Theoretical Investigations of

Radical-Mediated Protein Oxidation

Geoffrey P. F. Wood

BSc (Hons), University of Canterbury

A thesis submitted in partial fulfilment for the degree of Doctor of Philosophy of the University of Sydney

School of Chemistry, University of Sydney
July 2006

Te kai rapu, ko ia te kite
Tama tu, tama ora, tama moe, tama mate.
Takoto kau ana te whanau a Tane.
He manga wai koia kia kore e whitikia

Declaration

The work described in this thesis, to the best of my knowledge, is original and does not contain material, except where appropriately referenced, that has been published or presented by any other person, nor has it been submitted for a degree or diploma at any other University or College.

.....

Contents

Acknowledgements

The following is an excerpt from "Tout est Bien", an article in Voltaire's Philosophical Dictionary, which has hung over my workspace for the last three years. It reminds me to persevere; that things have an order and by design will come to natural fruition. A contradiction I hear you say! However, think of why the contrary choice is made and you will find yourself in an infinitely turning circle.

"Here we have a clear and fixed order among every kind of animal. There is order everywhere. When a stone is formed in my bladder it is by means of admirable mechanics: calculous juices pass little by little into my blood, they filter into the kidneys, pass through the ureters, deposit themselves in my bladder, and assemble there by an excellent Newtonian attraction; the stone is formed, gets bigger, I suffer pains a thousand times worse than death, by the most elegant arrangement in the world. A surgeon, having perfected the art invented by Tubalcain, comes to thrust a sharp and cutting iron into the perineum, and takes hold of my stone with his pincers. It breaks under his efforts by a necessary mechanism; and by the same mechanism I die in frightful torments. All this is good, all this is the evident consequence of inalterable physical principles. I agree with them, and I knew it as well as you did."

I have crossed the Rubicon! This journey has had its share of vicissitudes and I have needed good friends and excellent guidance to arrive where I am.

Firstly and foremostly, I would like to thank my supervisor, Professor Leo Radom. He has provided me with the problem, the tools, and the solution. Throughout the last four years, he has passed on his singular wisdom and direction to the extent that I have been able to combine all three of these aspects together.

The environment Leo has provided allowed me the privilege to work among highly-talented scientists, some of whom are on the threshold of successful academic careers and others who are already on that road. In particular, I would like to thank Dr Michelle Coote, who has introduced me into her "Leibniz-like" world, which I only aspire to replicate. Drs Tony Scott, David Henry, and Michael Sullivan were invariably patient enough in answering all of my early questions when I felt as though I was stranded on a highway with a flat tyre and a broken stiletto. I would like to extend my gratitude to them for stopping. I would also like to thank my

peer Gregory Sandala; he has helped in sharing the heavy load that a PhD can be. Drs Bo Durbeej, David Graham, Damian Moran, and Bun Chan have passed on their help and advice when I have needed it, and have always had their "doors" open when I came knocking. I would like to extend my sincerest thanks to them all.

I must thank all of the staff at the Supercomputing facility in Canberra, who have assisted with the technical difficulties arising throughout this PhD work. In this regard I would like to make particular mention of Dr Rika Kobayashi. I would like to thank the School of Chemistry in Sydney and the Research School of Chemistry in Canberra, which have provided me with space. I have had generous support from funding bodies, and, in particular, the Australian Postgraduate Award, the Agnes Campbell prizes, the Graduate School Scholarship from the Australian National University, the Supplementary Scholarship from the Research School of Chemistry, and the Alan Sargeson Merit scholarship.

Outside the Radom group and my place of work, I have been fortunate to meet a number of good friends and allies. They are: from Canberra, Adam, Valeska, Matt, Dave, and Colin, and, from Sydney, Matt and Kerry. I am most appreciative of my brother Philip A. C. Wood who has cooked me some of the most exciting and delicious meals that any top-class restaurant in Sydney could provide. I also want to thank Whaea Pauline and Matua Des who have constantly encouraged me to make free and informed choices and provided me with the love and support only parents are able to give.

Finally, I want to thank one of the most wonderful people in the world, who has challenged consistently the way I think and perceive the world, has offered me love and support and given me fun and laughter. She has fed me, clothed me, taken me to see the most wonderful places in Sydney and has driven me to and from University without a sigh or adverse remark. Alexandra I shall always be indebted to you.

List of Publications

Parts of this thesis have been published in peer-reviewed journals:

(1) Wood, Geoffrey P. F.; Henry, David J.; Radom, Leo. **Performance of the RB3-LYP, RMP2, and UCCSD(T) Procedures in Calculating Radical Stabilization Energies for NHX Radicals.** *Journal of Physical Chemistry A* **2003**, *107*, 7985–7990.

(2) Wood, Geoffrey P. F.; Moran, Damian.; Jacob, Rebecca,; Radom, Leo **Bond Dissociation Energies and Radical Stabilization Energies Associated with Model Peptide-Backbone Radicals.** *Journal of Physical Chemistry A* **2005**, *109*, 6318–6325.

(3) Wood, Geoffrey P. F.; Rauk, Arvi; Radom, Leo **Modeling beta-scission Reactions of Peptide Backbone Radicals: Backbone C–C Bond Fisson** *Journal of Chemical Theory and Computation*, **2005**, *1*, 889–899.

(4) Wood, Geoffrey P. F.; Barnes, Ericka C.; Petersson, George A.; Radom, Leo; Frisch, Michael J.; Montgomery Jr., John A. **Formulation and Evaluation of the ROCBS-QB3 Procedure** *Journal of Chemical Physics,* Submitted.

(5) Wood, Geoffrey P. F.; Easton, Christopher J.; Rauk, Arvi; Davies, Michael J. Radom, Leo **The Effect of Side Chains on Competing Pathways for β-Scission Reactions of Peptide-Backbone Alkoxyl Radicals** *Journal of Physical Chemistry A*, Submitted.

In addition, a number of parallel projects were carried out associated with the work presented in this thesis:

(6) Coote, Michelle L.; Wood, Geoffrey P. F.; Radom, Leo. **Methyl Radical Addition to C:S Double Bonds: Kinetic versus Thermodynamic Preferences.** *Journal of Physical Chemistry A* **2002**, *106*, 12124–12138.

(7) Moran, Damian; Jacob, Rebecca; Wood, Geoffrey P. F.; Coote, Michelle L.; Davies, Michael J.; O'Hair, Richard A. J.; Easton, Christopher J.; Radom, Leo **Rearrangements in Model Peptide-Type Radicals via Intramolecular Hydrogen-Atom Transfer** *Helvetica Chimica Acta,* Accepted.

(8) Wood, Geoffrey P. F.; Rintelman, Jamie, M.; Gordon, Mark S; Radom, Leo **Solvation of the Glycyl Radical** *in preparation.*

(9) Wood, Geoffrey P. F; Radom, Leo; Smith, David M. **The Nature of Glycine and its α-Carbon Radical in Aqueous Solution: A Theoretical Investigation** *in preparation.*

Summary

This thesis primarily details the application of high-level *ab initio* quantum chemistry techniques in order to understand aspects of free-radical mediated protein oxidation. Traditionally, product analysis and electron paramagnetic resonance (EPR) spectroscopy are the primary means for elucidating the chemistry of protein oxidation. However, in experiments involving relatively small proteins reacting with a controlled radical-flux, a vast array of compounds can be produced, which are often difficult to analyse. Quantum chemical techniques on the other hand, can calculate the properties of any particular species directly, without suffering from the problems associated with experiment, such as side-reactions and chain processes. The results presented in this thesis are aimed at elucidating mechanistic details of protein oxidation, which might otherwise be difficult to probe experimentally.

Chapter 1 gives an overview of the free-radical hypothesis of disease and ageing. Protein-derived radicals can undergo a variety of reactions, with the particular reaction that occurs depending on numerous aspects. Many types of reactions have been identified through radiolysis experiments of amino acids, and these are detailed in this chapter. In addition, the key reactive species are characterized and their different chemistries explained.

Chapter 2 details the theoretical tools used throughout this thesis. Species with unpaired electrons (radicals) present unique problems for quantum chemistry to handle, thus an appropriate choice of theoretical technique is needed. The approach taken in this thesis is to use high-level compound methods, many of which have been directly formulated to give improved results for radical species, to provide benchmark quality results by which other less demanding techniques can be assessed.

During the course of this study, it became apparent there was a void in the armoury of tools that could be used for the theoretical chemistry calculations. Chapter 3 details the formulation of a new tool in an attempt to fill this gap. Historically, the formulation of this new procedure came after much of the work in this thesis had been carried out. Thus, for the study of many of the reactions of this thesis the new method has not been used. However, it is most appropriate to place its formulation after summarizing the current status of techniques in common use today.

Chapters 4 and 5 detail computations carried out on models of peptides containing backbone carbon- and nitrogen-centered radicals. A number of different theoretical techniques are used in these chapters, ranging from the highly accurate and computationally intensive to the less reliable and less demanding. The highly accurate techniques are used to gauge the accuracy of the other less demanding theoretical techniques so that the latter can be used with confidence in larger systems. Not only is the choice of theoretical technique important but also the judicious choice of model is essential. With this in mind, models are incrementally built until convergence of the particular property of interest is reached.

Chapters 6 and 7 detail the calculations of β-scission reactions of alkoxyl radicals, which are a particular class of reaction known to occur on peptide backbones. Alkoxyl radicals are particularly difficult for theory to describe correctly. Therefore, Chapter 6 extensively assesses and then identifies the theoretical methods needed to portray them. Chapter 7 uses the techniques identified in the previous chapter in order to predict how the preference for a particular type of β-scission reaction changes.

List of Tables

List of Figures

List of Abbreviations

The following are the most common abbreviations used in this work.

Theoretical Procedures:

HF	Hartree-Fock theory
MP*n*	*n*th order Møller-Plesset perturbation theory
CCSD(T)	Coupled-cluster with singles and doubles and a perturbative treatment of triples
B3-LYP	Becke's three-parameter density functional model (B3), incorporating the Lee, Yang and Parr (LYP) gradient correction to the correlation energy
BMK	Density functional model formulated by Boese and Martin
MPWB1K	A combination of a modified Perdew and Wang exchange functional and the B1K correlation functional formulated by Truhlar and co-workers.
U/R	Each of the methods above have been used with either a restricted (designated by an R prefix) or unrestricted (designated by a U prefix) wave function.
G3X(MP2)-RAD	The RAD variant of the G3X(MP2) compound method
W1	The Weizmann-1 compound method for benchmark calculations
CBS	Complete basis set model chemistries formulated by Petersson et al.
CBS-QB3	The QB3 variant of Petersson's complete basis set methods
U-CBS-QB3	The QB3 variant of Petersson's complete basis set methods excluding the empirical correction for spin-contamination
ROCBS-QB3	The QB3 variant of Petersson's complete basis set methods based on restricted open-shell methods

Amino Acids:

Ala	Alanine	Leu	Leucine
Arg	Arginine	Lys	Lysine
Asp	Aspartate	Met	Methionine
Asn	Asparagine	Phe	Phenylalanine
Cys	Cysteine	Pro	Proline
Gln	Glutamine	Ser	Serine
Glu	Glutamic acid	Thr	Threonine
Gly	Glycine	Trp	Tryptophan
His	Histidine	Tyr	Tyrosine
Ile	Isoleucine	Val	Valine

CHAPTER 1

Radical-Mediated Protein Oxidation

1.1 Introduction

In 1977 the doublet electron paramagnetic resonance (EPR) spectrum of the ribonucleotide reductase (RR) protein from *Escherichia coli* was assigned to a tyrosine residue[1] following a number of failed experiments attempting to ascribe it to iron. Since then numerous other protein-bound free radicals have been identified. Many of these radicals are characterized by their involvement within a catalytic cycle. They can be persistent, as in the case of the tyrosine radical of RR, or transient, as in the case of the 5′-deoxyadenosyl radical involved in vitamin B_{12} chemistry.[2]

Great advances in the knowledge of biological processes involving free radicals have occurred since 1977. However, experimental investigation into free radicals can be plagued with technical problems. For instance, even the relatively well established but transient 5′-deoxyadenosyl radical of vitamin B_{12} chemistry has never been directly observed, having only been identified through stabilization effects of ring substitution.[3] A theoretical investigation of such a species on the other hand does not suffer from the same problems because properties may be directly calculated. In addition, theoretical studies of specific reactions can be directly modelled in the absence of side reactions. This gives theory a unique advantage over experiment and it has enabled significant insights to be obtained into the nature of such biological processes. Theory is now regarded by many as a necessary companion to experimentation for the understanding of such things as the structures, mechanisms and dynamics of not only biological systems but many facets of chemistry.

Radom and his co-workers have contributed significantly to the theoretical elucidation of vitamin B_{12} chemistry.[4] During the process of uncovering mechanistic aspects of B_{12} chemistry, Radom and his co-workers have adopted a 'model' approach to investigation, whereby highly accurate *ab initio* molecular orbital theory is used to model small systems, perhaps only 3 or 4 atoms from the first row of the periodic table, on the basis that such systems contain the essential chemistry. This allows theoretical

benchmarking to be undertaken so that other less computationally demanding theoretical techniques can be applied with confidence to larger model systems. The model system approach has been employed successfully to the chemistry of vitamin B_{12} and will be used throughout this thesis to assist understanding aspects of radical-mediated protein damage.

A substantial development in the understanding of free radicals involved in protein damage has taken place in the last quarter of the 20th century. Since the late 1960's and early 1970's there has been growing evidence that radical-mediated oxidative damage to proteins is of primary importance in many diseases such as Alzheimer's disease, atherosclerosis and diabetes as well as ageing.[5,6] Some researchers have argued that the occurrence of free radicals at the site of disease was peripheral to the cause of the disease but it is now understood that free radicals have a vital role to play in many disease pathologies.

Despite the immense advancement in knowledge, additional experimental problems can be associated with this particular area of biological free radicals. Radical species involved in protein damage are not only transient but exist in environments uncontrolled by enzymes. In addition, an initiating radical can cause damage to multiple sites on a protein because products of radical reactions are usually radicals themselves. Therefore a number of reactions may occur before the product radical is quenched. This can cause the examination of EPR experiments and product analysis, the two major techniques used in this area of research, to rapidly become unwieldy. Because experimental impediments have prevented the exact mechanistic understanding of radical-mediated protein oxidation it is hoped that theoretical chemistry calculations will supply some of the information that is currently lacking.

The remainder of this chapter will outline, in the first instance, the basis on which the importance of radical-mediated protein oxidation was established (§1.2). There will then be an outline of the current experimental and theoretical knowledge of

specific processes involved in radical-mediated protein oxidation (§1.3 and §1.4). Finally, there will be a brief overview of how radicals are involved in the pathology of some diseases (§1.5).

1.2 Superoxide Theory of Oxygen Toxicity

The importance of free radicals in disease has been attributed through the implications of the superoxide theory of oxygen toxicity. In aerobes, oxygen (O_2)[7] plays a vital role for the generation of ATP by utilizing the gradual release of energy in mitochondrial electron transport chains. However, overexposure to O_2 may cause serious problems. In humans the alveoli may be severely damaged by oxygen concentrations of over 50%. Anaerobes are more sensitive to the presence of O_2. In some cases brief exposure can kill anaerobes immediately (e.g. *Treponema denticola* involved in dental plaque), while in others a 10% concentration is able to be tolerated but inhibits growth (e.g. *Bacteroides fragilis*, bacterium in the colon).[8] The chief reason for oxygen toxicity is that it interferes with enzyme activity which, in turn, puts many organisms in a precarious position. On the one hand, O_2 is required for the production of energy in order to drive cellular processes yet O_2 can inhibit vital functioning. This is known as the oxygen paradox.[9] Over the last thirty years, there has been some controversy about the exact nature of oxygen toxicity and it is still not fully understood today. However, there is considerable evidence to support the aspects of oxygen toxicity outlined below.

1.2.1 Superoxide and Superoxide Dismutase (SOD)

Superoxide ($O_2^{\bullet -}$) is formed from a single electron reduction of O_2. This species can be created by a number of endogenous processes including *inter alia* the autoxidation of haemoglobin, myoglobin, and cytochrome *c*, and within the catalytic cycles of xanthine oxidase and aldehyde oxidase.[8] There have only been a few reports of the direct influence of superoxide on protein degradation, for example, glutathione

peroxidase.[10] It does not react with lipids or DNA but it can oxidize the heme group of haemoglobin[11] and the iron-sulfur cluster of the dehydratases, which subsequently releases Fe(II).[12] Even though superoxide itself does not detrimentally harm cellular processes, there is a group of enzymes abundant in all eukaryote cells known as superoxide dismutase (SOD) that catalyses the dismutation reaction of superoxide:[13]

$$O_2\bullet^- + O_2\bullet^- + 2H^+ \rightarrow O_2 + H_2O_2 \qquad (1.1)$$

This seems to be the *only* function this group of enzymes has. Catalysing the dismutation of superoxide at the expense of forming hydrogen peroxide presents strong evidence that it is a deleterious species and must be disposed of. Fortunately there are a number of other enzymes known as catalases and peroxidases that are dedicated to the elimination of H_2O_2. Even more interesting is the fact that the SODs and the catalases and peroxidases work synergistically in their removal of both species.

The SODs exist in a number of forms that may be characterized based on the metals ions resident in the active sites. CuZnSOD is found in all eukaryote cells and has the highest catalytic activity of all the SODs for the dismutase reaction. It is also remarkably stable, fairly resistant to heating, and safe from attack by proteases, and to denaturation by compounds such as urea, guanidinium chloride and sodium dodecyl sulphate.[14] Other forms include MnSOD, which is found predominately in bacteria such as *E. Coli* but is also widespread in plants and animals.[14] *E. Coli* contains two other SODs, an iron-containing one, FeSOD, and a cambialistic form that is active with either Fe or Mn.

The immediate question that arises from the discovery of SODs and their specificity for $O_2\bullet^-$ must be how can superoxide be cytotoxic when it is comparatively unreactive in physiological conditions? The protonated form of superoxide ($HO_2\bullet$) is more reactive and can initiate peroxidation of fatty acids[15] but with a pK_a of 4.8 the

concentration of HO_2• is relatively low. It has been argued, however, that mitochondrial membranes have large pH gradients allowing HO_2• to be formed in significant concentrations.[16] Nevertheless this still does not seem to account for the cytoxicity of superoxide.

1.2.2 Generation of •OH

It has been proposed that potentially one of the most damaging species derived from superoxide is the hydroxyl radical (•OH).[17] For reaction of •OH with amino acids, peptides and proteins the reaction rate constants are very large ($\sim 10^{7-10}$ dm^3 mol^{-1} s^{-1}).[5] Comparing, for instance, the rate constants of O_2•$^-$ and •OH reacting with free[18] valine which are <0.18 and 7.6×10^8 dm^3 mol^{-1} s^{-1}, respectively, it quickly becomes apparent that generation of •OH from superoxide could lead to significant damage to proteins, DNA, and lipids alike.

It is well known that •OH can be produced in large quantities via a catalytic cycle from O_2•$^-$ and H_2O using an iron(III) complex as the catalyst. This cycle is known as the iron catalysed Haber-Weiss reaction[19] or the superoxide-assisted Fenton reaction:

$$Fe^{3+} + O_2\bullet^- \rightarrow Fe^{2+} + O_2 \qquad (1.2)$$

$$Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + OH^- + \bullet OH \qquad (1.3)$$

Other metal ions could potentially catalyse this reaction and *in vitro* studies using Cr^{5+}/Cr^{6+},[20] Ni^{+}/Ni^{2+},[21] Co^{+}/Co^{2+},[21] Cu^{+}/Cu^{2+},[22] and V^{4+}/V^{5+} [23] have been undertaken. Such reactions may be important in metal poisoning but most studies for *in vivo* catalysis have focused on Fe^{2+}/Fe^{3+} and to a lesser extent Cu^{+}/Cu^{2+} systems. As long as these metals are available in oxidative/reductive complexes, it seems reasonable to suggest that •OH radicals generated in this fashion could potentially cause significant damage to biologically important species. Numerous studies[24] have debated the

possibility of available iron (or other metals) for this key step in the oxygen toxicity hypothesis.

Along with the question of metal availability a number of researchers have suggested it is not in fact the hydroxyl radical that is formed in the Fenton reaction but some sort of ferryl (Fe^{4+}) species[25] or a "crypto-•OH"[26] radical – a species that mimics free •OH. Although an intermediate ferryl species has yet to be ruled out, the weight of evidence suggests that the •OH radical must be the eventual reactive species because free •OH produced in radiolysis experiments gives identical results to Fenton-derived •OH experiments.[8]

1.2.3 Sources of Available Iron

Iron available for generation of •OH *in vivo* is very scarce. However the generation of $O_2•^-$ and H_2O_2 can cause situations where production of available iron is favorable. For instance $O_2•^-$ can oxidize iron–sulfur clusters releasing iron while H_2O_2 can degrade mammalian heme proteins also releasing iron. Quite often, however, the greatest source of catalytically available iron comes from situations where tissue has been damaged, a recurrent theme among many radical-dependent diseases.

There is a high concentration of iron in many biological systems the presence of which has to be extensively regulated. The protein transferrin transports iron in the blood plasma, binding it with a high affinity at physiological pH. There is usually more transferrin than there is iron in the plasma, effectively reducing the concentration of free iron in the plasma to zero. At pHs lower than normal (<6), iron can be mobilized from transferrin and it is thought that this may cause damage to cartilage in rheumatoid arthritis.[8] Otherwise transferrin sequesters iron so that it probably is not involved in Fenton chemistry. It may be that the high presence of these proteins in blood plasma and the concomitant low concentrations of SODs and catalases create an effective antioxidant defence system.

Haemoglobin is another obvious source of iron. It is known that bleeding into sites of inflammation exacerbates tissue injury, possibly through radical reactions involving •OH. It is not however the haemoglobin itself that is responsible for the catalytic formation of •OH but the iron released through reactions with chelating agents or with H_2O_2.[27]

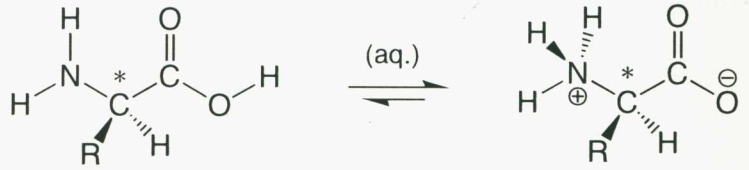
Most intracellular iron is bound up in proteins such as those described above. There is however, a small percentage present in a labile iron pool. The iron in the labile pool is usually chelated to low molecular weight molecules such as citrate, phosphate, ADP or ATP. These complexes can easily generate •OH through Fenton chemistry involving the $H_2O_2/O_2^{\bullet -}$ system.[28] It has been demonstrated that if the iron uptake mechanisms are altered so as to increase the concentration of iron in the pool, then there is also an increased susceptibility to injurious processes associated with radical-mediated oxidation.[29]

1.3 Damaging Reactions

The diverse range of chemical moieties present in the 20 common amino acids leads to an enormous variety of chemistry from initiation by a radical such as •OH. Reactions that have been documented include: hydrogen abstraction, electron transfer, addition, fragmentation and rearrangement, dimerization, disproportionation, and substitution (concerted addition and elimination). Many of these reactions have been studied intensively through radiolysis experiments of the free amino acids in aqueous solution and there are a large number of reviews on the subject.[5,6,30] The situation in peptides and proteins can differ from that of the free amino acids so this information is not always directly transferrable. The following section gives a brief overview of the important reactions and properties of species derived from •OH reacting with free amino acids. Where it is appropriate, similarities and differences will be noted for peptides and proteins.

1.3.1 Chemistry of the α-Carbon

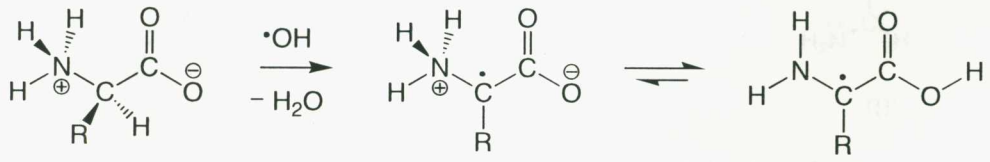
In an aqueous medium and at physiological pH, the free amino acids exist in a zwitterionic form. The equilibrium constant for free glycine ([zwitterion]/[neutral]) at 296 K is 2.7 x 10^5 which means that the zwitterion is about 30.5 kJ mol^{-1} more stable than the neutral (* is the α-C).[31]



neutral zwitterion

The presence of the positively charged amino group on the zwitterion causes the rate of α-C–H abstraction to be somewhat slower than those for peptides, where the zwitterion is prevented from forming. For instance the rate constant for •OH reacting with free glycine is 1.7 x 10^7 dm^3 mol^{-1} s^{-1} (pH range of 5.8 – 6.0).[5] By acetylating Gly, as in *N*-Ac-Gly, the rate constant for reaction with •OH becomes 4.0 x 10^8 dm^3 mol^{-1} s^{-1}, for a pH range of 6.6 – 8.7.[5] Further evidence for the deactivating effect of the positively charged free-amino group may be observed in dipeptides (e.g. *N*-Ac-Gly-Gly) where abstractions at the α-carbon next to the protected amino group (Ac–NH–αCH_2–) is preferred over the α-carbon next to the free amino group (–αCH_2–NH_3^+). The preference may be removed by derivatization of the *N*-terminal free amine group.

After abstraction, the zwitterionic form of the α-C-centered radical is no longer favored and the neutral predominates in solution. This could lead to an interesting reaction sequence, i.e.,



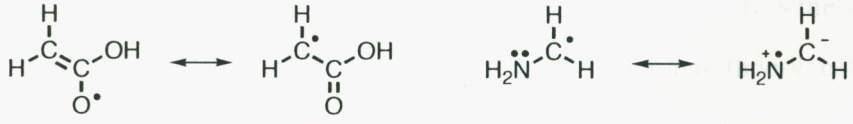
The solvent would have an important role to play in this sequence. It is not known if the solvent would cause the hydrogen-atom transfer to occur concurrently with the abstraction or subsequent to it. In the case of closed-shell glycine, Car-Parinello molecular dynamics studies have revealed that the free energy of activation for hydrogen-atom transfer from the zwitterion to the neutral is approximately 70.5 kJ mol^{-1} (40.0 kJ mol^{-1} in the reverse direction).[32] An equivalent study for the radical has not yet been reported.

1.3.2 The Captodative Effect

The predominance of the neutral amino-acid radical in solution is most likely because of the additional stability it gains from what has been termed the captodative effect.[33] This stability is gained from the synergistic delocalisation of the unpaired electron through the adjacent electron-withdrawing carbonyl and electron-donating amino groups. In valence bond theory, the origin of the captodative effect can be explained in terms of a greater number of resonance-hybrid contributors in the disubstituted radical. For example, the resonance contributor IV in the glycyl radical:

I II III IV

does not have a counterpart in the resonance contributors for the corresponding monosubstituted radicals, the carboxymethyl radical and the aminomethyl radical:



I' II' II'' III'

In terms of molecular orbital theory, the appropriate orbital interactions are given in Figure 1.1. Figure 1.1(a) demonstrates the interaction of the lone-pair orbital on the

nitrogen of NH_2–•CH_2 with the singly-occupied 2p(C•) orbital on carbon, lowering the energy of the lone-pair orbital and raising the energy of the 2p(C•) orbital, resulting in a net stabilization. Figure 1.1(b) shows that the raised 2p(C•) orbital interacts more effectively with the π^* orbital on the carbonyl giving an increased stabilization. Chapter 5 investigates how each group contributes to this stability and identifies the conditions required for the onset of the additional stability.

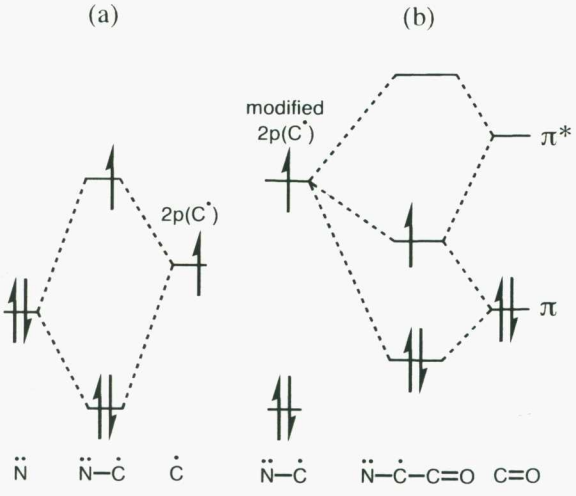


Figure 1.1: Orbital interaction diagram showing **(a)** the interaction of the singly-occupied 2p orbital on the carbon, 2p(C•), with the lone pair on the nitrogen in NH_2–•CH_2; and **(b)** the enhanced interaction of the modified 2p(C•) orbital with the π^* orbital on the carbonyl group in NH_2–•CH–CHO.

The captodative effect is maximized through co-planarity of the radical-center with the donor and acceptor groups. In a study by Rauk and co-workers,[34] the α-C–H bond dissociation energies (BDEs) for all 20 common amino acids were investigated. Using a truncated model peptide of the form $HCONHCH(R)CONH_2$ where R is the side chain, they determined all the α-C–H BDEs using density functional theory calculations. They found that the BDEs ranged from 326–358 kJ mol^{-1}. Figure 1.2 gives a representation of the model peptide containing an alanine residue used in that study.

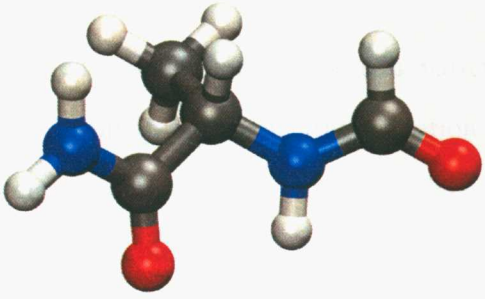


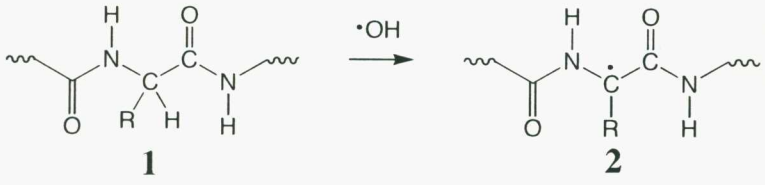
Figure 1.2: Model of peptide backbone containing an alanine residue that was used in the bond dissociation energy study of Rauk and co-workers.[34] (key: grey = C, white = H, blue = N, red = O).

The S-H bond in glutathione is considered to be reasonably weak (367 kJ mol^{-1}) but the BDEs of the α-C-H bonds were found to be less than this.[34] They are however, considerably higher than the O–O bond dissociation energy of H_2O_2 (214.1 kJ mol^{-1}).[35] A C–H BDE in a hydocarbon is much higher. For instance, the C–H BDE of the secondary carbon in C_3H_8 is 412 kJ mol^{-1}. All of the amino acids except for Ill, Val, Thr, and Pro were found[34] to have lower BDEs than Gly because the side-chains act as weak donors and thus stabilize the radical further. The higher BDEs of Ill, Val, Thr and Pro were attributed to the side chain preventing co-planarity of the backbone.

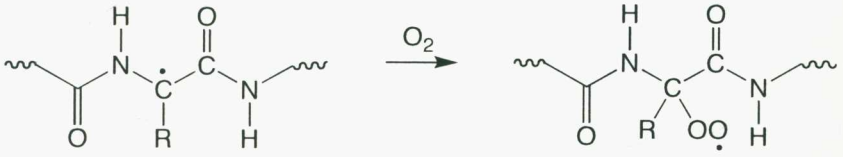
Interestingly, experiments carried out on acetylated dipeptides containing glycine and another residue showed that there was a preference for abstraction by an attacking radical for the α-C of glycine in contrast to the predicted order of stabilities of the resultant radicals based on BDEs.[36] For instance, product analysis of *N*-trifluoroacetyl-Ala-Gly showed a preference for the Gly α-C by a ratio of 7:1. If the residues appear in the reverse order, i.e., trifluoroacetyl-Gly-Ala then the preference ratio for glycine increases to 20:1.[36a] Theoretical calculations have concluded that the preference for abstractions from the α-C of Gly can be attributed to steric interactions preventing other residues from adopting an ideal planar configuration.[37]

1.3.3 Strand Breakage

Once formed, α-centered radicals can undergo a variety of reactions in the presence of oxygen that may result in backbone fragmentation. The consequence of these reactions occurring is significant in protein chemistry since strand breakage would cause major structural damage to a protein in one step. The first two reactions (1.4 and 1.5) show the initial formation of a peroxyl radical **3** from the reaction of oxygen with an α-centered radical.

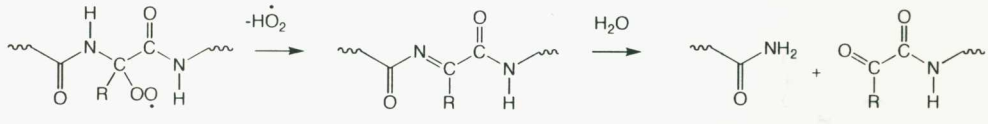


1 **2** (1.4)



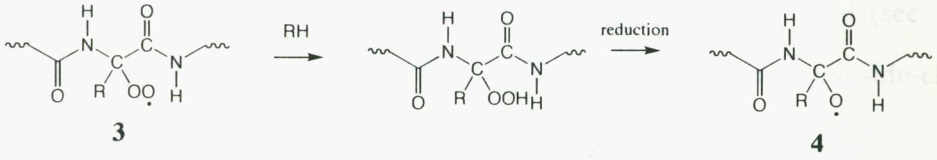
2 **3** (1.5)

The sequence of reactions in 1.6 show the first major pathway leading to strand breakage. The peroxyl radical **3** loses HO_2• to form an imine, and the imine is then hydrolysed to form an amide and an oxyamide. The net result of the reaction is to break the backbone α-C–N bond. Pulse radiolysis experiments on glycine-derived peptides have detected the presence of HO_2•, supporting the presence of this pathway.[38]



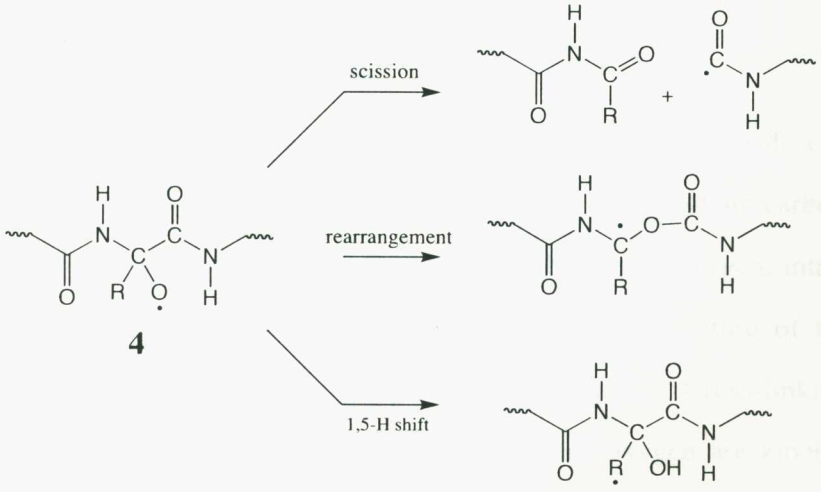
3 (1.6)

The second major pathway, involves the formation of an alkoxyl radical **4** from a hydroperoxide, as shown in the reaction sequence 1.7:



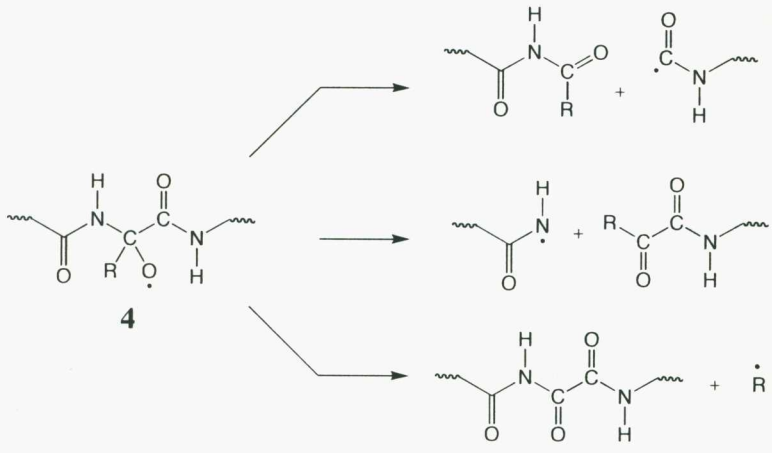
3 **4** (1.7)

Alkoxyl radicals can be involved in several classes of reactions. Experimental atmospheric and solution chemistry studies and theoretical computations have shown that alkoxyl radicals can *inter alia* undergo three classes of reactions, namely β-scission, skeletal rearrangements, and 1,5 hydrogen shifts.[39,40] It might be expected that peptide backbone alkoxyl radicals will undergo the same types of reactions, as shown in reactions 1.8, 1.9 and 1.10, respectively, with reaction 1.8 leading to strand breakage. Experimentally it is difficult to obtain quantitative information on the individual reactions, though some experimental evidence has been obtained for the occurrence of reaction 1.8. [40] This is because they are fast and may involve chain processes which makes it problematic to isolate the particular reaction of interest.



4 (1.8) (1.9) (1.10)

An additional degree of complexity arises because each of the three classes of reactions can, in principle, undergo three variations. For example, because there are three different groups joined to the α-carbon, reaction (1.8) is just one of three possible β-scission reactions that can occur, namely β-scission of the C–C bond (see 1.8 and repeated in 1.11), β-scission of the C–N bond (1.12), or β-scission of the side-chain C–R bond (1.13):



4

(1.11)

(1.12)

(1.13)

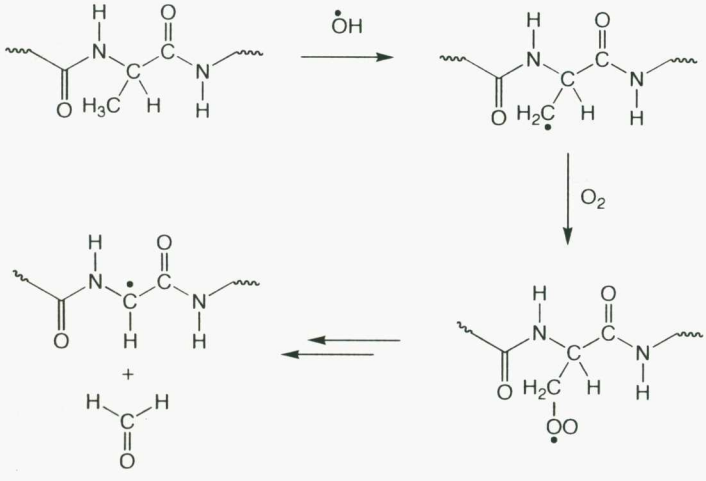
Previous EPR spin-trapping studies[40] have suggested that the process 1.11 predominates, though the occurrence of other reactions could not be discounted. This process is considered important for the fate of proteins under oxidative stress and is considered in detail in Chapters 6 and 7.

1.3.4 Reactions of the Side Chains

Aliphatic side-chains. The amino acids containing aliphatic side chains (Ala, Leu, Ile, and Val) usually react by hydrogen abstraction. The resultant carbon-centered radicals undergo the usual types of reactions. In the absence of oxygen, intramolecular or intermolecular re-abstraction may occur depending on the nature of the original carbon-centered radical, i.e., primary, secondary or tertiary. Cross-linking through radical-radical reactions can occur but in the absence of oxygen are kinetically slow because of low concentrations. The intramolecular abstractions can have consequences

for strand breakage since re-abstraction at the α-C could transfer damage to the backbone.

In the presence of oxygen, carbon-centered radicals react rapidly to form peroxyl species (ROO•) that undergo reactions to form alkoxyl radicals, which decompose to form carbonyl compounds and further radical species. The reaction sequence follows a similar scheme to the fate of α-C hydrogen abstractions as detailed in reactions 1.4, 1.5, 1.7 and 1.8. It has been demonstrated for alanine-containing peptides that hydrogen-atom abstraction at the side-chain methyl substituent (β-carbon) in the presence of oxygen leads to methanal (formaldehyde) formation and an α-centered radical (see reaction sequence 1.14).[41] In that study, methanal formation accounted for 80% of the initiating radicals, confirming this as a major process.



(1.14)

For free amino acids, the presence of the zwitterion and consequently the charged amine leads to hydrogen abstraction occurring distal from the α-carbon. This has been verified in solvent exchange experiments, where there was a direct relationship observed between the amount of exchanged solvent in methyl groups and their distance from the α-carbon, i.e., Leu > Ile > Val > Ala.[42]

Aliphatic heteroatom-containing side chains. For amino acids and peptides containing heteroatoms (Arg, Asn, Gln, Glu, Lys, Ser, and Thr) there is an enhanced selectivity for attack that leads to stabilized radical products, even if the abstraction is carried out by species such as •OH. In the cases of Thr and Ser the hydoxyl group stabilizes the incipient radical if abstraction occurs at the β-carbon. For the free amino acids, the existence of the zwitterion causes a subsequent deamination reaction to occur with loss of ammonium ions.[43]

With the presence of ionized carboxyl groups (e.g. Asp and Glu), there is also a similar effect for favored abstraction at carbons α to the carboxylate (see Figure 1.3).[44]

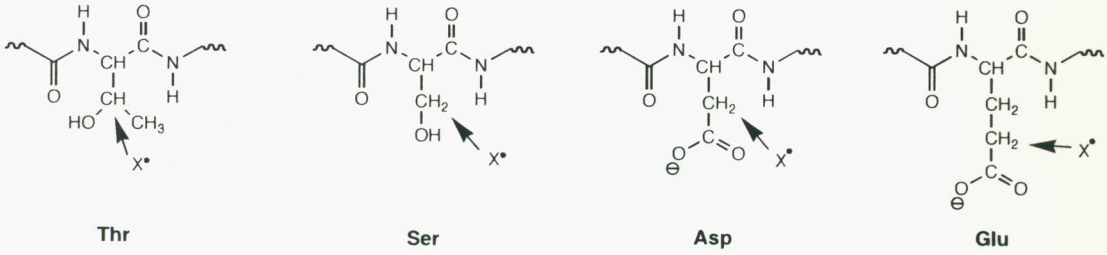


Figure 1.3: Site-directed hydrogen-atom abstraction is caused by the presence of hetero-atoms in the side chains of Thr, Ser, Asp and Glu, where the arrow indicates the preferred site.

Protonation of the amine side chain of Arg or Lys causes abstraction to be favored at sites remote from the site of the amino group (see Figure 1.4).[45] For the free amino acid forms of Arg and Lys, this leads to an interesting competitive effect between the positive charge resulting from protonation at the amino site adjacent to the α-carbon and the amino group of the side chain. The pK_a of the side chain sites are approximately one unit larger than the backbone site, which has a $pK_a \sim 9.5$. This leads to the product distributions favoring abstractions further away from the backbone amino group and nearer to the side-chain amino group. For example, Figure 1.4 gives the

product distributions in solvent exchange experiments involving hydrogen-abstraction by •OH on Arg and Lys.[42]

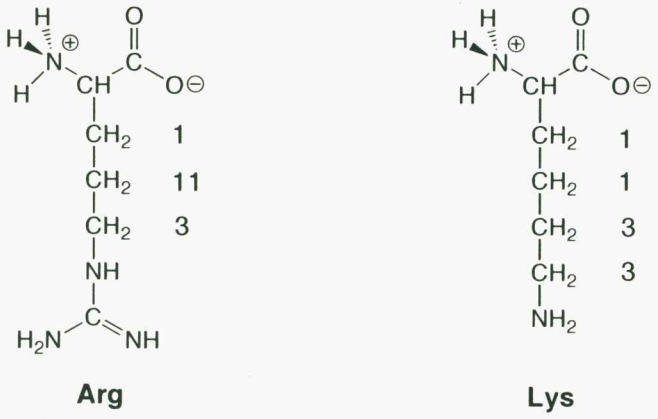
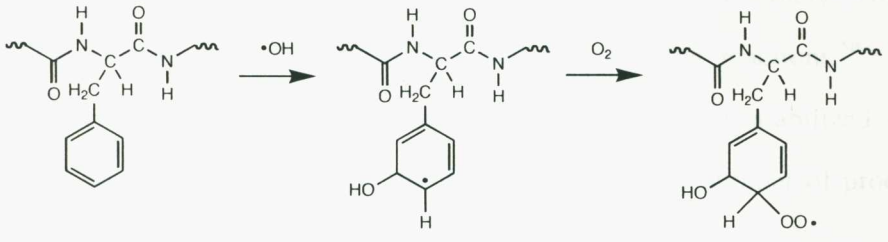


Figure 1.4: Product distributions found in solvent-exchange experiments involving hydrogen-atom abstraction by •OH on Arg and Lys.

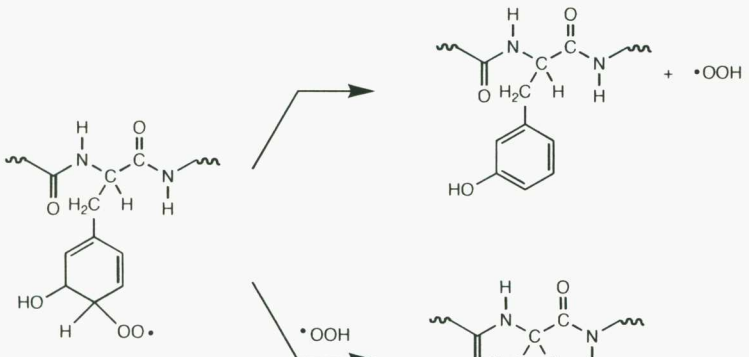
Aromatic side chains. The presence of aromatic side chains promotes addition reactions with •OH radicals, with abstraction becoming a minor process. For phenylalanine reacting with •OH, studies have found no directing effect of the ring substitution and the various proportions of ortho:meta:para adducts were reported as 2.1:1:1.5 and 2.3:1:1.2.[46] The resultant hydroxycyclohexadienyl radicals **5** react with oxygen to yield peroxyl radicals that can either decompose or react with HO_2• to produce tyrosine (see reactions 1.15, 1.16 and 1.17).

In the absence of oxygen, the major product of addition reactions of •OH radicals to phenylalanine is dimerization to give biphenyls.[47] This represents a major pathway for cross-linking reactions in small peptides.[48] At low pH, the hydroxycyclohexadienyl radicals **5**, can lose water to form radical cations. Intramolecular reactions transfer the radical damage to other sites (see reactions 1.18).



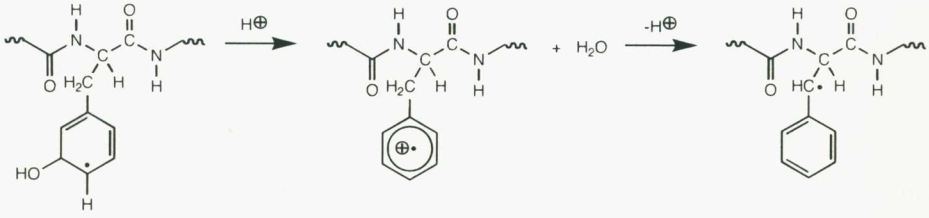
(1.15)

5



(1.16)

(1.17)



(1.18)

Unlike phenylalanine, tyrosine has a strong directing effect for addition reactions of •OH. The major product in the presence and absence of oxygen is DOPA (3,4-dihydroxyphenylalanine). In the presence of oxygen, rapid elimination of HO_2• occurs to yield one DOPA molecule per •OH adduct. In the absence of oxygen, two OH adducts are required to form one DOPA molecule via a disproportionation reaction. In accordance with this, protein oxidation gives DOPA concentrations with higher yields in the presence of oxygen.[49]

•OH additions to the rings of tryptophan and histidine are also known to occur. For Trp, addition can occur on the benzene ring and on the pyrrole moiety. Addition to the pyrrole has been shown to give subsequent ring-opening reactions.[50] For His, addition to the imidazole ring at carbons 2, 4 and 5 results in stabilized allyl-like radicals.[51] In the presence of oxygen, the His adducts form a number of products but these have been difficult to characterize.

Sulfur-containing side chains. The major locus of attack by an initiating radical for sulfur-containing side chains is the sulfur center, which leads to thiol radicals (RS•) via a hydrogen-atom abstraction reaction. In the presence of oxygen, thiol radicals react to form peroxyl species, which can be involved in a chain process with RSH species. At physiological pH, reactions of the anion (RS^-) can be important. The anions can react with the thiol radical to form disulfide linkages.[30]

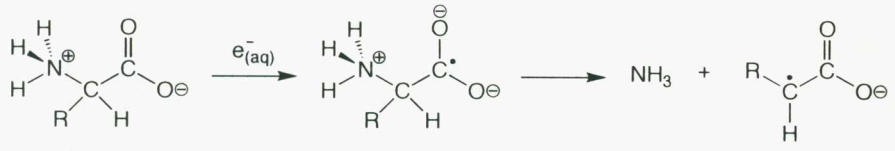
1.4 Other Damaging Species

Besides •OH, there are a number of other species formed *in vivo* via exogenous and endogenous processes that can oxidize proteins. The most important of these are: solvated electrons ($e^-_{(aq)}$), singlet oxygen ($^1O_2$) and hypochlorous acid (HOCl).

1.4.1 Solvated Electrons

Much of the discussion up to this point has focussed on initiating radicals and oxidants that have been generated by endogenous processes. Exogenous processes such as the interaction of high-energy radiation with water has the potential for creating damaging species. One such species is the solvated electron. Discovered in 1962,[52] the solvated electron has a reduction potential of –2.84 V, making it the most potent reducing agent found in biological systems. The chemistry of the solvated electron is dominated by addition reactions rather than abstraction reactions. Rate constants for reaction of the solvated electron with amino acids are large, ranging from $\sim 10^6$ to $\sim 10^{10}$ $dm^3\ mol^{-1}\ s^{-1}$.

For the free amino acid form of glycine, the •OH radical preferentially abstracts the hydrogen situated on the α-carbon. Introduction of alkyl side chains causes preferential abstraction to occur from hydrogens distal to the α-carbon, with methine carbons favored over methylene carbons, which are in turn favored over methyl carbons. The basic rationalization of this behavior can be thought of in terms of the electrophilic nature of •OH. Thus electron-poor sites such as α-carbons β to a protonated amino group are less favorable. For the solvated electron ($e^-_{(aq)}$), markedly different behavior is observed. The most favorable reaction of $e^-_{(aq)}$ with amino acids is addition to the carbonyl double bond yielding a radical anion that dissociates to form a carbon-centered radical,[53] i.e.,



(1.19)

Addition to the amine is also favorable,[54] which leads to a reductive deamination and equivalent products i.e.,

$$e^-_{(aq)} + NH_3{}^+CH(R)COO^- \rightarrow NH_3 + \bullet CH(R)COO^- \qquad (1.20)$$

These reactions are still favored for amino acids containing aliphatic side chains. For aromatic side chains, addition to the ring may also occur.

1.4.2 Singlet Oxygen

Singlet oxygen can exist in two states; the $^1\Sigma_g{}^+$ state, where the two highest electrons are in two different spatial orbitals with anti-parallel spins, and the $^1\Delta_g$ state, where the two highest electrons are in the same spatial orbital. Excitation of O_2 into singlet states can be achieved by several pigments, such as those within the lens of the

eye[55] or in chloroplasts.[56] Excessive generation of singlet oxygen can occur in some diseases,[57] where an overproduction of porphyrins in the skin can transfer excitation energy to O_2.

Singlet oxygen can react either chemically or through physical quenching of the excited state. With proteins, peptides, and amino acids, most reactions occur chemically, with only Trp giving significant physical quenching.[58] Chemical reaction with Trp occurs with a similar rate constant to physical quenching ~ 10^7 dm^3 mol^{-1} s^{-1}. The rate constants with the side chains of the other amino acids have more variance. At physiological pH, reaction with His, Tyr, Met, Cys, Trp, and cystine is fast with rate constants of ~ 10^7 dm^3 mol^{-1} s^{-1}. The side chains of the remaining residues react with rate constants of < 0.7 x 10^7 dm^3 mol^{-1} s^{-1} [59] and there is a strong dependence on pH.

1.4.3 Hypochlorous Acid

HOCl is a strong two-electron donor produced by the enzyme myeloperoxidase during respiratory bursts of neutrophils, i.e., release of reactive oxygen species in white blood cells.

$$H_2O_2 + Cl^- \rightarrow HOCl + OH^- \qquad (1.21)$$

HOCl has a pK_a of 7.54, and therefore at physiological pH it exists as a mixture of HOCl and OCl^-. It has been shown to inactivate the proteins α_1-antiproteinase and thrombomodulin by reacting with methionine residues. It can also oxidize thiols, ascorbate, NADH, NADPH, and lead to chlorination of DNA bases and tyrosine residues. It has also been shown that HOCl can react with superoxide or with Fe(II) to generate the hydroxyl radical in a reaction similar to the Fenton reaction.

HOCl reacts more slowly with peptides and amino acids than HO• but the variation in rate constants has been shown to be large.[60] For instance, it reacts quickly with the sulfur-containing Met and Cys residues with rate constants of approximately 3

x 10^7 M^{-1} s^{-1} but more slowly with Gln and Asn residues with rate constants of approximately 0.03 M^{-1} s^{-1}.

HOCl predominately reacts with free amino acids to form unstable chloramines. The chloramines undergo rapid decomposition to form nitrogen-centered radicals. It is postulated that inter- or intra- molecular hydrogen-abstractions would occur from the nitrogen-centered radicals to form more stable carbon-centered radicals and there has been some direct evidence for this process.[61]

1.5 Diseases and Disorders

Free radicals have been implicated in a number of human diseases. A limited list includes heart and cardiovascular diseases such as: atherosclerosis, myocardial infarction, angina and strokes, diseases of the brain and nervous system including: Alzheimer's disease, Parkinson's disease, multiple sclerosis and Huntington's chorea, and in other diseases such as diabetes and cancer, and they have been implicated in ageing itself.[8]

Many of these diseases commence as a result of tissue injury and not as a consequence of oxidative stress. However, the initial injury usually increases oxidative stress and whether this serves to promote the disease, through further tissue damage, or is no more than an epiphenomenon, appears to depend on the particular disease. The following diseases provide the greatest evidence for the direct involvement of free radicals in their pathologies.

1.5.1 Atherosclerosis

Atherosclerosis is a disease of the arteries characterised by the thickening of the arterial wall. In many instances it is the underlying contributor of many other conditions such as angina, myocardial infarction (heart attack), and stroke. In the western world it has a high prevalence. For example, in 2001 17% of Australia's

population (3.2 million people) were reported to have cardiovascular disease as a long-term condition.[62]

Atherosclerosis has an early onset starting in childhood and progressing with age. In general, three stages of arterial-wall thickening are recognized. Fatty streaks are caused by *foam cells* and are an early sign of atherosclerosis characterized by raised yellow areas in the arterial wall. The fatty streaks are usually precursors to fibrous plaques which are larger rounded lesions about a centimeter in diameter. Finally, there are complicated plaques. These plaques are modified fibrous plaques that are large and restrict blood flow to the tissues of vital organs.

It is believed that atherosclerosis begins by damage to the endothelium membrane, the membrane on the inside wall of the artery. Attachment by monocytes to the damaged area is then thought to proceed. The monocytes can develop into macrophages within the vessel wall. Further tissue damage is then caused by the activated macrophages and monocytes by release of $O_2\bullet^-$ and H_2O_2 into neighboring areas. It is thought that, once the secretion of these species occurs, oxidized low-density lipoprotein is taken up with enhanced efficiency by macrophages causing them to become the *foam cells* of fatty streaks.

1.5.2 Alzheimer's disease

Alzheimer's disease is classified differently from the normal dementia of ageing. It is characterized by the formation of plaques in the brain consisting of fibrils of a protein called β-amyloid.[63] Consequently much of the research surrounding Alzheimer's disease is focussed on how these plaques form in the brain and to what extent the β-amyloid protein plays in the disease pathology. It has been found that the β-amyloid protein can generate oxygen-derived radicals and it is thought that this leads to the eventual plaque formation. The mechanism of radical formation is not completely understood but it seems that the β-amyloid protein is able to reduce ions of iron and copper, forming a catalytic cycle.[64] It has been shown that the key residue in

plaque formation and radical generation is Met35, which implies that thiol radicals and/or anions are directly involved in the disease pathology.[65]

1.5.3 Ageing

The process of ageing, or senescence, is difficult to delineate from other processes that can determine life span. The original hypothesis that ageing might be related to oxidative stress was put forward in 1956 when it was noticed that the effects of radiation damage were akin to those of ageing.[66] The free-radical theory of ageing concentrates primarily on endogenous processes, whereby oxidants are produced *in vivo* by mechanisms such as electron-transport chain-leakage in the mitochondria. This correlates well with the gerontologists notion of "rate of living", which expresses a relationship between the rate of metabolism with that of the life span.[67] For instance, there is a direct inverse relationship between life span in species of drosophila fly and the amount of H_2O_2 produced in the mitochondria.[68]

1.6 References

(1) Sjöberg, B.-M.; Reichard, P.; Gräslund, A.; Ehrenberg, A. *J. Biol. Chem.* **1997**, *252*, 536.

(2) Banerjee, R., Ed. *Chemistry and Biochemistry of B_{12}*; John Wiley & Sons: New York, 1999.

(3) Magnusson, O. Th.; Frey, P. A. *J. Am. Chem. Soc.* **2000**, *122*, 8807.

(4) (a) Smith, D. M.; Golding, B. T.; Radom, L. *J. Am. Chem. Soc.* **1999**, *121*, 1383. (b) Smith, D. M.; Golding, B. T.; Radom, L. *J. Am. Chem. Soc.* **1999**, *121*, 5700. (c) Smith, D. M.; Golding, B. T.; Radom, L. *J. Am. Chem. Soc.* **1999**, *121*, 9388. (d) Smith, D. M.; Golding, B. T.; Radom, L. *J. Am. Chem. Soc.* **2001**, *123*, 1664. (e) Sandala, G. M.; Smith, D. M.; Coote, M. L.; Radom, L. *J. Am. Chem. Soc.* **2004**, *126*, 12206.

(5) Davies, M. J.; Dean, R. T. *Radical-Mediated Protein Oxidation: from Chemistry to Medicine*; Oxford University Press: Oxford; New York, 1997, and references therein.

(6) Dean, R. T.; Fu, S.; Stocker, R.; Davies, M. J. *Biochem. J.* **1997**, *324*, 1.

(7) The word oxygen in this section refers to the molecule dioxygen O_2.

(8) Halliwell, B.; Gutteridge, J. M. C. *Free Radicals in Biology and Medicine,* 3rd ed., Oxford University Press, Oxford (1999).

(9) Koppenol, W. H. In *Oxidases and Related Systems*, King, T. E.; Mason, H. S.; Morrison, M. (eds.), 4th ed., pp. 93–109, Alan. R. Liss, New York (1988).

(10) Blum, J.; Fridovich, I. *Arch. Biochem. Biophys.* **1985**, *240*, 500.

(11) Sutton, H. C.; Roberts. P. B.; Winterbourn, C. C. *Biochem. J.* **1976**, *155*, 503.

(12) Flint, D. H.; Tuminello, J. F.; Emptage, M. H. *J. Biol. Chem.* **1993**, *268*, 22369.

(13) McCord, J. M.; Fridovich, I. *J. Biol. Chem.* **1969**, *244*, 6049.

(14) Fridovich, I. *Annu. Rev. Biochem.* **1995**, *64*, 97.

(15) (a) Aikens, J.; Dix, T. A. *J. Biol. Chem.* **1991**, *266*, 15091. (b) Aikens, J.; Dix, T. A. *Arch. Biochem. Biophys.* **1993**, *305*, 516.

(16) Freeman, B. A.; Crapo, J. D. *Lab. Invest.* **1982**, *47*, 412.

(17) (a) Beauchamp, C.; Fridovich, I. *J. Biol. Chem.* **1970**, *245*, 4641. (b) McCord, J. M.; Day, E. D. *FEBS Lett.* **1978**, *86*, 139.

(18) Not bound in a protein.

(19) Haber, F.; Weiss, J. *Proc. R. Soc. London Ser. A*, **1934**, *147*, 332.

(20) Dalal, N. S.; Shi, X. *Arch. BioChem. Biophys.* **1990**, *277*, 342.

(21) Kasprzak, K. S. *Chem. Res. Toxicol* **1991**, *4*, 604.

(22) Halliwell, B.; Gutteridge, J. M. C. *Methods, Enzymol.* **1990**, *186*, 1.

(23) Liochev, S. I.; Fridovich, I. *Arch. Biochem. Biophys.* **1990**, *279*, 1.

(24) See for example: (a) Bilinski, T.; Krawiec, Z.; Liczmanski, A.; Litwinska, J. *Biochem. Biophys. Res. Commun* **1985**, 130, 533. (b) Halliwell, B.; Gutteridge, J. M. C. *Arch. Biochem. Biophys.* **1986**, *246*, 501.

(25) (a) Rush, J. D.; Koppenol, W. H. *J. Biol. Chem.* **1986**, *261*, 6730. (b) Lloyd, R.V.; Hanna, P. M.; Mason, R. P. *Free Radic. Biol. Med.* **1997**, *22*, 885.

(26) (a) Youngman, R. J.; Elstner, E. F. *FEBS Lett.* **1981**, *2*, 265. (b) Youngman, R. J. *Trends Biochem. Sci.* , **1984**, *9*, 280.

(27) (a) Doly, M.; Bonhomme, B.; Vennat, J. C. *Opthalmic Res.* **1986**, *18*, 21. (b) Gutteridge, J. M. C. *FEBS Lett.* **1986**, 201, 291.

(28) (a) Floyd, R. A. *Arch. Biochem. Biophys.* **1983**, *225*, 263. (b) Halliwell, B.; Gutteridge, J. M. C. *Mol. Aspects Med.* **1985**, *8*, 89.

(29) (a) Balla, G.; Vercellotti, G. M.; Eaton, J. W.; Jacob, H.S. *J. Lab. Clin. Med.* **1990**, *116*, 546. (b) Healing, G.; Gower, J.; Fuller, B.; Green, C. *Biochem. Pharmacol.* **1990**, *39*, 1239. (c) Voogd, A.; Sluiter, W.; van Eijk, H. G.; Koster, J. F. *J. Clin. Invest.* **1992**, *90*, 2050. (d) Baliga, R.; Ueda, N.; Shah, S. V. *Biochem, J.* **1993**, *291*, 901. (e) Sergent, O.; Morel, I.; Cogrel, P.; Chevanne, M.; Pasdeloup, N.; Brissot, P.; Lescoat, G. Clillard, P.; Cillard, J. *Biol. Trace Element Res.* **1995**, *47*, 185. (f) Breuer,W.; Epsztejn, S.; Cabantchik, Z. I. *J. Biol. Chem.* **1993**, *270*, 24209. (g) Öllinger, K.; Roberg, K. *J. Biol. Chem.* **1997**, *272*, 23707. (h) Stäubli, A.; Boelsteri, U. A. *Am. J. Physiol.* **1998**, *274*, G1031.

(30) (a) von Sonntag, C. *The Chemical Basis of Radiation Biology*, Taylor and Francis, London, 1987. (b) Garrison, W.M. *Radiat. Res. Rev.* **1972**, *3* 305. (c) Davies, K.J. *J. Biol. Chem.* **1987**, *262*, 9895–9901. (d) Garrison, W.M. *Chem. Rev.* **1987**, *87*, 381. (e) Stadtman, E.R. *Free Radic. Biol. Med.* **1990**, *9*, 315. (f) Easton, C.J. *Adv. Detail. React. Mech.* **1991**, *1* 83. (g) Stadtman, E.R. *Annu. Rev. Biochem.* **1993**, *62*, 797. (h) Stadtman, E.R. *Methods Enzymol.* **1995**, *258*, 379. (i) Easton, C.J. *Chem. Rev.* **1997**, *97*, 53.

(31) (a) M. Sheinblatt and H. S. Gutowsky, *J. Am. Chem. Soc.* **1964**, *86*, 4814. (b) G. Wada, E. Tamura, M. Okina, and M. Nakamura, *Bull. Chem. Soc. Jpn.* **1982**, *55*, 3064.

(32) Nagaoka, M.; Okuyama-Yoshida, N.; Yamabe, T. *J. Phys. Chem. A* **1998**, *102*, 8202.

(33) Viehe, H. G.; Janousek, Z.; Merenyi, R.; Stella, L. *Acc. Chem. Res.*, **1985**, *18*, 148.

(34) Rauk, A.; Yu, D.; Taylor, J.; Shustov, G. V.; Block, D. A.; Armstrong, D. A. *Biochemistry* **1999**, *38*, 9089.

(35) NIST chemistry Webbook *webbook.nist.gov*.

(36) (a) Sperling, J.; Elad, D. *J. Am. Chem. Soc.* **1971**, *93*, 967. (b) Schwarzberg, M.; Sperling, J.; Elad, D. *J. Am. Chem. Soc.* **1973**, *95*, 6418. (c) Easton, C. J. In *Advances in Detailed Reaction Mechanisms*, Coxon, J. M. (ed.), pp. 83–126, JAI Press, Greenwich, Connecticut (1991).

(37) (a) Croft, A. K.; Easton, C. J.; Radom, L. *J. Am. Chem. Soc.* **2003**, *125*, 4119. (b) Croft, A. K.; Easton, C. J.; Kociuba, K.; Radom, L. *Tetrahedron: Asymm.* **2003**, *14*, 2919.

(38) Abramovich, S. D.; Rabani, J. J. *Phys. Chem.* **1976**, *80*, 1562.

(39) See for example: (a) R. Atkinson, Int. J. Chem. Kinet., **1997**, 29, 99. (b) Mereau, R.; Rayez, M. T.; Caralp, F.; Rayez, J. C. *Phys. Chem. Chem. Phys.*, **2000**, *2,* 1919. (c) Mereau, R.; Rayez, M. T.; Caralp, F.; Rayez, J. C. *Phys. Chem. Chem. Phys.*, **2000**, *2*, 3765. (d) P. Devolder, J. Photochem. Photobiol. A, **2003**, 157, 137. (e) J. J. Orlando, G. S. Tyndall and T. J. Wallington, Chem. Rev. **2003**, 103, 4657. (f) R. Atkinson and J. Arey, *Chem. Rev.* **2003**, 103, 4605.

(40) Davies, M. J. *Arch. Biochem. Biophys.* **1996**, *336*, 163

(41) Headlam, H. A.; Mortimer, A.; Easton, C. J.; Davies, M. J. *Chem. Res. Toxicol.* **2000,** *13,* 1087.

(42) Nukuna, B. N.; Goshe, M. B.; Anderson, V. E. *J. Am. Chem. Soc.* **2001**, *123*, 1208.

(43) Behrens, G.; Koltzenburg, G. *Z. Naturforsch.* **1985**, *40C*, 785.

(44) (a) Sokol, H. A.; Bennett-Corniea, W.; Garrison, W. M. *J. Am. Chem. Soc.* **1965**, *87*, 1391. (b) Armstrong, W.A.; Humphreys, W.G. *Can. J. Chem.* **1967**, *45*, 2589. (c) Taniguchi, H.; Hatano, H.; Hasegawa, H.; Maruyama, T. *J. Phys. Chem.* **1970**, *74*, 3063. (d) Rustgi, S.; Joshi, A.; Moss, H.; Riesz, P. *Int. J. Radiat. Biol.* **1977**, *31*, 415.

(45) Hawkins, C. L.; Davies, M. J. *Biochim. Biophys. Acta.* **1997**, *1360*, 84.

(46) (a) Brodskaya, G. A.; Sharpatyi, V. A. *Russ. J. Phys. Chem.* **1967,** *41*, 583. (b) Wheeler, O.H.; Montalvo, R. *Radiat. Res.* **1969**, *40*, 1. (c) Balakrishnan, I.; Reddy, M. P. J. *Phys. Chem.* **1970**, *74*, 850. (d) Dizdaroglu, M.; Simic, M. G. *Radiat. Res.* **1980**, *83*, 437.

(47) Simic, M. G.; Gajewski, E.; Dizdaroglu, M. *Radiat. Phys. Chem.* **1984**, *24*, 465.

(48) Kim, H.-J.; Mee, L. K.; Adelstein, S. J.; Taub, I. A.; Carr, S. A.; Reinhold, V. N. *Radiat. Res.* **1984**, *100,* 30.

(49) Gieseg, S. P.; Simpson, J. A.; Charlton, T. S.; Duncan, M. W.; Dean, R. T. *Biochemistry* **1993**, *32*, 4780.

(50) Josimovic, L.; Jankovic I.; Jovanovic, S. V. *Radiat. Phys. Chem.* **1993**, *41*, 835.

(51) Kopoldova, J.; Hrneir, S. *Z. Naturforsch.* **1977,** *32C*, 482.

(52) Hart, E. J.; Boag, J. W. *J. Am. Chem. Soc.* **1962**, *84*, 4090.

(53) Hart, E.; Fielden, E. M.; Anbar, M. *J. Phys. Chem.* **1967**, *71*, 3993.

(54) Garrison, W. M. *Curr. Topics Radiat. Res.* **1968**, *4*, 43.

(55) Zigler, J. S.; Goosey, J. D. *Photochem. Photobiol.* **1981**, *33*, 869.

(56) B. Halliweil, *Chloroplast Metabolism: The Structure and Function of Chloroplasts in Green Leaf Cells*. Oxford Univ. Press (Clarendon), Oxford, 1984.

(57) Franck, B. *Angew. Chem. Int. Ed. Engl.* 21, 343 (1982).

(58) Matheson, I. B. C.; Etheridge, R.D.; Kratowich, N.R.; Lee, J. *Photochem. Photobiol.* **1975**, *21*, 165.

(59) Wilkinson, F.; Helman, W.P.; Ross, A.B. *J. Phys. Chem. Ref. Data* **1995**, *24*, 663.

(60) Davies, M. J. *Chem. Res. Toxicol.* **2001**, *14*, 1453.

(61) Hawkins, C. L.; Davies, M. J. *J. Chem. Soc., Perkin Trans. 2* **1998**, 1937.

(62) *Cardiovascular Disease in Australia: A Snapshot, 2001*, article number 4821.0.55.001, The Australian Bureau of Statistics, http://www.abs.gov.au/

(63) Mattson, M. P. *Nature Struct. Biol.* **1995**, *2*, 926.

(64) Huang, X.; Cuajungco, M. P.; Atwood, C. S.; Hartshorn, M. A.; Tyndall, J. D. A.; Hanson, G. R.; Stokes, K. C.; Leopold, M.; Multhaup,G.; Goldstein, L. E.; Scarpa, R. C.; Saunders: A. J.; Lim, J.; Moir, R. D.; Glabe, C.; Bowden, E. F.; Masters, C. L.; Fairlie, D. P.; Tanzi, R. E.; Bush, A. I. *J. Biol. Chem.* **1999**, *274*, 37111.

(65) (a) Pike, C. J.; Walencewicz-Wasserman, A. J.; Kosmoski, J.; Cribbs,D. H.; Glabe, C. J.; Cotman, C. W. *J. Neurochem.* **1995**, *64*, 253-265. (b) Yatin, S. M.; Varadarajan, S.; Link, C. D.; Butterfield, D. A. *Neurobiol. Aging* **1999**, *20*, 325–330. (c) Varadarajan, S.; Yatin, S.; Kanski,J.; Jahanshahi, F.; Butterfield, D. A. *Brain Res. Bull.* **1999**, *50*, 133.

(66) (a) Harman, D. *J. Gerontol.* **1956**, *2*, 298–300. (b) Hempelmann, L. H.; Hoffman. J. G.; *Annu. Rev. Nucl. Sci.* **1953**, *3*, 369.

(67) (a) Pearl, R. *The Rate of Living.* London, Univ. of London Press, 1928. (b) Sohal, R. S. in *Cellular Aging: Conceptsand Mechanisms*, ed Witler. R., Basel, Karger, pp. 25 (1976).

(68) Sohal, R. S.; Sohal, B. H.; Orr, W. C. *Free Radic. Biol. Med.* **1995**, *19*, 499.

CHAPTER 2

Theoretical Methods

2.1 Foundations of Quantum Chemistry

In 1929, Dirac famously subjugated the proud and erudite discipline of chemistry into the mechanical and formulated realm of applied mathematics with two sentences. His first statement was:

> *"The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble."*[1]

The second statement followed the first and says:

> *"It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, ...".*

Experimental chemistry is as important today as it was 77 years ago. The significance of these comments lies at the core of what is known today as quantum chemistry. They mark, not the subjugation of a discipline, but the birth of a new one.

Although the first statement is more of a philosophical one, the second has become a practicable tenet of quantum chemistry and has divided the discipline into two main areas. The first division applies currently available approximate methods to systems that are tractable by accessible computational means. As computers have become more powerful over the last seventy years, the realm of applicability has literally exploded exponentially in a Moores' law[2] fashion. The second area develops theoretical methods that either become more accurate or scale better computationally. This means the fundamental physical laws that Dirac speaks of can become soluble if a number of approximations are made. Therefore, one may attempt to build in less

approximations, which usually results in a greater amount of computational effort, or refine current approximations by re-formulating them.

The reader may ask what brought Dirac to make such a resounding statement? The reason lies in the rapid development of physical laws during the first part of the 20th century, which eventually were to solve some of the fundamental mysteries of classical physics. One of the most important of these, in terms of the development of quantum mechanics, was the "ultraviolet catastrophe". The heart of this problem is the power (energy per unit time) of an ideal electromagnetic emitter. Classical Maxwellian laws predict that the power tends to infinity for shorter wavelengths of radiation. This is certainly false; otherwise, skin cancer in Australia would be the least of our worries. The predicament was rectified by assuming that the electrons producing the radiation within a blackbody could only oscillate at fixed discrete frequencies. The successful prediction of blackbody radiation curves using this assumption led to the formulation by Bohr of a quantum theory of the electronic structure of atoms. In quick succession to this, Erwin Schrödinger formulated a wave equation, now bearing his name, which could in principle describe the electronic state of any particular system. If this could be done then any bond-making or bond-breaking event in chemistry could be described in its entirety.

2.2 Schrödinger's Equation

Simply put, Schrödinger's equation is a boundary-valued partial-differential equation that embodies Newton's laws of motion as applied to the nuclei and electrons of atoms and molecules. The significance of the wave equation is that it describes particles, which have an associated wavelength depending on their momentum, a duality first pointed out by de Broglie.[3] The equation can be written as:

$$i\hbar\frac{\partial}{\partial t}\Psi(q,t) = H(q,t)\Psi(q,t) \tag{2.1}$$

where, $\Psi(q,t)$ is the wave function describing the state of the system, $H(q,t)$ is the time-dependent Hamiltonian operator describing the total energy of the system, and q and t are the generalized position coordinates and time coordinate, respectively. Using

$$H = \frac{\mathbf{P}^2}{2m} + V(q) \tag{2.2}$$

as the Hamiltonian operator and converting the momentum operator into the quantum mechanical form, one obtains the famous time-dependent Schrödinger equation:

$$i\hbar \frac{\partial}{\partial t}\Psi(q,t) = -\left(\frac{\hbar^2}{2m}\right)\nabla^2\Psi(q,t) + V(q)\Psi(q,t) \tag{2.3}$$

In the special case where the Hamiltonian is time-independent, the solution of equation 2.1 may be reformulated into a product of time-independent and time-dependent functions.

$$\Psi(q,t) = \Psi(q)y(t) \tag{2.4}$$

Equation 2.4 may be used to separate equation 2.1 into two ordinary differential equations, one of which contains the time-dependency, i.e.,

$$i\hbar \frac{dy(t)}{dt} = Ey(t) \tag{2.5}$$

which has solutions of the form:

$$y(t) = Ae^{-\frac{iEt}{\hbar}} \tag{2.6}$$

The equation governing the time-independent variables is known as the time-*independent* Schrödinger equation and is given by:

$$-\left(\frac{\hbar^2}{2m}\right)\nabla^2\Psi(q) + V(q)\Psi(q) = E\Psi(q) \tag{2.7}$$

Erwin Schrödinger detailed this result in one of four monumental papers published in the first half of 1926 marking the birth of wave mechanics (or in his words the "undulatory theory of mechanics of atoms and molecules"[4] but this may be just a poor translation) and thus quantum chemistry. To solve equation 2.7 within the theory of partial-differential equations, one needs to state the appropriate boundary condition. For physical reasons, the condition states that the particles must be bound or confined within some region of space, i.e.,

$$\Psi(q) \to 0 \text{ as } |q| \to \infty \tag{2.8}$$

Imposing the boundary condition 2.8 and restricting the wave function so that it is single-valued, continuous and has continuous first derivatives gives non-trivial solutions to equation 2.7 with discrete values of E. Unfortunately for chemists, solutions to equation 2.7 subjected to the boundary condition 2.8 can only be found analytically in a small number of special cases and it is at this point Dirac's "approximate" methods must be invoked.

2.3 Standard Approximations

2.3.1 The Born-Oppenheimer Approximation

Born and Oppenheimer argued[5] that because nuclei are much heavier than electrons, the electronic response to movements by the nuclei must be nearly instantaneous. The consequence of this is a decoupling of electronic and nuclear motion. Mathematically this implies that the total wave function given in equation 2.7 can be separated into product wave functions of the electrons and nuclei, i.e., $\Psi(q) = \Psi^{elec}(q)\Psi^{nuc}(Q;q)$. Therefore one may solve equation 2.7 for the electronic

wave function in the mean field of the nuclei. The total electronic energy (E_{TOT}) one finds from 2.7 then becomes the sum of the electronic energy and the coulombic repulsion of the nuclei (nuclear repulsion).

$$E_{TOT} = E_{ELEC} + E_{NR} \tag{2.9}$$

If one writes the time-independent Hamiltonian in terms of operations occurring because of electron-electron, nucleon-nucleon and electron-nucleon interactions then equation 2.7 becomes;

$$\left(-\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}}\right)\Psi^{elec}(q)\Psi^{nuc}(Q;q) \tag{2.10}$$

$$= E_{TOT}\Psi^{elec}(q)\Psi^{nuc}(Q;q)$$

The first term in the brackets of equation 2.10 is the kinetic energy operator for the N electrons in the system, the second term is the coulomb attraction between the M nuclei and N electrons, the third is the coulomb repulsion between all pairs of electrons and the final term is the coulomb repulsion between all pairs of nuclei. Equation 2.10 introduces the use of atomic units, which are obtained by setting the mass of the electron (m_e), the charge on an electron (e) and Plank's constant divided by 2π ($\hbar$) to one. The units for energy, length and time become the Hartree, Bohr and Chronon, respectively. To solve the electronic problem one separates the equation into the operations that act on the electronic wave function, and then solve it for a particular arrangement of the nuclei. In this way the electronic wave function depends parametrically on the nuclear coordinates (Q). The nuclear coordinates can be changed and the electronic problem resolved for any desired nuclear configuration so that a potential energy hypersurface is built. Each time the electronic problem is solved, a countable infinite number of solutions is obtained and each of these solutions is termed an "electronic state".

2.3.2 The Orbital Approximation

The original time-independent Schrödinger equation has been reduced to the electronic Schrödinger equation by invoking a separation of variables through the Born-Oppenheimer approximation. Separation of variables is a common technique used to simplify partial-differential equations and it can be used again to reduce the complexity of the Schrödinger equation.

The electronic Schrödinger equation is a $3N$-dimensional problem, where N is the number of electrons, and the motions of the N electrons are highly coupled. This still leaves the resultant equation insoluble except in a few special cases, such as one-electron problems like the hydrogen atom. The coupling occurs through terms in the Hamiltonian involving inter-electronic distances, i.e. $1/r_{ij}$. By invoking the separation of variables technique once again, this time on the variables describing the motion of each electron, one obtains the independent-electron approximation, or the orbital approximation – so called because the electronic motion is now uncoupled and each electron moves in the average field of the others. Mathematically the wave function describing the highly coupled motions of all the electrons is separated into a product wave function of each electron:

$$\Psi(q) = \chi_i(q_1)\chi_j(q_2).....\chi_k(q_N) \tag{2.11}$$

where, each of the $\chi_k(q_N)$ is a one-electron wave function otherwise known as a spin orbital. Chemists intuitively regard electrons as spin orbitals but the latter should only be viewed as a purely mathematical construct. The spin orbitals, are comprised of a spatial function and a spin function. The spin functions dictate whether the electron has "up" or "down" spin and are orthogonal to one another. Equation 2.11 is also known as a Hartree product. If one takes the squared-norm of the Hartree product, *viz*:

$$|\chi_i(q_1)|^2|\chi_j(q_2)|^2 \ldots |\chi_k(q_N)|^2 \partial q_1 \partial q_2 \ldots \partial q_N \tag{2.12}$$

Then equation 2.12 indicates the simultaneous probability of finding electron 1 in volume element ∂q_1, electron 2 in volume element ∂q_2, etc.

Invoking this approximation allows the Schrödinger equation to become soluble by numerical methods. However, the wave function given by equation 2.11 has a serious flaw, namely it is not antisymmetric with respect to the exchange of two electrons – a physical requirement of indistinguishable fermionic particles. To keep the formalism as simple as possible, antisymmetry is introduced in an *ad hoc* way. This can be achieved by replacing the product wave function (2.11) with a determinant. First pointed out by J. C. Slater in 1929,[6] this type of determinant now bares his name.

$$\Psi(q) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_i(q_1) & \chi_j(q_1) & \cdots & \chi_k(q_1) \\ \chi_i(q_2) & \chi_j(q_2) & \cdots & \chi_k(q_2) \\ \vdots & \vdots & & \vdots \\ \chi_i(q_N) & \chi_j(q_N) & \cdots & \chi_k(q_N) \end{vmatrix} \tag{2.13}$$

This determinant easily satisfies the antisymmetry requirement because if two rows are exchanged, which is equivalent to exchanging two electrons, then the sign of the determinant, and thus the wave function is changed.

2.3.3 The Variational Principle

The variational principle states that an approximate (trial) wave function for a given non-relativistic time-independent Hamiltonian will have an energy expectation value greater than or equal to the exact ground state energy for the Hamiltonian in question. Mathematically this is usually represented in Dirac notation as:

$$E[\Psi] = \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle} \tag{2.14}$$

where $E[\Psi]$ is said to be the energy functional, and any Ψ that is a stationary point for the functional will generate an energy expectation value that is an upper bound to the exact ground state energy.

The variational method depends on the particular form of the trial wave function. If equation 2.14 is expanded in terms of the energy eigenstates of H, as given by equation 2.5, and the functional is minimized in terms of the expansion coefficients, then one can derive the matrix expression for the expansion coefficients ($\boldsymbol{C}$),

$$\boldsymbol{HC} = \boldsymbol{SCE} \tag{2.15}$$

which can be solved by diagonalization techniques of linear algebra. In this expression $\boldsymbol{H}$ is the matrix form of the Hamiltonian, $\boldsymbol{C}$ is the matrix of expansion coefficients, $\boldsymbol{S}$ is the matrix of overlap integrals ($\langle \phi_a | \phi_b \rangle$) and $\boldsymbol{E}$ is the vector of eigenvalues of $\boldsymbol{H}$, the lowest of which is identified with the ground-state energy.

In practice, a complete set of energy eigenstates is impossible to generate for any particular Hamiltonian and usually some sort of finite set is used. In quantum chemistry, the finite set is generated by expanding the energy eigenstates with a finite set of functions, known as a basis set.

2.3.4 Basis Sets

A basis set in quantum chemistry is a known finite set of functions that is used to expand the energy eigenstates of a particular Hamiltonian. The expansion is sometimes referred to as the linear combination of atomic orbitals method or the LCAO method. Since one basis function is used to describe one electron, a basis set is known as a one-electron basis set.

$$\phi_a = \sum_i c_{ai} \theta_i \tag{2.16}$$

The known set of functions, θ_i, should be chosen so they contain the essential physics of the problem at hand. For example, in molecular problems, they should tend to zero as the separation between the nucleus and the electron becomes very large. The functions should also be able to describe the electronic-cusp condition well. The cusp condition is the discontinuous derivative that occurs when the separation between electrons tends to zero.

Historically, the functions best describing the essential physics were exponential functions ($e^{-r\zeta}$), known as Slater-type orbtials (STOs). However, it soon became clear that integrals involving STOs were difficult to evaluate and instead Gaussian-type functions (GTFs) started to predominate in electronic structure calculations. Gaussian functions have a squared dependence in the exponential ($e^{-r^2\alpha}$) and although this makes the evaluation of integrals involving them simpler, it does not give a satisfactory description of the cusp condition.

Initially, an alternative way of dealing with the cusp condition was to fit a set of GTFs to Slater-type orbitals. The fixed linear combination of GTFs representing a Slater-type orbital became known as a contracted Gaussian function. However, as computers became more powerful it was sufficient to just add large numbers of easily evaluated GTFs as the bottleneck of a calculation was not usually caused this way.

John A. Pople and his co-workers, developed a range of basis sets formulated on Gaussian-type functions and his basis sets and nomenclature have become the standard in many electronic structure calculations. Many others such as Dunning have contributed and extended Pople's general theme.[7]

2.4 Ab Initio Methods

In this section, a hierarchy of approximate methods will be briefly described. They are known as *ab initio* methods because they only assume the Schrödinger equation and fundamental physical constants.

2.4.1 Hartree-Fock Theory

Hartree-Fock theory essentially takes all of the ideas of the previous section and combines them into a single formalism for calculating the ground-state energy of a given electronic-structure problem. That is, one takes a Slater-determinant (equation 2.13) and applies the variational principle to find the set of coefficients that minimizes the energy expectation value, subject to the constraint that the set of spin orbitals (χ_i) remain orthonormal. If this is done then one arrives at a set of integro-differential equations, known as the Hartree-Fock equations, where introduction of a basis set yields a set of matrix equations known as the Roothaan-Hall equations.[8] The Roothaan-Hall equations can be concisely written as

$$\sum_{\nu=1}^{N} c_{\mu i}(F_{\mu\nu} - \varepsilon_i S_{\mu\nu}) = 0 \quad \mu = 1,2,\cdots,N \tag{2.17}$$

where $c_{\mu i}$ is the set of coefficients that minimize the energy expectation value, $F_{\mu\nu}$ is an N by N matrix (where N is the number of basis functions) known as the Fock matrix, which is essentially a one-particle effective Hamiltonian operator, the ε_i's are the eigenvalues or the one-electron orbital energies of the Fock matrix, and $S_{\mu\nu}$ is an N by N matrix of overlap integrals between basis functions, which are in general normalized but not orthogonal to one another.

Importantly, the Fock matrix depends on the spin orbitals, which in turn are solutions to the eigenvalue equation. This means the Roothaan-Hall equations must be solved in an iterative manner. To accomplish this, an initial guess of the Fock operator is formed and the equations solved. The resultant coefficients are then used to reformulate the Fock matrix. This procedure is repeated until the set of coefficients used to form the Fock matrix are equivalent (within a numerical threshold) to the coefficients subsequently calculated. In this manner, the equations are solved until the point where the solutions are said to be self-consistent.

Until this point the equations have not considered if there is a difference between the spatial extent of spin "up" and spin "down" electrons, or equivalently α and β electrons. Treating the spatial orbitals of α and β electrons to be equivalent leads to the incorrect prediction of a zero Fermi-contact interaction in radical species. The Fermi-contact interaction is an energetic interaction, which splits the interaction energy of an unpaired electron in a spin-resonance experiment. In order for this to happen, there needs to be excess spin density at the nucleus. For example, in the methyl radical if the spatial orbitals of the α and β electrons are equivalent, then there can be no excess of α or β spin at the three methyl protons, leading to no Fermi-contact interaction. However, if the spatial orbitals of α or β spin are different then the unpaired electron is able to polarize the other electrons. Applying this idea to the methyl radical, the unpaired electron (say of α spin) in the $2p_z$ orbital of carbon would induce a higher probability of finding an α electron in the $2sp^2$ orbital, via Hund's rule. Because this orbital is part of the C–H bond the H 1s electron will have an enhanced probability of having β spin. This idea is represented in Figure 2.1.

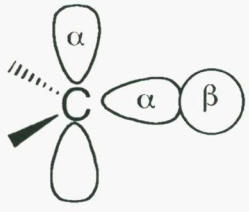


Figure 2.1: Spin Polarization in the Methyl Radical.

The Roothaan-Hall equations can be generalized to account for differing spatial orbitals of electrons of α or β spin. This was done in 1954 by Pople and Nesbet[9] and the corresponding equations are called the Pople-Nesbet equations. These equations are a set of coupled Roothaan-Hall equations for α and β electrons, i.e.,

$$\sum_{\nu=1}^{N} c_{\mu i}^{\alpha}(F_{\mu\nu}^{\alpha} - \varepsilon_i^{\alpha} S_{\mu\nu}) = 0$$
$$\sum_{\nu=1}^{N} c_{\mu i}^{\beta}(F_{\mu\nu}^{\beta} - \varepsilon_i^{\beta} S_{\mu\nu}) = 0 \quad \nu = 1,2,..,N \tag{2.18}$$

The Pople-Nesbet equations are coupled because the Fock operator in each equation depends on terms from the other equation. If the numbers of α and β electrons are equal, then these equations collapse to the Roothaan-Hall equations. A Hartree-Fock calculation carried out on an electronic system with unequal numbers of α and β electrons (open-shell systems) using the Pople-Nesbet formalism is called an *unrestricted* Hartree-Fock calculation, UHF for short. Open-shell systems treated by restricting the α and β spin electrons to be contained in the same spatial orbital are called *restricted* Hartree-Fock calculations (RHF).

For open-shell systems, UHF can successfully account for spin polarization but unfortunately this technique has a subtle but sometimes detrimental problem, that of spin-contamination. The two opposing problems of spin-polarization and spin-contamination have been referred to as the Charybdis and Scylla of quantum chemistry.[10]

Spin-contamination refers to the resultant wave function of a UHF calculation not being an eigenstate of the total spin-angular momentum operator ($\mathbf{S}^2$). This leads to expectation values of $\mathbf{S}^2$ that are larger than what they should ideally be. For example, the methyl radical has a spin multiplicity of two and therefore the expectation value of $\mathbf{S}^2$, or any other doublet, should be 0.75. However, a UHF calculation with the 6-31G(d) basis set (denoted UHF/6-31G(d)[11]) on an optimized structure of the methyl radical has an $\mathbf{S}^2$ expectation value 0.7615. Although this does not seem to be significant in itself, for species that have large spin-contamination this results in an over-estimation of spin polarization, or more seriously, to spurious results if this wave function is used as a starting point for correlation procedures (described below). A

detailed examination of the effects of spin-contamination in the formulation of *ab initio* methods is discussed in Chapter 3 where a new method for thermochemical calculations is introduced.

2.4.2 Full Configuration Interaction

The full configuration interaction method (full CI) and variants of it (§2.4.2–§2.4.3) are not used explicitly in this thesis but the concept of full CI introduces an important concept of electron correlation and so justifies a discussion of this method.

The Hartree-Fock determinant recovers approximately 99% of the total energy but the missing 1% is often important for describing chemical processes. Since the electrons in the Hartree-Fock problem move in the average field of the other electrons, the molecular orbitals found from an RHF calculation are contracted more than what they should be. Explicit electron-electron interactions would push, on average, the electrons further apart. Therefore, the correlation energy is defined to be the difference between the (restricted) Hartree-Fock energy and the true ground-state energy within the particular one-electron basis.

In order to improve the Hartree-Fock energy one needs to include explicit electron-electron correlation. It was suggested by Hylleraas[12] that one way of doing this is to include "antibonding" character by way of excited-state determinants. The RHF problem is variationally the best one-determinant solution for the N one-electron basis functions. However, there are $N!/n!(N!-n!)$ possible ways of distributing n electrons in N orbitals, and the Hartree-Fock determinant, although the most important, is only one of them. The remaining determinants are excited-state determinants because they correspond to the promotion of an electron from an occupied orbital (in the Hartree-Fock determinant) to a virtual orbital. This allows for a classification of determinants according to the number of electrons that are promoted in the Hartee-Fock determinant. If only one electron is promoted, this is known as a single (S) excitation, two electrons

are doubles (D), three triples (T) etc. until all N-n/2 unoccupied (virtual) orbitals are occupied by the n available electrons.

The different determinants are known as configurations and the method that ascertains how the determinants are related to one another is termed a configuration interaction. Mathematically, full configuration interaction is a linear expansion of determinants, where the coefficients are variationally optimized (the expansion coefficients are often referred to as amplitudes).

$$\Psi_{CI} = a_0\Psi_{HF} + \sum_S a_S\Psi_S + \sum_D a_D\Psi_D + \sum_T a_T\Psi_T \cdots \tag{2.19}$$

The full configuration interaction solution is termed: the best n-electron solution within the N one-electron basis, and recovers the remaining 1% of energy important for chemical processes. Quantum chemistry would stop at this point if full configuration interaction were computationally tractable. Unfortunately, it scales computationally as $N!$ (where ! is a factorial but should also entail an expletive), making the scope of its use limited to a small number of systems.

In order to overcome the computational expense of the full CI problem, a number of approximate methods have been developed. The methods can be classified into groups according to whether they involve truncation of the expansion, restriction of the correlation space or extrapolation of the energy. These methods are useful in that they reduce the computational expense of the calculation but sometimes the resultant method loses the important properties of *size consistency* and *size extensivity.*

Size consistency refers to the property that the energy of a non-interacting super-molecule is equal to the sum of energies of the fragments. *Size extensivity* refers to the property that the correlation energy recovered scales with the size of the system. Both full CI and Hartree-Fock methods have these properties.

2.4.3 Truncation of the Full CI Expansion

The full CI expansion as written in equation 2.19 leads to a natural truncation scheme that includes sets of determinants of a particular excitation type. Starting with the Hartree-Fock solution one could include only determinants with single excitations. However, within the formulation of the Hartree-Fock equations, a necessary condition known as Brillouin's theorem was imposed that prevented the interaction of single excitations with the Hartree-Fock determinant. Thus, the lowest level of truncation that improves the Hartree-Fock result is inclusion of all the doubly-excited determinants, where this method is known as CID (CI with doubles). The next level of truncation includes all single and double excitations, because though the singly excited determinants do not interact with the Hartree-Fock determinant they do interact with the doubly-excited determinants, i.e.,

$$\Psi_{CI} = a_0\Psi_{HF} + \sum_S a_S\Psi_S + \sum_D a_D\Psi_D \qquad (2.20)$$

This method is termed CISD for obvious reasons. The inclusion of higher excitations follows naturally from the CISD truncation to give methods such as CISDT, CISDTQ etc.

Truncation reduces the number of required determinants by decreasing the order of excitation. An alternative way of reducing the number of determinants is to decrease the number excitations within each order of excitation. Methods that invoke this type of reasoning use what is usually referred to as an active space. An active space is a certain number of occupied and virtual orbitals that are chosen in which excitations are to occur. Figure 2.2 demonstrates two common techniques for partitioning the active space. In the first, (Figure 2.2 a) the active space is constituted of two occupied orbitals and two virtual orbitals, all possible excited determinants are formed from the four total orbitals.

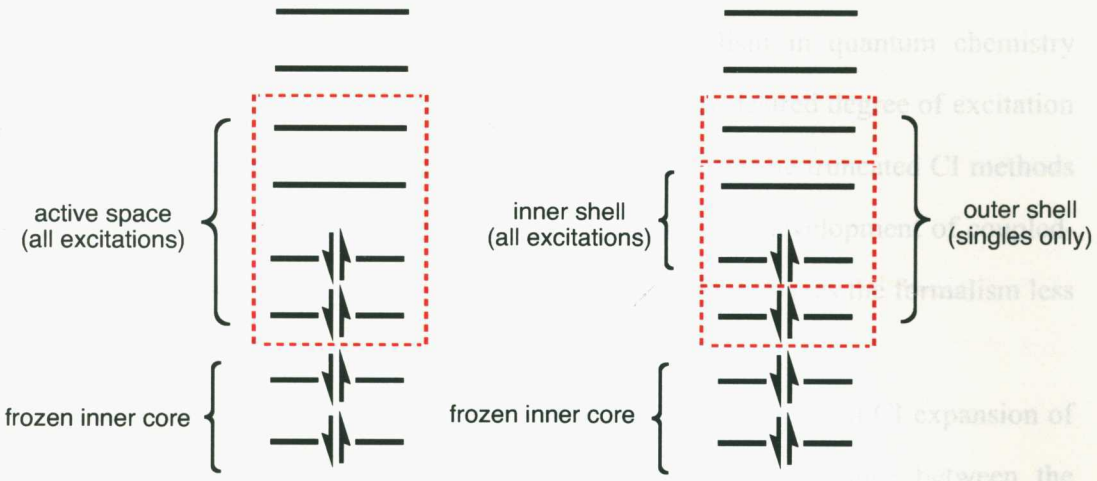


Figure 2.2: (a) Reducing a Full Configuration Interaction Calculation to a Subspace of Two Occupied and Two Virtual Orbitals. (b) A Further Partitioning of the Active Space.

In the second, (Figure 2.2 b) the active space is further divided so that there is an inner set of occupied and virtual orbitals and an outer-shell of orbitals containing a lower set of occupied orbitals and a higher set of virtual orbitals. In the inner set, all excitations are formed, and in the outer-shell a reduced number of excitations are formed.

In each case the energetically lowest set of occupied orbtials, below the active space, is referred to as the frozen core. In other correlation techniques, such as perturbation theory and coupled-cluster techniques, which will be discussed below, one may also invoke a frozen-core approximation. The frozen core is usually the chemically inert electrons, i.e. 1s electrons for first-row atoms, 2s and 2p electrons for second-row atoms. The frozen core approximation is widely used in quantum chemistry and the calculations in this thesis are no exception. If calculations have been performed with all electrons correlated, the designation of *full* is usually specified in association with the particular computational method and this practice will be followed in this thesis.

2.4.4 Coupled-Cluster Techniques

Coupled-cluster methods are an important formalism in quantum chemistry because they create a hierarchy of methods that include any desired degree of excitation while remaining size consistent. This may be contrasted with the truncated CI methods that are not size consistent for any degree of excitation. The development of coupled-cluster methods leads to a complicated set of equations, which makes the formalism less transparent than the CI methods.

The coupled-cluster methods can be motivated through the full CI expansion of two non-interacting hydrogen dimers (H_2-----H_2), where the distance between the dimers is arbitrary. In the full CI scheme the wave function is a linear expansion of determinants up to and including quadruple excitations, i.e.,

$$\Psi_{CI} = \Psi_{HF}^4 + \sum_D a_D \Psi_D^4 + a_Q \Psi_Q^4 \qquad (2.21)$$

In this expansion the '4' superscript indicates that the wave function is a four-electron wave function. Because the system is inherently linear single excitations are zero by symmetry. Excitations promoting an electron from one monomer to the other are also zero because the distance between monomers is arbitrary. This leaves only the reference Hartree-Fock determinant, two double excitations, and one quadruple excitation. Recognizing that the monomers are non-interacting, allows for the separation of the four-electron wave function into a direct product of two two-electron wave functions. In doing so, the quadruple excitation amplitude can be replaced exactly by a product of two two-electron excitation amplitudes $a_D a_{D'}$, i.e.,

$$\Psi_{CI} = \Psi_{HF}^2 \otimes \Psi_{HF}^2 + \sum_D a_D \Psi_D^2 \otimes \Psi_D^2 + a_D a_{D'} \Psi_D^2 \otimes \Psi_D^2 \qquad (2.22)$$

The truncated CI methods neglect the final term in equation 2.22, which is the cause of their size *in*consistency. The coupled-cluster methods include terms such as these and

they are known as disconnected excitations. The coupled-cluster methods were first introduced in this manner by Čižek[13] (although they have a longer history[14]) but today it is common to recast them into a formalism known as the "cluster-expansion". The result of this expansion is a hierarchy of methods that includes any degree of excitation in a size-consistent manner.

The cluster expansion of the wave function starts with an excitation operator $\boldsymbol{T}$, which is the sum of all types of excitations ($\boldsymbol{T}_1$ representing singles, $\boldsymbol{T}_2$ doubles, etc), i.e.,

$$\boldsymbol{T} = \boldsymbol{T}_1 + \boldsymbol{T}_2 + \boldsymbol{T}_3 + \boldsymbol{T}_4 + ,\cdots, + \boldsymbol{T}_N \tag{2.23}$$

When in an exponential form, the excitation operator generates all forms of disconnected excitations. To demonstrate this, consider the Taylor series expansion of the excitation operator that only includes single and double excitations, i.e., $\boldsymbol{T} = \boldsymbol{T}_1 + \boldsymbol{T}_2$, then,

$$e^{T} = 1 + \left(T_1 + T_2\right) + \frac{\left(T_1 + T_2\right)^2}{2!} + \cdots = \sum_{k=0}^{\infty} \frac{\left(T_1 + T_2\right)^k}{k!} \tag{2.24}$$

Terms can be regrouped into quantities depending on one particular level of excitation, as shown in equation 2.25.

$$1 + T_1 + \left(T_2 + \frac{1}{2}T_1^2\right) + \left(T_2T_1 + \frac{1}{6}T_1^3\right) + \left(\frac{1}{2}T_2^2 + \frac{1}{2}T_2T_1^2 + \frac{1}{24}T_1^4\right) + \cdots \tag{2.25}$$

Truncating the excitation operator in this fashion leads to the method known as CCSD (coupled-cluster with singles and doubles). In a similar fashion to full CI truncation, the lowest level of truncation with Hartree-Fock orbitals includes only double excitations and the method is abbreviated as CCD. Including higher excitation terms leads to

methods called CCSD, CCSDT, CCSDTQ, and so on. However, for modestly sized chemical systems the CCSD method is already computationally demanding.

Triple excitations in the cluster expansion can be included in an approximate manner by using perturbation theory. The most widely used technique for achieving this was formulated by Pople and his co-workers[15] and their method is designated (T). The CCSD(T) method is a very good approximation to the full CI expansion and has been called the 'gold standard' of correlation techniques.[16] The CCSD(T) method is used extensively in this thesis to benchmark other methods such as MP theory and density functional theory (discussed below).

Historically, Pople and his co-workers[15] solved the size-consistency problem of the truncated CI expansion by including terms explicitly in order to enforce size consistency. This resulted in a series of methods known as quadratic configuration interaction (QCI). Inclusion of doubles, singles and doubles, and singles, doubles and perturbative triples give methods known as QCID, QCISD and QCISD(T), respectively. Later, it was realised that these methods were essentially equivalent to the coupled-cluster techniques. In fact, QCID is identical to CCD, but CCSD and CCSD(T) include several extra terms that are relatively inexpensive to calculate.

2.4.5 Perturbation Theory

Perturbation theory is an alternative means of representing the correlation problem. As mentioned previously, the Hartree-Fock energy is approximately 99% of the total energy. Therefore, it may be reasonable to assume that the correlation energy could be included through a perturbation on the Hartree-Fock Hamiltonian, i.e.,

$$H = H_0 + \lambda V \tag{2.26}$$

where H is the full time-independent non-relativistic Hamiltonian, H_0 is the Hartree-Fock Hamiltonian, V is the perturbation and λ is a parameter that can be varied from

zero to one according to what extent the perturbation is switched on. The exact wave function and the energy are then expanded in a Taylor series of λ, i.e.,

$$\Psi = \Psi_{HF} + \lambda\Psi^{(1)} + \lambda^2\Psi^{(2)} + \lambda^3\Psi^{(3)} + \cdots + \lambda^i\Psi^{(i)}\cdots \quad (2.27)$$

$$E = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \lambda^3 E^{(3)} + \cdots + \lambda^i E^{(i)}\cdots \quad (2.28)$$

where the subscript *HF* denotes the Hartree-Fock wave function and the superscripted '*i*' represents the order-by-order corrections to the wave function and energy. Taking the perturbation operator as a two-electron operator, the series becomes known as Møller-Plesset (MP) theory, where the sum of the zeroth and first order energies is the Hartree-Fock energy. By solving the perturbation expressions one can obtain a sequence of order-by-order corrections to the Hartree-Fock wave function and energy. Inclusion of the second-order corrections results in MP2 theory, third-order corrections give MP3 theory, and so on. MP2 usually suffices in most calculations and often works well, inclusion of higher levels of excitation causes most calculations to converge quickly towards the full CI result. In some cases, however, the Møller-Plesset series has been shown to converge slowly and in others even diverge.[17]

The convergence behavior of the Møller-Plesset series depends on how good the zeroth-order wave function is as a starting point for the expansion. For open-shell species, the Hartree-Fock starting point may come from either a restricted or an unrestricted calculation. MP calculations based on a restricted formalism are denoted RMP*n* (n = 2,3,...) whereas if they are based on an unrestricted formalism they are denoted UMP*n* (n = 2,3,...). In cases where open-shell species suffer from significant spin-contamination, unrestricted Hartree-Fock starting-points usually give poor results. Some of the open-shell species in this thesis are no exception and the relative performance of RMP*n* and UMP*n* calculations are compared in subsequent chapters.

2.4.6 Ab Initio Hierarchy

The correlation methods outlined in the above sections are able to systematically improve the Hartree-Fock energy by including higher orders of excitation. In other words, they expand the many-electron basis. Full CI and the cluster expansion represent a complete expansion of the many-electron basis. The one-electron basis is the set of functions that is used to represent each molecular orbital. Examples include the Pople-style basis sets STO-3G, 6-31G(d), 6-311+G(d,p), and 6-311+G(3df,2p), and the Dunning-style basis sets cc-pVnZ (n = D, T, Q, 5, ...). If one uses an infinitely large one-electron basis set in conjunction with either full CI or the complete cluster expansion then the exact answer to the non-relativistic time-independent Schrödinger equation (equation 2.7) is obtained.

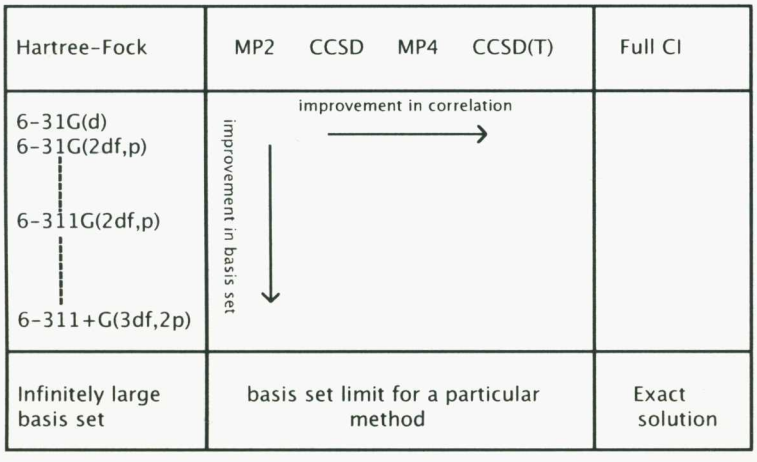


Figure 2.3: Pople Diagram Indicating how to Systematically Improve the Hartree-Fock Solution.

Figure 2.3 gives a Pople diagram,[18] which is a two-dimensional representation of systematic improvements to the Hartree-Fock energy in a minimal one-electron basis set. The horizontal direction classifies methods that systematically improve the many-electron basis while improvements in the one-electron basis set are given in the vertical direction.

2.5 Density Functional Theory

Density functional theory (DFT) is a theory that replaces the $3N$ variable problem ($4N$ including spin variables) of wave-function-based methods with a 3 variable problem, where the variable of interest is the electronic density ($\rho(\mathbf{x},\mathbf{y},\mathbf{z})$). Early progress was made in the development of density functional theory, most notably by Thomas and Fermi,[19] who derived an energetic expression dependent solely on the electronic density. Their expression was the sum of the classical coulomb-repulsion between electrons, the classical coulomb-attraction between electrons and nuclei, and a kinetic energy term derived from the uniform electron gas, which has the form:[20]

$$E_T^{TF27} = \frac{3}{10}(3\pi^2)^{2/3}\int \rho^{5/3}(r)dr \tag{2.29}$$

The subscript T refers to the kinetic energy and the superscript *TF27* refers to the authors initials and the year it was published. Dirac[21] later showed the exchange energy (E_X) of the uniform electron gas, which corrects for electrons interacting with themselves, is related to the density via 2.30.

$$E_X^{D30} = -\frac{3}{2}\left(\frac{3}{4\pi}\right)^{1/3}\int \rho^{4/3}(r)dr \tag{2.30}$$

If this term is added to the Thomas and Fermi expression, the resultant theory is known as Thomas-Fermi-Dirac theory.

Although Thomas-Fermi-Dirac theory made use of the electron density, there was no prescription to say that it was a theoretically valid approach to calculate ground state energies. DFT had to wait until 1964 before it was put onto a firm theoretical footing from two theorems by Hohenberg and Kohn.[22] The first was an existence theorem stating that there exists a unique ground-state electron density that determines

the Hamiltonian and thus all the properties of the system. Quoting from the original paper the theorem, in the words of the authors, is expressed as,

> *"... the external potential $V_{ext}(\mathbf{r})$ is (to within a constant) a unique functional of $\rho(\mathbf{r})$; since, in turn $V_{ext}(\mathbf{r})$ fixes H we see that the full many particle ground state is a unique functional of $\rho(\mathbf{r})$."*[22]

The second theorem states that a trial density delivers an energy that is an upper bound of the true ground-state energy, that is, a density can be variationally optimized in a similar manner to a trial wave function in Hartree-Fock theory.

In another contribution from Kohn,[23] this time in association with Sham, the Hohenberg-Kohn theorems were implemented in a practical formalism. In the Kohn-Sham equations, the kinetic energy is split into two terms. One term could be calculated exactly and represented the kinetic energy of an idealized non-interacting electronic system, i.e., that of Hartree-Fock theory. The other term corrects deficiencies in the first. The second term is usually folded into the exchange-correlation functional. Thus, the key result is the functional representation of the energy in terms of a sum of functionals (2.31):

$$E[\rho] = E_T[\rho] + E_V[\rho] + E_J[\rho] + E_{XC}[\rho] \qquad (2.31)$$

where $E_T[\rho]$ is the kinetic energy functional of an idealized non-interacting electronic system, $E_V[\rho]$ is the coulomb energy of the electronic attraction to the nuclei, $E_J[\rho]$ is the coulomb-repulsion functional, all of which are known exactly, and $E_{XC}[\rho]$ is the so-called exchange-correlation function the form of which is unknown. The main unknown of density functional theory is the exchange-correlation functional. It is common to split the exchange-correlation into terms representing the exchange energy

and the correlation energy, i.e., $E_{XC}[\rho] = E_X[\rho] + E_C[\rho]$. DFT methods are usually designated by the names given to the exchange and correlation functionals.

2.5.1 Local Density Approximations

The initial attempts by Thomas, Fermi and Dirac were based on the uniform electron gas, in which the electrons move on a positively charged background. Using the Kohn-Sham equations, closed-form expressions can be obtained for all terms except for the exchange and correlation functionals. In 1951,[24] Slater introduced a modified form of Dirac's exchange term containing an empirical constant modifier in front of the integral to give better agreement for atomic and molecular systems. Slater's exchange functional is usually abbreviated with an *S* and his method, which is a modification of Thomas-Fermi-Dirac theory, is called the $X\alpha$ method.

In subsequent years, highly accurate numerical results were generated for the correlation contribution of the uniform electron gas. Correlation functionals could then be obtained by fitting the results to analytical expressions with sophisticated interpolation schemes. The most well known of them was published by Vosko, Wilk and Nusair,[25] and their functional is abbreviated to VWN. The VWN and *S* functionals combined with the other terms, classical coluombic and the Hartree-Fock kinetic energy term, are now synonymous with local density approximations (LDA) or local spin density approximation (LDSA) if electrons of opposite spin are treated differently.

2.5.2 Generalized Gradient Approximation

Uniform electron densities are inherently deficient for describing atoms and molecules. To allow the density to be non-uniform, gradient corrections can be introduced through a Taylor expansion. However, this approximation violates some of the principal physics of the system (e.g. the so-called exchange-correlation "hole" needs to contain exactly one electronic charge). This frequently results in gradient-expanded methods giving poorer results than LDA. However, if the hole condition is enforced by

setting terms in the expansion to zero, then one obtains the generalized gradient approximation (GGA).

The first functional that appeared in the literature with dependence on the gradient came from a kinetic energy functional derived by von Weizsacker,[26] which was obtained by adding density ripples to the uniform electron gas:

$$E_T^{W35} = \frac{3}{10}(6\pi^2)^{2/3}\int \rho^{5/3}(r)dr \quad + \quad \frac{1}{8}\int \rho^{5/3}(r)x^2(r)dr$$
$$= E_T^{TF27} \quad + \quad \frac{1}{8}\int \rho^{5/3}(r)x^2(r)dr \tag{2.32}$$

where,

$$x(r) = \frac{|\nabla\rho(r)|}{\rho^{4/3}(r)} \tag{2.33}$$

In a similar fashion, a gradient-corrected form of Dirac's exchange functional was derived by Sham[27] and Kleinmann[28]

$$E_X^{SK31} = E_X^{D30} \quad - \quad \frac{5}{(36\pi)^{5/3}}\int \rho^{4/3}(r)x^2(r)dr \tag{2.34}$$

All of the preceding functionals have been derived from first principles without recourse to experiment or fitting procedures, so DFT at this stage may be termed an *ab initio* theory. However, combining the gradient-corrected functionals in the Kohn-Sham scheme gives less than satisfactory results. The breakthrough (or breakdown) of DFT was the introduction of parameterized functionals. The parameterization of functionals has been very successful for giving accurate molecular properties such as atomization energies, ionization energies, electron affinities, and proton affinities to within about 20 kJ mol^{-1} of experimentally derived numbers. Normally, "the pragmatists" shrug and are content because of the substantial computational gain in

obtaining such results. Some theoreticians on the other hand, argue that the essential physics becomes lost in a forest of parameters. No insight into the physics may be gained from these functionals thus leaving no firm outlook on where to go in the future. A tongue in cheek expression for this sentiment says "give me enough parameters and I can make an elephant, give me just one more and I can make his tail wag!"

The following paragraphs outline the most successful of these elephants, which although it may be relegated to a semi-empirical domain, have been highly successful. The first was a gradient-corrected form of the SK71 exchange functional, published in 1988 by Becke[29]

$$E_X^{B88} = E_X^{D30} - \beta \int \rho^{4/3}(r) \frac{x^2(r)}{1 + 6\beta x(r) \sinh^{-1} x(r)} dr \qquad (2.35)$$

The β parameter was determined to be 0.0042 by fitting equation 2.35 to the Hartree-Fock exchange energies of the inert gas elements. The introduction of the denominator dependent on $x(r)$ effectively damps the integrand as $x(r)$ grows. This was proved important in order to overcome the divergent behaviour of the SK71 functional. An alternative and simpler way of damping the denominator is to introduce an $x^{3/2}$(r) dependence in the integrand and reparameterizing β. Developed in 1996 by Gill,[30] this functional has similar accuracy to Becke's 1988 functional but is less widely used despite the simpler form:

$$E_X^{G96} = E_X^{D30} - \gamma \int \rho^{4/3}(r) x^{3/2}(r) dr \qquad (2.36)$$

where γ = 1/137. Perdew and Wang[31] have also proposed a gradient-corrected exchange functional (E_x^{PW86}) similar to Becke's.

In this section, the only mention of a correlation functional was the E_C^{VWN} functional, which is a complicated expression fitted to numerical results of the uniform

electron gas density. In 1988, Lee Yang and Parr[32] developed a correlation functional (E_C^{LYP}) based on the helium atom, which contains four parameters. Like E_C^{VWN} it is also a complicated expression.

2.5.3 Hybrid Functionals

The connection between the idealized non-interacting system and the fully interacting system can be related through an adiabatic correction, whereby the total exchange-correlation functional is integrated over a parameter that turns the electron interaction on. The simplest approximation to an integral is to average the sum between the end-points. When this is done for the exchange-correlation functional, it contains half the exact Hartree-Fock exchange (where the electrons are non-interacting) and half of an exchange-correlation functional where the electrons are interacting. If the interacting functional is approximated by LSDA, then the hybrid functional is called Becke's half and half method.[33] Becke went one step further and developed a three-parameter exchange-correlation functional. In conjunction with the LYP correlation function, the well-known B3-LYP combination of functionals is formed,[34] which has received widespread use in the literature and it is used extensively in this thesis. This functional has the form:

$$\begin{aligned} E^{B3LYP} = E_T \;+\; E_V \;+\; E_J \;+\; (1-c_1)E_X^{D30} \;+\; c_1 E_X \\ +\; c_2 \Delta E_X^{B88} \;+\; (1-c_3)E_C^{VWN} \;+\; c_3 E_C^{LYP} \end{aligned} \tag{2.37}$$

where the three c_i parameters are determined from experimental data.

2.5.4 What is Missing?

Success of the hybrid functionals over the gradient-corrected functionals suggests that some of the physics is still missing from DFT. The exchange energy can be expressed in terms of the interaction between the charge density ($\rho(r_1)$) of an electron

of spin σ with the exchange hole $h_x(r_1,r_2)$, where the exchange hole is identified with the probability density owing to antisymmetry. Thus the exchange energy is:

$$E_X = \frac{1}{2}\int \frac{\rho(r_1)h_X(r_1,r_2)}{r_{12}}dr \tag{2.38}$$

In 1983, Becke[35] derived a complicated Taylor expansion for the exchange hole $h_x(r_1,r_2)$ in spherically-averaged coordinates *viz*:

$$h_X(r_1,r_2) = \rho(r_1) + \frac{1}{6}\left(\nabla^2\rho(r_1) - 2\gamma\left(\tau(r_1) - \frac{1}{4}\frac{(\nabla\rho(r_1))^2}{\rho(r_1)}\right)\right)r_2 + \cdots \tag{2.39}$$

The important aspect of this expansion is there are four terms depending on the density, two of which have been met before, the density, and the gradient of the density. Two other terms have not yet been considered, the laplacian of the density $\nabla^2\rho(r_1)$ and the kinetic energy density $\tau(r_1)$, which is defined as:

$$\tau(r_1) = \sum_i |\nabla\Psi_i|^2 \tag{2.40}$$

Functionals including terms depending on the kinetic energy density are known as meta-gradient corrected (meta-GGA) functionals. Hybrid functionals containing these terms have also come into existence. In this thesis, two hybrid meta-GGAs have been used and go by the acronyms BMK[36] and MPWB1K.[37]

2.6 Compound Methods and Model Chemistries

One of the goals of computational quantum chemistry is to achieve highly accurate results for certain properties such as ionisation energies (IEs), electron

affinities (EAs), and total atomization energies (TAEs). To this end, chemical accuracy has been defined accordingly as ± 2 kcal mol^{-1}. In strategies to obtain a range of properties (such as IEs, EAs and TAEs) for various compounds, it becomes appropriate to define a theoretical-model chemistry. In order to compare different models for accuracy and predictive power models need to contain certain properties. First identified by Pople[18] these properties are:

(a) *Unique and well-defined:* Energies and wave functions obtained from approximating the Schrödinger equation should be only specified in terms of nuclear positions and the numbers and spins of electrons in the molecule.

(b) *Continuity:* Potential energy surfaces should be continuous with respect to nuclear displacements.

(c) *Unbiased*: Tthere should be no appeal to chemical intuition when setting up a calculation.

(d) *Size-consistent*: If different compounds are going to be compared, then the model must scale linearly with the number of particles (*size-extensive*). However, this can be difficult to achieve and is usually relaxed to size-consistency, whereby a property calculated for two infinitely-separated molecules should be the sum of the same property when calculated individually.

A number of model chemistries have been examined already. These include Hartree-Fock theory, full CI, coupled-cluster techniques, perturbation theory, and the density functional theories, with the proviso that a single basis set is used in association with these methods. Unfortunately chemical accuracy can be only be assured from methods whose computational cost scales poorly (CCSD(T) and full CI in association with a large basis set). Although density functional theory provides a cheap alternative to these methods, it suffers from not being systematically improvable. Therefore DFT needs to be calibrated, thus necessitating the use of methods such as CCSD(T) anyway. An alternative means of obtaining chemical accuracy without recourse to methods such

as CCSD(T) with an infinitely large basis set is to use compound methods. The aim of compound model-chemistries is to approximate a large-basis-set CCSD(T) calculation at a significantly reduced computational cost.

2.6.1 Gaussian-n (Gn) Methods

Gaussian-*n* methods (G*n*, *n* = 1, 2 or 3) are a series of methods that use an additive approximation for increasing basis set size. The first introduction of these methods was with the G1 method in 1989.[38] Later improvements were made to the initial G1 theory bringing about the introduction of G2[39] and G3[40] theories. The inherent approximation of basis set additivity is used among all these methods and since this thesis uses variants of G3 theory exclusively, the explicit details of only G3 will be explained.

Looking at the Pople diagram of Figure 2.3, one could reason that the energy lowering obtained from a particular theory when the basis set size is increased might be approximately the same for another method. For instance, the energy difference between two MP2 calculations with basis sets of different size could be approximately the same as the difference of two CCSD(T) calculations with the same basis sets. Therefore, one could make an approximation such as: CCSD(T)/6-311+G(3df,2p) ≈ CCSD(T)/6-31G(d) + [MP2/6-311+G(3df,2p) – MP2/6-31G(d)]. This is the essence of all the G*n* methods, though in practice an incremental introduction of correlation and basis-set size is used.

2.6.1.1 G3: Using the implicit additivity assumption of basis set size increase, the G3 energy at zero-kelvin can be expressed as:

$$E_0 = E\left[MP4/6-31G(d)\right] + \Delta E(+) + \Delta E(2df,p) \\ \Delta E(QCI) + \Delta E(G3large) + E(SO) + E(HLC) + E(ZPE) \qquad (2.41)$$

The base energy is an MP4/6-31G(d) calculation and added to that are two basis set corrections at the MP4 level for diffuse functions and for additional polarization functions:

$$\Delta E(+) = [\,MP4/6-31+G(d) \;-\; MP4/6-31G(d)\,] \tag{2.42}$$

$$\Delta E(2df,p) = [\,MP4/6-31G(2df,p) \;-\; MP4/6-31G(d)\,] \tag{2.43}$$

A QCISD(T) calculation is then added to account for higher levels of correlation:

$$E(QCI) = [\,QCISD(T)/6-31G(d) \;-\; MP4/6-31G(d)\,] \tag{2.44}$$

In addition, a correction for larger basis set effects and a correction for the non-additivity caused in the basis set extension are included:

$$\begin{aligned} E(G3large) = [\,&MP2(full)/G3large \;-\; MP2/6-31+G(d) \\ &-\; MP2/6-31G(2df,p) \;+\; MP2/6-31G(d)\,] \end{aligned} \tag{2.45}$$

E(SO) is a spin-orbit correction for atoms that has been found to be important for halide-containing coumpounds.[41] E(HLC) is an empirical "higher-level correction" based on the numbers of α and β electrons. For molecules it is $-An_\beta - B(n_\alpha - n_\beta)$, where A=6.386 and B=2.977 *m*Hartrees, and for atoms it is $-Cn_\beta - D(n_\alpha - n_\beta)$, where C=6.219 and D=1.185 *m*Hartrees. Finally, a zero-point vibrational energy is added from a frequency calculation on an HF/6-31G(d) equilibrium structure, which is scaled by 0.8929 to account for known deficiencies at this level.[42] All the other corrections have been calculated on an MP2(full)/6-31G(d) optimized structure. Thus, the final G3 energy is approximating a QCISD(T)/G3large//MP2(full)/6-31G(d) calculation.

2.6.1.2 G3X: The G3X method[43] is a variant of G3, which incorporates three changes. Firstly, geometries are obtained with B3-LYP/6-31+G(2df,p). Secondly, zero-point vibrational energies are obtained with B3-LYP/6-31+G(2df,p) and scaled by

0.9854. Finally, an additional basis set correction is included at the HF level for second-row atoms, HF/G3Xlarge – HF/G3large, where the G3Xlarge basis set incorporates a set of 7 pure *g* functions.

2.6.1.3 G3(MP2) methods: Both of the G3 and G3X methods can be approximated by replacing the computationally expensive MP4 calculations with cheaper MP2 calculations without a significant loss of accuracy.[43]

2.6.1.4 RAD variants of G3 methods: The G3 methods and their variants have been found to perform well for closed-shell systems but for some radicals they lead to spurious results. It was thought that the anomalous results were caused by basing these methods on an unrestricted formalism for MP calculations on open-shell species. The "RAD" variants of the G3 methods[44] introduce a number of changes to overcome the difficulties associated with an unrestricted formalism.

The principal features of the RAD methods include (a) B3-LYP geometries and scaled frequencies, (b) the use of URCCSD(T) for the inclusion of higher orders of correlation, replacing the UQCISD(T) calculation, and (c) the use of RMP in place of UMP for basis set extensions. These changes have led to the formulation of the G3-RAD, G3X-RAD, G3(MP2)-RAD and G3X(MP2)-RAD methods, which have shown to give improved accuracy for properties of radicals such as the their heat of formation.

2.6.2 Complete Basis Set Methods (CBS)

The complete basis set methods are another set of methods that attempt to use lower-level methods to approximate a highly correlated method, such as CCSD(T), with a complete basis set. The most popular variant of these methods is known by the acronym CBS-QB3 and is used widely in this thesis. Chapter 3 introduces a new method that is based on CBS-QB3. In a similar manner to the RAD variants of G3, the new method of Chapter 3 reformulates CBS-QB3 using a restricted formalism in order to improve the accuracy of calculations on radicals.

The main difference between the Gaussian methods and the CBS methods is that the CBS methods extrapolate the basis set extension of an MP2 calculation to the infinite limit. It had been observed by Schwartz[45] that the second-order correction to the energy followed an angular momentum dependence on the basis set size. This relationship was used to formulate an extrapolation scheme of the MP2 energy based on pair-natural orbitals by Petersson and several collaborators.[46] For the CBS-QB3 method the total energy is obtained from the following calculations:

(i) UB3-LYP/6-311G(2d,d,p) geometry optimization and frequencies,

(ii) UMP2/6-311+G(3d2f,2df,2p) energy and CBS extrapolation,

(iii) UMP4(SDQ)6-31+G(d(f),p) energy, and

(iv) UCCSD(T)/6-31+G† energy.

The calculations (ii)–(iv) are performed on the geometry obtained from step (i). The basis sets are modifications of the usual Pople-style sets, written with a more generalized notation to designate that different polarization functions are used on different atoms. Thus, 6-311+G(3d2f,2df,2p) means there are 3d2f polarization functions on second-row atoms, 2df polarization functions on first-row atoms, and 2p functions on hydrogen. Again, 6-31+G(d(f),p) indicates that there are d functions on first- and second-row atoms, an f function on selected second-row atoms, and a p function on hydrogen. The 6-31+G† basis set is a modification of 6-31+G(d) in which polarization function exponents are taken from 6-311G(d,p). The total CBS-QB3 energy is calculated from:

$$\begin{aligned} E(CBS\text{-}QB3) = E(UMP2) &+ \Delta E(CBS^{(2)}) + \Delta E(spin) \\ &+ \Delta E(CCSD(T)) + \Delta(CBS\text{-}int) \\ &+ \Delta E(emp) + \Delta E(ZPVE) \end{aligned} \tag{2.46}$$

Here, $E(UMP2)$ is the UMP2 energy calculated with the 6-311+G(3d2f,2df,2p) basis set, $\Delta E(CBS^{(2)})$ is obtained from the extrapolation of the second-order pair correlation energies to the complete-basis-set (CBS) limit, and $\Delta E(UMP34)$ and $\Delta E(CCSD(T))$ are given by:

$$\begin{aligned} \Delta E(UMP34) = E\left[UMP4(SDQ)/6\text{-}31+G(d(f),p)\right] \\ - E\left[UMP2/6\text{-}31+G(d(f),p)\right] \end{aligned} \tag{2.47}$$

and

$$\begin{aligned} \Delta E(UCCSD(T)) = E\left[UCCSD(T)/6\text{-}31+G^{\dagger}\right] \\ - E\left[UMP4(SDQ)/6\text{-}31+G^{\dagger}\right] \end{aligned} \tag{2.48}$$

The empirical correction to the higher-order correlation energy, $\Delta E(emp)$, in equation 2.43 is given by:

$$\Delta E(emp) = -5.79 m E_h \sum_{i=1}^{n_\beta} \left[\sum_{\mu=1}^{N_{virt}+1} C_{\mu_{ii}} \right]^2 |S|_{ii}^2 \tag{2.49}$$

where $|S|_{ii}$ is the absolute overlap integral:

$$|S|_{ii} = \int \left| {}^{\alpha}\varphi_i \, {}^{\beta}\varphi_i \right| d\tau \tag{2.50}$$

between the most similar α and β orbitals, and the interference factor, $\left[\sum_{\mu} C_{\mu}\right]^2$, is the square of the trace of the first-order wave function. This empirical higher-order correction is necessary to compensate for the very small basis sets employed in 2.47 and

2.48. The $\Delta E(CBS\text{-}int)$ correction to $\Delta E(MP34)$ and $\Delta E(CCSD)$ is obtained from the interference correction[47] to the pair energies in calculation (ii). This interference term corrects higher-order perturbation theory for near degeneracies.[47] These two higher-order corrections to $\Delta E(UMP34)$ and $\Delta E(UCCSD(T))$ are an attempt to estimate the CBS limit of these contributions.

No attempt is made to extrapolate the UHF/6-311+G(3d2f,2df,2p) energy to the CBS limit. The spin-contamination correction term, $\Delta E(spin)$, is given by,

$$\Delta E(spin) = -9.54 m E_h \Delta\left\langle S^2 \right\rangle \tag{2.51}$$

which adds a correction proportional to the deviation from ideal values of the spin-squared expectation value $\langle S^2 \rangle$ calculated for the UHF/6-311+G(3d2f,2df,2p) wave function. Finally, the $\Delta E(ZPVE)$ term is obtained from scaled (by 0.9900) frequencies calculated in (i). Chapter 3 examines the role of the empirical correction for spin-contamination.

2.6.3 Weizmann Methods (Wn)

The major pitfall of the previous sets of compound methods is that they contain some sort of parameterization against experimental data. This is appropriate for many first-row-containing compounds where there exists reliable experimental data but for second-row-containing compounds reliable experiments are sparse. In addition, though these methods quite often reach chemical accuracy they do give largest deviations that are an order of magnitude larger than the requisite precision. The Weizmann methods are a set of methods that have been established to offer benchmark quality results, defined to have an average deviation of 1 kJ mol^{-1}. They should also not have largest deviations exceeding chemical accuracy. In addition, they should be applicable to compounds that contain up to six heavy atoms (atoms not including hydrogen and helium) and finally they should not contain empirical parameters. The first two

methods developed with these criteria are known as the Weizmann-1 and Weizmann-2 (W1 and W2) theories.[48] More recently, the Weizmann-3 (W3) has been outlined in the literature but remains computationally impractical for regular use.[49]

The W1 and W2 methods rely on extrapolation schemes to give CCSD(T)/CBS energies. There are many similarities between the W1 and W2 methods, the main difference concerns the basis sets used in the extrapolation steps. The remaining paragraphs will therefore discuss the steps involved in the W1 method.

The W1 method requires the following calculations:

(1) Geometry optimization at B3-LYP/VTZ+1, where the "+1" denotes the addition to second-row atoms of the highest-exponent *d* function from the V5Z basis set. Frequencies obtained at the same level with zero-point energy scaled by 0.985.

(2) SCF single-point energy calculations using the aug′-cc-pVQZ+2d1f and aug′-cc-pVTZ+2d1f basis sets. Extrapolation of the SCF component to the energy using a two-point extrapolation formula ($E_{SCF}(l) = A + B/l^5$) where $l = 4$ and 3 for the quadruple- and triple-zeta basis sets that are used.

(3) CCSD single-point energy calculations using the aug′-cc-pVQZ+2d1f and aug′-cc-pVTZ+2d1f basis sets. Extrapolation of the CCSD component to the energy using a two-point extrapolation formula ($E_{CCSD}(l) = A + B/l^{3.22}$) where $l = 4$ and 3 for the quadruple- and triple-zeta basis sets that are used. The exponent of 3.22 is an empirical parameter determined for the W1 method determined from the results of W2. Although this seems to violate the principles of the Weizmann methods, there is still no reliance on experimental data.

(4) CCSD(T) single-point energy calculations using the aug′-cc-pVTZ+2d1f and aug′-cc-pVDZ+2d1f basis sets. Extrapolation of the (T) component to the energy using a two-point extrapolation formula ($E_{CCSD}(l) = A + B/l^{3.22}$) where $l = 3$ and 2 for the triple- and double-zeta basis sets that are used.

(5) CCSD(T)(*full*)/MTSmall and CCSD(T)/MTSmall single-point calculations, where the difference accounts for correlations between core electrons.

(6) Addition of a scalar relativistic correction using an ACPF/MTSmall calculation.

The W1 method using these steps has been found to give excellent accuracy for thermochemical calculations, and where possible has been used to benchmark other calculations in this thesis.

2.7 Software

The programmes used to perform the calculations in this thesis are: ACES II,[50] GAMESS US,[51] GAUSSIAN 03,[52] MOLPRO 2002.3,[53] and NWCHEM 4.7.[54]

2.8 References

(1) (a) Dirac, P. A. M. *Proc. Roy. Soc. (London)* **1929**, 123, 714. (b) Italics added.

(2) Moores' Law is attributed to Gordon E Moore and states that computing power increases in an exponential fashion. Interestingly, Moore was born in the same year that Dirac's statement was published.

(3) de Broglie, L. *Ann. Physik,* **1925**, *3*, 25.

(4) Schrodinger, E. *Phys. Rev.* **1926**, *28*, 1049.

(5) Born, M.; Oppenheimer, R. *Annalen der Physik* **1927**, *84*, 457.

(6) (a) Slater, J. C. *Phys. Rev.* **1929**, *34*, 1293 – 1323. (b) Slater, J. C. *Phys. Rev.* **1930**, *35*, 509.

(7) (a) Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1969**, *51*, 2657. (b) Collins, J. B.; Schleyer, P. v. R.; Binkley, J. S.; Pople, J. A. *J. Chem. Phys.* **1976**, *64*, 5142. (c) Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1980**, *102*, 939. (d) Gordon, M. S.; Binkley, J. S. ; Pople, J. A.; Pietro, W. J.; Hehre, W. J. *J. Am. Chem. Soc.* **1982**, *104*, 2797. (e) Pietro, W. J.; Francl, M. M.; Hehre, W. J.; Defrees, D. J.; Pople, J. A.; Binkley, J. S. *J. Am. Chem. Soc.* **1982**, *104*, 5039. (f) Dobbs K. D.; Hehre, W. J. *J. Comp. Chem.* **1986,** *7*, 359. (g) Dobbs K. D.; Hehre, W. J. *J. Comp. Chem.* **1987**, *8*, 861. (h) Dobbs K. D.; Hehre, W. J. *J. Comp. Chem.* **1987**, *8*, 880. (i) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724. (j) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257. (k) Hariharan, P. C.; Pople, J. A. *Mol. Phys.* **1974**, *27*, 209. (l) Gordon, M. S. *Chem. Phys. Lett.* **1980**, *76*, 163. (m) Hariharan, P. C.; Pople, J. A. *Theo. Chim. Acta* **1973**, *28*, 213. (n) Blaudeau, J.-P.; McGrath, M. P.; Curtiss, L. A.; Radom, L. *J. Chem. Phys.* **1997**, *107*, 5016. (o) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; DeFrees, D. J.; Pople, J. A.; Gordon, M. S. *J. Chem. Phys.* **1982**, *77*, 3654. (q) Binning Jr, R. C.; Curtiss, L. A. *J. Comp. Chem.* **1990**, *11*, 1206. (r) Rassolov, V. A.; Pople, J. A.; Ratner, M. A.; Windus, T. L. *J. Chem. Phys.* **1998**, 109, 1223. (s) Rassolov, V. A.; Ratner, M. A.; Pople, J. A.; Redfern, P. C.; Curtiss, L. A. *J. Comp. Chem.* **2001**, *22*, 976. (t) McLean, A. D.;

Chandler, G. S. *J. Chem. Phys.* **1980**, *72*, 5639. (u) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650.

(8) (a) Roothaan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69. (b) Hall, G. C. *Proc. Roy. Soc. (London)* **1951**, *A205*, 541.

(9) Pople, J. A.; Nesbet, R. K. *J. Chem. Phys.* **1954**, *22*, 571.

(10) Bally, T; Borden, W. T. In *Reviews in Computational Chemistry*, Lipkowitz, K. B.; Boyd, D. B. (eds.), **13**, pp. 1–97, John Wiley and Sons, Inc., New York (1999).

(11) In general, a calculation has a method and a basis set associated with it. The shorthand notation for describing the calculation is then *method/basis set.* The example given in the text is UHF/6-31G(d)

(12) (a) Hylleraas, E. A. *Z. Physik.* **1928**, *48*, 469. (b) Hylleraas, E. A. *Z. Physik.* **1929**, *54*, 347. (c) Hylleraas, E. A. *Z. Physik.* **1930**, *65*, 209.

(13) Čižek, J. *J. Chem. Phys.* **1966**, *45*, 4256.

(14) Paldus, J. in *Theory and Applications of Computational Chemistry The First Forty Years,* Dykstra, C. E.; Frenking, G.; Kim, K. S.; Scuseria, G. E. (eds), Elsevier, New York (2005).

(15) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. *J. Chem. Phys.* **1987**, *87*, 5968.

(16) Attributed to T. H. Dunning.

(17) (a) Gill, P. M. W.; Pople, J. A.; Radom, L.; Nobes, R. H. *J. Chem. Phys.* **1988**, *89*, 7307. (b) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007. (c) Olsen, J.; Jørgensen, P.; Helgakar, T.; Christiansen, O. *J. Chem. Phys.* **2000**, *112*, 9736. (d) Leininger, M. L.; Allen, W. D.; Schaefer, H. F., III; Sherrill, C. D. *J. Chem. Phys.* **2000**, *112*, 9213.

(18) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley, New York (1986).

(19) (a) Thomas, L. H. *Proc. Camb. Phil. Soc.* **1927**, *23*, 542. (b) Fermi, E. *Rend. Accad. Lincei*, **1927**, *6*, 602.

(20) (a) The labelling of this term follows that of Gill (b) Gill, P. M. W. In *The Encyclopaedia of Computational Chemistry,* Schleyer, P. v R.; Allinger, N. L.; Clark, T.; Gasteiger, J.; Kollman, P. A.; Schaefer, H. F. III; Schreiner, P. R. (eds.), **1**, pp. 678 – 689, Wiley, New York (1998).

(21) Dirac, P. A. M. *Proc. Camb. Phi. Soc.* **1930**, *26*, 376.

(22) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, 864.

(23) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, 1133.

(24) Slater, J. C. *Phys. Rev.* **1951**, *81*, 385.

(25) Vosko, S. J.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *88*, 322.

(26) von Weizsacher, C. F. *Z. Physik,* **1935**, *96*, 431.

(27) Sham, L. J. In *Computational Methods in Band Theory*, Marcus, P. J.; Janak, J. F.; Williams, A. R., (eds.), Plenum, New York (1971).

(28) Kleinmann, L.; Lee, S. *Phys. Rev. B,* **1988**, *37,* 4634.

(29) Becke, A. D. *Phys. Rev. A*, **1988**, *38*, 3098.

(30) Gill, P. M. W. *Mol. Phys.* **1996**, *86*, 433.

(31) Perdew, J. P.; Wang, Y. *Phys. Rev. B*, **1986**, *33*, 8800.

(32) Lee, C.; Yang, W.; Parr , R. G. *Phys. Rev. B,* **1988**, *37*, 785.

(33) Becke, A. D. *J. Chem. Phys.* **1992**, *98*, 1372.

(34) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(35) Becke, A. D. *Int. J. Quantum Chem.* **1983**, *13*, 1915.

(36) Boese, A. D.; Martin, J. M. L. *J. Chem. Phys.* **2004**, *121*, 3405.

(37) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6908.

(38) Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. *J. Chem. Phys.* **1989**, *90*, 5622.

(39) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, *94*, 7221.

(40) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764.

(41) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem Phys.* **1997**, *109*, 42.

(42) Pople, J. A.; Schlegel, H. B.; Krishnan, R.; Defrees, D. J.; Binkley, J. S.; Frisch, M. J.; Whiteside, R. A.; Hout, R. F.; Hehre, W. J. *Int. J. Quantum Chem.* **1981**, *15*, 269.

(43) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **2001**, *114*, 108.

(44) Henry, D. J.; Sullivan, M. B.; Radom, L. *J. Chem. Phys.* **2003**, *118*, 4849.

(45) Schwartz, C. *Phys. Rev.* **1962**, *126*, 1015.

(46) Ochterski, J. W.; Petersson, G. A.; Montgomery, J. A. *J. Chem. Phys.* **1996**, *104*, 2598.

(47) Petersson, G. A.; Nyden, M. R. *J. Chem. Phys.* **1981**, *75*, 3423.

(48) Martin, J. M. L.; de Oliveira, G. *J. Chem. Phys.* **1999**, *111*, 1843.

(49) Boese, A. D.; Oren, M.; Atasoylu, O.; Martin, J. M. L.; Kallay, M.; Gauss, J. *J. Chem. Phys.* **2004**, *120*, 4129.

(50) R. J. Bartlett, J. F. Stanton, J. Gauss, *et al.*, ACES II, Quantum Theory Project, (University of Florida, Gainsville, 1992).

(51) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N. GAMESS, version 12 dec 2003 (r2); Iowa State University, Ames, 2003.

(52) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M. Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A. Gaussian 03, Revision B.03; Gaussian, Inc.; Pittsburgh PA, 2003.

(53) Werner, H.-J.; Knowles, P. J.; Lindh, R.; Schu¨tz, M.; Celani, P.; Korona, T.; Manby, F. R.; Rauhut, G.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Lloyd, A. W.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T. *MOLPRO 2002.6, a package of ab initio programs*; University of Birmingham, Birmingham, U.K., 2003.

(54) (a) Apra, E.; Windus, T.L.; Straatsma, T.P.; Bylaska, E.J.; de Jong, W.; Hirata, S.; Valiev, M.; Hackler, M.; Pollack, L.; Kowalski, K.; Harrison, R.; Dupuis, M.; Smith, D.M.A; Nieplocha, J.;

Tipparaju V.; Krishnan, M.; Auer, A.A.; Brown, E.; Cisneros, G.; Fann, G.; Fruchtl, H.; Garza, J.; Hirao, K.; Kendall, R.; Nichols, J.; Tsemekhman, K.; Wolinski, K.; Anchell, J.; Bernholdt, D.; Borowski, P.; Clark, T.; Clerc, D.; Dachsel, H.; Deegan, M.; Dyall, K.; Elwood, D.; Glendening, E.; Gutowski, M.; Hess, A.; Jaffe, J.; Johnson, B.; Ju, J.; Kobayashi, R.; Kutteh, R.; Lin, Z.; Littlefield, R.; Long, X.; Meng, B.; Nakajima, T.; Niu, S.; Rosing, M.; Sandrone, G.; Stave, M.; Taylor, H.; Thomas, G.; van Lenthe, J.; Wong, A.; Zhang, Z.; "NWChem, A Computational Chemistry Package for Parallel Computers, Version 4.7" (2005), Pacific Northwest National Laboratory, Richland, Washington 99352-0999, USA. (b) Kendall, R.A.; Apra, E.; Bernholdt, D.E.; Bylaska, E.J.; Dupuis, M.; Fann, G.I.; Harrison, R.J.; Ju, J.; Nichols, J.A.; Nieplocha, J.; Straatsma, T.P.; Windus, T.L.; Wong, A.T.; Computer Phys. Comm. 2000, 128, 260–283.

CHAPTER 3

The ROCBS-QB3 Procedure

3.1 Introduction

A major goal of modern computational chemistry is the calculation of accurate molecular thermochemistry.[1] This has become a practical reality for a wide variety of molecules in recent years through the development of composite theoretical methods such as the Gaussian (G*n*) procedures of Curtiss, Raghavachari and Pople,[2] the complete-basis-set (CBS) methods of Petersson and co-workers,[3-6] and the Weizmann (W*n*) methods of Martin and co-workers.[7-9]

Particular problems in the determination of thermochemical properties can arise for open-shell systems if there is significant spin contamination in methods based on an unrestricted wave function. A number of procedures have been introduced to try to combat such problems.[10,11] For example, the use of B3-LYP geometries and frequencies in place of UHF and/or UMP2 is a useful step in this direction[11,12] and has been incorporated in the popular CBS-QB3 method[6] (see section 2.8.2). However, these improvements in geometries and zero-point energies do not address the errors that are associated with spin contamination in the electronic energies themselves.

CBS-QB3 includes a "spin-correction" term (equation 2.51 and repeated here as equation 3.1) that is proportional to the deviation in the expectation value of the spin-squared operator $\langle S^2 \rangle$ from the values appropriate for a pure spin state (e.g., 0.75 for a doublet).

$$\Delta E(spin) = -9.54 m E_h \Delta \langle S^2 \rangle \qquad (3.1)$$

In a highly spin-contaminated species such as the cyano radical, this spin correction, $-9.54 m E_h \Delta \langle S^2 \rangle$, reduces the error in the UCCSD(T) energy to a value comparable to that for species with little or no spin contamination (Figure 3.1). However, in a number of recent studies of species with less severe spin contamination,[13-16] it has been found that inclusion of the spin-correction term can sometimes lead to poorer agreement with

experiment or with the results of higher-level theoretical procedures such as W1 or W2. This has led to the use of CBS-QB3 without the spin-correction term, in a method denoted U-CBS-QB3.[13-16] The validity of this approach is re-examined in this Chapter.

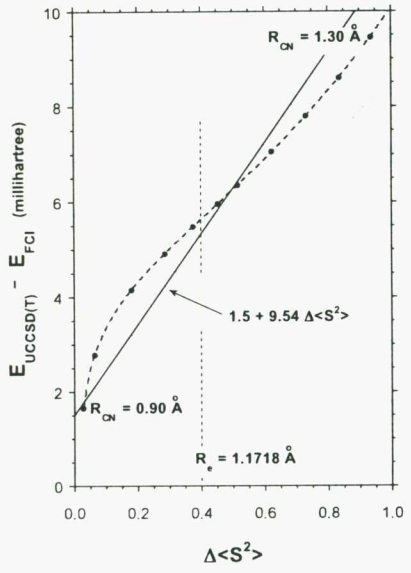


Figure 3.1: The minimum energy UHF wave function for the cyano radical, •C≡N, mixes the A^4S excited state with the X^2S ground state, leading to an error in $\langle S^2 \rangle$: $d\langle S^2 \rangle = \langle S^2 \rangle_{calc} - 0.75$. This UHF spin contamination and the associated error in the UCCSD(T) energy can be varied over a wide range by modest changes in the bond length, R_{CN}. The intercept of 1.5 mHartrees represents the CCSD(T) error in the absence of spin contamination. Calculations have been carried out in association with the cc-pVDZ basis set.

A potential alternative means of addressing the spin-correction problem is through the use of spin-restricted procedures in place of spin-unrestricted procedures, i.e., through a restricted-open-shell CBS-QB3 (ROCBS-QB3) procedure.[14] Unfortunately, spin-restricted methods introduce their own well-known problems.[17] For example, neither RCCSD(T) nor UCCSD(T) give an acceptable potential energy curve for the F_2 molecule (Figure 3.2). The RCCSD(T) energy dissociates to the wrong limit, whereas the UCCSD(T) energy is in error by more than 30 kJ mol^{-1} at a bond length (~1.9 Å or ~1.3 R_e) that might apply to a transition structure. The CBS-QB3 spin correction reduces the large errors at ~1.9 Å, but introduces a reduced version of the

kind of dissociation-limit problems associated with RCCSD(T). We note that while we define the new restricted-open-shell version of CBS-QB3 as ROCBS-QB3 we retain the prefix '*R*' with all other methods that make use of a restricted wave function irrespective of the numbers of unpaired electrons.

In the present Chapter, an ROCBS-QB3 procedure has been implemented, and we have compared its performance with that of standard CBS-QB3. The new ROCBS-QB3 model is found to perform comparably to or even slightly better than the standard CBS-QB3 model in a variety of spin-contaminated radicals and biradicals. The virtues and shortcomings of both methods are considered in some detail. Since no modifications are made to the CBS-QB3 model for closed-shell species, there are only minor changes in the calculated atomization energies (reported as heats of formation of these species), resulting from the use of ROCBS-QB3 energies for the atoms.

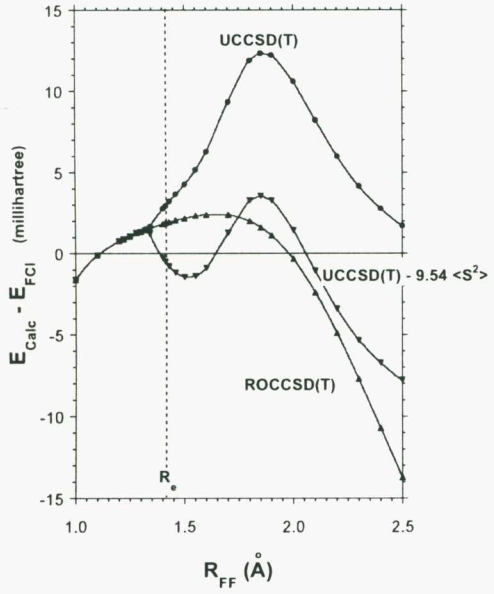


Figure 3.2: The errors in spin-restricted and spin-unrestricted CCSD(T)/cc-pVDZ energies for the F_2 $^1S_g^+$ ground state as a function of the bond length.

3.2 Theoretical Methods

The ROCBS-QB3 method retains most of the essential features of the standard CBS-QB3 method, including the use of UB3-LYP/6-311G(2d,d,p) geometries and

scaled frequencies, but has the following differences. In the first place, the extrapolation based on an unrestricted wave function is replaced with an extrapolation from a restricted wave function, i.e., RMP2/6-311+G(3d2f,2df,2p) energy and CBS extrapolation.

The pair-natural-orbital complete-basis-set (CBS) extrapolation of second-order aa- and ab-pair energies, and the application of the interference effect to extrapolate the contributions from higher orders of perturbation theory are carried over from the CBS-QB3 model without modification.[6] No attempt is made to extrapolate the contribution from single excitations to the second-order energy of open-shell species, since this is essentially a correction to the SCF energy, which is not extrapolated. The energy calculations based on unrestricted wave functions are also replaced with the corresponding calculations based on restricted wave functions, i.e., RMP4(SDQ)6-31+G(d(f),p) energy, and RCCSD(T)/6-31+G† energy. Because of the use of restricted wave functions, the empirical correction for spin contamination (equation 3.1) is always zero. Note that the new ROCBS-QB3 model gives total energies identical to those of the standard unrestricted CBS-QB3 model when applied to closed-shell species.

The GAUSSIAN suite of programs does not currently cater for the use of restricted wave functions for MP4 or CCSD(T) calculations on open-shell species, and the ACES II package was therefore used for such calculations. However, whereas GAUSSIAN supports the use of mixed spherical and cartesian basis sets, e.g., (6d,7f), ACES II does not. This creates a potential problem in comparisons of ROCBS-QB3 with standard CBS-QB3 since the latter uses (6d, 7f) combinations for the UMP4 and UCCSD(T) calculations (see section 2.8.2). In order to rectify this situation, UMP4 and UCCSD(T) calculations have been carried out using *both* (5d,7f) and (6d,7f) combinations. The difference is then added to the (5d,7f) RMP4 and RCCSD(T) components of ROCBS-QB3 as a basis set correction, yielding ROCBS-QB3 values that correspond approximately to the (6d,7f) treatment of standard CBS-QB3.

Table 3.1: Calculated and Experimental Bond Dissociation Energies (BDE's), Ionization Energies (IE's), and Electron Affinities (EA's) Involving Spin-Contaminated Species (kJ mol^{-1})

Species	Reaction	$\Delta<S^2>$ reactant	$\Delta<S^2>$ product	U-CBS-QB3	CBS-QB3	ROCBS-QB3	W1U	W1U + spin corr	W1(RO)[a]	Expt
σ-radicals										
BDE ΔH_{298}	H–C≡N → H + •C≡N	0.000	0.3822	540.9	531.3	529.8	536.4	530.1	531.8	528.4 ± 0.8[b]
	H–C≡CH → H + •C≡CH	0.000	0.3614	565.9	556.8	557.0	560.8	554.8	558.4	558.4 ± 0.1[c]
	H–CH=CH_2 → H + •CH=CH_2	0.000	0.1869	467.5	462.8	463.9	465.0	461.9	463.6	463.2 ± 2.5[b]
	H–C_6H_5 → H + •C_6H_5	0.000	0.5963	498.0	483.1	473.2				472.4 ± 2[b]
IE ΔE_0	C≡S → e^- + C≡S^+	0.000	0.7136	1106.2	1088.3	1092.7	1102.0	1090.2	1094.8	1093.2 ± 1[d]
	C≡O → e^- + C≡O^+	0.000	0.2081	1361.9	1356.7	1355.7	1356.0	1352.6	1353.0	1352.14±0.03[d]
EA ΔE_0	Cl^- → e^- + Cl	0.000	0.0104	355.5	355.2	353.6	350.2	350.0	350.0	348.6±0.0006[d]
	C≡N^- → e^- + C≡N	0.000	0.3822	386.4	376.8	375.3	380.1	373.8	375.6	372.6 ± 0.5[d]
π-radicals										
BDE ΔH_{298}	H–CH_2-CH=CH_2 → H + •CH_2-CH=CH_2	0.000	0.2014	369.9	364.9	364.8	369.4	366.1		371.5 ± 1.7[e]
	H–NH-CH=O → H + •NH-CH=O	0.000	0.1982	480.4	475.4	479.4			479.1	
	H–CH_2-C_6H_5 → H + •CH_2-C_6H_5	0.000	0.5851	393.8	379.1	375.7				375.7 ± 2.5[c]
Singlet biradicals										
BDE ΔH_{298}	F–F → F + F	0.3460	0.0076	151.0	159.5	157.0	152.0	157.6	156.0	159.0 ± 0.8[f]
	H–C≡C• → H + •C≡C•	0.3614	1.6142	520.6	489.2	470.8	511.2	490.5	479.9	488 ± 4[g]
	H–C_6H_4 → H + o-C_6H_4	0.5963	1.3056	364.1	346.3	341.6				326 ± 12[h]
	H–C_6H_4 → H + m-C_6H_4	0.5963	0.0000	378.9	393.8	403.7				393 ± 12[h]

Table 3.1 (continued): Calculated and Experimental Bond Dissociation Energies (BDE's), Ionization Energies (IE's), and Electron Affinities (EA's) Involving Spin-Contaminated Species (kJ mol^{-1})

Species	Reaction	$\Delta<S^2>$ reactant	$\Delta<S^2>$ product	U-CBS-QB3	CBS-QB3	ROCBS-QB3	W1U	W1U + spin corr	W1(RO)[a]	Expt
Singlet biradicals	$H–C_6H_4 \rightarrow H + p\text{-}C_6H_4$	0.5963	1.7508	479.8	450.9	446.9				456 ± 12[h]
	$•C{\equiv}C• \rightarrow C + C$	1.6142	0.02	548.4	588.3	603.2	569.1	595.4	602.7	596 ± 4[g]
EA ΔE_0	$:O–O–O• \rightarrow e^- + •O–O–O•$	0.0337	0.9434	247.3	224.5	214.6	235.6	220.6	209.2	202.9 ± 0.4[d]
Triplet biradicals										
EA ΔE_0	$H–C–C{\equiv}N^- \rightarrow e^- + H–C–C{\equiv}N$	0.1223	0.3653	198.3	192.3	192.4	195.3	191.3		193.3 ± 1.3[l]
RMS deviation[j]				21.0	7.0	5.8	14.2	5.3	4.1	± 1.9[k]
RMS deviation (excluding ozone)				18.7	5.0	5.2	11.3	2.5	3.8	
MAD[l]				16.1	4.8	3.6	9.8	3.2	3.2	± 1.5[k]
MAD (excluding ozone)				14.5	3.8	3.2	7.9	2.1	2.8	

[a] Standard restricted W1 method, refs 7 and 8. [b] Reference 19. [c] Reference 20. [d] Reference 21. [e] Reference 22. [f] Reference 23. [g] Reference 24. [h] Reference 25. [i] Reference 26. [j] RMS = root mean square deviation from experimental values. [k] Excludes those reactions involving C_6H_4 which have experimental uncertainties of ± 12 kJ mol^{-1}. [l] MAD = mean absolute deviation from experimental values

3.3 Results and Discussion

Firstly, the focus will be on how the performance of the new restricted-open-shell model differs from that of the standard unrestricted model for open shells.

3.3.1 Comparison of Restricted and Unrestricted Methods

The performance of the ROCBS-QB3 and standard CBS-QB3 model has been assessed for a variety of radicals and biradicals for which a direct comparisons with specific experiments (Table 3.1) can be made. The sample was selected to probe the effects of spin contamination – not to give a representative cross section of likely applications. The root mean square (RMS) and mean absolute deviations (MAD) from experiment for all methods are therefore larger than one might expect for typical applications. Included are results from some W1 calculations, both the standard version, W1,[7,8] which is denoted W1(RO) to emphasize the restricted treatment of open shells, and an unrestricted version, W1U[18] (Table 3.1). Firstly it is noted that the unrestricted version is in better agreement with both experiment and the standard restricted version if a spin correction is included, $-0.00628\ d\langle S^2\rangle$, analogous to equation 3.1. This correction approximates the difference between RCCSD(T) and UCCSD(T) energies (Figure 3.3). Table 3.1 compares the calculated numbers with the appropriate experimentally-determined results.[19–26]

3.3.1.1 σ-radicals

The radical reactions (Table 3.1) were selected to involve a single spin-contaminated open-shell product with a closed-shell reactant to avoid ambiguity in interpretation. Among the σ-radicals, the largest discrepancies between the standard unrestricted CBS-QB3 and the new ROCBS-QB3 model correspond to the phenyl radical (CBS-QB3 deviation from experiment 10.7 ± 2 kJ mol^{-1}) for which Equation 3.1 underestimates the correction, the chlorine atom (CBS-QB3 error 6.6 ± 0.0 kJ mol^{-1}) for which the spin

contamination is very small, and the CS^+ ion for which Equation 3.1 overestimates the correction (CBS-QB3 deviation from experiment –4.9 ± 1 kJ mol^{-1}). In each of these cases, ROCBS-QB3 gives energies closer to the experimental values. The result for the chlorine atom is indicative of the base error associated with extrapolations from modest basis sets even in the absence of spin contamination.

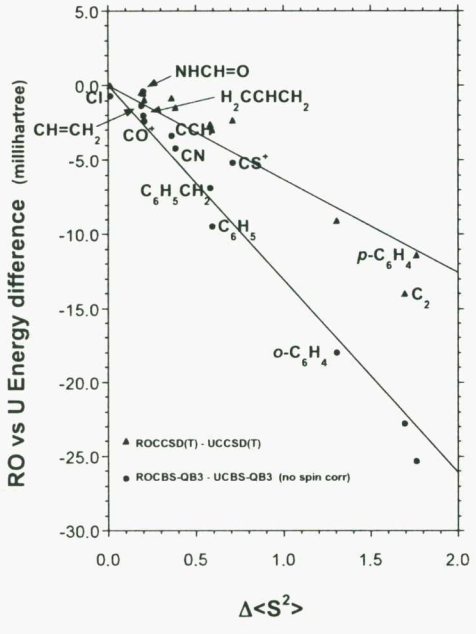


Figure 3.3: The difference between RCCSD(T) and UCCSD(T) energies and the difference between ROCBS-QB3 and unrestricted CBS-QB3 (before spin-contamination correction) energies are both proportional to the error in $\langle S^2 \rangle$, but the proportionality constant is greater for the CBS-QB3 energies.

The large CBS-QB3 error for the phenyl radical was unexpected, since the original CBS-Q model with UMP2 geometry gave 477.0 kJ mol^{-1} for the 298 K C–H bond dissociation enthalpy of benzene.[5] The good agreement of CBS-Q with experiment (472.4 ± 2 kJ mol^{-1}) must have been a bit fortuitous, since the improved B3-LYP geometry in CBS-QB3 gives 483.1 kJ mol^{-1}. The culprit here is probably the extreme sensitivity of the UHF spin contamination (and consequently the UCCSD(T) energy error) to small changes in geometry (Figure 3.1). The ROCBS-QB3 value

(473.2 kJ mol^{-1}) is certainly the beneficiary of error cancellation, but a reduced sensitivity to geometry is a potential advantage for the RO approach (Figure 3.2)

3.3.1.2 π-radicals

The limited number of π-radicals in Table 3.1 do not indicate any substantial difference between the standard unrestricted CBS-QB3 model and the new ROCBS-QB3 version. The spin correction overcorrects the formamide radical (by ~4 kJ mol^{-1} compared with the restricted version), but undercorrects the benzyl radical (by ~3.5 kJ mol^{-1}), so once again no evidence for a *systematic* error from equation 3.1 is seen. For the allyl radical, the restricted and unrestricted models have deviations from experiment of 6.6 and 6.7 kJ mol^{-1}, respectively. However, the consistency of the two models indicates that the deviation from experiment is a consequence of something other than spin contamination.

3.3.1.3 Biradicals

The quality of the results for the biradicals was a pleasant surprise, since these species should really be treated with a multi-configuration method. Although the specific strengths and weaknesses differ, the overall performance of the CBS-QB3 and ROCBS-QB3 models is comparable. This is consistent with the correlation between the spin-contamination correction in the standard CBS-QB3 model, equation 3.1, and the energy difference between RCCSD(T) and UCCSD(T) calculations, both before and after the CBS extrapolations (Figure 3.3).

The biradicals necessarily involve spin contamination in both reactant and product. The deviations from experiment for both restricted and unrestricted CBS-QB3 calculations are larger than the errors observed for the simple radicals considered previously, but smaller than might have been expected for such difficult species. It is obvious from Figure 3.2 that if the two radical centers are sufficiently separated, the ROCBS-QB3 model will fail badly, giving an energy that is too negative for the

biradical. This effect is evident in the ROCBS-QB3 underestimation of the C–H bond dissociation energy of H–C≡C• (by 17 ± 4 kJ mol^{-1}) and the overestimation of the bond dissociation energy of •C≡C• (by 7 ± 4 kJ mol^{-1}). Figure 3.2 also suggests that standard (unrestricted) CBS-QB3 could be unreliable for intermediate separations. This deficiency is illustrated by the large error (21.6 ± 0.4 kJ mol^{-1}) in the standard CBS-QB3 electron affinity of ozone. The similar error (17.7 ± 0.4 kJ mol^{-1}) for the spin-corrected W1Usc electron affinity of ozone indicates that this is an inherent deficiency in the UCCSD(T) description of ozone rather than an error in the CBS extrapolation. One might be tempted to treat ozone as a closed shell within the standard unrestricted CBS-QB3 model, thereby removing the largest error for this model chemistry. While this may be appropriate if the concern is with the molecule itself, treatment of O_3, F_2, or any other UHF-unstable species with restricted methods would create consistency problems in studies of *reactions* of these species with radicals (e.g., $F_2 + {}^{\bullet}CH_3 \rightarrow {}^{\bullet}F + CH_3F$) where the standard CBS-QB3 model would employ unrestricted methods along the reaction path. Therefore, the large errors in the energy of ozone for both the CBS-QB3 model (21.6 ± 0.4 kJ mol^{-1}) and the W1Usc model (17.7 ± 0.4 kJ mol^{-1}) must be accepted in assessing their performance.

The BDE of C_2 is a relatively clean example, since the product carbon atoms have little spin contamination. Neither standard unrestricted CBS-QB3 nor ROCBS-QB3 gives an especially good result. However, the correlation with the corresponding versions of W1 theory suggests that the problems are again inherent in using a single reference determinant. It is certainly clear from Figure 3.2 that CCSD(T)-based methods can be a poor choice for studies of radical-recombination reactions.

Both CBS models give results within the relatively large experimental uncertainties for the three isomers of benzyne. Note that m-benzyne closes to form a bicyclo[3,1,0]hexatriene[27] and thus is not open-shell. The difference between the CBS-QB3 and ROCBS-QB3 results for m-benzyne arises from the difference in the energy of

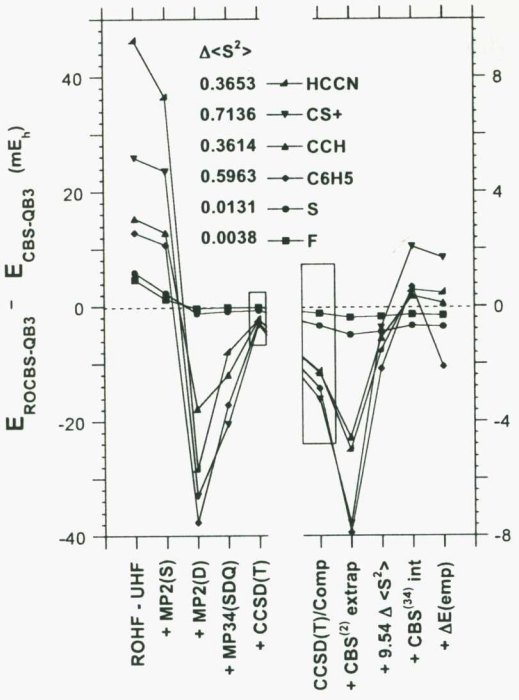


Figure 3.4: The cumulative difference between ROCBS-QB3 and standard unrestricted CBS-QB3 component by component. The composite CCSD(T) energy is repeated from the last point on the left-hand side of the figure as the first point on the higher resolution right-hand side of the figure, i.e., the points that are in the boxes.

the phenyl radical. Since it has been shown that the minimum energy RCCSD(T) geometry for p-benzyne is non-planar,[27] the good result for ROCBS-QB3 is perhaps fortuitous, just as for the allyl radical. On the other hand, these geometry errors may be of little consequence for thermochemistry, since the potential energy surface generally tends to be rather flat when restricted and unrestricted geometries differ. Finally, it is noted that the standard (unrestricted) version of CBS-QB3 gives a far better (probably fortuitously good) result than the new ROCBS-QB3 version for the second C–H bond dissociation energy for acetylene.

3.3.1.4 Order-by-order comparison

The surprisingly large difference (1.6 kJ mol^{-1}) between the restricted and unrestricted models for the chlorine atom (with $\Delta\langle S^2\rangle$ only 0.01) is a result of the mixing of contributions from different orders of perturbation theory, even though the CCSD(T) total energy is almost unchanged between restricted and unrestricted treatments ($\Delta E_{CCSD(T)}$ ~0.04 kJ mol^{-1}). The increased RO contribution from $E^{(2)}$ gives both a larger direct contribution with the large MP2 basis set (1.2 kJ mol^{-1}) and also a larger CBS extrapolation (0.4 kJ mol^{-1}).

The order-by-order differences between ROCBS-QB3 and standard unrestricted CBS-QB3 follow a consistent pattern (Figure 3.4). The energy lowering from RHF to UHF does not correlate very well with the error in $\langle S^2\rangle$, and the energy lowering from RMP2 single excitations is virtually independent of both (Figure 3.4). The second-order single excitation energy for triplet H–C–CN (9.8 mE_h) with two unpaired electrons is roughly twice as large as those for the doublet radicals (e.g. 3.7 mE_h for H–C–CN^-). The RMP2 – UMP2 double excitations follow the RHF – UHF differences, but overcorrect. The RMP2 doubles include simultaneous singles, which are clearly important in species with large RHF – UHF differences, as demonstrated by the failure of the second-order single excitations to adequately correct for the RHF – UHF difference (Figure 3.4).

Higher-orders compensate for the RMP2 overcorrection, making the composite CCSD(T) energies almost equal (Figure 3.4) and leaving only a small energy lowering of RCCSD(T) relative to UCCSD(T). In some cases this represents an inherent difference between restricted and unrestricted CCSD(T). For example, the RCCSD(T) and UCCSD(T) composite energies for phenyl radical differ by 2.84 mE_h while the CCSD(T)/6-31+$G^\dagger$ energies differ by 2.98 mE_h. In other cases, the basis-set dependence of the order-by-order contributions creates an artificial effect. For example, the cyanomethylene triplet has a composite energy difference of 2.29 mE_h, but a

difference of only 0.39 mE_h for the CCSD(T)/6-31+G† energies. In every case, the RCCSD(T) energy is lower than the UCCSD(T) energy. The CBS extrapolation of the second-order correlation energy favors the RMP2 version, since it is extrapolating from a larger calculated second-order energy. The spin-contamination correction, equation 3.1, compensates for both effects, making the differences between the restricted and unrestricted energies even smaller than the differences in the composite CCSD(T) energies (Figure 3.4).

Species with substantial UHF spin contamination may show excessive sensitivity of both restricted and unrestricted composite energies to small geometry changes, even when there is little difference between the RCCSD(T) and UCCSD(T) energies. The variation of the partition between RMP2(D) and RMP34(D) creates a potential instability for models such as CBS-QB3 that employ large basis sets for MP2, but small basis sets for MP34 and CCSD(T). Models such as CBS-APNO and W1 that employ large basis sets throughout should not show these problems, but may still fail for species that are poorly described by a single reference determinant (e.g., ozone in Table 3.1).

The CBS interference correction[28] to the higher-order contributions and the empirical correction, introduce more variation between the final restricted and unrestricted energies. However, it should be borne in mind that the differences between the R and U corrections are only a small fraction of the total interference and empirical corrections. The interference correction is necessary to handle near degeneracies[28] and the empirical correction is necessary to compensate for the very small basis sets used for the higher-order calculations. The large basis sets employed for CCSD and CCSD(T) calculations in the W1 model eliminate the need for these two corrections, but severely restrict the size of the species that can be studied.

3.3.1.5 Correlation between methods

The correlation between the results obtained with the six computational methods in Table 3.1 is explored in Table 3.2, where the RMS deviations of the models from experiment and from one another are examined. The results in Table 3.2 are overly pessimistic about the accuracy of W1 calculations due to a combination of the computational difficulty of the examples in Table 3.2 and the uncertainties in the experimental energies. The CBS-QB3 model is fairly close to the accuracy limit of CCSD(T) for these spin-contaminated species, so the W1 energies do not give as much improvement in accuracy as one would find in more typical applications of the model chemistries. However, the results are sufficiently reliable to reach some conclusions about the relationships between the various methods when applied to spin-contaminated species.

Table 3.2: The RMS Deviations (KJ Mol^{-1}) of the Computational Methods from Experiment and from one Another Based on the Reactions in Table 3.1.[a] The **Bold Faced Number** in each Column is the Minimum RMS Deviation, Indicating the Strongest Correlation Between Two Methods

	U-CBS-QB3	CBS-QB3	ROCBS-QB3	W1U	W1U + spin corr	W1(RO)[b]	Expt[c]
U-CBS-QB3	0.0	17.3	23.4	**8.0**	19.1	23.4	21.0
CBS-QB3	17.3	0.0	7.1	10.1	**3.1**	7.3	7.0
ROCBS-QB3	23.4	7.1	0.0	16.3	6.4	**3.5**	5.8
W1U	**8.0**	10.1	16.3	0.0	11.4	16.3	14.2
W1U + spin corr	19.1	**3.1**	6.4	11.4	0.0	5.6	5.3
W1(RO)[b]	23.4	7.3	**3.5**	16.3	5.6	0.0	4.1
Expt[c]	21.0	7.0	5.8	14.2	5.3	4.1	**1.9**[d]

[a] The deviations in the calculated energies for the benzynes relative to experiment were corrected for the 12 kJ mol^{-1} experimental uncertainty. [b] Standard restricted W1 method, ref. 7 and 8. [c] RMS error including ozone. [d] The experimental error estimates excluding the benzynes were used to calculate the RMS error (1.9 kJ mol^{-1}) for the experimental values.

1. The three types of calculation, unrestricted without spin correction, unrestricted with spin correction, and restricted open shell, correlate most strongly with other methods of the same "spin type". Specifically, the CBS-QB3 without spin correction correlates (U-CBS-QB3) best with the W1U without spin correction, the CBS-QB3 with spin correction correlates best with the W1Usc with spin correction, and the ROCBS-QB3 correlates best with the standard W1(RO) calculations.
2. The limited sample makes differences in RMS deviations (relative to experiment) of less than 25% statistically insignificant. Nevertheless, it can be concluded that W1Usc with spin correction ($-0.00628\Delta\langle S^2\rangle$) and standard W1(RO) are comparable in accuracy and are the most accurate methods of the group (hardly a surprise). The relative ranking of these two methods is entirely dependent on whether or not we include ozone in the test set (Table 3.1).
3. It can also be concluded that the new ROCBS-QB3 method is comparable in accuracy to (if not slightly more accurate than) standard (unrestricted) CBS-QB3 with spin correction. However, once again the relative ranking of the restricted *vs.* unrestricted methods is dependent on whether or not ozone is included in the test set (Table 3.1).
4. The unrestricted methods without spin correction are the least accurate. In fact, the W1U method without spin correction is less accurate than spin-corrected unrestricted CBS-QB3 and ROCBS-QB3. In addition, the unrestricted CBS-QB3 method without spin correction (U-CBS-QB3) shows the largest RMS

deviation from experiment, and therefore does not appear to represent a good general method, despite some earlier promising indications.[13-16]

5. One must exercise care in using a higher-level method to calibrate a computationally less demanding method. Comparison with W1(RO) would lead us to believe that the errors in the ROCBS-QB3 method are half as large as the errors in the spin-corrected unrestricted CBS-QB3 method. Similarly, comparison with spin-corrected W1Usc would lead us to believe that the errors in the spin-corrected unrestricted CBS-QB3 are half as large as the errors in the ROCBS-QB3 method. In fact, comparison with experiment shows the two methods to be of comparable accuracy. Methods of the same "spin type" have consistent (even if large) errors. Hence, U-CBS-QB3 correlates well with W1U without spin correction.

3.3.1.6 Summary

The new ROCBS-QB3 model has successfully eliminated one of the two empirical parameters in the earlier spin-corrected CBS-QB3 model and maintained comparable accuracy. There are not a sufficient number and variety of spin-contaminated examples to make a definitive judgment between the two methods, but it can be concluded that ROCBS-QB3 is at least a viable alternative if not slightly more accurate. Both standard (spin-corrected) CBS-QB3 and ROCBS-QB3 show good reliability for doublet radicals and triplet biradicals. Both have considerably larger errors for singlet biradicals and in general should only be applied to such species for qualitative purposes. The new restricted-open-shell model appears to be less sensitive to small changes in geometry, but cannot treat widely separated biradical centers and may have problems of incorrect localization in conjugated π-radicals (e.g. allyl radical).

3.3.2 Heats of Formation

Further assessment of the performance of the ROCBS-QB3 procedure has been carried out by using the 148 heats of formation at 298 K ($\Delta_f H_{298}$) (29 of which are for open-shell species) of the G2/97 test set of Curtiss et al.[29] The calculated molecular energies were used to derive heats of formation at 298 K ($\Delta_f H_{298}$) using the atomization method outlined by Nicolaides et al.,[30] (the calculated atomization energy being subtracted from the experimental heats of formation of the atoms, Table 3.3). The resulting "calculated" heats of formation for the molecules (Tables 3.4, 3.5, 3.6, and 3.7) are reported with uncertainties based on the total uncertainties in the experimental atomic $\Delta_f H_{298}$ values employed (Table 3.3).

The accuracy of *ab initio* calculations has improved to the point where it is now necessary to be more cautious in assessment of these methods. As a first step in this direction, both the uncertainty in the experimental heats of formation of the molecules in Tables 3.4, 3.5, 3.6, and 3.7, and also the uncertainty in the "calculated" heats of formation arising from the uncertainty in the experimental heats of formation of the constituent atoms (Table 3.3) are included. Both of these uncertainties are included in the uncertainties of the individual deviations from experiment:

$$\sigma_j(theory - experiment) = \left[\sigma_j^2(molecule) + \sum_A^{A\,in\,j} \sigma_A^2(atoms\ in\ j)\right]^{1/2} \qquad (3.2)$$

Note that these uncertainties exceed the actual deviations between theory and experiment for 15 of the 38 examples in the G2-1 test set, and 29 of the 81 examples in the G2-2 test set, reflecting the difficulty in finding sufficiently accurate thermodynamic data covering a broad range of chemistry. Both atomic and molecular uncertainties are also included in the uncertainties (σ_{MAD}) for the mean absolute deviations between theory and experiment in Tables 3.4, 3.5, 3.6, and 3.7:

$$\sigma_{MAD} = \frac{1}{n}\left[\sum_{j}^{n}\sigma_j^2(molecule)\right]^{1/2} + \left[\frac{1}{n}\sum_{j}^{n}\sum_{A}^{A\text{ in }j}\sigma_A^2(atoms\text{ in }j)\right]^{1/2} \quad (3.3)$$

Where *random errors in the molecular energies* and a "representative sample" have been assumed. The first term in equation 3.3 decreases as the size, *n*, of the test set increases, since the set of molecules is potentially unlimited and the errors are independent. However, the second term increases with increasing molecular size and is independent of the size of the test set, since the set of atoms is a fixed small number giving *systematic* errors, σ_A. It is the second term that limits the usefulness of $\Delta_f H_{298}$ data. For example, the first term contributes 0.2 kJ mol^{-1}, whereas the second term contributes 1.5 kJ mol^{-1}, to the uncertainty in the MAD of the calculated $\Delta_f H_{298}$ values for the closed-shell species in Table 3.4.

Supplementary error analyses are in Tables 3.4, 3.5 and 3.7, wherein heats of formation with total experimental uncertainties of 5 kJ mol^{-1} or greater are excluded from the statistics. Eliminating *those* large deviations between theory and experiment, *whose source lies in experimental values with large uncertainties*, allows for a more precise assessment of the performance of the theoretical methods. For instance, the largest deviations of CBS-QB3 and ROCBS-QB3 (Table 3.4) are −26 ± 5 kJ mol^{-1} and −20 ± 5 kJ mol^{-1}, respectively, if all species are included, compared with −23 ± 2 kJ mol^{-1} and +16 ± 2 kJ mol^{-1} when large experimental uncertainties are excluded. The latter eliminates the possibility that the larger error for the unrestricted method is a result of experimental error and thus is clearly a more unambiguous comparison.

3.3.2.1 Atoms

Table 3.3 presents the relative energies of atoms calculated with CBS-QB3 and ROCBS-QB3. The Table also includes $\langle S^2 \rangle$ values calculated with UHF/6-311+G(3d2f,2df,2p), and the experimental heats of formation of the atoms. Note that the recent pulsed field-ionization photoelectron-photoion coincidence measurement of

Table 3.3: Calculated Values of $<S^2>$, Total CBS-QB3 Energies (Hartree), Relative ROCBS-QB3 Energies (kJ mol^{-1}) and Experimental Heats of Formation (kJ mol^{-1}) for Atoms

atom	$<S^2>$[a]	CBS-QB3[b]	Δ(ROCBS-QB3)[c]	$\Delta_f H_{298}$ (Expt)[d]
H	0.750	–0.49982	0.0	217.998±0.006
Li	0.750	–7.43203	0.1	159.3± 1.0
Be	0.000	–14.62054	0.0	317.±3.[e]
B	0.761	–24.60182	–0.5	569.9± 0.8[f]
C	2.010	–37.78552	–0.5	716.7± 0.5
N	3.758	–54.52054	0.1	472.43±0.05[g]
O	2.009	–74.98799	–1.0	249.18±0.10
F	0.754	–99.64369	–0.8	79.4±0.3
Na	0.750	–161.84598	0.1	107.5±0.7
Al	0.770	–241.92916	–0.7	330.±4.
Si	2.015	–288.93198	–0.6	448.3±1.6[h]
P	3.751	–340.81718	0.4	316.5±1.0
S	2.013	–397.65825	–1.9	277.0±0.2
Cl	0.760	–459.68494	–1.6	121.30± 0.01

[a] $<S^2>$ values calculated with UHF/6-311+G(3d2f,2df,2p)//B3-LYP/6-311G(2d,d,p). [b] Standard unrestricted CBS-QB3 total energies. [c] Energies relative to the CBS-QB3 values (ROCBS-QB3 – CBS-QB3). [d] Reference 31, except as noted. [e] Reference 29. [f] Reference 32. [g] Reference 34. [h] Reference 33.

the heat of formation for a nitrogen atom (472.43 ± 0.05 kJ mol^{-1}) has been used.[34] Atomic energies calculated with ROCBS-QB3 are generally lower than the CBS-QB3 values, the exceptions being Li, N, Na and P, where the difference in any case is generally marginal. This implies that most of the atomization energies computed for closed-shell molecules with the ROCBS-QB3 procedure will be smaller than the corresponding standard unrestricted CBS-QB3 values.

As an example, the carbon atom is 0.5 kJ mol^{-1} lower in energy with ROCBS-QB3. As a consequence, the changes in the atomization energies computed by the ROCBS-QB3 procedure relative to the standard unrestricted CBS-QB3 procedure for closed-shell hydrocarbons will increase monotonically with the number of carbon atoms. This will be further examined in the following section.[35]

Table 3.4: Error Analysis (kJ mol^{-1}) for the Heats of Formation of the Subset of Closed-Shell Molecules in the G2/97 Test Set[a]

		deviation (theory – expt)[b]	
	Experiment[c]	CBS-QB3[d]	ROCBS-QB3[d,e]
MD		–0.4	+1.7
MAD	± 1.3	4.7 ± 1.7	4.7 ± 1.7
RMS	± 2.0	6.6 ± 1.7	6.3 ± 1.7
LD	± 8	–25.9 ± 5	–20.6 ± 5
Excluding species[f] with total experimental uncertainties ≥ 5 kJ mol^{-1}			
MD		0.1	+2.0
MAD	± 1.0	4.4 ± 1.5	4.4 ± 1.5
RMS	± 1.4	6.0 ± 1.5	5.8 ± 1.5
LD	± 4	–23.4 ± 2	+16.4 ± 2

[a] MD = mean deviation, MAD = mean absolute deviation, RMS = root mean square deviation, LD = largest deviation, from experimental values. [b] The uncertainty in the deviation from experiment reflects the uncertainty in both the experimental and the calculated values for the heat of formation of the molecule: $s_{dev} = [s_{exp}^2 + s_{calc}^2]^{1/2}$ [c] Reference 31, except where noted in Appendix 1 Table A1.1. [d] Conversion of CBS-QB3 energies and deviations (reference 40) from kcal mol^{-1} to kJ mol^{-1} introduces an uncertainty of ± 0.4 kJ mol^{-1}. [e] ROCBS-QB3 energies and deviations are obtained from the corresponding CBS-QB3 values (reference 40) by adding the Δ(ROCBS-QB3) values of the constituent atoms (see Table 3.3 footnote c). [f] SiH_2, LiF, ClF, AlF_3, $AlCl_3$, ClF_3, C_2F_4, CHF_3 and CH_3SiH_3.

3.3.2.2 Heats of formation for closed-shell species

The differences between ROCBS-QB3 and standard CBS-QB3 $\Delta_f H_{298}$ values for closed-shell molecules arise entirely from differences in the calculated atomic energies, since there is no change in the molecular calculations.

Table 3.4 gives a standard breakdown of the error analysis (Appendix 1 A1.1 gives the full list of molecules and their heats of formation) for this closed-shell subset in terms of mean deviations (MD's), mean absolute deviations (MAD's), root mean square deviations (RMS's) and largest deviations (LD's). The statistics for various sub-classes of molecules have also been calculated and are given in Table 3.5. The mean absolute deviations (MAD's) for the G2/97 subset of closed-shell molecules show that the performance of CBS-QB3 and ROCBS-QB3 are virtually identical (4.7 ± 1.7 kJ mol^{-1}). The uncertainties in the individual deviations (equation 3.2) and mean

Table 3.5: Error Analysis (kJ mol^{-1}) for the Heats of Formation of the Subset of Closed-Shell Molecules in the G2/97 Test Set, Broken Down into Various Subclasses of Molecules[a]

	CBS-QB3	ROCBS-QB3
Hydrocarbons		
MD	+6.5	+8.1
MAD	6.6 ± 2.1	8.1 ± 2.1
RMS	7.8 ± 2.1	9.2 ± 2.1
LD	14.6 ± 2	16.4 ± 2
Substituted Hydrocarbons		
MD	–0.7	+1.4
MAD	3.0 ± 1.6	3.1 ± 1.6
RMS	3.9 ± 1.6	3.8 ± 1.6
LD	10.2 ± 1.6	11.4 ± 1.6
Inorganic (all)		
MD	–3.1	–0.8
MAD	5.4 ± 1.8	4.6 ± 1.8
RMS	7.9 ± 1.8	6.6 ± 1.8
LD	–25.9 ± 5	–20.6 ± 5
Inorganic *(excluding species[b] with total experimental uncertainties ≥ 5 kJ mol^{-1})*		
MD	–2.9	–0.7
MAD	4.6 ± 1.4	3.8 ± 1.4
RMS	6.7 ± 1.4	5.3 ± 1.4
LD	–23.4 ± 2	–16.3 ± 2

[a] MD = mean deviation, MAD = mean absolute deviation, RMS = root mean square deviation, LD = largest deviation, from experimental values. [b] SiH_2, LiF, ClF, AlF_3, $AlCl_3$, ClF_3 and C_2F_4.

absolute deviations (equation 3.3) from experiment include contributions from both the uncertainties in the experimental molecular heats of formation and the uncertainties in the experimental atomic heats of formation used to determine the "calculated" heats of formation.

Interesting trends can be drawn from the mean deviations (MD's) of the sub-classes of molecules in the closed-shell set (Table 3.5). For example, the MD's for the hydrocarbons are the largest of all the sub-classes and they are positive, indicating that both methods overestimate the heats of formation.[36] The only difference between standard CBS-QB3 and ROCBS-QB3 in obtaining the heats of formation of

Table 3.6: Comparison of Calculated and Experimental Heats of Formation for Benzene, Naphthalene, and Anthracene (298 K, kJ mol^{-1}) Calculated with CBS-QB3 and ROCBS-QB3

		$\Delta H_f°$ (theory)[b]		Deviation (theory – expt)[c]	
Molecule	$\Delta H_f°$ (expt)[a]	CBS-QB3	ROCBS-QB3	CBS-QB3	ROCBS-QB3
C_6H_6 *(benzene)*	82.4 ± 0.5	90.7 ± 3	93.4 ± 3	8.3 ± 3	11.0 ± 3
$C_{10}H_8$ *(naphthalene)*	150.6 ± 1.1	159.5 ± 5	164.0 ± 5	8.9 ± 5	13.4 ± 5
$C_{14}H_{10}$ *(anthracene)*	230.8 ± 4.6	241.0 ± 7	247.3 ± 7	10.2 ± 8	16.5 ± 8

[a] Reference 31. [b] The uncertainty in the calculated value reflects the uncertainty in the experimental heats of formation of the constituent atoms (Table 3.3). [c] The uncertainty in the deviation from experiment reflects the uncertainty in both the experimental and the calculated values for the heat of formation of the molecule: $\sigma_{dev} = [\sigma_{exp}^2 + \sigma_{calc}^2]^{1/2}$

hydrocarbons arises from the relative energy of the carbon atom, as given in Table 3.3. Consequently, as the size of the hydrocarbon increases, the difference between the values given by CBS-QB3 and ROCBS-QB3 will also increase. This can be clearly seen by comparing the heats of formation for benzene, naphthalene and anthracene, which have 6, 10 and 14 carbons, respectively (Table 3.6).[37] Within the noise of the experimental data, the errors in the standard unrestricted CBS-QB3 model appear independent of molecular size, whereas the ROCBS-QB3 variation shows the linear increase in error with molecular weight expected for any size-consistent method.

The origins of the differences between the performance of CBS-QB3 and ROCBS-QB3 are apparent in the error analysis for the heats of formation of the hydrocarbons and inorganic molecules (Table 3.5). For the hydrocarbons, a larger MAD of 8.1 ± 2.1 kJ mol^{-1} is reported for ROCBS-QB3 (compared with 6.6 ± 2.1 kJ mol^{-1} for CBS-QB3), indicating that CBS-QB3 gives a better energy for the carbon atom. A smaller MAD of 4.6 ± 1.8 kJ mol^{-1} for the inorganic molecules for ROCBS-QB3 (compared with 5.4 ± 1.8 kJ mol^{-1} for CBS-QB3) and further inspection of the large deviations in the G2/97 test set confirm that ROCBS-QB3 gives a better energy for the chlorine atom. For instance, if the ROCBS-QB3 energy of Cl is used in calculating the

CBS-QB3 heats of formation of chlorine-containing molecules, the MAD for CBS-QB3 for the inorganic molecules decreases from 5.4 ± 1.8 kJ mol^{-1} to 5.0 ± 1.8 kJ mol^{-1}.

3.3.2.3 Heats of formation for open-shell species

Heats of formation at 298 K for the 29 radicals of the G2/97 test set, obtained with CBS-QB3 and ROCBS-QB3, are given in Table 3.7. These radicals were examined in a recent extensive assessment study[38] of a number of high-level theoretical techniques, including many of the G*n*, CBS and W*n* methods. Using the standard analysis, the best performing method in the previous work was G3X-RAD,[39] with a mean deviation (MD) of +0.1 kJ mol^{-1}, a mean absolute deviation (MAD) of 2.0 ± 1.4 kJ mol^{-1} and a largest deviation (LD) of –6.4 ± 0.2 kJ mol^{-1}. The uncertainties in the deviations from experiment include contributions from both the uncertainties in the experimental molecular heats of formation and the uncertainties in the experimental atomic heats of formation used to determine the "calculated" heats of formation [equation (3.2) and equation (3.3)]. The computationally more expensive W1 method has slightly larger MD, MAD and LD values of +0.5, 2.5 ± 1.4 and +6.7 ± 5.5 kJ mol^{-1}, respectively, but the difference in performance between W1 and G3X-RAD is small compared with the uncertainty in the experimental data.

The standard CBS-QB3 heats of formation for the 29 radicals in the G2/97 test set have been calculated previously[40] and are reproduced in Table 3.7. The MD, MAD, RMS and LD values are determined to be +0.5, 2.9 ± 1.4, 3.7 ± 1.4 and −7.9 ± 2.7 kJ mol^{-1}. The results obtained with ROCBS-QB3 are presented for the first time in Table 3.7. The method performs well overall, but slightly less well than the standard CBS-QB3 method, with MD, MAD, RMS and LD values of +0.5, 3.7 ± 1.4, 4.5 ± 1.4 and –9.6 ± 0.5 kJ mol^{-1}. However, the *uncertainty* in the deviation from experiment exceeds the actual deviation from experiment in 14 of the 29 examples, indicating the test set to be inappropriate for making fine distinctions.

Table 3.7: Comparison of Calculated and Experimental Heats of Formation (298 K, kJ mol^{-1}) for Open-Shell Species

			$\Delta H_f°$ (theory)[c]		deviation (theory – expt)[e]	
Molecule	$<S^2>$[a]	$\Delta H_f°$ (expt)[b]	CBS-QB3[d]	ROCBS-QB3	CBS-QB3[d]	ROCBS-QB3
BeH	0.752	341.8 ± 5.0	342.2 ± 3.0	342.4 ± 3.0	0.4 ± 5.8	0.5 ± 5.8
CH	0.759	596.2 ± 0.8	596.2 ± 0.5	596.0 ± 0.5	0.0 ± 0.9	–0.2 ± 0.9
CH_3	0.762	146.4 ± 0.3	148.9 ± 0.5	148.5 ± 0.5	2.5 ± 0.6	2.0 ± 0.6
NH_2	0.760	188.7 ± 0.3	188.3 ± 0.1	187.4 ± 0.1	–0.4 ± 0.3	–1.3 ± 0.3
OH	0.757	39.0 ± 0.4	37.7 ± 0.1	37.9 ± 0.1	–1.0 ± 0.4	–1.1 ± 0.4
SiH_3	0.755	200.4 ± 3.4	194.1 ± 1.6	195.2 ± 1.6	–6.3 ± 3.8	–5.2 ± 3.8
PH_2	0.770	138.5 ± 2.5	130.6 ± 1.0	128.9 ± 1.0	–7.9 ± 2.7	–9.6 ± 2.7
SH	0.764	143.1 ± 3.0	142.7 ± 0.2	142.9 ± 0.2	–0.4 ± 3.0	–0.2 ± 3.0
N_2^{+}[f]	0.768	1503.3 ± 0.1	1510.0 ± 0.1	1508.4 ± 0.1	6.7 ± 0.1	5.1 ± 0.1
NO	0.790	90.4 ± 0.2	87.9 ± 0.2	89.4 ± 0.2	–2.5 ± 0.3	–1.0 ± 0.3
NO_2	0.771	33.1 ± 0.8	27.2 ± 0.3	27.7 ± 0.3	–5.9 ± 0.8	–5.3 ± 0.8
O_2^{-}[f]	0.791	–43.2 ± 0.8	–45.1 ± 0.2	–46.2 ± 0.2	–1.9 ± 0.8	–3.0 ± 0.8
CN	1.132	438.9 ± 4.6	446.4 ± 0.6	445.1 ± 0.6	7.5 ± 4.6	6.2 ± 4.6
CO^{+}[f]	0.958	1241.4 ± 0.2	1245.6 ± 0.6	1246.0 ± 0.6	4.2 ± 0.6	4.6 ± 0.6
CS^{+}[f]	1.463	1368.2 ± 4.8	1368.2 ± 0.7	1374.8 ± 0.7	0.0 ± 4.9	6.6 ± 4.9
CCH	1.111	568.6[g] ± 0.8	573.2 ± 1.0	576.6 ± 1.0	4.6 ± 1.3	8.0 ± 1.3
HCO	0.766	41.8 ± 0.8	41.8 ± 0.6	42.5 ± 0.6	0.0 ± 1.0	0.7 ± 1.0
C_2H_3 ($^2A'$)	0.937	299.6 ± 3.3	300.9 ± 1.0	302.7 ± 1.0	1.3 ± 3.5	3.2 ± 3.5
C_2H_5 ($^2A'$)	0.763	120.9 ± 2.1	125.5 ± 1.0	125.7 ± 1.0	4.6 ± 2.3	4.7 ± 2.3
H_2COH (^2A)	0.762	–17.2 ± 1.3	–15.9 ± 0.6	–15.0 ± 0.6	1.3 ± 1.4	2.2 ± 1.4
CH_3O ($^2A'$)	0.760	17.2 ± 1.7	19.3 ± 0.6	19.6 ± 0.6	2.1 ± 1.8	2.5 ± 1.8
CH_3S ($^2A'$)	0.765	124.7 ± 2.1	122.6 ± 0.7	122.9 ± 0.7	–2.1 ± 2.2	–1.8 ± 2.2
CH_3CO ($^2A'$)	0.763	–10.0 ± 1.3	–8.7 ± 1.1	–7.3 ± 1.1	1.3 ± 1.7	2.8 ± 1.7
CH_2 ($^3B_1$)	2.017	392.0 ± 0.8	396.2 ± 0.5	395.7 ± 0.5	4.2 ± 0.9	3.7 ± 0.9
NH	2.017	356.5 ± 1.7	361.1 ± 0.1	359.2 ± 0.1	4.6 ± 1.7	2.8 ± 1.7
SiH_2 ($^3B_1$)	2.007	360.7 ± 4.9	357.4 ± 1.6	358.3 ± 1.6	–3.3 ± 5.2	–2.4 ± 5.2
O_2	2.047	0	–1.7 ± 0.2	–5.8 ± 0.2	–1.7 ± 0.2	–5.8 ± 0.2
S_2	2.061	128.4 ± 0.3	125.9 ± 0.4	118.9 ± 0.4	–2.5 ± 0.5	–9.6 ± 0.5
MD					+0.5	+0.5
MAD		± 1.8			2.9 ± 1.4	3.7 ± 1.4
RMS		± 2.5			3.7 ± 1.4	4.5 ± 1.4
LD		± 5.0			−7.9 ± 2.7	−9.6 ± 0.5
Excluding species[i] with total experimental uncertainties ≥ 5 kJ mol^{-1}						
MD					+0.2	–0.1
MAD		± 1.2			2.9 ± 1.0	3.6 ± 1.0
RMS		± 1.6			3.9 ± 1.0	4.6 ± 1.0
LD		± 3.4			−7.9 ± 2.7	−9.6 ± 0.5

[a] $<S^2>$ values calculated from MP2/6-311+G(3d2f,2df,2p)//B3-LYP/6-311G(2d,d,p). [b] Reference 31, except as noted. [c] Conversion of CBS-QB3 energies and deviations (reference 48) from kcal mol^{-1} to kJ mol^{-1} introduces an uncertainty of ± 0.4 kJ mol^{-1}. [d] The uncertainty in the calculated value reflects the uncertainty in the experimental heats of formation of the constituent atoms (Table 3.3). [e] The uncertainty in the deviation from experiment reflects the uncertainty in both the experimental and the calculated values for the heat of formation of the molecule: $\sigma_{dev} = [\sigma_{exp}^{2} + \sigma_{calc}^{2}]^{1/2}$ [f] Species is not part of the G2/97 test set. [g] References 20 and 45. [h] Reference [41]. [i] BeH, CN, CS, OOH, and SiH_2

3.3.3 Ionization Energies and Electron Affinities

Computing the atomization energy of a molecule is one of the more demanding tests of a theoretical method. However, assessments of high-level theoretical procedures using atomization energies can suffer from systematic errors that limit the usefulness of this approach. Ionization energies (IE's) and electron affinities (EA's) do not suffer from these systematic errors (i.e. the second term in equation 3.3) does not enter into the uncertainties in the calculated IE's and EA's), and the uncertainties in the experimental numbers are generally small, providing for a more precise set of comparisons. For example, in contrast to the large uncertainties in the experimental heats of formation, the experimental uncertainties exceed the deviation between theory and experiment for only 12 of the 88 ionization energies and 16 of the 58 electron affinities. Furthermore, this set of energy evaluations involves open-shell species in each case, which facilitates the assessment of the differences between the unrestricted and restricted versions of CBS-QB3.

The CBS-QB3 and ROCBS-QB3 ionization energies and electron affinities at 0 K of the G2/97 test set can be found in the Appendix 1 Tables A1.2 and A1.3, respectively. The ionization energies have been calculated by taking the difference in total energies at 0 K between the cation and the neutral, while the electron affinities have been obtained from the difference in total energies at 0 K between the neutral and the anion. The experimental ionization energies for some of these molecules have been recently updated in the NIST webbook,[42] and this leads to slightly better agreement between theory and experiment. The largest change is for BF_3, where the ROCBS-QB3 deviation using the old experimental IE is 18.8 kJ mol^{-1}, while the new experimental value of 1515 ± 30 kJ mol^{-1} results in a reduced deviation of 4 ± 30 kJ mol^{-1} (see Appendix 1 Table A1.2). However, the very large uncertainty assigned to the new experimental value raises doubts about the significance of the improved agreement. The updated IE for furan (C_4H_4O, 857 ± 1 kJ mol^{-1}) also leads to better agreement between

theory and experiment. The IE most recently recommended by NIST for NH_3 (971.6 ± 2 kJ mol^{-1})[42] has been superseded by the recent high resolution ZEKE measurement (982.8371 ± 0.0002 kJ mol^{-1}).[43] The new experimental values decrease the MAD between theory and experiment by less than 0.1 kJ mol^{-1} (see Appendix 1 Table A1.2).

The B_2F_4 molecule shows the largest deviation (–27.8 kJ mol^{-1}, see Appendix 1 Table A1.2) between ROCBS-QB3 (adiabatic) and experimental ionization energies. Similar large deviations have been found in other high-level theoretical studies,[2,5-9] suggesting that perhaps it is the experimental value that is at fault. Indeed, it has been proposed that the experimental number obtained for B_2F_4 corresponds to a vibrationally excited state,[5] since ionization induces a 90° rotation about the B-B bond. This uncertainty has led some researchers to omit B_2F_4 from their statistical comparisons and this is done likewise here.

It has been suggested[8] that the experimental IE value[44] of 1354 ± 2 kJ mol^{-1} for CN may correspond to the low-lying $^3\Pi$ state of CN^+. Taking this into consideration, IE's corresponding to both the $^1\Sigma$ and $^3\Pi$ cations for CN have been calculated. The deviations of theory from experiment of CBS-QB3 and ROCBS-QB3 are 4.4 ± 2 and 5.1 ± 2 kJ mol^{-1}, respectively, for the $^3\Pi$ cation, compared with −25.6 ± 2 and −24.1 ± 2 kJ mol^{-1} (see Appendix 1 Table A1.2), if the experiment is assumed to produce the $^1\Sigma$ cation. IE's generated from the $^3\Pi$ cation in calculating the deviation of theory from experiment for CN have been employed for the error analysis of Table 3.8.

ROCBS-QB3 performs slightly better than standard CBS-QB3 for the calculation of both ionization energies and electron affinities, with MAD's for the entire set of 4.1 ± 0.1 kJ mol^{-1} (compared with 4.4 ± 0.1) and 3.9 ± 0.2 kJ mol^{-1} (compared with 4.3 ± 0.2), respectively. The RMS errors also favor the spin-restricted method, but the largest deviation from experiment for the ROCBS-QB3 ionization energies (−13.5 ± 0.1 kJ mol^{-1}) exceeds the largest deviation for the CBS-QB3 ionization energies (12.6 ± 1.0 kJ mol^{-1}).

3.3.4 Error Analysis

A more comprehensive error analysis of results for the complete G2/97 test sets, including comparison with other popular model chemistries, is given in Table 3.8. The subcategories were selected to facilitate comparison with relevant data in the literature. It has become customary to evaluate models by comparison of the mean absolute deviation (MAD) (as a measure of the reliability) and the largest deviation (LD). Also included is the RMS deviation between theory and experiment as an additional measure of reliability, as well as the uncertainties in these quantities. On the basis of largest deviations alone, one might erroneously conclude from the G2-1 $\Delta_f H_{298}$ data that CBS-QB3 is competitive with W1 in accuracy. In particular, the largest deviation for W1 theory, 9.2 kJ mol^{-1} for $\Delta_f H_{298}$ (Si_2H_6), is greater than the largest deviation for CBS-QB3, −8.8 kJ mol^{-1} for $\Delta_f H_{298}$ (SiH_4). However, the uncertainty in the experimental heat of formation of Si_2H_6, 1.3 kJ mol^{-1}, and the uncertainty in the "calculated" value, 3.2 kJ mol^{-1}, give an uncertainty of ± 3.5 kJ mol^{-1} in the largest W1 deviation. The RMS deviation gives a measure of reliability that is not prone to such misinformation from one bad data point, and at least correctly indicates that W1 is more reliable than CBS-QB3 even if it underestimates the difference. Perhaps a more convincing example is the comparison of CBS-QB3 with G3 theory using the IE and EA data in the G2-1 test set. The largest deviations suggest that CBS-QB3 is much more reliable than G3, but the RMS deviations indicate that the two methods are more comparable in reliability. Nevertheless, the largest deviations from experiment for CBS-QB3 are generally smaller than those for G3 for the various subsets in Table 3.8. The uncertainty in the deviations for the G2-1 $\Delta_f H_{298}$ test set, ± 1.4 kJ mol^{-1}, indicates that the measured accuracy of W1 theory, 2.5 kJ mol^{-1}, is limited by both the accuracy of the calculations and the accuracy of the experimental data.

Table 3.8: Error Analysis (kJ mol^{-1}) for the Heats of Formation (298 K), Ionization Energies, and Electron Affinities of Molecules in the G2-1 and G2/97 Test Sets for Several Popular Model Chemistries

		W1[a]	G3X-RAD[b]	CBS-QB3[c]	ROCBS-QB3	G3[c]	CBS-4M[c]
G2-1 Test Set							
$\Delta_f H_{298}$	MAD	2.5±1.4	2.9±1.4	2.5±1.4	2.6±1.4	3.9±1.4	8.5±1.4
	RMS	3.2±1.4[a]	4.2±1.4	3.4±1.4	3.5±1.4	5.2±1.4	11.5±1.4[d]
	LD	9.2±3.5	−15.3±1.6	−8.8±2.2	−8.1±2.2	16.7±1.6	40.6±2
IE[d]	MAD	1.4±0.1[a,b]	4.1±0.1	4.4 ± 0.1	4.2±0.1	4.0±0.1	9.0±0.1
	RMS	2.2±0.1[a,b]	5.6±0.1	5.1±0.1	5.1±0.1	5.5±0.1	10.6±0.1
	LD	5.5±0.2[a]	15.5±0.1	10.0±1.0	11.9±0.02	−16.7±0.02	25.5±0.3
EA[e]	MAD	1.6±0.3[a]	4.2±0.3	5.1±0.3	4.6±0.3	4.7±0.3	10.6±0.3
	RMS	2.0±0.3[a]	5.7±0.3	5.9±0.3	5.7±0.3	6.3±0.3	13.1±0.3
	LD	4.9±2.8	−14.3±0.4	−10.5±0.4	9.9±0.4	17.6±0.4	−30.1±1
G2/97 Test Set							
$\Delta_f H_{298}$ (all)	MAD		3.1±1.5	4.3±1.5	4.3±1.5	3.9±1.5[b]	11.9±1.5
	RMS		4.5±1.5	6.1±1.5	6.0±1.5	5.1±1.5[b]	15.8±1.5
	LD		−15.3±1.4	−23.4±2	16.4 ±2	20.1±2[b]	52.7±2
$\Delta_f H_{298}$ (rad)	MAD	2.5±1.4[c]	2.0±1.4[f]	2.9±1.4[f]	3.7±1.4[g]	2.8±1.4[f]	
	RMS	3.1±1.4[c]	3.2±1.4[f]	3.7±1.4[f]	4.5±1.4[g]	3.6±1.4[f]	
	LD	6.7±5.5[c]	−6.4±0.2[f]	−7.9±2.7	−9.6±0.5[g]	8.5±1.0[f]	
IE[d]	MAD	1.9±0.1[a]	3.6±0.1	4.4±0.1	4.1±0.1	4.1±0.1[b]	15.8±1.5
	RMS	2.9 ±0.1[a]	6.8±0.1	5.3±0.1	5.2±0.1	5.5±0.1[b]	52.7±2
	LD	12.5±2	15.5±0.1	12.6±1	13.5±0.1	20.9±1[b]	174.1±1
EA[e]	MAD	1.7±0.2[a]	4.1±0.2	4.3±0.2	3.9±0.2	4.1±0.2[b]	12.8±0.2
	RMS	2.3±0.2[a]	5.9±0.2	5.5±0.2	5.0±0.2	5.7±0.2[b]	21.1±0.2
	LD	−6.4 ± 0.4[a]	16.3±0.0	11.3±0.4	9.9±0.4	17.6±0.4[b]	−121.8±0.6

[a] Reference 8. [b] From reference 38 unless otherwise noted. [c] From reference 40 unless otherwise noted. [d] Excludes BF_3, CN and B_2F_4. [e] Excludes Cl_2. [f] From reference 39, which includes some species not part of the G2/97 test set (see Table 3.7 footnote f). [g] Includes all species from Reference 38.

This is verified by the reduced W1 deviations for the G2-1 IE's (1.4 kJ mol^{-1}) and EA's (1.6 kJ mol^{-1}), for which the experimental data are more reliable (IE's ± 0.1 and EA's ± 0.3 kJ mol^{-1}). In contrast to W1 theory, the CBS-QB3 deviations are larger for the IE's and EA's than for the heats of formation. This may be the result of omitting the core effects that were part of the earlier CBS-APNO model.[4]

The problem of unreliable heats of formation is even worse for the expanded G2/97 test set (Table 3.8). None of the methods included in Table 3.8 has an MAD for the subset of neutral radicals (Table 3.7) that would be unaffected by the uncertainty in the experimental values (± 1.5 kJ mol^{-1}). In a pattern reminiscent of the comparison between CBS-QB3 and W1 for the G2-1 $\Delta_f H_{298}$ subset, CBS-QB3 and G3 appear competitive with W1 for the G2/97 $\Delta_f H_{298}$(rad) subset, and G3X-RAD appears to be slightly more accurate than W1 theory. However, examining the IE's and EA's, which necessarily also involve radicals and for which the experimental data are both more accurate and devoid of systematic errors, then W1 takes its rightful place as the most accurate (albeit by far the most expensive) of the models in Table 3.8. G3X-RAD is more comparable in accuracy to CBS-QB3 and G3, than to W1. G3X-RAD, a modification of G3X designed particularly for the treatment of open-shell systems, benefits in the case of spin-contaminated species from being based on restricted procedures, since G3 lacks a spin correction. Among the methods of comparable speed and MAD (i.e. CBS-QB3, ROCBS-QB3, G3, and G3X-RAD), ROCBS-QB3 has the smallest RMS errors for IE's and EA's, indicating the greatest reliability.

It has been observed[5] that the error distributions for the commonly used model chemistries are almost indistinguishable from Gaussian distributions, whose RMS error is $(\pi/2)^{1/2}$ times the MAD. Thus, the standard error function can be used to obtain confidence levels for each model. For example, the RMS error for the closed-shell heats of formation for ROCBS-QB3, 6.3 ± 1.7 kJ mol^{-1}, is slightly larger than $(\pi/2)^{1/2}$ times the MAD ($4.7(\pi/2)^{1/2}$ = 5.9 ± 1.7 kJ mol^{-1}), indicating that the "tails" of this error

distribution are slightly larger than those of a Gaussian curve. This implies that 10% of our heats of formation calculations will have errors exceeding twice the MAD (i.e. deviations greater than 9.4 kJ mol^{-1}). Furthermore, 99% confidence levels for the heats of formation, IE's and EA's for ROCBS-QB3 correspond to ±15.2, ±13.2 and ±13.9 kJ mol^{-1}, respectively (3.228 times the MAD for a Gaussian distribution). This indicates with 99% certainty that the error for a particular ROCBS-QB3 $\Delta_f H_{298}$, IE or EA calculation will be within these values. Conversely, there is a 1% probability that the error for the calculation will be *greater* than these values.

Within the accuracy of the G2/97 test data, the new ROCBS-QB3 version of CBS-QB3 appears to be competitive with standard unrestricted CBS-QB3 and G3 theory in accuracy. It is a little less accurate than G3X-RAD for neutral radicals and closed-shell molecules, but several times faster than G3 or G3X-RAD.

The G2 test sets are more than adequate to calibrate the accuracy of models of the order of CBS-4M and are for the most part sufficient to calibrate the accuracy of models of the order of CBS-QB3 or G3. However, the heats of formation in particular are clearly inadequate to calibrate a more refined model such as W1 theory. Using atomization energies rather than heats of formation would eliminate the uncertainty in the calculated quantities, but would not improve matters. Atomization energies would simply hide the problem by assigning the systematic atomic errors to the experimental molecular atomization energies. The fundamental problem is our uncertainty about the energies of single atoms for elements that are solids at 298 K.

Individual bond dissociation energies do not have the systematic errors encountered for atomization energies or heats of formation. It would therefore be desirable to develop a set of reliable test data of this type to calibrate computational models. In addition, the BDE's provide a direct comparison with experiment that better matches chemical problems. For example, it has been noted[39] that the CBS-APNO $\Delta_f H_{298}$ for the •CCH radical deviates from experiment by 6.8 ± 1.3 kJ mol^{-1} [575.4 kJ

mol^{-1} calculated (with an uncertainty of ± 1.0 kJ mol^{-1} reflecting the uncertainty in the heat of formation of carbon atoms) *vs.* 568.6 ± 0.8 kJ mol^{-1} experimental[20,45]]. However, there is a compensating deviation in the CBS-APNO $\Delta_f H_{298}$ for acetylene (232.9 ± 1.0 kJ mol^{-1} calculated *vs.* 228.2 ± 0.7 kJ mol^{-1} experimental[45]) giving an HCC–H bond dissociation enthalpy just 2.1 kJ mol^{-1} from the experimental value (560.5 kJ mol^{-1} calculated *vs.* 558.4 ± 0.1 kJ mol^{-1} experimental[20]). In fact, it was CBS-APNO calculations that were used to properly assign the experimental spectra.[46] Absolute heats of formation are uninformative about these important chemical energy *differences*.

As the standard test set has acquired increasing numbers of hydrocarbons of increasing molecular weight, the uncertainty in the heat of formation of a gas-phase carbon atom plays a disproportionate role in the evaluation of computational methods.[35] A method with several empirical parameters can adjust to fit the *error* in the heat of formation of a carbon atom and thus *appear* to be more accurate than a purely *ab initio* method. A glance at Table 3.6 would suggest that standard CBS-QB3 is more accurate than ROCBS-QB3 for closed-shell molecules, whereas the two methods are in fact identical for closed-shell molecules. The apparent difference arises because of the involvement of open-shell atoms in the calculation of the heats of formation. Clearly, this type of artificial effect needs to be given greater consideration in the testing of computational models.

3.4 Conclusions

In this Chapter, a new member of the CBS family of methods, ROCBS-QB3 has been formulated. The new method extrapolates correlation energies to the infinite-basis-set limit based on a *restricted* wave function. This results in the empirical correction for spin contamination in the standard CBS-QB3 method being zero for both closed- and open-shell systems. In addition, the correction for higher orders of

correlation based on UMP4 and UCCSD(T) calculations is replaced with one based on the equivalent restricted calculations.

The new ROCBS-QB3 model has successfully eliminated one of the two empirical parameters in the earlier unrestricted CBS-QB3 model and maintained comparable accuracy. There are not a sufficient number and variety of spin-contaminated examples with well-established thermochemistry to make a definitive judgment between the two methods, but it can be concluded that the new ROCBS-QB3 model is at least a viable alternative to, if not slightly more accurate than, standard CBS-QB3. Both standard CBS-QB3 and ROCBS-QB3 show good reliability for doublet radicals and triplet biradicals. Both have considerably larger errors for singlet biradicals and in general should only be applied to such species for qualitative purposes. The new restricted-open-shell model appears to be less sensitive to small changes in geometry, but cannot treat widely separated biradical centers and may have problems of incorrect localization in conjugated π-radicals.

Evaluation of the heats of formation, ionization energies and electron affinities for molecules in the G2/97 test set shows that the ROCBS-QB3 method provides good thermochemistry, with a mean absolute deviation (MAD) from experiment of 4.3 ± 0.9 kJ mol^{-1}, which is comparable to that for the standard formulation of CBS-QB3 (4.4 ± 0.9 kJ mol^{-1}). Based on the apparent reduced sensitivity to small changes in geometry, it can be tentatively recommended to use the ROCBS-QB3 method for reactions involving spin-contaminated products and transition structures.

3.5 References

(1) For recent reviews on high-level theoretical thermochemistry, see for example: (a) Helgaker, T.; Klopper, W.; Halkier, A.; Bak, K. L.; Jørgensen, P.; Olsen, J. in *Quantum-Mechanical Prediction of Thermochemical Data*, edited by Cioslowski, J. (Kluwer Academic: Dordrecht, The Netherlands, 2001), pp. 1–30. (b) Martin J. M. L.; Parthiban, S. in *Quantum-Mechanical Prediction of Thermochemical Data*, edited by Cioslowski, J. (Kluwer Academic: Dordrecht, The Netherlands, 2001), pp. 31–65. (c) Ragavachari, K.; Curtiss, L. A. in *Quantum-Mechanical Prediction of Thermochemical Data*, edited by Cioslowski, J. (Kluwer Academic: Dordrecht, The

Netherlands, 2001), pp. 67–98 (d) Petersson, G. A. in *Quantum-Mechanical Prediction of Thermochemical Data*, edited by Cioslowski, J. (Kluwer Academic: Dordrecht, The Netherlands, 2001), pp. 99–130. (e) Henry, D. J.; Radom, L. in *Quantum-Mechanical Prediction of Thermochemical Data*, edited by Cioslowski, J. (Kluwer Academic: Dordrecht, The Netherlands, 2001), pp. 161–197.

(2) (a) Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K. ; Curtiss, L. A. *J. Chem. Phys.* **1989**, *90,* 5622. (b) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, *94*, 7221. (c) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764.

(3) Petersson, G. A.; Bennett, A.; Tensfeldt, T. G.; Al-Laham, M. A.; Shirley, W. A. *J. Chem. Phys.* **1988**, *89,* 2193.

(4) Montgomery, Jr., J. A.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1994**, *101*, 5900.

(5) Ochterski, J. W.; Petersson, G. A.; Montgomery, Jr., J. A. *J. Chem. Phys.* **1996**, *104*, 2598.

(6) (a) Montgomery, Jr., J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1999**, *110*, 2822. (b) Montgomery, Jr., J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **2000**, *112*, 6532.

(7) Martin, J. M. L.; de Oliveira, G. *J. Chem. Phys.* **1999**, *111*, 1843.

(8) Parthiban S.; Martin, J. M. L. *J. Chem. Phys.* **2001**, *114*, 6014.

(9) Boese, A. D.; Oren, M.; Atasoylu, O.; Martin, J. M. L.; Kállay, M.; Gauss, J. *J. Chem. Phys.* **2004**, *120*, 4129.

(10) (a) Mebel, A. M.; Morokuma, K.; Lin, M. C. *J. Chem. Phys.* **1994**, *101*, 3916. (b) Mebel, A. M.; Morokuma, K.; Lin, M. C. *J. Chem. Phys.* **1995**, *103*, 7414.

(11) (a) Mayer, P. M.; Parkinson, C. J.; Smith, D. M.; Radom, L. *J. Chem. Phys.* **1998**, *108*, 604. (b) Smith, D. M.; Golding, B. T.; Radom, L. *J. Am. Chem. Soc.* **1999**, *121*, 5700. (c) Parkinson, C. J.; Mayer, P. M.; Radom, L. *Theor. Chem. Acc.* **1999**, *102*, 92. (d) Henry, D. J.; Parkinson, C. J.; Mayer, P. M.; Radom, L. *J. Phys. Chem. A* **2001**, *105*, 6750. (e) Henry, D. J.; Parkinson, C. J.; Radom, L. *J. Phys. Chem. A* **2002**, *106*, 7927. (f) Henry, D. J.; Sullivan, M. B.; Radom, L. *J. Chem. Phys.* **2003**, *118*, 4849.

(12) See for example: Bauschlicher, Jr., C. W.; Partridge, H. *J. Chem. Phys.* **1995**, *103*, 1788.

(13) Wood, G. P. F.; Henry, D. J.; Radom, L. *J. Phys. Chem. A* **2003**, *107*, 7985.

(14) Gómez-Balderas, R.; Coote, M. L.; Henry, D. J.; Radom, L. *J. Phys. Chem. A* **2004**, *108*, 2874.

(15) Coote, M. L. *J. Phys. Chem. A* **2004**, *108*, 3865.

(16) Wood, G. P. F.; Rauk, A.; Radom, L. *J. Chem. Theory and Comput.* **2005**, *1*, 889.

(17) See for example: Chen, W.; Schlegel, H. B. *J. Chem. Phys.* **1994**, *101*, 5957.

(18) Montgomery, Jr., J.; Frisch, M. J.; Petersson, G. A.; Martin, J. *unpublished.*

(19) Ervinand, K. M.; DeTuri, V. F. *J. Phys. Chem. A* **2002**, *106*, 9947.

(20) Mordaunt, D. H.; Ashfold, M. N. R. *J. Chem. Phys.* **1994**, *101*, 2630.

(21) Afeefy, H. Y.; Liebman, J. F.; Stein, S. E. "Neutral Thermochemical Data" in *NIST Chemistry WebBook*, NIST Standard Reference Database Number 69, edited by Mallard, W. G.; Linstrom, P. J. November 1998, National Institute of Standards and Technology, Gaithersburg, MD, 20899 (http://webbook.nist.gov/chemistry/).

(22) Ellison, G. B.; Davico, G. E.; Bierbaum, V. M.; DePuy, C. H. *Int. J. Mass Spectrom. Ion Processes* **1996**, *156*, 109.

(23) Cox, J. D.; Wagman, D. D.; Medvedev, V. A. *CODATA Key Values for Thermodynamics* (Hemisphere, New York, 1989).

(24) Blanksby, S. J.; Ellison, G. B. *Acc. Chem. Res.* **2003**, *36*, 255.

(25) Wenthold, P. G.; Squires, R. R. *J. Am. Chem Soc.* **1994**, *116*, 6401.

(26) Nimlos, M. R.; Davico, G.; Geise, C. M.; Wenthold, P. G.; Lineberger, W. C.; Blanksby, S. R.; Hadad, C. M.; Petersson, G. A.; Ellison, G. B. *J. Chem. Phys.* **2002**, *117*, 4323.

(27) (a) Crawford, T. D.; Kraka, E.; Stanton, J. F.; Cremer, D. *J. Chem. Phys.* **2001**, *114*, 10638. (b) Kraka, E.; Anglada, J.; Hjerpe, A.; Filatov, M. ; Cremer, D. *Chem. Phys. Lett.* **2001**, *348*, 115.

(28) Petersson, G. A.; Nyden, M. R. *J. Chem. Phys.* **1981**, *75*, 3423.

(29) (a) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **1997**, *106*, 1063. (b) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 42.

(30) Nicolaides, A.; Rauk, A.; Glukhovtsev, M. N.; Radom, L. *J. Phys. Chem.* **1996**, *100*, 17460.

(31) Lias, S. G.; Bartmess, J. E.; Liebman, J. F.; Holmes, J. L.; Levin, R. D.; Mallard, W. G. *J. Phys. Chem. Ref. Data* **1988**, *17*, Suppl 1.

(32) Ochterski, J. W.; Petersson, G. A.; Wiberg, K. B. *J. Am. Chem. Soc.* **1995**, *117*, 11299.

(33) Storms, E.; Mueller, B. *J. Phys Chem.* **1977**, *81*, 318.

(34) Tang, X.; Hou, Y.; Ng, C. Y.; Ruscic, B. *J. Chem. Phys.* **2005**, *123*, 74330.

(35) We note that it has recently been proposed that the standard heat of formation of $C_{(g)}$ is too low by 0.56 kJ mol^{-1}.; Tasi, G.; Izsák, R.; Matisz, G.; Császár, A. G.; Kállay, M.; Ruscic, B.; Stanton, J. F. *to be published.*; Stanton, J. F.; Kállay, M.; Császár, A. G.; Ruscic, B. *to be published.*

(36) We note that the use of a newly-proposed[35] $\Delta_f H_0$ for $C_{(g)}$ of 711.75 kJ mol^{-1} leads to a revised $\Delta_f H_{298}$ for $C_{(g)}$ of 718.285 kJ mol^{-1} and a slightly worse ROCBS-QB3 MAD (10.9 kJ mol^{-1}) for the hydrocarbons of Table V.

(37) The deviations in $\Delta_f H_{298}$ for both CBS-QB3 and ROCBS-QB3 increase by 3.4 (benzene), 5.6 (naphthalene) and 7.8 (anthracene) kJ mol^{-1} if the new $\Delta_f H_{298}$ for $C_{(g)}$ is used.

(38) (a) Henry, D. J.; Sullivan, M. B.; Radom, L. *J. Chem. Phys.* **2003**, *118*, 4849. (b) Henry, D. J. *personal communication.*

(39) (a) Henry, D. J.; Parkinson, C. J.; Radom, L. *J. Phys. Chem. A* **2002**, *106*, 7927. (b) Observe a deviation from experiment of 10.3 kJ mol^{-1}. However, this is reduced to 6.8 ± 1.3 kJ mol^{-1} if we compare with the more recent experimental value (reference 24 and reference 47).

(40) Montgomery, Jr., J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **2000**, *112*, 6532.

(41) "Experimental Thermochemical Data," NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101, Release 12, August 2005, edited by R.D. Johnson III, National Institute of Standards and Technology, Gaithersburg, MD, 20899 (http://srdata.nist.gov/cccbdb).

(42) Lias, S.G. "Ionization Energy Evaluation" in NIST Chemistry WebBook, NIST Standard Reference Database Number 69, Eds. P.J. Linstrom and W.G. Mallard, June 2005, National Institute of Standards and Technology, Gaithersburg MD, 20899 (http://webbook.nist.gov).

(43) Seiler, R.; Hollenstein, U.; Softley, T. P.; Merkt, F. *J. Chem. Phys.* **2003**, *118*, 10024.

(44) Berkowitz, J.; Chupka, W. A.; Walter, T. A. *J. Chem. Phys.* **1969**, *50*, 1497.

(45) Pedley, J. B. *Thermochemical Data and Structures of Organic Compounds, Volume 1* (Thermodynamics Research Center, Texas, 1994).

(46) Montgomery, Jr., J. A.; Petersson, G. A. *Chem. Phys. Lett.* **1990**, *168*, 75.

CHAPTER 4

Radical Stabilization Energies for •NHX Radicals

4.1 Introduction

Chapter 1 detailed a number of abstraction reactions that might occur in a protein under oxidative stress. Primarily, hydrogen atoms are abstracted by initiating radicals, such as •OH, to form carbon-centered radicals. However, oxidants such as hypochlorous acid preferentially form chloramines that subsequently decompose to generate nitrogen-centered radicals. In this Chapter, the effect of a substituent on the stability of a number of nitrogen-centered radicals relative to its effect in the parent closed-shell molecule will be scrutinized by using the radical stabilization energy or RSE.[1,2] For substituted methyl radicals, $•CH_2X$, the RSE is generally measured by the energy change in the isodesmic reaction:

$$•CH_2X + CH_4 \rightarrow •CH_3 + CH_3X \qquad (4.1)$$

This is equivalent to the difference in the bond dissociation energies of CH_4 and CH_3X:

$$RSE(•CH_2X) = BDE(CH_4) - BDE(CH_3X) \qquad (4.2)$$

Recent studies[2,3] have carried out extensive examinations of the RSEs of substituted methyl radicals at a number of levels of theory. The studies found that good results are generally obtained with high-level methods such as CBS-QB3 and G3X(MP2)-RAD but that caution needs to be exercised at lower theoretical levels where there is often incomplete cancellation of errors in reaction 4.1. In particular, unrestricted Møller-Plesset theory (e.g., UMP2) can lead to poor results when there is severe spin contamination in the $•CH_2X$ radical. However, *restricted* Møller-Plesset theory (specifically RMP2) is found to be much more reliable. [2,3]

In a subsequent investigation, Song, Cheng, Fu, Liu and Guo (SCFLG)[4] carried out an interesting study of the corresponding radical stabilization energies of substituted amino radicals, •NHX, measured as the energy change in the isodesmic reaction:

$$\bullet NHX + NH_3 \rightarrow \bullet NH_2 + NH_2X \quad (4.3)$$

This is equivalent to the difference in bond dissociation energies between NH_3 and NH_2X:

$$RSE(\bullet NHX) = BDE(NH_3) - BDE(NH_2X) \quad (4.4)$$

SCFLG[4] found good results with the G3 and CBS-Q methods, and also agreed with our finding that methods such as UMP2 could produce poor results in the cases of radicals showing high spin contamination. However, they also concluded that RMP2 is unreliable in several cases, specifically $\bullet NHCF_3$, $\bullet NHCOCH_3$ and $\bullet NHCONH_2$. They also found that RB3-LYP performs poorly for $\bullet NHCF_3$. In addition, they concluded that even UCCSD(T) fails for the •NHCHO radical. Although it is entirely possible that methods such as RMP2 and UCCSD(T) may have greater difficulty in describing •NHX radicals than $\bullet CH_2X$ radicals, we felt that it was a sufficiently important question to warrant additional examination.

In the present Chapter, a re-examination of the performance of selected levels of theory in calculating BDEs and RSEs will be carried out for the designated amino radicals, i.e., •NHX with X = CF_3, CHO, $COCH_3$ and $CONH_2$.

4.2 Theoretical Procedures

Calculations on radicals were performed either with a restricted-open-shell reference wave function, designated with an 'R' prefix (e.g., RHF, RMP2), or with an unrestricted-open-shell wave function, designated with a 'U' prefix (e.g., UHF, UMP2).

The frozen-core (fc) approximation was employed in all MP2 and CCSD(T) calculations. In order to facilitate comparisons with the results of SCFLG,[4] bond dissociation energies (BDEs) and radical stabilization energies (RSEs) were obtained from geometries optimized at the UB3-LYP/6-31+G(d) level followed by single-point energy calculations with the 6-31+G(d) basis set at the following theoretical levels: UHF, RHF, UB3-LYP, RB3-LYP, UMP2, RMP2, UCCSD(T) and URCCSD(T). Unscaled UB3-LYP/6-31+G(d) zero-point vibrational energies were used in these calculations, again to facilitate comparisons with the earlier work. Additional BDEs and RSEs were obtained through single-point calculations at all these theoretical levels with the 6-311+G(2df,p) basis set on UB3-LYP/6-31G(d) optimized geometries, incorporating UB3-LYP/6-31G(d) zero-point vibrational energies scaled by a factor of 0.9806.[5] Finally, high-level results were obtained with CBS-QB3[6,7] and G3X(MP2)-RAD,[8] and with W1,[9] the latter providing the ultimate benchmarks unless otherwise noted. We report CBS-QB3 values with and without the spin correction, the latter being denoted U-CBS-QB3 to distinguish it from the standard CBS-QB3 procedure.

4.3 Results and Discussion

4.3.1 States of the Amino and Substituted Amino Radicals

The $\bullet NH_2$ radical has two low-lying states that may be classified according to whether the unpaired electron is in a b_1 orbital ($^2B_1$ state in C_{2v} symmetry, equivalent to $^2A''$ in C_s symmetry) or in an a_1 orbital ($^2A_1$ state in C_{2v} symmetry, equivalent to $^2A'$ in C_s symmetry) (see Figure 4.1). It is well established that the ground state of $\bullet NH_2$ is $^2B_1$, with the $^2A_1$ state lying approximately 133 kJ mol^{-1} higher in energy.[10]

In examining substituted amino radicals, •NHX, it is important to bear in mind that substituents may affect this ordering of states, and that distortions from C_s symmetry will allow mixing of $^2A'$ and $^2A''$. Geometries of the •NHX radicals were

initially obtained at the UB3-LYP/6-31+G(d) level in C_s symmetry, for both ^{2}A" and ^{2}A' states.

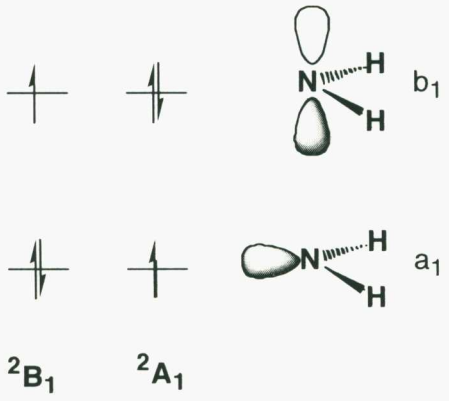


Figure 4.1: Occupation of orbitals in the ground ($^2B_1$) and first excited ($^2A_1$) states of the •NH_2 radical.

Where such C_s structures were indicated through frequency calculations to represent saddle points on the surface, re-optimizations in C_1 symmetry were performed. The global minimum was found in all cases to correspond either to a ^{2}A" state of C_s symmetry or to a pseudo-^{2}A" state of C_1 symmetry. Calculated energy differences between the ground-state structure, either ^{2}A" or pseudo-^{2}A", and the lowest energy ^{2}A' state are presented in Table 4.1.

The C_s ^{2}A' structures are found in several cases to be saddle points on the potential energy surface and so their relative energies represent an energy gap to constrained structures. Distortions from such structures lead monotonically downhill to the pseudo-A" minima. For the purpose of comparison with the results of SCFLG,[4] we have also carried out single-point calculations on C_1 structures with an A'-like wave function. These results are referred to as pseudo-A' but they do not have real physical significance.

Table 4.1: Energy Differences Between the $^2A''$ or Pseudo-$^2A''$ Ground State and the $^2A'$ Excited State of •NHX Radicals (0 K, kJ mol^{-1})[a]

X	UHF	RHF	CBS-QB3	G3X(MP2)-RAD	W1	Expt[b]
H	139.2	136.3	132.7	135.3	131.6	133.1
CF_3	127.8[c]	125.1[c]	121.8	124.0[c]	–	–
CHO	52.1	31.7	26.0[c]	28.9[c]	22.8[c]	–
$COCH_3$	45.1	27.1	18.2[c]	14.9[c]	–	–
$CONH_2$	71.2[c]	69.9[c]	89.2[c]	86.3[c]	–	–

[a] See text for details. [b] See reference 10. [c] Note that for these radicals the C_s $^2A'$ structure is a saddle point on the potential energy surface at the theoretical levels specified. See text.

Although it is not a primary emphasis in the current Chapter, we note that the calculated $^2B_1 - {}^2A_1$ splittings between the states for the parent $•NH_2$ system are in good agreement with the experimental value (133.1 kJ mol^{-1})[10] at all the levels of theory presented (Table 4.1). There is also good agreement between the HF and high-level values for the $^2A' - {}^2A''$ splittings for the $•NHCF_3$ radical. However, for the carbonyl-substituted radicals, the results change significantly in moving from HF to the high-level procedures, although it is reassuring that the latter agree well with one another. The splitting between the states varies greatly with substitution. Table 4.1 shows that the range is about 15 to 135 kJ mol^{-1}. The CHO and $COCH_3$ substituents provide significant relative stabilization to the $^2A_1$ state through delocalization of the nitrogen lone pair into the π^* orbital of the carbonyl, thus substantially lowering the $^2A' - {}^2A''$ splitting.

4.3.2 Bond Dissociation Energies (BDEs)

Table 4.2 presents a comparison of the N–H BDEs obtained by SCFLG[4] at the UHF, RHF, UB3-LYP, RB3-LYP, UMP2, RMP2 and UCCSD(T) levels (all with the 6-31+G(d) basis set), with values obtained at these levels in the present work. Table 4.3 presents results obtained with these theoretical procedures and the 6-311+G(2df,p) basis set. Also included in Table 4.3 are N–H BDEs obtained with the high-level W1 procedure of Martin et al.[9] which are used as a benchmark, as well as with the high-level CBS-QB3 and G3X(MP2)-RAD procedures.

Comparison with the Results of SCFLG.[4] Examination of the BDEs in Table 4.2 shows differences from the previously reported results of SCFLG[4] by up to 200 kJ mol^{-1}. For $•NHCF_3$, the differences observed with RHF, RB3-LYP and RMP2 between the present results and those of SCFLG[4] are in the range 175–200 kJ mol^{-1}. For •NHCHO, there are differences of 50–120 kJ mol^{-1} at the UHF, UMP2 and UCCSD(T) levels. For $•NHCOCH_3$, differences of 80–110 kJ mol^{-1} are observed with RHF and RMP2. Finally, for $•NHCONH_2$, differences of 70–180 kJ mol^{-1} are observed at the RHF and RMP2 levels.

What is responsible for the huge differences between the BDEs obtained in the study of SCFLG[4] and in the present work? SCFLG[4] attributed their discrepant results in these cases to poor performance by the theoretical procedures. However, we do not agree with this assessment. Rather, we have been able to trace the problem to convergence to the wrong SCF solution in each of the anomalous cases in the study by SCFLG.[4] Essentially, the incorrect solutions correspond to an A' state (for $•NHCF_3$) or to a pseudo-A' state (for •NHCHO, $•NHCOCH_3$ and $•NHCONH_2$). By appropriately manipulating the wave function, we are able to reproduce such results.[11]

Table 4.2: N–H Bond Dissociation Energies Associated with •NHX Radicals (0 K, kJ mol^{-1})[a,b]

X	UHF	RHF	UB3-LYP	RB3-LYP	UMP2	RMP2	UCCSD(T)	URCCSD(T)
H								
Previous work[c]	301.7	313.4	426.3	430.1	401.2	401.2	392.0	–
Present work	301.9	313.3	426.1	430.2	401.3	401.2	392.2	392.3
CF_3								
Previous work[c]	316.7	525.1	433.0	615.9	418.0	607.9	406.7	–
Present work	314.2	328.6	433.2	437.3	417.9	417.4	406.7	406.7
CHO								
Previous work[c]	443.9	370.3	452.3	461.1	537.2	442.7	485.3	–
Present work	328.2	370.1	452.0	461.1	480.1	442.7	421.8	421.1
$COCH_3$								
Previous work[c]	331.4	466.5	441.4	448.9	460.7	516.3	413.8	–
Present work	331.4	359.4	441.5	449.1	461.0	434.2	414.0	416.5
$CONH_2$								
Previous work[c]	322.6	515.9	423.4	429.3	417.6	484.1	400.4	–
Present work	322.4	337.3	423.6	429.2	417.7	415.6	400.5	400.4
MAD[d]	139.0	116.9	23.3	17.2	25.3	36.4	51.6	51.2
LD[e]	–146.1	–130.9	–28.0	–22.4	–42.9	–43.0	–53.6	–53.2

[a]Corresponding to the energy change for the reaction $NH_2X \rightarrow$ •NHX + •H. [b]Obtained from single-point calculations with the 6-31+G(d) basis set on the UB3-LYP/6-31+G(d) geometries. [c]From SCFLG.[4] [d]Mean absolute deviation from benchmark values (W1 unless otherwise specified). [e]Largest deviation from benchmark values (W1 unless otherwise specified).

Table 4.3: N–H Bond Dissociation Energies Associated with •NHX Radicals at Higher Theoretical Levels (0 K, kJ mol^{-1})[a]

X	UB3-LYP[b]	RB3-LYP[b]	UMP2[b]	RMP2[b]	UCCSD(T)[b]	URCCSD(T)[b]	G3X(MP2)-RAD	CBS-QB3[c]	W1
H	431.5	436.3	433.1	432.0	425.0	425.1	439.7	444.2 (444.4)	444.2
CF_3	438.4	443.8	447.1	445.6	436.9	437.0	451.0	455.8 (456.0)	455.3
CHO	460.7	470.6	514.3	476.0	457.1	455.9	467.6	470.2 (475.2)	474.3
$COCH_3$	449.3	457.6	496.5	468.7	450.4	449.3	463.1	463.4 (466.8)	–[d]
$CONH_2$	431.3	436.9	449.9	446.4	433.0	432.8	447.8	449.6 (450.4)	451.6
MAD[d]	16.4	9.6	18.0	6.0	18.1	18.6	4.8	2.2 (0.8)	0.0
LD[e]	–20.3	–14.7	+40.0	–12.2	–19.2	–19.1	–6.7	–4.2 (–1.2)	0.0

[a]Corresponding to the energy change for the reaction $NH_2X \rightarrow$ •NHX + •H. [b]Obtained from single-point calculations with the 6-311+G(2df,p) basis set on the UB3-LYP/6-31G(d) geometries. [c]Values in parenthesis obtained without the spin-correction term (U-CBS-QB3). [d]Benchmark value taken as 467.6 kJ mol^{-1}, see text.

Assessment of Results. It is convenient to use the high-level W1 benchmark results listed in Table 4.3 to assess the performance of simpler levels of theory for the calculation of the BDEs associated with the •NHX radicals.

•*NH_2.* Of the high-level methods, W1 and CBS-QB3 perform best and both give an N–H BDE for NH_3 of 444.2 kJ mol^{-1}, compared with the experimental BDE of 446.6 ± 1.3 kJ mol^{-1}.[10] G3X(MP2)-RAD yields a BDE of 439.7 kJ mol^{-1}. The results obtained with the direct methods listed in Table 4.2 in conjunction with the 6-31+G(d) basis set are in poorer agreement with experiment. At the extreme is UHF/6-31+G(d) which gives a BDE for NH_3 that is more than 140 kJ mol^{-1} less than the W1 benchmark value. The use of restricted wave functions with the 6-31+G(d) basis set gives little improvement over the unrestricted equivalents. RB3-LYP/6-31+G(d) gives the best result with the smaller basis set. Increasing the size of the basis set to 6-311+G(2df,p) leads to a noticeable improvement for all methods, and they all yield BDEs for NH_3 that are within 20 kJ mol^{-1} of the W1 value.

•*$NHCF_3$.* The benchmark W1 BDE for NH_2CF_3 to give the $^2A''$ ground state of the •$NHCF_3$ radical is 455.3 kJ mol^{-1}. The CBS-QB3 and G3X(MP2)-RAD BDEs agree well with the benchmark, giving values of 455.8 and 451.0 kJ mol^{-1}, respectively. The direct methods when used with the 6-31+G(d) basis set do not perform as well, but there is again considerable improvement in all cases in going to the larger 6-311+G(2df,p) basis set.

•*NHCHO.* The W1 procedure gives an N–H BDE for NH_2CHO of 474.3 kJ mol^{-1}. The CBS-QB3 and G3X(MP2)-RAD levels again agree well with the benchmark value, giving N–H BDEs of 470.2 and 467.6 kJ mol^{-1}, respectively. Even closer agreement with W1 is obtained when the spin-correction term in CBS-QB3 is omitted (U-CBS-QB3), giving 475.2 kJ mol^{-1}, but as pointed out in Chapter 3 this a fortuitous result. Of the direct methods, UMP2 unexpectedly performs the best with the 6-31+G(d) basis set, giving a BDE of 480.1 kJ mol^{-1}. In contrast to the other theoretical

methods, the UMP2 result actually worsens in going from 6-31+G(d) to 6-311+G(2df,p).

•*NHCOCH*$_3$. Due to the size of this species, it was not possible to calculate the N–H BDE at the W1 level with our existing computing resources. In order to provide a suitably reliable BDE, we have used the G3X(MP2)-RAD RSE of –23.4 kJ mol^{-1} (see below) in conjunction with the W1 BDE of 444.2 kJ mol^{-1} for •NH_2 to obtain a benchmark BDE of 467.6 kJ mol^{-1}. The CBS-QB3 and G3X(MP2)-RAD BDEs for •$NHCOCH_3$ of 463.4 and 463.1 kJ mol^{-1}, respectively, are slightly less than this benchmark value, as is also observed for •NHCHO. We can see from Table 4.2 that UMP2/6-31+G(d) again fortuitously gives the closest agreement with the benchmark, a result that becomes worse with the larger 6-311+G(2df,p) basis set. The other theoretical procedures all give improved BDEs with the 6-311+G(2df,p) basis set.

•*NHCONH*$_2$. The three composite methods all agree well for the N–H BDE of NH_2CONH_2. The benchmark W1 value is 451.6 kJ mol^{-1} while the CBS-QB3 and G3X(MP2)-RAD procedures give BDEs of 449.6 and 447.8 kJ mol^{-1}, respectively. U-CBS-QB3 gives 450.4 kJ mol^{-1}, but as pointed out in Chapter 3, this is a fortuitous result. There is improvement in going from 6-31+G(d) to 6-311+G(2df,p) with all the direct methods.

General Comparisons. Tables 4.2 and 4.3 include mean absolute deviations (MADs) and largest deviations (LDs) from benchmark values (W1 unless otherwise specified) for the five BDEs. Although the sample set is too small to be statistically significant, some useful general features do emerge. We note to begin that the $\mathcal{T}_1$ diagnostic values[12] calculated at the UCCSD/AVTZ//UB3-LYP/cc-pVTZ level of theory are 0.013, 0.034, 0.031 and 0.026 for •$NHCF_3$, •NHCHO, •$NHCOCH_3$ and •$NHCONH_2$, respectively. This suggests that •NHCHO, •$NHCOCH_3$ and •$NHCONH_2$ in particular represent challenging targets for calculations based on a single-reference configuration. In a related vein, the $<S^2>$ values found at UMP2/6-31+G(d)//UB3-

LYP/6-31+G(d) for •$NHCF_3$, •NHCHO, •$NHCOCH_3$ and •$NHCONH_2$ are 0.759, 0.959, 0.907 and 0.783, respectively. This suggests potential difficulties for UMP with •NHCHO and •$NHCOCH_3$, as is indeed observed.

We find that the composite W1, CBS-QB3 and G3X(MP2)-RAD methods give results that agree well with one another and with the only experimental BDE for the •NHX radicals. CBS-QB3 without the spin-correction term (U-CBS-QB3) unexpectedly performs particularly well, with an MAD from W1 of just 0.8 kJ mol^{-1} and an LD of –1.2 kJ mol^{-1}. Of the simpler levels of theory, RB3-LYP performs best with the 6-31+G(d) basis set, giving an MAD of 17.2 kJ mol^{-1}. When the larger 6-311+G(2df,p) basis set is used, RB3-LYP and RMP2 produce the best results, with MADs of 9.6 and 6.0 kJ mol^{-1}, respectively. We emphasize that care must be taken to converge to the lowest-energy SCF solutions to avoid spuriously anomalous results.

4.3.3 Radical Stabilization Energies (RSEs)

The radical stabilization energy of a substituted amino radical •NHX provides a measure of the effect of the substituent X on the stability of the •NHX radical relative to its effect in the parent closed-shell molecule NH_2X. It is given by the energy change in reaction 4.3 (above). Equivalently, it represents the difference in BDEs of NH_3 and NH_2X.

Table 4.4 presents a comparison of the RSEs for •NHX radicals obtained by SCFLG[4] at the UHF, RHF, UB3-LYP, RB3-LYP, UMP2, RMP2 and UCCSD(T) levels (all with the 6-31+G(d) basis set) with values obtained at these levels in the present work. Table 4.5 presents results obtained with these theoretical procedures and the 6-311+G(2df,p) basis set. Also included in Table 4.5 are the RSEs obtained with the high-level W1 procedure of Martin et al.[9] which are used as a benchmark unless otherwise noted, as well as with the high-level CBS-QB3 and G3X(MP2)-RAD procedures.

Table 4.4: Radical Stabilization Energies of •NHX Radicals (0 K, kJ mol^{-1})[a,b]

X	UHF	RHF	UB3-LYP	RB3-LYP	UMP2	RMP2	UCCSD(T)	URCCSD(T)
CF_3								
Previous work[c]	–14.6	–211.7	–7.1	–185.4	–16.7	–206.7	–14.6	–
Present work	–12.3	–15.3	–7.0	–7.1	–16.6	–16.2	–14.5	–14.5
CHO								
Previous work[c]	–142.2	–56.9	–25.9	–31.0	–136.0	–41.4	–93.3	–
Present work	–26.3	–56.8	–25.9	–30.9	–76.8	–41.4	–29.6	–28.8
$COCH_3$								
Previous work[c]	–29.7	–153.1	–15.1	–18.8	–59.4	–115.1	–24.3	–
Present work	–29.5	–46.1	–15.3	–18.9	–59.7	–33.0	–21.8	–24.3
$CONH_2$								
Previous work[c]	20.5	–202.5	2.5	0.8	–16.3	–82.8	–8.4	–
Present work	–20.5	–24.0	2.5	1.0	–16.5	–14.3	–8.3	–8.1
MAD[d]	6.1	17.5	6.6	4.4	24.4	8.2	1.6	1.6
LD[e]	–13.1	–26.6	+9.9	+8.4	–46.6	–11.2	–3.4	–3.4

[a]Corresponding to the energy change for the reaction $•NHX + NH_3 \rightarrow •NH_2 + NH_2X$. [b] Obtained from single-point calculations with the 6-31+G(d) basis set on the UB3-LYP/6-31+G(d) geometries. [c]From SCFLG.[4]

Table 4.5: Radical Stabilization Energies of •NHX Radicals at Higher Theoretical Levels (0 K, kJ mol^{-1})[a]

X	UB3-LYP[b]	RB3-LYP[b]	UMP2[b]	RMP2[b]	UCCSD(T)[b]	URCCSD(T)[b]	G3X(MP2)-RAD	CBS-QB3[c]	W1
CF_3	–6.9	–7.5	–14.0	–13.6	–11.9	–11.9	–11.3	–11.6 (–11.6)	–11.1
CHO	–29.2	–34.4	-81.2	–44.0	–32.1	–30.8	–27.9	–26.0 (–30.8)	–30.2
$COCH_3$	–17.8	–21.4	–63.4	–36.7	–25.4	–24.3	–23.4	–19.1 (–22.4)	–[d]
$CONH_2$	0.2	–0.7	–16.8	–14.4	–8.0	–7.7	–8.2	–5.4 (–6.0)	–7.4
MAD[d]	4.6	4.1	25.8	9.2	1.3	0.7	0.8	2.8 (0.9)	0.0
LD[e]	+7.6	+6.7	–51.0	–13.8	–2.0	–0.9	+2.3	+4.3 (+1.4)	0.0

[a]Corresponding to the energy change for the reaction $NH_2X \rightarrow$ •NHX + •H. [b]Obtained from single-point calculations with the 6-311+G(2df,p) basis set on the UB3-LYP/6-31G(d) geometries. [c]Values in parenthesis obtained without the spin-correction term (U-CBS-QB3).

Comparison with the Results of SCFLG.[4] Examination of the RSEs in Table 4.4 shows differences from the previously reported results of SCFLG[4] of up to 200 kJ mol^{-1}. These are not unexpected, given the discrepancies in the N–H BDEs discussed above. However, the calculated RSEs in Tables 4.4 and 4.5 make it clear that it is the SCFLG[4] values that are anomalous. For example, for $•NHCF_3$, all the RSEs determined in the present work lie within 6 kJ mol^{-1} of the W1 benchmark (Table 4.5). In contrast, the RHF, RB3-LYP and RMP2 values of SCFLG[4] differ by 170–200 kJ mol^{-1} from the W1 value. Likewise, anomalous RSEs from the previous work of SCFLG[4] can be seen for •NHCHO at the UHF, UMP2 and UCCSD(T) levels, for $•NHCOCH_3$ with RHF and RMP2, and for $•NHCONH_2$ with RHF and RMP2. The anomalous results all have their origins in convergence to the wrong SCF solution in the previous study, and not to poor performance of the theoretical procedures as had been previously concluded. [4]

Assessment of Results. As with the BDEs discussed in the previous section, the RSEs determined with the high-level W1 procedure generally represent the benchmark values for this work and are presented in Table 4.5. For $•NHCOCH_3$, for which we do not have a W1 RSE, we have taken the G3X(MP2)-RAD value as the benchmark. The benchmark RSEs are thus –11.1 ($•NHCF_3$), –30.2 (•NHCHO), –23.4 ($•NHCOCH_3$) and –7.4 ($•NHCONH_2$) kJ mol^{-1}, respectively. It can be seen from Tables 4.4 and 4.5 that for many of the theoretical procedures there is substantial cancellation of errors in the calculated RSEs, i.e., the MADs are significantly smaller than corresponding MADs for BDEs (Tables 4.2 and 4.3). This is not unexpected, given the nature of the equation defining the radical stabilization energies (equation 4.3). The basis set dependence of the RSEs is much smaller than for the BDEs, i.e., the RSEs determined with the 6-31+G(d) basis set are not substantially inferior to those obtained with the 6-311+G(2df,p) basis. The worst results (indicating less complete cancellation of errors) are observed with RHF and UMP2. This can be ascribed in the latter case to variable spin contamination in the reactant and product radicals of equation 4.3.

Both G3X(MP2)-RAD and CBS-QB3 give RSEs that compare well with the W1 benchmark values. The MADs and LDs are 0.8 and +2.3 kJ mol^{-1}, respectively, for G3X(MP2)-RAD and 2.8 and +4.3 kJ mol^{-1}, respectively, for CBS-QB3. For U-CBS-QB3, the MAD and LD from W1 decrease to just 0.9 and +1.4 kJ mol^{-1}, respectively.

Of the direct methods, best results are obtained with URCCSD(T)/6-311+G(2df,p). The MAD is just 1.2 kJ mol^{-1} while the LD is 3.0 kJ mol^{-1}. Although URCCSD(T)/6-311+G(2df,p) provides a direct means of obtaining accurate RSEs, it is computationally intensive and would not be tractable for systems of moderate to large size. A more economical, though still reliable, procedure would be desirable.

In previous studies on $•CH_2X$ radicals,[2,3] it was found that RMP2/6-311+G(2df,p)//UB3-LYP/6-31G(d) provides a cost-effective means of obtaining reasonable RSEs. The results of the present work suggest that although it is also a reasonable possibility for •NHX radicals, the discrepancies from benchmark values are somewhat larger. The MAD for RMP2/6-311+G(2df,p)//UB3-LYP/6-31G(d) is 9.2 kJ mol^{-1} with an LD of –13.8 kJ mol^{-1}. The MAD and LD for RMP2/6-31+G(d)//UB3-LYP/6-31+G(d) are 8.2 and –11.2 kJ mol^{-1}, respectively, which are slightly better than the RMP2/6-311+G(2df,p)//UB3–LYP/6-31G(d) results. These differences are not believed to be statistically significant.

The UMP2/6-31+G(d) and UMP2/6-311+G(2df,p) RSEs show large deviations from the benchmark values for •NHCHO and $•NHCOCH_3$. This supports the conclusion that UMP2 has difficulty with species that show large spin contamination. As noted above, the $<S^2>$ values found at UMP2/6-31+G(d)//UB3-LYP/6-31+G(d) for •NHCHO and $•NHCOCH_3$ are 0.959 and 0.907, respectively.

Both UB3-LYP and RB3-LYP show relatively small MADs and LDs for their calculated RSEs. With the 6-31+G(d) basis set, the MADs and LDs are 6.6 and +9.9 kJ mol^{-1}, respectively, for UB3-LYP, and 4.4 and +8.4 kJ mol^{-1}, respectively for RB3-LYP. With the 6-311+G(2df,p) basis set, the MADs and LDs are slightly better at 4.6

and +7.6 kJ mol^{-1}, respectively, for UB3-LYP and 4.1 and +6.7 kJ mol^{-1}, respectively for RB3-LYP. One shortcoming of UB3-LYP/6-31+G(d), RB3-LYP/6-31+G(d) and UB3-LYP/6-311+G(2df,p) is that they predict positive RSEs for $•NHCONH_2$ (i.e., a relative stabilizing effect in the radical), in contrast to the small negative benchmark value.

UHF/6-31+G(d) provides good RSEs for three of the •NHX radicals when compared with the benchmark values. The largest deviation at this level is –13.1 kJ mol^{-1} for $•NHCONH_2$. While RMP2 performs significantly better than UMP2, changing from an unrestricted to a restricted approach at the HF level actually leads to a deterioration in the results. Thus with the 6-31+G(d) basis set, the MADs and LDs are 6.1 and –13.1 kJ mol^{-1}, respectively, for UHF, and 17.5 and –26.6 kJ mol^{-1}, respectively, for RHF. With the 6-311+G(2df,p) basis set,[13] the MADs and LDs are slightly worse at 7.0 and –14.3 kJ mol^{-1}, respectively, for UHF and 19.8 and –30.8 kJ mol^{-1}, respectively for RHF.

In summary, there is good agreement among the high levels of theory (W1, G3X(MP2)-RAD and CBS-QB3) in the calculation of RSEs. U-CBS-QB3 gives results that are particularly close to W1. UCCSD(T) and URCCSD(T) also perform very well. RB3-LYP is the best of the computationally less demanding procedures. RMP2 performs somewhat less well for •NHX radicals than found previously for $•CH_2X$ radicals, though the results are still reasonable. Very poor results found previously[4] with RB3-LYP, RMP2 and UCCSD(T) may be attributed to convergence to the wrong SCF solutions. UMP2 sometimes gives large errors for RSEs and therefore should not be used when spin contamination is significant. RHF can also produce poor RSEs.

Substituent Effects in •NHX Radicals. The calculated radical stabilization energies in Tables 4.4 and 4.5 are all negative. This means that the •NHX radicals are destabilized relative to the corresponding closed-shell molecules NH_2X in each case. This effect is discussed in detail by SCFLG, based on G3 and CBS-Q results.[4] In brief,

the negative RSEs can be rationalized in terms of a reduction in the delocalization of the nitrogen lone pair that can take place in NH_2CHO, NH_2COCH_3 and NH_2CONH_2, when moving to the •NHX radicals. This reduction arises because the electronic ground state is now pseudo-^{2}A", and so it is the unpaired electron rather than the lone pair that is more strongly delocalized. The negative RSE for $•NHCF_3$ may be attributed largely to the electron-withdrawing nature of the CF_3 substituent destabilizing the electron-deficient radical center.

4.3.4 Spin Distributions

The discrepancies found in the BDEs and RSEs for the $•NHCF_3$, $•NHCOCH_3$ and $•NHCONH_2$ radicals at RMP2/6-31+G(d) were attributed by SCFLG[4] to a failure of RMP2 to correctly assign the spin distribution. They pointed out that for these three radicals RMP2 predicts an oxygen-centered radical, whereas the unpaired electron is mainly localized on the nitrogen with UB3-LYP and UCCSD(T). We find from the present work that the calculations of SCFLG[4] in these cases correspond to the excited ^{2}A' state for $•NHCF_3$, and to pseudo-^{2}A' states for $•NHCOCH_3$ and $•NHCONH_2$. For the A' and pseudo-A' states, the unpaired electron is indeed centered on the oxygen. On the other hand, we find that if the correct ground state A" or pseudo-A" state is examined, the spin is localized on nitrogen in all cases. Figure 4.2 displays the RMP2/6-31+G(d) spin distributions for the ^{2}A" ground state of $•NHCF_3$ and the pseudo-^{2}A" ground states of $•NHCOCH_3$ and $•NHCONH_2$. It is clear that the unpaired electron is localized on the nitrogen in all three radicals, in agreement with UB3-LYP and UCCSD(T).

Similarly, SCFLG[4] found that their UCCSD(T) calculation incorrectly assigns the unpaired electron in •NHCHO largely to the oxygen, in contrast to UB3-LYP and RMP2 which assign the electron to the nitrogen. In this case also the result may be attributed to their convergence to the higher-energy pseudo-^{2}A' state. Figure 4.2 includes the spin distribution for the pseudo-^{2}A" ground state of •NHCHO at

UCCSD(T)/6-31+G(d). Again it is clear that if the correct ground state is examined, the unpaired electron is mainly localized on nitrogen.

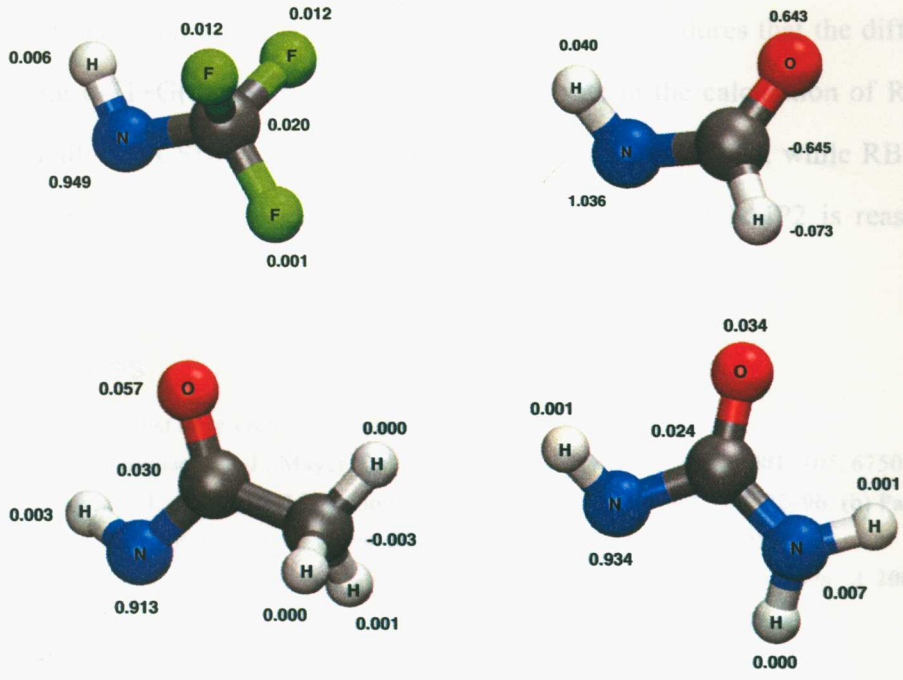


Figure 4.2: RMP2/6-31+G(d)//UB3-LYP/6-31+G(d) spin distributions for the •$NHCF_3$, •$NHCOCH_3$, and •$NHCONH_2$ radicals, and UCCSD(T)/6-31+G(d)//UB3-LYP/6-31+G(d) spin distribution for the •NHCHO radical.

4.4 Concluding Remarks

A primary purpose of the present Chapter has been to re-examine the conclusion reached by SCFLG[4] that for several •NHX radicals, specifically •$NHCF_3$, •NHCHO, •$NHCOCH_3$ and •$NHCONH_2$, RHF, RMP2 RB3-LYP and UCCSD(T) can lead to unrealistic spin localization and hence to poor BDEs and RSEs. They state that their results "justify many people's suspicion that the use of the spin-restricted wave function method is not always safe." This assessment is not correct. The poor performance is associated with calculations being performed on excited A' states or unphysical pseudo-A' solutions to the SCF equations for the •NHX radicals. Once the correct ground state is used in the BDE and RSE calculations, the very large errors observed previously disappear.

The high-level W1, G3X(MP2)-RAD and CBS-QB3 procedures all yield RSEs that agree closely with one another. CBS-QB3 performs particularly well when the spin-correction term is omitted (U-CBS-QB3), although this may be fortuitous (see Chapter 3). It has been found with the direct theoretical procedures that the difference between using 6-31+G(d) and 6-311+G(2df,p) basis sets in the calculation of RSEs is relatively small. UCCSD(T) and URCCSD(T) give very good RSEs, while RB3-LYP provides a very cost-effective alternative. The performance of RMP2 is reasonable though not as good as found previously for $•CH_2X$ radicals.

4.5 References

(1) For an extensive list of references, see reference 2.

(2) Henry, D. J.; Parkinson, C. J.; Mayer, P. M.; Radom, L. *J. Phys. Chem. A,* **2001**, *105*, 6750.

(3) (a) Parkinson, C. J.; Mayer, P. M.; Radom, L. *Theor. Chem. Acc.* **1999**, *102*, 92–96. (b) Parkinson, C. J.; Mayer, P. M.; Radom, L. *J. Chem. Soc. Perkin Trans. 2* **1999**, *11*, 2305.

(4) Song, K. S.; Cheng, Y. H.; Fu, Y.; Liu, L.; S., L. X.; Guo, Q. X. *J. Phys. Chem. A* **2002**, *106*, 6651.

(5) Scott, A. P.; Radom, L. *J. Phys. Chem.* **1996**, *100*, 16502.

(6) Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J.; Petersson, G. A. *J. Chem. Phys.* **2000**, *112*, 6532.

(7) Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1999**, *110*, 2822.

(8) Henry, D. J.; Sullivan, M. B.; Radom, L. *J. Chem. Phys.* **2003**, *118*, 4849.

(9) (a) Martin, J. M. L.; de Oliveira, G. *J. Chem. Phys.* **1999**, *111*, 1843. (b) Parthiban, S.; Martin, J. M. L. *J. Chem. Phys.* **2001**, *114*, 6014.

(10) Pd, Rajendra; Chandra, P. *J. Chem. Phys.* **2001**, *114*, 7450, and references therein.

(11) Note that the incorrect energies in Table 2 correspond to A' or pseudo-A' calculations carried out at the particular level on B3-LYP/6-31+G(d) structures that have been optimized for the A" or pseudo-A" state. They are therefore significantly higher than the energies that would be derived from the splittings in Table 1, in which the A' calculations have been carried out on optimized A' structures.

(12) (a) Lee, T. J.; Taylor, P. R. *Int. J. Quantum Chem. Symp.* **1989**, *23*, 199. (b) Jayatilaka, D.; Lee, T. J. *J. Chem. Phys.* **1993**, *98*, 9734. The $\mathcal{T}_1'$ diagnostic for open-shell systems is defined in ref 12b. A value of greater than 0.02 for the $\mathcal{T}_1'$ diagnostic has been proposed as an indication that a multireference electron correlation procedure may be necessary to recover a large fraction of the nondynamical electron correlation energy.

(13) The individual UHF/6-311+G(2df,p)/UB3-LYP/6-31G(d) (RHF/6-311+G(2df,p)/UB3-LYP/6-31G(d) values in parentheses) BDEs (in kJ mol^{-1}) are as follows: NH_3 306.8 (319.7), NH_2CF_3 321.6, (334.9) NH_2CHO 336.2, (380.6) NH_2COCH_3 339.3 (369.4), NH_2CONH_2 328.5 (344.9).

The corresponding RSEs are: $•NHCF_3$ –14.8 (–15.3), •NHCHO –29.4 (–61.0), $•NHCOCH_3$ –32.5 (–49.8), $•NHCONH_2$ –21.7 (–25.2).

CHAPTER 5

RSEs and BDEs Associated with Peptide-Backbone Radicals

5.1 Introduction

The previous chapter detailed RSEs for •NHX radicals, which might be important in reactions of amino acids with HOCl. However, as detailed in Chapter 1, abstractions by •OH primarily occur from backbone or side chain C–H bonds. Of particular interest are hydrogen abstractions from the α-C–H bond, as this could result in the formation of highly stabilized α-C-centered radicals by means of the captodative effect (see Chapter 1, section 1.3.2). The deviation from additivity in the RSE of a disubstituted carbon-centered radical is given by:

$$\Delta(\text{RSE}) = \text{RSE}(\text{X–•CH–Y}) - \text{RSE}(\text{X–•CH–H}) - \text{RSE}(\text{H–•CH–Y}) \qquad (5.1)$$

Defined in this way, a positive Δ(RSE) means that the two substituents have a reinforcing stabilizing effect in the radical, a zero value would indicate that the RSEs are additive, while a negative number would indicate a destabilizing effect occurring in the disubstituted radical compared with the corresponding monosubstituted radicals.

In a previous extensive study,[1] C–H BDEs leading to α-carbon radicals in a number of model systems containing a glycine peptide backbone were calculated with B3-LYP/6-31G(d). The calculated C–H BDEs for the model systems, $HCONH_2–CH_2–CONH_2$, $CH_3CONH_2–CH_2–CONH_2$, $HCONH_2–CH_2–CONHCH_3$ and $CH_3CONH_2–CH_2–CONHCH_3$ were approximately equal, suggesting that an amide substituent on both sides of the α-carbon is sufficient to satisfactorily model the mid-chain environment of the C–H dissociation process. The accuracy of the BDEs calculated using B3-LYP/6-31G(d) were assessed against values obtained using the G2(MP2) high-level procedure.

In another recent study,[2] RSEs of α-carbon radicals were examined for the free amino acids, glycine, alanine and valine, and their N-acetyl methyl esters. In that study, the ordering of RSEs for the N-acetyl methyl ester derivatives, which attempt to model

a peptide environment, was found to differ from that for the free amino acids. For example, the largest RSE for the model peptides occurred for glycine, while the largest RSE for the free amino acids was found for alanine. The change in the ordering was attributed to steric interactions of the side-chain with the amide carbonyl groups. The RSEs of that study were evaluated with RMP2/6-31G(d)//UB3-LYP/6-31G(d). In another detailed study,[3] C–H BDEs were calculated for all the naturally occurring amino acids in model systems representative of a peptide backbone. It was found that the BDEs ranged from about 345 to 400 kJ mol^{-1}.

In the present Chapter, we re-examine the C–H BDEs of model systems representing a glycine-containing peptide backbone using high-level procedures and a wider range of models. In addition, RSEs are calculated for the model peptides and compared with RSEs calculated for a series of monosubstituted analogs. In a similar manner, N–H BDEs and their associated RSEs are calculated for a similar series of model systems. The RSEs are used to examine the effect of substitution on a radical center and also to examine the deviation from additivity in the RSEs of the disubstituted systems.

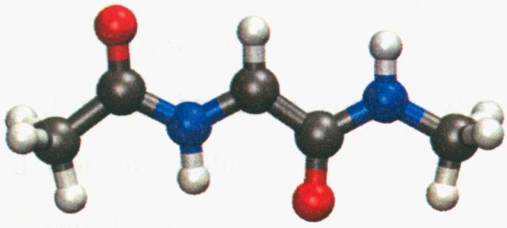
5.2 Theoretical Procedures

BDEs and RSEs were calculated with a number of theoretical procedures. Firstly, a series of simpler methods were used. These involved both MP2 and DFT single-point calculations in association with the 6-311+G(3df,2p) basis set carried out on UB3-LYP/6-31G(d) optimized geometries. Energies were adjusted to 0 K by adding a scaled (by 0.9806[4]) UB3-LYP/6-31G(d) zero-point vibrational energy (ZPVE). DFT single-point calculations were carried out with the B3-LYP combination of exchange and correlation functionals and with the BMK[5] and MPWB1K functionals.[6] We found that the performance of these two functionals are quite similar and so we only report on the BMK results throughout the text.

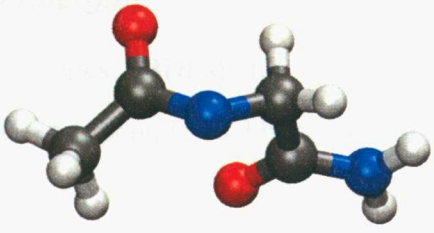
In addition to the simpler methods, standard high-level procedures were used to calculate BDEs and RSEs. The G3X(MP2)-RAD[7] and CBS-QB3[8] procedures were used for all species, while a selection were also examined with the W1 method of Martin et al.[9] Values obtained using U-CBS-QB3, which omits the spin correction term (see Chapter 3), are also reported.

5.3 Results and Discussion

The conformational spaces for all species were initially scanned using UB3-LYP/6-31G(d). In order to aid the analysis, extended conformations (corresponding to Ramachandran angles close to 180°) were used whenever possible. Figure 5.1(a) shows the optimized structure of the largest carbon-centered radical examined in this work ($CH_3CONH–•CH–CONHCH_3$) in an extended chain conformation. We found, however, that in a small number of cases, the lowest-energy conformations corresponded to folded structures due to internal hydrogen-bonding. These were handled as described below.



(a) $CH_3CONH–•CH–CONHCH_3$



(b) $CH_3CO–•N–CH_2CONH_2$

Figure 5.1: Optimized structures (B3-LYP/6-31G(d)) of the largest **(a)** carbon-centered radical ($CH_3CONH–•CH–CONHCH_3$) in an extended chain conformation and **(b)** nitrogen-centered radical ($CH_3CO–•N–CH_2CONH_2$) demonstrating the puckering that occurs at the nitrogen carrying the unpaired electron.

For the species involved in the series of carbon-centered radicals, the closed-shell molecule $CH_3CONH–CH_2–CONHCH_3$ is the only one in which the extended-chain conformation is not the global minimum. The difference in energy between the global minimum and the extended-chain conformation in this case is approximately 3 kJ mol^{-1} at the CBS-QB3 level of theory. For the sake of consistency, the numbers quoted are those derived from the extended-chain conformation of this species, but this does not affect the qualitative trends in our analysis

For the closed-shell counterparts of the nitrogen-centered radicals, the global minima all correspond to extended-chain conformations. For the nitrogen-centered radicals themselves, significant puckering occurs at the nitrogen carrying the unpaired electron. This makes it difficult in some cases to find conformations that do not contain some degree of internal hydrogen bonding. The species that are affected include the $CH_3CO–•NH–CH_2CHO$ and $CH_3CO–•NH–CH_2CONH_2$ radicals. Figure 5.1(b) shows the B3-LYP/6-31G(d) optimized structure of $CH_3CO–•NH–CH_2CONH_2$ that demonstrates the puckering that occurs on the nitrogen carrying the unpaired electron. The effect on the calculated BDEs and RSEs of the hydrogen bond that occurs in these species is approximately 5 kJ mol^{-1}, based on constrained-optimized rotational potentials. It is not expected that our qualitative conclusions will be affected by including in the analysis conformers that involve some internal hydrogen bonding.

5.3.1 Bond Dissociation Energies

Bond dissociation energies (BDEs) for various molecules that represent systematically larger models of a peptide-backbone containing a glycine residue are listed in Table 5.1 (C–H BDEs) and Table 5.2 (N–H BDEs). We have chosen W1 as the benchmark theoretical method as it has been shown to give excellent agreement with reliable experimental thermochemistry.[10]

Table 5.1: Comparison of Calculated C–H Bond Dissociation Energies (0 K, kJ mol^{-1}) Resulting in the Formation of α-Carbon-Centered Radicals on Small Models of a Glycine Peptide Backbone

radical	UB3-LYP[a]	RB3-LYP[a]	UBMK[a]	RBMK[a]	RMP2[a]	G3X(MP2)-RAD	CBS-QB3	U-CBS-QB3[b]	W1
H–•CH–H	424.7	428.8	431.0	435.1	417.3	429.1	434.3	434.5	432.3
H–•CH–CHO	379.4	387.3	390.1	399.5	385.1	393.7	393.3	397.6	395.7
H–•CH–$CONH_2$	394.9	399.8	406.3	411.5	397.0	407.4	408.7	409.8	
H–•CH–$CONHCH_3$	393.8	398.4	404.5	409.4	396.8	407.1	408.6	409.4	
NH_2–•CH–H	369.4	373.3	376.3	380.2	371.5	384.2	384.7	385.0	383.1
HCONH–•CH–H	371.4	375.9	379.4	384.2	376.1	387.4	389.4	390.0	
CH_3CONH–•CH–H	371.1	375.3	379.3	383.8	375.3	386.8	389.0	389.4	
NH_2–•CH–CHO	282.2	286.5	292.5	297.9	291.1	312.3	306.3	308.7	309.3
NH_2–•CH–$CONH_2$	313.7	317.2	319.8	323.5	327.0	339.1	336.9	337.5	
HCONH–•CH– CHO	304.5	309.2	321.4	327.2	314.8	335.4	332.2	335.3	
HCONH –•CH–$CONH_2$	330.8	334.5	338.8	342.9	341.3	353.5	351.7	352.5	
CH_3CONH–•CH–$CONHCH_3$	328.3	331.8	337.7	341.5	339.6	347.6	350.5	351.0	
MAD[c]	19.7	15.2	10.3	6.0	14.0	2.5	1.3		
MD[c]	–19.7	–15.2	–10.3	–5.3	–14.0	–1.4	–1.3		
LD[c]	–30.8	–26.1	–17.8	–14.0	–20.5	–5.4	–4.3		

[a] Calculations have been performed in association with the 6-311+G(3df,2p) basis set on UB3-LYP/6-31G(d) optimized geometries. Energies have been corrected to 0 K using UB3-LYP/6-31G(d) ZPVEs scaled by 0.9801.[4] [b] Standard CBS-QB3 but with the empirical spin-correction for spin contamination in the unrestricted wave function omitted. [c] Mean absolute deviation (MAD), mean deviation (MD) and largest deviation (LD) from the calculated U-CBS-QB3 results.

5.3.1.1 Carbon-Centered Radicals

W1 values for C–H BDEs have been calculated for four species, two of which have experimental values available for comparison. The calculated W1 BDEs at 0 K in these two cases, leading to the radicals H–•CH–H and H–•CH–CHO, are 432.3 and 395.7 kJ mol^{-1}, respectively, compared with experimental values of 432.3±0.4 and 389.5±9.2 kJ mol^{-1}.

The computational expense of W1 prohibits the use of this procedure beyond moderately-sized molecules. In a previous study of C–H BDEs,[11] the CBS-RAD procedure was found to provide a good secondary benchmark. In the present Chapter, we have carried out calculations with CBS-QB3 in place of CBS-RAD, the two procedures having been found previously to give similar results.[12] We also report results obtained with the U-CBS-QB3 method. The radicals in Table 5.1 that have been calculated with W1 have $<S^2>$ values of 0.7616 (H–•CH–H), 0.9221 (H–•CH–CHO), 0.7626 (NH_2–•CH–H) and 0.8462 (NH_2–•CH–CHO). In the two cases that have large spin contamination, i.e., H–•CH–CHO and NH_2–•CH–CHO, the calculated U-CBS-QB3 C–H BDEs are in closer agreement with the W1 values than are the standard CBS-QB3 values. Thus, for the purposes of this work we will use the U-CBS-QB3 BDEs as a secondary benchmark.

The largest spin-correction contribution to the CBS-QB3 procedure for the C–H BDEs examined in the present work amounts to –4.3 kJ mol^{-1} and occurs for H–•CH–CHO, which is the most highly spin-contaminated radical ($<S^2>$ = 0.9221). There is a relatively small mean absolute deviation (MAD) of 1.3 kJ mol^{-1} between the standard CBS-QB3 BDEs and the U-CBS-QB3 BDEs. The MAD and the mean deviation (MD) for CBS-QB3 are the same magnitude because the spin-correction term is always negative.

The other high-level composite method in Table 5.1 is G3X(MP2)-RAD. The MAD from the BDEs calculated with U-CBS-QB3 is 2.5 kJ mol^{-1}, which demonstrates

the good overall performance for this method. Unfortunately, as pointed out previously,[11] the largest deviation for this method (–5.4 kJ mol^{-1}) turns out to occur for the BDE of methane which, as will be seen, causes the radical stabilization energies (RSEs) to consistently be slightly underestimated.

In a previous study,[11] it was found that RB3-LYP/6-311+G(3df,2p)//RB3-LYP/6-31G(d) yielded good C–H BDEs. In that study, only monosubstituted species were considered. In the present work, RB3-LYP/6-311+G(3df,2p) single-point calculations on UB3-LYP/6-31G(d) optimized geometries lead to an MAD of 15.2 kJ mol^{-1}. The poorer overall performance of RB3-LYP in the present work arises because of the poorer results for the disubstituted radicals. Thus the smallest absolute deviation in the case of disubstuted radicals is 19.2 kJ mol^{-1} while the largest absolute deviation in the monosubstituted radicals is 14.1 kJ mol^{-1}. Using an unrestricted reference wave function instead of a restricted open-shell reference further increases the MAD, a trend also pointed out in the previous study.[11]

RMP2/6-311+G(3df,2p)//UB3-LYP/6-31G(d) produces an MAD of 14.0 kJ mol^{-1}. This is a small improvement over RB3-LYP. For this method, the deviations from the U-CBS-QB3 values are equally large for the disubstituted radicals and the monosubstituted radicals.

Single-point calculations with the recently formulated BMK functional of Boese and Martin[5] give good results for C–H BDEs. For UBMK, the MAD is 10.3 kJ mol^{-1}, while RBMK gives an MAD of 6.0 kJ mol^{-1}. Considering the modest cost of this method, these results are very encouraging. As with RB3-LYP, larger absolute deviations occur for disubstituted systems compared with monosubstituted systems. It should be noted that this does not necessarily represent a correlation in the error with the size of the molecules involved since some of the disubstituted systems are smaller than the monosubstituted species.

5.3.1.2 Nitrogen-Centered Radicals

Listed in Table 5.2 are N–H BDEs that lead to nitrogen-centered radicals on small models of a peptide backbone containing a glycine residue. We again use U-CBS-QB3 values as a secondary benchmark and compare values obtained with other methods with the U-CBS-QB3 results.

Standard CBS-QB3 shows only a small deviation from the U-CBS-QB3 values, with an MAD of 1.9 kJ mol^{-1}. The largest deviation occurs for the HCO–•N–H radical, which has an $<S^2>$ value of 0.9482.

The BDEs found with the high-level G3X(MP2)-RAD procedure also compare well with the BDEs found with U-CBS-QB3 for nitrogen-centered radicals. The MAD is 4.4 kJ mol^{-1}. In a previous study,[13] it was found that RMP2 did not perform as well for the calculation of N–H BDEs as it did for C–H BDEs. This does not appear to be the case in the present work. The MAD for the RMP2 N–H BDEs is 3.9 kJ mol^{-1}, which is a more satisfactory result than for the current C–H BDEs.

As with the C–H BDEs, the DFT methods that are examined in Table 5.2 are improved by using a restricted reference-wave-function instead of an unrestricted reference-wave-function. For example, the MAD for the N–H BDEs obtained with UB3-LYP is 15.2 kJ mol^{-1} compared with the RB3-LYP value of 9.5 kJ mol^{-1}. The BMK results are also affected in the same way. The N–H BDEs calculated with UBMK have an MAD of 5.8 kJ mol^{-1}, compared with 3.3 kJ mol^{-1} for RBMK. The results found with RBMK for the N–H BDEs are very encouraging, as was also the case for C–H BDEs, especially when considering the modest computational cost of this method. In fact, the performance of RBMK for N–H BDEs is comparable to that of the high level G3X(MP2)-RAD procedure.

Table 5.2: Comparison of Calculated N–H Bond Dissociation Energies (0 K, kJ mol^{-1}) Resulting in the Formation of Nitrogen-Centered Radicals on Small Models of a Glycine Peptide Backbone

radical	UB3-LYP[a]	RB3-LYP[a]	UBMK[a]	RBMK[a]	RMP2[a]	G3X(MP2)-RAD	CBS-QB3	U-CBS-QB3[b]	W1
H–•N–H	432.6	443.2	438.9	443.1	437.6	439.8	444.2	444.4	444.2
HCO–•N–H	461.6	471.6	473.0	483.7	481.1	469.3	470.2	475.2	474.3
CH_3CO–•N–H	449.8	458.3	463.4	472.4	473.0	462.4	463.2	466.7	
NH_2CH_2CO–•N–H	449.4	456.8	469.2	471.5	465.7	462.9	465.1	467.8	
H–•N–CH_3	395.9	401.0	403.2	407.8	408.2	409.4	412.4	412.7	412.2
H–•N–CH_2CHO	401.7	406.5	409.2	413.9	414.2	415.5	418.2	418.5	
H–•N–CH_2CONH_2	397.8	402.5	405.8	410.5	411.0	412.3	415.5	415.8	
HCO–•N–CH_3	434.9	441.6	445.4	453.0	456.0	447.8	447.7	451.3	
CH_3CO–•N–CH_3	426.1	431.9	438.3	444.2	448.2	441.4	442.4	444.6	
HCO–•N–CH_2CHO	461.6	454.0	458.9	465.5	465.9	458.8	461.1	463.4	
CH_3CO–•N–CH_2CHO	438.8	445.4	449.2	454.9	455.6	450.2	455.4	456.4	
CH_3CO–•N–CH_2CONH_2	440.6	446.4	452.1	457.1	456.6	451.3	455.7	456.7	
MAD[c]	15.2	9.5	5.8	3.3	3.9	4.4	1.9		
MD[c]	–15.2	–9.5	–5.6	+0.3	0.0	–4.4	–1.9		
LD[c]	–18.5	–13.3	–10.0	+8.5	–6.8	–6.2	–5.0		

[a] Calculations have been performed in association with the 6-311+G(3df,2p) basis set on UB3-LYP/6-31G(d) optimized geometries, energies have been corrected to 0 K using UB3-LYP/6-31G(d) ZPVEs scaled by 0.9801.[4] [b] Standard CBS-QB3 but with the empirical spin-correction for spin contamination in the unrestricted wave function omitted. [c] Mean absolute deviation (MAD), mean deviation (MD) and largest deviation (LD) from the calculated U-CBS-QB3 results.

Apart from RBMK, the DFT methods examined in Table 5.2 generally underestimate N–H BDEs. This is reflected in the MDs and MADs having the same magnitude. We noted above that the C–H BDEs calculated with DFT procedures for disubstituted systems are not as good as for monosubstituted radicals. In the case of the N–H BDEs, the larger absolute deviations from the benchmark values occur for species that are larger in size (rather than for disubstituted versus monosubstituted).

5.3.2 Radical Stabilization Energies

Tables 5.3 and 5.4 present radical stabilization energies (RSEs) for carbon- and nitrogen-centered radicals, respectively, associated with small models of a glycine-containing peptide backbone. Systematic errors in the calculation of BDEs can cancel in the evaluation of RSEs. This can mean that methods that perform less well in predicting BDEs may still produce acceptable RSEs.

5.3.2.1 Carbon-Centered Radicals

RSEs calculated with U-CBS-QB3 are taken as a secondary benchmark because of the enhanced agreement with W1. Most of the methods displayed in Table 5.3 perform better in determining RSEs than BDEs, as shown by the smaller MADs and MDs in this Table compared with the MADs and MDs in Table 5.1. Standard CBS-QB3 compares well with U-CBS-QB3, with an MAD of 1.2 kJ mol^{-1}.

The G3X(MP2)-RAD method is the only method to have a worse MAD for the RSEs than for the BDEs though, even in this case, the MAD for the RSEs is an acceptable 4.3 kJ mol^{-1}. The origin of the problem in calculating RSEs with G3X(MP2)-RAD lies with the greater-than-expected deviation of this method for the BDE of methane, leading to the observed underestimation of the RSEs. The RMP2 RSEs of Table 5.3 also are quite acceptable, with an MAD of 4.2 kJ mol^{-1}.

Table 5.3: Comparison of Calculated Radical Stabilization Energies (0 K, kJ mol^{-1}) of α-Carbon Radicals on Small Models of a Glycine Peptide Backbone

radical	UB3-LYP[a]	RB3-LYP[a]	UBMK[a]	RBMK[a]	RMP2[a]	G3X(MP2)-RAD	CBS-QB3	U-CBS-QB3[b]	W1
H–•CH–CHO	45.3	41.5	43.6	40.9	32.2	35.4	41.0	36.9	36.6
H–•CH–$CONH_2$	29.8	29.0	27.0	25.5	20.3	21.7	25.6	24.7	
H–•CH–$CONHCH_3$	30.9	30.4	27.7	26.7	20.5	22.0	25.7	25.1	
NH_2–•CH–H	55.3	55.5	55.1	54.7	45.8	44.9	49.6	49.5	49.2
HCONH–•CH–H	53.3	53.0	50.2	49.1	41.2	41.7	44.9	44.5	
CH_3CONH–•CH–H	53.6	53.5	50.7	49.9	42.0	42.3	45.3	45.1	
NH_2–•CH–CHO	142.5	142.3	140.1	137.2	126.2	116.8	128.0	125.8	123.0
NH_2–•CH–$CONH_2$	111.0	111.6	107.8	107.4	90.3	90.0	97.4	97.0	
HCONH–•CH– CHO	120.2	119.7	116.3	112.9	102.5	93.7	102.1	99.2	
HCONH –•CH–$CONH_2$	93.9	94.4	89.3	88.6	76.0	75.6	82.6	82.0	
CH_3CONH–•CH–$CONHCH_3$	96.4	97.0	91.9	91.5	77.7	81.5	83.8	83.5	
MAD[c]	10.8	10.4	7.4	6.5	4.2	4.3	1.2		
MD[c]	+10.8	+10.4	+7.4	+6.9	–3.5	–4.3	+1.2		
LD[c]	+21.0	+20.5	+14.2	+14.6	–6.7	–9.0	+4.1		

[a] Calculations have been performed in association with the 6-311+G(3df,2p) basis set on UB3-LYP/6-31G(d) optimized geometries. Energies have been corrected to 0 K using UB3-LYP/6-31G(d) ZPVEs scaled by 0.9801.[4] [b] Standard CBS-QB3 but with the empirical spin-correction for spin contamination in the unrestricted wave function omitted. [c] Mean absolute deviation (MAD), mean deviation (MD) and largest deviation (LD) from the calculated U-CBS-QB3 results.

Table 5.4: Comparison of Calculated Radical Stabilization Energies (0 K, kJ mol^{-1}) of Nitrogen-Centered Radicals on Small Models of a Glycine Peptide Backbone

radical	UB3-LYP[a]	RB3-LYP[a]	UBMK[a]	RBMK[a]	RMP2[a]	G3X(MP2)-RAD	CBS-QB3	U-CBS-QB3[b]	W1
HCO–•N–H	–29.0	–28.4	-34.1	-40.6	–43.5	–29.5	–26.0	–30.8	–30.1
CH_3CO–•N–H	–17.2	–15.1	-24.5	-29.3	–35.4	–22.6	–19.0	–22.3	
NH_2CH_2CO–•N–H	–16.8	–13.6	-30.3	-28.4	–28.1	–23.1	–20.9	–23.4	
H–•N–CH_3	36.7	42.2	35.7	35.3	29.4	30.4	31.8	31.7	32.0
H–•N–CH_2CHO	30.9	36.7	29.7	29.2	23.4	24.3	26.0	25.9	
H–•N–CH_2CONH_2	34.8	40.7	33.1	32.6	26.6	27.5	28.7	28.6	
HCO–•N–CH_3	–2.3	1.6	-6.5	-9.9	–18.4	–8.0	–3.5	–6.9	
CH_3CO–•N–CH_3	6.5	11.3	0.6	-1.1	–10.6	–1.6	1.8	–0.2	
HCO–•N–CH_2CHO	–29.0	–10.8	-20.0	-22.4	–28.3	–19.0	–16.9	–19.0	
CH_3CO–•N–CH_2CHO	–6.2	–2.2	-10.3	-11.8	–18.0	–10.4	–11.2	–12.0	
CH_3CO–•N–CH_2CONH_2	–8.0	–3.2	-13.2	-14.0	–19.0	–11.5	–11.5	–12.3	
MAD[c]	5.5	9.1	2.7	3.8	7.4	1.0	1.8		
MD[c]	+3.7	+9.1	+0.1	–1.8	–7.4	–0.3	+1.8		
LD[c]	–10.6	+12.1	–6.9	–9.8	–13.1	–1.6	+4.8		

[a] Calculations have been performed in association with the 6-311+G(3df,2p) basis set on UB3-LYP/6-31G(d) optimized geometries. Energies have been corrected to 0 K using UB3-LYP/6-31G(d) ZPVE scaled by 0.9802.[4] [b] Standard CBS-QB3 but with the empirical spin-correction for spin contamination in the unrestricted wave function omitted. [c] Mean absolute deviation (MAD), mean deviation (MD) and largest deviation (LD) from the calculated U-CBS-QB3 results.

Of the DFT methods, RBMK performs the best with an MAD of 6.9 kJ mol^{-1}. Using UBMK makes the agreement less good for the RSEs (MAD = 7.4 kJ mol^{-1}) but the effect is not as pronounced as for the BDEs where the MAD is almost doubled. Both RBMK and UBMK systematically overestimate the RSEs, resulting in MADs and MDs that are the same in magnitude. These effects are also seen with RB3-LYP and UB3-LYP where the MADs are 10.4 and 10.8 kJ mol^{-1}, respectively.

The U-CBS-QB3 results in Table 5.3 show that the carbonyl-substituted radicals have RSEs between 24.7 and 36.9 kJ mol^{-1}. The origin of the stabilization of a carbon radical center by a CX=O group has been explained in Chapter 1, section 1.3.2. Briefly, the π and π^* orbitals on the carbonyl can interact with the singly-occupied 2p carbon orbital, 2p(C•), producing a net stabilization. The smaller the energy separation between the 2p(C•) orbital on the carbon and the π^* orbital on the carbonyl, the greater the stabilizing effect.

The formyl group (CH=O) has a greater stabilizing effect on the radical than a formamidyl group ($C{=}ONH_2$), as reflected in the U-CBS-QB3 RSEs of 36.9 and 24.7 kJ mol^{-1} for H–•CH–CHO and H–•CH–$CONH_2$, respectively. This arises because the NH_2 substituent decreases the π-accepting ability of the CH=O group. Further replacement of an amino hydrogen on the formamidyl group with a methyl group, to give the H–•CH–$CONHCH_3$ radical, has only a minor further effect on stabilization, evident in the small change in the RSE calculated with U-CBS-QB3 from 24.7 kJ mol^{-1} to 25.1 kJ mol^{-1} for H–•CH–$CONH_2$ and H–•CH–$CONHCH_3$, respectively. This suggests that the stability of a monosubstituted α-carbon radical will not generally be significantly affected by substituents beyond the amide bond, and supports previous conclusions to this effect.[1]

Amine (NHY) functionalities also stabilize carbon-centered radicals. The calculated U-CBS-QB3 RSE for aminomethyl radical (NH_2–•CH–H) is 49.5 kJ mol^{-1}. The large stabilizing effect arises from the interaction of the nitrogen lone-pair orbital with the 2p(C•) orbital on the carbon.

Formally replacing an amino hydrogen with a formyl group results in the HCONH–•CH–H radical. The calculated U-CBS-QB3 RSE for this radical is 44.5 kJ mol^{-1}. This amounts to a change in the RSE of 5.0 kJ mol^{-1} in going from NH_2–•CH–H to CHONH–•CH–H, resulting from a decreased π-electron-donating ability of CHONH compared with NH_2. This change is relatively minor compared with the change in the RSE of the carbonyl-substituted radicals, H–•CH–CHO and H–•CH–$CONH_2$, of 12.2 kJ mol^{-1}. Thus, formation of an amide bond affects the π-donating ability of the nitrogen lone pair in this environment significantly less than the π-accepting ability of the carbonyl. Although there are many factors involved in the stabilization of each radical, calculated spin densities at the radical center are consistent with this result. The UB3-LYP/6-31G(d) spin densities on the carbon formally carrying the unpaired electron are +0.903 (NH_2–•CH–H), +0.904 (CHONH–•CH–H), +0.864 (H–•CH–CHO) and +0.983 (H–•CH–$CONH_2$). These results show that there is only a small change in spin density accompanying formyl substitution in NH_2–•CH–H, but there is a larger positive change accompanying amino substitution in H–•CH–CHO.

An additional methyl substituent, giving the CH_3CONH–•CH–H radical, results in a calculated U-CBS-QB3 RSE of 45.1 kJ mol^{-1}, a change of just 0.6 kJ mol^{-1} from the RSE of CHONH–•CH–H. As with the situation for the carbonyl-substituted radicals, substitution beyond the initial amide bond has little effect on the RSE for monosubstituted α-carbon radicals, again supporting previous conclusions.[1]

The calculated U-CBS-QB3 RSEs for the disubstituted carbon-centered radicals range between 82.0 and 125.8 kJ mol^{-1}, indicating much larger stabilizing effects than observed in the monosubstituted radicals (Table 5.5). The simplest disubstituted radical (NH_2–•CH–CHO) has a calculated U-CBS-QB3 RSE of 125.8 kJ mol^{-1}, which is the largest value in the table. The deviation from additivity evaluated with equation (5.1),

Table 5.5: Comparison of Radical Stabilization Energies of Carbon-Centered Radicals (0 K, kJ mol–1) Calculated with U-CBS-QB[3a] for Disubstituted Radicals (X– •CH–Y) with Values for the Corresponding Monosubstituted Radicals (X–•CH–H and H– •CH–Y)

X–•CH–Y	X–•CH–H	H–•CH–Y	X–•CH–Y	Δ(RSE)[b]
NH_2–•CH–CHO	49.5	36.9	125.8	39.4
NH_2–•CH–$CONH_2$	49.5	24.7	97.0	15.7
HCONH–•CH–CHO	44.5	36.9	99.2	17.8
HCONH–•CH–$CONH_2$	44.5	24.7	82.0	12.8
CH_3CONH–•CH–$CONHCH_3$	45.1	25.1	83.5	13.3

[a] Standard CBS-QB3 but with the empirical spin-correction for spin-contamination in the unrestricted wave function omitted. [b] The estimated synergistic stabilization (captodative effect) of two groups working in combination, measured by taking the difference of the RSE calculated for the disubstituted radical with the sum of the RSEs calculated for the analogous monosubstituted radicals, i.e., RSE(X–•CH–Y) – RSE(X–•CH–H) – RSE(H–•CH–Y).

given in the final column of Table 5.5, is 39.4 kJ mol^{-1}. For the monosubstituted radicals a change from H–•CH–CHO to H–•CH–$CONH_2$ decreases the RSE by 15.3 kJ mol^{-1}. For the equivalent change in the disubstituted radicals, resulting in NH_2–•CH–$CONH_2$, the RSE decreases by 28.8 kJ mol^{-1}. This indicates that not only is there a decrease in stability arising from the poorer π-accepting ability of the C=ONH_2 group but there is also a decrease in the synergistic interaction, as demonstrated by the reduced deviation from additivity from 39.4 to 15.7 kJ mol^{-1}.

Formyl substitution in NH_2–•CH–CHO leads to CHONH–•CH–CHO which has a U-CBS-QB3 RSE of 99.2 kJ mol^{-1}. This represents a decrease of 26.6 kJ mol^{-1}. The deviation from additivity in this case is 17.8 kJ mol^{-1}. Thus, the decreased π-donating ability of the nitrogen lone-pair associated with formyl substitution also causes the loss of approximately half the synergistic stabilization.

Further losses in synergistic stabilization for radicals derived from molecules beyond the first amide bond are small, as reflected in the Δ(RSE) for HCONH–•CH–$CONH_2$ (12.8 kJ mol^{-1}) and CH_3CONH–•CH–$CONHCH_3$ (13.3 kJ mol^{-1}).

5.3.2.2 Nitrogen-Centered Radicals

The RSEs for nitrogen-centered radicals in a model peptide backbone are presented in Table 5.4.

The standard CBS-QB3 RSEs compare well with the values calculated with U-CBS-QB3, with an MAD of just 1.8 kJ mol^{-1}. This is expected because of the small size of the empirical spin-correction term.

The RSEs calculated with G3X(MP2)-RAD are also in good agreement with the RSEs calculated with U-CBS-QB3, with an MAD of 1.0 kJ mol^{-1}. In the case of nitrogen-centered radicals, the RSEs do not suffer the problem encountered with the RSEs of carbon-centered radicals, i.e., there is a smaller error in the calculated BDE of ammonia compared with the error in the calculated BDE of methane. RMP2 performs reasonably well for the calculation of RSEs of nitrogen-centered radicals, with an MAD of 7.4 kJ mol^{-1}.

The DFT methods also perform reasonably well. However, the previously observed trend in which the restricted reference-wavefuction improves the calculated BDEs compared with those obtained with an unrestricted reference-wave-function is reversed. For example, the MAD for RB3-LYP is 9.1 kJ mol^{-1}, compared with 5.5 kJ mol^{-1} for UB3-LYP. The trend is the same for the BMK functional, with MADs of 3.8 kJ mol^{-1} (RBMK) and 2.7 kJ mol^{-1} (UBMK).

Table 5.6 compares RSEs calculated with the U-CBS-QB3 procedure for disubstituted nitrogen-centered radicals and their monosubstituted analogues. The monosubstituted radicals of this Table can be grouped according to whether a CX=O group or a CH_2Z group is attached to the radical center. For the CH_2Z substituents, the RSEs range between 25.9 and 31.7 kJ mol^{-1}, indicating that they are stabilizing substituents. On the other hand, the RSEs for the carbonyl (CX=O) substituents range between –22.3 and –30.8 kJ mol^{-1} representing a destabilizing effect for this substituent compared with their effect in the closed-shell parents. The reason for the

Table 5.6: Comparison of Radical Stabilization Energies of Nitrogen-Centered Radicals (0 K, kJ mol^{-1}) Calculated with U-CBS-QB3[a] for Disubstituted Radicals (X– •CH–Y) with Values for the Corresponding Monosubstituted Radicals (X– •CH–H and H– •CH–Y)

X– •N–Y	X–•N–H	H–•N–Y	X–•N–Y	Δ(RSE)[b]
$HCO–•N–CH_3$	–30.8	31.7	–6.9	–7.8
$CH_3CO–•N–CH_3$	–22.3	31.7	–0.2	–9.6
$HCO–•N–CH_2CHO$	–30.8	25.9	–19.0	–14.1
$CH_3CO–•N–CH_2CHO$	–22.3	25.9	–12.0	–15.6
$CH_3CO–•N–CH_2CONH_2$	–22.3	28.6	–12.3	–18.6

[a] Standard CBS-QB3 but with the empirical spin-correction for spin-contamination in the unrestricted wave function omitted. [b] The calculated deviation from additivity of two groups working in combination, measured by taking the difference of the RSE calculated for the disubstituted radical with the sum of the RSEs calculated for the analogous monosubstituted radicals, i.e., RSE(X–•CH–Y) – RSE(X–•CH–H) – RSE(H–•CH–Y).

stabilization in the case of the methyl group and the destabilization in the case of the carbonyl group has been explained in detail elsewhere.[13,14] Briefly, stabilization of the radical by CH_2Z can occur through a hyperconjugative interaction, which is not present to the same extent in the parent closed-shell species. For the carbonyl substituents, the observed destabilization occurs because in the closed-shell parent species it is the nitrogen lone pair that is delocalized, whereas in the radical, it is the single unpaired electron that interacts with the π* orbital. In both cases, the stabilizing or destabilizing effect of the substituent is attenuated as the chain becomes longer.

For disubstituted nitrogen-centered radicals, the RSEs range between –19.0 and –0.2 kJ mol^{-1}, indicating a net destabilization in all cases. This contrasts with the behavior of the carbon-centered radicals, which benefit through a captodative stabilization from disubstitution. For example, the RSE of $H–•N–CH_3$ is calculated to be 31.7 kJ mol^{-1} and the calculated RSE of HCO–•N–H is –30.8 kJ mol^{-1}, whereas the

calculated RSE of HCO–•N–CH_3 is –6.9 kJ mol^{-1}. If the effect of the two substituents were exactly additive, then the RSE of HCO–•N–CH_3 would be 0.9 kJ mol^{-1}. Similar behavior is seen for the larger nitrogen-centered radicals where the Δ(RSE) values range from –18.6 to –14.1 kJ mol^{-1}. These negative values possibly reflect additional stabilization in the disubstituted closed-shell parent species. In order to distinguish between these two possibilities, we have calculated the enthalpies of reactions 5.2 and 5.3:

$$X–NH–Y + NH_3 \rightarrow X–NH_2 + Y–NH_2 \quad (5.2)$$

$$X–N•–Y + •NH_2 \rightarrow X–•NH + Y–•NH \quad (5.3)$$

These provide a measure of the respective stabilizing effects of disubstitution in the closed-shell parent species and the analogous N-centered radicals. We find that disubstitution has a greater stabilizing effect in the closed-shell parent species, and this in turn leads to the observed negative Δ(RSE) values (Table 5.7).

Table 5.7: Comparison of Enthalpies of Reactions Measuring Stabilizing Interactions at N and N• (U-CBS-QB3,[a] 0 K, kJ mol^{-1})

X–(N)–Y[b]	X–NH–Y[c]	X–•N–Y[d]	ΔΔH[e]
HCO–(N)–CH_3	25.5	17.6	–7.8
CH_3CO–(N)–CH_3	25.9	16.4	–9.6
HCO–(N)–CH_2CHO	25.3	11.2	–14.1
CH_3CO–(N)–CH_2CHO	28.2	12.7	–15.6
CH_3CO–(N)–CH_2CONH$_2$	31.2	12.5	–18.6

[a] Standard CBS-QB3 but with the empirical spin-correction for spin-contamination in the unrestricted wave function omitted. [b] (N) = NH or •N. [c] Enthalpy change for X–NH–Y + NH_3 → X–NH_2 + Y–NH_2. [d] Enthalpy change for X–•N–Y + •NH_2 → X–•NH + Y–•NH. [e] Equivalent to Δ(RSE).

5.4 Concluding Remarks

Bond dissociation energies (BDEs) and radical stabilization energies (RSEs) have been calculated for a series of molecules that represent the backbone of a small glycine-containing peptide. The resultant radicals are either carbon-centered, at the α-carbon, or nitrogen-centered, on the backbone nitrogens. We find both for BDEs and RSEs that the high-level theoretical procedures G3X(MP2)-RAD, CBS-QB3 and U-CBS-QB3 give results that are in good agreement with one another, and with W1 values in the few cases where the latter are available.

Of the less computationally demanding methods, RBMK/6-311+G(3df,2p), RMPWB1K/6-311+G(3df,2p) and RMP2/6-311+G(3df,2p), carried out with UB3-LYP/6-31(d) geometries, perform well for BDEs. For RSEs, the restricted and unrestricted forms of BMK/6-311+G(3df,2p)//UB3-LYP/6-31G(d) and MPWB1K/6-311+G(3df,2p)//UB3-LYP/6-31G(d) give results in very good agreement with the U-CBS-QB3 values.

Positive RSEs (indicating a relative stabilization) are found for monosubstituted carbon-centered radicals for both amine (NHY) and carbonyl (CX=O) substituents, with the NHY substituents being the more effective stabilizing groups. The disubstituted radicals are further stabilized through a large captodative interaction.

For monosubstituted nitrogen-centered radicals, we find positive RSEs for methyl and related (CH_2Z) substituents, while we find negative RSEs for the CX=O substituents. An additional destabilization occurs for the disubstituted radicals

5.5 References

(1) Armstrong, D. A.; Yu, D.; Rauk, A. *Can. J. Chem.* **1996**, *74*, 1192.

(2) Croft, A. K.; Easton, C. J.; Radom, L. *J. Am. Chem. Soc.* **2003**, *125*, 4119.

(3) Rauk, A.; Yu, D.; Taylor, J.; Shustov, G. V.; Block, D. A.; Armstrong, D. A. *Biochemistry* **1999**, *38*, 9089.

(4) Scott, A. P.; Radom, L. *J. Phys. Chem.* **1996**, *100*, 16502.

(5) Boese, D. A.; Martin, J. M. L. *J. Chem. Phys.* **2004**, *121*, 3405.

(6) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6908.

(7) Henry, D. J.; Sullivan, M. B.; Radom, L. *J. Chem. Phys.* **2003**, *118*, 4849.

(8) Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1999**, *110*, 2822.

(9) Martin, J. M. L.; De Oliveira, G. J. *J. Chem. Phys.* **1999**, *111*, 1843.

(10) Parthiban, S.; Martin, J. M. L. *J. Chem. Phys.* **2001**, *114*, 6014.

(11) See, for example: Henry, D. J.; Parkinson, C. J.; Mayer, P. M.; Radom, L *J. Phys. Chem. A* **2001**, *105*, 6750.

(12) Henry, D. J.; Parkinson, C. J.; Radom, L *J. Phys. Chem. A* **2002**, *106*, 7927.

(13) Wood, G. P. F.; Henry, D. J.; Radom, L. *J. Phys. Chem. A* **2003**, *107*, 7985

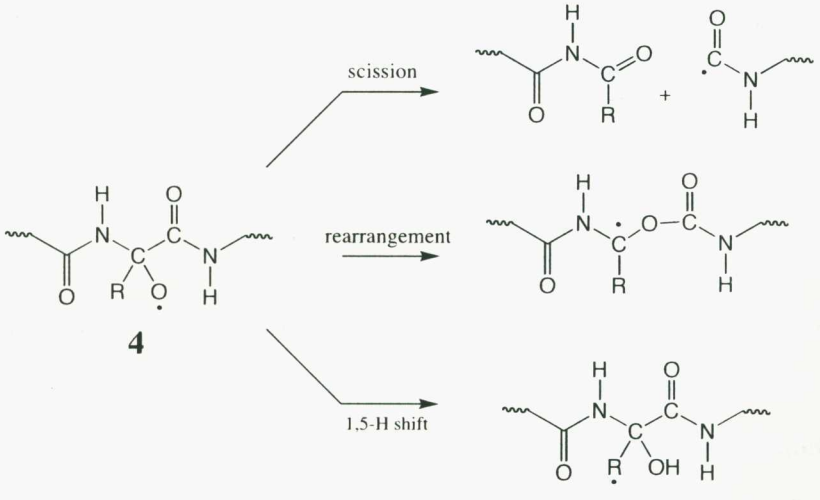
(14) Song, K. S.; Cheng, Y. H.; Fu, Y.; Liu, L.; Li, X. S.; Guo, Q. X. . *J. Phys. Chem. A* **2002**, *106*, 6651.

CHAPTER 6

Modeling β-Scission Reactions of Peptide Backbone Alkoxyl Radicals: Backbone C–C Bond Fission

6.1 Introduction

Alkoxyl radicals are a particular class of short-lived reactive oxygen species (ROSs) that have been postulated to form through initial hydrogen abstractions on alkyl side chains or on the α-carbon of an amino acid residue (see Chapter 1). Their involvement in radical reactions on proteins has been determined through the detection and characterization of hydroperoxides,[1] from which they can be formed via a tetroxide or via one-electron reduction reactions. It has been established, both through experimental atmospheric and solution chemistry studies and through theoretical computations, that alkoxyl radicals can *inter alia* undergo three classes of reactions, namely β-scission, skeletal rearrangements, and 1,5-hydrogen shifts.[1b,2] It might be expected that peptide backbone alkoxyl radicals will undergo the same types of reactions, as shown in reactions 6.1, 6.2 and 6.3, respectively.



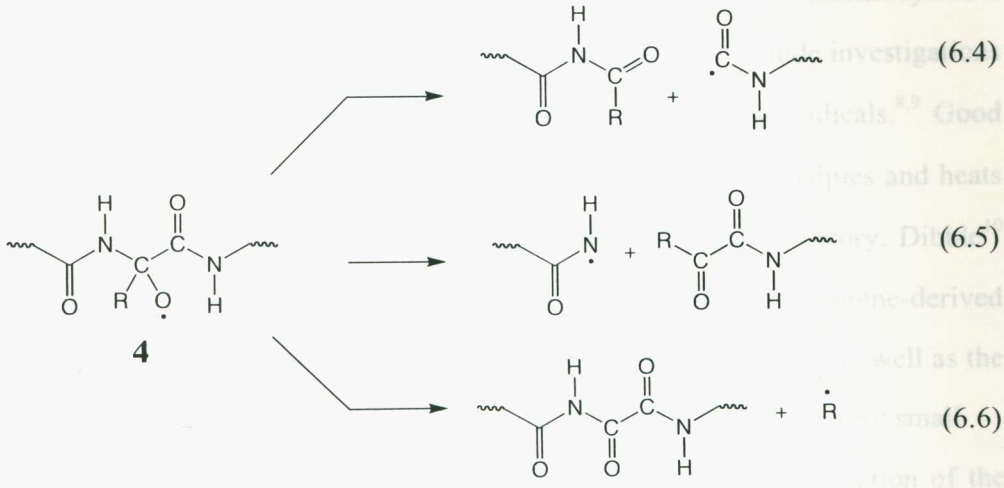
(6.1)

(6.2)

(6.3)

Experimentally it is difficult to obtain quantitative information on the individual reactions, though some experimental evidence has been obtained for the occurrence of reaction (6.1).[1b,3] This is because the reactions are fast and can involve chain processes, which makes it problematic to isolate the particular reaction of interest. An additional

degree of complexity arises because each of the three classes of reactions can in principle undergo three variations. For example, because there are three different groups joined to the α-carbon, reaction (6.1) is just one of three possible β-scission reactions that can occur, namely β-scission of the C–C bond (6.1, repeated in 6.4), β-scission of the C–N bond (6.5), or β-scission of the side-chain C–R bond (6.6):



(6.4)

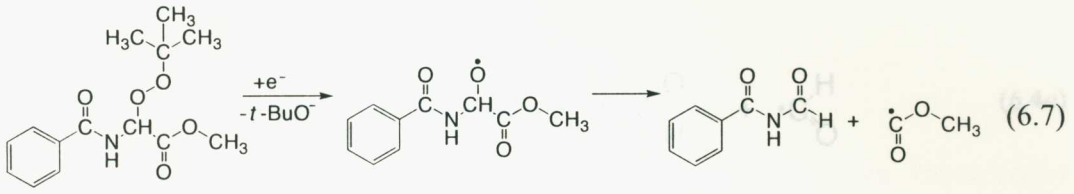
(6.5)

(6.6)

4

Previous EPR spin-trapping studies have suggested that the process (6.4) predominates, though the occurrence of other reactions could not be discounted.[1b] This was confirmed recently for a Gly-containing peptide through product studies.[3] The investigations showed that the α-alkoxyglycyl radical (**6**) derived from α-tert-butylperoxy-N-benzoylglycine methyl ester (**5**), through either photolysis or reduction with tris-(triphenylphosphine) dichlororuthenium, underwent β-scission of the α-carbon-carbonyl bond to give N-formylbenzamide (**7**) in good yield,[3] see equation 6.7.

In general, alkoxyl cleavage reactions have been found to be endothermic,[4] and the preferred pathway is generally determined by the stability of the radical that is formed.[5]



(6.7)

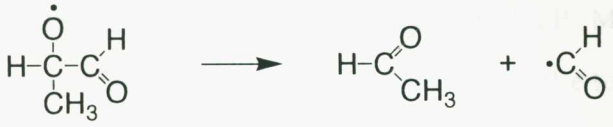
5 **6** **7**

However, this is not always the case and for one particular class of ring-opening reactions,[6] that has been studied computationally,[7] the most stable thermodynamic products are not formed. Other computational studies in this area include investigations of the associated β-scission reactions of a general series of alkoxyl radicals.[8,9] Good agreement between theory and experiment was found for reaction enthalpies and heats of formation at the B3-LYP/SVP, CBS-RAD and G2(MP2) levels of theory. Dibble[10] has also carried out theoretical studies on cleavage reactions of isoprene-derived alkoxyl radicals. Using B3-LYP with small and medium-sized basis sets, as well as the CBS-4 and CBS-q model chemistries, he found that the cleavage barriers were small.

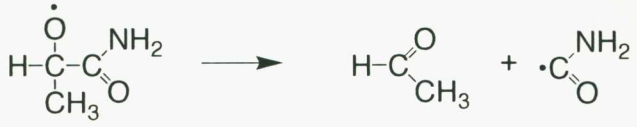
In a recent study, Huang and Rauk[11] performed an initial investigation of the three possible fragmentation reactions 6.4, 6.5 and 6.6. In that study, the B3-LYP/6-31G(d) method was used to determine free energies of fragmentation for model systems containing an alanine and glycine residue. In the present Chapter, we examine in more detail the pathway leading to C–C bond fission (6.4) using four models (6.4*a*–6.4*d*). We use both high-level *ab initio* procedures and simpler calculations to investigate their reaction enthalpies and barriers.

The first reaction, 6.4*a*, represents the smallest system that can be used to model reaction 6.4. It is clearly a very simplified model but it allows the assessment of a wide range of theoretical procedures. At the other extreme, reaction 6.4*d* represents a more realistic model, but because of the increased size it is more restricted in the range of theoretical procedures that can be applied.

(6.4*a*)



(6.4*b*)



(6.4*c*)

(6.4*d*)

We also make high-level predictions of the heats of formation at 0 K for the alkoxyl radicals and for the cleavage products of reactions 6.4*a*-6.4*d*. Heats of formation for some of these species have been previously determined experimentally and computationally, and we use the new calculated values in these cases to further assess the performance of the various theoretical techniques. In other cases, we predict heats of formation for species for which the values are not currently known. For example, reactions 6.4*c* and 6.4*d* involve substituted alkoxyl radicals that have not been studied theoretically before, and for which no thermodynamical data are currently available. On the other hand, the radical product $NH_2C(\bullet)=O$, formed in the cleavage reaction 6.4*c*, has been previously studied theoretically in the context of combustion

chemistry[12] and also in the examination of N–H bond energies.[13] However, an accurate heat of formation is lacking.

6.2 Theoretical Procedures

A number of levels of theory were used to optimize geometries. These include the HF, MP2, and QCISD ab initio methods, and the B3-LYP, MPW1K and BMK density functional theory methods, using a range of basis sets. We found that, for some of the alkoxyl radicals, the lowest-energy conformer varies between optimization levels. To resolve these discrepancies, URCCSD(T) single-point calculations were carried out on the structures optimized at the various levels, with the lowest energy indicating a preferred structure.

A number of levels of theory were also used to determine the energy profiles for reactions 6.4*a*-6.4*d*. These include simpler procedures such as HF, MP2, QCISD, MPW1K, BMK and B3-LYP with a variety of basis sets. Energies at 0 K were obtained by incorporating zero-point vibrational energies (ZPVEs) using appropriate scale factors.[14] In cases for which a scale factor is not available, these were taken from a closely related method. Specifically, for HF and MP2 in association with the 6-311++G(3df,2p) and 6-31G(2df,p) basis sets, the corresponding scale factors from the 6-311(d,p) basis set were used. In the case of B3-LYP in association with the 6-311+G(3df,2p) and cc-pVTZ basis sets, the 6-31G(2df,p) scale factor[15] was used, and in the case of MPWIK/6-31+G(3df,2p), the MPWIK/6-31+G(d,p) scale factor[16] was used. For the QCISD/6-31G(d) energy profiles, we used scaled UB3-LYP/6-31G(d) ZPVEs.

We also examined the performance of simple methods in which the energies and geometries were determined at different levels of theory. These include UB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d), RB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d), UBMK/6-311+G(3df,2p)//UB3-LYP/6-31G(d), RBMK/6-311+G(3df,2p)//UB3-LYP/6-31G(d), UBMK/6-311+G(3df,2p)//UBMK/6-31G(d),

RBMK/6-311+G(3df,2p)//UBMK/6-31G(d), UMP2/6-311+G(3df,2p)//UB3-LYP/6-31G(d) and RMP2/6-311+G(3df,2p)//UB3-LYP/6-31G(d), corrected in each case with either scaled UB3-LYP/6-31G(d) ZPVEs or unscaled UBMK/6-31G(d) ZPVEs where appropriate.

Finally, we used high-level composite methods such as CBS-QB3[17] and U-CBS-QB3[18] from the CBS family of methods, G3(MP2)//B3-LYP,[19] G3X(MP2)-RAD,[20] G3-RAD[20] and G3//B3-LYP[19] from the G3 group of methods, and also the W1 method of Martin et al.[21]

6.3 Results and Discussion

6.3.1 Determination of Preferred Conformations

Our initial task was to determine the preferred conformations of the alkoxyl radicals, the cleavage products and the transition structures connecting reactants to products in each of the reactions 6.4*a* to 6.4*d*. These were initially screened with geometry optimizations at the B3-LYP/6-31G(d) level of theory. Further optimizations were carried out with B3-LYP and the larger 6-311+G(3df,2p) basis set and with the HF and MP2 methods in conjunction with the small and large basis sets. In order to examine the sensitivity of the relative energies of alternative possible conformations to the level of geometry optimization, single-point URCCSD(T) calculations were carried out on the various structures. The calculated relative energies are presented in Table 6.1. Based on the results of this analysis, discussed in more detail below, preferred structures for the alkoxyl radicals (Figure 6.1, **a1**–**a4**), fragmentation products (Figure 6.2, **fp1**–**fp5)** and fragmentation (addition) transition structures (Figure 6.1, **ts1**–**ts4**) were identified. Figures 6.1 and 6.2 also include some of the more interesting structural parameters obtained at the B3-LYP/6-311+G(3df,2p) level of theory.

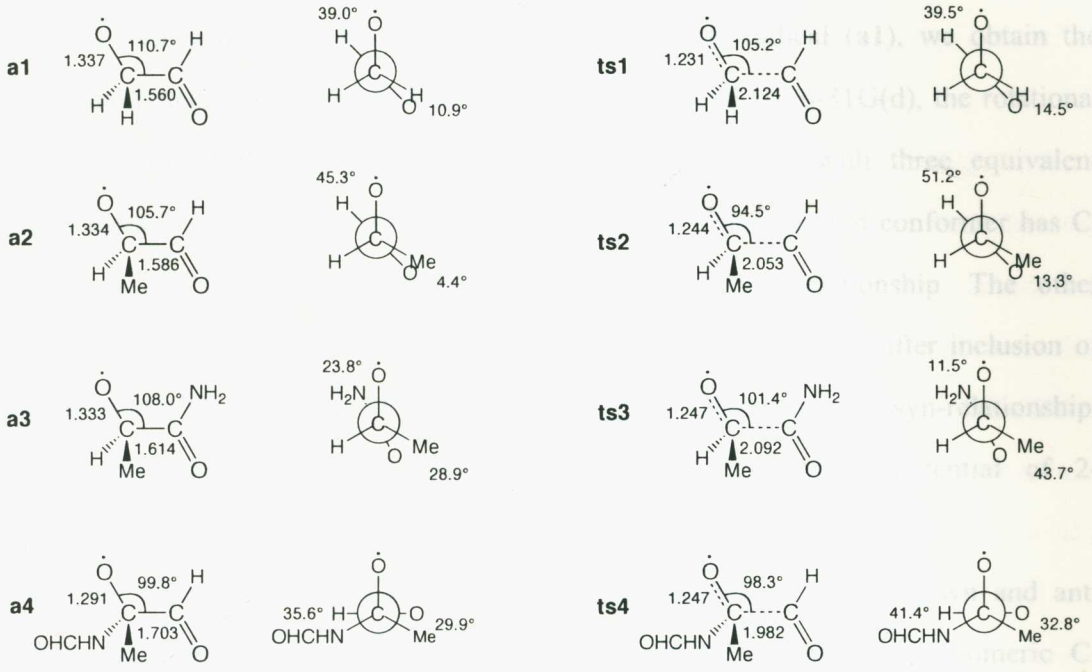


Figure 6.1: Optimized B3-LYP/6-311+G(3df,2p) structures of preferred conformers of the alkoxyl radicals (**a1–a4**) and their associated transition structures (**ts1–ts4**).

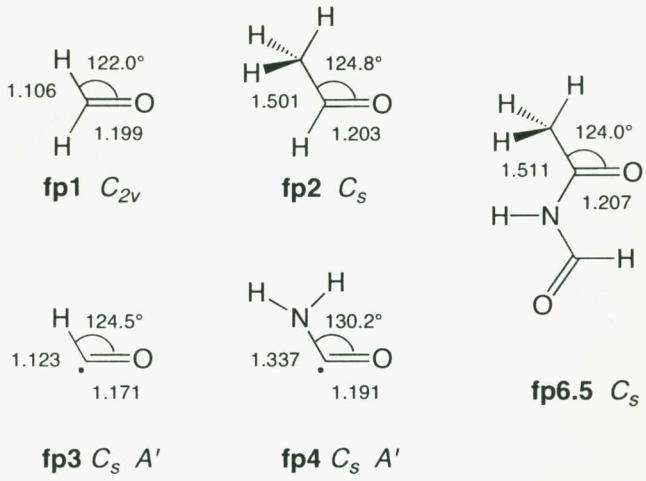


Figure 6.2: Optimized B3-LYP/6-311+G(3df,2p) structures of preferred conformers of the fragmentation products **fp1–fp5** of reactions 6.4*a*–6.4*d*.

We begin our discussion by noting that acetaldehyde ($CH_3CH=O$) is well known to have three equivalent conformational minima. Each of these conformers has a C–H bond eclipsing the C=O bond. If one of the methyl hydrogens of acetaldehyde is formally replaced by oxygen to give the $•OCH_2CH=O$ radical (**a1**), we obtain the rotational potential of Figure 6.3. It can be seen that with HF/6-31G(d), the rotational potential around the central C–C bond changes from one with three equivalent conformers to one with just two distinct conformers. The preferred conformer has C_s symmetry (A" state) with the oxygen atoms in an anti-relationship. The other conformer, which is approximately 12.5 kJ mol^{-1} higher in energy after inclusion of ZPVE, is also C_s symmetric (A" state) but with the oxygen atoms in a syn-relationship. This behavior bears a qualitative similarity to the rotational potential of 2-fluoroacetaldehyde.[22]

At the B3-LYP/6-31G(d) level, there is a distortion from the syn and anti structures. In total, four stable conformers are found, a pair of enantiomeric C_1 structures that have OCCO dihedral angles of 145.6 and 214.4° (C–O• bond approximately anti to the C=O bond), respectively, and another enantiomeric pair that

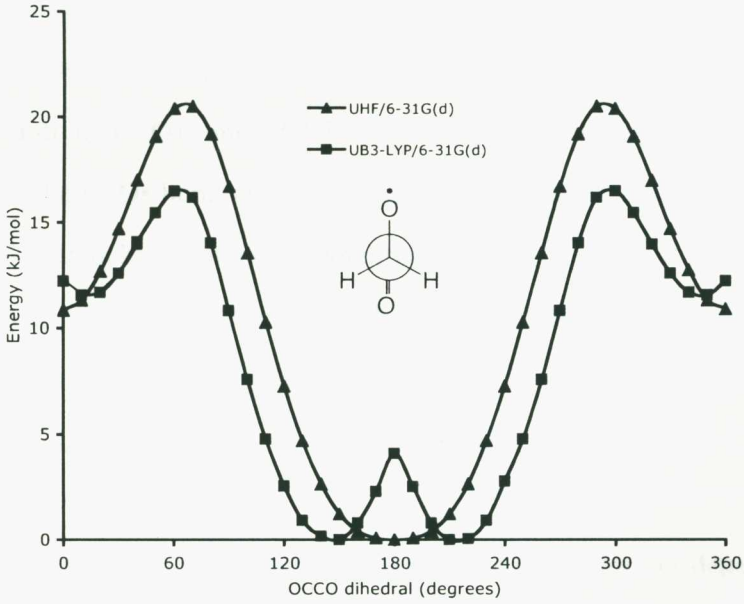


Figure 6.3: Rotational potential about the central C–C bond of the $O=CH–CH_2O•$ radical (**a1**) with HF/6-31G(d) and B3-LYP/6-31G(d).

are 10 kJ mol^{-1} higher in energy, with dihedral angels of 13.3 and 346.7° (C–O• bond approximately syn to the C=O bond). Improving the basis set in the B3-LYP calculations to 6-311+G(3df,2p) has a relatively small effect on the rotational potential.

The B3-LYP rotational potential of Figure 6.3 suggests that alkoxyl radical **a1** suffers from second-order Jahn-Teller (SOJT) effects.[23] It has been noted[23c] that density functional theory and coupled-cluster theory can reliably treat SOJT effects, while Hartree-Fock and MP theory do not. In the context of the present work, because of the possibility of SOJT effects, caution will be exercised when considering vibrational frequencies. We find that the energy differences between conformers that are distorted through SOJT effects are quite small. For example, at the B3-LYP/6-31G(d) level, the difference in energy between the constrained C_s (A″ state) conformer with C–O• anti to C=O (which has one imaginary frequency) and the energetically lower C_1 structure (no imaginary frequencies) is just 1.7 kJ mol^{-1}.

The •$OCH(CH_3)CH=O$ radical (**a2**) is obtained formally by replacing one of the oxy-carbon hydrogens of **a1** with a methyl group. The possible symmetric structures obtained with HF/6-31G(d) for **a1** cannot be achieved by alkoxyl radical **a2**. At this level, three conformers are found for **a2** with OCCO dihedral angles of 5.3° (C–O• bond approximately eclipsing C=O bond), 166.2° (a distortion of the structure with the C–O• bond anti to C=O), and 202.3° (a distortion in the opposite direction of the structure with the C–O• bond anti to C=O), which can be interpreted as a modification of the fluoro-acetaldehyde rotational potential.

The conformers with OCCO dihedral angles of 166.2° ($\mathbf{a2_a}$) and 202.3° ($\mathbf{a2_b}$) at HF/6-31G(d) were reoptimized at a number of levels of theory. URCCSD(T)/6-311+G(3df,2p) single points were then carried out on a selection of the optimized geometries.

The results in Table 6.1 show that the preferred conformer depends on the level of theory, but the energy differences are small. We have taken our best level for

Table 6.1: Differences in Energy (kJ mol^{-1}) Between Conformers of Alkoxyl Radicals **a2**, and **a4** Calculated at Various Levels of Theory[a]

Level of Theory	$\mathbf{a2_a - a2_b}$	$\mathbf{a4_a - a4_b}$	$\mathbf{a4_a - a4_c}$
HF/6-31G(d)	–0.7	–0.4	9.6
HF/6-311++G(3df,2p)	0.6	0.8	9.5
MP2/6-31G(d)	–0.9	–1.1	4.9
MP2/6-31+G(d)	–1.0	–	–
MP2/6-31+G(2df,p)	2.4	–	–
B3-LYP/6-31G(d)	–3.0	–2.1	–10.2
B3-LYP/6-311+G(3df,2p)	–3.0	–1.4	–12.6
//HF/6-31G(d)	–0.4[b]	2.1[c]	1.5[c]
//MP2/6-31+G(2df,p)	0.9[b]	–	–
//B3-LYP/6-31G(d)	0.6[b]	–0.7[c]	–5.2[c]
//B3-LYP/6-311+G(3df,2p)	–0.7[b]	–1.3	–8.2[c]

[a] Energy differences are not corrected for zero-point vibrational energy. [b] URCCSD(T)/6-311+G(3df,2p) calculations on the indicated geometry. [c] URCCSD(T)/6-311+G(d,p) calculations on the indicated geometry.

geometry optimization throughout this Chapter to be B3-LYP/6-311+G(3df,2p) and the conformation that gives the lowest coupled-cluster energy at this level was therefore taken to be the preferred structure. For •$OCH(CH_3)CH{=}O$, $\mathbf{a2_a}$ is thus our predicted structure. At the B3-LYP/6-311+G(3df,2p) level, the OCCO dihedral angle in $\mathbf{a2_a}$ is 134.2°.

For •$OCH(CH_3)C(NH_2){=}O$ (**a3**), analogies with fluoroacetaldehyde are not apparent. In this case, only one conformation of significance was found (Figure 6.1) and so no further analysis was carried out.

For the •$OC(CH_3)(NHCH{=}O)CH{=}O$ radical (**a4**), three conformers were examined in detail, namely $\mathbf{a4_a}$, which has the C=O bond approximately eclipsing the C–CH_3 bond, and $\mathbf{a4_b}$, and $\mathbf{a4_c}$, both of which have the C=O bond approximately eclipsing the C–N bond but which differ in the orientation of the NHCH=O group. The

energetic ordering at various levels of theory of the three conformers are summarized in Table 6.1. The conformer that gives the lowest coupled-cluster energy for the B3-LYP/6-311+G(3df,2p) geometry is conformer **a4**$_a$, shown as **a4** in Figure 6.1, and so this was used in the subsequent calculations.

The geometries of fragmentation products were also obtained by initially scanning for the lowest-energy conformations. Unlike the situation for the alkoxyl radicals, there was general agreement as to the preferred conformations at the various levels of theory. These are displayed in Figure 6.2.

Transition structures were found by initially determining single-point energy profiles at the B3-LYP/6-31G(d) level as the cleaving bond was stretched, starting with the preferred conformers of the alkoxyl radicals. This procedure identified approximate transition structures, which were refined by full geometry optimizations. Using IRC calculations, we confirmed that the transition structures were connected to the initial alkoxyl radicals. However, these transition structures do not necessarily connect directly to the global conformational minima of the products. It is assumed that the energy required for the products to subsequently rearrange to the preferred conformer through torsional motions is small compared with the energy for the initial bond cleavage.

In summary, assignment of the preferred conformations for the alkoxyl radicals **a2**, **a3**, and **a4** is sensitive to the level of theory used for geometry optimization. We have taken the preferred conformation to be the one that gives the lowest CCSD(T) energy with B3-LYP/6-311+G(3df,2p) geometries. The theoretical determination of the preferred geometries of the transition structures and products is found to be more straightforward than for the alkoxyl radical reactants.

6.3.2 Effect of Level of Geometry Optimization on Calculated Reaction Enthalpies and Barriers

Having selected the preferred conformations of each of the alkoxyl radicals, cleavage products (addition reactants), and transition structures, we next sought to determine the sensitivity of the calculated reaction enthalpies and barriers to the level of geometry optimization. To do this, we calculated the enthalpies and barriers for reactions 6.1*a*, 6.1*b* and 6.1*c* by obtaining CCSD(T) energies for a variety of optimized structures. For reaction 6.1*a*, calculations were carried out with UCCSD(T)/6-311+G(3df,2p). For reactions 6.1*b* and 6.1*c*, calculations were carried out with URCCSD(T)/6-311+G(d,p) or URCCSD(T)/6-311+G(2df,p). The results of this analysis are summarized in Tables 6.2–6.4.

Table 6.2: Effect of Geometry on Calculated Enthalpy for Reaction 6.4*a* and on Cleavage and Addition Barriers (UCCSD(T)6-311+G(3df,2p), kJ mol^{-1})[a]

Geometry	ΔH	$\Delta H^{\ddagger}_{cleavage}$	$\Delta H^{\ddagger}_{addition}$
//UHF/6-31G(d)	20.3	42.4	22.1
//UMP2/6-31G(d)	19.4	42.0	22.6
//UMPW1K/6-31+G(d,p)	20.1	40.5	20.4
//UMPW1K/6-31+G(3df,2p)	21.0	41.2	20.2
//UBMK/6-31G(d)	19.6	40.6	20.9
//UB3-LYP/6-31G(d)	19.7	41.5	21.8
//UB3-LYP/6-31G(2df,p)	18.3	40.4	22.1
//UB3-LYP/6-311+G(3df,2p)	18.3	40.5	22.2
//UB3-LYP/cc-pVTZ	18.6	40.9	22.3
//UQCISD/6-31G(d)	20.1	43.8	23.7

[a] Enthalpies and barriers are not corrected for zero-point vibrational energy.

Table 6.3: Effect of Geometry on Calculated Enthalpy for Reaction 6.4*b* and on Cleavage and Addition Barriers (URCCSD(T)6-311+G(d,p), kJ mol^{-1})[a]

Geometry	ΔH	$\Delta H^{\ddagger}_{cleavage}$	$\Delta H^{\ddagger}_{addition}$
//UHF/6-31G(d)	1.7	32.7	31.0
//UHF/6-311++G(3df,2p)	5.8	34.0	28.2
//UMP2/6-31G(d)	0.6	37.1	36.5
//UMP2/6-31+G(d)	1.0	36.4	35.4
//UMP2/6-31+G(2df,p)	1.2	39.8	38.6
//UBMK/6-31G(d)	–0.2	33.2	33.5
//UB3-LYP/6-31G(d)	–0.2	33.5	33.7
//UB3-LYP/6-31G(2df,p)	–1.7	32.4	34.1
//UB3-LYP/6-311+G(3df,2p)	–2.5	32.4	34.9

[a] Enthalpies and barriers are not corrected for zero-point vibrational energy.

Table 6.2 shows that the reaction enthalpies and barriers for reaction 6.4*a* are not very sensitive to the method used to optimize the structures. In this case, the enthalpies and barriers all agree to within 3.5 kJ mol^{-1}. The enthalpies and barriers for reactions 6.4*b* and 6.4*c* are somewhat more sensitive to the level of geometry optimization, although for the most part the effects are still not very large. The greater sensitivity may in part be due to the reduced basis set size. It is noteworthy that the basis set effects appear smaller for B3-LYP than for HF or MP2.

Table 6.5 lists the lengths of the fragmenting (forming) bonds in the cleavage (addition) transition structures of reactions 6.4*a*, 6.4*b*, 6.4*c* and 6.4*d*, as calculated at a variety of theoretical levels. It can be seen that the B3-LYP and MPW1K DFT methods tend to predict longer bonds than the corresponding HF and MP2 bonds, consistent with previous observations.[24] The BMK functional predicts bond lengths similar to those of the other DFT methods, despite its formulation as a functional to give improved kinetic parameters.

In summary, assessment of the effect of geometry on the calculated enthalpies and barriers indicates that the level of theory used for geometry optimization does not have a great effect on these properties. This lends confidence to the use of B3-LYP geometries in calculating improved energies with composite methods.

Table 6.4: Effect of Geometry on Enthalpies for Reaction 6.4*c* [a] and on Cleavage and Addition Barriers (URCCSD(T)6-311+G(2df,p), kJ mol^{-1})

Geometry	ΔH	$\Delta H^{\ddagger}_{cleavage}$	$\Delta H^{\ddagger}_{addition}$
//UHF/6-31G(d)	–17.7	5.5	23.3
//UHF/6-311++G(3df,2p)	–28.3	3.5	31.8
//UMP2/6-31+G(d)	–25.2	2.7	27.8
//UB3-LYP/6-31G(d)	–20.8	5.4	26.2
//UB3-LYP/6-31G(2df,p)	–19.7	6.3	26.0
//UB3-LYP/6-311+G(3df,2p)	–19.3	6.8	26.1

[a] Enthalpies and barriers are not corrected for zero-point vibrational energy.

Table 6.5: Calculated Lengths of Cleaving (Forming) Bonds (Angstroms) in Transition Structures for Reactions 6.4*a*, 6.4*b*, 6.4*c*, and 6.4*d*

Level of Theory	**ts1**	**ts2**	**ts3**	**ts4**
HF/6-31G(d)	2.078	2.052	2.044	1.975
HF/6-311++G(3df,2p)	2.046	2.020	2.015	–
MP2/6-31G(d)	1.961	1.926	1.884	1.801
MP2/6-31+G(d)	1.954	1.917	1.875	–
MP2/6-311+G(3df,2p)	1.930	–	–	–
MPW1K/6-31+G(d,p)	2.174	2.096	2.159	–
MPW1K/6-31+G(3df,2p)	2.170	–	–	–
BMK/6-31G(d)	2.138	2.071	2.165	2.047
B3-LYP/6-31G(d)	2.148	2.068	2.155	2.024
B3-LYP/6-31G(2df,p)	2.103	2.045	2.125	1.998
B3-LYP/6-311+G(3df,2p)	2.124	2.040	2.092	1.981
QCISD/6-31G(d)	2.025	–	–	–

6.3.3 Assessment of Reaction Energies and Barriers

Enthalpies and barriers for reactions 6.4*a* to 6.4*b*, and barriers for the reverse addition reactions at various levels of theory, are given in Tables 6.6, 6.7, 6.8, and 6.9. In this section, we initially describe the results obtained with high-level composite methods. We then examine the results from several computationally more efficient levels that are more amenable to application to larger systems, and gauge their performance against that of the high-level methods. All of the high-level composite methods we use prescribe B3-LYP geometries. We found in the previous section that this should not have a large effect on the calculated relative energies.

The highest-level method used in the present Chapter is W1, which was restricted to the calculation of the addition barrier for reaction 6.4*a*. The highest-level method used for all four reactions is G3//B3-LYP. We will use this procedure to compare the performance of the other methods. Other high-level methods used for all four reactions are standard CBS-QB3, U-CBS-QB3, G3(MP2)//B3-LYP, and G3X(MP2)-RAD. Mean absolute deviations (MADs) from G3//B3-LYP for all enthalpies and barriers (cleavage and addition) are 2.2 (CBS-QB3), 1.3 (U-CBS-QB3), 1.5 (G3(MP2)//B3-LYP), and 2.8 (G3X(MP2)-RAD) kJ mol^{-1}. The MADs show that the high-level procedures agree well with one another, although it can be argued that a sample size of 12 may not be statistically significant. G3(MP2)//B3-LYP and G3X(MP2)-RAD are considerably cheaper than the other methods but give comparable performance.

The standard G*n* methods are less reliable when open-shell species have large spin contamination in the unrestricted wave functions. The RAD variants of the Gn methods[20] were designed specifically to alleviate this situation. In the present Chapter, we used G3-RAD to calculate enthalpies and barriers for reactions 6.4*a* and 6.4*b* (Tables 6.6 and 6.7) and G3X(MP2)-RAD (Tables 6.6, 6.7, 6.8 and 6.9) to calculate enthalpies and barriers for reactions 6.4*a*–6.4*d*.

Table 6.6: Effect of Level of Theory on Calculated Enthalpy of Reaction 6.4*a* and on Barriers for Fragmentation and Addition (0 K, kJ mol^{-1})

Level of theory	ΔH	$\Delta H^{\ddagger}_{cleavage}$	$\Delta H^{\ddagger}_{addition}$
W1	—	—	26.6
G3//B3-LYP	2.9	27.5	24.6
G3-RAD	4.2	30.4	26.1
CBS-QB3	3.1	24.4	21.2
U-CBS-QB3	3.2	27.9	24.7
G3X(MP2)-RAD	0.4	29.4	29.0
G3(MP2)//B3-LYP	0.5	26.1	25.7
UB3-LYP/6-311+G(3df,2p)//B3 [a]	5.4	23.7	18.2
RB3-LYP/6-311+G(3df,2p)//B3 [a]	9.6	23.7	14.1
UBMK/6-311+G(3df,2p)//B3 [a]	13.1	35.9	22.7
RBMK/6-311+G(3df,2p)//B3 [a]	11.8	37.8	26.0
UBMK/6-311+G(3df,2p)//BM [b]	13.1	36.5	23.3
RBMK/6-311+G(3df,2p)//BM [b]	9.8	36.5	26.5
UMP2/6-311+G(3df,2p)//B3 [a]	–23.8	38.9	62.7
RMP2/6-311+G(3df,2p)//B3 [a]	–26.9	–0.6	26.3
UQCISD/6-31G(d)	12.5	53.9	41.4
UB3-LYP/cc-pVTZ	10.8	29.0	18.2
UB3-LYP/6-311+G(3df,2p)	9.7	28.3	18.6
UB3-LYP/6-31G(2df,p)	17.0	30.7	13.8
UB3-LYP/6-31G(d)	24.8	36.7	12.0
UMPW1K/6-31+G(3df,2p)	46.2	61.0	14.8
UMPW1K/6-31+G(d,p)	54.9	70.3	15.4
UMP2/6-311+G(3df,2p)	–21.9	46.4	68.3
UMP2/6-31G(2df,p)	–13.7	53.2	66.8
UMP2/6-31G(d)	–5.9	75.1	81.0
UHF/6-311++G(3df,2p)	38.6	96.5	58.0
UHF/6-31G(d)	60.1	107.8	47.7

[a] The notation //B3 is used to designate that the calculations were carried out on UB3-LYP/6-31G(d) optimized structures. [b] The notation //BM is used to designate that the calculations were carried out on UBMK/6-31G(d) optimized structures.

Table 6.7: Effect of Level of Theory on Calculated Enthalpy of Reaction 6.4*b* and on Barriers for Fragmentation and Addition (0 K, kJ mol^{-1})

Level of theory	ΔH	$\Delta H^{\ddagger}_{cleavage}$	$\Delta H^{\ddagger}_{addition}$
G3//B3-LYP	–12.3	21.1	33.4
G3-RAD	–12.7	21.6	34.3
CBS-QB3	–10.8	19.2	30.0
U-CBS-QB3	–10.7	22.6	33.3
G3X(MP2)-RAD	–13.6	23.5	37.1
G3(MP2)//B3-LYP	–14.5	19.9	34.4
UB3-LYP/6-311+G(3df,2p)//B3 [a]	–14.1	20.4	34.5
RB3-LYP/6-311+G(3df,2p)//B3[a]	–16.2	19.7	35.9
UBMK/6-311+G(3df,2p)//B3 [a]	–7.2	28.9	36.1
RBMK/6-311+G(3df,2p)//B3 [a]	–8.5	30.2	38.7
UBMK/6-311+G(3df,2p)//BM [b]	–6.9	29.3	36.2
RBMK/6-311+G(3df,2p)//BM [b]	–8.7	30.0	38.7
UMP2/6-311+G(3df,2p)//B3 [a]	–42.4	27.5	69.9
RMP2/6-311+G(3df,2p)//B3 [a]	–45.7	–13.0	32.7
UB3-LYP/6-311+G(3df,2p)	–13.9	19.3	35.0
UB3-LYP/6-31G(2df,p)	–2.9	23.9	28.7
UB3-LYP/6-31G(d)	4.0	28.7	26.7
UMP2/6-31+G(2df,p)	–37.9	33.8	71.8
UMP2/6-31+G(d)	–53.5	29.6	85.3
UMP2/6-31G(d)	–30.3	51.6	84.1
UHF/6-31G(d)	34.2	97.5	63.4
UHF/6-311++G(3df,2p)	12.2	87.5	75.3

[a] The notation //B3 is used to designate that the calculations were carried out on UB3-LYP/6-31G(d) optimized structures. [b] The notation //BM is used to designate that the calculations were carried out on UBMK/6-31G(d) optimized structures.

Table 6.8: Effect of Level of Theory on Calculated Enthalpy of Reaction 6.4*c* and on Barriers for Fragmentation and Addition (0 K, kJ mol^{-1})

Level of theory	ΔH	$\Delta H^{\ddagger}_{cleavage}$	$\Delta H^{\ddagger}_{addition}$
G3//B3-LYP	12.7	14.1	1.4
CBS-QB3	15.1	13.9	–1.2
U-CBS-QB3	15.1	16.5	1.4
G3X(MP2)-RAD	13.4	19.1	5.6
G3(MP2)//B3-LYP	10.8	13.5	2.7
UB3-LYP/6-311+G(3df,2p)//B3 [a]	5.1	14.4	9.3
RB3-LYP/6-311+G(3df,2p)//B3 [a]	3.1	14.1	11.0
UBMK/6-311+G(3df,2p)//B3 [a]	16.4	23.8	7.5
RBMK/6-311+G(3df,2p)//B3 [a]	14.2	24.4	10.2
UBMK/6-311+G(3df,2p)//BM [b]	15.4	22.7	7.4
RBMK/6-311+G(3df,2p)//BM [b]	9.4	19.6	10.2
UMP2/6-311+G(3df,2p)//B3 [a]	–16.0	8.9	24.9
RMP2/6-311+G(3df,2p)//B3 [a]	–19.4	–18.6	0.8
UB3-LYP/6-311+G(3df,2p)	2.0	11.6	9.6
UB3-LYP/6-31G(2df,p)	21.8	18.2	–3.6
UB3-LYP/6-31G(d)	26.8	23.2	–3.5
UMP2/6-31+G(d)	–5.7	41.3	47.0
UMP2/6-31G(d)	–2.0	38.6	40.6
UHF/6-311++G(3df,2p)	33.4	89.8	56.4
UHF/6-31G(d)	52.3	96.5	44.2

[a] The notation //B3 is used to designate that the calculations were carried out on UB3-LYP/6-31G(d) optimized structures. [b] The notation //BM is used to designate that the calculations were carried out on UBMK/6-31G(d) optimized structures.

Table 6.9: Effect of Level of Theory on Calculated Enthalpy of Reaction 6.4*d* and on Barriers for Fragmentation and Addition (0 K, kJ mol^{-1})

Level of Theory	ΔH	$\Delta H^{\ddagger}_{cleavage}$	$\Delta H^{\ddagger}_{addition}$
G3//B3-LYP	–49.8	–4.6	45.2
CBS-QB3	–45.8	–4.2	41.6
U-CBS-QB3	–46.3	–2.2	44.1
G3X(MP2)-RAD	–51.9	–2.9	49.1
G3(MP2)//B3-LYP	–52.2	–5.8	46.3
UB3-LYP/6-311+G(3df,2p)//B3 [a]	–60.8	–1.1	59.7
RB3-LYP/6-311+G(3df,2p)//B3 [a]	–62.9	–2.1	60.8
UBMK/6-311+G(3df,2p)//B3 [a]	–54.3	4.6	58.9
RBMK/6-311+G(3df,2p)//B3 [a]	–57.7	3.8	61.4
UBMK/6-311+G(3df,2p)//BM [b]	–55.7	1.8	57.6
RBMK/6-311+G(3df,2p)//BM [b]	–60.2	0.0	60.2
UMP2/6-311+G(3df,2p)//B3 [a]	–91.5	–18.8	72.7
RMP2/6-311+G(3df,2p)//B3 [a]	–98.0	–55.2	42.8
UB3-LYP/6-311+G(3df,2p)	–60.5	0.4	60.8
UB3-LYP/6-31G(2df,p)	–49.1	2.8	51.9
UB3-LYP/6-31G(d)	–44.4	6.0	50.3
UMP2/6-31G(d)	–83.4	9.0	92.5
UHF/6-31G(d)	–43.4	66.7	110.1

[a] The notation //B3 is used to designate that the calculations were carried out on UB3-LYP/6-31G(d) optimized structures. [b] The notation //BM is used to designate that the calculations were carried out on UBMK/6-31G(d) optimized structures.

G3-RAD produces enthalpies and barriers that are similar to those from the G3//B3-LYP method. For the enthalpies of reactions 6.4*a* and 6.4*b*, this is not unexpected. The spin contamination of the alkoxyl radicals and the product radicals formed in the cleavage reactions is very small. At the UMP2/6-31G(d) level, the $<S^2>$ values are less than 0.77, compared with 0.75 for a pure doublet. The transition structures for reactions 6.4*a* and 6.4*b* have larger $<S^2>$ values (0.89 and 0.88, respectively) and so it is more surprising that the difference between G3-RAD and

G3//B3-LYP barriers for these two reactions is small. With G3X(MP2)-RAD, enthalpies and barriers for reactions 6.4*a* and 6.4*b*, as well as for 6.4*c* and 6.4*d*, compare well with G3//B3-LYP. This is characterised by the relatively small MAD of 2.8 kJ mol^{-1}, as quoted above.

Enthalpies and barriers calculated from UB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d) and RB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d) energies give fairly good agreement with those of the high-level methods. MADs from the G3//B3-LYP results are 5.1 and 5.6 kJ mol^{-1} and largest deviations (LDs) are 15.6 and 14.5 kJ mol^{-1} for UB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d) and RB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d), respectively. The relatively small MADs and LDs make these methods candidates for use with larger systems when G3(MP2)//B3-LYP and G3X(MP2)-RAD become too costly. The very small difference between the two methods is not surprising when considering the $<S^2>$ values for the alkoxyl radicals, product radicals, and transition structures. For UB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d), $<S^2>$ values are always less than 0.76.

We note that the UB3-LYP/6-311+G(3df,2p) enthalpies and barriers obtained from fully optimized UB3-LYP/6-311+G(3df,2p) structures are close to those obtained from UB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d). The MAD for the fully optimized UB3-LYP/6-311+G(3df,2p) structures increases slightly to 5.9 kJ mol^{-1} and has an LD of 15.6 kJ mol^{-1}. B3-LYP enthalpies and barriers obtained with smaller basis sets, *viz.* 6-31G(d) and 6-31G(2df,p), are somewhat poorer. Thus, for B3-LYP/6-31G(d), the MAD and LD are 10.8 and 21.9 kJ mol^{-1}, while for B3-LYP/6-31G(2df,p) they are 6.4 and 14.1 kJ mol^{-1}, respectively.

The MADs of 6.9 and 7.6 kJ mol^{-1} for UBMK/6-311+G(3df,2p)//UB3-LYP/6-31G(d) and RBMK/6-311+G(3df,2p)//UB3-LYP/6-31G(d), respectively, demonstrate that these two methods, like the analogous B3-LYP calculations, are also potentially cost-effective alternatives to the more sophisticated composite methods. The MADs are

improved to 6.5 kJ mol^{-1} for UBMK/6-311+G(3df,2p) and to 6.9 kJ mol^{-1} for RBMK/6-311+G(3df,2p) if UBMK/6-31G(d) geometries are used in place of the UB3LYP/6-31G(d) geometries.

Enthalpies and barriers calculated with HF and MP2 in general are not very satisfactory. The MADs for UMP2/6-311+G(3df,2p)//UB3-LYP/6-31G(d) and RMP2/6-311+G(3df,2p)//UB3-LYP/6-31G(d) are 24.1 and 23.7 kJ mol^{-1}, with LDs of 41.7 and 48.2 kJ mol^{-1}, respectively. For UMP2/6-311+G(3df,2p)//UB3-LYP/6-31G(d), it is interesting that the addition barriers agree reasonably well with G3//B3-LYP with an LD of 14.2 kJ mol^{-1}, while the cleavage barriers are poor with an LD of 38.0 kJ mol^{-1}. On the other hand, the addition barriers are poor with RMP2/6-311+G(3df,2p)//UB3-LYP/6-31G(d) with an LD of 50.7 kJ mol^{-1}, but the cleavage barriers are good with an LD of 2.4 kJ mol^{-1}. For both UMP2/6-311+G(3df,2p)//UB3-LYP/6-31G(d) and RMP2/6-311+G(3df,2p)//UB3-LYP/6-31G(d), the reaction enthalpies are similar and consistently too exothermic.

The differences between the RMP2 and UMP2 results can be rationalized in terms of the general performance of MP2 theory in describing alkoxyl radicals, and the $<S^2>$ values of the alkoxyl radicals and the transition structures. As noted in the section on conformers, MP2 theory predicts higher than expected energies for the alkoxyl radicals. However, MP2 theory performs reasonably in the calculation of the fragmentation products. This leads to a greater than expected calculated exothermicity. The UMP2 and RMP2 reaction enthalpies are similar because of low spin contamination in both the alkoxyl radicals and their associated products, i.e., the errors in the enthalpies are associated with MP2 rather than spin contamination. Large spin contamination becomes a problem in the transition structures at the UMP2 level. UMP2/6-31G(d) $<S^2>$ values for the transition structures of reactions 6.4*a*, 6.4*b*, 6.4*c*, and 6.4*d* are 0.89, 0.87, 0.84, and 0.81, respectively. This leads to a fortuitous cancellation of errors when calculating fragmentation barriers with UMP2 but poor

fragmentation barriers with RMP2. For the same reasons, addition barriers are poor at the UMP2 level but good with RMP2.

Additional UMP2 and UHF enthalpies and barriers were calculated from optimizations at these levels using a variety of basis sets. These include 6-31G(d), 6-31+G(d), 6-31G(2df,p), 6-31+G(2df,p), 6-31+G(3df,2p), and 6-311++G(3df,2p). As expected from the discussion above, enthalpies and addition barriers are poor with no substantial improvement with basis set size.

The remaining methods that were examined involved full geometry optimizations at the UMPW1K and UQCISD levels of theory with various basis sets for reaction 6.4*a*. The results are summarised in Table 6.6. For UMPW1K with small and large basis sets, the deviations from the G3//B3-LYP results are large. UQCISD/6-31G(d) also does not perform well in comparison with the high-level methods. Contributing factors to this result are most probably the small basis set and the omission of the perturbative triples.

In summary, high-level composite methods are in generally good agreement with one another for the reaction enthalpies and barriers (both cleavage and addition). The best composite methods in terms of accuracy and cost are G3(MP2)//B3-LYP and G3X(MP2)-RAD, which give enthalpies and barriers in good agreement with the more expensive G3//B3-LYP, G3-RAD, CBS-QB3, and U-CBS-QB3 methods. Relatively accurate but low-cost methods suitable for application to larger systems are provided by single-point energy calculations with UB3-LYP/6-311+G(3df,2p), RB3-LYP/6-311+G(3df,2p), UBMK/6-311+G(3df,2p) or RBMK/6-311+G(3df,2p) on UB3-LYP/6-31G(d) geometries. Using UBMK/6-31G(d) geometries in place of UB3-LYP/6-31G(d) geometries has only a small effect. UMP2/6-311+G(3df,2p)//UB3-LYP/6-31G(d) calculations appear to give good results for the cleavage barriers but not for addition barriers. On the other hand RMP2/6-311+G(3df,2p)//UB3-LYP/6-31G(d) calculations

appear to give good results for the addition barriers but not for cleavage barriers. Neither of these two methods is suitable for the calculation of reaction enthalpies.

6.3.4 Heats of Formation

Heats of formation at 0 K for the fragmentation products (**fp1**–**fp5**) and the four substituted alkoxyl radicals (**a1**–**a4**) of reactions 6.4*a*, 6.4*b*, 6.4*c* and 6.4*d* were calculated at various levels of theory. These results are compared with available experimental values in Table 6.10.

In general, the high-level methods agree well with one another and with available experimental data. B3-LYP gives varied results, with discrepancies of up to 30 kJ mol^{-1} in some cases, and only a few kJ mol^{-1} in others. It has previously been found that B3-LYP/6-311+G(3df,2p) enthalpies of formation for species in the G3/99 test set have an MAD from reliable experimental values of 22.6 kJ mol^{-1}.[25] The B3-LYP results of this section are better than expected on this basis since all but one of the enthalpies of formation have deviations less than this value. It has also previously been found that errors accumulate in the B3-LYP heats of formation as the size of the system increases. This is why the G3/99 test set, with an average of approximately 17 electron pairs per species, has a larger MAD (22.6 kJ mol^{-1}) for heats of formation compared with the G2/97 test set (13.7 kJ mol^{-1}), which has an average of approximately 11 electron pairs per species. We do not find such a correlation for the small number of species in the present work, indicating that other effects are having a more important influence in these cases. For example, the B3-LYP/6-311+G(3df,2p)//B3-LYP/6-31G(d) heat of formation of •$OC(CH_3)(NHCH{=}O)CH{=}O$ (**a4**), which has 22 electron pairs, deviates by 5.5 kJ mol^{-1} from the G3//B3-LYP value. On the other hand, the B3-LYP/6-311+G(3df,2p)//B3-LYP/6-31G(d) heat of formation of $NH_2(C\bullet){=}O$ (**fp4**), which has 8 electron pairs, deviates by 28.9 kJ mol^{-1} from the G3//B3LYP value.

Table 6.10: Effect of Level of Theory on Calculated Heats of Formation (0 K, kJ mol^{-1}) of Fragmentation Products (**fp1**, **fp2**, **fp3**, **fp4**, and **fp5**) and of Alkoxyl Radicals (**a1**, **a2**, **a3** and **a4**)[a]

Level of Theory	**fp1**	**fp2**	**fp3**	**fp4**	**fp5**	**a1**	**a2**	**a3**	**a4**
B3-LYP/6-311+G(3df,2p)//B3	–107.2	–152.4	31.3	–36.4	–351.8	–81.4	–101.5	–193.9	–259.8
B3-LYP/6-311+G(3df,2p)	–108.4	–154.6	30.3	–37.2	–355.3	–87.9	–109.9	–192.7	–264.6
CBS-QB3	–109.1	–155.7	41.3	–7.6	–354.9	–70.9	–102.1	–178.3	–267.7
U-CBS-QB3	–110.9	–158.1	39.9	–8.9	–359.0	–74.2	–106.1	–182.1	–272.8
G3(MP2)//B3-LYP	–108.9	–156.3	37.4	–5.5	–349.5	–71.9	–101.4	–172.6	–259.9
G3X(MP2)-RAD	–108.3	–155.8	40.1[f]	–3.2	–349.1	–68.6	–100.3	–172.5	–257.1
G3//B3-LYP	–108.7	–157.0	39.1	–7.5	–354.1	–72.6	–102.7	–177.2	–265.3
G3-RAD(6d,10f)	–107.4	–156.2	41.0[f]	–2.8	–350.2	–70.7	–102.6	–	–
W1	–107.3	–	40.5	–	–	–	–	–	–
Expt.	–104.6±0.7[b]	–155.0[b]	41.3±0.8[c]/ 44.6±0.4[d,e]	–	–	–	–	–	–

[a] See Figures 6.1 and 6.2 for B3-LYP/6-311+G(3df,2p) structures of **a1**-**a4** and **fp1**-**fp5**, respectively. [b] From Lias et al.[29] [c] From Berkowitz et al.[33] [d] From Becerra et al.[32] [e] $\Delta_f H_{298}$ back-corrected to 0 K using theoretical temperature correction from B3-LYP/6-31G(d) frequencies (scaled by 0.9989[14]): see Henry et al.[31] [f] from Henry et al.[31]

Surprisingly the heats of formation calculated with BMK/6-311+G(3df,2p)//B3-LYP/6-31G(d)[26] are quite different from the corresponding B3-LYP/6-311+G(3df,2p) values and it is recommended that this method not be used for calculating heats of formation for the alkoxyl radicals.

*CH_2=O (**fp1**).* Formaldehyde is a well-studied molecule and is primarily included in Table 6.10 for the sake of completeness. Our high-level methods agree in their predicted heats of formation for formaldehyde to within 1.8 kJ mol^{-1}, with W1, the highest level used in this Chapter, predicting a value of –107.3 kJ mol^{-1}. It is evident that both the current theoretical predictions and other recent high-level theoretical predictions, e.g. the W2 and W3 results of Martin,[27] are consistently more negative than the experimental heat of formation, –104.6 ± 0.7 kJ mol^{-1},[28] by about 3 kJ mol^{-1}. The reason for this apparent discrepancy is not clear.

*CH_3CH=O (**fp2**).* Acetaldehyde, also has a well-characterised experimental heat of formation (–155.0 kJ mol^{-1}).[29] The high-level theoretical values are all in good agreement with this experimental value. The largest deviation from experiment is 3.1 kJ mol^{-1} at the U-CBS-QB3 level.

*•CH=O (**fp3**).* The heat of formation of the formyl radical has been theoretically investigated in several high-level studies. Most recently, Feller et al.[30] have used various extrapolation schemes of coupled-cluster energies to obtain a value of 43.5 ± 0.8 kJ mol^{-1}. The associated error has been estimated from the spread of data given by the different extrapolations. Earlier, Henry et al.[31] used a number of high-level composite methods to calculate $\Delta_f H^\circ_0$(•CHO). Some of the values from their work are noted in Table 6.10. Two experimental values are available, the more recent value of 44.6 ± 0.4 from Becerra et al.[32] and an earlier value of 41.3 ± 0.8 kJ mol^{-1} from Berkowitz et al.[33] The W1, G3-RAD, and G3X(MP2)-RAD estimates all fall within the experimental uncertainty of the Berkowitz et al.[39] value. However, the theoretical value of Feller et al.[30] is in better agreement with the more recent value of Becerra et al.[32]

*$NH_2C(\bullet)=O$ (**fp4**).* A heat of formation for the $NH_2C(\bullet)=O$ fragmentation product of reaction A3 has not been reported in the literature. The heats of formation predicted by the high-level methods range between –3.2 kJ mol^{-1} from G3X(MP2)-RAD to –8.9 kJ mol^{-1} from U-CBS-QB3. The previous comparisons with experimental values suggest that U-CBS-QB3 might slightly underestimate heats of formation and so this value may be too low. The B3-LYP values are more negative by 30 kJ mol^{-1} compared with those from the high-level composite methods. If we average the high-level methods, i.e. all the values in Table 6.10 excluding the B3-LYP estimates, we obtain –5.9 kJ mol^{-1} as the 0 K heat of formation for **fp4**.

*$CH(=O)NHC(CH_3)=O$ (**fp5**).* The previous observation that U-CBS-QB3 seems to underestimate heats of formation also applies in the case of **fp5**. Compared with the other high-level methods, U-CBS-QB3 gives the most negative heat of formation of –359.0 kJ mol^{-1}. In contrast to **fp4**, the B3-LYP values compare well with the remaining high-level methods. By averaging the results of all the high-level methods in Table 6.10, we obtain –352.8 kJ mol^{-1} for the 0 K heat of formation for $CH(=O)NHC(CH_3)=O$ (**fp5**).

*Alkoxyl radicals (**a1**, **a2**, **a3**, and **a4**).* As in the case of the fragmentation products discussed in the preceding paragraphs, the high-level methods are in good agreement with one another. Another continued trend is the underestimation of the heats of formation by U-CBS-QB3. The B3-LYP methods give mixed results. In the case of **a1** and **a3,** the values are somewhat outside the ranges of the high-level methods. However, for alkoxyl radicals **a2** and **a4** the B3-LYP heats of formation are in good agreement with the high-level results. If we average the values from the high-level methods in Table 6.10, we obtain heats of formation at 0 K for the alkoxyl radicals of –71.4 (**a1**), –102.5 (**a2**), –176.6 (**a3**) and –264.6 (**a4**) kJ mol^{-1}, respectively.

In summary, heats of formation at 0 K calculated by high-level methods, agree well with one another. U-CBS-QB3 seems to underestimate the heats of formation

when compared with the other high levels of theory, but the deviation is only small. In recommending heats of formation, we average the high-level results (including U-CBS-QB3). Alkoxyl radical **a4** gives the largest range of such values (10.4 kJ mol^{-1}) but for the other species, the range of values is much less (5.8 kJ mol^{-1}). B3-LYP gives varied results, with discrepancies of up to 30 kJ mol^{-1} in some cases, and only a few kJ mol^{-1} in others.

6.4 Concluding Remarks

The purpose of the present Chapter has been to establish methodologies for investigating the C–C beta-scission reactions of substituted alkoxyl radicals that might serve as models for peptide backbone alkoxyl radicals. We seek methods that provide a suitable compromise between the accuracy in predicting enthalpies and barriers for the beta-scission reactions, and computational cost.

We find that determination of the lowest-energy conformers of the $•OCH_2CH=O$ (**a1**), $•OCH(CH_3)CH=O$ (**a2**), $•OCH(CH_3)C(NH_2)=O$ (**a3**), and $•OC(CH_3)(NHCH=O)CH=O$ (**a4**) substituted alkoxyl radicals is sensitive to the method used to optimize geometries. We have selected the conformer optimized with B3-LYP/6-311+G(3df,2p) that gives the lowest CCSD(T) energy for use in subsequent calculations.

In the calculation of enthalpies and barriers, we find that G3(MP2)//B3-LYP and G3X(MP2)-RAD give similar accuracy to more expensive high-level methods. However, these methods are still computationally too demanding to be used routinely with currently available resources for systems significantly larger than the largest models used in the present Chapter, i.e. reactions 6.4*c* and 6.4*d*. Cost-effective alternative methods are UB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d) and RB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d) or UBMK/6-311+G(3df,2p) and RBMK/6-311+G(3df,2p) with either UB3-LYP/6-31G(d) or UBMK/6-31G(d) geometries. These

give reasonable estimates of enthalpies and barriers that are within about 15 kJ mol^{-1} of the reference G3//B3-LYP values.

Heats of formation for each of the species involved in the four reactions under study were also calculated. The high-level methods perform well for the three species for which experimental values are available for comparison. U-CBS-QB3 seems to underestimate the heats of formation when compared with the other high-level methods. Predicted heats of formation at 0 K for the alkoxyl radicals are –71.4 (•OCH_2CH=O, **a1**), –102.5 (•$OCH(CH_3)CH$=O, **a2**), –176.6 (•$OCH(CH_3)C(NH2)$=O, **a3**), and –264.6 (•$OC(CH_3)(NHCH$=O)CH=O, **a4**) kJ mol^{-1}. For the fragmentation products, NH_2C(•)=O (**fp4**), and CH(=O)$NHC(CH_3)$=O (**fp5**), we predict the respective values of –5.3 kJ mol^{-1} and –351.6 kJ mol^{-1}.

6.5 References

(1) (a) Davies, M. J.; Fu, S.; Dean, R. T. *Biochem. J.* **1995**, *305*, 643. (b) Davies, M. J. *Arch. Biochem. Biophys.* **1996**, *336*, 163. (c) Gebicki, J. M. *Redox Rep.* **1997**, *3*, 99.

(2) See for example: (a) R. Atkinson and J. Arey, *Chem. Rev.* **2003**, 103, 4605. (b) R. Atkinson, *Int. J. Chem. Kinet.*, **1997**, 29, 99. (c) Mereau, R.; Rayez, M. T.; Caralp, F.; Rayez, J. C. *Phys. Chem. Chem. Phys.*, **2000**, *2*, 1919. (d) Mereau, R.; Rayez, M. T.; Caralp, F.; Rayez, J. C. *Phys. Chem. Chem. Phys.*, **2000**, *2*, 3765. (e) P. Devolder, *J. Photochem. Photobiol. A*, **2003**, 157, 137. (f) J. J. Orlando, G. S. Tyndall and T. J. Wallington, *Chem. Rev.* **2003**, 103, 4657.

(3) Mortimer, A.; Easton, C. J. *unpublished data.*

(4) Gray, P.; Williams, A. *Chem. Rev.* **1959**, *59*, 239.

(5) Kochi, J. K. *J. Am, Chem. Soc.* **1962**, *84*, 1193.

(6) Zhang, W.; Dowd, P. *Tetrahedron* **1993**, *49*, 1965.

(7) Wilsey, S.; Dowd, P.; Houk, K. N. *J. Org. Chem.* **1999**, *64*, 8801.

(8) Rauk, A.; Boyd, R. J.; Boyd, S. L.; Henry, D. J.; Radom, L. *Can. J. Chem.* **2003**, *81*, 431.

(9) (a) Hippler, H.; Striebel, F.; Viskolcz, B. *Phys. Chem. Chem. Phys.* **2001**, *3*,2450. (b) Fittschen, C.; Hippler, H.; Viskolcz, B. *Phys. Chem. Chem. Phys.* **2000**, *2*, 1677.

(10) Dibble, T. S. *J. Phys. Chem A* **1999**, *103*, 8559.

(11) Huang, M. L.; Rauk, A. *J. Phys. Org. Chem.* **2004**, *17*, 777.

(12) Shapley, W. A.; Bacskay, G. B. *J. Phys. Chem A* **1999**, *103,* 4514.

(13) Song, K-S.; Cheng, Y-H.; Fu, Y.; Liu, L.; Li, X-S.; Guo, Q-X. *J. Phys. Chem. A* **2002**, *106,* 6651.

(14) Scott, A. P.; Radom, L. *J. Phys. Chem.* **1996**, *100,* 16502.

(15) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **2001**, *114*, 108.

(16) Lynch, B. J.; Truhlar, D. A. *J. Phys. Chem A* **2001**, *105*, 2936.

(17) Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1999**, *110*, 2822.

(18) U-CBS-QB3 is a variant of CBS-QB3 that omits the empirical correction for spin contamination in the unrestricted wave function. See for example: (a) Gómez-Balderas, R.; Coote, M. L.; Henry, D. J.; Radom, L. *J. Phys. Chem. A,* **2004**, *108*, 2874. (b) Wood, G. P. F.; Henry, D. J.; Radom, L. *J. Phys. Chem. A*, **2003**, *107*, 7985. (c) Coote, M. L. *J. Phys. Chem. A,* **2004**, *108*, 3865.

(19) Baboul, A. G.; Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **1999**, *110*, 7650.

(20) Henry, D. J.; Sullivan, M. B.; Radom, L. *J. Chem. Phys.* **2003**, *118,* 4849.

(21) Martin, J. M. L.; De Oliveira, G. J. *J. Chem. Phys.* **1999**, *111*, 1843.

(22) Bataev, V. A.; Abramenkov, A. V.; Godunov, I. A. *Russ. Chem. Bull.* **2001**, *50*, 945.

(23) (a) Von Neumann, J.; Wigner, E. *Z. Phys.* **1929**, *30*, 467 (b) Stanton, J. F. *J. Chem. Phys.* **2001**, *115*, 10382–10393, and references therein. (c) Russ, N. J.; Crawford, D. T.; Tschumper, G. S. *J. Chem. Phys.* **2004**, *120*, 7298.

(24) See, for example: Coote, M. L.; Wood, G. P. F.; Radom, L. *J. Phys. Chem. A* **2002,** *106*, 12124.

(25) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **2000**, *112*, 7374.

(26) Heats of formation in (0 K, kJ mol^{-1}) calculated with UBMK/6-311+G(3df,2p)//UBMK/6-31G(d) are: –113.4 (CH_2=O, **fp1**), –114.1 (CH_3CH=O, **fp2**), 25.7 (•CH=O, **fp3**), –42.6 (NH_2C(•)=O, **fp4**), –393.4 (CH(=O)NHC(CH_3)=O, **fp5**), –100.8 (•OCH_2CH=O, **a1**), –131.5 (•OCH(CH_3)CH=O, **a2**), –222.0 (•OCH(CH_3)C(NH2)=O, **a3**), –312.0 (•OC(CH_3)(NHCH=O)CH=O, **a4**).

(27) (a) Parthiban, S.; Martin, J. M. L. *J. Chem. Phys.* **2001**, *114*, 6014. (b) Boese, D. A.; Oren, M.; Atasoylu, O.; Martin, J. M. L.; Kállay, M.; Gauss. J. *J. Chem. Phys.* **2004**, *120*, 4129.

(28) Fletcher, R. A.; Pilcher, G. *Trans. Faraday Soc.* **1970**, *66*, 794.

(29) Lias, S. G.; Bartmess, J. E.; Liebman, J. F.; Holmes, J. L.; Levin, R. D.; Mallard, W. G. *J. Phys. Chem. Ref. Data* **1988**, *17*, Suppl 1.

(30) Feller, D.; Dixon, D. A.; Francisco, J. S. *J. Phys. Chem. A* **2003**, *107,* 1604.

(31) Henry, D. J.; Parkinson, C. J.; Radom, L. *J. Phys. Chem. A* **2002**, *106,* 7927.

(32) Becerra, R.; Carpenter, I. W.; Walsh, R. *J. Phys. Chem. A* **1997**, *101,* 4185. The reported $\Delta_f H_{298}$ value has been back-corrected to 0 K: see Henry et al.[31]

(33) Berkowitz. J.; Ellison, G. B.; Gutman, D. *J. Phys. Chem.* **1994**, *98,* 2744.

CHAPTER 7

Preferred β-Scission Reaction of Peptide Backbone Alkoxyl Radicals: C–C, C–N or C–R?

7.1 Introduction

In this Chapter, we examine in more detail the β-scission pathways of a peptide backbone alkoxyl radical, firstly by using higher levels of theory than previously employed on the glycine and alanine peptide models, and secondly by examining a wider range of peptide models to see how the preference for C–C, C–N and C–R β-scission changes with the side chain. It has previously been found that the activation energies for β-scission reactions are often determined largely by the stabilities of the daughter radicals.[1] Therefore, we examine whether the large gap previously found[2] between the barriers for C–C and C–R β-scission reactions decreases if the side-chain reactions involve residues that give rise to highly stabilized fragment radicals. We find that the C–R β-scission can indeed become competitive with backbone C–C β-scission in such circumstances. Attempts to test this prediction experimentally have so far proved problematical.

7.2 Theoretical Procedures

In the previous Chapter we identified a number of procedures that would be suitable for theoretical investigations of C–C β-scission reactions. B3-LYP/6-31G(d) geometries were found to be satisfactory for single-point energy evaluations with higher levels of theory. For the calculation of enthalpies and barriers, G3(MP2)//UB3-LYP and G3X(MP2)-RAD gave results with an accuracy similar to that of more expensive high-level methods, such as CBS-QB3 and W1. Cost-effective alternative procedures included UB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d), RB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d), UBMK/6-311+G(3df,2p)//UB3-LYP/6-31G(d) and RBMK/6-311+G(3df,2p)//UB3-LYP/6-31G(d). The assessment of theoretical methods is extended in the present Chapter to the C–N β-scission pathway and the C–R β-scission pathway. Geometries have been optimized and harmonic frequencies, used to obtain zero-point vibrational energies (with a scale factor of 0.9806[3]), calculated

using B3-LYP/6-31G(d). Energies were calculated using a selection of the methods noted above to have performed best for the C–C pathway. Radical stabilization energies (RSEs) for radicals •R have been calculated as the energy change in the formal reaction:

$$\bullet R + CH_4 \rightarrow RH + \bullet CH_3$$

i.e., the difference in bond dissociation energies for CH_4 and RH. The RSEs provide a measure of the stabilities of the radicals •R relative to their closed-shell counterparts RH.

Standard transition state theory employing the harmonic approximation was used to obtain Arrhenius parameters (log k, log A and E_a) at 298 K.

7.3 Results and Discussion

In the previous Chapter we examined in detail the effect of level of theory on reaction enthalpies and barriers for the C–C β-scission pathway in models for Ala peptide radicals. A primary conclusion of that study was that large-basis-set calculations with the hybrid density functionals BMK and B3-LYP (either unrestricted or restricted) give reasonable estimates of enthalpies and barriers when compared with high-level composite methods such as W1, G3X(MP2)-RAD and CBS-QB3.

A notable result was the poor performance of MP2 (with both restricted and unrestricted reference wave functions) for these calculations. This might be associated with the presence of a low-lying excited state in alkoxyl radicals. Because the CBS-QB3 and G3X(MP2)-RAD composite methods involve a number of MP2 calculations, the reliability of these methods could be called into question as a consequence of the poor performance of MP2. In this regard, we found in the previous Chapter that the CBS-QB3 and G3-type procedures gave consistent results, which suggests that these

procedures *are* giving reliable enthalpies and barriers. However, only a minimal number of external checks were made with high-level procedures (such as W1 and CCSD(T)) that do not involve MP2 calculations. This aspect is therefore examined further here. We also examine the effect of level of theory on reaction enthalpies and barriers for the C–N and C–R pathways of an Ala-containing peptide.

7.3.1 C–C β-Scission Reactions

Table 7.1 presents reaction enthalpies and barriers for six model peptide-backbone alkoxyl radicals reacting via the C–C β-scission pathway, as presented in Figure 7.1.

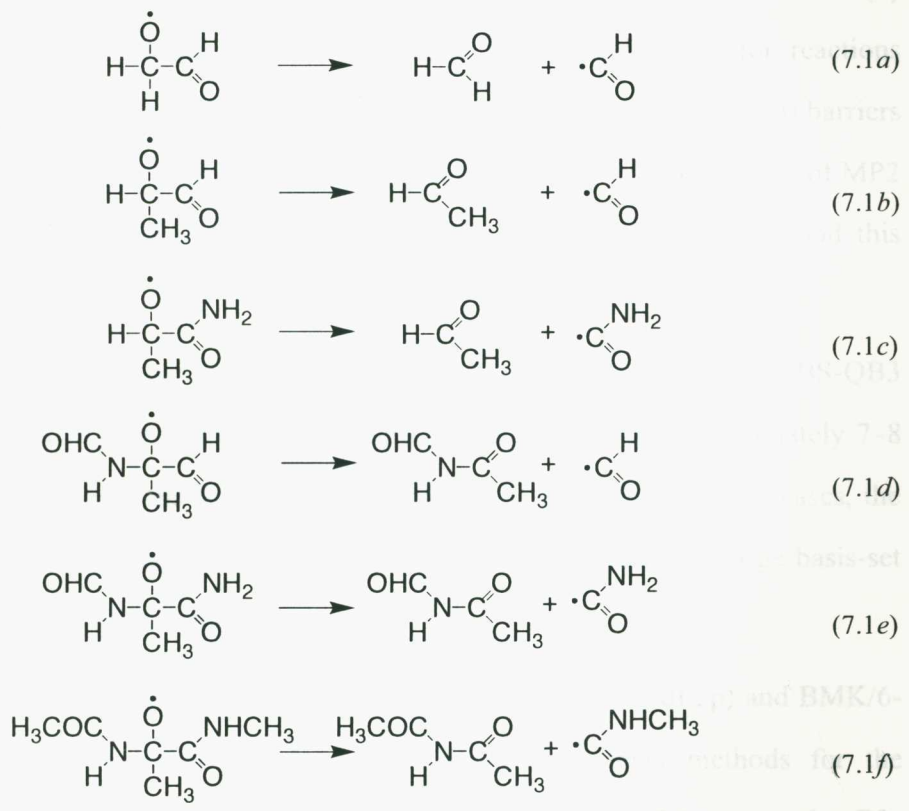


Figure 7.1: Six model peptides for the study of the β-scission of the C–C backbone bond from an α-C-centered alkoxyl radical.

The six reactions have been selected to demonstrate the effect on the thermochemical parameters of building up the peptide chain so as to increasingly simulate the Ala radical center, leading up to the final reaction, which represents our largest model of the peptide backbone (see Figure 7.4 below). For reactions 7.1*a*–7.1*c*, the CBS-QB3 and G3X(MP2)-RAD results are compared with those of a large-basis-set (cc-pVTZ) URCCSD(T) calculation. The enthalpy and barriers for reaction 7.1*a* are also calculated with UBD(T)/cc-pVTZ//UB3-LYP/6-31G(d). In general, the high-level results are in good agreement with one another, suggesting that the basis-set corrections based on MP2 calculations in the case of G3X(MP2)-RAD and the pair-energies extrapolation in CBS-QB3 are not suffering from the same problems in the computation of enthalpies and barriers as does MP2 in isolation. Comparison of the RMP2/6-31G(d) and URCCSD(T)/6-31G(d) reaction enthalpies and β-scission barriers for reactions 7.1*a*–7.1*d* (Table 7.1) shows marked differences, whereas the reverse (addition) barriers are in reasonable agreement with one another. This implies a poor performance of MP2 for the reactant alkoxyl radical species, suggested in the previous Chapter, and this needs to be borne in mind.

In some instances, there is a larger-than-expected difference between CBS-QB3 and G3X(MP2)-RAD. For example, the addition barriers differ by approximately 7–8 kJ mol^{-1} for all four reactions for which a comparison is possible. In these cases, the G3X(MP2)-RAD results tend to be in better agreement with those for the large-basis-set URCCSD(T) and UBD(T) calculations.

As we reported in the previous Chapter, B3-LYP/6-311+G(3df,2p) and BMK/6-311+G(3df,2p) offer cost-effective alternatives to the high-level methods for the calculation of reaction enthalpies and barriers. The results in Table 7.1 suggest that B3-LYP generally underestimates the C–C backbone β-scission barriers while BMK tends to overestimate them. Table 7.1 also includes MPWB1K/6-311+G(3df,2p) results, which show a more severe overestimation of the β-scission barriers than the other DFT

methods when compared with the G3X(MP2)-RAD values. Interestingly, the barriers for the reverse (addition) reactions show less sensitivity to level of theory.

Table 7.1: Reaction Enthalpies and Barriers (0 K, kJ mol^{-1}) Calculated with Various Theoretical Techniques for Six Model C–C Backbone β-Scission Reactions[a]

Reaction[a]	ΔH	$\Delta H^{\ddagger}_{scission}$	$\Delta H^{\ddagger}_{addition}$
Reaction 7.1*a*			
CBS-QB3	3.1[b]	24.4[b]	21.2[b]
G3X(MP2)-RAD	0.4[b]	29.4[b]	29.0[b]
UBD(T)/cc-pVTZ//B3[c]	5.0	32.2	28.7
URCCSD(T)/cc-pVTZ//B3[c]	3.2	32.2	29.0
UB3-LYP/6-311+G(3df,2p)//B3[c]	5.4[b]	23.7[b]	18.2[b]
UBMK/6-311+G(3df,2p)//B3[c]	13.1[b]	35.8[b]	22.6[b]
UMPWB1K/6-311+G(3df,2p)//B3[c]	28.6	48.0	19.3
URCCSD(T)/6-31G(d)	22.4	48.2	25.9
RMP2/6-31G(d)	–6.0	16.1	22.1
Reaction 7.1*b*			
CBS-QB3	–10.8[b]	19.2[b]	30.0[b]
G3X(MP2)-RAD	–13.6[b]	23.5[b]	37.1[b]
URCCSD(T)/cc-pVTZ//B3[c]	–11.9	24.1	36.1
UB3-LYP/6-311+G(3df,2p)//B3[c]	–14.1[b]	20.4[b]	34.5[b]
UBMK/6-311+G(3df,2p)//B3[c]	–7.2[b]	27.7[b]	34.9[b]
UMPWB1K/6-311+G(3df,2p)//B3[c]	7.5	40.2	32.7
URCCSD(T)/6-31G(d)	8.0	42.8	34.8
RMP2/6-31G(d)	–24.1	5.3	29.4
Reaction 7.1*c*			
CBS-QB3	15.1[b]	13.9[b]	–1.2[b]
G3X(MP2)-RAD	13.4[b]	19.1[b]	5.6[b]
URCCSD(T)/cc-pVTZ//B3[c]	13.7	19.8	6.0
UB3-LYP/6-311+G(3df,2p)//B3[c]	5.1[b]	14.1[b]	9.3[b]
UBMK/6-311+G(3df,2p)//B3[c]	14.5[b]	22.0[b]	7.5[b]
UMPWB1K/6-311+G(3df,2p)//B3[c]	27.3	30.5	3.2
URCCSD(T)/6-31G(d)	29.7	32.0	2.3
RMP2/6-31G(d)	–3.4	–7.8	–4.4
Reaction 7.1*d*			
CBS-QB3	–45.8[b]	–4.2[b]	41.6[b]
G3X(MP2)-RAD	–51.9[b]	–2.9[b]	49.1[b]
UB3-LYP/6-311+G(3df,2p)//B3[c]	–60.8[b]	–1.1[b]	59.7[b]
UBMK/6-311+G(3df,2p)//B3[c]	–55.8[b]	3.2[b]	59.1[b]
UMPWB1K/6-311+G(3df,2p)//B3[c]	–45.8	9.1	54.9
URCCSD(T)/6-31G(d)	–40.3	9.9	50.2
RMP2/6-31G(d)	–85.9	–43.7	42.2
Reaction 7.1*e*			
G3X(MP2)-RAD	–37.4	–0.2	37.6
UB3-LYP/6-311+G(3df,2p)//B3[c]	–31.6	–2.5	34.1
Reaction 7.1*f*			
UB3-LYP/6-311+G(3df,2p)//B3[c]	–25.0	0.2	24.8

[a] See Figure 7.1 for details of the reactions. [b] Previous Chaapter. [c] Calculation have been performed on UB3-LYP/6-31G(d) (abbreviated B3) optimized geometries and include scaled (by 0.9806)[3] zero-point vibrational energy.

7.3.2 C–N β-Scission Reactions

Table 7.2 presents reaction enthalpies and barriers for six model peptide-backbone alkoxyl radicals shown in Figure 7.2 reacting via the C–N β-scission pathway. Again, the six reactions have been chosen to incrementally build up the peptide about the radical center in order to increasingly approximate an Ala radical center.

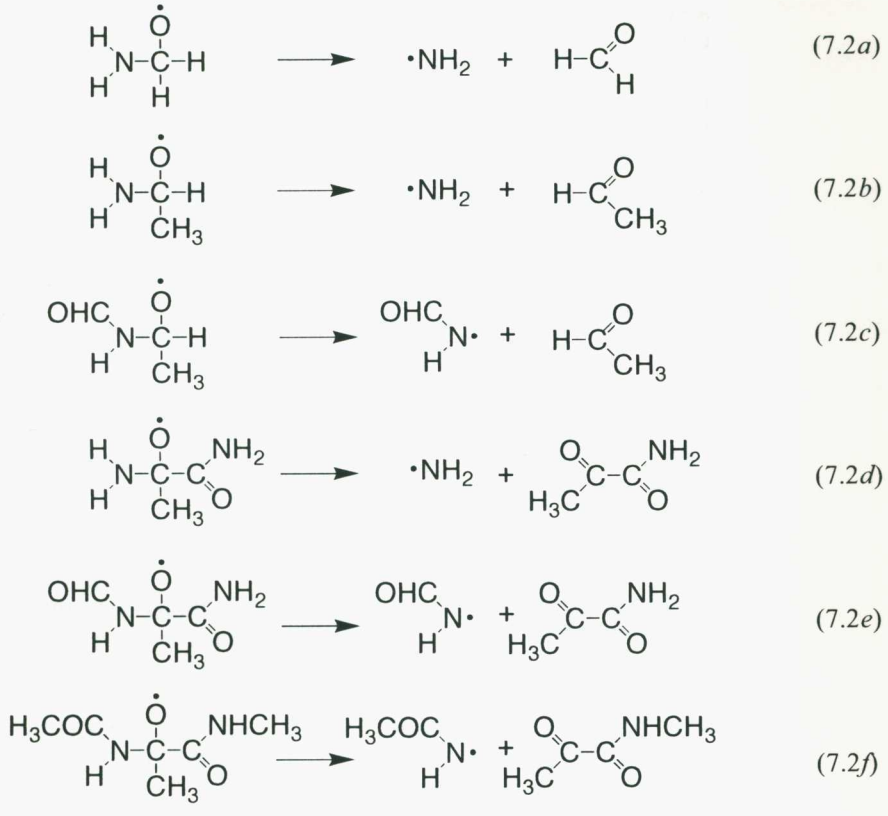


Figure 7.2: Six model peptides for the study of the β-scission of the C–N backbone bond from an α-C-centered alkoxyl radical.

Table 7.2: Reaction Enthalpies and Barriers (0 K, kJ mol^{-1}) Calculated with Various Theoretical Techniques for Six Model C–N Backbone β-Scission Reactions[a]

Reaction [a]	ΔH	$\Delta H^{\ddagger}_{scission}$	$\Delta H^{\ddagger}_{addition}$
Reaction 7.2*a*			
CBS-QB3	47.2	59.6	12.3
G3X(MP2)-RAD	42.7	60.4	17.7
UBD(T)/cc-pVTZ//B3[b]	44.7	60.4	15.7
URCCSD(T)/cc-pVTZ//B3[b]	43.9	59.6	15.7
UB3-LYP/6-311+G(3df,2p)//B3[b]	48.1	61.0	12.9
UBMK/6-311+G(3df,2p)//B3[b]	59.0	74.6	15.6
UMPWB1K/6-311+G(3df,2p)//B3[b]	73.1	85.0	11.9
Reaction 7.2*b*			
CBS-QB3	28.6	46.0	17.4
G3X(MP2)-RAD	21.9	46.1	24.2
UBD(T)/cc-pVTZ//B3[b]	23.3	46.3	23.0
URCCSD(T)/cc-pVTZ//B3[b]	22.2	45.2	23.0
UB3-LYP/6-311+G(3df,2p)//B3[b]	22.8	47.3	24.5
UBMK/6-311+G(3df,2p)//B3[b]	31.5	57.7	26.2
UMPWB1K/6-311+G(3df,2p)//B3[b]	42.9	65.3	22.4
Reaction 7.2*c*			
CBS-QB3	70.4	90.8	20.4
G3X(MP2)-RAD	66.3	93.5	27.2
URCCSD(T)/cc-pVTZ//B3[b]	65.2	91.5	26.2
UB3-LYP/6-311+G(3df,2p)//B3[b]	56.9	86.9	30.0
UBMK/6-311+G(3df,2p)//B3[b]	72.9	103.2	30.3
UMPWB1K/6-311+G(3df,2p)//B3[b]	88.1	114.9	26.8
Reaction 7.2*d*			
CBS-QB3	24.5	41.1	16.5
G3X(MP2)-RAD	–	–	23.7
UB3-LYP/6-311+G(3df,2p)//B3[b]	10.6	42.3	31.7
UBMK/6-311+G(3df,2p)//B3[b]	19.5	52.9	33.4
UMPWB1K/6-311+G(3df,2p)//B3[b]	29.4	56.0	26.6
Reaction 7.2*e*			
G3X(MP2)-RAD	56.4	98.4	42.0
UB3-LYP/6-311+G(3df,2p)//B3[b]	43.3	81.8	38.5
Reaction 7.2*f*			
UB3-LYP/6-311+G(3df,2p)//B3[b]	32.7	82.0	49.3

[a] See Figure 7.2 for details of the reactions. [b] Calculations have been performed on UB3-LYP/6-31G(d) (abbreviated B3) optimized geometries and include scaled (by 0.9806)[3] zero-point vibrational energy.

As found previously for the C–C β-scission pathway, the high-level composite methods, G3X(MP2)-RAD and CBS-QB3, generally give results in reasonable agreement with one another. Table 7.2 also includes results with URCCSD(T)/cc-pVTZ//B3-LYP/6-31G(d) for reactions 7.2*a*–7.2*c*, and results with UBD(T)/cc-pVTZ//B3-LYP/6-31G(d) for reactions 7.2*a* and 7.2*b*. The good agreement with the results from the composite methods suggests, as for the C–C β-scission pathway (Table 7.1), that the composite methods do not suffer from problems associated with MP2 calculations. Where there are discrepancies between G3X(MP2)-RAD and CBS-QB3, the CCSD(T) and BD(T) results tend to be closer to those obtained from G3X(MP2)-RAD.

The DFT functionals, B3-LYP, BMK and MPWB1K in association with the 6-311+G(3df,2p) basis set, were also assessed for the calculation of reaction enthalpies and barriers of model peptide-backbone alkoxyl radicals undergoing the C–N β-scission pathway. As for the C–C β-scission pathway, B3-LYP and BMK offer cost-effective alternatives for achieving good accuracy. The results of Table 7.2 suggest that BMK tends to overestimates C–N β-scission barriers as does B3-LYP, albeit to a lesser extent.

7.3.3 C–R β-Scission Reaction

Table 7.3 presents enthalpies and barriers for five peptide-backbone alkoxyl radicals undergoing reaction via a C–R β-scission pathway, as shown in Figure 7.3. Observations on the sensitivity of the reaction enthalpies and barriers to level of theory are similar to those made for the C–C and C–N β-scission pathways and therefore will not be discussed in detail.

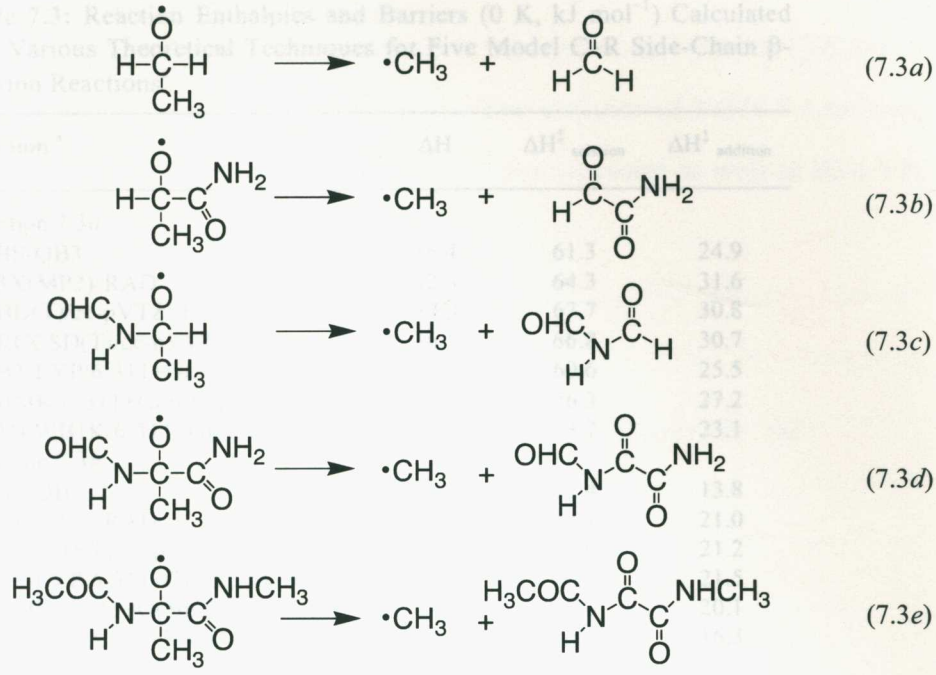


Figure 7.3: Five model peptides for the study of the β-scission of the C–R side-chain bond from an α-C-centered alkoxyl radical.

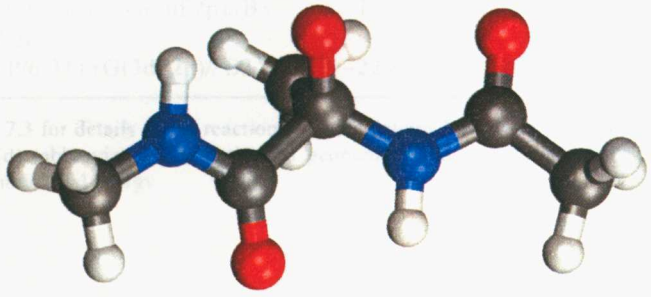


Figure 7.4: B3-LYP/6-31G(d) optimized structure of the peptide-backbone alkoxyl-radical reactant of reactions 7.1*f*, 7.2*f* and 7.3*e* ($CH_3CONH–C(O•)CH_3–CONHCH_3$, see Figures 7.1, 7.2 and 7.3) which represents our largest model system.

Table 7.3: Reaction Enthalpies and Barriers (0 K, kJ mol^{-1}) Calculated with Various Theoretical Techniques for Five Model C–R Side-Chain β-Scission Reactions[a]

Reaction [a]	ΔH	$\Delta H^{\ddagger}_{scission}$	$\Delta H^{\ddagger}_{addition}$
Reaction 7.3*a*			
CBS-QB3	36.4	61.3	24.9
G3X(MP2)-RAD	32.6	64.3	31.6
UBD(T)/cc-pVTZ//B3[b]	37.0	67.7	30.8
URCCSD(T)/cc-pVTZ//B3 [b]	36.1	66.8	30.7
UB3-LYP/6-311+G(3df,2p)//B3 [b]	35.2	60.6	25.5
UBMK/6-311+G(3df,2p)//B3	49.1	76.3	27.2
UMPWB1K/6-311+G(3df,2p)//B3 [b]	50.7	73.7	23.1
Reaction 7.3*b*			
CBS-QB3	38.7	52.4	13.8
G3X(MP2)-RAD	33.1	54.1	21.0
URCCSD(T)/cc-pVTZ//B3	35.7	57.0	21.2
UB3-LYP/6-311+G(3df,2p)//B3 [b]	38.5	60.0	21.5
UBMK/6-311+G(3df,2p)//B3 [b]	47.8	68.0	20.1
UMPWB1K/6-311+G(3df,2p)//B3 [b]	60.3	76.6	16.3
Reaction 7.3*c*			
CBS-QB3	–4.3	37.2	41.5
G3X(MP2)-RAD	–8.9	42.1	50.9
URCCSD(T)/cc-pVTZ//B3	–8.6	42.2	50.8
UB3-LYP/6-311+G(3df,2p)//B3 [b]	–15.8	35.1	50.9
UBMK/6-311+G(3df,2p)//B3 [b]	–7.8	41.1	48.9
UMPWB1K/6-311+G(3df,2p)//B3 [b]	10.4	55.1	44.7
Reaction 7.3*d*			
G3X(MP2)-RAD	–30.1	29.2	59.3
UB3-LYP/6-311+G(3df,2p)// B3 [b]	–29.9	31.9	61.8
UBMK/6-311+G(3df,2p)//B 3 [b]	–19.8	38.8	58.6
UMPWB1K/6-311+G(3df,2p)//B3 [b]	–18.2	36.2	54.4
Reaction 7.3*e*			
UB3-LYP/6-311+G(3df,2p)// B3 [b]	–22.8	38.3	61.1

[a] See Figure 7.3 for details of the reactions. [b] Calculations have been performed on UB3-LYP/6-31G(d) (abbreviated B3) optimized geometries and include scaled (by 0.9806)[3] zero-point vibrational energy.

7.3.4 Performance Evaluation

To quantify the observations regarding the performance of the various levels of theory, Table 7.4 presents the mean absolute deviation (MAD), mean deviation (MD), and largest deviation (LD) from the G3X(MP2)-RAD results for each of these procedures in calculating the reaction enthalpies and barriers for all the reactions of the

β-scission pathways. These statistics indicate that UB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d) provides an attractive cost-effective alternative to G3X(MP2)-RAD in the prediction of reaction barriers and enthalpies. The statistics of Table 7.4 indicate that, at least for the present systems, BMK appears not to perform as well as B3-LYP, in contrast to previous observations.

Table 7.4: Performance of Various Theoretical Methods for the Calculation of β-Scission Reaction Enthalpies and Barriers[a]

Theoretical Procedure	ΔH	$\Delta H^{\ddagger}_{scission}$	$\Delta H^{\ddagger}_{addition}$
MAD[b]			
CBS-QB3	4.3	2.9	7.2
UB3-LYP/6-311+G(3df,2p)//B3[b]	5.6	4.8	4.0
UBMK/6-311+G(3df,2p)//B3[b]	9.0	8.3	3.2
UMPWB1K/6-311+G(3df,2p)//B3[b]	19.9	16.0	5.0
MD[c]			
CBS-QB3	+4.3	–2.9	–7.2
UB3-LYP/6-311+G(3df,2p)//B3[b]	–2.6	–2.6	–0.4
UBMK/6-311+G(3df,2p)//B3[b]	+8.3	+8.1	–0.2
UMPWB1K/6-311+G(3df,2p)//B3[b]	+19.9	+16.0	–3.9
LD[d]			
CBS-QB3	+6.7	–5.2	–9.4
UB3-LYP/6-311+G(3df,2p)//B3[b]	–13.1	–16.6	–10.8
UBMK/6-311+G(3df,2p)//B3[b]	+16.5	+14.2	+10.0
UMPWB1K/6-311+G(3df,2p)//B3[b]	+30.4	+24.6	–9.7

[a] Calculations have been performed on UB3-LYP/6-31G(d) (abbreviated B3) optimized geometries and include scaled (by 0.9806)[3] zero-point vibrational energy. Energy units are kJ mol^{-1}. [b] Mean absolute deviation from G3X(MP2)-RAD. [c] Mean deviation from G3X(MP2)-RAD. [d] Largest deviation from G3X(MP2)-RAD.

7.3.5 Model Size Considerations

The results for the model systems of Figures 7.1, 7.2 and 7.3 for each of the pathways provide a means of assessing what system size is suitable for reliably representing the cleavage reactions. It has been noted previously[4] that convergence of calculated RSEs for α-C-centered radicals is achieved once the backbone formally

contains an amide functionality on both sides of the radical center. The results given in Tables 7.1–7.3 indicate that the backbone needs to be of *at least* this size to obtain convergence in the enthalpies and barriers for β-scission reactions. The final two reactions in each sequence of models involve the formal substitution of hydrogen with methyl capping groups. For the C–C β-scission pathway, this substitution is accompanied by a small change in the β-scission barrier height (of 2.7 kJ mol^{-1} between 7.1*e* and 7.1*f* with UB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d)). The changes in the reaction enthalpy and the addition barrier are slightly greater at 6.6 and 9.3 kJ mol^{-1}, respectively. In an equivalent formal substitution for the models of the C–N β-scission pathway, the differences between reactions 7.2*e* and 7.2*f* are similar: 0.2 kJ mol^{-1} for $\Delta H^{\ddagger}_{scission}$ and 10.6 and 10.8 kJ mol^{-1} for ΔH and $\Delta H^{\ddagger}_{addition}$, respectively. For the C–R β-scission pathway, the changes in going from reaction 7.3*d* to 7.3*e* are 6.4, 7.1 and 0.7 kJ mol^{-1} for $\Delta H^{\ddagger}_{scission}$, ΔH and $\Delta H^{\ddagger}_{addition}$, respectively.

7.3.6 Competing Pathways C–C, C–N, and C–R

The calculations in this section allow us to compare the barriers for the three possible β-scission reactions of peptide-backbone alkoxyl radicals at an Ala residue. These indicate an order of preference C–C > C–R > C–N; this preference for the C–C β-scission pathway is in accord with previous experimental[5,6] and theoretical data.[7] For the largest model systems, the thermochemical parameters indicate that the C–C backbone β-scission reaction is essentially barrierless (0.2 kJ mol^{-1}). The β-scission reaction leading to C–CH_3 side-chain fragmentation has a higher barrier of 38.3 kJ mol^{-1}, while the C–N backbone β-scission reaction has a still higher barrier (82.0 kJ mol^{-1}).

7.3.7 Side-Chain Dependence

Our discussion up to now has focussed on the β-scission reactions of an alkoxyl radical located at an Ala residue of a peptide. Calculated barriers indicate that the C–C

pathway is favored in this case. Experimental product analysis of a Gly-containing peptide found that the C–C pathway predominates also for this residue.[6] However, because the rates of β-scission reactions of alkoxyl radicals are sensitively dependent on the stability of the radical that is formed,[8] and because more highly stabilized radical products could potentially be formed for other side chains, this may result in the C–R β-scission pathway becoming competitive with the C–C β-scission pathway. In order to probe these ideas, we begin by examining whether a correlation exists between the calculated radical stabilization energy (RSE) of the fragment radical on the one hand, and the barriers and enthalpies of the side-chain β-scission reactions of α-C-centered alkoxyl radicals on the other.

7.3.8 Radical Stabilization Energies

Table 7.5 shows the radical stabilization energies calculated with G3X(MP2)-RAD of radicals formed from the homolytic bond cleavage of the α-C–R bond, where R is the side chain. The Table compares the RSEs for the R groups of all the natural amino acids except for Gly and Pro.

Table 7.5: Radical Stabilization Energies of Side-Chain Fragment Radicals (•R) Formed from the Homolytic Cleavage of the α-C–R Bond (R = Side Chain) of the Natural Amino Acids[a]

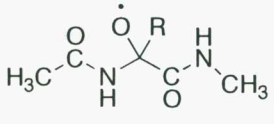
Ala[b]	Glu[c]	Gln[d]	Leu[e]	Met[f]	Lys[g]	Arg[h]	Asp[i]	Ile[j]
0.0	6.1	11.2	11.5	12.8	13.1	13.5	22.4	24.1
Asn[k]	**Val[l]**	**Ser[m]**	**Thr[n]**	**Cys[o]**	**Phe[p]**	**Tyr[q]**	**His[r]**	**Trp[s]**
24.4	25.5	33.5	37.3	39.4	59.5	61.5	61.9	62.2

[a] Calculated with G3X(MP2)-RAD at 0 K in units of kJ mol^{-1}. [b] •CH_3. [c] •CH_2CH_2COOH. [d] •$CH_2CH_2CONH_2$. [e] •$CH_2CH(CH_3)_2$. [f] •$CH_2CH_2SCH_3$. [g] •$CH_2CH_2CH_2CH_2NH_2$. [h] •$CH_2CH_2CH_2NHC(NH)NH_2$. [i] •CH_2COOH. [j] •$CH(CH_3)CH_2CH_3$. [k] •CH_2CONH_2. [l] •$CH(CH_3)_2$. [m] •CH_2OH. [n] •$CH(OH)CH_3$. [o] •CH_2SH. [p] •$CH_2–C_6H_5$. [q] •$CH_2–C_6H_4–OH$. [r] •CH_2–"imidazole". [s] •CH_2–"indole".

The Ala side-chain fragment radical is $\bullet CH_3$, which has an RSE of zero by definition. The remaining amino acid side chains give RSEs that are all positive, i.e., the radicals are relatively more stable that the $\bullet CH_3$ of alanine. The ordering of RSEs for the different R groups is Ala < Glu < Gln ~ Leu ~ Met ~ Lys ~ Arg < Asp ~ Ile ~ Asn ~ Val < Ser ~ Thr ~ Cys < Phe ~ Tyr ~ His ~ Trp.

We expect that the rates of reaction for side-chain cleavage will increase as the RSE of the side-chain fragment increases. In order to test this hypothesis, we begin by calculating barriers for the three β-scission pathways of model systems containing an Ala, Leu, Val or Phe residue. These residues were chosen to span the range of RSEs given in Table 7.5, thus providing a sufficient spectrum of enthalpies and barriers to give reasonable insights into the side-chain dependence of the reaction rates. We note that preliminary conformational studies of model systems containing other residues such as Asp and Asn, showed that intramolecular hydrogen bonding made it difficult to choose an appropriate conformation that represented a realistic protein-bound peptide fragment. This would have hampered a consistent comparison of β-scission enthalpies and barriers. In contrast, the Ala, Leu, Val and Phe residues all have alkyl side chains that do not make strong internal hydrogen bonds to the backbone amide functionalities.

Based on the results of the previous section, the model systems that we use to investigate the β-scission reactions include the backbone that incorporates amide linkages with methyl capping groups formally substituted on each side of the α-C, i.e.,



This should represent a reasonable model for estimating rates of reaction in a protein-bound peptide. Unfortunately, this system is already sufficiently large to prevent the

use of methods such as G3X(MP2)-RAD with our currently available resources. We have therefore used B3-LYP/6-311+G(3df,2p)//B3-LYP/6-31G(d), which our assessment study suggests should provide a reliable cost-effective alternative to G3X(MP2)-RAD.

7.3.9 Calculated Rate Parameters

Table 7.6 presents the Arrhenius activation energies (E_a), pre-exponential factors (given as log A) and rate constants (given as log k) for each of the fragmentation pathways, obtained on the basis of B3-LYP/6-311+G(3df,2p)//B3-LYP/6-31G(d) calculations using the harmonic approximation. Also included are the calculated reaction enthalpies at 0 K (ΔH). The species investigated include model peptides containing an Ala, Leu, Val or Phe residue.

Table 7.6: Calculated Arrhenius Parameters[a] and Reaction Enthalpies for the Three Possible β-Scission Reactions of an α-C-Centered Alkoxyl Radical on a Model Peptide Containing Specific Amino Acid Residues

Pathway	Ala	Leu	Val	Phe
C–C β-Scission				
E_a	0.9	11.5	3.7	10.7
log A	14.1	13.7	13.4	13.9
log k	13.2	11.7	12.8	12.0
ΔH	–25.0	–33.7	–31.9	–44.0
C–N β-Scission				
E_a	85.7	101.1	96.7	89.7
log A	14.8	13.2	13.3	14.4
log k	–0.7	–4.5	–3.7	–1.3
ΔH	32.7	24.4	22.1	49.0
C–R β-Scission				
E_a	39.5	22.6	9.0	7.7
log A	14.3	13.2	13.0	13.3
log k	6.6	9.3	11.4	11.9
ΔH	–22.8	–45.4	–61.9	–79.8

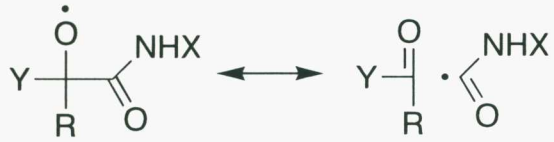
[a] Calculated within the harmonic approximation at 298 K using UB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d).

We note in the first place that the A factors for the three pathways are all very similar, with log A lying between 13.0 and 14.8, which is consistent with results of other theoretical studies of alkoxyl β-scission reactions.[9] Although our calculated rates are only expected to be of moderate accuracy because of the theoretical procedures and peptide models used, [10,11] they do provide insight into the dependence of β-scission rates on residue type. The C–C β-scission rates are all very fast and essentially independent of the nature of the side chain, with log k lying between 11.7 and 13.2. The C–N β-scission rates on the other hand are all very slow. There is a variation in rate constants as a function of side chain of four orders of magnitude, with log k ranging from –0.7 to –4.5. The rate constants for the side-chain cleavage show an even stronger dependence on the residue type, with the rate constant increasing rapidly as the RSE of the side-chain fragment increases. The calculated RSEs provide an approximate basis on which to select residues that are most likely to have competitive C–R β-scission reactions. For Val, the rate of β-scission of the side chain is 1.4 orders of magnitude less than the C–C β-scission reaction of the backbone. Thus, side-chain fragment radicals with RSEs greater than approximately 30 kJ mol^{-1}, are likely to have fast side-chain β-scission reactions, with log k ~ 12 ± 1. As the RSE increases above 30 kJ mol^{-1}, it is likely that the side-chain β-scission reaction will increasingly compete with C–C β-scission of the backbone. With an RSE of 59.5 kJ mol^{-1}, the fragment produced from the C–R β-scission reaction of Phe is associated with a side-chain β-scission rate constant that is approximately the same as that for the C–C β-scission reaction. Fragments that have RSEs greater than 59.5 kJ mol^{-1} are expected to provide even stronger candidates for competitive side-chain cleavage. The natural amino acids that give rise to side-chain fragments with calculated RSEs greater than 30 kJ mol^{-1} include Ser, Thr, Cys, Phe, Tyr, His and Trp. Attempts to provide direct experimental evidence for the theoretical prediction of competitive side-chain cleavage have not yet been successful.

7.3.10 Why Are the C–C β-Scission Reactions Fast?

As indicated in the previous section, side-chain fragment radicals that have large RSEs result in C–R β-scission reactions with large rate constants. Similar considerations apply to the C–C β-scission reactions, where the radical product (•$CONHCH_3$) has a large RSE of 37.4 kJ mol^{-1},[12] consistent with the large calculated rates. Significantly smaller rate constants are found for β-scission reactions of non-peptide-related alkoxyl radicals. For instance, in an experimental study of 11 simple alkoxyl radicals,[13] the largest rate constant was found to be approximately 10^5 s^{-1} for the decomposition of 2-methyl-2-butoxyl ($CH_3CH_2(CH_3)_2CO$•) radical to give the ethyl radical, consistent with the RSE for the ethyl radical of just 16.3 kJ mol^{-1}.

While the RSE provides a quick and convenient indication of how fast the reaction might go, it is a less direct measure than the reaction enthalpy. The C–C β-scission reactions of the present Chapter are all exothermic, with reaction enthalpies ranging from –25.0 to –44.0 kJ mol^{-1} (see Table 7.6). This shows that the C–C bond adjacent to the alkoxyl radical is weak and easily ruptured, which is reflected not only in the β-scission rate but also in long reactant bond lengths. For instance, in the model system for Phe, the reactant C–C bond length is 1.606 Å compared with normal C–C lengths of ~1.54 Å. This can be rationalized in terms of a strong contribution from the non-bonded resonance hybrid to the overall reactant wave function, i.e.,



The effect of stabilized radical products on the enthalpy of reaction and hence the rate of reaction can be more clearly seen for the series of C–R β-scission reactions.

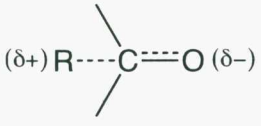
The rates of these reactions vary by 6 orders of magnitude depending on the stability of the radical that is formed, and follow the reaction enthalpies (–22.8 to –79.8 kJ mol^{-1}, see Table 7.6) as well as the RSEs (0.0 to 39.5 kJ mol^{-1}, Table 7.5). Again, the larger the rate constant, the weaker the C–R bond and the more easily it is ruptured. This is clearly demonstrated by the progression in the C–R bond lengths for the different residues investigated in this Chapter, i.e., Ala 1.538 Å, Leu 1.583 Å, Val 1.603 Å, and Phe 1.600 Å.

7.3.11 Empirical Rate Relationships

Experimental studies have led to the derivation of empirical relationships to predict Arrhenius parameters of the β-scission reactions of alkoxyl radicals. For example, Choo and Benson[14] have proposed that the activation energy E_a is given by:

$$E_a = a + b\Delta H$$

where ΔH is the reaction enthalpy, b is a constant with the value 1.58, and a is a parameter that depends on the ionization energy of the alkyl radical fragment: $a = 8.8 \times IE - 25.9$, where the IEs are in eV, and we have converted all the remaining energy units to kJ mol^{-1}. Since the initial proposal of this equation,[14] improved values for the parameters have been developed by a number of research groups. For example, Atkinson[15] refined the parameters with a larger data set and proposed $a = 10.0 \times IE - 33.9$ and $b = 0.36$. The large dependence on the ionization energy reflects the tight transition structure with significant polarization, i.e.,



Other researchers have developed structure–activity relationships (SARs) that only depend on the ionization energy of the products[13] or on the nature of the alkoxyl radical.[16]

Table 7.7: Comparison of Empirical and Directly Calculated Arrhenius Activation Energies (E_a, kJ mol^{-1})

Pathway	IE	E_a(emp)[a]	E_a(calc)
C–C β-Scission			
Ala	6.9	26.7	0.9
Leu	6.9	23.6	11.5
Val	6.9	24.2	3.7
Phe	6.9	19.9	10.7
C–N β-Scission			
Ala	9.7	75.0	84.1
Leu	9.7	72.0	101.1
Val	9.7	71.1	96.7
Phe	9.7	80.9	89.7
C–R β-Scission			
Ala	9.8	56.3	39.5
Leu	7.3	22.6	22.6
Val	7.5	18.9	9.0
Phe	7.3	10.6	7.7

[a] Derived from the calculated reaction enthalpy (ΔH, kJ mol^{-1}, Table 7.6) and the ionization energy (IE, eV) of the radical fragment using the empirical relationship of Atkinson.[15]

We have investigated the applicability of a number of these correlations to the alkoxyl radicals of the present Chapter, using theoretical ionization energies (calculated with CBS-QB3) and enthalpies (calculated with UB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d)). In general, the empirical methods give poor agreement with the directly calculated kinetic parameters for the alkoxyl radicals of this Chapter. For example, although the empirical activation energies derived from the Atkinson parameters (Table 7.7) show the general pattern of the calculated values, i.e., high for C–N β-scission reactions and lower for C–C backbone and C–R side-chain β-scission reactions, deviations between the explicitly-calculated and empirically-estimated activation energies can be large. The best agreement is seen for the C–R side-chain β-

scission reactions. This may be attributed to the radical fragments in these reactions being alkyl-type radicals, and the empirical relationships being derived from alkyl-substituted alkoxyl radicals. However, the empirical relationships do not provide a practical alternative means for deriving activation energies from computations, as the saving in computational cost is minimal and the resultant activation energies are generally in poor agreement with the directly calculated values.

7.4 Conclusions

Arrhenius parameters have been calculated for models of the three types of β-scission reactions involving an alkoxyl radical located at the α-carbon position of a peptide. The parameters have been obtained for the Ala, Leu, Val and Phe residues to illustrate the effect of variation in the side chain. We find that the rates of C–C backbone β-scission are all fast, with rate constants on the order of 10^{12} s^{-1}, consistent with previous EPR data.[5] The C–N backbone β-scission reactions are all found to have very small rate constants that range from $10^{-4.5}$ to $10^{-0.7}$ s^{-1}. The side-chain C–R β-scission reactions have a range of rate constants that depend on the side chain (R). The rate constants are found to increase as the stability of the daughter radical (•R) resulting from the side-chain β-scission increases. We predict that alkoxyl radical β-scission reactions from an α-C-centered radical on a peptide involving side-chain daughter radicals that have radical stabilization energies greater than approximately 30 kJ mol^{-1} are likely to compete with the backbone C–C β-scission reaction. The residues identified in this Chapter that are likely to display this behavior include Ser, Thr, Cys, Phe, Tyr, His and Trp. Experiments are currently in progress to see if this prediction can be verified.

7.5 References

(1) Rauk, A.; Boyd, R. J.; Boyd, S.; Henry, D. J.; Radom, L *Can. J. Chem.* **2003**, *81*, 431.

(2) Huang, M. L.; Rauk, A. *J. Phys. Org. Chem.* **2004**, *17*, 777.

(3) Scott, A. P.; Radom, L. *J. Phys. Chem.* **1996**, *100*, 16502.

(4) Wood, G. P. F.; Moran, D.; Jacob, R.; Radom, L. *J. Phys. Chem. A* **2005**, *109*, 6318.

(5) Davies, M. J. *Arch. Biochem. Biophys.* **1996**, *336*, 163.

(6) Mortimer, A.; Easton, C. J. *unpublished data.*

(7) Huang, M. L.; Rauk, A. *J. Phys. Org. Chem.* **2004**, *17*, 777.

(8) Gray, P.; Williams, A. *Chem. Rev.* **1959**, *59*, 239.

(9) Fittschen, C.; Hippler, H.; Viskolcz, B. *Phys. Chem. Chem. Phys.* **2000**, *2*, 1677.

(10) A variation of approximately 5 kJ mol^{-1} in the activation energy results in a change in the calculated rate constant at 298 K of an order of magnitude. Based on the comparison of UB3-LYP/6-311+G(3df,2p)//UB3-LYP/6-31G(d) results with those of high-level composite methods, we expect our calculated activation energies to have an uncertainty of about ±10 kJ mol^{-1}. Thus, when we refer to our calculated rates as being "moderately accurate", we mean that they should be accurate to within about two orders of magnitude. In has previously been found[11] that A factors are much less sensitive than E_as to the level of theory used in their calculation. It is therefore the accuracy of the activation energies that is dominant in determining the accuracy of the calculated rates.

(11) See for example: Gómez-Balderas, R.; Coote, M. L.; Henry, D. J.; Radom, L. *J. Phys. Chem. A* **2004**, *108*, 2874.

(12) The BDE has been determined through G3X(MP2)-RAD calculations.

(13) Johnson, D.; Carr, S.; Cox, A. R., *Phys. Chem. Chem. Phys.* **2005**, *7*, 2182.

(14) Choo, K. Y.; Benson, S. W. *Int. J. Chem. Kinet.* **1981**, *13*, 833.

(15) R. Atkinson, *Int. J. Chem. Kinet.* **1997**, *29*, 99.

(16) Peeters, J.; Fantechi, G.; Vereechken, L. *J. Atmos. Chem.* **2004**, *48*, 59.

APPENDIX 1

Tables of Heats of Formation, Ionization Energies and Electron Affinities for the G2/97 set

Table A1.1

Comparison of Calculated and Experimental Heats of Formation (298 K, kJ mol^{-1}) for Closed-Shell Species

		$\Delta H_f°$(theory)[b]		deviation (theory – expt)[e]	
Molecule	$\Delta H_f°$(expt)[a]	CBS-QB3[c]	ROCBS-QB3[c,d]	CBS-QB3[c]	ROCBS-QB3[c,d]
G2-1 test set					
LiH	139.3±1.3	140.6±1	140.5±1	1.3±1.6	1.2±1.6
CH_2 ($^1A_1$)	430.1±4	429.3±0.5	429.7±0.5	–0.8±4	–0.4±4
CH_4 (methane)	–74.5±0.4	–74.1±0.5	–73.6±0.5	0.4±0.6	0.9±0.6
NH_3	–46.0±0.4	–44.2±0.1	–44.3±0.1	1.8±0.4	1.7±0.4
H_2O	–241.83±0.04	–242.6±0.1	–241.6±0.1	–0.8±0.1	0.2±0.1
HF	–272.4±0.7	–274.5±0.3	–273.7±0.3	–2.1±0.8	–1.3±0.8
SiH_2 ($^1A_1$)	272.8±5	265.3±1.6	265.9±1.6	–7.5±5	–6.9±5
SiH_4	34.3±1.5	25.5±1.6	26.2±1.6	–8.8±2.2	–8.1±2.2
PH_3	5.4±1.7	1.6±1	1.2±1	–3.8±2	–4.2±2
H_2S	–20.5±0.5	–23.4±0.2	–21.6±0.2	–2.9±0.5	–1.1±0.5
HCl	–92.5±0.2	–93.8±0.01	–92.2±0.01	–1.3±0.2	0.3±0.2
Li_2	215.9±3	215.5±2	215.4±2	–0.4±4	–0.5±4
LiF	–335.1±8	–334.3±1.3	–333.5±1.3	0.8±8	1.6±8
C_2H_2 (acetylene)	228.2±0.7[f]	234.7±1.0	235.6±1.0	6.7±1.2	7.6±1.2
C_2H_4 (ethylene)	52.3±0.3	56.1±1.0	57.0±1.0	3.8±1.0	4.7±1.0
C_2H_6 (ethane)	–83.7±0.3	–82.0±1.0	–81.1±1.0	1.7±1.0	2.6±1.0
HCN	128.9±4	132.8±0.6	133.1±0.6	3.9±4	4.2±4
CO	–110.5±0.2	–111.3±0.6	–109.8±0.6	–0.8±0.6	0.7±0.6
H_2CO	–108.8±0.5	–113.0±0.6	–111.5±0.6	–4.2±0.8	–2.7±0.8
CH_3OH	–201.7±0.5	–203.0±0.6	–201.5±0.6	–1.3±0.8	0.2±0.8
N_2	0	2.8±0.1	2.6±0.1	2.8±0.1	2.6±0.1
N_2H_4	95.4±0.5	99.1±0.1	98.8±0.1	3.7±0.5	3.4±0.5
HOOH	–136.0±0.2	–138.9±0.2	–136.8±0.2	–2.9±0.3	–0.8±0.3
F_2	0	0.0±0.6	1.6±0.6	0.0±0.6	1.6±0.6
CO_2	–393.7±0.2	–398.3±0.7	–395.7±0.7	–4.6±0.7	–2.1±0.7
Na_2	142.3±0.2	141.0±1.4	140.8±1.4	–1.3±1.4	–1.5±1.4
P_2	143.5±2	142.2±2	141.5±2	–1.3±3	–2.0±3
Cl_2	0	–0.4±0.02	2.7±0.02	–0.4±0.02	2.7±0.02
NaCl	–182.4±1	–182.4±0.7	–181.0±0.7	0.0±1.2	1.4±1.2
SiO	–102.9±3	–100.8±1.7	–99.1±1.7	2.1±3.4	3.8±3.4
SC	279.9±0.4	279.9±0.7	282.2±0.7	0.0±0.8	2.3±0.8
SO	5.0±1.3	3.7±0.3	6.6±0.3	–1.3±1.3	1.6±1.3
ClF	–55.2±5	–51.4±0.3	–49.1±0.3	3.8±5	6.1±5
Si_2H_6	79.9±1.3	71.5±3.2	72.8±3.2	–8.4±3.5	–7.1±3.5

Table A1.1 (continued) Comparison of Calculated and Experimental Heats of Formation (298 K, kJ mol^{-1}) for Closed-Shell Species

		$\Delta H_f°$(theory)[b]		deviation (theory – expt)[c]	
Molecule	$\Delta H_f°$(expt)[a]	CBS-QB3[c]	ROCBS-QB3[c,d]	CBS-QB3[c]	ROCBS-QB3[c,d]
CH_3Cl	–82.0±2	–83.7±0.5	–81.7±0.5	–1.7±2	0.3±2
CH_3SH	–23.0±0.6	–23.8±0.7	–21.5±0.7	–0.8±0.9	1.5±0.9
HOCl	–74.5±2	–77.0±0.1	–74.4±0.1	–2.5±2	0.1±2
SO_2	–296.6±0.2	–293.3±0.4	–289.3±0.4	3.3±0.4	7.3±0.4
G2-2 test set					
BF_3	–1135.5±0.8	–1129.6±1.7	–1126.7±1.7	5.9±1.9	8.8±1.9
BCl_3	–404±2	–408.6±0.8	–403.4±0.8	–4.6±2.2	0.6±2.2
AlF_3	–1209±2.5[g]	–1217.4±5	–1214.3±5	–8.4±6	–5.3±6
$AlCl_3$	–585±2.9[g]	–610.9±4	–605.6±4	–25.9±5	–20.6±5
CF_4	–933.0±1	–941.4±1.7	–937.7±1.7	–8.4±2.0	–4.7±2.0
CCl_4	–97.1±3	–112.1±0.5	–105.5±0.5	–15.0±3.0	–8.4±3.0
COS	–138.5±2	–148.5±0.8	–145.2±0.8	–10.0±2.2	–6.7±2.2
CS_2	117.2±1	106.7±0.9	110.9±0.9	–10.5±1.3	–6.3±1.3
CF_2O	–609.2±4[h]	–612.1±1.2	–609.0±1.2	–2.9±4.2	0.2±4.2
SiF_4	–1615.0±0.8	–1613.4±2.8	–1609.6±2.8	1.6±2.9	5.4±2.9
$SiCl_4$	–662.7±0.8[g]	678.6±1.6	–671.7±1.6	–15.9±1.8	–9.0±1.8
N_2O	82.0±0.4	77.3±0.2	78.1±0.2	–4.7±0.4	–3.9±0.4
ClNO	51.9±0.4	52.9±0.2	55.3±0.2	1.0±0.4	3.4±0.4
NF_3	–132.2±1.3	–140.9±1.0	–138.6±1.0	–8.7±1.6	–6.4±1.6
PF_3	–958.6±3.8[g]	–946.9±1.9	–944.9±1.9	11.7±4.2	13.7±4.2
O_3	142.7±1.7	144.8±0.30	147.9±0.30	2.1±1.7	5.2±1.7
F_2O	24.7±1.7	20.9±0.5	23.6±0.5	–3.8±1.8	–1.1±1.8
ClF_3	–159±5.0[g]	–143.1±0.9	–139.2±0.9	15.9±5	19.8±5
C_2F_4	–672±6[i]	–682.5±2.2	–678.4±2.2	–10.5±6.4	–6.4±6.4
C_2Cl_4	12.6±2	–10.8±1.0	–3.7±1.0	–23.4±2	–16.3±2
CF_3CN	–499.8±1.2	–504.1±2.0	–500.9±2.0	–4.3±2.3	–1.1±2.3
C_3H_4 (propyne)	185.4±0.9	192.0±1.5	193.4±1.5	6.6±1.8	8.0±1.8
C_3H_4 (allene)	190.4±1	195.4±1.5	196.8±1.5	5.0±1.8	6.4±1.8
C_3H_4 (cyclopropene)	277±3	290.8±1.5	292.2±1.5	13.8±3.4	15.2±3.4
C_3H_6 (propene)	20.2±0.4	26.1±1.5	27.9±1.5	5.9±1.6	7.7±1.6
C_3H_6 (cyclopropane)	53.3±0.5	60.4±1.5	61.8±1.5	7.1±1.6	8.5±1.6
C_3H_8 (propane)	–104.5±0.3	–100.7±1.5	–99.4±1.5	3.8±1.6	5.1±1.6
C_4H_6 (butadiene)	110.0±0.8	120.0±2.0	121.8±2.0	10.0±2	11.8±2
C_4H_6 (2-butyne)	145.6±1	155.2±2.0	157.0±2.0	9.6±2	11.4±2
C_4H_6(methylene cyclopropane)	201±2	201.0±2.0	202.8±2.0	0.0±3	1.8±3
C_4H_6 (bicyclobutane)	217.0±0.8	231.6±2.0	233.4±2.0	14.6±2	16.4±2
C_4H_6 (cyclobutene)	156.7±1.5	170.9±2.0	172.7±2.0	14.2±3	16.0±3

Table A1.1 (continued) Comparison of Calculated and Experimental Heats of Formation (298 K, kJ mol^{-1}) for Closed-Shell Species

		$\Delta H_f°$(theory)[b]		deviation (theory – expt)[c]	
Molecule	$\Delta H_f°$(expt)[a]	CBS-QB3[c]	ROCBS-QB3[c,d]	CBS-QB3[c]	ROCBS-QB3[c,d]
C_4H_8 (cyclobutane)	28.4±0.5	34.7±2.0	36.5±2.0	6.3±2	8.1±2
C_4H_8 (isobutene)	–16.7±1.1	–9.2±2.0	–7.4±2.0	7.5±2	9.3±2
C_4H_{10} (butane)	–125.5±0.7	–120.5±2.1	–118.7±2.1	5.0±2	6.8±2
C_4H_{10} (isobutane)	–134.3±0.6	–128.4±2.1	–126.6±2.1	5.9±2	7.7±2
C_5H_8 (spiropentane)	185.4±0.8	193.3±2.5	195.6±2.5	7.9±2.7	10.2±2.7
C_6H_6 (benzene)	82.4±0.5	91.2±3.0	93.9±3.0	8.8±3	11.5±3
CH_2F_2	–450.6±0.8	–456.5±1.1	–454.4±1.1	–5.9±1.4	–3.8±1.4
CHF_3	–697.1±8	–703.4±1.4	–700.5±1.4	–6.3±8.1	–3.4±8.1
CH_2Cl_2	–95.7±0.8	–99.0±0.5	–95.5±0.5	–3.3±1.0	0.2±1.0
$CHCl_3$	–104.8±2	–110.8±0.5	–105.7±0.5	–6.0±2.1	–0.9±2.1
CH_3NH_2 (methyl amine)	–23.0±0.4	–20.0±0.6	–19.6±0.6	3.0±0.7	3.4±0.7
CH_3CN (methyl cyanide)	74.0±0.4	77.5±1.1	78.3±1.1	3.5±1.1	4.3±1.1
CH_3NO_2 (nitromethane)	–74.5±1	–81.9±0.8	–79.5±0.8	–7.4±1.3	–5.0±1.3
CH_3ONO (methyl nitrite)	–66.5±0.9	–73.1±0.8	–70.7±0.8	–6.6±1.2	–4.2±1.2
CH_3SiH_3 (methyl silane)	–29±4	–29±2.1	–28±2.1	0.0±5	1.1±5
HCOOH (formic acid)	–378.8±0.5	–382.6±0.7	–380±0.7	–3.8±0.9	–1.2±0.9
$HCOOCH_3$ (methyl formate)	–355.6±0.7	–364.8±1.2	–361.8±1.2	–9.2±1.4	–6.2±1.4
CH_3CONH_2 (acetamide)	–238.3±0.8	–235.3±1.2	–233.4±1.2	3.0±1.4	4.9±1.4
C_2H_4NH (aziridine)	126.5±0.9	130.4±1.1	131.2±1.1	3.9±1.4	4.7±1.4
NCCN (cyanogen)	306.7±0.7	312.9±1.1	313.5±1.1	6.2±1.3	6.8±1.3
$(CH_3)_2NH$ (dimethyl amine)	–18.5±0.4	–15.5±1.1	–14.7±1.1	3.0±1.2	3.8±1.2
$CH_3CH_2NH_2$ (ethyl amine)	–47.5±0.7	–46.5±1.1	–45.8±1.1	1.0±1.3	1.7±1.3
CH_2CO (ketene)	–47.7±2.5	–47.3±1.1	–45.8±1.1	0.4±2.7	1.9±2.7
C_2H_4O (oxirane)	–52.6±0.6	–54.7±1.1	–52.7±1.1	–2.1±1.3	–0.1±1.3
CH_3CHO (acetaldehyde)	–165.8±0.4	–165.8±1.1	–163.9±1.1	0.0±1.2	1.9±1.2
HCOCOH (glyoxal)	–211.9±0.8	–217.3±1.1	–214.4±1.1	–5.4±1.4	–2.5±1.4
CH_3CH_2OH (ethanol)	–234.8±0.2	–234.8±1.1	–232.9±1.1	0.0±1.2	1.9±1.2
CH_3OCH_3 (dimethyl ether)	–184.0±0.5	–187.8±1.1	–185.8±1.1	–3.8±1.2	–1.8±1.2
C_2H_4S (thiooxirane)	82.1±1.2	77.9±1.2	80.7±1.2	–4.2±1.7	–1.4±1.7
$(CH_3)_2SO$ (dimethyl sulfoxide)	–151.3±0.8	–148.4±1.3	–144.6±1.3	2.9±1.6	6.7±1.6
C_2H_5SH (ethane thiol)	–46.3±0.6	45.0±1.2	–42.3±1.2	1.3±1.4	4.0±1.4
CH_3SCH_3 (dimethyl sulfide)	–37.5±0.5	–38.3±1.2	–35.6±1.2	–0.8±1.3	1.9±1.3
CH_2=CHF (vinyl fluoride)	–138.8±1.7	–141.7±1.3	–140.0±1.3	–2.9±2.2	–1.2±2.2
C_2H_5Cl (ethyl chloride)	–112.1±0.5	–112.1±1.0	–109.6±1.0	0.0±1	2.5±1
CH_2=CHCl (vinyl chloride)	23±2[j]	25.1±1.0	27.5±1.0	2.1±2	4.5±2
CH_2=CHCN (acrylo nitrile)	184±	194.2±1.6	195.4±1.6	10.2±1.6	11.4±1.6
CH_3COCH_3 (acetone)	–217.2±0.4	–217.2±1.6	–214.8±1.6	0.0±1.7	2.4±1.7
CH_3COOH (acetic acid)	–432.1±0.4	–432.9±1.2	–429.9±1.2	–0.8±1.3	2.2±1.3
CH_3COF (acetyl flouride)	–442.2±3	–442.6±1.4	–439.9±1.4	–0.4±3.3	2.3±3.3
CH_3COCl (acetyl chloride)	–243±1	–243.8±1.1	–240.3±1.1	–0.8±1.5	2.7±1.5

Table A1.1 (continued) Comparison of Calculated and Experimental Heats of Formation (298 K, kJ mol^{-1}) for Closed-Shell Species

		$\Delta H_f°$(theory)[b]		deviation (theory – expt)[e]	
Molecule	$\Delta H_f°$(expt)[a]	CBS-QB3[c]	ROCBS-QB3[c,d]	CBS-QB3[c]	ROCBS-QB3[c,d]
$CH_3CH_2CH_2Cl$ (propyl chloride)	–132.4±0.6	–132.4±1.6	–129.5±1.6	0.0±1.7	2.9±1.7
$(CH_3)_2CHOH$ (isopropanol)	–272.5±0.4	–272.1±1.6	–269.7±1.6	0.4±1.7	2.8±1.7
$C_2H_5OCH_3$ (methyl ethyl ether)	–216.4±0.6	–221.0±1.6	–218.6±1.6	–4.6±1.8	–2.2±1.8
$(CH_3)_3N$ (trimethyl amine)	–23.7±0.6	–24.4±1.6	–23.2±1.6	–0.7±1.7	0.5±1.7
C_4H_4O (furan)	–34.8±0.4	–31.5±2.1	–28.6±2.1	3.3±2.2	6.2±2.2
C_4H_4S (thiophene)	115.1±1	116.8±2.2	120.4±2.2	1.7±2.4	5.3±2.4
C_4H_5N (pyrrole)	108.4±0.5	103.1±2.1	104.8±2.1	–5.3±2.1	–3.6±2.2
C_5H_5N (pyridine)	140.6±1.5	144.5±2.6	146.6±2.6	3.9±3.0	6.0±3.0
H_2	0	–4.6±0.01	–4.6±0.01	–4.6±0.01	–4.6±0.01
MD[k]				–0.4	+1.7
MAD[k]	±1.3			4.7±1.7	4.7±1.7
RMS[k]	±2.0			6.6±1.7	6.3±1.7
LD[k]	±8			–25.9±5	–20.6±5
Excluding species[l] with total experimental uncertainties ≥ 5 kJ mol^{-1}					
MD[k]				–0.1	+2.0
MAD[k]	±1.0			4.4±1.5	4.4±1.5
RMS[k]	±1.4			6.0±1.5	5.8±1.5
LD[k]	±4			–23.4±2	+16.4±2

[a] Reference 1, except as noted. [b] The uncertainty in the calculated value reflects the uncertainty in the experimental heats of formation of the constituent atoms (Table 3.3). [c] Conversion of CBS-QB3 energies and deviations (reference 2) from kcal mol^{-1} to kJ mol^{-1} introduces an uncertainty of ± 0.4 kJ mol^{-1}. [d] ROCBS-QB3 energies and deviations are obtained from the corresponding CBS-QB3 values (reference 2) by adding the Δ(ROCBS-QB3) values of the constituent atoms (see Table 3.3 footnote c). [e] The uncertainty in the deviation from experiment reflects the uncertainty in both the experimental and the calculated values for the heat of formation of the molecule: $s_{dev} = [s_{exp}^2 + s_{calc}^2]^{1/2}$ [f] Reference 3. [g] Reference 4. [h] Reference 5. [i] Reference 6. [j] Reference 7. [k] MD = mean deviation, MAD = mean absolute deviation, RMS = root mean square deviation, LD = largest deviation, from experimental values. [l] SiH_2, LiF, ClF, AlF_3, $AlCl_3$, ClF_3, C_2F_4, CHF_3 and CH_3SiH_3.

Table A1.2

Comparison of Calculated and Experimental Ionization Energies (0 K, kJ mol^{-1})

Molecule	IE(expt)[a]	deviation (theory – expt)[b]	
		CBS-QB3	ROCBS-QB3
G2-1 test set			
Li	520.2± 0.001	–5.0	–5.2
Be	899.5± 0.001	5.0	4.7
B	800.6± 0.002	–7.1	–6.6
C	1086.4± 0.01	–6.7	–6.6
N	1402.3± 0.1	–3.3	–4.5
O	1313.9 ± 0.1	–1.3	–1.8
F	1681.1 ± 0.1	5.0	5.0
Na	495.8± 0.1	–1.3[c]	–1.1[c]
Mg	737.7 ± 0.1	–5.4	–4.7
Al	577.6± 0.1	–5.4	–6.2
Si	786.5± 0.003	–8.4	–8.0
P	1011.8± 0.001	–7.9	–3.9
S	999.6± 0.1	–8.8	–8.7
Cl	1251.2±0.1	–5.0	–3.5
CH_4	1216.7±1.0	10.0±1.0	9.3±1.0
NH_3	982.8371±0.0002[d]	–0.9	–1.8
OH	1255.9±0.02	1.3	0.0
H_2O	1217.9±0.03	3.3	2.0
FH	1548.0±0.3	8.8±0.3	7.6±0.3
SiH_4	1061.3±1.9	0.8±1.9	1.2±1.9
PH	979.2±0.8	–4.2±0.8	–3.4±0.8
PH_2	947.9±0.2	–5.9±0.2	–4.9±0.2
PH_3	952.3±0.2	–1.7±0.2	–1.3±0.2
SH	1005.6±0.04	–2.9	–7.5
SH_2 ($^2B_1$ cation)	1008.6±0.1	–2.5	–1.9
SH_2 ($^2A_1$ cation)	1231.1±3.5	1.3±3.5	4.7±3.5
HCl	1229.9±0.2	2.5±0.2	1.4±0.2
C_2H_2	1100.2±0.03	3.8	3.0
C_2H_4	1014.4±0.1	3.3	2.9
CO	1352.2±0.03	5.0	3.5
N_2 ($^2\Sigma$ cation)	1503.3±0.8	2.9±0.8	1.9±0.8
N_2 ($^2\Pi$ cation)	1611.2±0.1	5.0	4.3
O_2	1164.5±0.02	8.8	11.9
P_2	1019.6±0.2	1.7±0.2	–2.8±0.2
S_2	902.7±0.2	4.2±0.2	9.2±0.2
Cl_2	1107.7±0.3	3.8±0.3	0.3±0.3

Table A1.2 (continued) Comparison of Calculated and Experimental Ionization Energies (0 K, kJ mol^{-1})

Molecule	IE(expt)[a]	deviation (theory – expt)[b]	
		CBS-QB3	ROCBS-QB3
ClF	1221.5±1.0	3.3±1.0	2.1±1.0
SC	1093.2±1.0	–5.0±1.0	–0.5±1.0
G2-2 test set			
H	1312.1±0.1	0.0	0.2
He	2372.4±0.1	2.9	2.9
Ne	2080.7±0.1	11.3	13.5
Ar	1520.6±0.1	–2.1	2.3
BF_3	1515±30[e,f]	5.2±30	4.4±30
BCl_3	1119.2[e]±2	1.7±2	0.3±2
B_2F_4	1164.4[e] (vertical)	–26.8 (adiabatic)	–27.8(adiabatic)
CO_2	1329.4±0.2	0.4±0.2	0.5±0.2
CF_2	1101.9±1.0	5.4±1.0	5.0±1.0
OCS	1079.2±0.2	2.1±0.2	1.6±0.2
CS_2	972.6±0.2	–2.5±0.2	–2.7±0.2
CH_2 ($^3B_1$)	1003.1±0.3	–6.7±0.3	–5.5±0.3
CH_3	949.1±0.5	–3.8±0.5	–2.8±0.5
C_2H_5	783.2±0.8	2.1±0.8	3.2±0.8
C_3H_4	932.8±0.5	–1.7±0.5	–1.8±0.5
CH_2CCH_2	934.7±0.2	2.1±0.2	2.4±0.2
Sec-C_3H_7	711.1±1.9	10.0±1.9	10.9±1.9
C_6H_6	891.9±0.006	8.4	8.5
$C_6H_5CH_3$	851.9[e]±0.1	8.4	
CN	1354±2 ($^3\Pi$)[g]	4.4±2	5.1±2
HCO	785.4±3.9	1.3±3.9	1.9±3.9
CH_2OH	729.4±1.0	–2.5±1.0	–3.2±1.0
CH_3O	1034.3±1.0	2.9±1.0	4.6±1.0
H_3COH	1046.9±2.9	6.7±2.9	6.9±2.9
CH_3F	1203.3±1.0	12.6±1.0	11.8±1.0
CH_2S	904.6±0.3	–2.9±0.3	–2.0±0.3
CH_2SH	727.1±0.3	–3.8±0.3	–3.0±0.3
CH_3SH	911.1±0.2	0.8±0.2	–0.7±0.2
CH_3Cl	1086.9±0.3	6.3±0.3	0.5±0.3
C_2H_5OH	1010.0[e]±7	–8.4±7	–6.0±7
CH_3CHO	986.7±0.5	4.2±0.5	3.6±0.5
CH_3OF	1094.1±0.8	6.7±0.8	7.3±0.8
C_2H_4S	873.3±0.6	2.5±0.6	0.9±0.6
NCCN	1290.4±0.8	0.8±0.8	1.6±0.8
C_4H_4O	856.8±1.0[e,f]	3.8±1.0	2.9±1.0
C_4H_4NH	792.0[e]±0.5	4.2±0.5	3.3±0.5
C_6H_5OH	820.9[e]±2	5.0±2	

Table A1.2 (continued) Comparison of Calculated and Experimental Ionization Energies (0 K, kJ mol^{-1})

Molecule	IE(expt)[a]	deviation (theory – expt)[b]	
		CBS-QB3	ROCBS-QB3
$C_6H_5NH_2$	744.8±5	0.8±5	2.9±5
B_2H_4	935.9±1.9	–12.1±1.9	–12.6±1.9
NH	1301.6±1.0	–5.4±1.0	–4.5±1.0
NH_2	1074.8±1.0	3.8±1.0	2.7±1.0
N_2H_2	925.2±0.7	2.1±0.7	0.4±0.7
N_2H_3	734.2±1.0	–0.4±1.0	–1.1±1.0
HOF	1226.3±1.0	6.3±1.0	6.1±1.0
SiH_2 ($^1A_1$)	882.8±1.9	–4.2±1.9	–3.5±1.9
SiH_3	784.9±0.5	–5.9±0.5	–5.4±0.5
Si_2H_2	791.2±1.9	–3.8±1.9	–2.9±1.9
Si_2H_4	780.6±2.9	1.3±2.9	0.3±2.9
Si_2H_5	733.5[c]	2.9	2.5
Si_2H_6	939.8±1.9	–11.3±1.9	–11.2±1.9
MD[h]		+0.4	+0.4
MAD[h]	±0.6	4.4±0.1	4.1±0.1
RMS[h]	±1.0	5.3±0.1	5.2±0.1
LD[h]	±3.9	12.6±1.0	–13.5±0.1

[a]Experimental values have been reproduced from reference 8 unless otherwise noted. [b]Experimental uncertainties greater than 0.1 kJ mol^{-1} are included explicitly. [c] Includes Na core-valence correlation energy correction ($10.29q_{Na} + 7.43q^2_{Na}$ kJ mol^{-1}, see reference 9). [d] Reference 10. [e] Reproduced from reference 11. [f] Reference 12. [g] Reference 13. [h] Excludes BF_3, B_2F_4, $C_6H_5CH_3$ and C_6H_5OH.

Table A1.3

Comparison of Calculated and Experimental Electron Affinities (0 K, kJ mol^{-1})

Molecule	EA(expt)[a]	deviation (theory – expt)[b]	
		CBS-QB3[c]	ROCBS-QB3
G2-1 test set			
C	121.9[d]±0.03	–7.5	–8.2
O	141.0[d]±0.0003	–8.4	–8.1
F	328.2[d]±0.0004	–3.3	–2.9
Si	134.1[e]±0.01	–4.2	–3.0
P	72.0[d]±0.03	–5.9	–1.0
S	200.4[d]±0.0001	1.3	1.8
Cl	348.6[d]±0.01	6.3	8.5
CH	119.4±0.8	–10.5±0.8	–9.6±0.8
CH_2	62.9±0.6	–2.9±0.6	–0.2±0.6
CH_3	7.7±2.9	–8.4±2.9	–9.4±2.9
NH	35.7±0.4	–10.5±0.4	–9.9±0.4
NH_2	74.9±3.6	–5.9±3.6	–7.6±3.6
OH	176.3±0.004	–6.3	–6.8
SiH	123.2±0.8	–7.9±0.8	–6.3±0.8
SiH_2	108.4±2.1	–10.5±2.1	–8.6±2.1
SiH_3	135.7±1.4	–3.8±1.4	–3.1±1.4
PH	99.2±1.0	–2.9±1.0	–2.3±1.0
PH_2	122.6±1.0	–0.4±1.0	–2.0±1.0
SH	223.6±0.2	0.8±0.2	3.6±0.2
O_2	43.5±0.7	0.8±0.7	–2.8±0.7
NO	2.5±0.5	–0.8±0.5	1.5±0.5
CN	372.6±0.5	4.2±0.5	2.6±0.5
PO	105.4±1.0	–3.3±1.0	–1.8±1.0
S_2	160.5±3.9	3.3±3.9	–2.0±3.9
Cl_2[f]	232±20	11.3±20	12.2±20
G2-2 test set			
Li	59.6±0.02	4.2	4.5
B	27.0[g]±0.003	–10.9	–8.3
Na	52.9±0.004	8.4[h]	8.7[h]
Al	41.8[i]±0.0.05	–9.6	–8.7
C_2	315.8±0.8	–8.4±0.8	–7.8±0.8
C_2O	220.9±1.7	2.9±1.7	1.1±1.7
CF_2	17.3±0.5	–2.9±0.5	–1.4±0.5
NCO	348.2±0.5	0.0±0.5	0.6±0.5
NO_2	219.3±0.5	0.4±0.5	–1.1±0.5
O_3	213.3±0.4	11.3±0.4	1.3±0.4
OF	219.2±0.6	0.8±0.6	0.0±0.6

Table A1.3 (continued). Comparison of Calculated and Experimental Electron Affinities (0 K, kJ mol^{-1})

Molecule	EA(expt)[a]	deviation (theory – expt)[b]	
		CBS-QB3[c]	ROCBS-QB3
SO_2	106.8±0.8	5.0±0.8	7.4±0.8
S_2O	181.1±0.8	10.0±0.8	9.1±0.8
C_2H	286.5±0.6	–1.7±0.6	–1.3±0.6
C_2H_3	64.4±2.3	0.0±2.3	1.3±2.3
C_3H_2	173.1±0.8	2.9±0.8	2.3±0.8
C_3H_3	86.2±0.5	–0.8±0.5	0.3±0.5
C_3H_5	46.4±0.8	–3.3±0.8	–4.2±0.8
HCO	30.2±0.5	–6.3±0.5	–6.8±0.5
HCF	52.3±0.5	–4.2±0.5	–3.3±0.5
CH_3O	151.5±0.5	–0.4±0.5	–1.1±0.5
CH_3S	179.6±0.4	2.5±0.4	1.8±0.4
CH_2S	44.9±2.2	2.5±2.2	3.6±2.2
CH_2CN	149.0[j]±0.6	–2.9±0.6	–1.5±0.6
CH_2NC	102.1[j]±2.5	5.4±2.5	4.5±2.5
CHCO	226.7±2.1	–0.4±2.1	6.5±2.1
CH_2CHO	176.0±0.004	–1.7	–0.5
CH_3CO	40.8±3.6	0.4±3.6	–0.5±3.6
CH_3CH_2O	165.3±0.5	7.1±0.5	–0.3±0.5
CH_3CH_2S	188.4±0.4	5.4±0.4	2.4±0.4
LiH	33.0±1.2	–1.3±1.2	–1.6±1.2
HNO	32.6±1.5	–1.7±1.5	0.7±1.5
HOO	104.0±1.6	–1.3±1.6	–1.9±1.6
MD		–1.3	–1.3
MAD	±0.9	4.3±0.2	3.9±0.2
RMS	±1.3	5.5±0.2	5.0±0.2
LD	±3.9	11.3±0.4	–9.9±0.4

[a]Experimental values have been reproduced from reference 8 unless otherwise noted. [b] Experimental uncertainties greater than 0.1 kJ mol^{-1} are included explicitly. [c] Reproduced from Ref 9. [d] Reference 14. [e] Reference 15. [f] Excluded from all statistics. [g] Reference 16. [h] Includes Na core-valence correlation energy correction ($10.29q_{Na} + 7.43q^2_{Na}$ kJ mol^{-1}, see reference 9). [i] Reference 17. [j] Reference 18.

A1.4 References

(1) Lias, S. G.; Bartmess, J. E.; Liebman, J. F.; Holmes, J. L.; Levin, R. D.; Mallard, W. G. *J. Phys. Chem. Ref. Data* **1988**, *17*, Suppl 1.

(2) J. A. Montgomery, Jr., M. J. Frisch, J. W. Ochterski, and G. A. Petersson, *J. Chem. Phys.* **2000**, *112*, 6532.

(3) J. B. Pedley, *Thermochemical Data and Structures of Organic Compounds*, *Volume I* (Themodynamics Research Center, Texas, 1994).

(4) "Experimental Thermochemical Data," NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101, Release 12, August 2005, edited by R.D. Johnson III, National Institute of Standards and Technology, Gaithersburg, MD, 20899 (http://srdata.nist.gov/cccbdb).

(5) Montgomery, Jr., J. A.; Michels, H. H.; Francisco, J. S. *Chem. Phys. Lett.* **1994**, *220*, 320.

(6) Dixon, D. A.; Feller, D.; Sandrone, G. *J. Phys. Chem. A* **1999**, *103*, 4744.

(7) Colegrove B. T.; Thompson, T. B. *J. Chem. Phys.* **1997**, *106*, 1480.

(8) Afeefy, H. Y.; Liebman, J. F.; Stein, S. E. ''Neutral Thermochemical Data'' in *NIST Chemistry WebBook*, NIST Standard Reference Database Number 69, edited by W. G. Mallard and P. J. Linstrom, November 1998, National Institute of Standards and Technology, Gaithersburg, MD, 20899 (http://webbook.nist.gov/chemistry/).

(9) Ochterski, J. W.; Petersson, G. A.; Montgomery, Jr., J. A. *J. Chem. Phys.* **1996**, *104*, 2598.

(10) Seiler, R.; Hollenstein, U.; Softley, T. P.; Merkt, F. *J. Chem. Phys.* **2003**, *118*, 10024.

(11) Boese, A. D.; Oren, M.; Atasoylu, O.; Martin, J. M. L.; Kállay, M.; Gauss, J. *J. Chem. Phys.* **2004**, *120*, 4129.

(12) Lias, S. G."Ionization Energy Evaluation" in NIST Chemistry WebBook, NIST Standard Reference Database Number 69, Eds. P.J. Linstrom and W.G. Mallard, June 2005, National Institute of Standards and Technology, Gaithersburg MD, 20899 (http://webbook.nist.gov).

(13) Berkowitz, J.; Chupka, W.A.; Walter, T.A. *J. Chem. Phys.* **1969**, *50*, 1497.

(14) *CRC Handbook of Chemistry and Physics*, edited by H. P. R. Frederikse and D. R. Lide, 78th Ed., CRC Press, Boca Raton, FL, 1997.

(15) Thøgersen, J.; Steele, L. D.; Scheer, M.; Brodie, C. A.; Haugen. H. K. *J. Phys. B.* **1996**, *29*, 1323.

(16) Scheer, M.; Bilodeau, R. C.; Haugen, H. K. *Phys. Rev. Lett.* **1998**, *80*, 2562.

(17) Scheer, M.; Bilodeau, R. C.; Thøgersen, J.; Haugen, H. K. *Phys. Rev. A* **1998**, *57*, 1493.

(18) Berkowitz, J.; Ellison, G.B.; Gutman, D. *J. Phys. Chem.* **1994**, *98*, 2744.