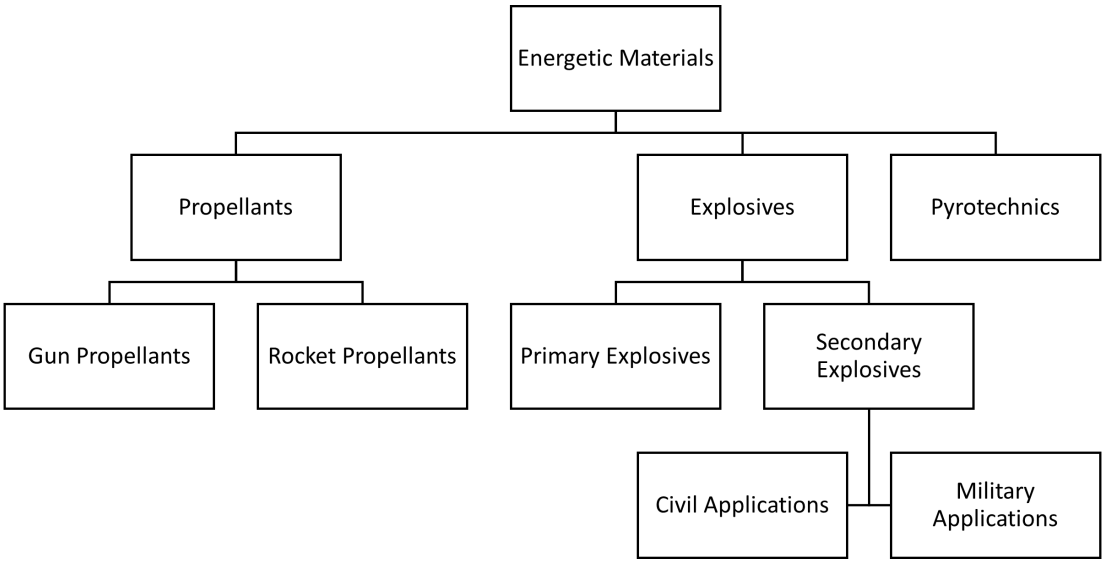


Fig. 1.1. The classification of energetic materials and their applications.

operations (Figure 1.1).[2] The origin of EMs can be traced back to a report in China around 220 BC that involved black powder, also known as gunpowder.[3] Chinese alchemists attempted to separate gold and silver by adding sulfur and potassium nitrate to gold ore in a furnace and neglected to add charcoal in the first step of the reaction. Charcoal was then added to the reaction to rectify this oversight, resulting in an explosion and marking the first reported encounter of an EM. Gunpowder remained overlooked until the 13th century when its properties were researched by Roger Bacon, an English monk, and Berthold Schwarz, a German monk, who introduced it into the military domain at the end of the century.[2] However, it was not until gunpowder was refined by corning, the process of compressing, crushing, and screening gunpowder based on particle size to produce uniform grains that enhance its stability and performance, that it was introduced as propellant charges in small and large calibre guns.[3]

The synthesis of nitroglycerine (NG) in 1846 by Sobrero[4] was the next notable advancement of EMs, which was investigated due to the difficulty in mining and tunnelling operations of black powder.[3] Immanual and Alfred Nobel addressed NG's tendency to accidentally initiate by developing metal blasting cap detonators and replaced black powder with mercury fulminate to initiate NG. Following a devastating explosion, Alfred reduced the sensitivity of NG by mixing it with an absorbent clay called Kieselguhr to produce a mixture called "Guhr Dynamite".[5] While researching NG, Schönbein and Böttger carried out the nitration of cellulose to make nitrocellulose (NC). Subsequently, Alfred Nobel observed that mixing NC with NG formed a gel and refined it to create blasting gelatine, gelatine dynamite, and then the first smokeless powder called ballistite, which is still used today as a rocket propellant.[2]

The two most commonly used polynitroaromatic compounds used in World War I were trinitrophenol, also known as picric acid (PA), and 2,4,6-trinitrotoluene (TNT).[6] Glauber developed PA in 1742, and its extremely explosive nature surpassed even that of TNT.[2] A limitation of pure PA is its propensity to form metal salts that are impact-sensitive when in

contact with shell walls, which was addressed with the introduction of TNT and led to its widespread use due to its cost-effective production and low sensitivity to shock and friction. During World War II, cyclotrimethylene trinitramine, also referred to as hexogen or Royal Demolition Explosive (RDX), and pentaerythritol tetranitrate (PETN) were the most commonly used explosives. The military application of PETN has widely been replaced by RDX due to its sensitivity in its purest form and comparatively low chemical stability. Following World War II, research and development into new and more powerful explosive compositions resulted in cyclotetramethylenetetranitramine, also referred to as octogen or High Melting Explosive (HMX). To date, most high-energy material compositions used for military applications are based on TNT, RDX, and HMX (Figure 1.2). For example, octol is a heterogeneous solid explosive that consists of roughly 76 % HMX and 24 % TNT.[7]

1.1.1 Applications of Energetic Materials

To meet the high demand to improve the performance of existing products, many new EMs have been developed.[8] High-energy density materials (HEDMs) refer to EMs that derive their energy either through the combustion of the carbon backbone, in the same manner as traditional explosives, or the high positive enthalpy of formation.[9] Hexanitrohexaazaisowurtzitane, or CL-20, is an example of an HEDM that derives its energy from its high

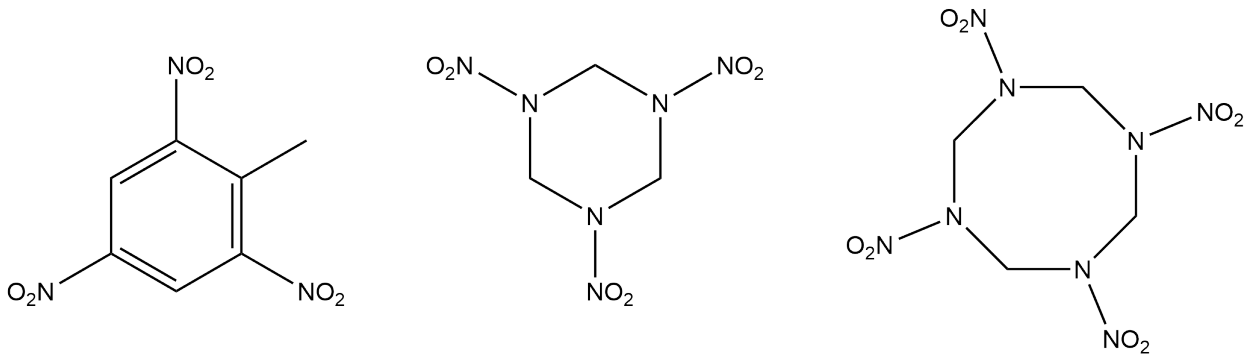


Fig. 1.2. Structures of 2,4,6-Trinitrotoluene (TNT), hexogen (RDX), and octogen (HMX).

enthalpy of formation and close packing of the constituent atoms in a cage-like structure, which increases the density of the HEDM.[10] Density is a critical property of HEDMs that significantly affects the detonation velocity, and the detonation pressure increases proportionally to the square of the density (see Section 1.1.2). In CL-20, the six $N-NO_2$ groups present also contribute to its high energy content. High-nitrogen EMs belong to a class of advanced HEDMs that derive their energy from the large number of N−N and C−N bonds that exhibit large enthalpies of formation coupled with the availability of adjacent nitrogen atoms set to form nitrogen gas (N_2).[11] The formation of N_2(g) releases a large amount of energy, which is driven by the large difference in average bond energies of N−N ($160\,kJ\,mol^{-1}$) and N=N ($418\,kJ\,mol^{-1}$) in comparison to N≡N ($954\,kJ\,mol^{-1}$).[12] The application of HEDMs depends on the energy-releasing process: either deflagration or detonation.[2] Propellants, also known as deflagrating explosives, undergo decomposition in a closed chamber through a thermal process called deflagration, converting stored chemical energy into hot gas to produce sufficiently high temperatures and pressure imparting a propulsive force for the acceleration of an object. Propellants, within the context of energetic materials, have applications as gun propellant charges and as components of rocket fuels.[13,14] Detonation explosives, more frequently referred to simply as explosives, decompose through a supersonic shock wave on a microsecond timescale, whereas deflagration is sub-sonic. Explosives have applications in industry, such as cratering, blasting, and the manipulation of metals.[15–17] Pyrotechnics decompose through a self-sustained exothermic chemical reaction to produce an effect by light, sound, heat, gas, or smoke.[11] While explosives function at the highest speed of exothermic reaction to produce gaseous products and propellants function at comparatively slower speeds, pyrotechnics function at a noticeably observable rate and produce mainly solid as well as gaseous products.[2]

1.1.2 Properties of Energetic Materials

There are two classes of explosives: primary and secondary explosives. Primary explosives are very sensitive to external stimuli, while secondary explosives exhibit a rapid transition from deflagration to detonation.[2] Primary explosives create either a large amount of heat or a shockwave, which facilitates the detonation of a less sensitive secondary explosive. While primary explosives are more sensitive, they generally exhibit lower detonation velocities, detonation pressures, and heats of explosion than that of secondary explosives. The most frequently used primary explosives are lead azide ($Pb(N_3)_2$) and lead styphnate ($C_6HN_3O_8Pb$).[18,19] Secondary explosives, or high explosives, are more powerful than primary explosives but less sensitive to external stimuli and cannot be initiated through heat or shock. The most frequently used single-component secondary explosives used for military applications are TNT, RDX, HMX, NG, and NC. Primary explosives are necessary for the initiation of secondary explosives as their shockwave activates the secondary explosive. However, secondary explosives have greater performance than primary explosives (Table 1.1), where the performance criteria are detonation pressure (P_D), detonation velocity (D), and the heat of explosion (Q). Kamlet and Jacobs[20] devised a semi-empirical approach to determine D and P_D using,

$$D = 1.01\phi^{1/2}(1 + 1.30\rho) \tag{1.1}$$

$$P_D = 15.58\phi\rho \tag{1.2}$$

where the Kamlet-Jacobs parameter is given by,

$$\phi = NM_{\text{ave}}^{1/2}Q^{1/2} \tag{1.3}$$

where ρ is the density at which the explosive is loaded, N is the moles of gaseous detonation products per gram of explosive, and M_{ave} is the average molecular mass of gaseous det-

Table 1.1. Representative performance data for primary and secondary explosives.[a]

Performance data	Typical primary explosive	$Pb(N_3)_2$	Typical secondary explosive	RDX
Detonation velocity	3500-5500	4600-5100	6500-9000	8750
Detonation pressure	N/A	343	210-390	347
Heat of explosion	1000-2000	1639	5000-6000	5277

[a] Detonation velocity in $m\,s^{-1}$, detonation pressure in kbar, and heat of explosion in $kJ\,g^{-1}$.

onation products.[21] Following detonation, an explosive with mass, M, generates n moles of gaseous products, such that,

$$N = \frac{n}{M}, \tag{1.4}$$

and,

$$M_{\text{ave}} = \frac{M}{n}. \tag{1.5}$$

However, CHNO explosives may produce solid carbon following detonation, such that,

$$M_{\text{ave}} = \frac{M - C}{n}, \tag{1.6}$$

where C is mass of condensed carbon (graphite) per mole of explosive.[22]

The heat of detonation (Q) is typically considered to be the negative of the enthalpy change (ΔH) for the overall decomposition of products following detonation,

$$Q = -\frac{1}{M_y}\left[\sum_i n_i \Delta_{\text{f}} H(i) - \Delta_{\text{f}} H(Y)\right], \tag{1.7}$$

where M_Y is the molecular mass of the explosive, n_i is the total number of moles of the detonation products i that has a molar enthalpy of formation $\Delta_{\text{f}} H(i)$, and $\Delta_{\text{f}} H(Y)$ is the enthalpy of formation of explosive Y (see Section 1.2).[21,23] Understanding the composition of the detonation products is necessary to determine the detonation parameters, and a key property of secondary explosives is their oxygen balance.

In the design of explosives, oxygen balance (Ω) measures the extent to which the oxygen content in an explosive can be oxidised and is used to determine whether there is a desirable balanced ratio between the oxidiser and the combustible components.[22] A CHNO explosive with a balanced oxygen balance ($\Omega = 0$) is entirely converted into H_2O(g), CO_2(g), and N_2(g). The oxygen balance for explosives with the general formula $C_aH_bN_cO_d$ can be calculated using,

$$\Omega = \Omega_{CO_2} = \frac{\left[d - 2a - \frac{b}{2}\right] \times 1600}{M}, \tag{1.8}$$

where M is the molecular mass of the explosive.[2] However, if secondary explosives are oxygen deficient (Ω<0 %), the detonation products will include N_2 and H_2O, along with a combination of CO(g), CO_2(g), C(s), and H_2(g).[22] To counteract the oxygen deficit, an oxidant may be added to create a formulation with an overall improved Ω.

The oxygen balance alone is not sufficient to estimate the thermodynamics of a reaction. The detonation products must be known or estimated, which can be done approximately using the modified Springall-Roberts rules.[2] The rule is as follows: (i) Carbon and oxygen are converted into gaseous carbon monoxide, (ii) The remaining oxygen atoms oxidise hydrogen gas to gaseous water, (iii) Any further oxygen atoms that remain oxidise the carbon monoxide to carbon dioxide gas, (iv) all nitrogen atoms are converted to dinitrogen gas, (v) A third of the carbon monoxide formed is converted into solid carbon atoms and carbon dioxide gas, (vi) A sixth of the carbon monoxide formed in Step (i) is converted into solid carbon atoms and water.

1.2 Introduction to Enthalpy of Formation

The enthalpy change (ΔH) for a process that only involves PV work is quantified by measuring the flow of heat at a constant pressure (q_P),

$$\Delta H = H_f - H_i = q_P, \tag{1.9}$$

where f and i are final and initial states, respectively. The enthalpy of reaction ($\Delta_\mathrm{r}H$) is the change in enthalpy during a chemical reaction at constant temperature and pressure. Furthermore, the standard enthalpy of reaction ($\Delta_\mathrm{r}H^\ominus$) is the enthalpy of reaction observed when the reactants and products are under standard state conditions of 298.15 K and 1 bar. While determining $\Delta_\mathrm{r}H^\ominus$ experimentally can be done, compiling a complete table for all possible chemical reactions would be a monumental task. Consider a reaction between solid magnetite (Fe_3O_4) and hydrogen gas (H_2) to produce solid iron and liquid water,

$$\mathrm{Fe_3O_4(s)} + 4\mathrm{H_2(g)} \longrightarrow 3\mathrm{Fe(s)} + 4\mathrm{H_2O(l)}, \tag{1.10}$$

as illustrated in the top reaction equation in Figure 1.3. In a suitable reaction vessel, the heat flow can be measured and equated to the enthalpies of the reactants and products,

$$\begin{aligned} \Delta_\mathrm{r}H^\ominus &= H^\ominus_\text{products} - H^\ominus_\text{reactants} \\ &= 3H^\ominus_\mathrm{m}(\mathrm{Fe,s}) + 4H^\ominus_\mathrm{m}(\mathrm{H_2O,l}) - H^\ominus_\mathrm{m}(\mathrm{Fe_3O_4,s}) - 4H^\ominus_\mathrm{m}(\mathrm{H_2,g}), \end{aligned} \tag{1.11}$$

where m refers to the molar enthalpies of the species. However, there is no unique reference zero from which to experimentally measure the absolute enthalpies of these species. As opposed to H, only ΔH can be experimentally determined. A more practical form of Equation 1.11 is required by introducing the enthalpy of formation, defined as the enthalpy change of a reaction where the reactants are in their standard reference state and the product is 1 mol of the target species, as shown by the species in boxes in Figure 1.3. A standard reference state is defined as pure elements in their most stable physical form under standard state conditions (e.g., H_2O(l) for water), as can be seen by the bottom reaction equation in Figure 1.3. The standard state should not be mistaken for the standard conditions, which IUPAC defines as a pressure of 1 bar and a temperature of 273.15 K or 0 °C. Although the pressure under standard conditions aligns with the standard pressure in thermodynamics, the temperature holds little significance in thermochemistry because

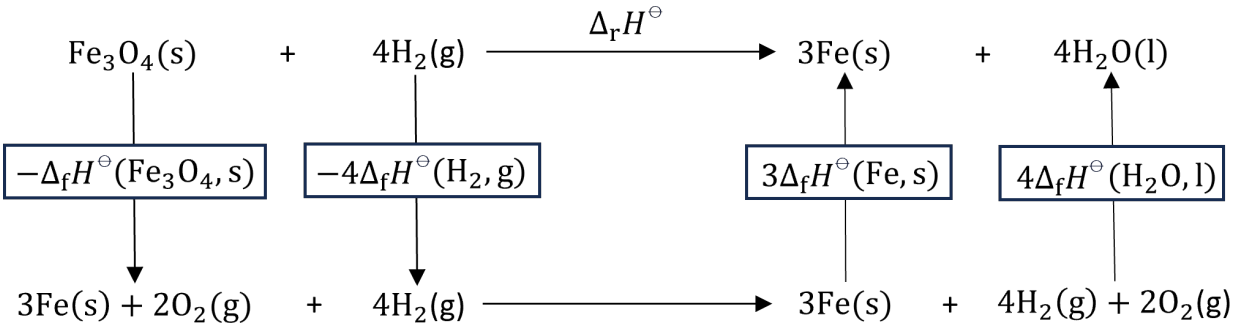


Fig. 1.3. Thermodynamic cycle for the reaction between solid magnetite (Fe_3O_4) and hydrogen gas (H_2) to produce solid iron and liquid water.

$\Delta_f H^\circ$ can be calculated at various temperatures, if the temperature of the products and reactants is the same.[24]

Based on the definition of enthalpy, it follows that since H_2(g) and Fe(s) are in their standard reference state in Equation 1.11, $\Delta_f H^\circ$ is zero for these species because the reactants and products are the same. Therefore, $\Delta_r H^\circ$ in terms of the formation reactions for the remaining reactant and product from enthalpy values are as shown,

$$\begin{gathered} \mathrm{Fe(s)} + 2\mathrm{O_2(g)} \longrightarrow \mathrm{Fe_3O_4(s)} \\ \Delta_r H^\circ = \Delta_f H^\circ(\mathrm{H_2O,l}) = H_m^\circ(\mathrm{H_2O,l}) - H_m^\circ(\mathrm{H_2,g}) - \frac{1}{2} H_m^\circ(\mathrm{O_2,g}), \end{gathered} \tag{1.12}$$

$$\begin{gathered} \mathrm{H_2(g)} + \frac{1}{2}\mathrm{O_2(g)} \longrightarrow \mathrm{H_2O(l)} \\ \Delta_r H^\circ = \Delta_f H^\circ(\mathrm{H_2O,l}) = H_m(\mathrm{H_2O,l}) - H_m(\mathrm{H_2,g}) - \frac{1}{2} H_m^\circ(\mathrm{O_2,g}), \end{gathered} \tag{1.13}$$

where each $\Delta_f H^\circ$ is the enthalpy difference between the compound and its elemental components, as opposed to an absolute enthalpy. Consequently, Equation 1.11 can be expressed in terms of the enthalpies of formation,

$$\Delta_r H^\circ = 4\Delta_f H^\circ(\mathrm{H_2O,l}) - \Delta_f H^\circ(\mathrm{Fe_3O_4,s}). \tag{1.14}$$

Expressing $\Delta_r H^\circ$ in terms of formation enthalpies significantly simplifies the determination compared to the collection of experimentally measured values for the reaction enthalpies. To optimise the research and development of novel energetic salts, computational methods have been developed to accurately predict the properties of proposed structures and display the potential of enhanced performance, reduced sensitivity, or a reduced environmental impact.

1.3 Azoles

Azoles are a class of five-membered heterocyclic ring compounds that contain one nitrogen atom and at least one other heteroatom, which have received considerable attention in the design of new energetic salts and ILs with their large number of N−N and C−N bonds that exhibit large positive enthalpies of formation. The azoles include pyrrole, pyrazole, imidazole, 1H-1,2,3-triazole, 1,2,4-triazole, tetrazole, and pentazole.

Pyrrole is characterized by an aromatic five-membered heterocyclic ring, consisting of four sp^2-hybridised carbon atoms, each contributing one π electron in a p orbital, and one sp^2-hybridised nitrogen atom, which contributes two π electrons from its lone pair into a p orbital perpendicular to the ring. The resonance structures of 1H-pyrrole are shown in Figure 1.4. Since nitrogen's lone pair forms part of the aromatic sextet, it is referred to as a "pyrrole-like" nitrogen (Figure 1.5). In comparison, pyridine contains five sp^2-hybridised carbon atoms, each containing a p orbital with a π electron, and an sp^2 hybridised nitrogen that contributes the remaining electron in the p orbital to the aromatic sextet. The lone pair of electrons on the nitrogen are in an sp^2 orbital in the plane of the ring and do not form part of the aromatic sextet (Figure 1.5). Therefore, this nitrogen is referred to as a "pyridine-like" nitrogen.

Diazole is defined by a five-membered aromatic ring with three carbon atoms and two nitrogen atoms. The two isomeric forms of diazole are 1,2-diazole, more commonly re-

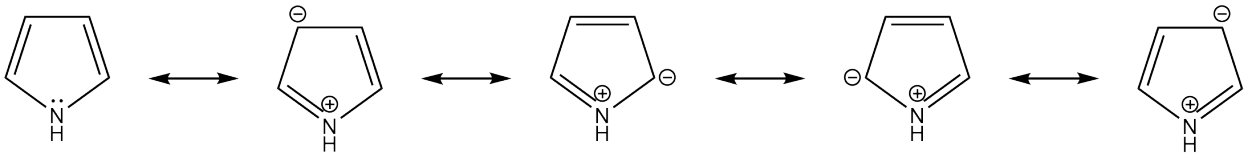


Fig. 1.4. Resonance structures of 1H-pyrrole.

ferred to as pyrazole, and 1,3-diazole, or imidazole, where the systematic name indicates the positions of the nitrogens in the ring. Diazoles contain both a "pyridine-like" nitrogen that contributes one π electron and a "pyrrole-like" nitrogen that contributes two π electrons, as shown in Figure 1.6.

As nitrogen-containing five-membered heterocyclic compounds, triazoles exist in four isomeric forms: 1H-1,2,3-triazole, 2H-1,2,3-triazole, 1H-1,2,4-triazole, and 4H-1,2,4-triazole. The four isomeric forms are separated into two pairs of tautomers. The 1,2,3-triazole isomers have three adjacent nitrogen atoms, while the 1,2,4-triazoles contain an interstitial carbon that separates a nitrogen atom from two adjacent nitrogen atoms (Figure 1.7). The 1,2,3-triazole parent molecule exists as 1H-1,2,3-triazole and 2H-1,2,3-triazole, which interconverts through a proton transfer between the neighbouring nitrogen atoms. The two tautomers of the 1,2,4-triazole parent molecule are 1H-1,2,4-triazole and 4H-1,2,4-triazole.

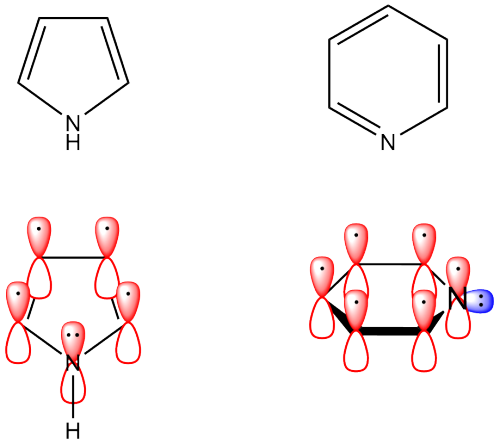


Fig. 1.5. Orbital structure diagram of 1H-pyrrole and pyridine showing the role of the nitrogen lone pair in aromaticity.

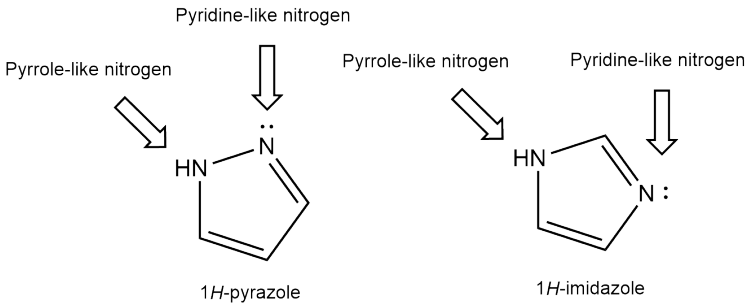


Fig. 1.6. Structures of 1H-pyrazole and 1H-imidazole showing the positions of the pyrrole-like and pyridine-like nitrogens.

The practical and theoretical relevance of tetrazoles, along with the diversity of their properties, makes them a key component of energetic materials. Composed of four nitrogen atoms and one carbon atom, the tetrazole ring is among the stable azoles with the highest number of nitrogen, whereas pentazole is highly explosive even at low temperatures.[25] Tetrazole has three isomers; $1H$-, $2H$-, and $5H$-tetrazole. With six π electrons, $1H$-tetrazole and $2H$-tetrazole are aromatic, while $5H$-tetrazole is nonaromatic.

Pentazoles are aromatic compounds that feature a five-membered ring comprised entirely of nitrogen atoms. The synthesis of pentazole has faced significant challenges, including a widely disputed initial synthesis report in 1915. Decades later, in 2003, was the parent pentazole synthesized by Butler, Stephens, and Burke.[26] Since then, significant

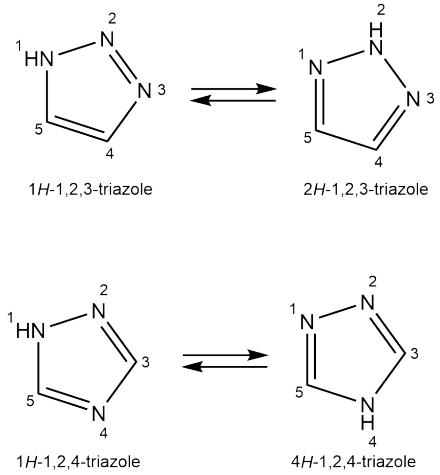


Fig. 1.7. Tautomeric forms of 1,2,3-triazole and 1,2,4-triazole.

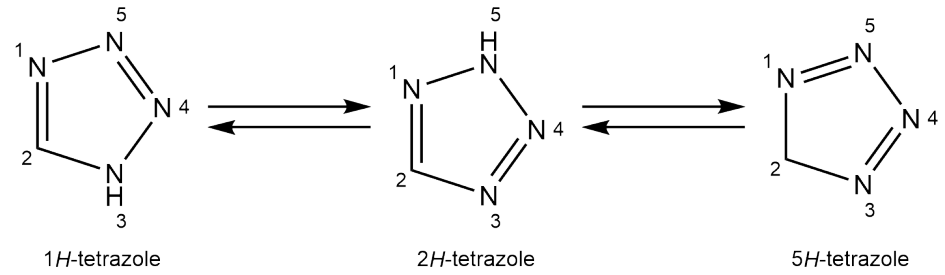


Fig. 1.8. Tautomeric forms of tetrazole.

advancements have been made in synthesising pentazole and pentazole derivatives. Due to the difficulty in synthesizing and detection of these compounds, there have been many theoretical investigations. In 2001, Mull et al.[27] studied the decomposition rate calculated across a temperature range to assess their kinetic stability. The findings suggest that pentazole is unlikely to serve as a HEDM as it would require temperatures close to 200 K to achieve sufficient stability, which indicates that pentazole's thermal stability may not be sufficient for practical application in energetic materials under standard conditions.

The physical properties (e.g., enthalpy of formation, density, thermal stability, and oxygen balance) and detonation parameters of these azoles are based on the ring structure (Figure 1.9), which can be altered by substituting the hydrogen atoms with explosophoric functional groups (e.g., $-CN$, $-N_3$, $-NO_2$, $-NH_2$, and $-NH-NO_2$).[11] A series of nitropyrrole derivatives were designed by substituting the hydrogen atoms on the pyrrole ring with nitro groups.[28] The detonation velocity and pressure of the derivatives were calculated using the Kamlet-Jacobs equation, and some (e.g., trinitropyrrole, tetranitropyrrole, and pentanitropyrrole derivatives) have superior detonation performance in comparison to RDX. Notably, the position of the nitro substituents significantly affects the stability of the molecules due to steric hindrance, while the presence of the nitro functional group on the nitrogen decreases the overall stability of the molecule.

Due to the compactness, inherent stability, and modifiability of pyrazole, functionalisation of pyrazole is relatively easy to accomplish and research into pyrazole derivatives has led to the synthesis of nitrated pyrazole.[29] The nitro functionalisation of the pyrazole

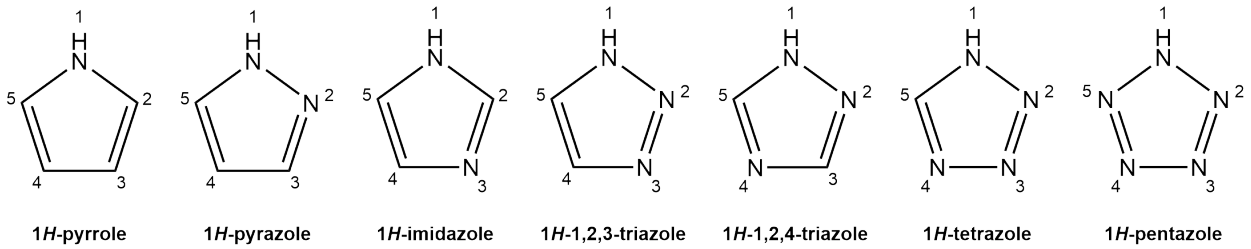


Fig. 1.9. Structures of 1H-pyrrole, 1H-pyrazole, 1H-imidazole, 1H-1,2,3-triazole, 1H-1,2,4-triazole, 1H-tetrazole, and 1H-pentazole.

ring increases both the density and nitrogen content, while improving the oxygen balance. Some of these compounds exhibit superior detonation properties across multiple properties, including density, enthalpy of formation, detonation pressure and velocity, and impact sensitivity in comparison to TNT. An example of the development and synthesis of imidazole-based energetic materials is the synthesis and characterisation of 1-methyl-2,4,5-trinitroimidazole (MTNI). The theoretical evaluation of the explosive performance of MTNI found that it was comparable to RDX.[30]

Triazoles are frequently studied for their high thermal stability and high enthalpy of formation. One example of such derivatives involves synthesising energetic polymers containing furazan, 1,2,3-triazole rings, and a nitramine group into a polymer backbone. The polymer was analysed using spectral and physico-chemical methods, showing superior performance in terms of thermal stability, heat of formation, and glass transition temperature when compared to the standard benchmark polymer, nitrocellulose (NC).[31] Another notable example is the design and synthesis of a novel family of nitrogen-rich salts that contain a (1,2,4-triazolyl) furoxan core. These salts exhibit superior performance in terms of experimental density, high enthalpy of formation, and detonation performance, surpassing that of conventional explosives such as TNT and PETN.[32]

Due to the planar structure of the tetrazole ring and its high nitrogen content, tetrazole compounds exhibit high density and are capable of releasing a significant amount of energy and gas upon decomposition or explosion, which accounts for the superior explosive properties of many tetrazole derivatives. The high enthalpy of formation in tetrazole

compounds is due to the nitrogen-nitrogen bonds, ring strain, and their density.[11] A number of promising tetrazole-based energetic derivatives have been identified, one of which includes a family of 5-hydrazino-1H-tetrazolium salts. Compared to azido groups, which increase the energy content but reduce the handling safety due to increased impact sensitivity, hydrazine groups increase the enthalpy of formation while also increasing the intra- and intermolecular hydrogen bonding, which increases the density and reduces the sensitivity to impact and friction. These salts, composed of various nitrogen-rich anions (e.g., NO_3) and 5-hydrazino-1H-tetrazolium cations, demonstrate notable energetic properties. Specifically, some of these compounds exhibit high densities, high enthalpies of formation, and detonation pressures. These properties contribute to their superior detonation parameters, making them superior to RDX and comparable to HMX.[33]

1.4 Predicting Enthalpy of Formation of Energetic Azoles

This section introduces the theoretical approach and group contribution methods available to estimate the enthalpy of formation of energetic salts. It explores the hierarchy of approximations within group contribution methods, discussing their underlying principles and gives examples to illustrate their application in predicting the enthalpy of formation of energetic salts.

1.4.1 Theoretical Approach

Physical properties can be estimated through theoretical, semi–empirical or empirical models.[34] Molecular modelling is part of theoretical estimation methods, while empirical approaches involve correlated functions representing a particular dataset. Semi–empirical methods use parameters, with values previously determined through regression, together with equations that represent the relationship between the structure and physical property of a molecule.[35] Computational methods can compute $\Delta_f H^\ominus$ within chemical accu-

racy, which has generally been defined as a 95 % confidence interval of 1 kcal mol^{-1} or $\pm$4 kJ mol^{-1}.[24] Weizmann-n family,[36–38] HEAT family,[39,40] and the ANL-n family[41] are examples of computational methods that belong to the high-level approaches and can compute $\Delta_f H^\circ$ within chemical accuracy. These families of methods are classified as "fixed recipe" methods, which means that each computational step is extensively documented. In comparison, other high–level approaches such as Feller-Petersen-Dixon[42] (FPD) and the focal point approach[43] are flexible methods, meaning they only present guidance for the computational steps that depend on the target species.[24] To a large degree, these high-level approaches compute thermochemical data from first principles and are not dependent on empirical and semi–empirical parameters. However, most variants do make use of experimental data, such as spectroscopic spin–orbit terms for atoms. These high–level approaches are composite methods, which combine the result of multiple computational steps to achieve an accurate estimation of the non-relativistic full configuration interaction electronic energy at a complete basis set (FCI/CB). Additionally further computations that account for significant energy contributions (e.g., scalar relativistic effects and non–Born–Oppenheimer corrections) and accurate vibrational zero–point energy (ZPE) and its anharmonic refinement are done, which are necessary terms to account for any contribution at the kJ mol^{-1} level to the atomization energy. In all the methods previously mentioned, the FCI/CBS energy uses the coupled clusters method. First, the complete basis set limit is determined using coupled clusters with single, double, and quasi–perturbative triple excitations (CCSD(T)/CBS). Then, computations for the post–CCSD(T) contributions, such as connected triple excitations and quasi–perturbative quadruple excitations.

Mid–level approaches achieve $\Delta_f H^\circ$ accuracies of $\pm$8–30kJ mol^{-1}. Considering the accuracy of the Weizmann-n family, HEAT family, and the ANL-n methods, the methods that were previously described as high–level approaches are now described as mid–level approaches.[24] The following mid–level approaches discussed are fixed-recipe family of composite approaches. The Gaussian-n family is a set of mid–level methods by Curtiss et al.[44]

The first generation, G1,[44] and its successor G2,[45] as well as the G2(MP2)[46] variant, are no longer widely used because of the more accurate and less computationally expensive G3[47] and G3(MP2)[48] methods. Rather than the fourth–order Møller–Plesset (MP4), the Gn(MP2) variants use second-order Møller–Plesset perturbation theory for the basis set extension corrections, which is less computationally expensive but also less accurate. The fourth and newest generation, G4,[49] and its G4(MP2)[50] variant are more accurate than any of the previous generations at a comparatively larger computational cost. In comparison to G1, G2, and G3, which used the Hartree-Fock geometry optimisation, variants of G3 and G4 uses the more reliable density–functional theory (DFT), more specifically the B3LYP functional for the initial geometry optimisation. The Gn approaches have an empirical component contained in the higher-level corrections (HLC). While the HLC used in G1 used two parameters, one for unpaired electrons that were derived from correcting the electronic energy of a hydrogen atom, and one for electron pairs, which was derived from correcting the dissociation energy error of H_2. In G2, the same two parameters were obtained by fitting to minimise the difference between computation and benchmark data. The number of empirical parameters increased to four in G3 and six in G4, which increased the accuracy but has lost the original physical meaning. The correlation-consistent composite approaches[51] (ccCA) are a related family of methods that were derived to provide an alternative to the Gn family that do not contain empirical parameters. The CBS family includes the popular mid-level CBS-QB3[52] and CBS-APNO[53] methods. The older CBS-APNO method is more computationally expensive and can only be used for molecules containing the first-row atoms to compute $\Delta_f H^\circ$ that are more accurate than G4. The CBS-QB3 includes spin–orbit corrections, uses the B3LYP-based geometry optimisation, and is less expensive at the cost of less accurate $\Delta_f H^\circ$. Despite the accuracy of these methods, the calculations require extensive computational resources and time.[24]

1.4.2 Group Contribution Methods

Semi–empirical methods are most effectively used for predictive modelling and screening a large amount of molecules. There are two types of semi–empirical methods that can be used for the prediction of azole properties: quantitative structure property relationship models and group contribution methods.[35] Quantitative structure-property relationships (QSPR) models use quantum calculations to determine the value of descriptors, which are generally chemical, physical, or physicochemical properties. A group contribution method (GCM) is built on the principle that the property of a molecule can be predicted using the sum of all the individual contributions associated with its smaller structural units, or groups, present in the structure of a molecule. The underlying assumption of a GCM is that the contribution of a group is the same in all compounds that contain that group. The physical principle stems from the observation that the forces between atoms, in the same or different molecules, are short-range and are only significant over distances greater than 1–3 Å.[54] A trivial example of this additive method is the relationship between the molecular weight of a molecule and the atomic weights of the constituent atoms in that molecule. Benson and Buss[55] proposed that the interaction between atoms or groups in a molecule becomes negligible as a consequence of increasing separation, justifying the additive nature of molecular properties. Quantitatively, if XNX and YNY are molecules that contain X and Y groups, respectively, N is a molecular framework, and Φ is the molecular property, then as the distance between X groups and Y groups significantly increases in their respective molecules, the effect of their interactions becomes negligible. In this limit, $\Delta\Phi$ for the disproportionation reaction,

$$\mathrm{XNX} + \mathrm{YNY} \rightleftharpoons 2\mathrm{XNY}, \tag{1.15}$$

is the contribution attributed to the change in symmetry or optical isomerism alone.[55] For a framework that is unsymmetrical, the two different substitution sites may be denoted N

and N′. Due to the constraints of the presence of symmetry in N, the disproportionation reaction for an unsymmetrical framework NN′ is written as,

$$\text{XNN'X} + \text{YNN'Y} \rightleftharpoons \text{XNN'Y} + \text{YNN'X}. \tag{1.16}$$

For these disproportionation reactions, $\Delta C_p \rightarrow 0$, $\Delta H \rightarrow 0$, and $\Delta S = R \ln K_\sigma$, where σ is the symmetry number including both internal and external symmetry and $K = \sigma_{\text{XNN'X}}\sigma_{\text{YNN'Y}}/\sigma_{\text{XNN'Y}}\sigma_{\text{YNN'X}}$.[55] The general formula to determine the value of a property, f, through a GCM is given by,

$$f = \sum n_i \Delta f_i + c, \tag{1.17}$$

where n_i is the number of additive contributions of type i, Δf_i is the i–th contribution, and c is a fitting constant.[56]

There have been various GCMs that have been devised to predict a variety of molecular properties including Joback and Reid,[57] Benson and Buss,[55] Constantinou and Gani,[58] Marrero and Gani,[59] Bruce et al.,[34] and Hukkerikar et al.[61] The advantages of using a GC method is the simplicity and ease of use, as properties are estimated by extracting group parameters, or contributions, from a reference table and applying them in a mathematical model.[62] GCMs are particularly beneficial when qualitatively accurate property estimates are required and/or quantitative accuracy is not necessary (e.g., during the product design or early stage of process synthesis). For example, if a GCM can correctly predict which solvent has preferable properties between two different solvents, the model can also be applied to a wide search space of solvents. Once a solvent is selected for additional model-based studies, the GCM can be required to provide a quantitatively accurate prediction within an acceptable margin while taking into consideration that the properties used during the synthesis may differ from the design verification step.

Although GCMs are straightforward and easy to use, their application is limited to compounds and/or chemical systems that can be represented by the corresponding group contribution parameters available.[63] The accuracy of previous GCMs for large multifunctional molecules, for example, increasing the number of aromatic rings or acid groups, has come into question. However, this limitation has been addressed by the addition of new groups and/or correction terms.[64] New groups and/or correction terms have also been introduced to account for the inability of GCMs to distinguish between isomers (e.g., cis and $trans$ isomers).[58] The addition of new groups and/or correction terms generally requires experimental data to define and estimate their contributions. Where experimental data is not available, virtual experimental data can be generated using computational methods to estimate the parameters. In the development of property estimation using GCMs, the hierarchy of approximations serves as a framework and the majority of that use experimental data to determine parameters.

1.4.2.1 Hierarchy of Approximations.

Benson and Buss[55] described a hierarchical system of additivity laws in which GCMs are classified as zero-order, first-order, and second order approximations.

Zero-order Approximation. The zero-order approximation, or the law of additivity of atom properties, estimates a molecular property as a sum of the atomic contributions. The disproportionation reaction that describes a zero-order approximation is given by,

$$X_2 + Y_2 \rightleftharpoons 2XY. \tag{1.18}$$

To illustrate the zero-order approximation, consider the carbon compounds of type CA_iB_{4-i}, $C_2A_jB_{6-j}$, and A_kB_{2-k}, where $i = 0, 1, \ldots, 4$, $j = 0, 1, \ldots, 6$, and $k = 0, 1, 2$. This amounts to a total of 18 compounds, with six of them being symmetrical, A_2, B_2, A_2B_6, C_2A_6, C_2B_6, CA_2B—CA_2B, and CA_2—CAB_2. For each pair of the six compounds, an

equation analogous to Equation 1.18 can be written, e.g.,

$$\mathrm{CA_3—CA_3 + Y—Y \rightleftharpoons 2CA_3—Y}, \tag{1.19}$$

which additively connects the properties of the three compounds that appear in the equation. Among the 18 compounds, there are 15 unique pairs of the six symmetrical compounds and 15 disproportionation equations. Therefore, by selecting any three independent compounds as standards, the properties of the remaining compounds can be additively determined. For example, by choosing A_2, B_2, and C_2A_6 as standards, then A, B, and C can be determined by,

$$\begin{aligned} \mathrm{A} &= \frac{1}{2}\,(\mathrm{A_2}) \\ \mathrm{B} &= \frac{1}{2}\,(\mathrm{B_2}) \\ \mathrm{C} &= \frac{1}{2}\,(\mathrm{C_2A_6}) - 3(\mathrm{A}). \end{aligned} \tag{1.20}$$

First-order Approximation. The next order of approximation is first-order, which is the law of additivity of bond properties. In the disproportionation reaction shown in Equation 1.15, the molecular framework N is a single atom or partially substituted atom (e.g., NH). To illustrate the first-order approximation, consider the five carbon compounds, CA_4, CAB_3, CA_2B_2, CA_3B, and CB_4, where three of these species are symmetrical. By choosing two of the three symmetrical species as standards, the properties of the five compounds can be determined additively. For example, if CA_4 and CB_4 are chosen as standards then,

$$\begin{aligned} \mathrm{C—A} &= \frac{1}{4}(\mathrm{CA_4}) \\ \mathrm{C—B} &= \frac{1}{4}(\mathrm{CA_4}). \end{aligned} \tag{1.21}$$

The first successful GCMs was developed in 1955 by Lydersen[65] who estimated the critical temperature, pressure, and volume of organic compounds through a first-order approximation. After an increase in experimental values since, first-order groups have been

improved. The most notable was by Joback and Reid[57] (JR) who reevaluated Lydersen's GCM to estimate a further eight properties of organic compounds using 41 groups and reparameterised the group contributions using linear regression. The group contributions were optimised by minimising the sum of the absolute errors between the predicted and experimental values for each property. This approach was favoured over minimising the sum of squares of the errors since outliers were too heavily weighted. However, while this gave a better estimation for most compounds, there were larger errors for these outliers. The JR method assumes no interaction between groups and the structurally-dependent parameters are calculated by summing the frequency of each group in a molecule multiplied by the groups contribution to the property. To illustrate the JR method, $\Delta_{\mathrm{f}}H^{\ominus}_{298.15K}$ will be estimated for 2-ethylphenol. As seen in Figure 1.10, 2-ethylphenol contains four unsubstituted benzylic carbons, represented by =CH, two substituted benzylic carbons, =C, and has been functionalized with an alcohol, −OH(phenol), a terminal methyl $-CH_3$, and a secondary methylene $-CH_2-$, summarised in Table 1.2. The formula for estimating $\Delta_{\mathrm{f}}H^{\ominus}_{298.15K}$ is given by,

$$\Delta_{\mathrm{f}}H^{\ominus}_{298.15K} = 68.29 + \sum_i N_k \Delta H_{fk}. \tag{1.22}$$

Consequently, by substituting the sum of the groups into Equation 1.22, the predicted value of $\Delta_{\mathrm{f}}H^{\ominus}_{298.15K}$ for 2-ethylphenol is $-149.23\,\mathrm{kJ\,mol^{-1}}$, which compares well with the experimental value of $-145.23\,\mathrm{kJ\,mol^{-1}}$. The law of bond additivity performs well for most

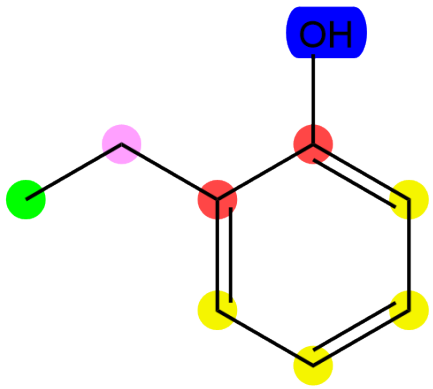


Fig. 1.10. Structure of 2-Ethylphenol.

Table 1.2. Functional groups (k) present in 2-ethylphenol and their contribution to the enthalpy of formation ($N_k \Delta H_{fk}$).[a]

Group k	N_k	$N_k \Delta H_{fk}$
$-CH_3$	1	-76.45
$-CH_2-$	1	-20.64
$=CH$	4	8.36
$=C$	2	92.86
$-OH$(phenol)	1	-221.65
$\sum_{k=1}^{5} N_k \Delta hfk$		-217.52

[a] Contribution to enthalpy of formation in $kJ\,mol^{-1}$. Values based on Joback-Reid group contribution method.[57]

molecules except heavily branched compounds and fails to distinguish between isomers (e.g., n-butene and isobutene).

Second-order Approximation. The second-order approximation attempts to resolve some of the shortcomings of the law of bond additivity by treating the molecular property as a sum of contributions from constituent groups. Using Equation 1.15, the rule of additivity for group properties assumes that the molecular framework, N, can be two atoms or structural elements. An example of a disproportionation reaction that illustrates second-order additivity is,

$$F{-}CH_2{—}CH_2{-}F + H{-}CH_2{—}CH_2{-}H \rightleftharpoons 2H{-}CH_2{—}CH_2{-}F, \tag{1.23}$$

where it should be noted that the fluorine atoms are bound to carbon atoms, which have two hydrogen atoms and one carbon atom as neighbouring ligands, and the disproportionation reaction preserves this relationship. Therefore, the substitution of groups happens on the adjacent atoms instead of the same atom, like the first-order approximation, and the disproportionation reaction keeps the nearest neighbouring atoms or groups constant. According to Benson and Buss,[55] a second-order group is interpreted as a polyvalent atom along with its connectivities written as $X-(A)_m(B)_n(C)_o(D)_p$, where X is the central atom

that is attached to m atoms of type A, n atoms of type B, o atoms of type C, and p atoms of type D. With this definition of a group, the rule of group additivity is equivalent to the second-order approximation. To illustrate the second-order approximation, consider the compounds $C_2A_iB_{6-i}$ where i =0,1,...,6 to produce a total of ten compounds, for which there are seven unique formulas and three isomers. The disproportionation reactions can be composed for six unique frameworks, namely, the symmetrical nuclei $-(C_2A_4)-$, $-(CAB-CAB)-$, $-(C_2B_4)-$ and the unsymmetrical nuclei $-(CA_2—CB_2)-$, $-(CA_2—CAB)-$, $-(CB_2—CAB)-$. There are six disproportionation reactions that can be constructed with these symmetrical and unsymmetrical nuclei. For example, the disproportionation reaction for $-(C_2A_4)-$ is,

$$A(C_2A_4)A + B(C_2A_4)B \rightleftharpoons 2A(C_2A_4)B. \tag{1.24}$$

Therefore, there are six additive equations that govern the properties of the ten compounds and as a result four independent groups, C—(C)$(A)_3$, C—(C)$(B)_3$, C—(C)(A)$(B)_2$, and C—(C)$(A)_2$(B). By knowing the molecular properties of four compounds and given that at least one of the independent groups is contained within those compounds, then the properties can be predicted using the disproportionation reactions. Alternatively, the faster and more direct method of group additivity, where the property is estimated using the sum of groups in the molecule. For example, C_2A_6 can be written as,

$$(C_2A_6) = 2[C—(C)(A)_3]. \tag{1.25}$$

This rule of group additivity for molecular properties is limited to compounds that contain at least two polyvalent atoms, meaning that they contain a minimum of two groups.

The GCM proposed by Benson and Buss[55] in 1958 became the foundation for many subsequent advances in second-order additivity schemes, and estimated the heat capacity, enthalpy of formation, and entropy of hydrocarbons and their derivatives. Benson and Buss started with saturated hydrocarbons and linear chain hydrocarbons consist of

only the terminal methyl group, C—$(H)_3(C)$, and the secondary methylene group, C—$(H)_2(C)_2$.[55] To accommodate branched aliphatic compounds, only two additional groups are required: the tertiary group, C—$(H)(C)_3$, and the quaternary group, C—$(C)_4$. By assigning values to these four groups, it is possible to estimate the properties of all saturated hydrocarbon molecules. The expansion of the GCM to include noncyclic, olefinic compounds requires GAVs for the following newly defined groups; C_d—$(C_d)(H)_2$, C_d—$(C_d)(C)(H)$, C_d—$(C_d)(C)_2$, C—$(C_d)(H)_3$, C—$(C_d)(C)(H)_2$, C—$(C_d)(C)_2(H)$, C—$(C_d)(C)_3$. The derivatives included alkynes, polysubstituted benzene, where C_B represents a carbon atom in a benzene ring, oxygen-containing compounds (e.g., alcohols and glycols), organosulfur compounds, and halogen-containing compounds, among others. The GCM is obeyed to within $\pm 2.50\,\mathrm{kJ\,mol^{-1}}$ with some errors reaching $12.55\,\mathrm{kJ\,mol^{-1}}$.

The GCM is capable of differentiating isomers when the isomerism arises from variations in the groups present within molecules. For example, 3–ethylhexane (C_8H_{18}) consists of three C$-(H)_3(C)$ groups, four C$-(H)_2(C)_2$ groups, and one C$-(H)(C)_3$ group. In contrast, the 2,2–dimethylhexane (C_8H_{18}), consists of four C$-(H)_3(C)$ groups, three C$-(H)_2(C)_2$ groups, and one C$-(C)_4$ group. However, 2–methylheptane, 3–methylheptane, and 4–methylheptane are positional isomers that all have three C$-(C)(H)_3$ groups, four C$-(C)_2(H)_2$ groups, and one C$-(C)_3(H)$ group, which prevents differentiation among all octane isomers based solely on Benson's groups. Addressing this limitation can be achieved by including next-nearest neighbour interactions, which Benson and Buss has described as third-order approximations.[55] In order to distinguish between cis-trans isomers, the GAVs are the average of the isomeric values and a correction term is added. Another interaction between next-nearest neighbours involves a correction for gauche interactions for atoms bigger than hydrogen. This is obtained by counting the number of gauche configurations of each group and applying a correction to each of these configurations. This is relevant for molecules such as linear paraffins, where the most stable conformation occurs when the heavy groups are trans to each other.[54] Lastly, a correction

that accounts for each pair of substituted groups on benzene that are ortho to each other is included.

Benson et al.[66] extended the original GCM to include cyclic compounds and revised previously reported GAVs due to an increase in experimental data. The approach to cyclic compounds is to use GAVs derived from open-chain compounds and a ring strain correction (RSC). The ring strain correction used to estimate the enthalpy of formation is called the strain energy, E_{RS}. The GCM estimates unbranched hydrocarbons and their derivatives with relatively low errors and can be assigned as "unstrained" standards. Therefore, for each ring, the E_{RS} is determined by subtracting the unstrained GAVs from the observed enthalpy of formation of the unsubstituted ring. Data for the RSC of saturated nitrogen-containing includes only ethyleneimine, azetidine, pyrrolidine, piperidine, succininimide, and pyrazine. Benson et al.[66] noted that for unsaturated, nitrogen containing heterocyclic compounds, a RSC is not easily attainable since these compounds also display resonance. Building on previous work, Benson[54] then compiled all previous GAVs and further expanded to include free radicals and polycyclic structures.

In 1988, Domalski and Hearing extended Benson's GCM to predict the enthalpy of formation, heat capacity, and entropy of hydrocarbons in both liquid and solid phases at 298.15 K.[67] The development of groups and GAVs started with alkanes, then alkenes, alkynes, aromatic hydrocarbons, cycloalkanes, and other derivatives. Additionally, rather than a gauche correction, the repulsive interaction of hydrogen atoms in methyl groups for tertiary and quaternary carbons was corrected. The $\Delta_{\mathrm{f}}H^{\circ}(\mathrm{g})$ was reevaluated with estimation accuracies comparable to the GAVs determined by Benson.[54] From the dataset, $\Delta_{\mathrm{f}}H^{\circ}$ for 559 hydrocarbon molecules in the gas, liquid, and solid phases are compared and have a mean absolute error of 2.6 kJ mol^{-1}. In further work by Domalski and Hearing[68], the GCM was extended to include compounds that contain carbon, hydrogen, oxygen, nitrogen, sulfur, and halogens across the gas, liquid, and solid phases. Apart from the enthalpy of formation, heat capacity, and entropy, the entropy of formation, Gibbs free

energy of formation, and the natural logarithm of the equilibrium constant of formation were also calculated as auxiliary properties. A notable advancement was the extension of the GCM to estimate the enthalpy of formation of pyridine and substituted pyridine molecules, which was subsequently extended to include five-membered aromatic heterocycles (e.g., furan, pyrrole, and thiophene). Benson et al.[66] originally divided benzene by six to determine C_B—(H)$(C_B)_2$, where each group also contained a sixth of the resonance energy of benzene. Domalski and Hearing[68] introduced the N_I—(C_B) group using pyridine, which also accounts for the conjugation energy inherent to pyridine. In the extension to five-membered aromatic heterocycles, the carbon atoms were treated as C_B—(H)$(C_B)_2$ groups. Their work included 1512 organic molecules and over 3700 comparisons between literature and estimated values were conducted. Of these comparisons, the estimation of $\Delta_f H^\circ$ showed that 67 % of the residuals were less than $\pm 4\,\mathrm{kJ\,mol^{-1}}$, 16 % were between $\pm 4\,\mathrm{kJ\,mol^{-1}}$ and $\pm 8\,\mathrm{kJ\,mol^{-1}}$, and 17 % were greater than $\pm 8\,\mathrm{kJ\,mol^{-1}}$.

Under contract with the National Institute of Standards and Technology (NIST), Cohen[69] started a program to evaluate and refine group additivity values required for predicting thermochemical properties of CHO-containing compounds in the gas, liquid, and solid phase.[69] This program focused on updating these values within the framework of the Benson group contribution method to establish a uniformly derived set of GAV estimations, enhancing the accuracy and reliability of the prediction of thermochemical properties. In addition to revised GAVs, the report includes ring strain corrections and contributions from non-nearest neighbour interactions, extrapolated from the database. The GAVs presented in this report are derived using only experimental data and exclude all computationally predicted values. Additionally, the reliability of the experimental data could not be independently verified, they were still accepted to determine GAVs, unless errors were readily apparent. In the gas phase, the average error for 1028 compounds was approximately $5.5\,\mathrm{kJ\,mol^{-1}}$, the liquid phase had an average error of $5.6\,\mathrm{kJ\,mol^{-1}}$ for 941 compounds, and the solid phase had an average error of $9.1\,\mathrm{kJ\,mol^{-1}}$ for 538 compounds.

After earlier work on hydrocarbons by Sumathi and Green,[70] Ashcraft and Green worked to address the limited amount of GAVs available to estimate nitrogen-containing compounds by deriving 49 atom-centered GAVs to predict the enthalpy of formation, entropy, and heat capacity using CBS-QB3 QM calculations of 105 noncyclic CHNO-containing molecules.[71] The GAVs were determined using multiple linear regression analysis and included nitro, nitroso, nitrite, nitrate, amine, imino, and azo groups. In the fitting, it was noted that 17 groups showed inherent linear dependencies, as each group is consistently adjacent to another. The solution to resolve the dependency is to arbitrarily divide the thermochemical contribution between these two interdependent groups, thus preserving the overall additive nature of the structure. The average signed error for the enthalpy of formation is $0.08\,\mathrm{kJ\,mol^{-1}}$ and the mean absolute deviation is $10.25\,\mathrm{kJ\,mol^{-1}}$. The predictions for the gas-phase enthalpy of formation were compared to experimental values. The main challenge in these comparisons was the missing group additivity values of specific groups adjacent to aromatic rings necessary to build the target molecules that were not parametrised in this work, leading to the use of substitute groups. This substitution and the lack of steric effects not reflected in the group values likely account for much of the errors.

Holmes and Aubry[72] revised both prior determinations and established new group additivity values from Benson et al.[66] and Cohen[69] to estimate the enthalpy of formation of molecules containing C,H, and O atoms, with an extended analysis on C, H, N, O, S atoms and halogens.[73] The first publication provides updates for factors such as cis-isomer effects, gauche interactions, double-bond positioning, ring strain, conjugation, and steric hindrance in aromatic compounds. Both articles demonstrate the thermodynamic consequence of substituting one functional group for another. For example, replacing an exocyclic methylene group with oxygen in a cyclic alkane typically has a negligible impact on stability. However, in specific cases such as cyclopentadienyl and furan, the stability increases from the quasiaromatic character of the rings, assisted by oxygen's lone-pair

of electrons. This revised GCM yields enthalpy of formation values with an accuracy of $\pm 4\,\mathrm{kJ\,mol^{-1}}$.

Sabbe et al.[74] reported a comprehensive, internally consistent set of 95 Benson GAVs for determining the standard enthalpy of formation of hydrocarbons and hydrocarbon radicals at 298.15 K and 1 bar. These values were derived from a database of 223 *ab initio* standard enthalpies of formation, calculated using the CBS-QB3 level of theory. Among the 95 GAVs, 16 apply to hydrocarbons, 25 to hydrocarbon radical groups, and multiple new corrections to account for ring strain. Additionally, previously published non-next-nearest neighbour interactions were assessed, enhancing the method's prediction and consistency and a novel approach to account for non-next-nearest neighbour interactions in radicals was proposed. Hydrogen bond increments (HBI) were determined to calculate radical standard enthalpies of formation, with particular emphasis on resonance stabilized radicals. The HBI method calculates an increment that represents the enthalpy change when a radical forms from its parent molecule through the loss of a hydrogen atom, using the known enthalpy of formation of its parent molecule. This increment is added to the enthalpy of formation of the parent to estimate enthalpy of formation of the radical.[75] Although the HBI method potentially better accounts for resonance effects beyond the group region, it faces challenges in choosing the appropriate increment group, as mentioned by Sabbe et al..[74] Since the authors indicated no obvious relationship between ring strain corrections and the structural features of each ring, a ring strain correction was necessary for each specific ring, factoring in the ring size and the number of endocyclic double bonds it contains. The approach achieved an accuracy within $2\,\mathrm{kJ\,mol^{-1}}$ by including a bond additive correction and using the updated parameters. A set of 60 Benson GAVs was reported by Paraskevas et al.[76] for oxygenate molecules and 97 GAVs were reported for oxygenate radicals to predict the enthalpy of formation, entropy, and heat capacities. Nearly half of the GAVs for oxygenated molecules and the majority of oxygenate radical GAVs were newly derived in this work. The GAVs were determined from a set of 202 oxygenate molecules

and 248 radicals calculated using *ab initio* methods at the CBS-QB3 level of theory. A set of 12 molecules and 11 radicals, alongside experimental data, was used to validate the GAVs determined. A comparison between the *ab initio* values and the group additive predicted values shows that the method predicted the enthalpy of formation within 5 kJ mol^{-1}. Ince et al.[77] published a set of seven Benson GAVs alongside 15 correction terms for non-next-nearest neighbour interactions. This set was used to predict gas-phase enthalpies of formation, heat capacities, and entropies for monocyclic aromatic compounds, which contain substituents methyl, ethyl, vinyl, formyl, hydroxyl, and methoxy groups. The GAVs were derived using least-squares regression on a dataset of 143 molecules, which were computed with the post-Hartree-Fock G4 composite method. Of the 15 non-nearest neighbour interactions incorporated, 13 represent new substituent effects that are specific to aromatic compounds. The GCM developed to predict the enthalpies of formation are in strong agreement with experimental values with an error of less than 4 kJ mol^{-1} for all but two molecules. This method was expanded to include nitrogen-containing compounds by Pappijn et al.[78] A comprehensive set of 91 GAVs and three non-nearest neighbour interactions was derived from CBS-QB3 calculations on 300 molecules, including 104 radicals. The resulting GCM predicts the standard enthalpy of formation, entropy at 298.15 K, and the heat capacities over 300 K–1500 K. A training set of 274 species was used to determine GAVs, while the test set included 26 species to assess the accuracy of the GAVs. Notably, a general rule for the Benson GCM in radicals was applied, opting to use the canonical structure with the lowest standard enthalpy of formation in the linear regression. The GCM was validated by compiling a test set of 11 nitrogen-containing compounds, which had a mean absolute deviation (MAD) between the experimental and GCM estimated enthalpy of formation of 2.8 kJ mol^{-1}.

Bjorkman et al.[75] used G4 quantum mechanical calculations and isodesmic reactions, which preserve the number of each bond type (e.g., single, double, and triple C−C bonds) in reactants and products, to calculate the enthalpies of formation of 57 cyclic and acyclic

carbenium ions, including allylic carbenium ions. With the enthalpy of formation values, the Benson-type GAVs were derived through multiple linear regression. The GCM compares well with both experimental and quantum mechanical data, as well as expands the scope of species that can be described using a GCM. The deviations from experimental values for this GCM give a mean absolute deviation of 11.12 kJ mol^{-1}, while the root mean square error is 12.38 kJ mol^{-1}. This GCM was further extended to analyse oxygenates, oxonium ions, and oxygen-containing carbenium ions.[79] This involved regression of 71 GAVs, 65 of which were not previously reported and six newly re-estimated based on the existing dataset of 65 groups. To calculate the enthalpies of formation used in the multiple linear regression, isodesmic reactions for 195 species were constructed using reference molecules. The result indicated an average deviation of 8.16 kJ mol^{-1} between the GCM estimated values and the G4 computed values.

In pursuit of greater accuracy, Constantinou and Gani developed a new GCM that significantly improved the prediction for properties of pure organic compounds.[58] This new method proposes that the molecular structure is a collection of two types of groups: first- and second-order groups. The first-order groups reported by Joback and Reid are not sufficient to estimate the properties of mixtures; therefore, a set of groups commonly used to estimate mixture properties were used. The second-order groups account for proximity effects that are only partially described by first-order groups and enable the GCM to differentiate between isomers. For example, $\Delta_{\mathrm{f}}H^{\circ}$(g) can be estimated using the model given by,

$$\Delta_{\mathrm{f}}H^{\circ}(\mathrm{g}) = 10.835 + \sum_{k} N_k(hf1k) + W\sum_{j} M_j(hf2j), \tag{1.26}$$

where N_k is the number of first-order groups of type k, $gf1k$ is the contribution of the first-order group labelled $1k$, N_j is the number of second-order groups of type j, $hf2j$ is the contribution of the second-order group labelled $2j$, and $W = 0$ for first-order calculations or $W = 1$ for second order calculation. Using the groups developed by Constantinou and Gani, 2-ethylphenol consists of one CH_3, four ACH, one $ACCH_2$, and one ACOH first-order

groups. There are no second order groups present; therefore, $W = 0$ and Equation 1.26 can be simplified to,

$$\Delta_{\mathrm{f}} H^{\ominus}(\mathrm{g}) = 10.835 + \sum_{k} N_k (hf1k). \tag{1.27}$$

While the methods above have benefits, their scope remains limited. Existing approaches may be insufficient predict properties of large, complex, and multifunctional compounds. These methods mainly rely on small datasets of simpler compounds, which can limit their accuracy when estimating large, polycyclic or multifunctional molecules. Moreover, most GCMs lack the necessary groups to represent complex molecules. In response to these limitations, Marrero and Gani[80] constructed a more robust GCM, increasing the predictive accuracy giving reliable estimations across a broader range of complex chemical molecules. The property estimation model to predict $f(X)$, a simple function of property X,

$$f(X) = \sum_{i} N_i C_i + w \sum_{j} M_j D_j + z \sum_{k} O_k E_k, \tag{1.28}$$

where C_i represents the contribution from the first-order group of type i occurring N_i times, D_j is the contribution from the second-order group of type j with M_j occurrences, and E_k represents the third-order group fo type k appearing O_k times in a molecule. The constants w and z are set to zero using only first-order groups in the first level of estimation. For the second level, $w = 1$ and $z = 0$, while in the third level of estimation, both w and z are set to one. This GCM estimates various properties including normal boiling and melting points, pressure, volume, critical temperature, standard enthalpy of formation, vaporisation, fusion and Gibbs energy. The GAVs were developed through regression analysis on a dataset containing over 2000 compounds, which also includes complex polycyclic molecules.

Meier[81] developed a GCM for estimating the enthalpy of formation of organic molecules achieving a maximum error of 4.20 kJ mol^{-1} from experimental values while ensuring a minimal amount of GAVs. Reducing parameters is important for GCMs as ex-

cessive complexity can reduce model efficiency. The group parameters were systematically derived, starting with alkanes alone, and then adding one additional group type for each subsequent molecule class. The model equation used to predict n-alkanes is,

$$\Delta_{\mathrm{f}}H^{\ominus} = 2\text{*}\mathrm{GC}_{\mathrm{CH_3}} + \mathrm{N}_{\mathrm{CH_2}}\text{*}\mathrm{GC}_{\mathrm{CH_2}}, \tag{1.29}$$

where $\mathrm{GC}_{\mathrm{CH_3}}$ is the CH_3 GCM parameter, $\mathrm{GC}_{\mathrm{CH_2}}$ is the CH_2 GCM parameter, and $\mathrm{N}_{\mathrm{CH_2}}$ is the number of times CH_2 is present in the molecule. The GCM was extended to mono-methylalkanes by introducing a CH_2 group, in the equation,

$$\Delta_{\mathrm{f}}H^{\ominus} = 3\text{*}\mathrm{GC}_{\mathrm{CH_3}} + \mathrm{N}_{\mathrm{CH_2}}\text{*}\mathrm{GC}_{\mathrm{CH_2}} + \mathrm{GC}_{\mathrm{CH}}. \tag{1.30}$$

This GCM was further expanded to include oxygenated compounds, alkenes, alkynes, nitrogen-containing compounds, benzene derivatives, and naphthalene structures. Subsequent publications continued to broaden the GCM application across a wide range of molecular classes.[82–85]

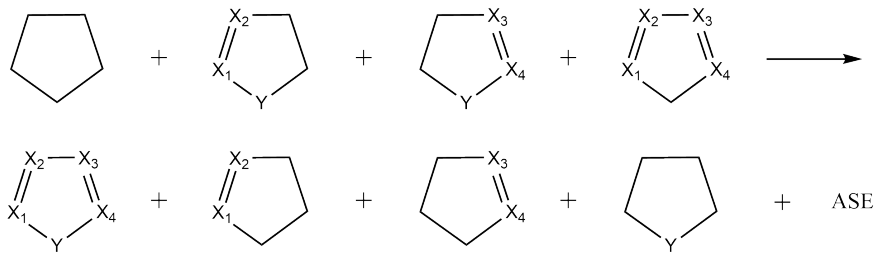
1.5 Aromatic Stabilisation Energy

As discussed previously, each group derived from an aromatic molecule retains a component of the aromaticity of the molecule from which it was derived, rather than treating aromaticity as a correction in GCMs. Additionally, ring strain was quantified by subtracting the sum of the "unstrained" groups from the enthalpy of formation of the strained molecule.

Aromaticity cannot be directly measured or computed. The concept of aromaticity is generally assessed using geometric, energetic, and magnetic criteria. This estimation often involves comparisons with non-aromatic reference compounds or relies on the non-additive nature of data derived from such reference species that are deemed appropriate.

Aromaticity was initially quantitatively described by resonance energy (RE) and accounted for the increased stability of an aromatic compound in comparison to its virtual olefinic counterpart.[86] For example, three single and three double bonds are the virtual olefinic equivalent of benzene. Following RE, aromatic stabilisation energy (ASE) more accurately described aromaticity through the use of homodesmotic reactions.[87] Homodesmotic reactions are reactions that meet the criteria of an equal number of carbon atoms in their various states of hybridisation in reactions and products and a matching of carbon-hydrogen bonds in terms of the number of hydrogen atoms joined to individual carbon atoms in reactants and products.[88] A drawback of homodesmotic reactions is ASE can depend significantly on using appropriate reference structures, which is observed less in hydrocarbons but more in heterocyclic systems.[87]

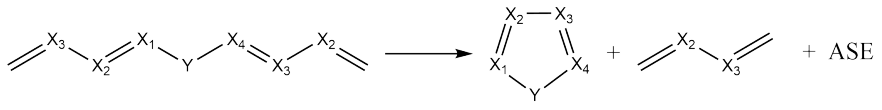
Cyrañski et al.[89] reported a homodesmotic reaction,



(1.31)

which was developed by modifying an isodesmotic reaction previously reported by Schleyer et al.[90] This was used to quantify the ASE, RE, magnetic susceptibility exaltation, nucleus independent chemical shift (NICS), and harmonic oscillator model of aromaticity for 75 five-membered heterocyclic aromatic systems and 30 endo-monosubstituted compounds. The perturbation caused by additional effects, such as charge stabilisation, interactions between heteroatoms and strain, were diminished due to the choice of reference structures used, which were all five-membered heterocycles. This homodesmotic reaction was further explored by Cyrański et al.[91] The additional homodesmotic reaction based on

acyclic analogues, as shown,



(1.32)

was explored, as well as four isodesmic reactions. The acyclic reference compounds gave less accurate ASE results than its cyclic analogue, largely due to the effects of ring strain. Therefore, while Equation 1.31 can be used to quantify ASE alone, Equation 1.32 can be used to quantify RS and ASE.

1.6 Aims and Objectives

This project aims to extend Benson's group contribution method to accurately predict the gas phase enthalpy of formation of energetic azoles. **This includes the following objectives:**

1. Provide a quantitative assessment of the uncertainty and bias in the G4(MP2)-6X composite quantum mechanical method for the calculation of the gas-phase enthalpies of formation.

2. Parameterise group additive values (GAVs) for groups in unsubstituted azoles using acyclic molecules, ensuring consistency with the group definitions employed by Benson in the original group additivity scheme.

3. Apply the groups developed for acyclic molecules to calculate a ring strain and aromatic stabilisation energy correction for each azole framework.

4. Model each azole framework using groups with ring strain and aromatic stabilisation energy explicitly incorporated, allowing comparison with the previous model.

5. Extend both group contribution methods to azoles functionalised with explosophoric groups and assess the accuracy and limitations of both approaches.

To address these objectives, the remainder of the thesis is structured as follows. Chapter 2 outlines the procedure for quantifying the uncertainty and bias associated with gas-phase enthalpies of formation for CHN-containing molecules computed using the G4(MP2)-6X composite method. In Chapter 3, two approaches for extending Benson's group contribution method are examined. The first approach parameterises group additivity values (GAVs) using acyclic molecules, together with ring strain and aromatic stabilisation energy corrections, to estimate the enthalpy of formation of azoles. The alternative approach is subsequently outlined, in which aromatic GAVs are regressed directly from the ten unfunctionalised azoles. This is followed by the regression of azoles functionalised with N-methyl, N-amine, N-azide, N-nitro, and N-nitraminde groups, with the fitting performed individually and then collectively for comparison. Chapter 4 provides the theoretical foundation of thermochemistry, outlining the principles underlying the enthalpy of formation. It then summarises methods for calculating thermochemical quantities, discusses the role of the Active Thermochemical Tables in data validation, and introduces linear regression as a tool for estimating enthalpies of formation. Lastly, Chapter 5 reviews the electronic structure methods used in this work, including Hartree-Fock theory, Density Functional Theory, Møller-Plesset perturbation theory, and Coupled-Cluster theory, with particular emphasis on the G4(MP2)-6X composite method used throughout this project.

2 Quantifying the Uncertainty in G4(MP2)-6X

The initial development of group contribution methods by Benson and Buss[55] and the subsequent determination of GAVs relied on experimental target data. Consequently, the extent of the fitting set would be limited by the availability of data. However, the development of highly accurate computational chemistry methods that are comparable in accuracy to experimental methods (e.g., *ab initio* composite methods), has enabled the practice of developing GAVs using computational data only.[74–79] This practice also allows for broader coverage of target molecules.

When reporting experimental results, an error is routinely reported, although the same practice is not often followed for computationally determined results. To compare the value of any computational measurement result, whether it is against reference values or between other measurements, a quantitative expression of the associated uncertainty of the result should be given.[92]

This chapter presents the procedure for the quantitative measure of the uncertainty associated with the enthalpy of formation for quantum chemistry calculations based on the notation used and method developed by Irikura et al.[92] Then, the uncertainty associated with the enthalpy of formation for CHN-containing molecules using the G4(MP2)-6X method is quantified.

2.1 Uncertainty associated with virtual measurements

The International Organisation of Standardisation (ISO) Guide to the Expression of Uncertainty in Measurement (GUM) outlines general rules to evaluate and express the uncertainty of physical measurements.[93] Physical measurements arise from experimental methods, while virtual measurements refer to the result of *ab initio* calculations. A virtual measurement of a molecular property consists of a measurand and its associated uncertainty, which predominantly emerges from the systematic error present, often referred to

as bias. Based on the procedure published in GUM, an approach for quantifying the uncertainty of a measurement determined from a computational quantum chemistry model was developed by Irikura et al.[92] from the National Institute of Standards and Technology (NIST).[94] This method determines a corrected virtual measurement, which includes a correction for bias. Given that the bias is unknown, it introduces uncertainty that is quantified and incorporated into the uncertainty associated with the corrected measurement.

Following the notation applied by Irikura et al., suppose that $x_{(t,b)}$ is a virtual measurement of *Y*, where *t* and *b* represent the formal theory (e.g., HF, MP2, CCSD(T), or a DFT method) and basis set, respectively. The mean of a sampling distribution over multiple repetitions of $x_{(t,b)}$ is denoted by $X_{(t,b)}$ and the associated additive bias $B_{(t,b)}$ is determined by,

$$B_{(t,b)} = X_{(t,b)} - Y. \tag{2.1}$$

Notably, when the same formal theory, basis set, and computational chemistry software package (e.g., Gaussian) are used, with consistent selection of settings such as convergence criteria, then $x_{(t,b)}$ is always equivalent to $X_{(t,b)}$. Following Equation 2.1, the measurement equation that includes a correction term to counter the bias associated with $x_{(t,b)}$ is,

$$Y = X_{(t,b)} + C_{(t,b)}, \tag{2.2}$$

where $C_{(t,b)}$ is the negative of the bias. Therefore, $x_{(t,b)}$ can now be referred to as an uncorrected virtual measurement for *Y* that is subsequently corrected by $c_{(t,b)}$, which is an expected value for the variable $C_{(t,b)}$, yielding a corrected virtual measurement *y* for *Y*,

$$y = x_{(t,b)} + c_{(t,b)}. \tag{2.3}$$

Since the uncorrected virtual measurement ($X_{(t,b)}$) and the correction ($C_{(t,b)}$) are independent, the covariance between these variables is zero and the standard uncertainty associ-

ated with *y*, denoted by $u(y)$, is given by,

$$u(y) = \left[u^2(x_{(t,b)}) + u^2(c_{(t,b)})\right]^{1/2}, \tag{2.4}$$

where $u(x_{(t,b)})$ and $u(c_{(t,b)})$ is the standard uncertainty associated with the uncorrected virtual measurement $x_{(t,b)}$ and $c_{(t,b)}$, respectively.

2.1.1 Uncertainty associated with bias

Quantifying an estimated correction for bias in *ab initio* calculations requires a highly accurate benchmark set with errors consistent with the target molecules. For example, since the target molecules are nitrogen-containing heterocycles (the azoles), a list of nitrogen-containing organic molecules with biases expected to be similar to that of methanimine is compiled to quantify the correction. The high-quality measurements for the standard enthalpy of formation ($\Delta_f H^\ominus$) and their associated uncertainties of the benchmark set were therefore collected from Active Thermochemical Tables, or ATcT (Section 4.3). Note that in the following, the term "physical measurement" refers to any high quality determination, whether experimentally measured or by using highly accurate computational methods. Virtual measurement refers to the computational method of which the uncertainty is being quantified.

Suppose that $Y_1,\ldots,Y_m$ are the values of $\Delta_f H^\ominus$ for a class of *m* molecules in the ATcT that have biases consistent with the target molecule, which have corresponding virtual measurements (x_i) with standard uncertainties ($u(x_i)$) and physical measurements (r_i) with standard uncertainties ($u(r_i)$), for $i = 1,\ldots,m$. The bias in x_i is determined by taking the difference between the virtual and physical measurements ($x_i - r_i$), and the negative of the bias can be taken as the correction for each molecule in the class, c_i, given by,

$$c_i = r_i - x_i. \tag{2.5}$$

The corrections are expected to have an approximately normal distribution since they are treated as a set of randomly distributed values that cluster around the arithmetic mean (μ). The normality and spread of the corrections are evaluated by examining the standard deviation (σ),

$$\sigma = \left[\sum (c_i - \mu)^2/m\right]^{1/2}, \tag{2.6}$$

and the coefficient of skewness (η_3),

$$\eta_3 = \left[\sum (c_i - \mu)^3/m\right]/\sigma^3, \tag{2.7}$$

where η_3 is zero in a normal distribution. The covariance between the physical measurement (R_i) and the virtual measurement (X_i) is zero since R_i and X_i are independently determined; therefore, the standard deviation of c_i, denoted by $u(c_i)$, is given by,

$$u(c_i) = \left[u^2(r_i) + u^2(x_i)\right]^{1/2}. \tag{2.8}$$

As mentioned earlier, the uncertainty associated with the *ab initio* calculations ($u(x_i)$) is negligible relative to the uncertainty of the physical measurement ($u(r_i)$) and Equation 2.8 can be approximated as $u(c_i) \approx u(r_i)$. The correction ($c_{(t,b)}$) and uncertainty ($u(c_{(t,b)})$) are specified as,

$$c_{(t,b)} = \mu = \frac{1}{m}\sum_i c_i, \tag{2.9}$$

and

$$\begin{aligned} u(c_{(t,b)}) &= \left[\frac{1}{m}\sum_i u^2(c_i) + \frac{1}{m}\sum_i (c_i - \mu)^2\right]^{1/2} \\ &= \left[\frac{1}{m}\sum_i u^2(r_i) + \sigma^2\right]^{1/2}, \end{aligned} \tag{2.10}$$

respectively. As noted above, in Equation 2.10, $u(c_i)$ is approximated by the uncertainty associated with the high-quality physical measurement $u(r_i)$. Therefore, $u(c_{(t,b)})$ is the combined standard deviation of the uncertainty associated with the physical measurements and the estimated corrections.

2.1.2 The Bias in Nitrogen-containing Organic Molecules

In this work, the benchmark dataset consisted of 47 uncharged molecules selected from the ATcT (Version 1.130) by searching for permutations of CHN-containing molecules of no more than four non-hydrogen atoms and selecting species that had available data for both $\Delta_{\mathrm{f}}H^{\ominus}$ at 298.15 K and the corresponding uncertainty. The Artificial Bee Colony (ABC) algorithm, implemented in the ABCluster software, was used to generate 100 low-energy conformers using GFN2-xTB Hamiltonian,[95] and only the lowest energy conformer for each molecule was considered thereafter. The lowest-energy conformer was further optimised as part of the G4(MP2)-6X calculation. Where multiple spin states were available, only singlets were selected. The target azole molecules all have singlet ground states with no unpaired electrons, and consequently systems with unpaired electrons were excluded from the benchmarking. The molecules contained within the benchmark dataset are shown in Table 2.1 along with their associated $\Delta_{\mathrm{f}}H^{\ominus}$ values using G4(MP2)-6X. Quantum mechanical calculations are determined in atomic units, which have zero uncertainty because they are defined.[92] However, the conversion to conventional units (e.g., kJ mol^{-1}) and empirical parameters included in QM computations carries uncertainty, which is negligible relative to the uncertainty of the correction. Therefore, the standard uncertainty of the G4(MP2)-6X QM method is negligible relative to the uncertainty of the correction.

Subsequently, Equation 2.3 and Equation 2.4 can be used to determine the G4(MP2)-6X corrected $\Delta_{\mathrm{f}}H^{\ominus}$ and its associated uncertainty, which amounts to the uncertainty in the bias, or correction. A histogram of the corrections for bias in $\Delta_{\mathrm{f}}H^{\ominus}$ is shown in Figure 2.1 from which it can be seen that a slight positive skew is present. The summary

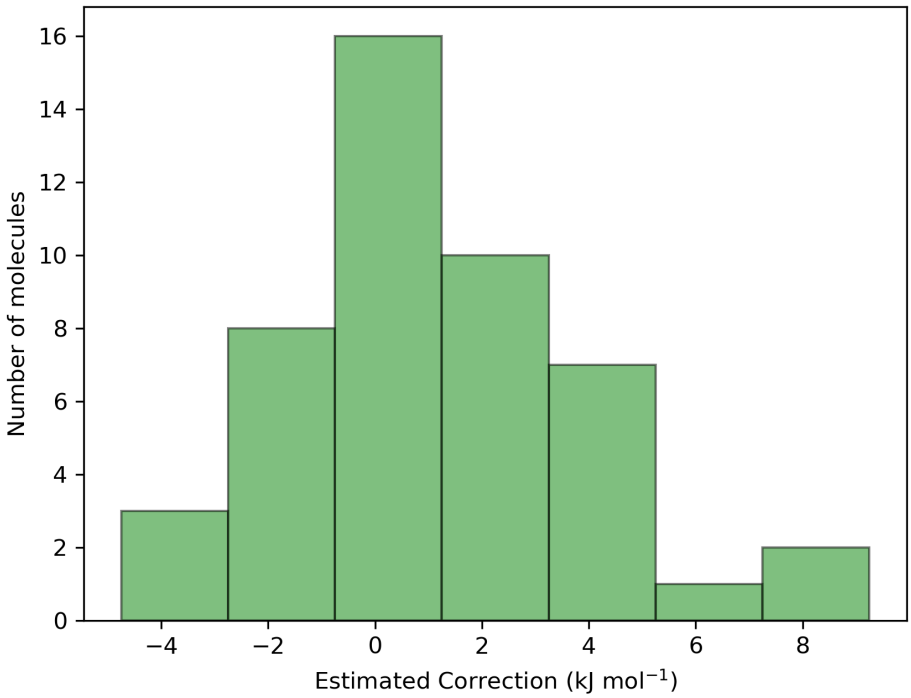


Fig. 2.1. Estimated correction for the enthalpies of formation for CHN-containing molecules, as computed using G4(MP2)–6X.

statistics of the estimated corrections are as follows: the correction is 1.11 kJ mol^{-1} (Equation 2.9), the correction uncertainty is 3.04 kJ mol^{-1} (Equation 2.10), the standard deviation is 2.85 kJ mol^{-1} (Equation 2.6), and the coefficient of skewness is 0.40 kJ mol^{-1} (Equation 2.7). The application of the correction and associated uncertainty can be demonstrated using ethenamine (C_3H_5N). The uncorrected $\Delta_f H^\ominus$ using G4(MP2)-6X is $x_{(t,b)} = 115.04\,\text{kJ mol}^{-1}$. Using the correction and uncertainty, the corrected virtual measurement for ethenamine is $y = x_{(t,b)} + c_{(t,b)} = 116.15\,\text{kJ mol}^{-1}$. However, in thermochemistry, uncertainties are expressed as estimates of the 95 % confidence intervals ($u_{95\%}$) and are calculated as twice the standard deviation.[96] The $u_{95\%}$ uncertainty quantifies the expected accuracy and if properly quantified, the true value should fall within the error bounds 95 % of the time. Therefore, the result of the measurement using G4(MP2)-6X QM for ethenamine should be expressed as (116.15 ± 6.08)kJ mol^{-1}.

Table 2.1 Comparison between the enthalpy of formation of CHN-containing molecules obtained from the Active Thermochemical Tables and computed using the G4(MP2)-6X QM method.[a]

	Molecule	$\Delta_f H^{\circ}_{298.15}$(ATcT)	$\Delta_f H^{\circ}_{298.15}$(G4(MP2)-6X)
1	1-Cyanoethylidene	470.50 ± 1.10	468.15
2	1-Isocyanoethylidene	523.20 ± 1.10	518.87
3	1-Propanamine	−69.95 ± 0.42	-69.49
4	2-Cyclopropen-1-imine	394.10 ± 1.40	395.26
5	2-Methylene-2H-azirine	396.90 ± 1.70	394.42
6	2-Propanamine	−83.45 ± 0.52	-83.19
7	3H-Diazirin-3-ylidene	579.40 ± 2.70	571.72
8	3H-Diazirin-3-ylidene-methylene	839.90 ± 2.40	831.10
9	Acetonitrile	74.06 ± 0.24	74.41
10	Acetylene	228.32 ± 0.13	229.20
11	Acrylonitrile	187.05 ± 0.66	186.50
12	Allene	189.95 ± 0.23	186.41
13	Azete	461.30 ± 1.50	456.82
14	Aziridine	126.72 ± 0.89	127.20
15	Cyanamide	134.50 ± 1.50	139.25
16	Cyanic azide	498.20 ± 1.90	500.42
17	Cyanoacetylene	373.87 ± 0.62	376.26

Continued on next page

Table 2.1 –continued from previous page

	Molecule	$\Delta_f H^\ominus_{298.15}$(ATcT)	$\Delta_f H^\ominus_{298.15}$(G4(MP2)-6X)
18	Cyanogen	310.12 ± 0.41	310.08
19	Cyanomethylene	530.50 ± 1.30	528.31
20	Cyanovinylidene	584.20 ± 1.20	580.64
21	Cyclopropane	53.88 ± 0.38	53.73
22	Cyclopropene	283.63 ± 0.45	282.71
23	Cyclopropenylidene	496.11 ± 0.45	494.65
24	Diisocyanogen	613.20 ± 1.70	606.46
25	Dimethylamine	−17.43 ± 0.40	−16.44
26	Ethane	−84.02 ± 0.12	−84.60
27	Ethanimine	41.07 ± 0.73	41.15
28	Ethenamine	54.71 ± 0.69	57.06
29	Ethylamine	−49.88 ± 0.46	−48.27
30	Ethylene	52.38 ± 0.12	50.16
31	Hydrazine	97.57 ± 0.42	102.19
32	Hydrazoic acid	291.58 ± 0.49	290.53
33	Hydrogen cyanide	129.30 ± 0.09	128.47
34	Hydrogen isocyanide	192.44 ± 0.32	191.43
35	Isocyanoacetylene	486.90 ± 1.00	487.29
36	Isocyanoethene	278.69 ± 0.94	275.67

Continued on next page

Table 2.1 –continued from previous page

	Molecule	$\Delta_f H^{\ominus}_{298.15}$(ATcT)	$\Delta_f H^{\ominus}_{298.15}$(G4(MP2)-6X)
37	Isocyanogen	414.40 ± 1.30	411.70
38	Isocyanomethane	177.30 ± 0.48	173.39
39	Isocyanomethylene	583.60 ± 1.70	579.52
40	Methanimine	88.51 ± 0.36	86.82
41	Methylamine	−21.25 ± 0.23	−19.51
42	N-Methylenemethanamine	79.58 ± 0.68	78.16
43	N-Methylethanamine	−45.80 ± 1.10	−42.17
44	Propane	−105.00 ± 0.16	−105.94
45	Propene	20.02 ± 0.19	18.01
46	Triazirine	457.60 ± 2.60	452.84
47	Trimethylamine	−26.62 ± 0.75	−27.44

[a] All data in kJ mol^{-1}.

In Table 2.1, the largest correction of 8.80 kJ mol^{-1} is for to 3H-diazirin-3-ylidene-methylene, while 3H-diazirin-3-ylidene has the next largest correction of 7.29 kJ mol^{-1}. This may be attributed to the presence of a lone pair on a carbon atom in both structures, which leads to electron correlation effects and the molecules exhibiting multi-reference character. G4(MP2)-6X is paramterised with the G2/97 training set, a set of 148 molecules used to benchmark QM methods, the standard deviation from the G2/97 test set is 5.85 kJ mol^{-1}. Notably, this set contains some of the same molecules in the current benchmark set, such as aziridine, allene, dimethylamine, and propene, among others.[97] For comparison, in this work, the standard deviation was 2.85 kJ mol^{-1}, which shows that G4(MP2)-6X is ideally suited to calculate the enthalpies of formation of CHN molecules.

3 Development of the Group Contribution Method

This section first examines the difference between linear regression and a systematic approach to determining GAVs for n-alkanes. Then two linear regression approaches are used to estimate the enthalpy of formation of azoles. In the first approach, acyclic molecules containing the groups corresponding to those in azoles are calculated. This is followed by the determination of the ring strain (RS) and aromatic stabilisation energy (ASE). The alternative approach is subsequently outlined where aromatic GAVs are regressed using the ten unfunctionalised azoles. This is followed by the regression of azoles functionalised with N-methyl, N-amine, N-azide, N-nitro, and N-nitramide groups, where the fitting is done individually and then collectively for comparison.

3.1 Comparing Systematic Assignment and Regression Fitting

Alkanes provide the largest database for the determination of GAVs and serve as a suitable starting point for their systematic analysis. Benson and Buss[55] used the best available thermochemical data in 1958 to estimate GAVs for several classes of organic compounds. Since then, several GAVs have been revised given more accurate experimental data. The GAVS that form the basis of all n-alkanes, namely: C—$(H)_3$(C) and C—$(H)_2(C)_2$, which exists in two subtly different environments and can be determined using two distinct methods. The first method systematically divides ethane and propane into their constituent groups to determine their GAVs. Ethane consists of two C—$(H)_3$(C) groups and $\Delta_f H^\ominus = -83.34\,\mathrm{kJ\,mol^{-1}}$ using G4(MP2)-6X. By halving $\Delta_f H^\ominus$ for ethane, each C—$(H)_3$(C) group is assigned the value of $-41.67\,\mathrm{kJ\,mol^{-1}}$. It is reasonable to infer that in 1958, Benson and Buss[55] used this method to estimate C—$(H)_3$(C) since there is a $0\,\mathrm{kJ\,mol^{-1}}$ difference between the experimental and estimated values of ethane. Following on, the presence of two C—$(H)_3$(C) groups in propane is used to determine the value of the C—$(H)_2(C)_2$ group, which has a value of $-21.50\,\mathrm{kJ\,mol^{-1}}$.

The second method to determine the GAVs in n-alkanes uses a multiple linear regression (MLR) model applied to molecules of the form C_nH_{2n+2}, where n=2,3,...,16. The mean absolute error (MAE) was used as the evaluation metric to train the model and estimate the coefficients. The regressed GAVs have a value of $-41.34\,\mathrm{kJ\,mol^{-1}}$ for C—$(H)_3(C)$ and $-22.02\,\mathrm{kJ\,mol^{-1}}$ for C—$(H)_2(C)_2$, with a MAE of $0.16\,\mathrm{kJ\,mol^{-1}}$. Domalski and Hearing[67] adopted a similar approach, performing a "global least squares, least sums, or regression-type analysis" with the aid of a desktop electronic calculator. Figure 3.1 compares the estimated errors between the experimental and predicted enthalpies of formation, using both the systematic and regression analysis. If the principle of second-order

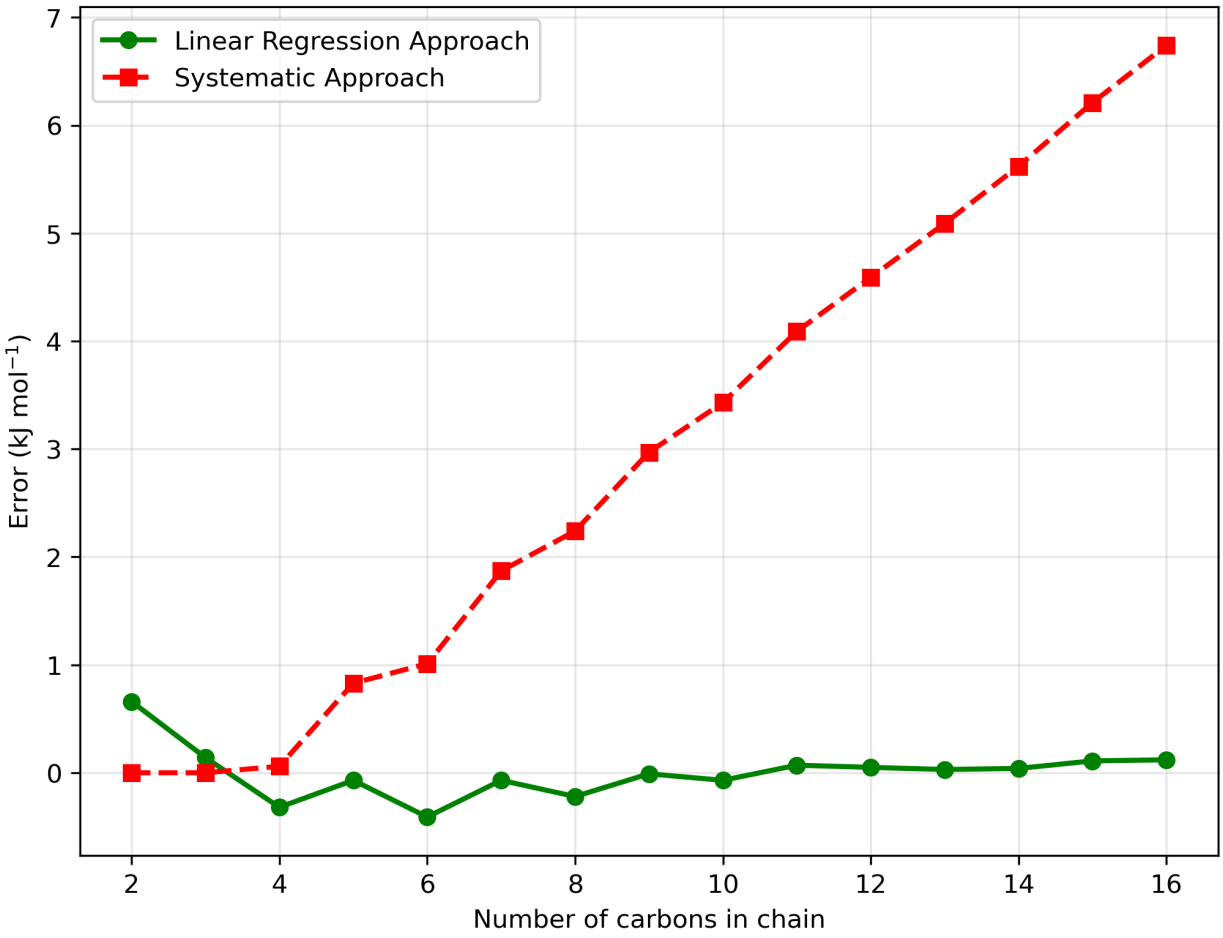


Fig. 3.1. Comparison of estimated errors between experimental and predicted enthalpy of formation for C_2H_6 to $C_{16}H_{34}$ using systematic and regression analyses.

group additivity were exactly followed, then it would be possible to use the exact same GAV for the C—$(H)_2(C)_2$ group irrespective of whether it is bonded to the same or different groups. However, Figure 3.1 shows that as the systematic approach encounters longer carbon chains the errors increase significantly. In contrast, applying a linear regression model $\Delta_f H^\circ$ of 15 alkanes, distributes the error associated with these groups in larger molecules across the fitting set, leading to a more accurate prediction for longer chains.

3.2 Development of a Group Contribution Method for Azoles Using Acyclic Groups

The first GCM explored to estimate $\Delta_f H^\circ$ for unsubstituted azoles used acyclic reference molecules and a correction term, consistent with Benson et al.[66] and Domalski and Hearing[68] treatment of cyclic compounds. The approach was as follows: (i) a database of acyclic molecules that contain GAVs analogous to those present in azoles was developed, (ii) the calculated $\Delta_f H^\circ$ values of the molecules in the database were then used to regress the Benson-type GAV, (iii) the acyclic equivalents were summed up for each azole, and (vi) the difference between the G4(MP2)-6X computed $\Delta_f H^\circ$ and the summation of the acyclic equivalents was assigned as a correction term. In addition to unsubstituted azoles, the dataset contained azoles functionalised with a methyl group and explosophoric functional groups such as amine, azide, nitro, and nitramide on the pyrrole-like nitrogen atom.

At this point, it is necessary to point out that in this work, a distinction is made between carbon and nitrogen atoms depending on the hybridisation and ligands attached, which predominantly adheres to the convention introduced by Benson and Buss.[55] In this notation, there are carbon atoms in an alkene (C=C), referred to as C_d, carbon atoms in imine groups (C=N), referred to as C_I, nitrogen atoms in imine groups, referred to as N_I, and nitrogen atoms in azo groups (N=N), referred to as N_A. Figure 3.2 illustrates the notation of the different carbon and nitrogen groups. It is implied that a sp^2 C_d atom is

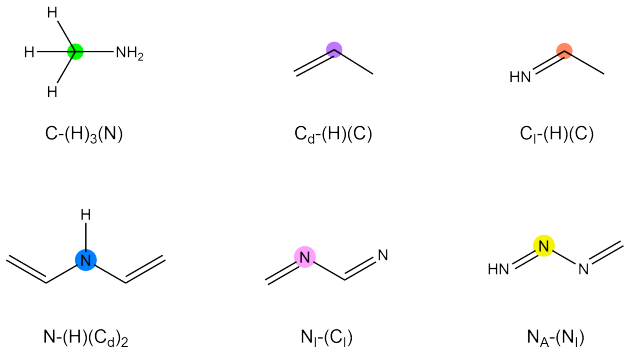
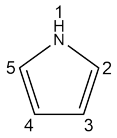
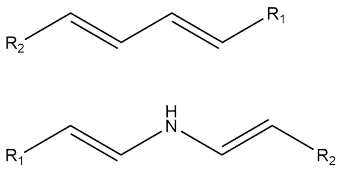
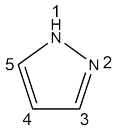
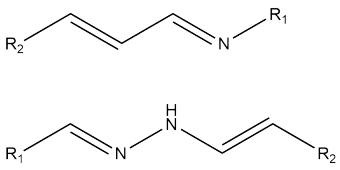
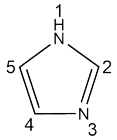
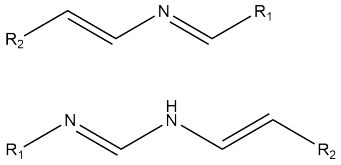
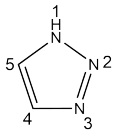
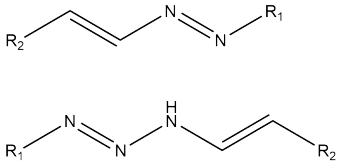
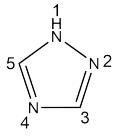
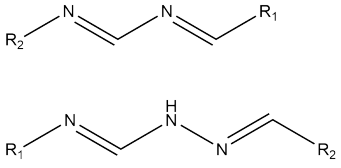
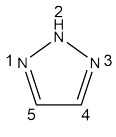
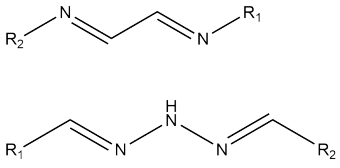
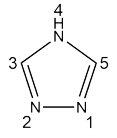
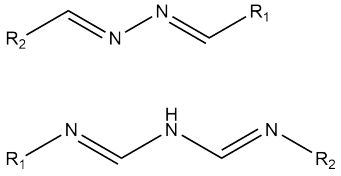
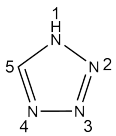
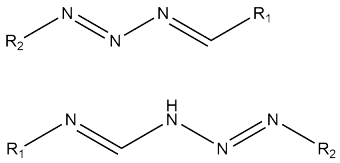
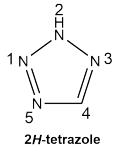
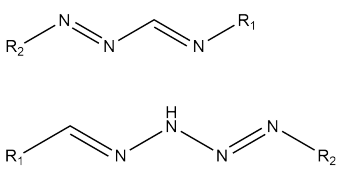
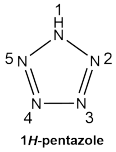
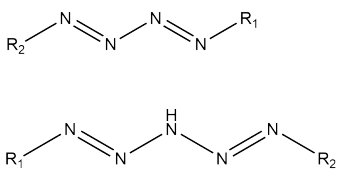


Fig. 3.2. Distinction between notation of the carbon and nitrogen atoms in acyclic groups based on their hybridisation and ligands.

always bonded to another sp^2 C_d atom, and by convention, the ligand C_d atom is not noted in the notation.[66] For example, the notation C_d—(H)(C) in Figure 3.2 is used rather than C_d—(H)(C)(C_d) since the C_d ligand is implied by the subscript of the central atom. This convention is also used for the C_I, N_I, and N_A central atoms so that the respective ligand N_I, C_I, and N_A atoms are not shown. The C_I atom type was not initially used by Benson et al.[66] as the C_d atom type described a carbon atom bonded to either a carbon or nitrogen atom. However, to the best of our knowledge the C_I atom type was introduced by Pappijn et al.[78] to differentiate the bonding partner of the sp^2 carbon.

In order to determine the acyclic group values, two distinct molecular frameworks were compiled for each of the ten unsubstituted azoles, as shown in Table 3.1. Rather than taking each atom in the azole and building a small molecule to determine their GAVs, the pair of frameworks contained all GAVs necessary to determine the azole from which they were derived. The fitting set was assembled by considering R_1 and R_2 as either hydrogen atoms or methyl groups, which results in a total of 72 acyclic molecules that collectively contain 42 groups. 1H-Pyrazole can be used to demonstrate how the frameworks were generated. The first acyclic molecule was determined by eliminating the pyrrole-like nitrogen atom and functionalising the terminal atoms with R groups. In 1H-pyrazole, this framework allows for GAVs to be fitted to unique groups at position 3 and 4 of the ring, which always appear as conjugate pairs. Conjugate pairs refer to groups whose individual contributions

Table 3.1 The acyclic frameworks associated with each azole framework, where R_1 and R_2 are either a hydrogen atom or methyl group.

Azole	Frameworks	Azole	Frameworks
 1*H*-pyrrole		 1*H*-pyrazole	
 1*H*-imidazole		 1*H*-1,2,3-triazole	
 1*H*-1,2,4-triazole		 2*H*-1,2,3-triazole	
 4*H*-1,2,4-triazole		 1*H*-tetrazole	
 2*H*-tetrazole		 1*H*-pentazole	

cannot be isolated, as one group is inherently linked to the other and always appears next to it. For example, the first acyclic molecule of 1H-pyrazole shows that C_d—(H)(C_I) always appears next to C_I—(H)(C_d), no matter what R_1 and R_2 are and prevents their separate GAV determination. In a system with a conjugate pair, values were arbitrarily assigned to one of the pairs or their occurrences were merged with another group that it is set equivalent to prior to the fitting, making the estimation easier without influencing the molecule's predicted property. The second acyclic molecular framework was obtained by breaking the single bond opposite the pyrrole-like nitrogen in the ring. This is necessary to account for the neighbouring sp^2 carbon and sp^2 nitrogen atoms in 1H-pyrazole at positions 2 and 3, respectively.

The lowest energy conformers were calculated using the GFN2-xTB semiempirical method by the Global Optimization Algorithm (GOAT), followed by a further optimization using the low-cost HF-3c QM method to confirm the global minimum. It should be noted that in their lowest energy structures, these molecular frameworks (as drawn) have double bonds in a s-trans conformation, whereas in the corresponding azoles the double bonds are s-cis. However, the choice was made to fit to each molecule in its (global) minimum energy conformation to ensure reproducibility and avoid an arbitrary choice of higher energy conformations. The enthalpies of formation for the 72 acyclic molecules were then calculated using G4(MP2)-6X, and these values were corrected using a bias of $1.11\,\mathrm{kJ\,mol^{-1}}$. Initially, multiple linear regression was performed on the total set of 72 acyclic molecules using the 42 groups. However, there are only 24 independent variables, which indicates the intercorrelation between 18 GAVs, also referred to as multicollinear variables.[98] Multiple linear regression models with mutlicollinearity are not well-defined, as the system of equations has infinitely many solutions. To make a tractable system, the 18 multicollinear GAVs must therefore be assigned values or made equivalent to other groups. The latter can be achieved by initially assigning and counting the occurrences of

the group as usual, but for the regression analysis, these occurrences are then merged with the number of another group to which it is set equivalent. An example of the former, in Section 3.1, C—$(H)_3$(C) has a value of $-41.49\,\mathrm{kJ\,mol^{-1}}$. Benson later proposed a convention where C—$(H)_3$(X) is made equivalent to C—$(H)_3$(C) for any polyvalent atom X, such as C_d, O, S, N, and N_I.[54] Benson and Buss initially argued that the group C—$(H)_3(C_d)$ should not be assigned the same value as C—$(H)_3$(C), as this could lead to inaccurate and potentially misleading results if applied to other groups.[55] Subsequently, additional data would be needed to differentiate between the two GAVs. However, most GCMs typically apply the convention that the value of a methyl group remains constant regardless of its attachments.[54,68,71,73] Therefore, in the current fitting, C—$(H)_3(C_d)$, C—$(H)_3(C_I)$, C—$(H)_3(N_I)$, and C—$(H)_3(N_A)$ are all assigned the value of $-41.49\,\mathrm{kJ\,mol^{-1}}$. This leaves 14 multicollinear variables that must be addressed.

Ethene (CH_2=CH_2) has a value of $51.00\,\mathrm{kJ\,mol^{-1}}$ for $\Delta_f H^\ominus$ that can be halved, due to the symmetry of the molecule, to determine the value of the group C_d—$(H)_2$ as $25.95\,\mathrm{kJ\,mol^{-1}}$. Since the GAVs in methanimine (CH_2=NH) form a conjugate pair, Benson et al.[66] assumed the value of C_I—$(H)_2$ is equivalent to C_d—$(H)_2$ to allow for the unique determination of N_I—(H). The $\Delta_f H^\ominus$ value of methanimine is $87.94\,\mathrm{kJ\,mol^{-1}}$, resulting in N_I—(H) having a value of $61.99\,\mathrm{kJ\,mol^{-1}}$. Additionally, $\Delta_f H^\ominus$ for diazene (NH=NH) is $200.76\,\mathrm{kJ\,mol^{-1}}$. Therefore, each N_A—(H) group has a value of $100.38\,\mathrm{kJ\,mol^{-1}}$. These four GAVs were assigned values based on these calculations, while the 10 remaining multicollinear GAVs were increasingly more complex to do through the systematic approach and required analysis of the correlation matrix.

A correlation matrix illustrates the strength and direction of the correlation between all possible pairwise combinations of GAVs, which ranges from -1 to $+1$, where -1 represents a perfect negative correlation, 0 represents no correlation, and $+1$ represents a perfect positive correlation. A perfect positive correlation signifies a direct linear relationship between two variables, which means that a change in one variable has a proportional

effect in the same direction on another variable. In a perfect negative correlation, the increase in one variable results in a proportional decrease in another. The correlation matrix (Figure 3.3) indicates six perfect positive correlations between the GAVs. The N_A—(C_d) group is perfectly correlated with C_d—(H)(N_A). This can be resolved either by assigning a value to one of two groups, or by combining it with a different group and assigning

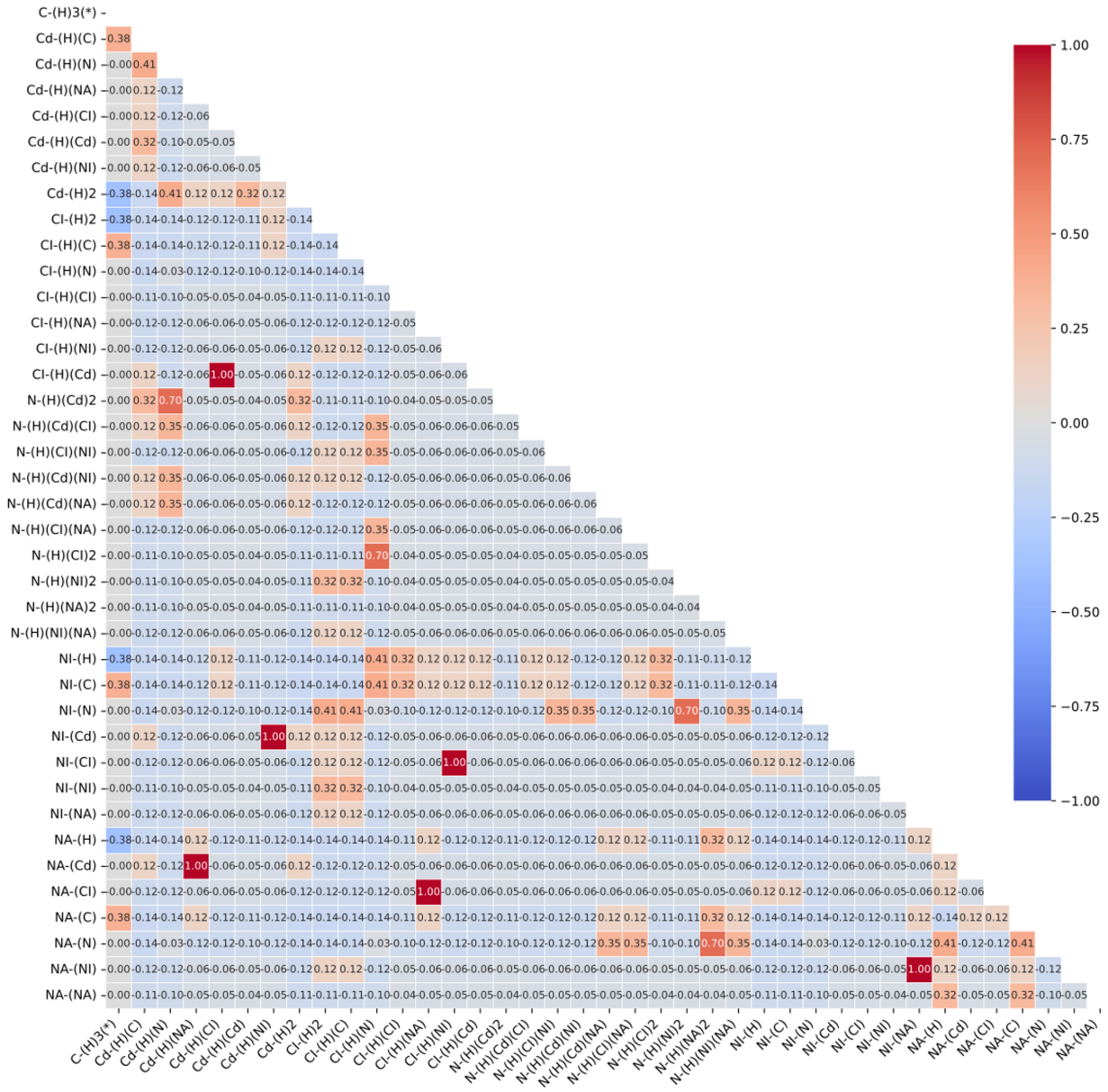


Fig. 3.3. Correlation matrix showing all groups in acyclic molecules. Blue shading represents negative correlation, while red shading denotes positive correlation.

equivalent values. Since the central atom is different between these two groups, it is more chemically consistent to assign C_d—(H)(N_A) to a group that contains C_d as the central atom with one hydrogen atom as a ligand (e.g., C_d—(H)(C_I), C_d—(H)(C_d), or C_d—(H)(N_I)). A similar assumption was made by Benson, in which C_d—(H)(C_d) was set equivalent to C_d—(H)(N_I).[54] More recently, Pappijn et al.[78] treated C_d—(H)(C_d) as equivalent to C_d—(N_I) and C_d—(H)(C_I). Therefore, in the present fitting, C_d—(H)(N_A) is fitted equivalent to C_d—(H)(C_d) to resolve the perfect correlation. Another perfect correlation is between C_I—(H)(C_d) and C_d—(H)(C_I) that can be addressed by making C_d—(H)(C_I) equivalent to C_d—(H)(C_d) as well. The perfect correlation between N_I—(C_d) and C_d—(H)(N_I) is resolved by also making C_d—(H)(N_I) equivalent to C_d—(H)(C_d). The same approach is taken for groups that have C_I as the central atom to address the perfect correlation between N_I—(C_I) and C_I—(H)(N_I), such that C_I—(H)(N_I) is made equivalent to C_I—(H)(C_d). In the correlation matrix, N_A—(C_I) is perfectly correlated with C_I—(H)(N_A). Considering the available groups that have N_A as the central atom, the most suitable group that it can be made equivalent to is N_A—(C_d) since the ligands are both sp^2 carbon atoms. Similarly, the perfect correlation between N_A—(N_I) and N_I—(N_A) is resolved by making N_A—(N_I) equivalent to N_A—(N_A) because both ligands are sp^2 nitrogen atoms. After resolving these perfect correlations, four multicollinear groups remain.

In a correlation matrix, the closer a coefficient is to 1, or -1, the stronger the linear relationship between variables, which indicates that one variables provides some predictive power for another. The remaining groups that must be assigned to resolve the multicollinearity of the system all have positive correlation coefficients larger than 0.7. Another equivalence that Pappijn et al.[78] employed is between C_d—(H)(C) and C_d—(H)(N). This was also used in the current fitting to address the strong positive correlation between N—(H)$(C_d)_2$ and C_d—(H)(N). The strong positive correlation between N—(H)$(C_I)_2$ and C_I—(H)(N) is treated analogously by making C_I—(H)(N) equivalent to C_I—(H)(C). Lastly, the strong positive correlation between N_I—(N) and N—(H)$(N_A)_2$, as well as N_A—(N) and

N—(H)$(N_A)_2$ are resolved by making N_I—(N) equivalent to N_I—(C) and N_A—(N) equivalent to N_A—(C), respectively. The summary of all the equivalences made are shown in Table 3.2. With multicollinearity addressed, linear regression was repeated to determine the 24 unique contributions and the full list of groups and their GAVs are shown in Table 3.3. The MAE of this fitting is 3.31 kJ mol^{-1} and the MAPE is 1.70 %.

Table 3.2 Summary of the group equivalences implemented in the acyclic group regression model.

Group equivalences
$C—(H)_3(C) \equiv C—(H)_3(C_d) \equiv C—(H)_3(C_I) \equiv C—(H)_3(N_I) \equiv C—(H)_3(N_A)$
$C_d—(H)_2 \equiv C_I—(H)_2$
$C_d—(H)(C_d) \equiv C_d—(H)(C_I) \equiv C_d—(H)(N_A) \equiv C_d—(H)(N_I)$
$C_d—(H)(C) \equiv C_d—(H)(N)$
$C_I—(H)(C_d) \equiv C_I—(H)(N_I)$
$C_I—(H)(C) \equiv C_I—(H)(N)$
$N_I—(C) \equiv N_I—(N)$
$N_A—(C) \equiv N_A—(N)$
$N_A—(N_I) \equiv N_A—(N_A)$

Table 3.3 Group additivity values derived from G4(MP2)-6X calculated enthalpy of formation data of the acyclic molecular frameworks shown in Table 3.1.[a]

Group	GAV
C_d—(H)(C)	38.99
C_d—(H)(N)	38.99
C_d—(H)(C_d)	26.67
C_d—(H)(C_I)	26.67
C_d—(H)(N_I)	26.67
C_d—(H)(N_A)	26.67
C_I—(H)(C)	26.37
C_I—(H)(N)	26.37
C_I—(H)(C_d)	20.94
C_I—(H)(N_I)	20.94
C_I—(H)(C_I)	23.90
C_I—(H)(N_A)	30.00
N—(H)(C_d)$_2$	14.28
N—(H)(C_d)(C_I)	−13.39
N—(H)(C_I)(N_I)	43.01
N—(H)(C_d)(N_A)	44.05
N—(H)(C_I)(N_A)	19.85
N—(H)(C_I)$_2$	−29.04
N—(H)(N_I)$_2$	111.18
N—(H)(N_A)$_2$	78.64
N—(H)(N_I)(N_A)	96.89
N_I–(C)	97.40
N_I—(N)	97.40

Continued on next page

Table 3.3 –continued from previous page

Group	GAV
N_I—(C_d)	83.08
N_I—(C_I)	77.16
N_I—(N_I)	105.60
N_I—(N_A)	104.06
N_A—(C_d)	121.40
N_A—(C_I)	121.40
N_A—(C)	116.81
N_A—(N)	116.81
N_A—(N_I)	128.58
N_A—(N_A)	128.58

[a] All data in $\mathrm{kJ\,mol^{-1}}$

The enthalpy of formation of the azoles (without "ring strain") can now be calculated using the corresponding constituent acyclic GAVs shown in Table 3.3. Considering that the molecular frameworks used for fitting have conjugated double bonds, but no aromatic stabilisation energy (ASE) and ring strain (RS), the difference between the G4(MP2)-6X calculated $\Delta_f H^\circ$ of the azole and the sum of acyclic GAVs can be attributed to the sum of these effects. For example, the $\Delta_f H^\circ$ of 1H-pyrrole can be calculated using acyclic groups as shown below,

$$\begin{aligned}\Delta_f H^\circ(\text{1H-pyrrole}) &= \text{N—(H)(C}_d\text{)(N}_I\text{)} + 2[\text{C}_d\text{—(H)(N)}] + 2[\text{C}_d\text{—(H)(C}_d\text{)}] \\ &= 145.60\mathrm{kJ\,mol^{-1}}.\end{aligned} \tag{3.1}$$

The G4(MP2)-6X calculated $\Delta_f H^\ominus$ is 109.44 kJ mol^{-1}. The difference of -36.16 kJ mol^{-1} is attributed to the sum of RS and ASE. Table 3.3 provides the RS and ASE for each azole calculated in a similar way.

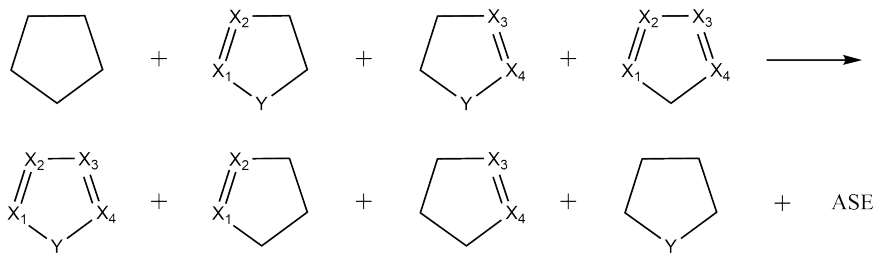
It is clear from Table 3.4 that the sum of RS and ASE differ vastly among the azoles and it might be reasonable to question the qualitative accuracy of the described approach to determine these values. Consequently, the ASE and RS contributions were also determined from first principles using homodesmotic reactions, for comparison. A homodesmotic reaction is a reaction that conserves the same type and number of bonds between heavy atoms while maintaining the number of heavy atoms with attached hydrogen atoms in both the reactants and products (see Section 4.2.2 for further details). Cyrański et al.[91] investigated 105 five-membered aromatic, non-aromatic, and anti-aromatic systems using

Table 3.4 Comparison between the G4(MP2)-6X calculated and GAV-estimated enthalpy of formation for azoles, along with the corresponding ring strain and aromatic stabilisation energy correction term.[a]

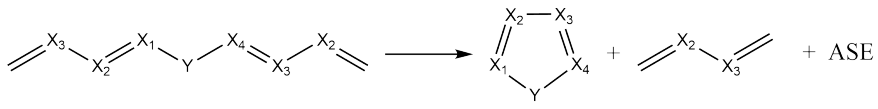
Azole	G4(MP2)-6X	$\sum$GAVs	RS + ASE
1H-Pyrrole	109.44	145.60	36.16
1H-Pyrazole	178.08	249.82	71.74
1H-Imidazole	132.62	161.72	29.10
1H-1,2,3-Triazole	263.62	347.92	84.30
1H-1,2,4-Triazole	193.42	264.88	71.46
2H-1,2,3-Triazole	248.34	353.78	105.44
4H-1,2,4-Triazole	217.67	234.90	17.23
1H-Tetrazole	333.53	395.67	62.14
2H-Tetrazole	325.48	462.50	137.02
1H-Pentazole	451.80	569.42	117.62

[a] All data in kJ mol^{-1}

homodesmotic and isodesmic reactions to determine aromatic stabilisation energy (ASE). The two homodesmotic reactions of interest are,



(3.2)



(3.3)

In these reactions, X_1, X_2, X_3, and X_4 are carbon or nitrogen atoms, while Y is a nitrogen atom for azoles, and the ASE is calculated as,

$$\text{ASE} = \sum \Delta_\text{f} H^\ominus[\text{reactants}] - \sum \Delta_\text{f} H^\ominus[\text{products}]. \tag{3.4}$$

The first homodesmotic reaction is a revision of the isodesmic reaction presented by Schleyer et al.[90] In the construction of the first homodesmotic reaction, the system of interest is shown as the first product molecule is considered to have three unsaturated units that are reflected in the reactants through the corresponding unsaturated reference molecules. The remaining reference molecules preserve the homodemsotic conditions of the reaction. Due to the presence of only five-membered rings in their most stable conformations in the reaction, the strain effects cancel so that the reaction accounts for the ASE resulting from cyclic conjugation alone. High positive values indicate aromaticity, whereas negative values indicate anti-aromatic systems.[90] The second homodesmotic reaction is derived from schemes often used to estimate the ASE of benzene.[99,100] In contrast to the first homodesmotic reaction, the second uses acyclic analogues in their lowest energy conformations. Consequently, the calculated "ASE" of the reference molecules also includes ring strain. Another included effect is the contribution of having double bonds s-*cis* in the cyclic

molecule as opposed to s-*trans* in the acyclic reference. A further four isodesmic reactions were considered by Cyrański et al.;[91] however, homodesmotic reactions provide greater accuracy and isodesmotic reactions should not be expected to give reliable ASE values. In applying these two homodesmotic reactions to the azoles, the lowest energy conformers of the reference molecules were again calculated using the GFN2-xTB semiempirical method by the Global Optimization Algorithm (GOAT), followed by a further optimisation using the low-cost HF-3c QM method to confirm the global minimum. The results of the homodesmotic reactions are given in Table 3.5 along with a comparison of the results obtained

Table 3.5 Comparison between ring strain and aromatic stabilisation energy term with two homodesmotic reactions.[a]

	This work			Cyrański et al.[b]	
Azole	RS + ASE	ASE 1	ASE 2	ASE 1	ASE 2
1H-Pyrrole	36.16	75.45	35.99	75.50	22.01
1H-Pyrazole	71.74	87.10	71.18	85.60	55.44
1H-Imidazole	29.10	69.35	32.39	67.70	14.56
1H-1,2,3-Triazole	84.30	88.20	68.61	84.56	46.48
1H-1,2,4-Triazole	71.46	77.83	82.87	75.35	60.58
2H-1,2,3-Triazole	105.44	93.28	98.60	92.93	93.97
4H-1,2,4-Triazole	17.23	53.97	12.20	51.00	-5.10
1H-Tetrazole	62.14	65.67	69.31	59.12	43.30
2H-Tetrazole	137.02	92.66	119.81	92.68	97.19
1H-Pentazole	117.62	N/A	135.54	N/A	95.90

[a] All data in $kJ\,mol^{-1}$

[b] Reference 91.

by Cyrański et al.,[91] where ASE 1 corresponds to the first homodesmotic reaction (Equation 3.2) and ASE 2 corresponds to the second homodesmotic reaction (Equation 3.3).

As noted previously, the values for ASE 1 reflect only the ASE as the reference molecules are all cyclic, while ASE 2 uses acyclic reference molecules that are additionally perturbed by strain effects, thus comparing better to the corrections determined using GAVs. Since the current fitting derives values from acyclic reference molecules, RS + ASE can be directly compared to ASE 2, yielding a Pearson correlation coefficient of 0.96. There is a discrepancy between the values determined in this work and those reported by Cyrański et al.[91] despite using the same homodesmotic frameworks. This is likely due to a difference in the conformer selection used between both methods.

Lastly, an investigation to assess whether the RS and ASE terms are transferable to functionalised azoles was conducted. The azoles were functionalised on the pyrrole-like nitrogen with a methyl group as well as explosophoric groups, which included amine ($-NH_2$), azide ($-N_3$), nitro ($-NO_2$), and nitramide ($-NHNO_2$). This required the determination of GAVs for the explosphoric groups. The 10 acyclic frameworks shown in Figure 3.4 was used to determine the necessary groups, which corresponds to the second framework for each azole in Table 3.1 with functionalisation on the pyrrole-like nitrogen, and $R_3 = -CH_3$, $-NH_2$, $-N_3$, $-NO_2$, $-NHNO_2$. This resulted in a total of 180 acyclic molecules. Again, the lowest energy conformers were calculated using the GFN2-xTB semi–empirical method by the Global Optimisation Algorithm (GOAT), followed by a further optimisation using the low-cost HF-3c QM method to confirm the global minimum and the enthalpy of formation was calculated using G4(MP2)-6X with a bias of $1.11\,\mathrm{kJ\,mol^{-1}}$ was applied. The N-functionalised groups were regressed while the groups in Table 3.3 were kept constant. Again, the value of the methyl group was $-41.49\,\mathrm{kJ\,mol^{-1}}$, while the contribution of N—$(H)_2$(N) $= 51.65\,\mathrm{kJ\,mol^{-1}}$ was determined by halving $\Delta_f H^\circ$ for hydrazine (N_2H_4). The remainder of the explosophoric groups were determined by subtracting the value of the N—$(H)_2$(N) from aminoazide (H_2N_4) results in $371.31\,\mathrm{kJ\,mol^{-1}}$ for N_3—(N). Similarly,

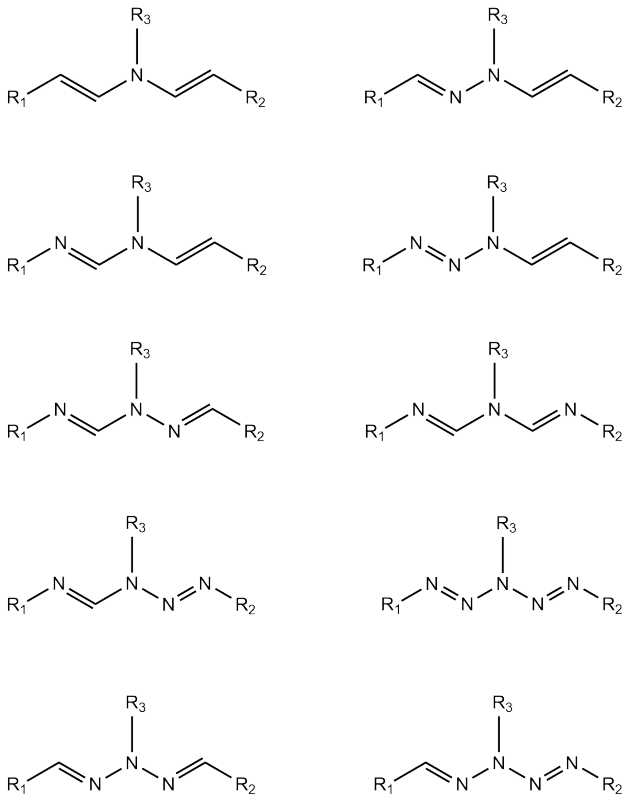


Fig. 3.4. N-functionalised frameworks where R_1 and R_2 are a hydrogen atom or methyl group, while R_3 is $-CH_3$, $-NH_2$, $-N_3$, $-NO_2$, or $-NHNO_2$.

subtracting N—$(H)_2$(N) from nitramide ($H_2N_2O_2$) gives a value of $-39.26\,\mathrm{kJ\,mol^{-1}}$ for NO_2—(N). Finally, removing N—$(H)_2$(N) from nitric hydrazide ($H_3N_3O_2$) obtains a value of $58.48\,\mathrm{kJ\,mol^{-1}}$ for $NHNO_2$—(N). Multicollinearity was not observed in the fit, as each molecule was modelled by fitting only the pyrrole-like nitrogen, which are all independent, while the other groups were assigned values based on previous fittings. The results for the groups and their GAVs is given in Table 3.6. Then, using the data in Tables 3.4, 3.3, 3.6, and the definition of RS + ASE in Equation 3.4, the N-functionalised azoles can be estimated. For example, 1-methylpyrrole and pyrrole-1-amine can be calculated using

the acyclic groups as shown below,

$$\begin{aligned}\Delta_{\mathrm{f}}H^{\circ}(\text{1-Methylpyrrole}) &= \text{C—(H)}_3\text{(N)} + \text{N—(C)(C}_\text{d}\text{)}_2 + 2[\text{C}_\text{d}\text{—(H)(N)}] \\ &\quad + 2[\text{C}_\text{d}\text{—(H)(C}_\text{d}\text{)}] - [\text{RS + ASE}]_{\text{1H-Pyrrole}} \\ &= 99.48\mathrm{kJ\,mol^{-1}},\end{aligned} \tag{3.5}$$

$$\begin{aligned}\Delta_{\mathrm{f}}H^{\circ}(\text{Pyrrol-1-amine}) &= \text{N—(H)}_2\text{(N)} + \text{N—(N)(C}_\text{d}\text{)}_2 + 2[\text{C}_\text{d}\text{—(H)(N)}] \\ &\quad + 2[\text{C}_\text{d}\text{—(H)(C}_\text{d}\text{)}] - [\text{RS + ASE}]_{\text{1H-Pyrrole}} \\ &= 215.23\mathrm{kJ\,mol^{-1}}.\end{aligned} \tag{3.6}$$

The remainder of the G4(MP2)-6X computed values, GCM estimated values, AE and APE are given in Table 3.7. The MAE for this fitting was 18.05 $\mathrm{kJ\,mol^{-1}}$ and the MAPE was 4.69 %.

Table 3.6 Group additivity values for N-functionalised acyclic molecules derived from G4(MP2)-6X calculated enthalpy of formation data.[a]

Group	GAV
N—(C)$(C_d)_2$	45.81
N—(C)$(C_d)(C_I)$	15.46
N—(C)$(C_I)(N_I)$	68.06
N—(C)$(C_d)(N_I)$	96.23
N—(C)$(C_d)(N_A)$	69.34
N—(C)$(C_I)(N_A)$	43.99
N—(C)$(C_I)_2$	0.46
N—(C)$(N_I)_2$	141.76
N—(C)$(N_A)_2$	109.13
N—(C)$(N_I)(N_A)$	121.84
N—(N)$(C_d)_2$	68.42
N—(N)$(C_d)(C_I)$	42.33
N—(N)$(C_I)(N_I)$	79.34
N—(N)$(C_d)(N_I)$	108.52
N—(N)$(C_d)(N_A)$	78.37
N—(N)$(C_I)(N_A)$	77.29
N—(N)$(C_I)_2$	20.69
N—(N)$(N_I)_2$	154.19
N—(N)$(N_A)_2$	124.16
N—(N)$(N_I)(N_A)$	133.95
N—$(N_3)(C_d)_2$	88.17
N—$(N_3)(C_d)(C_I)$	70.81
N—$(N_3)(C_I)(N_I)$	111.2

Continued on next page

Table 3.6 –continued from previous page

Group	GAV
N—$(N_3)(C_d)(N_I)$	126.68
N—$(N_3)(C_d)(N_A)$	107.29
N—$(N_3)(C_I)(N_A)$	104.55
N—$(N_3)(C_I)_2$	51.73
N—$(N_3)(N_I)_2$	171.29
N—$(N_3)(N_A)_2$	154.31
N—$(N_3)(N_I)(N_A)$	160.19
N-$(NO_2)(C_d)_2$	101.23
N-$(NO_2)(C_d)(C_I)$	93.33
N-$(NO_2)(C_I)(N_I)$	154.28
N-$(NO_2)(C_d)(N_I)$	161.08
N-$(NO_2)(C_d)(N_A)$	164.78
N-$(NO_2)(C_I)(N_A)$	158.64
N-$(NO_2)(C_I)_2$	88.57
N-$(NO_2)(N_I)_2$	210.08
N-$(NO_2)(N_A)_2$	220.22
N-$(NO_2)(N_I)(N_A)$	216.42
N-$(NHNO_2)(C_d)_2$	87.05
N-$(NHNO_2)(C_d)(C_I)$	58.70
N-$(NHNO_2)(C_I)(N_I)$	102.09
N-$(NHNO_2)(C_d)(N_I)$	131.40
N-$(NHNO_2)(C_d)(N_A)$	110.11
N-$(NHNO_2)(C_I)(N_A)$	98.42
N-$(NHNO_2)(C_I)_2$	38.24

Continued on next page

Table 3.6 –continued from previous page

Group	GAV
N-$(NHNO_2)(N_I)_2$	173.70
N-$(NHNO_2)(N_A)_2$	157.72
N-$(NHNO_2)(N_I)(N_A)$	159.52

[a] All data in $kJ\,mol^{-1}$

Table 3.7 Comparison between the G4(MP2)-6X calculated and group contribution method ($\sum$GAVs) estimated enthalpy of formation of functionalised azoles, along with corresponding absolute error (AE) and absolute percentage error (APE).[a]

Azole	G4(MP2)-6X	$\sum$GAVs[b]	AE	APE
1-Methylpyrrole	101.85	99.48	2.37	2.33
1-Methylimidazole	123.60	119.98	3.62	2.92
1-Methylpyrazole	161.02	167.00	5.98	3.71
1-Methyl-1,2,3-triazole	244.94	247.42	2.48	1.01
1-Methyl-1,2,4-triazole	175.10	176.98	1.88	1.07
2-Methyl-1,2,3-triazole	226.41	237.43	11.02	4.87
4-Methyl-1,2,4-triazole	209.42	205.68	3.74	1.79
1-Methyltetrazole	312.67	316.18	3.51	1.12
2-Methyltetrazole	299.00	308.94	9.94	3.32
1-Methylpentazole	420.43	440.80	20.37	4.85
Pyrrol-1-amine	224.73	215.23	9.50	4.23
Imidazole-1-amine	250.18	239.99	10.19	4.07
Pyrazol-1-amine	277.09	272.43	4.66	1.68
Triazol-1-amine	364.14	349.59	14.55	4.00

Continued on next page

Table 3.7 –continued from previous page

Azole	G4(MP2)-6X	∑GAVs	AE	APE
1,2,4-Triazol-1-amine	292.99	281.40	11.59	3.96
Triazol-2-amine	353.50	343.00	10.50	2.97
1,2,4-Triazol-4-amine	337.13	319.05	18.08	5.36
Tetrazol-1-amine	436.55	442.62	6.07	1.39
Tetrazol-2-amine	429.26	414.19	15.07	3.51
Pentazol-1-amine	556.98	548.97	8.01	1.44
1-Azidopyrrole	568.62	554.64	13.98	2.46
1-Azidoimidazole	599.31	588.13	11.18	1.87
1-Azidopyrazole	629.62	610.25	19.37	3.08
1-Azidotriazole	720.70	698.17	22.53	3.13
1-Azido-1,2,4-triazole	650.89	632.92	17.97	2.76
2-Azidotriazole	707.81	679.76	28.05	3.96
4-Azido-1,2,4-Triazole	691.02	669.75	21.27	3.08
1-Azidotetrazole	799.44	789.54	9.90	1.24
2-Azidotetrazole	788.06	760.09	27.97	3.55
1-Azidopentazole	920.40	898.78	21.62	2.35
1-Nitropyrrole	172.03	157.13	14.90	8.66
1-Nitroimidazole	209.94	200.08	9.86	4.70
1-Nitropyrazole	256.30	234.08	22.22	8.67
1-Nitrotriazole	359.65	345.09	14.56	4.05
1-Nitro-1,2,4-triazole	285.22	265.43	19.79	6.94
2-Nitrotriazole	348.70	307.98	40.72	11.68
4-Nitro-1,2,4-triazole	310.42	296.02	14.40	4.64
1-Nitrotetrazole	443.78	433.06	10.72	2.42

Continued on next page

Table 3.7 –continued from previous page

Azole	G4(MP2)-6X	$\sum$GAVs	AE	APE
2-Nitrotetrazole	440.92	405.75	35.17	7.98
3-Nitropentazole	582.04	554.12	27.92	4.80
N-Pyrrol-1-ylnitramide	263.21	240.69	22.52	8.56
N-Imidazol-1-ylnitramide	296.38	263.19	33.19	11.20
N-Pyrazol-1-ylnitramide	320.41	302.14	18.27	5.70
N-Triazol-1-ylnitramide	417.37	388.16	29.21	7.00
N-1,2,4-Triazol-1-ylnitramide	344.27	310.98	33.29	9.67
N-1,2,3-Triazol-2-ylnitramide	403.67	369.34	34.33	8.50
N-1,2,4-Triazol-4-ylnitramide	390.64	343.43	47.21	12.09
N-Tetrazol-1-ylnitramide	497.45	470.58	26.87	5.40
N-Tetrazol-2-ylnitramide	489.83	446.59	43.24	8.83
N-Pentazol-3-ylnitramide	626.70	589.36	37.34	5.96

[a] All data in kJ mol^{-1}.

[b] Includes respective RS + ASE correction.

3.3 Analysis of Group Contribution Method Using Aromatic Groups

The second GCM investigated to estimate $\Delta_f H^\ominus$ used MLR of azoles as the fitting set, but differs from the previous model as it moves away from a "ring strain" correction, towards an approach where the contributions of RS and ASE are directly parameterised into each GAV. In the second GCM, the notation presented in Figure 3.5 was used, which distinguishes between the pyrrole-like nitrogen, denoted as N, and the "pyridine-like" nitrogen, denoted as N_p. Additionally, the convention introduced by Benson and Buss[55] was followed, in which the C_B atom type is used for all aromatic carbon atoms, where the subscript represents a "benzene–type" carbon. This notation captures how the C_B—$(H)(C_B)_2$ group was initially derived by Benson and Buss,[55] which was to divide $\Delta_f H^\ominus$ of benzene by six.

The determination of these GAVs can be examined using two strategies: (i) fitting the set of unsubstituted azoles, and fixing their GAVs in subsequent fittings as functional groups are added, or (ii) fitting the full set of azoles that incorporates all unsubstituted and

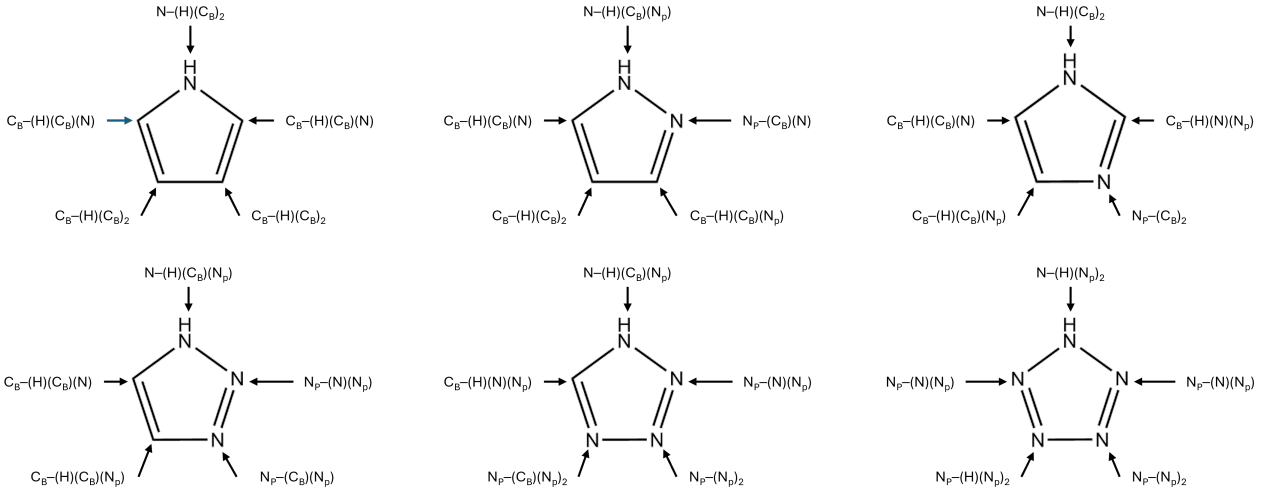


Fig. 3.5. Aromatic group assignments shown for 1H-pyrrole (top left), 1H-pyrazole (top centre), 1H-imidazole (top right), 1H-1,2,3-triazole (bottom left), 1H-tetrazole (bottom centre), and 1H-pentazole (bottom right).

functionalised groups simultaneously. The first strategy discussed is the "conventional" method that was done by Benson et al.,[66] who first determined the groups in alkanes and kept these values fixed in the subsequent determinations of the other groups. Similarly, the first approach reported here, the GAVs for unsubstituted azoles are determined and kept fixed in the functionalised azole regression.

As before, the lowest energy conformers were calculated for the full fitting set using the GFN2-xTB semiempirical method by the Global Optimisation Algorithm (GOAT) followed by a further optimisation using the cost-effective HF-3c method to confirm the global minimum and the enthalpy of formation was calculated using G4(MP2)-6X with a bias of 1.11 kJ mol^{-1} was applied. In the unsubstituted azoles, there are a total of 13 groups present and five exhibit multicollinearity that need to be made equivalent to other groups or removed from the fitting. Using the correlation matrix shown in Figure 3.6 as a guide, the N_p—$(N_p)_2$ and N_p—(N)(N_p) groups are made equivalent prior to the fitting to resolve their strong correlation of 0.81. Furthermore, C_B—(H)$(C_B)_2$ has a strong positive correlation of 0.81 with C_B—(H)(C_B)(N). This correlation can be resolved by fixing the value of C_B—(H)$(C_B)_2$ to that of a sixth of benzene (13.42 kJ mol^{-1}) and assigning the C_B—(H)(C_B)(N) the same value as C_B—(H)$(C_B)_2$. To address the next highest correlation of 0.56 between C_B—(H)(C_B)(N_p) and N_p—(C_B)(N_p), as well as C_B—(H)(N)(N_p) and C_B—(H)$(C_B)_2$, C_B—(H)(C_B)(N_p) and C_B—(H)(N)(N_p) were fixed to the same value as C_B—(H)$(C_B)_2$. Notably, Domalski and Hearing[68] treated *all* aromatic carbon atoms in five-membered heterocyclic aromatic compounds (e.g., pyrrole, furan, and thiophene) as C_B—(H)$(C_B)_2$ groups, which was also derived by dividing the enthalpy of formation of benzene by six. With multicollinearity resolved, the eight groups were fitted against the set of ten azoles, and the results are shown in Table 3.8. The fitting has an MAE of 0.11 kJ mol^{-1} and an MAPE of 0.04 %, having 1H-tetrazole with a maximum error of 0.27 kJ mol^{-1}. This low error is evidence that the azoles can be effectively modelled using the groups that have

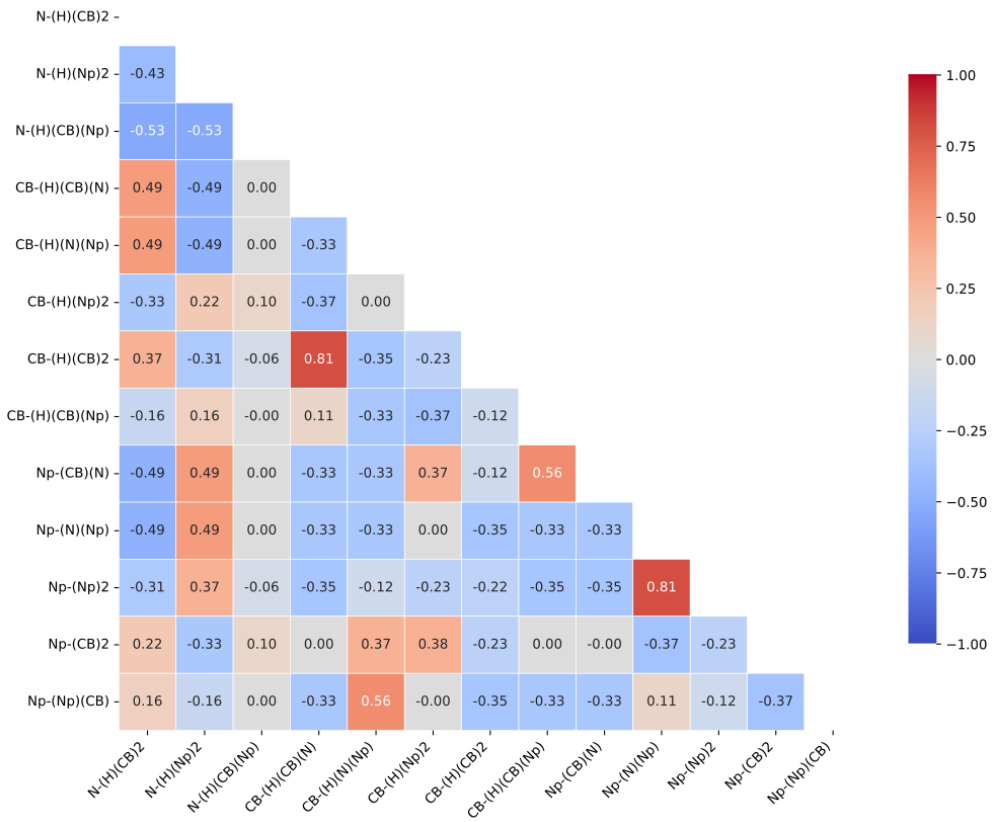


Fig. 3.6. Correlation matrix for groups in the unsubstituted azoles. Blue shading represents negative correlation, while red shading denotes positive correlation.

RS and ASE parameterised into them, without the need to resort to additional RS and ASE corrections.

Following on from this, the GAVs for the unsubstituted azoles were kept fixed for the MLR of the methyl, amine, azide, nitro, and nitramide functionalised azoles. Additionally, the five functional groups on the pyrrole-like nitrogen have been assigned values. Again, the C—$(H)_3$(N), N—$(H)_2$(N), N—$(H)_2$(N), N_3—(N), NO_2—(N), and $NHNO_2$—(N) groups were taken from the previous section. The resulting GAVs are given in Table 3.9. However, whereas the fitting of unsubstituted azoles had a remarkably low error, fitting the substituted azoles had an MAE of 19.33 kJ mol^{-1} and an MAPE of 5.68 %, with a maximum error of 47.21 kJ mol^{-1} for N-1,2,4-triazol-4-ylnitramide.

Table 3.8 Group additivity values for unsubstituted azoles derived from G4(MP2)-6X computed enthalpy of formation data.[a]

Group	GAV
C_B—(H)$(C_B)_2$	13.42
C_B—(H)(C_B)(N)	13.42
C_B—(H)$(C_B)(N_p)$	13.42
C_B—(H)(N)(N_p)	13.42
C_B—$(N_p)_2$	5.30
N—(H)$(C_B)_2$	55.78
N—(H)$(C_B)(N_p)$	85.46
N—(H)$(N_p)_2$	116.75
N_p—$(C_B)_2$	36.73
N_p—(C_B)(N)	52.38
N_p—$(C_B)(N_p)$	67.46
N_p—(N)(N_p)	83.73
N_p—$(N_p)_2$	83.73

[a] All data in $kJ\,mol^{-1}$

Table 3.9 Group additivity values for N-functionalised azoles derived from G4(MP2)-6X calculated enthalpy of formation data by fixing the values in Table 3.8.[a]

Group	GAV
N—(C)$(N_p)_2$	131.64
N—(C)$(C_B)(N_p)$	108.22
N—(C)$(C_B)_2$	88.99
N—$(C_B)_2$(N)	121.57

Continued on next page

Table 3.9 –continued from previous page

Group	GAV
N—(N)$(N_p)_2$	309.35
N—(C_B)(N)(N_p)	176.21
N—$(N_p)_2(N_3)$	208.99
N—$(C_B)(N_p)(N_3)$	172.15
N—$(C_B)_2(N_3)$	150.88
N—$(C_B)_2(NO_2)$	172.60
N—$(N_p)_2(NO_2)$	271.35
N—$(C_B)(N_p)(NO_2)$	218.80
N—$(C_B)_2(NHNO_2)$	160.80
N—$(N_p)_2(NHNO_2)$	223.13
N—$(C_B)(N_p)(NHNO_2)$	179.69

[a] All data in $kJ\,mol^{-1}$

To improve on the model, the full set of GAVs were next regressed using the full set of functionalised and unfunctionalised azoles. The resulting GCM has 33 GAVs of which ten groups are multicollinear. Using the same equivalences and assignments as the previous fitting since the same groups are present, the resulting groups regressed are given in Table 3.10. The estimation of the N-functionalised azoles are given in Table 3.11, and has an MAE of 3.49 $kJ\,mol^{-1}$ and the MAPE is 1.28 %, with a maximum error of 10.95 $kJ\,mol^{-1}$ for 1-nitropyrazole, which is a significant decrease from the previous fitting that had an MAE of 19.33 $kJ\,mol^{-1}$ and an MAPE of 5.68 %. The conclusion made in the previous section, that the ASE cannot simply be carried over to N-functionalized azoles, is again evident in this model. Furthermore, if the influence of the explosophoric group on the ASE was independent of the azole framework, one might have expected that this change could be parameterised into the connecting pyrrole-like groups alone, which does not seem to be the case.

Table 3.10 Group additivity values for unsubstituted, amine, azide, nitro, and nitramide N-functionalised azoles derived from G4(MP2)-6X computed enthalpy of formation data.[a]

Group	GAV
C_B—(H)$(C_B)_2$	13.42
C_B—(H)(C_B)(N)	13.42
C_B—(H)$(C_B)(N_p)$	13.42
C_B—(H)(N)(N_p)	13.42
C_B—$(N_p)_2$	3.70
N—(H)$(C_B)(N_p)$	75.09
N—(H)$(N_p)_2$	101.69
N—(C)$(C_B)_2$	83.17
N—(C)$(C_B)(N_p)$	97.85
N—(C)$(N_p)_2$	116.59
N—$(C_B)_2$(N)	115.75
N—(C_B)(N)(N_p)	123.97
N—(N)$(N_p)_2$	154.75
N—$(C_B)_2(N_3)$	145.06
N—$(C_B)(N_p)(N_3)$	161.78
N—$(N_p)_2(N_3)$	193.93
N—$(C_B)_2(NO_2)$	166.78
N—$(C_B)(N_p)(NO_2)$	208.43
N—$(N_p)_2(NO_2)$	256.30
N—$(C_B)_2(NHNO_2)$	154.98
N—$(C_B)(N_p)(NHNO_2)$	169.32
N—$(N_p)_2(NHNO_2)$	208.07
N_p—$(C_B)_2$	42.44

Continued on next page

Table 3.10 –continued from previous page

Group	GAV
N_p—$(C_B)(N)$	57.83
N_p—$(C_B)(N_p)$	73.33
N_p—$(N)(N_p)$	88.64
N_p—$(N_p)_2$	88.64
C—$(H)_3(N)$	-41.49
N—$(H)_2(N)$	51.65
N_3—(N)	371.31
NO_2—(N)	−39.26
$NHNO_2$—(N)	58.48

[a] All data in $kJ\,mol^{-1}$

Table 3.11 The G4(MP2)-6X computed and group contribution method ($\sum$GAVs) estimated enthalpy of formation of unsubstituted, methyl, amine, azide, nitro, nitramide functionalised azoles, along with corresponding absolute error (AE) and absolute percentage error (APE).[a]

Name	G4(MP2)-6X	$\sum$GAVs	AE	APE
1H-Pyrrole	109.44	103.62	5.82	5.32
1H-Pyrazole	178.08	173.17	4.92	2.76
1H-Imidazole	132.62	132.65	0.02	0.02
1H-1,2,3-Triazole	263.62	263.89	0.27	0.10
1H-1,2,4-Triazole	193.42	192.48	0.94	0.49
2H-1,2,3-Triazole	248.34	244.18	4.16	1.68
4H-1,2,4-Triazole	217.67	223.45	5.78	2.66
1H-Tetrazole	333.53	339.12	5.59	1.67
2H-Tetrazole	325.48	325.19	0.29	0.09
1H-Pentazole	452.80	456.25	4.45	0.98
1-Methylpyrrole	101.85	95.34	6.51	6.39
1-Methylpyrazole	161.02	154.44	6.59	4.09
1-Methylimidazole	123.60	124.37	0.76	0.62
1-Methyl-1,2,3-triazole	244.94	245.16	0.22	0.09
1-Methyl-1,2,4-triazole	175.10	173.75	1.36	0.77
2-Methyl-1,2,3-triazole	226.41	217.59	8.82	3.90
4-Methyl-1,2,4-triazole	209.42	215.17	5.75	2.75
1-Methyltetrazole	312.67	320.39	7.71	2.47
2-Methyltetrazole	299.00	298.60	0.40	0.13
1-Methylpentazole	420.43	429.66	9.23	2.20
Pyrrole-1-amine	224.73	221.06	3.67	1.63
Pyrazol-1-amine	277.09	273.70	3.39	1.23

Continued on next page

Table 3.11 –continued from previous page

Name	G4(MP2)-6X	$\sum$GAVs	AE	APE
Imidazol-1-amine	250.18	250.09	0.10	0.04
Triazole-1-amine	364.14	364.42	0.28	0.08
1,2,4-Triazol-1-amine	292.99	293.01	0.01	0.01
Triazol-2-amine	353.50	348.89	4.61	1.30
1,2,4-Triazol-4-amine	337.13	340.89	3.76	1.12
Tetrazol-1-amine	436.55	439.65	3.09	0.71
Tetrazol-2-amine	429.26	429.90	0.64	0.15
Pentazol-1-amine	556.98	560.96	3.98	0.71
1-Nitropyrrole	172.03	181.18	9.15	5.32
1-Nitropyrazole	256.30	267.25	0.95	4.27
1-Nitroimidazole	209.94	210.21	0.26	0.13
1-Nitrotriazole	359.65	357.97	1.68	0.47
1-Nitro-1,2,4-triazole	285.22	286.56	1.33	0.47
2-Nitrotriazole	348.70	359.53	10.83	3.11
4-Nitro-1,2,4-triazole	310.42	301.01	9.41	3.03
1-Nitrotetrazole	443.78	433.20	10.58	2.39
2-Nitrotetrazole	440.92	440.54	0.38	0.09
3-Nitropentazole	582.04	571.60	10.44	1.79
N-Pyrrol-1-ylnitramide	263.21	267.12	3.91	1.49
N-Pyrazol-1-ylnitramide	320.41	325.88	5.47	1.71
N-Imidazol-1-ylnitramide	296.38	296.15	0.23	0.08
N-Triazol-1-ylnitramide	417.37	416.60	0.77	0.18
N-1,2,4-Triazol-1-ylnitramide	344.27	345.19	0.92	0.27
N-1,2,3-Triazol-2-ylnitramide	403.67	409.04	5.37	1.33

Continued on next page

Table 3.11 –continued from previous page

Name	G4(MP2)-6X	$\sum$GAVs	AE	APE
N-1,2,4-Triazol-4-ylnitramide	390.64	386.95	3.69	0.94
N-Tetrazol-1-ylnitramide	497.45	491.83	5.62	1.13
N-Tetrazol-2-ylnitramide	489.83	490.05	0.22	0.04
N-Pentazol-3-ylnitramide	626.70	621.11	5.59	0.89
1-Azidopyrrole	568.62	570.03	1.41	0.25
1-Azidopyrazole	629.62	621.17	1.54	0.25
1-Azidoimidazole	599.31	599.06	0.26	0.04
1-Azidotriazole	720.70	721.89	1.19	0.17
1-Azido-1,2,4-triazole	650.89	650.48	0.42	0.06
2-Azidotriazole	707.81	707.73	0.08	0.01
4-Azido-1,2,4-triazole	691.02	689.86	1.16	0.17
1-Azidotetrazole	799.44	797.12	2.33	0.29
2-Azidotetrazole	788.06	788.74	0.68	0.09
1-Azidopentazole	920.40	919.80	0.60	0.07

[a] All data in kJ mol^{-1}

Given that the MLR model improved when using the complete fitting set of unsubstituted and functionalised azoles, the effect of functionalisation with implicit ASE and RS was examined. As opposed to regressing the GAVs in the unsubstituted azoles and fixing them in subsequent fittings, this was performed by independently regressing the GAVs for unsubstituted, methyl, amine, azide, nitro, and nitramide functionalised azoles, while addressing multicollinearity in each fit. Since each azole in the fit contains a single pyrrole-like nitrogen bonded to either a hydrogen or a functional group, the frequency of the groups in each MLR model remains unchanged. Only the ligand attached to the pyrrole-like nitrogen varies, allowing the same groups to be assigned values or treated as equivalent across all six fittings to address multicollinearity. Again, the multicollinear-

ity is addressed by making C_B—(H)$(C_B)_2 \equiv C_B$—(H)(C_B)(N) $\equiv C_B$—(H)$(C_B)(N_p) \equiv C_B$—(H)(N)$(N_p) = 13.42\,kJ\,mol^{-1}$, and the N_p—(N)(N_p) is made equivalent to N_p—$(N_p)_2$. Additionally, the C—$(H)_3$(N) group is $-41.49\,kJ\,mol^{-1}$, N—$(H)_2$(N)=$51.65\,kJ\,mol^{-1}$, N_3—(N)=$371.31\,kJ\,mol^{-1}$, NO_2—(N)=$-39.26\,kJ\,mol^{-1}$, and $NHNO_2$—(N)=$58.48\,kJ\,mol^{-1}$. Once the regression is performed for the N-functionalised azoles separately, the GAVs, MAE, MSE, and MAPE are shown in Table 3.12.

Table 3.12 Comparison of the group additivity values for the unsubstituted ($-$H), methyl ($-CH_3$), amine ($-NH_2$), azide ($-N_3$), nitro ($-NO_2$), and nitramide ($-NHNO_2$) functionalised azoles derived from G4(MP2)-6X computed enthalpy of formation Data.[a]

Group	$-$H	$-CH_3$	$-NH_2$	$-N_3$	$-NO_2$	$-NHNO_2$
C_B—(H)$(N_p)_2$	5.30	3.91	3.00	3.91	3.36	3.36
N—$(C_B)_2$(X)	55.78	89.34	119.15	143.33	157.32	150.79
N—$(C_B)(N_p)$(X)	85.46	110.20	130.38	156.57	191.97	161.87
N—$(N_p)_2$(X)	116.75	135.22	164.04	185.07	232.91	197.34
N_p—$(C_B)_2$	36.73	35.91	39.17	44.30	51.71	46.82
N_p—(C_B)(N)	52.38	52.76	55.35	62.14	63.96	60.37
N_p—$(C_B)(N_p)$	67.46	67.34	69.87	75.00	82.88	77.43
N_p—(N)$(N_p) \equiv N_p$—$(N_p)_2$	83.73	81.66	85.38	91.12	97.16	92.80
MAE	0.11	0.32	0.22	0.36	0.25	0.26
MAPE	0.04	0.18	0.07	0.05	0.08	0.07

[a] All data in $kJ\,mol^{-1}$

The data in Table 3.12 is graphically represented in Figure 3.7, which shows the average value and range of the GAVs. Following the independent regression of the azoles, the range of the GAVs with a pyrrole-like nitrogen central atom is relatively large across the fittings due to the presence of the functional groups attached to it. In contrast, the GAVs with a pyridine-like nitrogen central atom have a similar range. If the range of one of the pyridine-like nitrogen GAVs were larger, this would warrant a third-order groups where

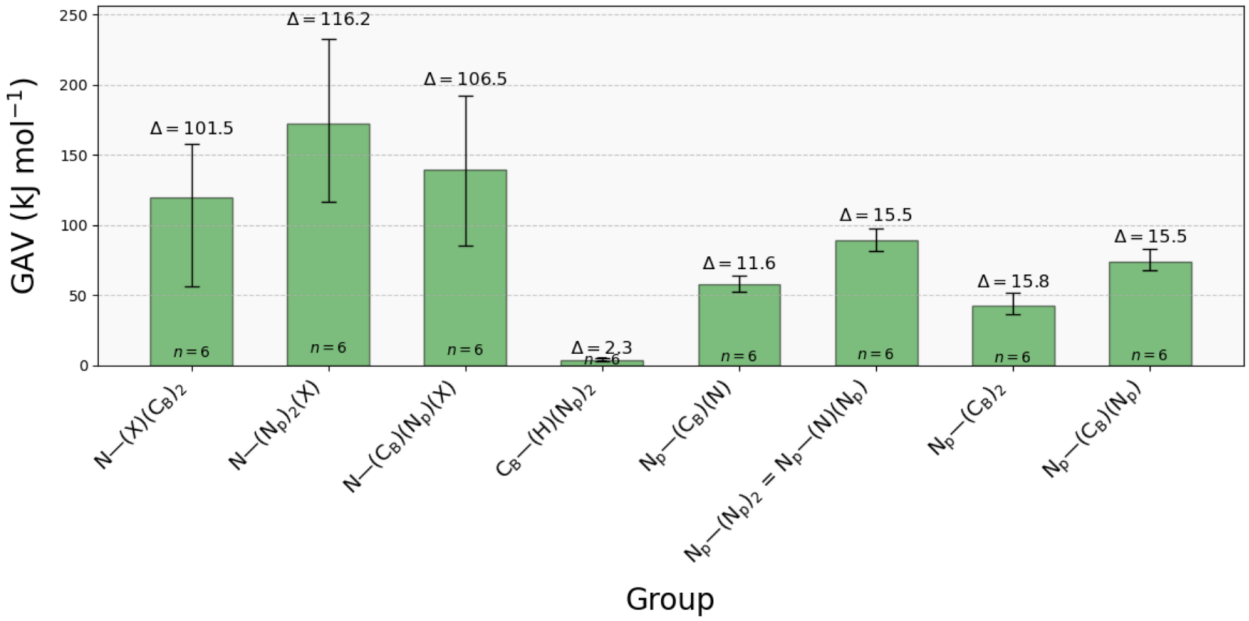


Fig. 3.7. Comparison of the group additivity values derived by separate fitting of unsubstituted, methyl, amine, azide, nitro, and nitramide functionalised azoles where Δ indicates the range of the value of the group.

the explicit consideration of the groups attached to the ligands would also be considered. However, since the pyridine-like nitrogen GAVs have similar values, including a third-order effect is not justified.

3.4 Summary of Findings

Section 3.1 demonstrates that although a systematic approach that uses a minimal set of small molecules aligns well with early thermochemical reasoning and performs accurately for small alkanes, its errors grow with increasing molecular size. In contrast, a regression-based approach, which accounts for a wide range of molecular environments, provides more accurate and reliable GAV determination. The regression model has a low MAE of 0.16 kJ mol^{-1}, reducing systematic errors and providing more robust predictions for longer chains.

In Section 3.2, the first group contribution method for azoles was developed using acyclic reference molecules. After addressing multicollinearity, 42 initial groups were re-

duced to 24 GAVs. For unsubstituted azoles, with an MAE of 3.31 kJ mol^{-1} and an MAPE of 1.70 %. This approach also allowed the estimation of the combined ring strain (RS) and aromatic stablisation energy (ASE) for each azole framework. These RS + ASE values aligned strongly with those obtained from homodesmotic reactions, with a correlation coefficient of 0.96, which confirmed that the approach qualitatively captured the correct aromatic trends. However, when applied to the larger set of 180 N-functionalised azoles, transferring the RS + ASE correlation from the unsubstituted azole to all substituted systems resulted in comparatively poor performance. The resulting model had an MAE of 18.05 kJ mol^{-1} and an MAPE of 4.69 %. This indicates that the framework-based RS + ASE correction, which was calibrated only on unsubstituted azoles, is not transferable to substituted systems.

Figure 3.8 evaluates the effect of including the RS + ASE term in the GCM. Incorporating the RS + ASE correction term lowers the predicted enthalpies of formation, resulting in systematically lower enthalpies of formation than the uncorrected values. The GCM estimated values without a RS + ASE correction displays larger deviations from the line of parity. In contrast, incorporating a RS + ASE term produces points that cluster more tightly around the line of parity. Figure 3.9 compares the MAEs for the functional groups, calculated over all ten substituted azole in every class, using acyclic groups and the RS + ASE correction. The error increases in the order $-CH_2 < -NH_2 < -N_3 \approx -NO_2 < -NHNO_2$. This trend aligns with established electronic behaviour: methyl is a weak σ-donor, while amino is a stronger π-donor that perturbs the ring more noticeably. Azide and nitro groups act as strong π-acceptors and significantly alter the π-electron distribution. Nitramide substituents are even more strongly electron-withdrawing and highly conjugated, and structural and electronic studies show that they couple very strongly to the ring while preserving planarity and ring currents,[101] which matches the larger aromaticity changes observed.

Section 3.3 developed a second GCM that incorporated the effects of aromaticity and ring strain directly into the GAVs, rather than relying on a separate correction term.

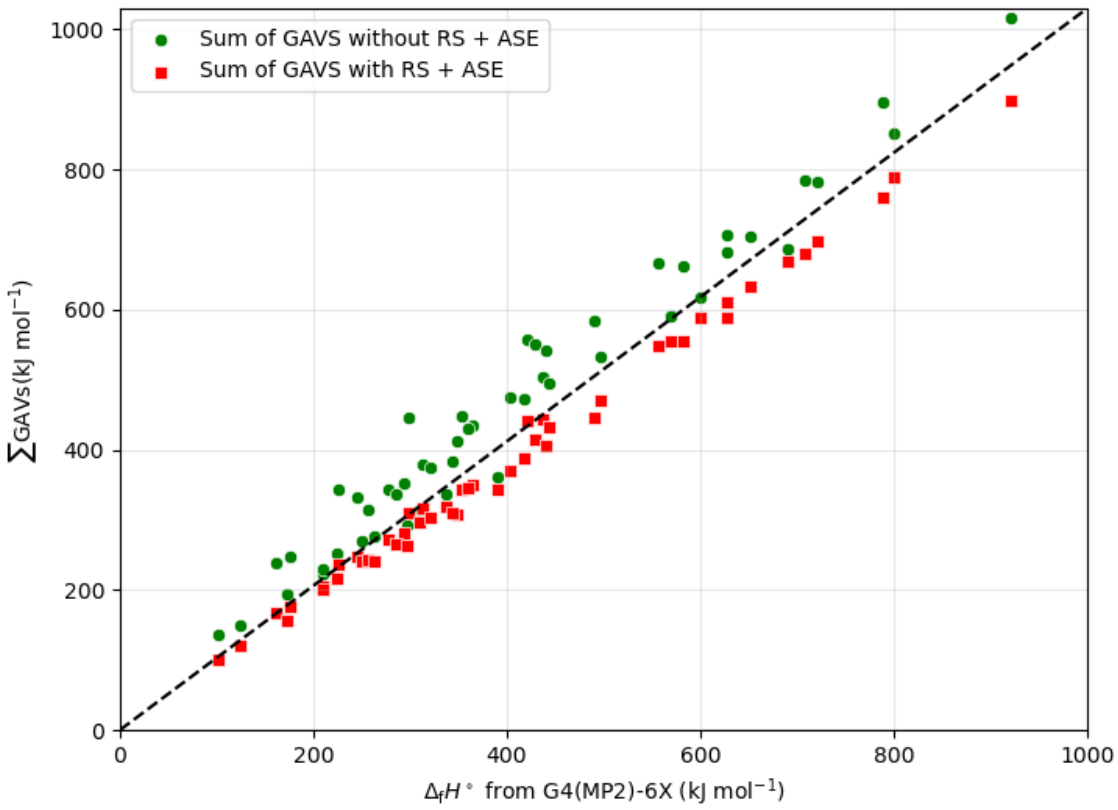


Fig. 3.8. Enthalpies of formation obtained from the GCM based on acyclic groups ($\sum$ GAVs), both with and without the ring strain (RS) and aromatic stablisation energy (ASE) term, compared to G4(MP2)-6X computed values of functionalised azoles. A line of parity is plotted for comparison.

When this model was fitted to the unsubstituted azoles alone, it performed extremely well, with an MAE of 0.11 kJ mol^{-1} and an MAPE of 0.04 %. When extended to functionalised azoles, fixing the unsubstituted GAVs caused errors to increase significantly (MAE of 19.33 kJ mol^{-1} and an MAPE of 5.68 %). However, refitting the unsubstituted and substituted azoles simultaneously produced much better accuracy (MAE of 3.49 kJ mol^{-1} and an MAPE of 1.28 %).

Taken together, the findings across the three sections show a clear progression: methods that treat aromaticity and ring strain as a separate contribution, and rely on fixed GAV-based corrections derived for unsubstituted rings do not generalise well to function-

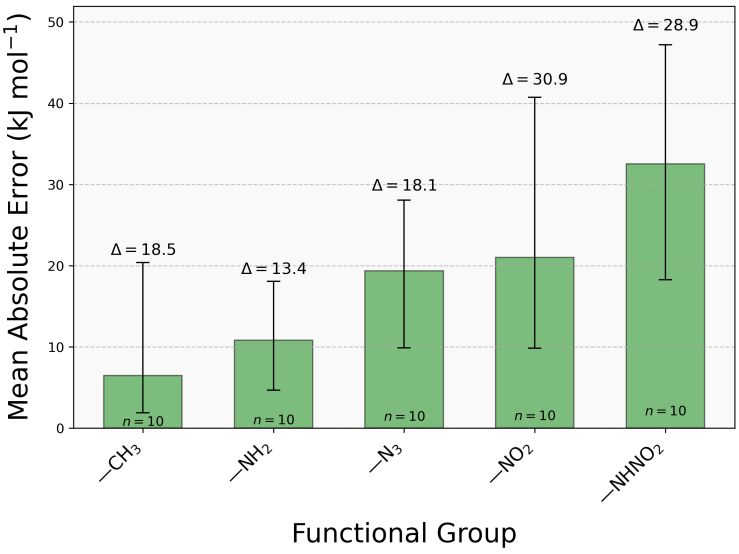


Fig. 3.9. Comparison of the MAE for the ten azoles associated with each functional group, where Δ indicates the range of the absolute error of the group.

alised systems. In contrast, the fully aromatic regression model in Section 3.3 is the most accurate and transferable, as it fits the entire dataset in a single regression and embeds aromatic effects directly into the GAVs, resulting in substantially lower errors across diverse chemical environments (Figure 3.10).

The first GCM remains advantageous in its ease of extension, since only new functional groups and their connecting ring atoms require parametrisation. However, this also means that the full effect of functionalisation must be carried solely by those added groups. The second GCM, although requiring refitting of the full dataset for each new functional group, distributes the impact of substitution across all GAVs and can therefore be more inherently transferable.

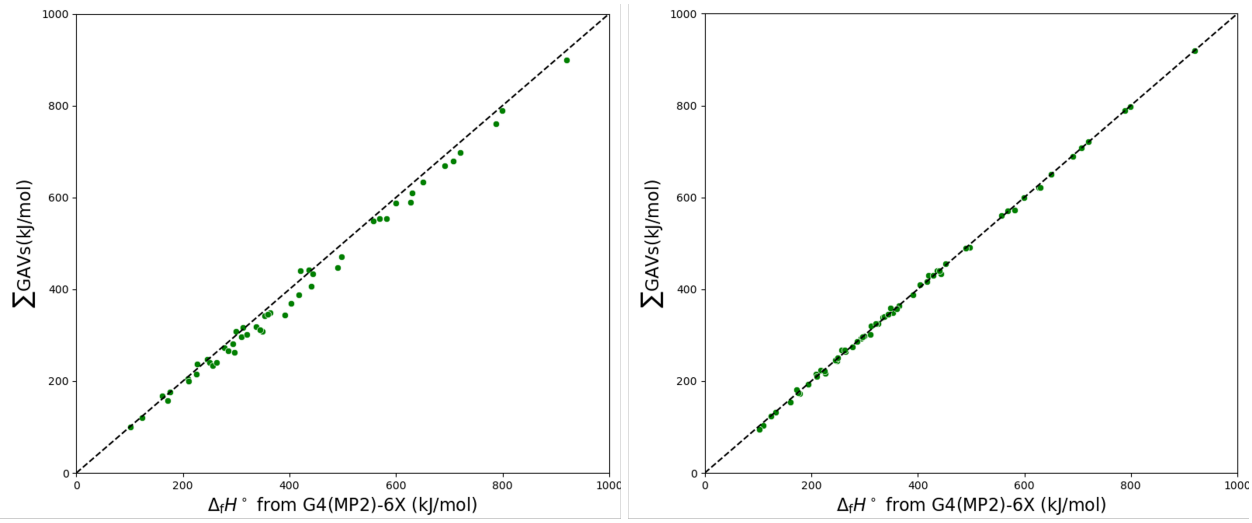


Fig. 3.10. Enthalpies of formation obtained from the GCM ($\sum$ GAVs) parametrised using (a) acyclic groups and (b) aromatic groups compared to G4(MP2)-6X computed values for functionalised azoles. A line of parity is plotted for comparison.

4 Computational Thermochemistry

This chapter starts by outlining the fundamental principles in thermochemistry with a focus on the enthalpy of formation. It then explores the methods for calculating thermodynamic quantities by exploring the atomisation method and homodesmotic reactions. The framework for validating thermochemical data is discussed next by explaining the role of Active Thermochemical Tables. Finally, the chapter introduces linear regression to estimate the enthalpy of formation.

4.1 Basic Thermochemistry

The first law of thermodynamics asserts that the internal energy (U) of an isolated system is constant. That is, when a change in U occurs to a system that is in contact with its surroundings, then the total energy change (ΔU_{total}) is given by,

$$\Delta U_{\text{total}} = \Delta U_{\text{system}} + \Delta U_{\text{surroundings}} = 0, \tag{4.1}$$

which can also be simplified to,

$$\Delta U_{\text{system}} = -\Delta U_{\text{surroundings}}, \tag{4.2}$$

where ΔU_{system} and $\Delta U_{\text{surroundings}}$ are the change in energy of the system and surroundings, respectively. In a closed system, in which phase changes or chemical reactions take place, the internal energy of the system is a result of the flow of heat, work, or a combination of both, across the boundary between the system and its surroundings. This leads to a second expression of the first law,

$$\Delta U = q + w, \tag{4.3}$$

where q is heat, w is work, and ΔU is the change in internal energy of the system. Work is the energy exchange across the boundary between the system and its surroundings that results from a force (**F**) that moves an object through a distance (**x**),

$$w = \int_{x_i}^{x_f} \mathbf{F} \cdot d\mathbf{x}. \tag{4.4}$$

Work can also be expressed using the definition of pressure as the force per unit area (*A*),

$$w = -\int_{x_i}^{x_f} P_{\text{external}} A dx = -\int_{V_i}^{V_f} P_{\text{external}} dV, \tag{4.5}$$

where $P_{external}$ is an external pressure on the system, V_i is the initial volume, and V_f is the final volume. Heat refers to the exchange of energy across the boundary between the system and its surroundings due to a temperature difference. The change in internal energy under constant volume conditions is given by,

$$\Delta U = q_V, \tag{4.6}$$

since $w = -\int PdV = 0$. Therefore, ΔU can be experimentally measured from the heat flow between the system and its surroundings under constant volume. Generally, chemical reactions are performed under constant pressure, rather than constant volume, and consequently, the change in internal energy under constant pressure conditions is given by,

$$dU = \not{d}q_P - PdV. \tag{4.7}$$

By integrating Equation 4.7 between the initial and final states,

$$U_f - U_i = q_P - (P_f V_f - P_i V_i), \tag{4.8}$$

which is rearranged to obtain,

$$(U_f + P_f V_f) - (U_i + P_i V_i) = q_P, \tag{4.9}$$

where U, P, and V are state functions. Therefore, $U + PV$ is a state function. This new state function is enthalpy, denoted by H, and is defined as,

$$H \equiv U + PV. \tag{4.10}$$

Enthalpy is also be referred to as the "heat content" of a system; however, this term is ambiguous since heat is a process rather than a characteristic of a system.[102] The enthalpy change, ΔH, for a process that only involves PV work is quantified by measuring the heat exchange between a system and its surroundings under constant pressure,

$$\Delta H = q_P. \tag{4.11}$$

4.2 Direct Methods for Calculating Thermodynamic Quantities

In the field of computational thermochemistry, the "holy grail" is to arrive at chemical accuracy, which is generally considered to be $1\,\text{kcal}\,\text{mol}^{-1}$ or $4\,\text{kJ}\,\text{mol}^{-1}$.[81] Computational methods are generally the most practical route to compute thermochemical quantities within chemical accuracy. However, these robust methods are very computationally demanding and are only viable for small molecules. An alternative approach to predicting thermochemical quantities with high accuracy is the use of systematic cancellation of errors. In contrast to error-cancelling reactions, there has been work to compute thermochemical properties within "subchemical accuracy", which is approximately $0.1\,\text{kcal}\,\text{mol}^{-1}$ or $0.4\,\text{kJ}\,\text{mol}^{-1}$.[88]

The standard enthalpy of formation, $\Delta_f H^\ominus$, is a measure of the enthalpy change that accompanies the formation of one mole of a species from its chemical elements in their reference states. An elements reference state is defined as its most thermodynamically stable form at the standard state conditions of 298.15 K and 1 bar. The direct computation of $\Delta_f H^\ominus$ is not possible because the standard state of some elements (e.g. carbon) cannot be calculated with standard electronic structure methods. To approach $\Delta_f H^\ominus$ with subchemical accuracy favours the atomization method.

4.2.1 Atomization Method

The atomization method is a thermodynamic cycle that connects the target species to the constituent elements in their standard state. The approach is as follows. Calculate $\Delta_f H_0^\ominus$ and then correct to 298.15 K to get $\Delta_f H_{298.15}^\ominus$, which will be shown by example using 1H-imidazole (ImH), which has the molecular formula $C_3H_4N_2$, as shown in Figure 4.1. At 0 K, ImH is atomized into C(^{3}P), H(^{2}S), and N(^{4}S) in their ground electronic states. Moreover, ImH can be made from the elements in their reference states; namely, graphite (C(s)), molecular hydrogen (H_2(g)), and molecular nitrogen (N_2(g)). The 0 K enthalpy of

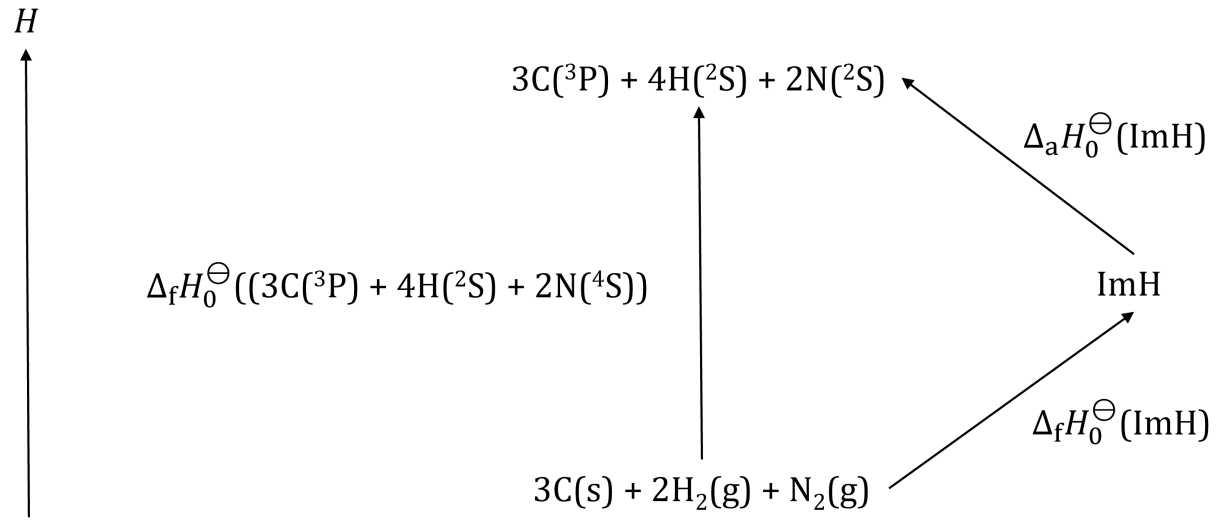


Fig. 4.1. The basis of the atomization method to obtain the *ab initio* enthalpy of formation of 1H-imidazole.

formation of ImH, $\Delta_f H_0^\circ(\text{ImH})$, follows from equating the enthalpy required to generate the atoms from their reference states via ImH (indicated by the two arrows on the right of Figure 4.1) to the enthalpy required to generate the atoms directly from the elements in their reference states (indicated by the arrow on the left), as shown by,

$$\Delta_f H_0^\circ(\text{ImH}) + \Delta_a H_0^\circ(\text{ImH}) = \Delta_f H_0^\circ(3\text{C}(^3\text{P}) + 4\text{H}(^2\text{S}) + 2\text{N}(^4\text{S}))$$

i.e.

$$\Delta_f H_0^\circ(\text{ImH}) = \Delta_f H_0^\circ(3\text{C}(^3\text{P}) + 4\text{H}(^2\text{S}) + 2\text{N}(^4\text{S})) - \Delta_a H_0^\circ(\text{ImH}). \quad (4.12)$$

The *ab initio* atomization enthalpy of ImH is $\Delta_a H_0^\circ(\text{ImH})$, which represents the enthalpy difference between the atoms in their ground states and ImH. The 0 K enthalpy of formation of C, H, and N atoms, i.e., the atomization enthalpy of C(graphite), H_2(g), and N_2(g) denoted $\Delta_f H^\circ(\text{C}(^3P))$, $\Delta_f H^\circ(\text{H}(^2S))$, and $\Delta_f H^\circ(\text{N}(^4S))$ cannot be determined via *ab initio* calculations due to the number of unpaired electrons of C(s). Therefore, values from the Active Thermochemical Tables, or ATcT (see Section 4.3), are used to maintain consistency and are shown in Table 4.1. The $\Delta_a H_0^\circ(\text{ImH})$ is calculated as,

$$\Delta_a H_0^\circ(\text{ImH}) = \Delta U_0(\text{C}(^3\text{P}) + \text{H}(^2\text{S}) + \text{N}(^4\text{N})) - \Delta U_0(\text{ImH}), \quad (4.13)$$

Table 4.1. Experimental enthalpy of formation ($\Delta_f H_{298.15}^\circ$) of atoms in ground electronic states.[a]

Element	$\Delta_f H_{298.15}^\circ$
H	216.03
C	711.39
N	470.58
O	246.84

[a] Enthalpy of formation in kJ mol^{-1}. Numbers are obtained from the Active Thermochemical Tables.[103]

where ΔU_0(C(^{3}P), H(^{2}S), and N(^{4}N)) is the *ab initio* enthalpy of formation for C, H, and N atoms in their electronic states and ΔU_0(ImH) is the ZPE-corrected *ab initio* enthalpy of formation of ImH. These values were calculated using the G4(MP2)-6X *ab initio* method and the results are given in Table 4.2.

Table 4.2. Atomization enthalpy ($\Delta_a H_0^\ominus$(ImH(g)) of atoms in ground electronic states.[a]

Element	$\Delta_f H_0^\ominus$
H	-0.50
C	-37.80
N	-54.54
O	-75.01

[a] Atomization enthalpy in kJ mol^{-1}. Numbers are calculated using G4(MP2)-6X.

To correct $\Delta_f H_0^\ominus$(ImH) to the enthalpy of formation at 298.15 K, $\Delta_f H_{298.15}^\ominus$(ImH), the increase in enthalpy of ImH from 0 K to 298.15 K, $\Delta\Delta H^\ominus$(ImH), is added and the corresponding increases for the elements in their reference states is subtracted,

$$\begin{aligned}\Delta_f H_{298.15}^\ominus(\text{ImH}) = \Delta_f H_0^\ominus(\text{ImH}) + \Delta\Delta H^\ominus\text{ImH} - (3\Delta\Delta H^\ominus(\text{C}) \\ + 2\Delta\Delta H^\ominus(\text{H}_2) + \Delta\Delta H^\ominus(\text{N}_2)),\end{aligned} \tag{4.14}$$

where $\Delta\Delta H^\ominus(C)$, $\Delta\Delta H^\ominus$(H$_2$), and $\Delta\Delta H^\ominus$(N$_2$) are the experimental increases in enthalpy from 0 K to 298.15 K for C(s), H$_2$(g), and N$_2$(g), respectively, and are given in Table 4.3. The value for $\Delta\Delta H^\ominus$ImH is the difference between G4(MP2)-6X calculated $\Delta_f H_{298.15}^\ominus$ and $\Delta_f H_0^\ominus$ provided in the thermochemical summary of the G4(MP2)-6X calculation.

Alternatively, $\Delta_f H_{298.15}^\ominus$ can be calculated directly since the 0 K *ab initio* enthalpy of ImH in Equation 4.12 and Equation 4.14 can be subtracted out, and it follows that,

$$\begin{aligned}\Delta_f H_{298.15}^\ominus(\text{ImH}) = [\Delta_f H_0^\ominus(\text{C}) + \Delta_f H_0^\ominus(\text{H}) + \Delta_f H_0^\ominus(\text{O})] \\ - [\Delta U_0(\text{C}) + \Delta U_0(\text{H}) + \Delta U_0(\text{N})] \\ - [\Delta\Delta H^\ominus(\text{C}) + \Delta\Delta H^\ominus(\text{H}_2)]\end{aligned} \tag{4.15}$$

Table 4.3. Experimental enthalpy correction ($\Delta\Delta H^{\ominus}$) of elements.[a]

Atom	$\Delta\Delta H^{\ominus}$
H_2	8.467
C(s)	1.051
N_2	8.670
O_2	8.68

[a] Enthalpy increases in kJ mol^{-1}. Numbers are obtained from NIST-JANAF thermochemical tables.[94]

This more direct calculation of $\Delta_f H^{\ominus}_{298.15}$ will be implemented.

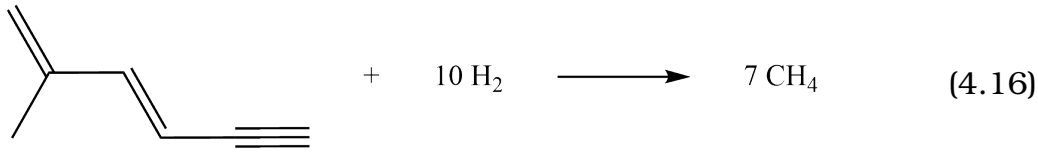
4.2.2 Homodesmotic Reactions

Chemical concepts such as ring strain, π-conjugation, aromaticity, and hyperconjugation are not directly measurable.[88] Aromaticity was introduced to account for the low reactivity, unusually high electronic stability, and structure of benzene and benzenoid derivatives, which later became known as the "benzene problem".[104] Erich Hückel explained the benzene problem, based on MO theory, by introducing the Hückel rule whereby conjugated and monocyclic planar organic compounds that have (4n+2) π-electrons ($n = 0, 1, 2, 3, \cdots$) display relatively high stability and aromatic character.[105]

An alternative to computationally expensive ab initio methods to calculate the aromatic stabilization energy (ASE) and the ring strain of N-heterocycles, is the use of error balanced reactions. Error-balanced reactions use the systematic cancellation of errors in the predicted reaction enthalpies were introduced before major advancements in computational hardware, wider accessibility to correlated *ab initio* methods, and the full development of density functional methods (DFT) to accurately predict thermochemical properties of molecules, irrespective of their size.

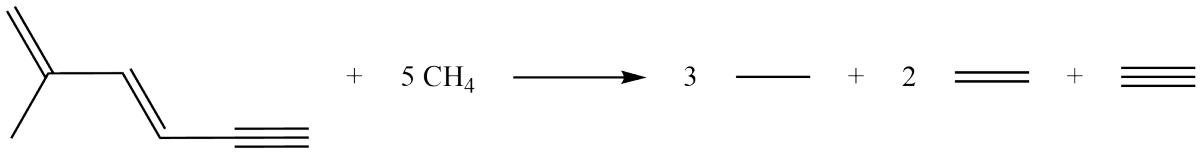
Isogyric reactions (IG) conserve the number total number of electron pairs in the reactants and products. For example, Equation 4.16 has zero unpaired electrons in the

reactants and products.



(4.16)

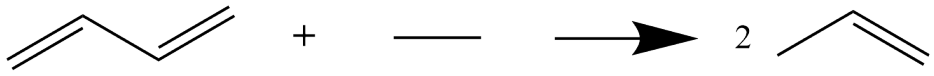
A subset of isogyric reactions for which the number of each type of bond (single, double, and triple) between two heavy atoms is conserved in the products and reactants, is defined as isodesmic reactions (ID). In Equation 4.17, there are an equal number of C−C, C=C, and C≡C bonds in the reactants and products.



(4.17)

To predict stabilisation energy in cyclic conjugated hydrocarbons, George et al.[106] proposed homodesmotic reactions (HD1), a subset of isodesmic reactions. Homodesmotic reactions were originally defined as reaction that have an equal number of heavy atoms in their states of hybridisation in the reactants and products, and the number of bonds between a heavy atom and hydrogen is conserved in the products and reactants. However, this definition is not applicable for systems that contain a combination of sp^3, sp^2, and sp atoms and further refinement of HD1 was necessary. George et al.[107] provided a new definition for homodesmotic reactions (HD2) where the type and number of bonds between heavy atoms (e.g., $C_{sp^3}-N_{sp^3}$, $C_{sp^3}-N_{sp^2}$, $C_{sp^2}=N_{sp^2}$, $C_{sp}=N_{sp^2}$, or $C_{sp}=N_{sp}$) is conserved, and there are an equal number of each type of heavy atoms (sp^3, sp^2, and sp) with zero, one, two, and three hydrogens attached in the reactants and products.[88] The benzene transformation shown in Equation 4.18 is an example of a reaction that satisfies the criteria for HD1 and not HD2; however, the second criterion of HD2 encompasses both

requirements for an HD1 reaction.



(4.18)

Hess and Schaad[108] released a study of the resonance energies of benzene and cyclobutadiene in which they reported that homodesmotic reactions, based on definition HD1, were insufficient in their prediction of the necessary key interactions, and went on to define hyperhomodesmotic reactions. Hyperhomodesmotic reactions were defined as reactions in which there are eight carbon-carbon bond types ($H_2C{=}CH$, $HC{=}CH$, $H_2C{=}C$, $HC{=}C$, $C{=}C$, $HC{-}CH$, $HC{-}C$, and $C{-}C$) that are conserved in the products and reactants. In addition, the $H_3C{-}CH_2$, $H_3C{-}CH$, $H_2C{-}CH_2$, $H_3C{-}C$, $H_2C{-}CH$, $H_2C{-}C$, $HC{\equiv}C$, and $C{\equiv}C$ bond types can be added to include all hydrocarbons. However, these transformations originate from and satisfy the HD1 definition, but do not meet the requirements for HD2.

In literature, HD1 and HD2 have been used interchangeably, prompting Wheeler et al.[88] to publish a hierarchy of homodesmotic reactions. In order to resolve the confusion between HD1 and HD2, Wheeler et al.[88] was proposed that reactions fulfilling the criteria for HD1 be termed hypohomodesmotic reactions, while homodesmotic refers to reactions that satisfy the definition for HD2. Additionally, hyperhomodesmotic reactions were redefined as a subset of homodesmotic reactions such that, (a) there are an equal number of heavy atom types (e.g., $H_3C{-}CH_3$, $H_3C{-}CH$) and (b) an equal number of carbon atoms in their hybridized states with zero, one, two, and three hydrogen atoms attached. A consistent hierarchy of reactions classes can now be assembled: isogyric $\supseteq$ isodesmic $\supseteq$ hypohomodesmotic $\supseteq$ homodesmotic $\supseteq$ hyperhomodesmotic.

4.3 Active Thermochemical Tables

The availability of high-quality thermochemical values for a wide range of chemical species is central to many areas of chemistry, vital in many industries, and can be used as benchmark values in the development and assessment of quantum chemistry methods. Thermochemical tables are tabulations of thermochemical properties, arranged by chemical species, in which the enthalpy of formation is the central type of thermochemical property listed, and is regularly accompanied by the Gibbs energy of formation, entropy, heat capacity, and enthalpy increment.[103] The listed properties are most frequently reported at the reference temperature of 298.15 K, occasionally also be reported at 0 K, and some tables report at a wide range of temperatures.

Since the listed thermochemical properties are derived from basic determinations, they and can be classified as either species–specific or species–interrelating.[109] Species–specific information accounts for determinations that relate to the property of one species and are independent of the properties of any other species. Species–interrelating information accounts for determinations that relate to a property relative to one or more species. Thermochemical quantities that relate to the partition function; such as, entropy, heat capacity, and enthalpy increment are species–specific determinations while thermodynamic quantities such as enthalpy of formation, atomization enthalpy, bond dissociation energy, and enthalpy of chemical reactions are species–interrelating determinations. The traditional approach to thermochemical tables provides a change in enthalpy or energy for the overall reaction, which by definition requires the involvement of two or more chemical species. The complications arising from the multiple relationships as a result of using species–interrelating data can be resolved using the sequential approach. The sequential approach consists of consecutive steps that only centre on one target chemical species using the available "best" species-interrelating data, according to a critical evaluation process by an evaluator, that have been established during prior steps. However, this method only works if the best determinations are known, which restricts the number of available species–

interrelating determinations. While using the best species–interrelating determinations may improve the value of the property of the target species, it introduces inconsistencies due to the number of progenitor-progeny dependencies, of which it is not always clear which species needs to be updated. To avoid the interdependencies associated with the sequential approach, a Thermochemical Network (TN) approach can be used.

Active Thermochemical Tables (ATcT), a recently developed software suite, are a novel approach for deriving reliable, accurate, and internally consistent thermochemical quantities, which constructs, analyses, corrects, and solves a Thermochemical Network (TN).[96,109] Figure 4.2 is a graphical representation of a simple TN in which the primary vertices, depicted as rectangles, represent the enthalpy of formation of chemical species that need to be determined while secondary vertices, depicted as ovals, represent the chemical reactions that have available relevant measurements. Secondary vertices may have multiple degeneracies, represented by the number and letter combinations, which indicate competing species-interrelating determinations of the same chemical reaction at a certain temperature. The graph edges, depicted as arrows, are directed to indicate participation of the chemical species in certain chemical reactions and are generally weighted to reflect stoichiometry, the latter of which is not explicitly shown in Figure 4.2. The graph edges always connect a primary and secondary vertex, all first neighbours are of the same type while all second neighbours are of the same type. In Figure 4.2, there are 10 primary vertices, 13 secondary vertices, and a sum degeneracy of 34, which corresponds to the chemical species, chemical reactions and number of determinations, respectively. In comparison, the Core (Argonne) Thermochemical Network, which is the TN maintained in the central library of the ATcT database, contains >600 primary vertices and >3200 secondary vertices.[103]

There are two variants of TN: ab ovo TN, also referred to as a global TN, and a local TN. Apart from reference elements in standard states, primary vertices are treated as unknowns

in a global TN; whereas, in a local TN the primary vertices are set to pre–selected values and removed using information gathered during prior considerations.

On a visual examination of Figure 4.2, between any two arbitrarily chosen primary vertices, there are multiple allowed paths. The sequential approach consists of a subset of paths, follows sequential steps by selecting the "best" path. This trivial graph corresponds to a linear equation with one unknown that is solved and produces the value of the primary vertex; however, it is statistically better to consider all possible paths rather than arbitrarily choosing one. This can be achieved by solving for the simultaneous solution of the whole system, conventionally via the minimisation of χ^2, but this is reliant on whether the associated uncertainties represent the species-interrelating determination in the TN. Statistical analysis to detect and correct any "optimistic" uncertainties is necessary to find the simultaneous solution that is weighted in space. The competing determinations and multiple alternative pathways of a TN creates redundancy of information that can be used to isolate "offender".

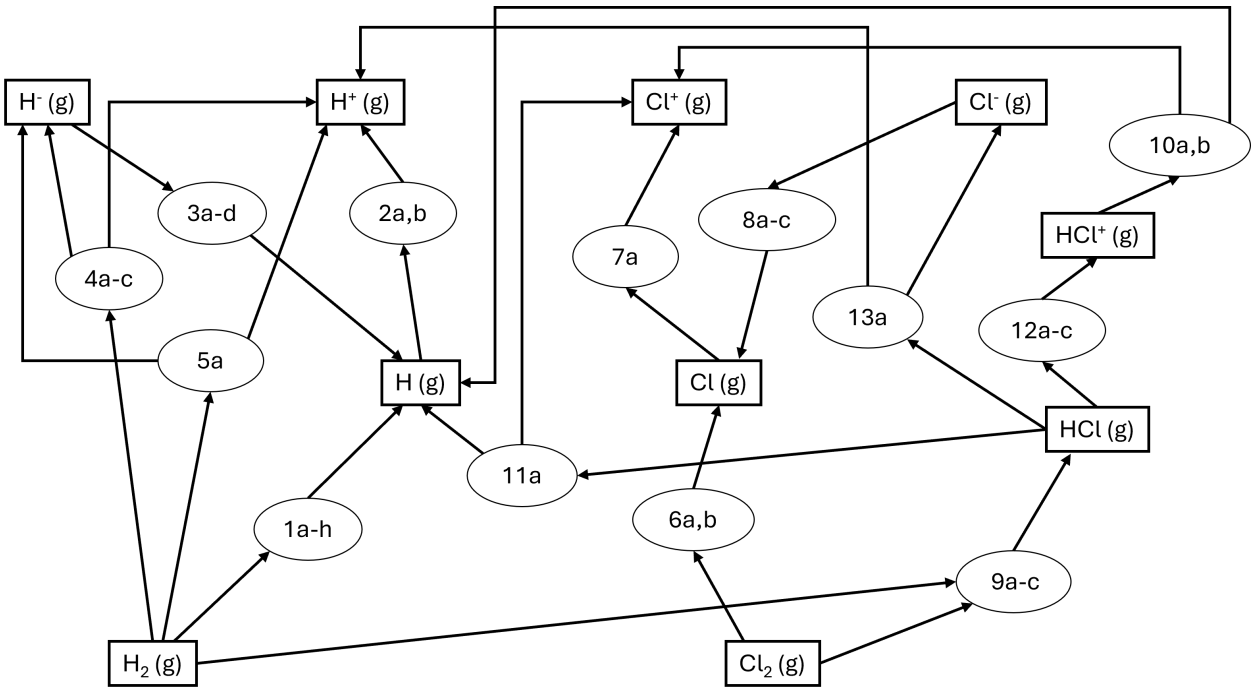


Fig. 4.2. Graphical representation of a simple thermochemcial table where primary vertices are depicted as rectangles and secondary vertices are depicted as ovals.

4.4 Linear regression

A linear regression model computes a weighted sum of input features plus an intercept term,

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n, \tag{4.19}$$

where $\hat{y}$ is the predicted value, n is the number of features, β_j is the j^{th} model parameter, which includes the intercept term β_0 and the feature weights $\beta_1, \beta_2, \ldots, \beta n$. This equation can be more succinctly written using the vectorized form,

$$\hat{y} = h_{\boldsymbol{\beta}}(\mathbf{x}) = \boldsymbol{\beta} \cdot \mathbf{x}, \tag{4.20}$$

where $\boldsymbol{\beta}$ is the parameter vector for the model that contains the feature weights $\beta_1, \ldots, \beta_n$ and the intercept term β_0, $\mathbf{x}$ is the feature vector that contains $x_0 \ldots, x_n$, and h_β is the hypothesis function. Training the linear regression model entails computing the parameters that best fit the model to the training set; however, this requires a measure of the performance to fit the training data.

4.4.1 Model Evaluation

The simplest scoring function to evaluate model performance is the mean absolute error (MAE). The MAE of the linear regression model's predictive function, h_β, is

$$\text{MAE}(\mathbf{X}, h_\beta) = \frac{1}{N} \sum_{i=1}^{N} |h_\beta \mathbf{x}^{(i)} - y^i|, \tag{4.21}$$

where $\mathbf{X}$ is the training set matrix, N is the number of instances in the dataset, $\mathbf{x}^{(i)}$ is the vector that contains the feature value of the i^{th} instance in the dataset, and y^i is the target value.

Another performance measure of a regression model is the mean square error (MSE),

$$\text{MSE}(\mathbf{X}, h_\beta) = \frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{\beta}^T\mathbf{x}^{(i)} - y^{(i)}\right)^2, \tag{4.22}$$

where $\boldsymbol{\beta}$ is the transpose of the model's parameter vector.

4.4.2 The Normal Equation

With a measure of performance outlined, the linear regression model is trained by computing a value of $\boldsymbol{\beta}$ that minimises the MAE. The value of $\boldsymbol{\beta}$ that minimises the cost function is given by the mathematical equation that gives the direct result (i.e., a closed-form solution),

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}, \tag{4.23}$$

where $\hat{\boldsymbol{\beta}}$ is the value of $\boldsymbol{\beta}$ that minimizes the cost function, $\mathbf{X}^T$ is the transpose of $\mathbf{X}$, and $\mathbf{y}$ is the vector of target values.

4.4.3 Multicollinearity

Exact collinearity describes a perfect linear relationship between two variables.[98] A high degree of linear intercorrelation between more than two variables in a regression model is described as multicollinearity. The coefficient of determination, R^2,

$$R^2 = \frac{\sum_i(\mathbf{y} - \hat{\mathbf{y}}_i)^2}{\sum_i(\mathbf{y}_i - \bar{y})^2}, \tag{4.24}$$

where $\hat{\mathbf{y}}_i$ is a vector of the predicted values and $\bar{y}$ is the mean of the observed data, such that the numerator represents the residual sum of squares and the denominator represents the total sum of squares of the data. A value of $R^2 = 0$ represents no multicollinearity between variables, while $R^2 = 1$ denotes an exact multicollinearity between them.

5 Computional Methodology

This chapter gives an overview of electronic structure methods used in computational quantum chemistry. The G4(MP2)-6X *ab initio* composite method (see Section 5.4) was predominantly used in this work; however, this method consists of computations using Density Functional Theory (DFT) and post-Hartree-Fock (HF) methods such as Møller-Plesset (MP) perturbation theory and Coupled-Cluster (CC) theory. Hence, these methods are first described below, starting with HF theory. Citations of original work can be found in the textbooks used to compile this section, which was *Introduction to Computational Chemistry* by Frank Jensen and *Essentials of Computational Chemistry* by Christopher Cramer.[110,111]

5.1 Hartree-Fock Theory

The Hartree-Fock (HF) method is an approximate attempt to computationally solve Schrödinger's equation for a multi-electron system by reducing the problem to a series of one-electron problems. The Schrödinger equation characterises the change of a quantum mechanical system over time; however, time is not a variable for molecular systems and the general form of the time-independent Shrödinger equation is given by,

$$\hat{H}\Psi = E\Psi, \tag{5.1}$$

where the Hamiltonian operator $\hat{H}$ is the sum of the potential and kinetic energy operators, E is the energy of the system, and Ψ is the wavefunction. Since nuclei are much heaver than electrons, the Born-Oppenheimer approximation can be used, which assumes that the nuclear positions can be frozen and the energy and wavefunction for the molecular system can be solved using only the coordinates of the electrons. Applying the Born-Oppenheimer

approximation to the Schrödinger equation results in,

$$\hat{H}_{\mathrm{e}}\Psi_{\mathrm{e}}(\mathbf{r}) = E_{\mathrm{e}}\Psi_{\mathrm{e}}(\mathbf{r}), \tag{5.2}$$

where **r** denotes the electronic coordinates and $\hat{H}_e$ is the electronic Hamiltonian operator for the time-independent, non-relativistic Schrödinger equation given by,

$$\hat{H}_{\mathrm{e}} = -\frac{1}{2}\sum_i \nabla_i^2 - \sum_{A,i}\frac{Z_A}{r_{iA}} + \sum_{i<j}\frac{1}{r_{ij}}, \tag{5.3}$$

∇_i^2 denotes the Laplacian operator for electron i, Z_A is the charge of nucleus A, r_{iA} is the distance between electron i and nucleus A, and r_{ij} is the distance between electrons i and j. The Laplacian operator applied to the cartesian coordinates of electron i is given by,

$$\nabla_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}. \tag{5.4}$$

In order to solve Equation 5.2, each electron is described by a one-electron spin orbital and the total N-electron wavefunction is constructed from a linear combination of these to form N molecular orbitals (MOs), where each spin orbital is constructed from a spatial orbital and spin function. However, to obey the antisymmetry principle, which states that the exchange of any two electrons' coordinates must be associated with a change in sign of the wavefunction, the N-electron wavefunction (Φ) must be expressed as a Slater determinant as shown below,

$$\Phi_{\mathrm{SD}} = \frac{1}{\sqrt{N!}}\begin{vmatrix} \psi_1(\mathbf{r}_1) & \psi_1(\mathbf{r}_2) & \dots & \psi_1(\mathbf{r}_N) \\ \psi_2(\mathbf{r}_1) & \psi_2(\mathbf{r}_2) & \dots & \psi_2(\mathbf{r}_N) \\ \vdots & \vdots & \dots & \vdots \\ \psi_N(\mathbf{r}_1) & \psi_N(\mathbf{r}_2) & \dots & \psi_N(\mathbf{r}_N) \end{vmatrix} \tag{5.5}$$

where ψ_i is the i^{th} MO. The energy of the Slater determinant wavefunction can be derived by applying the Hamiltonian operator to Equation 5.5 giving,

$$E = \sum_{i=1} h_i + \frac{1}{2} \sum_{i=1} \sum_{j>i} (J_{ij} - K_{ij}), \tag{5.6}$$

where h_i is a one-election integral describing the kinetic and potential energy of electron i in the field of all nuclei, J_{ij} is the Coulomb integral that represents the electrostatic repulsion between an electron in ψ_i and an electron in ψ_j, and K_{ij} is the exchange integral resulting from the antisymmetric nature of the Slater determinant expansion terms but has no classical analogue. Expressions for these integrals are

$$h_i = \int \psi_i(1) \left[-\frac{1}{2}\nabla_1^2 - \sum_A \frac{Z_A}{r_{1A}} \right] \psi_i(1) d\mathbf{r}_1 = \langle \psi_i | \hat{h}_i | \psi_i \rangle, \tag{5.7}$$

$$J_{ij} = \int \psi_i(1)\psi_j(2) \frac{1}{r_{12}} \psi_i(1)\psi_j(2) d\mathbf{r}_1 d\mathbf{r}_2 = \langle \psi_i \psi_j | \psi_i \psi_j \rangle, \tag{5.8}$$

$$K_{ij} = \int \psi_i(1)\psi_j(2) \frac{1}{r_{12}} \psi_j(1)\psi_i(2) d\mathbf{r}_1 d\mathbf{r}_2 = \langle \psi_i \psi_j | \psi_j \psi_i \rangle, \tag{5.9}$$

where the expressions on the right are the associated integrals in bra-ket notation.

The variation theorem, which states that the energy of an approximate wavefunction must be greater than or equal to the exact ground state energy, can be applied to Equation 5.6 to determine the MOs that give rise to a minimum energy, subject to the constraint of the MOs remaining orthogonal. This results in the Hartree-Fock equations, which must be solved for each MO, ψ_i,

$$\hat{F}_i \psi_i = \epsilon_i \psi_i, \tag{5.10}$$

where ϵ_i are the orbital energies and $\hat{F}_i$ is the Fock operator given by,

$$\hat{F}_i = \hat{h}_i + \sum_j (\hat{J}_{ij} - \hat{K}_{ij}). \tag{5.11}$$

in which the $\hat{h}_i$, $\hat{J}_{ij}$ and $\hat{K}_{ij}$ operators generate the corresponding integrals shown in Equations 5.7, 5.8 and 5.9, respectively. In order to solve the HF equations for molecules, the MOs are expressed as a linear combination of atomic orbitals (χ_k),

$$\psi_i = \sum_k c_{ik}\chi_k, \tag{5.12}$$

where c_{ik} is the coefficient of atomic orbital k in MO i. A set of atomic orbitals that are combined in linear combination is referred to as a basis set. The set of HF equations can then be combined in matrix form by substituting Equation 5.12 into Equation 5.11 to give the Roothaan-Hall equations,

$$\mathbf{FC} = \mathbf{SC}\epsilon, \tag{5.13}$$

where $\mathbf{F}$ is the Fock matrix, $\mathbf{C}$ is a matrix of coefficients, $\mathbf{S}$ is the overlap matrix, and ϵ is the matrix of orbital energies, which can be made diagonal by convention. Fock matrix diagnolization is required to determine the orbital coefficients and orbital energies; however, knowledge of the orbital coefficient matrix is necessary to calculate the J_{ij} and K_{ij} integrals. Therefore, the Roothaan-Hall equations are solved using the self-consistent field (SCF) procedure, whereby an initial guess is made for the orbital coefficients and then solved iteratively until convergence.

5.2 Post-HF Methods

Solving the Schrödinger equation using the HF method involves approximating the total molecular wavefunction as a single Slater determinant and using the Fock operator to calculate the interaction of each electron with the static electric field created by all other electrons in an average way. However, each electron actually moves under the influence of the repulsion of individual electrons rather than an average electron cloud at any moment. A measure of the extent to which an *ab initio* calculation accounts for electron correlation is called the "correlation energy" (E_{corr}) and may be defined as the difference between the

exact solution (E) and the HF energy (E_{HF}),

$$E_{\text{corr}} = E - E_{\text{HF}}. \tag{5.14}$$

To incorporate correlation energy, the HF results can be improved by incorporating additional Slater determinants in a broad range of calculations called post-HF methods, which take the general form,

$$\Psi = a_0\Phi_{\text{HF}} + \sum_i a_i\Phi_i, \tag{5.15}$$

where a_0 and a_i are coefficients that reflect the weight of the ground state HF determinant Φ_{HF} and all other Slater determinants Φ_i, respectively. A set of determinants can be constructed by replacing occupied MOs in the HF determinant with unoccupied MOs and are denoted by the number of HF MOs that have been replaced. This leads to Slater determinants that are singly, doubly, triply, quadruply, etc., excited, and are referred to as Singles (S), Doubles (D), Triples (T), Quadruples (Q), etc., determinants (see Figure 5.1). Three *ab initio* approaches that are typically used to calculate electron correlation are Configuration Interaction (CI), Coupled Cluster (CC), and Many-Body Perturbation Theory (MBPT).

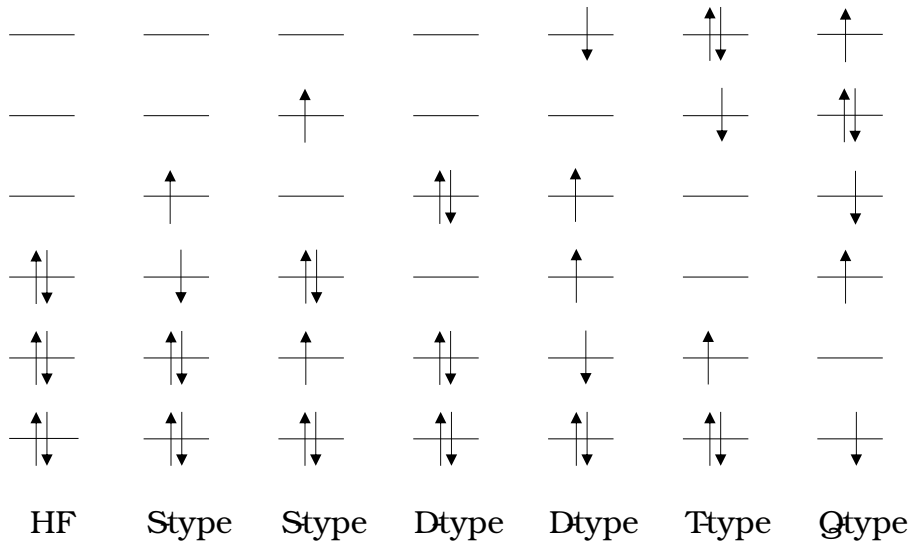


Fig. 5.1. Excited Slater determinants derived from a Hartree-Fock reference.

5.2.1 Many-Body Perturbation Theory

Perturbation theory allows for approximating the Hamiltonian for a system of interest that differs slightly from a Hamiltonian that has been well studied. Mathematically, the Hamiltonian operator is comprised of two parts; a reference $\hat{H}_0$ referred to as the unperturbed Hamiltonian, and a perturbation $\hat{H}'$,

$$\hat{H} = \hat{H}_0 + \lambda \hat{H}' \tag{5.16}$$

$$\hat{H}_0 \Phi_i = E_i \Phi_i, i = 0, 1, 2, \cdots, \infty, \tag{5.17}$$

where λ is a parameter that determines the perturbation strength. Solutions to the unperturbed Hamiltonian should be calculable and form a complete set. The perturbed Schrödinger equation is given by

$$\hat{H}\Psi = \left(\hat{H}_0 + \lambda \hat{H}'\right) = W\Psi, \tag{5.18}$$

where W is the energy of the perturbed system. The perturbed energy and wavefunction can then be written in powers of λ,

$$W = \lambda^0 W_0 + \lambda^1 W_1 + \lambda^2 W_2 + \cdots \tag{5.19}$$

$$\Psi = \lambda^0 \Psi_0 + \lambda^1 \Psi_1 + \lambda^2 \Psi_2 + \cdots \tag{5.20}$$

where W_0 and Ψ_0 are the zeroth-order or unperturbed energy (E_0) and wavefunction (Φ_0), respectively, and W_1, W_2, etc. and Ψ_1, Ψ_2, etc. are first-, second-, etc. order corrections to the unperturbed energy and wavefunction, respectively. Equations 5.19 and 5.20 can then be substituted into Equation 5.18 and expressions for the corrections can be derived in terms of the solutions to the unperturbed Hamiltonian (Equation 5.17). The latter is possible because the solutions to the unperturbed system form a complete set and each

correction can be written as

$$\Psi_n = \sum_i a_i \Phi_i. \tag{5.21}$$

In Møller-Plesset (MP) perturbation theory, the sum of the one-electron Fock operators (Equation 5.11) is taken as the unperturbed Hamiltonian. The zeroth-order energy is then the sum of MO energies (MP0); adding the first-order correction leads to the HF energy (MP1); and addition of the second-order correction, which is the first improvement on the HF energy, gives the MP2 energy as

$$E_{\text{MP2}} = E_{\text{HF}} + \sum_{i<j}^{\text{occ}} \sum_{a<b}^{\text{vir}} \frac{\left(\langle \psi_i \psi_j | \psi_a \psi_b \rangle - \langle \psi_i \psi_j | \psi_b \psi_a \rangle\right)^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}, \tag{5.22}$$

where ψ_i and ψ_j are occupied MOs, ψ_a and ψ_b are virtual or empty orbitals, and ϵ_i and ϵ_a are the corresponding MO energies. Therefore, in MP2 theory the electron correlation is described as a sum over doubly excited determinants (D) by promoting two electrons from the occupied orbitals i and j to the virtual orbitals a and b. It can be shown that the MP3 energy is expressed using only doubly excited determinants, whereas to calculate the MP4 energy, triply and quadruply excited determinants (T and Q) are also included.

5.2.2 Coupled Cluster Theory

Perturbation theory adds *all* types of corrections (e.g., D, T, Q, etc.) to a reference wave function to a given order. On the other hand, Coupled Cluster (CC) methods include all corrections of a given type to *infinite* order. This is achieved by constructing excitations using the cluster operator shown below,

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \ldots + \hat{T}_N, \tag{5.23}$$

where the $\hat{T}_i$ operator acting on an HF reference Slater determinant generates all determinants having i excitations. For example, the $\hat{T}_2$ operator is given by

$$\hat{T}_2 = \sum_{i<j}^{\text{occ}} \sum_{a<b}^{\text{vir}} t_{ij}^{ab} \Phi_{ij}^{ab}, \tag{5.24}$$

where the amplitudes t_{ij}^{ab} are equivalent to the expansion coefficients a_i in Equation 5.15 and Φ_{ij}^{ab} is the determinant created by exciting electrons from orbitals i and j to the virtual orbitals a and b, respectively. The CC wavefunction is then defined as

$$\Psi_{\text{CC}} = e^{\hat{T}} \Phi_0, \tag{5.25}$$

where $\hat{T}$ is the excitation operator given by Equation 5.23. The exponential operator can be expanded as a Taylor series, which if truncated to include only single and double excitations (i.e., coupled cluster with single and double excitations, or CCSD) becomes

$$\begin{aligned} e^{\hat{T}} &= e^{\hat{T}_1 + \hat{T}_2} \\ &= 1 + \left(\hat{T}_1 + \hat{T}_2\right) + \frac{1}{2!}\left(\hat{T}_1 + \hat{T}_2\right)^2 + \frac{1}{3!}\left(\hat{T}_1 + \hat{T}_2\right)^3 + \cdots \\ &= 1 + \hat{T}_1 + \left(\hat{T}_2 + \frac{1}{2}\hat{T}_1^2\right) + \left(\hat{T}_2\hat{T}_1 + \frac{1}{6}\hat{T}_1^3\right) + \cdots, \end{aligned} \tag{5.26}$$

where the first parenthesis in the last line includes *all* double excitations that may either be *connected* ($\hat{T}_2$) or *disconnected* ($\hat{T}_1^2$). The connected or "true double" excitation corresponds to two electrons interacting simultaneously, whereas the disconnected or "product double" excitation corresponds to two non-interacting single excitations. The next parenthesis shows that despite truncating the $\hat{T}$ operator at double excitations, disconnected triple excitations are also formed as products of single ($\hat{T}_1^3$) or double and single ($\hat{T}_2\hat{T}_1$) excitations. This property elevates CCSD above MP2 in computing the correlation energy,

albeit at an increased computational cost. The corresponding CCSD energy is given by

$$E_{\text{CCSD}} = \langle \Phi_0 | \hat{H} e^{\hat{T}_1 + \hat{T}_2} | \Phi_0 \rangle. \tag{5.27}$$

Including triple excitations in the $\hat{T}$ operator to form CCSDT comes at a substantial increase in computational cost and is too expensive for all but the smallest of molecules. Alternatively, the contribution of $\hat{T}_3$ can be estimated using perturbation theory, leading to the CCSD(T) method, which is considered the "gold standard" in *ab initio* computational quantum chemistry.

5.3 Density Functional Theory

Hohenberg and Kohn published two theorems that form the basis of Density Functional Theory (DFT). The first theorem is an existence theorem that says all ground state molecular properties are a functional of the ground state electron density. However, while this theorem guarantees that molecular properties can be calculated from the electron density, it does not offer a way to predict the density of the system. The second Hohenberg-Kohn theorem shows that, just as in MO theory, the density obeys a variation principle such that any trial electron density function will give an energy more than or equal to the true ground state energy, i.e.,

$$E_v[\rho_t] \geq E_0[\rho_0], \tag{5.28}$$

where $E_v[\rho_t]$ is the E_v functional of the trial electronic density ρ_t and $E_0[\rho_0]$ is the ground state energy that corresponds to the true ground state electronic density ρ_0.

To progress in developing an energy functional, the energy can be divided into the kinetic energy, $T[\rho]$, electron-electron repulsion, $E_{\text{ee}}[\rho]$, and the attraction between nuclei and electrons, $E_{\text{ne}}[\rho]$. Similarly to HF theory, $E_{\text{ee}}[\rho]$ can be divided into Coulomb and exchange functionals, $J[\rho]$ and $K[\rho]$, which both implicitly include the correlation energy.

For example, the Thomas-Fermi (TF) functional describes the energy using

$$\begin{aligned} E_{\text{TF}}[\rho] &= T_{\text{TF}}[\rho] + E_{\text{ne}}[\rho] + J[\rho] \\ &= \frac{3}{10}\left(3\pi^2\right)^{2/3} \int \rho^{5/3}\left(\mathbf{r}\right) d\mathbf{r} - \sum_{A}^{N_{\text{nuclei}}} \int \frac{Z_A}{|\mathbf{r}-\mathbf{r}_A|} \rho\left(\mathbf{r}\right) d\mathbf{r} + \frac{1}{2} \iint \frac{\rho\left(\mathbf{r}_1\right)\rho\left(\mathbf{r}_2\right)}{|\mathbf{r}_2-\mathbf{r}_1|} d\mathbf{r}_1 d\mathbf{r}_2 \end{aligned} \tag{5.29}$$

while the addition of $K[\rho]$ constitutes the Thomas-Fermi-Dirac (TFD) model,

$$E_{\text{TFD}}[\rho] = E_{\text{TF}}[\rho] + K[\rho] = E_{\text{TF}}[\rho] - \frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} \int \rho^{4/3}\left(\mathbf{r}\right) d\mathbf{r}. \tag{5.30}$$

Earlier attempts to deduce $T[\rho]$ and $K[\rho]$, such as the TFD model, assumed a uniform electron gas for valence electrons, which behaved well for metallic systems, but worked poorly for atoms and molecules.

The Kohn-Sham (KS) approach rather assumes that the kinetic energy can be written as an exact term and a small correction, with the former obtained as in HF theory assuming a sum of non-interacting electrons expressed as a Slater determinant of molecular orbitals, ψ_i,

$$T_{\text{S}} = \sum_{i}^{N_{\text{elec}}} \langle \psi_i \left| -\frac{1}{2}\nabla^2 \right| \psi_i \rangle. \tag{5.31}$$

In this case, the MOs are referred to as KS MOs. However, electrons do interact and Equation 5.31 does not account for the real kinetic energy, hence the need for a correction term, $\Delta T[\rho]$. With the reintroduction of molecular orbitals, $E_{\text{ne}}[\rho]$ and $J[\rho]$ can be recast into an orbital representation,

$$E_{\text{ne}}[\rho] + J[\rho] = \sum_{i}^{N_{\text{elec}}} \left(- \left\langle \psi_i \left| \sum_{A}^{N_{\text{nuclei}}} \int \frac{Z_A}{|\mathbf{r}_i - \mathbf{r}_A|} \right| \psi_i \right\rangle + \left\langle \psi_i \left| \frac{1}{2} \int \frac{\rho(\mathbf{r})}{|\mathbf{r}_i - \mathbf{r}|} d\mathbf{r} \right| \psi_i \right\rangle \right), \tag{5.32}$$

since $\rho = \sum_i^{N_{\text{elec}}} |\psi_i|^2$. Furthermore, the classic electrostatic repulsion must be corrected for all non-classical effects and electron correlation, $\Delta J[\rho]$, i.e.,

$$E_{\text{DFT}}[\rho] = (T_{\text{S}}[\rho] + \Delta T[\rho]) + E_{\text{ne}}[\rho] + (J[\rho] + \Delta J[\rho])\,. \tag{5.33}$$

These corrections are formally combined into an *exchange-correlation* functional, $E_{\text{xc}}[\rho]$, so that the general expression for a DFT functional is

$$E_{\text{DFT}}[\rho] = T_{\text{S}}[\rho] + E_{\text{ne}}[\rho] + J[\rho] + E_{\text{xc}}[\rho]. \tag{5.34}$$

Conventionally, E_{xc} can be separated as shown,

$$E_{\text{xc}}[\rho] = E_{\text{x}}[\rho] + E_{\text{c}}[\rho] \tag{5.35}$$

$$E_{\text{x}}[\rho] = E_{\text{x}}^{\alpha}[\rho_\alpha] + E_{\text{x}}^{\beta}[\rho_\beta] \tag{5.36}$$

$$E_{\text{c}}[\rho] = E_{\text{c}}^{\alpha\alpha}[\rho_\alpha] + E_{\text{c}}^{\beta\beta}[\rho_\beta] + E_{\text{c}}^{\alpha\beta}[\rho_\alpha, \rho_\beta]. \tag{5.37}$$

where ρ_α and ρ_β are the α and β spin densities. The mechanics of HF theory can then be used to determine optimal non-interacting molecular orbitals and thus an approximate electron density so that the challenge in DFT transitions to the development and parameterisation of an exchange-correlation functional.

5.3.1 Exchange-Correlation Functionals

The *Local Density Approximation* (LDA) is the simplest method to approximate E_{xc}, in which the basic assumption is that the density can be treated as a uniform electron gas where the density has the same value at every position, or equivalently that the density is

a slowly varying function,

$$E_{\text{xc}}^{\text{LDA}}[\rho] = \int \rho(\mathbf{r}) \varepsilon_{\text{xc}}^{\text{LDA}}\left(\rho(\mathbf{r})\right) d\mathbf{r}, \tag{5.38}$$

where ε_{xc} is the exchange-correlation energy density. In general, α and β spin densities are not equal; therefore, the LDA has mostly been replaced by the *Local Spin Density Approximation* (LSDA), which allows for a separate treatment of the densities. For the special case of a uniform electron gas, the LSDA method is an exact DFT method. However, when applied to molecular systems, LSDA underestimates the exchange energy by approximately 10 %, which exceeds the errors for the whole correlation energy. Furthermore, the electron correlation is overestimated by a factor close to two, which consequently causes an overestimation of the bond strengths. Despite basic fundamental assumptions, LSDA methods often yield results that are similar in accuracy to the HF method.

The LSDA approach can be improved by considering a non-uniform gas. This can be achieved by making the correlation and exchange energy dependent on the electron density as well as the derivatives of the density resulting in the *Generalised Gradient Approximation* (GGA),

$$E_{\text{xc}}^{\text{GGA}}[\rho] = \int \rho(\mathbf{r}) \varepsilon_{\text{xc}}^{\text{GGA}}\left(\rho(\mathbf{r}), \nabla\rho(\mathbf{r})\right) d\mathbf{r}. \tag{5.39}$$

GGA and other higher-order functionals are typically constructed such that the correction is added to the energy density of a suitable LDA functional for either exchange or correlation, e.g.,

$$E_{\text{x/c}}^{\text{GGA}}[\rho] = \varepsilon_{\text{x/c}}^{\text{LDA}}[\rho] + \Delta\varepsilon_{\text{x/c}}\left[\frac{|\nabla\rho(\mathbf{r})|}{\rho^{4/3}(\mathbf{r})}\right], \tag{5.40}$$

where the correction depends on the dimensionless reduced gradient rather than the absolute gradient.

Extension of the GGA involves the incorporation of higher-order derivatives of the electron density into the exchange and correlation functionals. In the resulting *meta-GGA*

functionals, the kinetic energy density (τ) or Weizsäcker kinetic energy (τ_{W}) is included,

$$E_{\mathrm{xc}}^{\text{meta-GGA}}[\rho] = \int \rho(\mathbf{r}) \varepsilon_{\mathrm{xc}}^{\text{meta-GGA}} \left(\rho(\mathbf{r}), \nabla\rho(\mathbf{r}), \tau(\mathbf{r})\right) d\mathbf{r}, \tag{5.41}$$

$$\tau(\mathbf{r}) = \frac{1}{2} \sum_{i}^{\text{occ}} \left|\nabla\psi_i(\mathbf{r})\right|^2, \tag{5.42}$$

$$\tau_{\mathrm{W}}(\mathbf{r}) = \frac{\left|\nabla\rho(\mathbf{r})\right|^2}{8\rho(\mathbf{r})}. \tag{5.43}$$

GGA and meta-GGA functionals are also referred to as *semi-local functionals*; although higher-order "non-local" derivates of the electron density are used, the electron density information is still independently evaluated only a local point.

Hybrid or *Hyper-GGA* functionals utilise the fact that the exchange energy can be calculated exactly in the HF theory due to the antisymmetric nature of the Slater determinant. However, since the KS orbitals are not the exact orbitals of the system, mixing of empirical DFT exchange and exact exchange is still warranted, i.e., the hybrid exchange-correlation energy is typical expressed in the form

$$E_{\mathrm{xc}}^{\text{Hybrid}} = (1-a)\, E_{\mathrm{x}}^{\text{DFT}} + a E_{\mathrm{x}}^{\text{HF}} + E_{\mathrm{c}}^{\text{DFT}} \tag{5.44}$$

where a is the mixing parameter.

The last formal level of improvement is to use not only the occupied KS orbitals in constructing the exchange-correlation energy, but also the virtual or empty orbitals. The simplest way to achieve this is to use the MP2 expression (the last term in Equation 5.22) but with KS orbitals and energies instead of HF orbitals and energies. The resulting *double hybrid* DFT (DHDFT) exchange correlation energy is then given by

$$E_{\mathrm{xc}}^{\text{DHDFT}} = (1-a)\, E_{\mathrm{x}}^{\text{DFT}} + a E_{\mathrm{x}}^{\text{HF}} + (1-b)\, E_{\mathrm{c}}^{\text{DFT}} + b E_{\mathrm{c}}^{\text{MP2}} \tag{5.45}$$

where a and b are mixing parameters.

5.4 Gaussian-*n* Theory

Gaussian-*n* (G*n*) theories ($n = 1, 2, 3, 4$)[44,45,47,49] were developed with the purpose of approaching the exact molecular energy through calculations that are based on *ab initio* molecular orbital theory and employ different levels of accuracy and basis sets. The Gn methods are therefore composite approaches, in which high-level correlation methods and moderate-sized basis sets are used in combination with energies from low-level calculations with larger basis sets. Any remaining deficiencies are accounted for by including an empirical high-level correction (HLC).

The guiding objectives of the G*n* methods are its: (a) applicability to all molecular systems, (b) computational efficiency, and (c) ability to reproduce experimental results to a prescribed accuracy. G3 theory replaced the G1 theory, the first in the Gn series, as well as its successor, the G2 theory, to eliminate several deficiencies that were identified. G3 theory uses a series of well-defined *ab initio* calculations to get the total energy, given as,[112]

$$\begin{aligned} \text{G3} = {} & E(\text{QCISD(T,FC)/6-31G(d)}) + E(\text{plus}) + E(\text{2df,p}) + E(\Delta\text{G3L}) \\ & + E(\text{SO}) + E(\text{HLC}) + E(\text{ZPE}), \end{aligned} \tag{5.46}$$

where *E*(QCISD(T,FC)/6-31G(d,p)) is the correlated energy determined using an MP2/6-31G(d) optimised geometry. The frozen core (FC) approximation, which excludes valence electrons from the correlation treatment, is used in the QCISD calculation, whereas the MP2 calculation includes all electrons. *E*(plus) is the effect of diffuse functions that is calculated as the difference between MP4(FC)/6-31+G(d) and MP4(FC)/6-31G(d) energies. The difference between MP4(FC)/6-31G(d) and MP4(FC)/6-31G(2df,p) accounts for the effect of the higher polarisation functions, *E*(2df,p). A correction for larger basis set effects and the nonadditivity resulting from the assumption of separate basis set extensions for higher polarisation and diffuse functions is denoted *E*(ΔG3L) = *E*(MP2/G3Large)

$-$ E(MP2/6-31G(2df,p)) $-$ E(MP2/6-31+G(d)) $+$ E(MP2/6-31G(d)). Atomic species require a spin-orbit correction term E(SO), which is taken from experimental results if available or theoretical calculations if not. E(HLC) is a high-level correction given by $-6.386n_\beta - 2.977n_\alpha - n_\beta$ for molecules and $-6.219n_\beta - 1.185n_\alpha - n_\beta$ for atomic species, where n_α and n_β are the number of valence α and β electrons. Lastly, E(ZPE) is the zero point energy calculated from harmonic frequencies at the HF/6-31G(d) level of theory, scaled by a factor of 0.8929.

G3(MP2) is a modified version of G3 that employs a reduced order of Møller-Plesset perturbation theory, making the Gn(MP2) procedures applicable to a wider range of systems since they are less computationally demanding. In this theory, the energy is given by,[112]

$$\begin{aligned} \text{G3(MP2)} = {} & E(\text{QCISD(T,FC)/6-31G(d)}) + E(\text{G3MP2L}) + E(\text{SO}) \\ & + E(\text{HLC}) + E(\text{ZPE}), \end{aligned} \tag{5.47}$$

where the E(plus), E(2df,p), and E(ΔG3L) terms from Equation 5.46 have been replaced with E(G3MP2L), which is determined as the difference in energy between an MP2/G3MP2Large and MP2(FC)/6-31G(d) calculation. Furthermore, the HLC coefficients were redetermined by fitting to the G2/97 test set.

The Gaussian-4 theory modifies the G3 theory in several ways,[112]

$$\begin{aligned} \text{G4} = {} & E(\text{CCSD(T,FC)/6-31G(d)}) + E(\text{plus}) + E(\text{2df,p}) + E(\Delta\text{G3LXP}) \\ & + E(\text{HF/limit}) + E(\text{SO}) + E(\text{HLC}) + E(\text{ZPE}), \end{aligned} \tag{5.48}$$

The equilibrium structure is obtained at the DFT level of theory using B3LYP/6-31G(2df,p). The next modification is that the Hartree-Fock basis set limit, E(HF/limit) is calculated through a linear two-point extrapolation scheme and Dunning's aug-cc-pVnZ basis sets such that $E_{\text{HF/aug-cc-pVnZ}} = E_{\text{HF/limit}} + B\ \exp(\alpha n)$, where the number of contractions in the

valence shell of the basis set is denoted by *n* and α is equal to 1.63. A further modification is that because of the dramatic failures of some molecules, the QCISD(T,FC) method is replaced by CCSD(T,FC)/6-31G(d). The scaled ZPE is calculated at the B3LYP level of theory with a 6-31G(2df,p) basis set, with a scaling factor of 0.9854. Another modification is the use of the G3LargeXP basis set, which extends on the G3Large basis set, in the *E*(ΔG3LXP) term. Lastly, two new empirical parameters are added to the HLC of G3.

As before, the G4(MP2) theory drops some of the terms and reduces the maximum order of the perturbation theory to MP2. However, in this work, the more recent G4(MP2)-6X method was used and is given by,[97]

$$\begin{aligned}\text{G4(MP2)-6X} = \text{HF/CBS} + E^{\text{corr}}_{\text{SCS-MP2}}/\text{G3MP2LargeXP} + \Delta E_{\text{S-CCSD}}/6-31\text{G(d)} + \\ E^{\text{corr}}_{\text{S-(T)}}/6-31\text{G(d)} + \text{HLC} + \text{ZPVE} + E_{\text{SO}}.\end{aligned} \tag{5.49}$$

A description of G4(MP2)-6x, including relevant contrasts with G4(MP2), is as follows:

- The equilibrium structure is obtained at the BMK/6-31+G(2df,p) level of theory. This differs from the G4(MP2) theory that calculates the geometries at the B3LYP-6-31G(2df,p) level.

- HF/CBS estimates the Hartree-Fock energy limit and is calculated by extrapolating to the complete basis set limit using aug-cc-pV(n+d)Z(n = T,Q) basis sets that were adjusted by minimising the number of diffuse and polarisation functions,

$$E_{\text{CBS}} = \frac{[E_{\text{Q}} - E_{\text{T}} \exp(-1.63)]}{1 - \exp(-1.63)}. \tag{5.50}$$

- a_1, a_2, a_3, a_4, a_5, and a_6 are parameters optimized using the E2 training set.

- $\Delta E_{\text{SCS-MP2}}$ is a correction term for the correlation effects at the MP2 level and is calculated as,

$$E^{\text{corr}}_{\text{SCS-MP2}}/\text{G3MP2LargeXP} = a_3 \cdot E^{\text{corr}}_{\text{MP2OS}}\prime + a_4 \cdot E^{corr}_{\text{MP2SS}}\prime, \tag{5.51}$$

where $a_3 \cdot E^{\text{corr}}_{\text{MP2OS}}\prime$ are the scaled opposite-spin (OS) and $a_4 \cdot E^{\text{corr}}_{\text{MP2SS}}\prime$ are the scaled same-spin (SS) contributions to the MP2/G3MP2LargeXP correlation energy, respectively.

- The additional corrections that account for higher-order correlation effects are given by,

$$\Delta E_{\text{scal-CCSD}}/\text{6-31G(d)} = a_5 \cdot E^{\text{corr}}_{\text{CCSD}} - (a_1 \cdot E^{\text{corr}}_{\text{MP2OS}} + a_2 \cdot E^{\text{corr}}_{\text{MP2SS}}), \tag{5.52}$$

where $a_1 \cdot E^{\text{corr}}_{\text{MP2OS}}$ are the scaled OS and $a_2 \cdot E^{\text{corr}}_{\text{MP2SS}}$ are the scaled SS contributions to the MP2/6-31G(d) correlation energy, and $a_5 \cdot E^{\text{corr}}_{\text{CCSD}}$ is the scaled CCSD contribution to the CCSD(T)/6-31G(d) correlation energy. Finally,

$$E^{\text{corr}}_{\text{S-(T)}}/\text{6-31G(d)} = a_6 \cdot E^{\text{corr}}_{\text{(T)}}, \tag{5.53}$$

where $a_6 \cdot E^{\text{corr}}_{\text{(T)}}$ is the scaled perturbative triples contribution to the CCSD(T)/6-31G(d) correlation energy.

- The frozen-core approximation is used for all correlation calculations consistent with the G4(MP2)-6X method. The largest noble-gas core is frozen, apart from the following that are treated as valence orbitals:
  - 3d orbitals on third-row main-group elements (Ga-Kr)
  - 2s and 2p orbitals on Na and Mg
  - 3s and 3p orbitals on K and Ca

- A higher-level-correction (HLC) term is dependent on the number of valence electrons and is given as parameters that have been optimized using the E2 set:

$$\text{HLC} = \begin{cases} -\text{A}n_\beta & \text{for closed-shell molecules} \\ -\text{A}n_\beta - B(n_\alpha - n_\beta) & \text{for open-shell molecules} \\ -\text{C}n_\beta - D(n_\alpha - n_\beta) & \text{for atomic species} \\ -\text{E}n_\beta & \text{for “single electron pair” species, such as } Li_2 \end{cases}$$

  where n_α and n_β are the number of valence α and β electrons, respectively. In contrast, the G3/05 training set is used to optimize the same parameters for the G4(MP2) procedure.

- Scaled BMK/6-31+G(2df,p) vibrational frequencies are used to get the zero-point vibrational energies (0.9770), thermal corrections to the enthalpy at 298.15 K (0.9627), and the entropy at 298.15 K (0.9695) for G4(MP2)-6X while scaled B3LYP/6-31G(2df,p) vibrational frequencies are employed for G4(MP2).

- The same spin-orbit correction (E_{SO}) is included as in G4(MP2), which is calculated through experiment or high-level computations.

The G4(MP2)-6X has a mean absolute deviation (MAD) of 3.64 kJ mol^{-1}, which is an improvement compared to 4.42 kJ mol^{-1}, across the diverse set of 526 energies in the E2 dataset.

6 Conclusion

The project aimed to extend Benson's second-order group contribution method (GCM) to accurately predict the gas phase enthalpy of formation ($\Delta_f H^\circ$) of energetic azoles. To achieve this, a quantitative assessment of the uncertainty and bias in G4(MP2)-6X calculed gas-phase enthalpies of formation was done. This involved compiling a list of 47 CHN-containing molecules, which were selected as the biases were expected to be similar to those of the azoles. The resulting correction was determined to be 1.11 kJ mol^{-1}, with an associated standard uncertainty of 3.04 kJ mol^{-1}. The accuracy of G4(MP2)-6X was therefore found to be sufficient and well within chemical accuracy of approximately 4 kJ mol^{-1} (1 kcal mol^{-1}).

The first GCM considered used 72 *acyclic* molecules to regress 42 group additive values (GAVs) that describe ten azoles (1H-pyrrole, 1H-pyrazole, 1H-imidazole, 1H-1,2,3-triazole, 1H-1,2,4-triazole, 2H-1,2,3-triazole, 4H-1,2,4-triazole, 1H-tetrazole, 2H-tetrazole, 1H-pentazole). The MAE of this fitting was 3.31 kJ mol^{-1} with a MAPE of 1.70 %. Subsequently, $\Delta_f H^\circ$ was calculated for each of the azoles using these GAVs, and the difference between this GCM value and the G4(MP2)-6X computed $\Delta_f H^\circ$ was attributed to the combined effect of ring strain and aromatic stabilisation energy (RS + ASE). To assess whether the resulting RS + ASE terms are transferable to functionalised azoles, a fitting set of 180 acyclic molecules, functionalised at the pyrrole-like equivalent nitrogen atom with a methyl ($-CH_3$) group, and explosophoric groups amine ($-NH_2$), azide ($-N_3$), nitro ($-NO_2$), and nitramide ($-NHNO_2$), were regressed. The GAVs for groups obtained in the previous fitting were kept constant, and only the values of explosophoric groups and those groups containing pyrrole-like nitrogen as the central atom were determined. This fitting had an MAE of 18.05 kJ mol^{-1} and a MAPE of 4.69 %, demonstrating limitations in the transferability of the RS + ASE correction. In transferring the correction to functionalised azoles while keeping the GAVs fitted for unsubstituted azoles unchanged, the substituent

effect must be carried by the new functional group and pyrrole-like nitrogen groups only. However, the results show that these groups are not sufficient to describe the substitiuent effect uniformly across all ten azole frameworks. Nonetheless, the advantage of this GCM is its flexibility, as new functional groups can be incorporated without refitting the entire dataset.

The alternative GCM used linear regression of the ten azoles to determine 13 GAVs that have the RS + ASE directly included, therefore avoiding this contribution as a correction. The fitting resulted in a low MAE of 0.11 kJ mol^{-1} and an MAPE of 0.04 %. Following this, these GAVs for the unsubstituted azoles were kept fixed, while remaining groups in the methyl and explosophoric N-functionalised azoles were regressed. The error for this fitting increased significantly, with an MAE of 19.33 kJ mol^{-1} and an MAPE of 5.68 %. This model was then improved by regressing the full set of functionalised and unfunctionalised azoles, leading to a much lower MAE of 3.49 kJ mol^{-1} and an MAPE of 1.28 %. The advantage of this GCM is that each of the 13 general azole GAVs together with the functional group and pyrrole-like nitrogen atom group now carry the effect of functionalisation, on average. However, this approach requires refitting the dataset every time a new functional group is added, which makes practical implementation cumbersome. To assess whether the substituent effect can be sufficiently described by fewer than the total number of groups, the subsets of unsubstituted, methyl, amine, azide, nitro, and nitramide functionalised azoles were regressed separately. However, the results demonstrated that this is not possible with variations up to 12 kJ mol^{-1} in the five general groups used in the subsets. Therefore, the only way to improve would be to include multiple third-order groups that would have to be parameterised together with every newly added functional group.

Overall, this work has demonstrated that the GCM can be extended to azoles, which can estimate $\Delta_f H^\circ$ to within chemical accuracy. Building on the success of the second GCM model would be the functionalisation of the remainder of the atoms on the ring with more explosophoric groups.

References

1. Q. Zhang and J. M. Shreeve, *Chem. Rev.*, 2014, **114**, 10527–10574.
2. T. M. Klapötke, *Chemistry of High-Energy Materials*, De Gruyter, Munich, Germany, 4th edn, 2017.
3. J. Akhavan, *The Chemistry of Explosives*, The Royal Society of Chemistry, Cambridge, UK, 2nd edn, 2004.
4. A. Sobrero, *C.R.*, 1847, **24**, 247–248.
5. T. Lauder Brunton, *Lancet*, 1867, **90**, 97–98.
6. E. E. Benli, O. K. Koç, A. Üzer and R. Apak, *Talanta Open*, 2023, **8**, 100245.
7. Y. Han, Q. Liu, Y. Duan, Y. Zhao and X. Long, *Def. Technol.*, 2025, **44**, 83–97.
8. D. M. Badgujar, M. B. Talawar, S. N. Asthana and P. P. Mahulikar, *J. Hazard. Mater.*, 2008, **151**, 289–305.
9. K. E. Gutowski, R. D. Rogers and D. A. Dixon, *J. Phys. Chem. B*, 2007, **111**, 4788–4800.
10. U. R. Nair, R. Sivabalan, G. M. Gore, M. Geetha, S. N. Asthana and H. Singh, *Combust. Explos. Shock Waves.*, 2005, **41**, 121–132.
11. H. Gao and J. M. Shreeve, *Chem. Rev.*, 2011, **111**, 7377–7436.
12. M. B. Talawar, R. Sivabalan, T. Mukundan, H. Muthurajan, A. K. Sikder, B. R. Gandhe and A. S. Rao, *J. Hazard. Mater.*, 2009, **161**, 589–607.
13. D. Xiaoli, R. Xiaoting and L. Chao, *Int. Commun. Heat Mass Transf.*, 2023, **144**, 106788.

14. J. Nie, T. Jia, L. Pan, X. Zhang and J.-J. Zou, *Trans. Tianjin Univ.*, 2022, **28**, 1–5.

15. R. D. Ambrosini, B. M. Luccioni, R. F. Danesi, J. D. Riera and M. M. Rocha, *Shock Waves*, 2002, **12**, 69–78.

16. I. Onederra, V. Bailey, G. Cavanough and A. Torrance, *Trans. Inst. Min. Metall. A: Min. Technol.*, 2012, **121**, 151–159.

17. A. S. Bahrain and B. Grassland, *Proc. Inst. Mech. Eng.*, 1964, **179**, 264–305.

18. A. R. J. P. Ubbelohde, P. Woodward, J. L. Copp, S. E. Napier, T. Nash, W. J. Powell, H. Skelly, A. R. J. P. Ubbelohde, P. Woodward and R. Robertson, *Philos. trans., Math. phys. eng. sci.*, 1948, **241**, 238–248.

19. H. C. J. Saint and J. Hewson, *Analyst*, 1959, **84**, 183–187.

20. M. J. Kamlet and S. J. Jacobs, *J. Chem. Phys.*, 1968, **48**, 23–35.

21. P. Politzer and J. S. Murray, *Propellants, Explos., Pyrotech.*, 2019, **44**, 844–849.

22. P. Politzer and J. S. Murray, *Cent. Eur. J. Energ. Mater.*, 2014, **11**, 459–474.

23. M. Keshavarz, *Propellants, Explos., Pyrotech.*, 2008, **33**, 448–453.

24. B. Ruscic and D. H. Bross, *Comput. Aided Chem. Eng.*, 2019, **45**, 3–114.

25. C. G. Neochoritis, T. Zhao and A. Dömling, *Chem. Rev.*, 2019, **119**, 1970–2042.

26. R. N. Butler, J. C. Stephens and L. A. Burke, *Chem. Commun.*, 2003, 1016–1017.

27. H. F. Mull, J. M. Turney, G. E. Douberly and I. Schaefer, Henry F., *J. Phys. Chem. A*, 2021, **125**, 9092–9098.

28. B. Li, L. Li and X. Li, *Mol. Simul.*, 2019, **45**, 1459–1464.

29. S. Zhang, Z. Gao, D. Lan, Q. Jia, N. Liu, J. Zhang and K. Kou, *Molecules*, 2020, **25**, 3475.

30. J. R. Cho, K. J. Kim, S. G. Cho and J. K. Kim, *J. Heterocycl. Chem.*, 2002, **39**, 141–147.

31. P. S. Gribov, N. N. Kondakova, N. N. Il'icheva, E. R. Stepanova, A. P. Denisyuk, V. A. Sizov, V. D. Dotsenko, D. B. Vinogradov, P. V. Bulatov, V. P. Sinditskii, K. Y. Suponitsky, M. M. Il'in, M. L. Keshtov and A. B. Sheremetev, *Int. J. Mol. Sci.*, 2023, **24**, 9645.

32. A. A. Larin, A. N. Pivkina, I. V. Ananyev, D. V. Khakimov and L. L. Fershtat, *Front. Chem.*, 2022, **10**, 1012605.

33. Q. Lin, Y. Li, C. Qi, W. Liu, Y. Wang and S. Pang, *J. Mater. Chem. A*, 2013, **1**, 6776–6785.

34. G. Rafiqul, *Curr. Opin. Chem. Eng.*, 2019, **23**, 184–196.

35. D. K. Mital, P. Nancarrow, S. Zeinab, N. A. Jabbar, T. H. Ibrahim, M. I. Khamis and A. Taha, *Molecules*, 2021, **26**, 2454.

36. J. M. L. Martin and G. de Oliveira, *J. Chem. Phys.*, 1999, **111**, 1843–1856.

37. A. D. Boese, M. Oren, O. Atasoylu, J. M. L. Martin, M. Kàllay and J. Gauss, *J. Chem. Phys.*, 2004, **120**, 4129–4141.

38. A. Karton, E. Rabinovich, J. M. L. Martin and B. Ruscic, *J. Chem. Phys.*, 2006, **125**, 144108.

39. M. E. Harding, J. Vázquez, B. Ruscic, A. K. Wilson, J. Gauss and J. F. Stanton, *J. Chem. Phys.*, 2008, **128**, 114111.

40. A. Tajti, P. G. Szalay, A. G. Császár, M. Kállay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vázquez and J. F. Stanton, *J. Chem. Phys.*, 2004, **121**, 11599–11613.

41. S. J. Klippenstein, L. B. Harding and B. Ruscic, *J. Phys. Chem. A*, 2017, **121**, 6580–6602.

42. D. A. Dixon, D. Feller and K. A. Peterson, in *Chapter One - A Practical Guide to Reliable First Principles Computational Thermochemistry Predictions Across the Periodic Table*, ed. R. A. Wheeler, Elsevier, Amsterdam, 2012, vol. 8, pp. 1–28.

43. A. G. Császár, W. D. Allen and I. Schaefer, Henry F., *J. Chem. Phys.*, 1998, **108**, 9751–9764.

44. L. A. Curtiss, C. Jones, G. W. Trucks, K. Raghavachari and J. A. Pople, *J. Chem. Phys.*, 1990, **93**, 2537–2545.

45. L. A. Curtiss, K. Raghavachari and J. A. Pople, *J. Chem. Phys.*, 1993, **98**, 1293–1298.

46. L. A. Curtiss, K. Raghavachari and J. A. Pople, *J. Chem. Phys.*, 1993, **98**, 1293–1298.

47. L. A. Curtiss, K. Raghavach Ari, P. C. Redfern, V. Rassolov and J. A. Pople, *J. Chem. Phys.*, 1998, **109**, 7764–7776.

48. L. A. Curtiss, P. C. Redfern, K. Raghavachari, V. Rassolov and J. A. Pople, *J. Chem. Phys.*, 1999, **110**, 4703–4709.

49. L. A. Curtiss, P. C. Redfern and K. Raghavachari, *J. Chem. Phys.*, 2007, **126**, 084108.

50. L. A. Curtiss, P. C. Redfern and K. Raghavachari, *J. Chem. Phys.*, 2007, **127**, 124105.

51. N. J. DeYonker, B. Mintz, T. R. Cundari and A. K. Wilson, *J. Chem. Theory Comput.*, 2008, **4**, 328–334.

52. J. Montgomery, J. A., M. J. Frisch, J. W. Ochterski and G. A. Petersson, *J. Chem. Phys.*, 2000, **112**, 6532–6542.

53. J. W. Ochterski, G. A. Petersson and J. Montgomery, J. A., *J. Chem. Phys.*, 1996, **104**, 2598–2619.

54. S. W. Benson, *Thermochemical Kinetics*, John Wiley & Son, New York, 2nd edn, 1976.

55. S. W. Benson and J. H. Buss, *J. Chem. Phys.*, 1958, **29**, 546–572.

56. C. Shi and T. B. Borchardt, *ACS Omega*, 2017, **2**, 8682–8688.

57. K. G. Joback and R. C. Reid, *Chem. Eng. Commun.*, 1987, **57**, 233–243.

58. L. Constantinou and R. Gani, *AIChE J.*, 1994, **40**, 1697–1710.

59. J. Marrero and R. Gani, *Fluid Ph. Equilib.*, 2001, **183-184**, 183–208.

60. M. Bruce, R. Jürgen and R. Deresh, *J. Mol. Liq.*, 2008, **143**, 52–63.

61. A. S. Hukkerikar, R. J. Meier, G. Sin and R. Gani, *Fluid Ph. Equilib.*, 2013, **348**, 23–32.

62. V. van Speybroeck, R. Gani and R. J. Meier, *Chem. Soc. Rev.*, 2010, **39**, 1764–1779.

63. G. Rafiqul, *Curr. Opin. Chem. Eng.*, 2019, **23**, 184–196.

64. R. Gani, P. M. Harper and M. Hostrup, *Ind. Eng. Chem. Res.*, 2005, **44**, 7262–7269.

65. A. L. Lydersen, *Estimation of Critical Properties of Organic Compounds by the Method of Group Contributions, Report 3*, Engineering Experiment Station, Madison, WI, 1955.

66. S. W. Benson, F. R. Cruickshank, D. M. Golden, G. R. Haugen, H. E. O'Neal, A. S. Rodgers, R. Shaw and R. Walsh, *Chem. Rev.*, 1969, **69**, 279–324.

67. E. S. Domalski and E. D. Hearing, *J. Phys. Chem. Ref. Data*, 1988, **17**, 1637–1678.

68. E. S. Domalski and E. D. Hearing, *J. Phys. Chem. Ref. Data.*, 1994, **23**, 157–159.

69. N. Cohen, *J. Phys. Chem. Ref. Data.*, 1996, **25**, 1411–1481.

70. R. Sumathi and W. H. Green, *J. Phys. Chem. A*, 2002, **106**, 11141–11149.

71. R. W. Ashcraft and W. H. Green, *J. Phys. Chem. A*, 2008, **112**, 9144–9152.

72. J. L. Holmes and C. Aubry, *J. Phys. Chem. A*, 2011, **115**, 10576–10586.

73. J. L. Holmes and C. Aubry, *J. Phys. Chem. A*, 2012, **116**, 7196–7209.

74. M. K. Sabbe, M. Saeys, M.-F. Reyniers, G. B. Marin, V. Van Speybroeck and M. Waroquier, *J Phys Chem A.*, 2005, **109**, 7466–7480.

75. K. R. Bjorkman, C.-Y. Sung, E. Mondor, J. C. Cheng, D.-Y. Jan and L. J. Broadbelt, *Ind. Eng. Chem. Res.*, 2014, **53**, 19446–19452.

76. P. D. Paraskevas, M. K. Sabbe, M.-F. Reyniers, N. Papayannakos and G. B. Marin, *Chem. Eur. J.*, 2013, **19**, 16431–16452.

77. A. Ince, H. H. Carstensen, M. F. Reyniers and G. B. Marin, *AIChE J.*, 2015, **61**, 3858–3870.

78. C. A. R. Pappijn, R. Van de Vijver, M.-F. Reyniers, M. K. Sabbe, G. B. Marin and K. M. Van Geem, *Phys. Chem. Chem. Phys.*, 2024, **26**, 19021–19034.

79. L. D. Dellon, C.-Y. Sung, D. J. Robichaud and L. J. Broadbelt, *Ind. Eng. Chem. Res.*, 2017, **56**, 10259–10270.

80. J. Marrero and R. Gani, *Fluid Phase Equilib.*, 2001, **183-184**, 183–208.

81. R. J. Meier, *ChemEngineering*, 2021, **5**, 24.

82. R. J. Meier, *AppliedChem*, 2021, **1**, 111–129.

83. R. J. Meier, *AppliedChem*, 2022, **2**, 213–228.

84. R. J. Meier and P. R. Rablen, *Thermo*, 2023, **3**, 289–308.

85. R. J. Meier and P. R. Rablen, *Applied sciences*, 2024, **14**, 1929.

86. L. Pauling and J. Sherman, *J. Chem. Phys.*, 1933, **1**, 606–617.

87. T. M. Krygowski and H. Szatylowicz, *ChemTexts*, 2015, **1**, 12.

88. S. E. Wheeler, K. N. Houk, P. v. R. Schleyer and W. D. Allen, *J. Am. Chem. Soc.*, 2009, **131**, 2547–2560.

89. M. K. Cyrañski, T. M. Krygowski, A. R. Katritzky and P. v. R. Schleyer, *J. Org. Chem.*, 2002, **67**, 1333–1338.

90. P. v. R. Schleyer, H. Jiao, B. Goldfuss and P. K. Freeman, *Angew. Chem. Int. Ed. Engl.*, 1995, **34**, 337–340.

91. M. K. Cyrański, P. v. R. Schleyer, T. M. Krygowski, H. Jiao and G. Hohlneicher, *Tetrahedron*, 2003, **59**, 1657–1665.

92. K. K. Irikura, R. D. Johnson and R. N. Kacker, *Metrologia*, 2004, **41**, 369–375.

93. B. N. Taylor and C. Keyatt, *Guide to the Expression of Uncertainty in Measurement, corrected and reprinted 1995 International Organisation for Standardisation Geneva Switzerland*, 1995 (accessed February 2021).

94. P. Linstrom and E. W.G. Mallard, *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*, National Institute of Standards and Technology, Gaithersburg, MD, 20899, https://ilthermo.boulder.nist.gov, (accessed January 2021).

95. J. Zhang and M. Dolg, *Phys. Chem. Chem. Phys.*, 2015, **17**, 24173–24181.

96. B. Ruscic, *Int. J. Quantum Chem.*, 2014, **114**, 1097–1101.

97. B. Chan, J. Deng and L. Radom, *J. Chem. Theory Comput.*, 2011, **7**, 112–120.

98. J. H. Kim, *Korean J Anesthesiol*, 2019, **72**, 558–569.

99. R. C. Haddon and K. Raghavachari, *J. Am. Chem. Soc.*, 1985, **107**, 289–298.

100. P. v. R. Schleyer, H. Jiao, N. J. R. v. E. Hommes, V. G. Malkin and O. L. Malkina, *J. Am. Chem. Soc.*, 1997, **119**, 12669–12670.

101. A. Dippold, T. Klapoetke, F. Martin and S. Wiedbrauk, *Eur. J. Inorg. Chem.*, **2012**, 2429–2443.

102. E. Lewars, *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*, Springer, New York, US, 2nd edn, 2011.

103. B. Ruscic, R. E. Pinzon, G. v. Laszewski, D. Kodeboyina, A. Burcat, D. Leahy, D. Montoy and A. F. Wagner, *J. Phys. Conf. Ser.*, 2005, **16**, 561–570.

104. L. Zhao, R. Grande-Aztatzi, C. Foroutan-Nejad, J. M. Ugalde and G. Frenking, *ChemistrySelect*, 2017, **2**, 863–870.

105. S. Kikuchi, *J. Chem. Educ.*, 1997, **74**, 194.

106. P. George, M. Trachtman, C. W. Bock and A. M. Brett, *Theor. Chim. Acta*, 1975, **38**, 121–129.

107. P. George, M. Trachtman, C. W. Bock and A. M. Brett, *J. Chem. Soc. Perkin Trans. 2.*, 1976, 1222–1227.

108. J. Hess, B. A. and L. J. Schaad, *J. Am. Chem. Soc.*, 1983, **105**, 7500–7505.

109. B. Ruscic, R. E. Pinzon, M. L. Morton, G. von Laszevski, S. J. Bittner, S. G. Nijsure, K. A. Amin, M. Minkoff and A. F. Wagner, *J. Phys. Chem. A*, 2004, **108**, 9979–9997.

110. F. Jensen, *Introduction to Computational Chemistry*, John Wiley & Son, Ltd, Chichester, West Sussex, UK, 3rd edn, 2017.

111. F. Jensen, *Essentials of Computational Chemistry*, John Wiley & Son, Ltd, Chichester, West Sussex, UK, 2nd edn, 2004.

112. L. A. Curtiss, P. C. Redfern and K. Raghavachari, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2011, **1**, 810–825.