Department of Chemistry
Faculty of Science
University of Helsinki

COMPUTATIONAL STUDIES OF REACTION CHANNELS FOR ^{3}(RO…OR') INTERMEDIATES FORMED IN PEROXY RADICAL SELF AND CROSS-REACTIONS.

Galib Hasan

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Science of the University of Helsinki, for public examination in lecture room A129, Chemicum, Kumpula, on 15th December 2023, at 12 noon.

Helsinki 2023

Supervisors:

Professor Theo Kurtén, University of Helsinki
University Researcher Rashid Valiev, University of Helsinki

Pre-examiners:

Professor Heikki M. Tuononen, University of Jyväskylä, Finland
Professor Klaus Braagaard Møller, Technical University of Denmark

Opponent:

Professor Keith Kuwata, Macalester College, USA

ISBN 978-951-51-9554-8 (print)
ISBN 978-951-51-9555-5 (online)
ISSN 2954-2898 (print)
ISSN 2954-2952 (online)

PunaMusta, Joensuu 2023

Abstract

Gas phase accretion reactions ("dimer formation") have recently received attention as a significant process in the atmosphere, because they can lead to the formation of very-low volatility highly oxygenated organic molecules (HOMs), which later form aerosol particles. On the other hand, organic peroxy radical (RO_2) are key intermediates in the chemistry of the atmosphere. One of the main sink reactions of RO_2 especially in clean conditions is the recombination reaction $RO_2 + R'O_2$. This reaction goes through an intermediate complex $^1(RO...^3O_2...OR')$. The $^3O_2$ is very weakly bound, and evaporates from the system, giving a $^3(RO...OR')$ cluster. This complex cluster has three major reaction channels, leading to the following products: 1) $R_{-H}=O + R'OH$, 2) ROOR', and 3) RO + R'O. This thesis presents a series of quantum chemical studies on the properties and fates of the $^3(RO...OR')$ intermediate clusters. The starting point of the work is a systematic conformational search of the intermediate clusters on the triplet potential energy surface, using a cost-effective model based on a combination of multiple state-of-the-art quantum chemical methods. The next step is to calculate the reaction rate of the three channels: evaporation or dissociation to radical products RO + R'O, a hydrogen shift reaction forming carbonyl + alcohol products $R_{-H}=O + R'OH$, and a "spin-flip" (intersystem crossing; ISC) leading to the formation of ROOR' peroxide "dimer" accretion products. This work shows that for small and medium-sized RO_2, all these channels are competitive with each other in the atmosphere, with typical rates greater than 10^6 s^{-1}, and many rates exceeding 10^{10} s^{-1}. However, for complex R or R', the binding energy effect may render not only the dissociation channel, but also the H-shift channel, uncompetitive with the ISC. The overall results suggests that the computationally proposed novel $RO_2 + R'O_2$ reaction mechanism is qualitatively compatible with experimental results on accretion product formation and may thus play a key role in the formation of organic aerosols.

Acknowledgements

I would like to express my gratitude to the Academy of Finland for providing the funding for my PhD and to the Department of Chemistry for giving me the opportunity to work here. I am especially grateful to my supervisor, Professor Theo Kurten, whose invaluable guidance, and support made this thesis possible. I am also grateful to Rashid Valiev, my co-supervisor, we discussed lot about multireference method during my PhD work. I am thankful to the reviewers Heikki M. Tuononen and Klaus Braagaard Møller who spent their valuable time on my thesis. I would also like to extend my thanks to all present and past co-workers in the computational atmospheric chemistry research group, particularly Vili, Siddharth, Thomas, Noora, Lauri, Chris, Christian, Robert. I am also thankful to all colleagues on the fourth floor, especially Usman, Atif, Kajsa, Nanna, Emelda, Feda and Dominika. I would also like to thank to the Simu group people from the physics side, especially Jakub, you helped me lot during my early stage of PhD. I am also grateful to all my friends outside of academia whose companionship and encouragement have played a significant role in my motivation and growth. I am thankful to Shahjan, Saiful, Shanto, Mamun, Ranjan for your companionship since my bachelor studies.

I am deeply thankful to my parents and father-in-law, whose unwavering support and prayers have been a source of strength throughout my journey. I would also like to express my gratitude to my siblings Dr. Mehedi Hasan, Asif Hasan, and Mahmuda Sultana. Finally, I have no words to express my gratitude to my wife, Asma, and my son, Ridwan. You have selflessly sacrificed your time and support for me during my PhD journey.

List of abbreviations

ABC	Artificial bee colony algorithm
AOs	Atomic orbitals
BO	Born-Oppenheimer approximation.
BVOC	Biogenic volatile organic compound
CASSCF	Complete active space self consistent field
CC	Coupled cluster
CHARMM	Chemistry at HARvard Molecular Mechanics
CI	Configurational interaction
sCIs	stabilized Criegee intermediates
CVT	Canonical variational theory
DFT	Density functional theory
DFTB	Density functional tight binding
FC	Franck-Condon
GDA	Generalized gradient approximation
GTO	Gaussian type orbital
HF	Hartree-Fock
HOMs	Highly oxygenated organic molecules
ISC	Inter system crossing
KS-DFT	Kohn-Sham density functional theory
LM	Local minima
MMFF	Merck molecular force field
MM	Molecular Mechanics
MCSCF	Multi-configurational self-consistent field
MOs	Molecular orbitals
MSD	Mean square deviation
NBO	Natural bond orbital
PT	Perturbation theory

PES	Potential energy surface
POZ	Primary ozonide
RASSCF	Restricted active space self-consistent field
RRHO	Rigid rotor harmonic oscillator
RP-VTST	Reaction path variational state theory
SOA	Secondary organic aerosol
SOC	Spin orbit coupling
SOCME	Spin orbit coupling matrix elements
SCF	Self-consistent-field
SD	Slater determinant
STO	Slater type orbital
SCT	Small curvature tunneling
TS	Transition state
TST	Transition state theory
VTST	Variational transition state theory
VRC-VTST	Variable reaction coordinate variational transition state theory
VOC	Volatile organic compounds
WF	Wave function
XMC-QDPT2	Extended multi-configuration quasi-degenerate perturbation theory
ZDO	Zero differential overlap
ZCT	Zero curvature tunneling

Index

Abstract iii

Acknowledgements iv

List of abbreviations v

List of original publications ix

1 Introduction 1

2 Accretion product formation through RO_2 recombination 4

2.1 Formation of peroxy radicals (RO_2) 4

2.2 Atmospheric chemistry of RO_2 6

2.3 Molecular- level mechanism of the RO_2 + R'O_2 reaction 8

3 Objective of the thesis 12

4 Theoretical background 13

4.1 Background of quantum chemistry 13

4.2 Quantum chemical methods 14

4.2.1 Wave-function based theory 14

4.2.2 Density functional theory 19

4.2.3 Semiempirical and molecular mechanics approaches 20

4.3 Basis sets 22

4.4 Thermochemical and vibrational analysis 24

4.5 Conformational sampling challenges 26

4.5.1 Conformational sampling of monomers (isolated alkoxy radicals) 26

4.5.2 Configurational sampling of triplet dimer clusters of two alkoxy radicals 27

4.5.3 ABC cluster algorithm 28

4.5.4 Details of the sampling 29

4.6 Uniqueness and filtering 34

4.7 Reaction rate calculation 36

4.7.1 Transition state theory (TST) 36

4.7.2 Variational transition state theory (VTST) 38

4.7.3 ISC rate constant calculation 41

5 Results and discussion 45

5.1 Configurational sampling and structures of the triplet ^{3}(RO…OR') cluster 45

5.2 Stability of the ^{3}(RO…OR') cluster with respect to RO + R'O 47
5.3 VTST rate for the RO + R'O ⇒ ^{3}(RO…OR') reaction..... 53
5.4 H-shift rate calculation 54
5.5 Inter system crossing rate constant 59
6 Conclusions and outlook..... 63
7 References 67

List of original publications

List of publications included in the thesis:

I. **Hasan, G.;** Salo, V. T.; Valiev, R. R.; Kubečka, J.; Kurtén, T. Comparing Reaction Routes for ^{3}(RO…OR′) Intermediates Formed in Peroxy Radical Self- and Cross-Reactions. Journal of Physical Chemistry A, **2020**, 124, 8305–8320.

II. **Hasan, G.;** Valiev, R. R.; Salo, V. T.; Kurtén, T. Computational Investigation of the Formation of Peroxide (ROOR) Accretion Products in the OH- and NO_3-Initiated Oxidation of α-Pinene. Journal of Physical Chemistry A, **2021**, 125, 10632–10639.

III. **Hasan, G**.; Salo, V.-T.; Golin Almeida, T.; Valiev, R. R.; Kurtén, T. Computational Investigation of Substituent Effects on the Alcohol + Carbonyl Channel of Peroxy Radical Self- and Cross-Reactions. Journal of Physical Chemistry A, **2023**, 127, 1686–1696.

List of other publications not included in the thesis:

IV. Valiev, R. R.; **Hasan, G**.; Salo, V.-T.; Kubečka, J.; Kurten, T. Intersystem Crossings Drive Atmospheric Gas-Phase Dimer Formation. Journal of Physical Chemistry A, **2019**, 123, 6596–6604.

V. Peräkylä, O.; Berndt, T.; Franzon, L.; **Hasan, G**.; Meder, M.; Valiev, R. R.; Daub, C. D.; Varelas, J. G.; Geiger, F. M.; Thomson, R. J.; Rissanen, M.; Kurtén, T.; Ehn, M. Large Gas-Phase Source of Esters and Other Accretion Products in the Atmosphere. Journal of the American Chemical Society, **2023**, 145, 7780–7790.

VI. Collins, S.; **Hasan, G**.; Joshi, A.; McIndoe, J. S.; Linnolahti, M. Are Methylaluminoxane Activators Sheets? ChemPhysChem, **2021**, 22, 1326–1335.

VII. D'Ambro, E. L.; Schobesberger, S.; Gaston, C. J.; Lopez-Hilfiker, F. D.; Lee, B. H.; Liu, J.; Zelenyuk, A.; Bell, D.; Cappa, C. D.; Helgestad, T.; Li, Z.; Guenther, A.; Wang, J.; Wise, M.; Caylor, R.; Surratt, J. D.; Riedel, T.; Hyttinen, N.; Salo, V.-T.; **Hasan, G**.; Kurtén, T.; Shilling, J. E.; Thornton, J. A. Chamber-Based Insights into the Factors Controlling Epoxydiol (IEPOX) Secondary Organic Aerosol (SOA) Yield, Composition, and Volatility. Atmospheric Chemistry and Physics, **2019**, 19, 11253–11265.

The publications are referred to in the text by their roman numerals.

Author's contribution:

In papers **I**, **II** and **III**, I was responsible for the vast majority of the quantum chemical calculations, especially the conformer sampling, cluster thermodynamics and H-shift rate calculation. The intersystem rate constant (ISC) was calculated in association with Rashid Valiev, my co-supervisor. I also performed the data analysis, and the writing of the initial draft of paper **I**, **II** and **III**. The final manuscript was written in collaboration with all co-authors. In papers **IV-VII**, I made contributions to some part of the quantum chemical calculations, specifically conformer sampling of triplet clusters ^{3}(RO…OR'), data analysis and the writing of the manuscript. I am solely responsible for writing the thesis introduction.

1 Introduction

The atmosphere is one of the most complex and vital systems on Earth. It stretches 10,000 kilometres above the Earth's surface and blends into space. However, all the complex chemistry occurs in the troposphere, which extends only about 8-15 kilometres above the ground. The clouds we see in the sky, the rain that falls from the clouds and the wind we feel are all components of Earth's atmosphere. Even the sunlight we feel on our bodies is scattered and filtered through Earth's atmosphere. The essential element in Earth's atmosphere is oxygen, but it constitutes only 21 percent, whereas nitrogen accounts for 78 percent, and the remaining percent includes argon, carbon dioxide, water vapor, neon, and all other trace gases.

The air we breathe is an essential part of our life. We can't imagine our life for a minute without its existence. But the air doesn't consist of only gas molecules. Each cubic centimeter of air also contains thousands of tiny aerosol particles. Aerosols are defined as solid or liquid airborne particles suspended in gas[1], and atmospheric aerosols significantly impact both living species and the climate. There are two types of sources of aerosol particles in the atmosphere: natural sources (e.g., forests, oceans, deserts), and human activities (e.g., combustion processes and traffic). Especially fine (<1 μm in diameter) and ultrafine particles (<100 nm) have significant health effects when they are inhaled. For example, aerosol particles cause approximately 2.9 million deaths annually, whereas the primary gas-phase pollutant, ozone, causes an additional 217000 deaths.[2] Aerosols also decrease visibility and damage properties and landmarks when deposited on their surfaces.

On a global scale, aerosols directly affect the Earth's energy budget by scattering and absorbing solar radiation, leading on average to cooling and warming effects, respectively.[3] They also indirectly impact the climate by acting as nuclei for cloud and fog formation.[4] Primary coarse particles, such as dust, significantly impact the

total mass distribution of atmospheric aerosols. Ultrafine (submicron) aerosol particles, while representing only a small fraction of the total mass, are nevertheless responsible for many of the effects of air pollution.[5] A substantial fraction (90 %) of these particles are organic, and are formed by clustering of vapor molecules in the atmosphere. They are therefore denoted secondary organic aerosol (SOA).[6] Many experimental and computational studies have been performed over the last few decades to understand the reaction mechanism of various atmospheric elements related to aerosols, but especially the formation process of covalently bonded accretion products recently observed in the gas phase is still unknown.[7]

Understanding SOA-forming chemistry is essential for better air quality and climate change predictions. The main sources of SOA are the biogenic and anthropogenic emissions of volatile organic compounds, particularly hydrocarbons, into the atmosphere. Current atmospheric chemistry models generally underestimate the observed SOA in the atmosphere.[8] On the other hand, current models also overestimate the volatility of the precursors of SOA.[9] This leads to incorrect predictions of SOA condensation onto aerosols of different sizes.

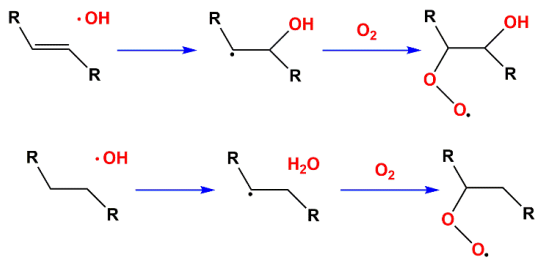
The essential chemical reaction converting volatile compounds such as hydrocarbons into condensable vapors is oxidation. While many different oxidant molecules can initiate this process, the main atmospheric oxidants are OH, O_3, and NO_3.[10,11] Next, a subsequent O_2 addition forms a peroxy radical intermediate, denoted RO_2. Recent computational and mass spectrometric studies demonstrated that RO_2 with a suitable functional group[12] or resonance structure[13] can undergo hydrogen shift isomerization reactions (autoxidation). This allows the rapid formation of highly oxygenated organic molecules (HOM). HOMs are very low volatility molecules that help explain the SOA formation missing from atmospheric chemistry models.[14] However, in addition to autoxidation, recent experiments indicate the existence of an even more efficient pathway leading directly to extremely low-volatility SOA components: the recombination of two RO_2 to form accretion products.

In this thesis work, we studied the recombination reaction of several different peroxy radicals. Our results indicate that the reaction does indeed form accretion products, and likely occurs very efficiently for large systems (large R groups). However, this is a complex reaction with multiple competing channels and several reaction steps. Understanding the process thus requires extensive computational modelling. Therefore, we used various state-of-the-art computational methods for thermodynamic and kinetics studies of the proposed channels.

2 Accretion product formation through RO_2 recombination

2.1 Formation of peroxy radicals (RO_2)

In general, the reaction of OH with hydrocarbons proceeds via addition to an unsaturated site (C=C-double bond), or abstraction of a H-atom, initially forming a carbon-centered (alkyl) radical. Rapid addition of O_2 to the carbon-centred radical then produces an organic peroxy radical (RO_2):[10,11]

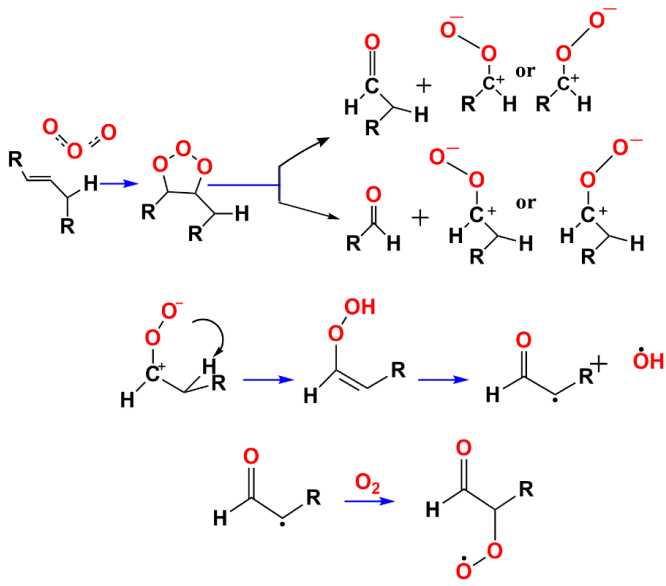


Scheme 1: Formation of peroxy radical by OH oxidation of alkenes (top) and alkanes (bottom).

The formation of peroxy radicals from the OH initiated oxidation of unsaturated hydrocarbons (alkenes) is faster than that from saturated hydrocarbons (alkanes). Both mechanisms are roughly similar (see Scheme 1**)**, but the difference is that for alkanes, the R radical is formed through a loss of hydrogen rather than through addition to the double bond. Additionally, unlike the case of alkanes, the OH oxidation of alkenes always forms new functionalities, specifically a OH group. As a result, the products derived from alkanes have higher saturation vapour pressures

than the corresponding alkene-derived products. Therefore, they have a lower likelihood of forming highly oxygenated molecules (HOMs). The formation of peroxy radicals from NO_3 – initiated oxidation follows the same mechanism as that from OH, but instead of producing alcohols and water, it produces nitrate esters and nitric acid.

Peroxy radical formation from ozone oxidation (ozonolysis) follows a more complicated pathway.[15] Ozonolysis proceeds by first forming primary ozonide intermediates (Scheme 2). These decompose to a carbonyl compound, and a carbonyl oxide known as a Criegee Intermediate (CI). Depending on the parent hydrocarbon, the Criegee Intermediate may have two possible conformers: syn conformers (where the carbonyl oxide groups are adjacent to a H atom) and anti-conformers (where they are not). The interconversion between the two conformers has a high barrier and is thus generally not expected to occur.[16] Due to the substantial excess energy of the Criegee Intermediates, the syn-conformers can undergo very fast unimolecular hydrogen shifts, forming an unstable vinyl hydroperoxide.



Scheme 2: Formation of peroxy radicals from ozone oxidation of alkenes.

Subsequent OH decomposition forms a resonance-stabilized carbon-centered (vinoxy) radical. Finally, the vinoxy radical forms peroxy radicals by adding oxygen. The anti-conformers cannot undergo fast hydrogen shifts, and their chemistry does not directly lead to peroxy radical formation. Criegee intermediates can also be collisionally stabilised, and undergo bimolecular reactions, for example with water, and potentially also with other trace gases.[17,18] However, in the atmosphere, the yields of these reactions are limited because the concentration of stabilized Criegee Intermediates (sCI) is very low. This is mainly due to the fact that also thermal unimolecular decomposition of sCIs is rather rapid, especially for syn-isomer sCIs which have accessible α-hydrogens.[19] Bimolecular reactions with water also tend to remove sCIs before they can react with other species.

2.2 Atmospheric chemistry of RO_2

Peroxy radicals play a significant role in the radical cycles that control the oxidative capacity of the Earth's lower atmosphere and form secondary pollutants, such as ozone, organic nitrates, organic acids, and particulate matter.[11,20] Compared to other radical species (e.g., alkyl and alkoxy radicals), peroxy radicals have relatively long lifetimes in the atmosphere. In most conditions, alkyl radicals tend to only add O_2. Similarly, when considering a particular alkoxy radical, one dominant reaction channel typically prevails, though there can be occasional exceptions. In contrast, the behaviour of a peroxy radical can vary significantly depending on the specific conditions, with a wide range of possible outcomes. They can, for example, undergo both uni- and bimolecular reaction.[21–23] It is this versatility of reaction mechanisms that leads to RO_2s "controlling" the radical cycles. In pristine conditions, RO_2 react either unimolecularly, or with HO_2 and other RO_2 molecules, while in polluted conditions they react predominantly with NO_x. The RO_2 + NO_x reaction has both termination or propagation channels, yielding organic nitrates or peroxy nitrates and alkoxy radicals, respectively:[11]

$$RO_2 + NO + M \rightarrow RONO_2 + M \quad (1)$$

$$RO_2 + NO \rightarrow RO + NO_2 \quad (2)$$

$$RO_2 + NO_2 + M \rightleftharpoons RO_2NO_2 + M \quad (3)$$

where M denotes bath gas molecules (air) that collisionally stabilize the reaction products.

Peroxy radicals can also react with HO_2 yielding a hydroperoxide, ROOH, or (as minor channels), RO or ROH:[11]

$$RO_2 + HO_2 \rightarrow ROOH + O_2 \quad (4)$$

$$RO_2 + HO_2 \rightarrow RO + OH + O_2 \quad (5)$$

$$RO_2 + HO_2 \rightarrow ROH + O_3 \quad (6)$$

Finally, and most important for this thesis, RO_2 molecules can react with another RO_2 (self- or cross-reaction). This reaction has (at least) three important channels:[11]

$$RO_2 + R'O_2 \rightarrow RO + R'O + O_2 \quad (7)$$

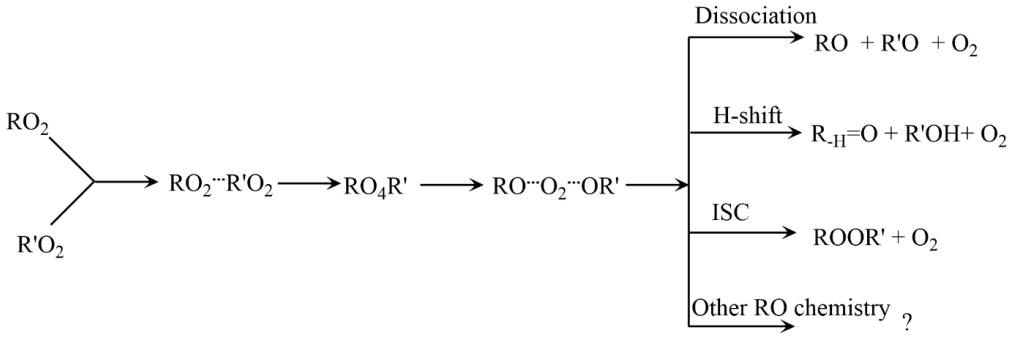
$$RO_2 + R'O_2 \rightarrow ROH + R'_{-H} = O + O_2 \quad (8)$$

$$RO_2 + R'O_2 \rightarrow ROOR' + O_2 \quad (9)$$

The first reaction is a propagation channel, while the second and third are termination reactions. The second reaction involves a hydrogen transfer from R'O to another RO, forming an alcohol (ROH) and a carbonyl product (R'_{-H}=O). The third reaction is an accretion reaction, and is believed to be important in the formation of the largest and lowest-volatility HOM.[20,24] As the formation of ROOR is the key subject of this thesis, it will be discussed in greater detail in the next chapter.

2.3 Molecular- level mechanism of the RO_2 + $R'O_2$ reaction

The formation of covalently bound accretion products (containing around twice the number of carbon atoms compared to the parent hydrocarbon) is an important process in the atmospheric chemistry leading to secondary organic aerosol. Until about a decade ago,[25] no-one thought that accretion product formation could happen in the gas phase - all known accretion reactions were restricted to the liquid phase, for example, closed-shell esterification and other condensation processes.[26] While potentially important, these reactions do not explain how gas-phase compounds decrease in volatility to form or grow nanoparticles. However, recent high-resolution mass spectrometric studies confirmed that accretion reactions are indeed taking place also in the gas phase.[24] Unfortunately most of the proposed mechanisms do not align with the observed data. One postulated gas-phase accretion mechanism is the reaction of sCIs with other organics (peroxy radicals, hydroperoxides or acids) in laboratory conditions. However, accretion product formation has been observed also in purely OH- and NO_3- initiated oxidation, where no sCI are formed.[27]



Scheme 3. Reaction between two peroxyl radicals (RO_2) and their fates.

Peroxy radical recombination (RO_2 + $R'O_2$ → ROOR' + O_2) reactions provide an alternative gas phase accretion product formation mechanism, which phenomenologically fits the observational gas-phase data.[28] However,

experiments[29] and computational studies[30] on simple peroxy radicals suggests[11] that this channel is negligible when compared to the competing channels leading to RO + R'O + O_2 or R_{-H}=O + R'OH + O_2 **(**Scheme 3**)**.

The "alcohol + carbonyl" (R_{-H}=O + R'OH, also called "molecular") route is the dominant channel for the simplest RO_2 + RO_2 reaction; MeO_2 + MeO_2, where the product yield is 65 %. Adding one more alkyl group (EtO_2 + EtO_2) leads to the dissociation channel becoming dominant (with a product yield of 60 %).[31] For $AceO_2$ + $AceO_2$, the branching ratio between these two channels is 50:50. However, few molecular-level details about any RO_2 + RO_2 reaction have been experimentally resolved, and the factors determining the branching ratios for different peroxy systems are poorly investigated.[11] Some experimental data, such as findings of triplet carbonyl formation[32,33], indirectly provide evidence on the mechanism for "alcohol + carbonyl" channel, and suggest that it involves an H-shift on the triplet potential energy surface.

Although much literature data exists on unimolecular alkoxy H-shifts,[34] there are almost no studies on intermolecular alkoxy H-shifts, particularly on the triplet surface. Experimental studies similarly suggest that the dissociation channel is quite feasible, but it is difficult to quantify the dissociation rate due to the reactivity of the product alkoxy radicals, and the multitude of potential products.[11]

A computational study by Lee *et al.*[33] has provided a new explanation for RO_2 + $R'O_2$ reactions. According to this study, all these reactions proceed through the same singlet RO_4R tetroxide structure, which undergoes two-bond cleavage to form a singlet RO…O_2….OR' "cage," where the O_2 is in its triplet ground state. The two RO radicals must also be coupled as a triplet to maintain the overall spin multiplicity as a singlet. This study partially contradicts the long-established Russell mechanism, which postulates that RO_2 + $R'O_2$ reactions first form metastable RO_4R' tetroxide intermediates, which then undergo various rearrangements to yield the three product channels: RO + R'O + O_2, R_{-H}=O + R'OH + O_2, or ROOR' + O_2. The complex RO_2 formed in autoxidation might potentially form more stable and long-lived RO…O_2…OR' or ^{3}(RO…OR') complexes due to

hydrogen bonding interactions, thereby enhancing the relative yield of the dimer-forming (ROOR') channel. However, the formation of ROOR' is prevented by the Pauli principle, which does not allow two radicals with the same spin to recombine. As a result, ROOR formation from the ^{3}(RO...OR') triplet pairs requires an intersystem crossing (ISC, "spin-flip"), which is an effect described by relativistic quantum mechanics (QM). Unfortunately, the Lee *et al.*[33] study did not provide any values for the ISC rates. Moreover, the rate-limiting barriers in their data are far too high compared to the experimental overall RO_2 + $R'O_2$ rates.

More complex RO_2 may also have additional competitive reaction channels. A very recent experimental and theoretical work by Peräkylä *et al.*[35] (Paper **V**) confirmed a new gas-phase mechanism, where alkoxy radicals with suitable structures can undergo rapid C-C beta scission inside the ^{3}(RO...OR') cluster, with subsequent recombination of the complex to a (non-peroxide) accretion product, such as an ester.

There is little information in the literature on the branching ratio of the accretion product ROOR' channel. The only explicitly reported non-zero branching ratio for a ROOR' channel is from the Leibniz Institute for Tropospheric Research (TROPOS) for the self-reaction of $CH_3C(O)CH_2O_2$; 16 %.[24] None of the same group's work directly reports others branching ratios for the ROOR' channel. However, we can estimate some branching numbers based on the bimolecular rate coefficient given for this channel for some larger RO_2 species. It is important to note that the overall bimolecular RO_2 + $R'O_2$ reaction rate (including all the product channels) cannot be larger than the gas-kinetic collision rate, which is on the order of 3×10^{-10} (cm^3 $molecule^{-1}$ s^{-1}), and is likely slightly below that. For example, in a cross-reaction between two large OH-derived systems, the reported rate for ROOR' formation is 2.6×10^{-10} (cm^3 $molecule^{-1}$ s^{-1}), which is quite close to the gas-kinetic collision rate. This suggests that the branching ratio is at least around 85 %, and potentially close to 100 %.[24] Furthermore, the reactivity of RO_2 for the formation of accretion products increases with increasing RO_2 size. Another study by the same group at TROPOS demonstrates accretion product formation from multiple RO_2 formed in the ozonolysis and OH oxidation of α-Pinene. They did not provide any direct

branching ratios for the ROOR' channel. However, by comparing the accretion reaction rate coefficients presented in that work to the gas-kinetic collision rate, we can speculate that the branching ratio is almost 100 % for all the larger accretion products.[36]

3 Objective of the thesis

The overall objective of this thesis work was to investigate the competition between the reaction channels of the RO_2 + $R'O_2$ reaction in the atmosphere. We were interested in the thermodynamics and kinetics of the reaction, and especially on the fate of the triplet intermediate complex ^{3}(RO…OR'). Modelling this requires a combination of quantum chemical and kinetics calculations. First, we wanted to develop a cost-effective conformational sampling approach to generate thousands of triplet ^{3}(RO…OR') clusters for our thermodynamic studies, and to have representative cluster conformers for further kinetic studies. The aim was also to build an automatic workflow that can save computational and especially human time. Then, we wanted to use the generated data to evaluate the effects of especially stereoisomer and substituents (functional groups) on the thermodynamics of the reaction channels. Next, we wanted to determine the (relative or absolute) rate constants of the three known channels: 1) evaporation or dissociation reaction leading to RO +R'O, 2) hydrogen shift reaction forming R_{-H}=O + R'OH, and 3) "spin-flip" (intersystem crossing) leading to the formation of ROOR' "dimer" accretion products. We were especially interested in understanding the factors controlling the reaction rates and branching ratios for a given RO_2 + $R'O_2$ system. Finally, we wished to understand the implications of the results for the atmospheric role of RO_2 + RO_2 in forming the least-volatile SOA components missing from current atmospheric models.

4 Theoretical background

In this thesis work, we applied a large variety of pre-existing quantum chemical methods as tools, with no actual method development. These methods are therefore described relatively briefly, with limited details. More emphasis is given to the approaches where I made a substantial contribution to the field, in particular, conformational sampling of triplet radical clusters.

4.1 Background of quantum chemistry

All thermodynamic and kinetic calculations in this thesis work were based on quantum chemical simulations. Quantum chemistry basically entails solving the Schrödinger equation (SE) for a molecule or a set of molecules. Many desired ground-state properties of molecules can be obtained from the ground state wave function Ψ by solving the time-independent Schrödinger equation (or for relativistic methods, the Dirac equation[37]).[38]

$$\hat{H}\Psi(R,r) = \left(\hat{T}_e + \hat{T}_n + \hat{V}_{ne} + \hat{V}_{ee} + \hat{V}_{nn}\right)\Psi(R,r) = E\Psi(R,r) \tag{10}$$

Here $\hat{H}$ is the Hamiltonian operator, which describes the potential and kinetic energy of the system, i.e. the kinetic energy of the electrons $\hat{T}_e$ and nuclei $\hat{T}_n$, as well as the electrostatic interactions between nuclei and electrons $\hat{V}_{ne}$, interactions between electrons $\hat{V}_{ee}$, and interactions between nuclei $\hat{V}_{nn}$. The wave function $\Psi(R,r)$ depends explicitly on the position of all nuclei *(R)* and all electrons *(r)*, and E is the total ground state energy of the system with the wave function $\Psi(R,r)$.

However, the exact solution of the equation above is not possible for many-electron atoms or molecules. Therefore, multiple approximations need to be applied. The first is the Born-Oppenheimer (BO) approximation.[39] This approximation is based on the fact that the motion of nuclei is slow compared to the motion of the electrons, as electrons are much lighter. The wavefunction is therefore separated into its electronic and nuclear components, and the electronic part can be solved separately, while keeping the nuclear position fixed.

$$\Psi(R_1, R_2 \dots R_N, r_1, r_2 \dots r_n) = \varphi(R_1, R_2 \dots R_N)\psi(r_1, r_2 \dots r_n) \quad (11)$$

Here $R_1, R_2 \dots R_N$ are the coordinates of the nuclei and $r_1, r_2 \dots r_n$ are the coordinates of the electrons. ψ and φ are the electronic and nuclear wave functions, respectively. The electronic wave function includes the spin coordinates which account for the intrinsic angular momentum property known as "spin.". However, solving the Schrödinger equation even with the BO approximation is not possible for many-electron systems and, therefore, further approximations are required. These approximations are made both to the Hamiltonian (in perturbation theory) and to the wave function.

4.2 Quantum chemical methods

The quantum chemical methods in use today can be divided into two groups: wave function methods (which attempt to solve the electronic wave function in the Schrödinger equation subject to various approximations), and density functional theory (DFT) methods (which instead solve a related set of equations for the electron density).

4.2.1 Wave-function based theory

In wave function methods, the wave function is expressed in terms of so-called Slater determinants, formed by a combination of one-electron wave functions (in turn expressed in terms of a basis set; see section 4.3). One of the oldest and most

common approximations to the SE is the Hartree-Fock method (HF).[40–42] In this approach, the wave function is represented as a single Slater determinant:

$$\psi^{SD}(\mathrm{r}_1, \mathrm{r}_2, \ldots, \mathrm{r}_\mathrm{N}) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1) & \phi_2(1) & \cdots & \phi_N(1) \\ \phi_1(2) & \phi_2(2) & \cdots & \phi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(N) & \phi_2(N) & \cdots & \phi_N(N) \end{vmatrix} \tag{12}$$

Here, N is a normalization constant, and each of the single-electron wave functions ϕ_1 is a spin-orbital. Slater determinants are used instead of *e.g.* simple products of one-electron wave functions as they obey the Pauli exclusion principle, which states that the total wavefunction must be antisymmetric with respect to the interchange of two electrons. The limitation of the HF approach is that it can only account for the interactions between electrons on average, and thus excludes electron correlation. This means that it is unable to find the exact energy of the system. Although the HF method can predict the absolute electronic energy of a system to within 99 %, the missing 1 % (due to electron correlation), is unfortunately crucial for chemistry and physics. Electron correlation arises due to the instantaneous interaction between electrons. Correlation can be dynamic or static. Static correlation is the correlation calculated for a specific set of multi-reference orbitals, and dynamic correlation is the correlation outside of those orbitals.[43] Electron correlation can be accounted for by expressing the wavefunction as a linear combination of several Slater determinants, each with a specific electron configuration. These electronic configurations can be obtained by exciting electrons from occupied orbitals to unoccupied (virtual) orbitals. All single, double, or triple (etc.) excitations can be included. There are many ways of treating electron correlation (in principle, many ways of making these multi-determinant wave functions), collectively known as post-Hartree-Fock methods. The main method classes include Configurational Interaction (CI), Coupled Cluster (CC), and various perturbation theories (such as MP2 or MP4). These are usually denoted “single-reference” methods, as they are based on a Hartree-Fock reference (a single Slater determinant). Multireference methods, based on a MCSCF reference wave function containing multiple determinants, form their own family of wave function – based

methods, which includes both CI approaches (MRCI) and perturbation theories (e.g. CASPT or XMC-QDPT).

In the CI methods, the wave function is constructed as a linear combination of the HF Slater determinant and determinants that describe excited states related to the HF Slater determinant. In recent decades, CI-based approaches have been superseded by Coupled Cluster as the ''gold standard'' post-HF method for estimating the electron correlation energy.[44,45] Coupled Cluster accounts for dynamical correlation effects through an exponential operator acting on the Hartree-Fock wavefunction (Ψ^{HF}) as:

$$\Psi_{\mathrm{CC}} = e^{\hat{T}}\Psi^{\mathrm{HF}} \tag{13}$$

Where Ψ_{CC} is the coupled-cluster wave function. $\hat{T}$ is the cluster operator, containing the sum of all possible excitation operators:

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \cdots + \hat{T}_n \tag{14}$$

where $\hat{T}_1$, $\hat{T}_2$, $\hat{T}_3$ etc stand for single, double, and triple excitations. A Taylor expansion of $e^{\hat{T}}$ yields:

$$e^{\hat{T}} = 1 + \hat{T} + \frac{\hat{T}^2}{2!} + \frac{\hat{T}^3}{3!} + \cdots = \sum_{k=0}^{\infty} \frac{\hat{T}^k}{k!} \tag{15}$$

For a given basis set, the couple cluster method provides the exact solution for the electronic energy when all excitations are considered. However, the computational cost of this is prohibitive. Practically applicable CC methods are truncated: the expansion is not carried out to infinite excitation order, but to a certain level. For example, the commonly used CCSD method includes only single and double excitations. This is not accurate enough for many applications as the triple (and in some cases quadruple) excitations are still important. CCSDT also includes triple excitations, and gives more accurate results, but it is computationally much more expensive than CCSD. A common and cost-effective compromise is provided by the

CCSD(T) method, which includes single S, double D, and perturbative triple (T) excitations.[46] The (T) term means that an estimation of the connected triples contribution is calculated non-iteratively using perturbation theory. Even this method is only applicable to relatively small molecular systems as the calculation time scales as N^6 for the CCSD part, while the perturbative part scales as N^7, where N is the system size, as defined (roughly) by the number of basis functions. However, CCSD(T) is still much more economical than a full CCSDT calculation, which scales as N^8.

The main coupled-cluster method used in this thesis is explicitly correlated CCSD(T), denoted CCSD(T)-F12a.[47] The designation "F12" in the method indicates that the interelectronic distance (usually denoted r_{12}) is explicitly considered in the calculation, albeit using more complicated functions than simply including $1/r_{12}$ terms in the Hamiltonian. The practical advantage of using explicitly correlated CCSD(T)-F12 is that it provides CCSD(T) - level results, but with much faster basis set convergence compared to regular CCSD(T).[46]

Static electron correlation methods are needed when the HF solution is qualitatively wrong. This usually happens when multiple electronic configurations are close in energy (near degenerate). In these cases, choosing the lowest-energy configuration leads to an incomplete description of the electronic wavefunctions. Examples include biradical systems such as ozone, Criegee intermediates, transition metal elements and non-adiabatic reaction processes (such as ISCs). For such systems, a single Slater determinant cannot describe the electronic structure of the ground state, and therefore correlated methods based on single-reference wave functions (e.g. coupled cluster) also suffer. In this situation, multireference methods are needed.

While also static correlation could be described by full CI calculations, this would require far too high computational resources for most systems. A more affordable alternative is provided by multi-configuration self-consistent field (MCSCF) methods, where one simultaneously optimizes both the expansion coefficient of the different configurations, and the orbitals for each determinant. Two determinants

are in principle sufficient to describe the two-resonance structure of ozone, while more configurations are typically needed for more complex systems. The MCSCF wave functions are typically generated as a full CI from the selected orbitals, collectively known as the active space. Because a full CI is even expensive for a small number of orbitals, the active space is usually kept as small as possible. This approach is denoted as the complete active space SCF (CASSCF) method. Another variant of MCSCF, restricted active space SCF – RASSCF – differs from CASSCF by having restrictions on the excitations within the active orbital space. In MCSCF, orbitals are thus divided into three classes: [43]

1. A set of inactive orbitals
2. A set of active orbitals
3. A set of virtual orbitals

The inactive orbitals are doubly occupied, and the virtual orbitals unoccupied, in all configurations during the building of MCSCF wave functions. Then the active electrons occupy the active orbitals with different occupation numbers. In principle, the active space should describe all the electron configurations relevant to the reaction or process of interest. However, in many cases the choice of active space is a complex task. It requires good chemical knowledge and often several trials and errors. For example, in the context of bond formation or breaking, the active space should consist of (at least) the bonding and antibonding molecular orbitals (MOs) associated with the bond. Although CASSCF is able to recover some extent of the total correlation energy, the small size of the active space means that only the static correlation part is efficiently recovered. However, dynamic electron correlation is usually needed to describe chemical reaction. Methods such as CASPT2[48] and XMC-QDPT2[49] combine a CASSCF reference with dynamic correlation recovered by second order perturbation theory. CASPT2 stands for complete active space second order perturbation theory, where several electronic states are coupled at second order via an effective Hamiltonian.[48] XMCQDPT stands for the extended multi-configuration quasi-degenerate perturbation theory by Granovsky A,[49] which is improved variant of MCQDPT (Multi-Configuration Quasi-Degenerate Perturbation Theory) approach suggested by Nakano[50]. Its updated version is

XMC-QDPT2, which guarantees the calculation of electronic states without intruder states (configurations that do not correspond to the ground or excited states of the molecule), and at the same time gives the opportunity to calculate energies of quasi-degenerate electronic states with high accuracy. The MCQDPT family differs from pure CASPT2 in the building procedure of the second order Hamiltonian contribution (H(2)). In MCQDPT, the effective operator in the second order of perturbations is constructed specifically taking into account only non-zero integral contributions.[49] This allows significant calculation time savings. In this thesis work, we used the XMC-QDPT2 method (with CASSCF reference wavefunctions) to evaluate especially the spin orbit couplings required for calculating intersystem rate constants in Papers **I- III**, and the work by Valiev *et al.* (Paper **IV**),[51] and Peräkylä *et al.* (Paper **V**).[35]

4.2.2 Density functional theory

Density Functional Theory (DFT) is a very popular and widely used method in computational chemistry due to its accuracy, efficiency, and versatility. The DFT methods currently used in quantum chemistry are based on the theorems given by Hohenberg and Kohn.[52] According to their first theorem, the electron density (ρ) determines the Hamiltonian operator and thus the ground-state wave function, energy and all associated properties of the system. This has since been extended to excited states by time dependent (TD)-DFT. The second theorem states that the Hohenberg-Kohn functional, which gives the ground state energy of the system, gives the lowest energy if and only the input electron density is the true ground state density. Later Kohn and Sham[53] extended the applicability of DFT by introducing a set of fictitious non-interacting electrons, often referred to as Kohn-Sham (KS) electrons. These electrons move in an effective potential, which is determined by the electron density of the real electrons. The key idea is to map the interacting many-electron system onto a non-interacting system that is mathematically more tractable. The vast majority of all DFT work nowadays is based on KS-DFT methods, in which the energy is related to the electron density as follows:[53]

$$E_{DFT}[\rho(r)] = T_S[\rho(r)] + J_{EE}[\rho(r)] + E_{Ne}[\rho(r)] + E_{xc}[\rho(r)] \quad (16)$$

Where T_S describes the kinetic energy of a hypothetical system of non-interacting electrons, E_{Ne} is the potential energy between electrons and atomic nuclei, J_{EE} is the Coulomb part of the electron-electron repulsive interaction, and E_{xc} is the exchange-correlation (XC) energy of electrons. E_{xc} is the only unknown part of the functional and must be approximated.[53] The accuracy (and cost) of different DFT approaches thus mainly depends on the applied exchange-correlation functional. The most popular hybrid function is B3LYP, which combines the Becke-3-parameter exchange function with the Lee-Yang-Parr correlation function.[54–56] Hybrid functionals include some part of exact exchange (from Hartree-Fock theory). Another popular more recent functional is the range-separated hybrid functional ωB97X-D[57] developed by the Head-Gordon group, which contains ten parameters, and accounts for nonlocal correlation effects.[58] The term D is added for long-range non-local correlation (dispersion) corrections.[58] In this thesis work (Paper **I-III**) and a many previous works,[35,51,59–61] this method performed very well for describing the intermolecular interactions of molecular clusters.

4.2.3 Semiempirical and molecular mechanics approaches

Semiempirical methods provide a computationally very cheap alternative to the quantum chemical methods presented above. Traditionally, these methods have been based on the Hartree-Fock formalism, but with several additional approximations, and some parameters obtained from empirical data. The central assumption in such semi-empirical models – applied to varying degrees in different methods - is the zero-differential overlap (ZDO) approximation, which disregards product terms of basis functions that rely on the same electron coordinates, but are located on different atoms.[62] The errors incurred by the approximations in semi-empirical methods are compensated by adjusting the parameters based on a comparison between the calculations and experimental data. Recently, density functional tight binding (DFTB) semiempirical programs have also become very popular due to their computational efficiency.[63,64] While DFTB methods are called semi-empirical, they have very little in common with WF-based semi-empirical methods. The DFTB approach involves expanding the Kohn-Sham DFT expression

for energy by considering density fluctuations in relation to the densities of isolated atoms around a superposition of (neutral) atomic reference densities. Within the DFTB framework, various methods can be used to solve for the energy via a self-consistent procedure, where the electronic density may undergo charge fluctuations as it converges to the stable state that corresponds to the lowest energy.[63] GFN2-xTB is a cost-effective DFTB method for providing geometries (G), vibrational frequencies (F), and noncovalent interaction (including dispersion and electrostatic interaction) energies (N) of molecular clusters.[63] Version 2 of the method differs from the first by way of implementation, for example, the empirical parameters are tailored or customized for each element individually, considering their unique characteristics and properties.[63] Nevertheless, we used this method only for pre-optimization, since the optimized geometries and formation energies need to be more accurate for our purposes.

A computationally even cheaper (albeit even less accurate) alternative is provided by force-field based molecular mechanics (MM), in which molecules are treated as point-like atoms connected by bond springs, with molecules interacting with each other via non-bonded interactions. Electrons are thus not treated explicitly at all. In MM, the interactions between atoms are usually treated based on classical potentials, such as harmonic springs, classical point charges, etc. In this thesis, MM was performed only for conformational sampling of clusters using rigid body molecules. For this application, the only relevant parameters are the non-bonded two-body interactions. For these, we used the very common Lennard–Jones (LJ) potential,[65] which forms one of the building blocks of most force fields:

$$U_{LJ} = \sum_{i=1}^{N} \sum_{i<j}^{N} 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right) \quad (17)$$

Here, ϵ and σ are the pair well depth and the equilibrium distance respectively, and r is the distance of separation between two particles, typically atoms. The r_{ij}^{-12} term corresponds to the repulsive part of the force field, and r_{ij}^{-6} to the attractive part. In this work, CHARMM force field LJ parameters have been extensively used in the molecular cluster generation. In addition to the LJ parameters, the rigid-body MM

requires partial charges, as we are simulating polar molecules. These charges are distributed to each atom, leading to an overall neutral charge (in case of neutral molecules). Electrostatic interactions between partial charges are modelled using the Coulomb potential from classical physics:

$$U_{ij}^{C}(r_{ij}) = k_e \frac{q_i q_j}{r_{ij}} \tag{18}$$

here k_e is the Coulomb constant, and q_i *and* q_j are the partial charges. The Coulomb interaction between partial charges on two molecules is very important in determining the configurations of molecular clusters. However, standard molecular mechanics (MM) typically relies on predefined charges, and thus quantum chemistry methods are required for calculating more accurate or refined partial charges for molecules not included in standard force-field libraries. Unfortunately, there is no single unambiguous way to assign atom-centered (or other) partial charges (as they are not quantum mechanical observables, unlike the total charge), so many different methods have been developed for determining them. In this thesis, we used the Natural Bond Orbital (NBO) method[66], as it is widely used in force-field construction, and seems to work for our purposes.

4.3 Basis sets

The spin-orbitals used in the Slater determinants of wave-function methods, and the one-electron orbitals of DFT, are both typically represented as a linear combination of specific functions known as basis functions, below denoted X:[67]

$$\phi = \sum_r C_r X_r \tag{19}$$

In practice, for example a Hartree-Fock calculation then entails finding the optimal values for the coefficients C_r for each spin-orbital. In calculations on molecular systems, one of two major types of atom-centered basis functions are usually used: Slater-type orbitals (STO), and Gaussian-type orbitals (GTO).[68] Other types of

functions, e.g. those used in calculations for solids and surfaces, include plane waves, numerical grids, and wavelets. The Slater-type atomic orbitals, known as STO, have exponential radial forms.[69] They are constructed based on the hydrogenic atomic orbitals (i.e., the analytical solutions of the Schrödinger equation for one-electron atoms). The general expression for STO-type basis functions is:

$$\chi_{\zeta,n,l,m}(r,\theta,\phi) = NY_{l,m}(\theta,\phi)r^{n-1}e^{-\zeta r} \tag{20}$$

Where N is a normalization constant, $Y_{l,m}$ are the spherical harmonic functions which depend on the angular momentum quantum number l and the magnetic quantum number m, n is the principal quantum number, ζ controls the width of the orbital and r, θ and ϕ are atom-centered polar coordinates.[68] Gaussian-type functions resemble STOs but have r^2 in the exponent (making calculations much more efficient despite a lower accuracy per basis function):

$$\chi_{\zeta,n,l,m}(r,\theta,\phi) = NY_{l,m}(\theta,\phi)r^{2n-2-l}e^{-\zeta r^2} \tag{21}$$

Basis sets are classified as (valence) double-ζ, (valence) triple-ζ etc, depending on the number of basis functions used to describe the (valence) orbitals of an atom. Larger basis sets generally lead to more accurate results. Most calculations in this study use one of the three following basis sets: 6-31++G(d,p), a valence double-ζ basis, VDZ-F12, a valence double-ζ basis specifically designed for explicitly correlated calculations, or aug-cc-pVTZ, a valence-triple-ζ basis. In the context of weakly bonded clusters, a common error source is basis set superposition error (BSSE),[70–72] which may lead to artificially strong binding. However, the problem tends to disappear if a large enough basis set is used. Myllys *et al.*[59] showed that when double-zeta basis sets are used for geometry optimizations, and triple-zeta basis sets for coupled-cluster corrections, BSSE is not a major source of error in molecular cluster calculations.

4.4 Thermochemical and vibrational analysis

In typical quantum chemical calculations, thermodynamical properties, including free energies and enthalpies, are obtained (at 298 K or any other temperature) through a vibrational analysis, after the ground state geometries and electronic energies have first been calculated (at T = 0 K).[73] The total energy is often expressed as the sum of the temperature-independent zero-point - corrected electronic energy ϵ_0 and the temperature-dependent thermal energy $\epsilon_T\,(T)$:

$$\epsilon_{tot}(T) = \epsilon_0 + \epsilon_T\,(T) \tag{22}$$

The zero point (vibrational) energy, ZP(V)E, is the vibrational energy present at absolute zero temperature, and is typically included with the electronic energy, as in the above equation. The thermal energy encompasses all temperature-dependent contributions. Any thermodynamic properties of a system can be calculated from the partition function (q), which is defined as the sum of the Boltzmann factors of all the available quantum energy states:

$$q(N,V,T) = \sum_{j\in\left(\substack{quantum\\states}\right)} e^{-\frac{\epsilon_T(N,V)}{k_B T}} \tag{23}$$

Where k_B is the Boltzmann constant. Here, we consider the canonical ensemble (N,V,T), which has a constant number of particles N, volume V, and temperature T. The partition function includes the contribution from translational, rotational, vibrational, and electronic degrees of freedom. Assuming that their contributions are separable, the total partition function can be written as[73]

$$q_{total} = q_{trans}\, q_{rot}\, q_{vib}\, q_{elec} \tag{24}$$

The thermal correction to the Gibbs energy G_{corr} of the system is related to the thermal enthalpy correction H_{corr} and the total entropy S_{tot} as:

$$G_{corr} = H_{corr} - TS_{tot} \tag{25}$$

The total (corrected) enthalpy and entropy of the systems are in turn expressed as the sums of the contributions from each of the four factors. Each component of the entropy (S_x), enthalpy (H_x) and thermal energy (E_x) can be obtained from the corresponding partition function as follows:[73]

$$E_x = k_B T^2 \left(\frac{\partial \, lnq_x}{dT} \right)_{N,V} \tag{26}$$

$$S_X = k_B T \left(\frac{\partial \, lnq_x}{dT} \right)_{N,V} + k_B \ln q_x \tag{27}$$

$$H_X = k_B T^2 \left(\frac{\partial \, lnq_x}{dT} \right)_{N,V} + k_B T \tag{28}$$

Here, q_x is the partition function of either electronic, translational, vibrational, or rotational degree of freedom, k_B is the Boltzmann constant, T is the temperature and V is the volume.

Vibrational frequencies are usually obtained using the rigid rotor harmonic oscillator (RRHO) approximation.[74] RRHO treats molecules (and molecular clusters) as equilibrated ideal gas particles, vibrating harmonically about their equilibrium geometries, while assuming all rotational modes are rigid. It is the simplest (and crudest) way to account for the effect of temperature on molecular properties, e.g., thermodynamics. The RRHO treatment can be improved by various post-hoc corrections. For our purposes, the most important of these are the quasi-harmonic corrections.[75,76] These remove the worst errors, especially to the entropy, associated with treating the lowest-frequency modes of weakly bound molecular clusters as harmonic vibrations.

Let us next consider a clustering reaction, and relate it to the thermochemical properties explained above:

$$A + B \rightarrow \text{AB} \tag{29}$$

Where A and B are the atoms or molecules that form a large cluster, or dissociate into individual molecules in the reverse reaction.

The formation free energy of the cluster AB is given by:

$$\Delta G = \Delta H - T\Delta S \tag{30}$$

where ΔH is the formation enthalpy (which includes both the zero-point corrected electronic energy and the thermal entropy correction) and ΔS is the formation entropy. Two important things happen during the clustering process. The enthalpy decreases ($\Delta H < 0$) due to the formation of H-bonds and other intermolecular interactions, which releases heat energy, while entropy decreases ($\Delta S < 0$) due to the reduction of the number of translational and rotational degrees of freedom. Therefore, the spontaneity of the reaction (sign of ΔG) depends on the temperature.

4.5 Conformational sampling challenges

The exploration of a potential energy surface (PES) using higher level quantum chemical methods (DFT or coupled cluster) is computationally very expensive. In this thesis, we present a systematic alternative approach for configurational sampling based on a "building up" approach to search for optimal configurations of molecular clusters. This approach can overcome the typical problems associated with searching for global minima on a multidimensional PES. In our study, ^{3}(RO...OR') clusters with different binding patterns were formed from a relatively small set of different monomers (RO), with proper configuration sampling and global minima search carried out for both the monomers and the clusters. We also evaluate and apply energy cutoffs, which can significantly reduce the computational cost with negligible loss of accuracy.

4.5.1 Conformational sampling of monomers (isolated alkoxy radicals)

In the articles included in this thesis, systematic conformer searches of the

"monomer" RO were carried out using different versions of the Spartan program (16 and 20).[77,78] In these searches, every nonterminal bond is rotated a certain number of times, and an optimization is performed using the molecular-mechanics force field (MMFF) to determine if the resulting structure represents a distinct minimum on the potential energy surface. The default number of rotations in Spartan (also used in this study) is 3 for bonds between sp^3 hybridized atoms and 2 for bonds between sp^2 hybridized atoms. All the conformer searches were performed using the keywords "searchmethod=systematic" to ensure that the systematic conformer search algorithm was used (rather than the Monte Carlo – algorithm which is the default for larger systems). The atom type of the radical oxygen was manually set (using the additional keyword "ffhint = X~~Y") to avoid Spartan treating the system as an anion in the earlier versions, for example, Spartan versions 14 and 16. However, manually setting the type of the radical oxygen atom is not required from Spartan 18 onwards.

4.5.2 Configurational sampling of triplet dimer clusters of two alkoxy radicals

The systematic configurational sampling of clusters has many different steps. For each step of configurational sampling, we have developed an automated Bash-wrapped Python script (Jammy Key for Configurational Sampling, JKCS) approach which has saved both human and computing time.[79] JKCS is coupled with the ABCluster program which uses the Artificial Bee Colony (ABC) algorithm to generate initial structures for the ^{3}(RO…OR') cluster. This algorithm performs rigid-body molecular-mechanics dynamics.[80,81] Thus, to generate cluster conformers, it requires three essential parameter sets for the reactant molecules: Cartesian coordinates, Lennard-Jones potentials, and atomic partial charges of all atoms present in the clustering molecule. In principle, other force-field potentials, for example, Coulomb-Born-Mayer, Morse, Z, and Gupta potentials, could also be used. In this study, the classic "12-6" Lennard Jones was adopted, as it is widely used in chemistry to describe the dispersion interactions due to its computational expediency. The possible structures of the "monomer" RO were taken from Spartan.

As discussed in section 4.2.3, single point Natural Bond Orbital (NBO) calculations were performed to obtain partial atomic charges for each structure, while Lennard-Jones parameters were collected from the CHARMM force-field database.[82,83] The details of the ABC algorithm are explained below.

4.5.3 ABC cluster algorithm

The artificial bee colony (ABC) algorithm can solve nonlinear optimization problems, finding the optimal values for a set of variables. This approach uses an evolutionary algorithm (EA),[84] based on the original idea of honeybee swarm behavior as proposed by Karaboga and Basturk.[85] Various other configurational sampling approaches uses similar algorithms, for example OGOLEM[86] and CLUSTER[87]. We chose ABCluster due to its simple implementation and thorough documentation. ABC is inspired by the intelligent foraging behavior of honeybees looking for a quality nectar source. The honeybees in the ABC algorithm are computational agents divided into three groups: employed bees, onlooker bees, and scout bees.[85] This description refers to the analogy of real honeybees. In this analogy, the overall nectar source (meadow with multiple flowers) is the configurational space of a studied molecular cluster. Each individual nectar source (flower) corresponds to an energy minimum configuration of a molecular cluster. The energy of the configuration corresponds to the nectar source quality. Generally, the employed bees explore the area of nectar sources based on previous knowledge (e.g., by using quality-weighted combinations of several sources). This step corresponds to evolutionary crossover. If they find a better nectar source, they fly home to update the memory of other bees. Hence, the bad nectar sources are in time forgotten and replaced by better sources. At home, the employed bees pass the information to the onlooker bees about the quality and position of the nectar by a so-called waggle dance. The onlooker bees use information from several employed bees and fly to collect nectar from the best source, and eventually also its surrounding. Hence, they may potentially find new nectar sources. Finally, since the nectar sources can be depleted (a particular configuration and its surroundings is fully explored), they are left after several iteration, and the scout bees keep looking for new nectar sources, This guarantees diversity of the nectar, or in other

words, ensures that the exploration is not stuck in a single configuration for too long. The overall hive knowledge corresponds to the all energy minima found during the simulation.

4.5.4 Details of the sampling

In this section, the details of the algorithm implemented in ABCluster program are explained in the context of searching for global minimum conformers on a potential energy surface.[80,81]

There key parameters in the application of the ABC algorithm are:,

SN - population size (i.e., *SN* employed bees and up to~3×*SN* of total bees)

g_{max} - maximum number of cluster generations (length of simulation)

g_{limit}- survival generation limit until scout bees are called

L- simulation box size (typically 1.5-3 times the cluster diameter)

LM- number of energetically low-lying local minimum conformers to be saved

The product $SN \times g_{max}$ (population size × length of simulation) is the prime factor for determining the quality of the configurational exploration. The survival limit (g_{limit}) is a parameter that controls the number of times a particular candidate (each nectar source, i.e., an individual cluster configuration) can be visited before it is abandoned as potential candidate.

Step 1: Initialize the population

SN random molecular cluster structures (population) are generated at the beginning of the conformational sampling. The initial geometry guesses are then locally optimized to their nearest energy minima. In terms of the ABC analogy, the trial solution (cluster geometry) *X* is the nectar source, and its quality is the potential energy *U(X)*.

$$U(X) = \sum_{i<j}^{N} u_2(r_{ij}) + \sum_{i<j<k}^{N} u_3(r_{ij}, r_{ik}, r_{jk}) + \cdots + u_N\,(r_{12},\cdots, r_N) \quad (31)$$

$$X \equiv [X_1, \ldots, X_{SN}] \quad (32)$$

where X_i describes the i-th minimum conformer at the current simulation step, and u_N represents the higher-order potential energy terms involving interactions between N particles within the molecular cluster. In our work, the sampled (triplet) clusters consisted of two rigid monomers. Here, 'rigid' means that all the internal degrees of freedom of the monomer molecule, for example, the bond lengths, bond angles, and dihedral angles, were kept constant. In our two-body cluster only the first term of the equation 31 is essential, as we only use two-body LJ and Coulomb potential energy terms. Each cluster configuration is fully described by 12 coordinates, which encompass both the position $r = (x, y, z)$ and orientation $o = (\alpha, \beta, \gamma)$ of each rigid molecule.

$$X_i \equiv [(\vec{r}_1, \vec{o}_1), \ldots, (\vec{r}_N, \vec{o}_N)] \quad (33)$$

$$\equiv [(x_1, y_1, z_1, \alpha_1, \beta_1, \gamma_1), (x_2, y_2, z_2, \alpha_2, \beta_2, \gamma_2)] \quad (34)$$

The trial minima evolve during the simulation or iteration steps, but the number of minima remains the same. Here, the number of configurations *SN* equals the number of employed bee tasks. This also determines the overall number of bees (3**SN*). Larger colonies (larger number of bees) and more generations always improve the sampling- however using too large colonies and long generation might be a waste of computational resources.

Step 2: Modeling employed bees

During the simulation loop for each generation, the employed bee algorithm of the ABC approach tries to find new minimum conformers and replaces the previous ones if better minima are found. The new trial configuration for X_i, denoted $X_{i-trial}$, is generated as a linear combination of two or more other random minima. For example, for $[X_j, X_k, X_l]$ where $(i \neq j \neq k \neq l)$

$$X_{i-trial} \coloneqq \sum_{m \in j,k,l} \frac{|E_m|}{|E_j|+|E_k|+|E_l|} \cdot X_m \tag{35}$$

Here, E_m is the minimum energy of conformer X_m is the minimum configuration. The trial solution is again optimized to the nearest minima before its energy $E_{i-trial}$ is evaluated. This process corresponds to employed bees finding a new nectar source (minimum configuration). This follows the genetic algorithm terminology, where a crossover happens between two parents' conformers and produces new offspring. In equation 35, the information is shared among three parents, (i.e., an evolutionary crossover), resulting in the mutation of the old X_i into $X_{i-trial}$. This is therefore often called a 'trigonometric mutation'.[88]

However, the trial solution is only accepted if it is better than the previous solution in a "greedy" way. This guarantees natural selection, and only better configurations can appear in the next generation.

$$X_i \coloneqq \begin{cases} X_{i-trial} & if\ E_{i-trial} < E_i, \\ X_i & otherwise \end{cases} \tag{36}$$

Step 3: Modeling onlooker bees

After step 2, the onlooker bees attempt to find the best candidate for a subsequent mutation. This process can happen in two ways: by roulette wheel or by tournament selection.[84] In the roulette wheel method, a configuration X_i is chosen from the whole population with a probability proportional to its energy. However, this method has a strong bias toward finding the best (lowest-energy) solution of X_i, and thus reduces the efficiency of the algorithm. Therefore, the tournament selection method is also used in the ABCluster simulation. In this method, several players (5 in this case) are randomly selected from the population $[X_i, X_j, X_k, X_l, X_m]$, where $i \neq j \neq k \neq l \neq m$,, and the best one is chosen for mutation.

$$X_i \coloneqq \begin{cases} X_i + F\left(X_j + X_k - X_l - X_m\right) & if\ \eta < 0.5 \\ X_{best} + F\left(X_j + X_k - X_l - X_m\right) & otherwise \end{cases} \tag{37}$$

where X_{best} represents the best candidate configuration present in current generation, and it is used in situations where η is greater than or equal to 0.5. η is a random number between 0 and 1, while F is a function that assigns all its arguments a random number weighting between 0 and 1. These numbers play a role in shaping how the crossover operation blends information from different candidate configurations to create new candidate configurations.

Step 4: Modeling scout bees

After the onlooker bee searches, each of the X_i is checked. If it did not change over g_{limit} cycles, it is replaced by a random trial solution $X_{i-trial}$ regardless of whether this solution is better than X_i or not. In the absence of this scout bee algorithm, the ABC simulation may end up trapped in a local minimum conformer very fast. The task of the scout bee algorithm is thus to ensure that diverse local minimum conformers on the PES landscape are sampled.

Step 5: If $g_{limit} > g_{max}$ the algorithm is finished, otherwise go back to step 2

In the initial ABC sampling, the monomer structures were kept rigid, but all the combinations of monomer structures were explored for each of the RO + R'O combinations. Initially, three thousand clusters were assigned for each combination of "monomer" RO using our JKCScommand. For instance, if we have 11 alkoxy ("RO") conformers, there can be a total of 66 distinct pairs, giving 198000 initial conformers for the ^{3}(RO…OR') cluster. The details of the workflow is given in the flowchart (**Figure 2**).

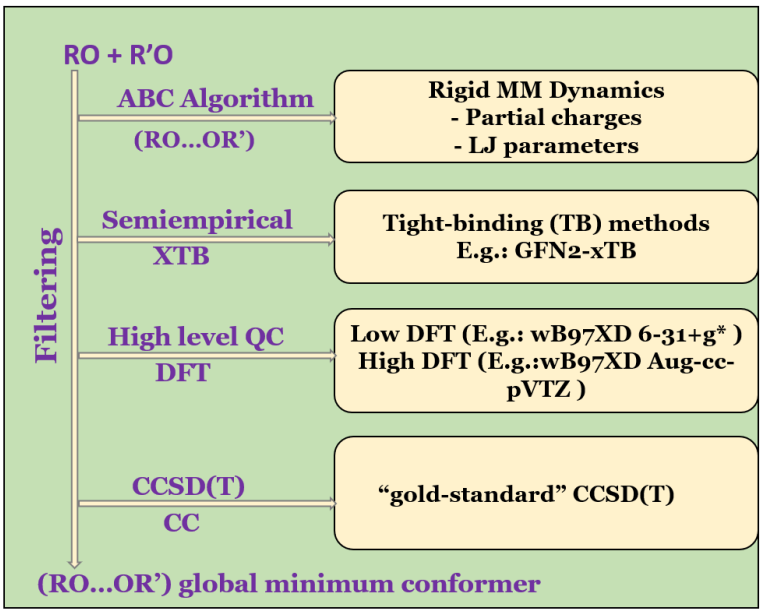


Figure 2: Flowchart of describing the overall process of the conformational sampling.

Many clusters may be redundant, meaning that further optimization will lead to identical structures (local minima). It is not possible to run DFT optimizations for the enormous number of clusters generated by ABC. To reduce computational cost, ABC-generated structures were optimized using the semi-empirical (GFN2-xTB) level of the theory, which is very efficient and well parallelized.

However, various types of unwanted reactions were observed at this step. (By definition, any reaction making or breaking chemical bonds is unwanted, since the objective was to sample intact ^{3}(RO...OR') clusters.) These reactions were typically combinations of H-shifts and C–C bond scissions, which likely occurred due to a combination of strained input geometries and (likely artificially) low reaction barriers at the GFN2-xTB level. While some of the reactions occurring during the xTB optimization may be real, such conformers are not useful for the conformational sampling. We first tested different optimization criteria available in the xTB program to produce a maximum number of intact conformers for the DFT calculations. However, our attempt was unsuccessful, and we ended up still having

many unwanted conformers that had undergone various reactions. Therefore, we used a new systematic approach to remove them. We considered not only the radius of gyration or electronic energies as collective coordinates, but also specific bond lengths (for example H…C, O…C, H…O bonds formed or broken in the unwanted reactions), using the Goodvibes program implemented as a Python module.[89] All the duplicate and unwanted structures were removed, as well as local minima with energies more than 41.84 kJmol^{-1}(10 kcal mol^{-1}) above the lowest-energy structure. DFT calculations on the remaining set of structures were first carried out at a lower level of theory (smaller basis set), to reduce the overall computational cost. Again, duplicate structures were checked and removed with some energy threshold (with different thresholds used in papers **I-III** and the work by Valiev *et al*,[51] and Peräkylä *et al*.[35]). Then a high level DFT method (large basis set) was applied to the remaining set of clusters for optimization and vibrational frequency analysis. Finally, the energetically lowest (global minimum) conformer was taken, and a coupled cluster single-point energy calculation was performed on it.

4.6 Uniqueness and filtering

At each step of the configurational sampling protocol, we excluded redundant structures and applied various filtering to save computational time. One common strategy for filtering out redundant structures is calculating mean-squared-displacement (MSD, or Root-MSD (RMSD))[90] of two different configurations, A and B (centered, ordered, and oriented in the same manner).[91,92]

$$MSD_{A,B} = \frac{1}{N} \sum_{i=1}^{N} | \vec{r}_{i,A} - \vec{r}_{i,B} |^2 \tag{38}$$

where $\vec{r}_{i,X}$ is the position of atom i in configuration X, which has overall N atoms. This approach becomes slow for too many configurations. Hence to save further computational time, we only computed MSD for structures, which had similar values for a collective coordinate, called the radius of gyration R_g, defined as follows:

$$R_g^2 = \frac{\sum_{i=1}^{N} m_i |\vec{r}_i - \vec{r}_{COM}|^2}{\sum_{i=1}^{N} m_i} \quad (39)$$

Where m_i is the mass of atom i, $\vec{r}_i$ is its position and $\vec{r}_{com}$ represents the center of mass of the whole cluster. Other possible collective co-ordinates include the electronic energy, dipole moment, and number of hydrogen bonds. Two clusters are considered as different if one of the collective coordinates differ by more than a given threshold. For example, if two molecules have the same radius of gyration (R_g) and energy (E), then the molecules are considered identical. We used thresholds of 0.01 Å for R_g and 0.001 Hartree for energy. While these are arbitrary numbers, the combination worked well for our cluster sampling to filter out the redundant conformers. We note that for other systems of interest, optimal parameter values may be different. The reason for using thresholds was to filter out structures that are too similar to other structures. Much coarser (larger) thresholds can also be used to select representative conformations, either to further reduce the computational effort, or for visualisation purposes. Figure 3 illustrates how a uniform selection of conformations is applied to a large set of structures.

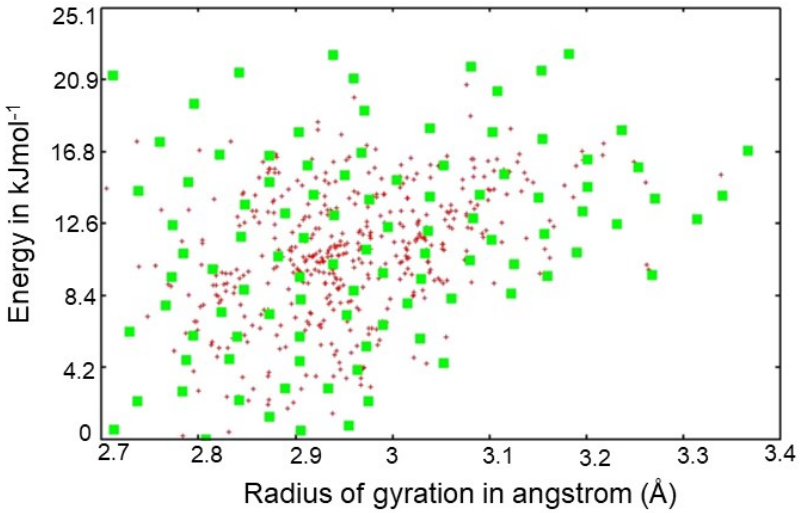


Figure 3: The relative single point GFN2-xTB energies of several hundreds of cluster structures (produced by the ABCluster program) are plotted with respect to their radius of gyration (red crosses). The green points show the 100 uniform structures selected for further calculations.

This approach does not always ensure that the global minimum structure will be discovered. However, if the selected points are not too far from each other, then very low-lying structures (close to the global minima) will certainly be found. Therefore, the final lowest-energy structure obtained by this approach will be a good approximation of the true global minimum. The uniform selection approach was used in Valiev *et al.* [51] (Paper **IV**) but in the actual thesis papers (**I-III**), we exhaustively explore all possible cluster structures by the approach mentioned above, without applying a uniform selection step.

4.7 Reaction rate calculation

4.7.1 Transition state theory (TST)

Thermodynamics tells us whether a reaction happens spontaneously or not, but does not give any information about how fast the reaction can happen: for this kinetics is needed. Reaction kinetics explicitly deals with reactions mechanisms, and provides information on the time scale of a reaction, for example in terms of its reaction rate constant. The starting point of computational kinetics studies is to construct a potential energy surface depicting the reactants, products, transition states and intermediates along the reaction path. With the transformation from reactant to the products, the reactant's geometry gets distorted along a minimum energy path that connect both minima. The species at the highest position of this surface is called the transition state (first order saddle point).[93,94] The definition of a transition state can also be extended to barrierless reactions which do not have actual saddle points, this broader definition is called a generalized-TS (GTS).[95] The GTS is the point of minimal flux across the potential energy surface (minimized rate for the reaction).[96]

For reactions with a well-defined first order saddle point, transition state theory

(TST) can be applied quite straightforwardly. TST was developed by Eyring, Evans & Polanyi in 1935.[93,94] In TST, the rate constant k for a unimolecular reaction is given by: [93,94]

$$k = \kappa \frac{k_B T}{h} \frac{Q_{TS}}{Q_R} \, exp^{\left(-\frac{E_{TS}-E_R}{k_B T}\right)} \tag{40}$$

where κ is is the quantum-mechanical tunneling coefficient, k_B is Boltzmann's constant, T is the temperature in Kelvin (K), h is Planck's constant, Q_{TS} and Q_R are the partition functions of the transition state and the reactant, respectively, and E_{TS} and E_R are the corresponding zero-point-corrected electronic energies of the transition state and the reactant. The difference in energy between the reactant and transition state is the energy barrier. In quantum mechanics there is a certain probability that a reactant having less energy than the activation energy may still reach the product side through tunneling. The magnitude of the tunneling depends on the mass being transferred during the reaction, as well the width (shape) of the barrier, and its height in both forward and reverse directions. Tunneling is often an important parameter for hydrogen shift reactions due to the low mass of H atoms. Typically, slower reactions, such as unimolecular peroxy radical H-shifts, have large tunneling corrections. In our calculations using elementary TST, we used the Eckart tunneling approach to calculate κ. This is a one-dimensional approach where the tunneling coefficient is calculated by solving the Schrödinger equation for an asymmetrical one-dimensional Eckart potential fitted to the forward and backward barrier heights and the imaginary vibrational modes (which gives information on the barrier shape).[97] The imaginary frequency was calculated by a high level DFT method (i.e., ωB97X-D/aVTZ), and the barriers were calculated using F12 single point electronic energies with ωB97X-D/aVTZ zero-point vibrational corrections. The Eckart tunneling correction was then calculated using a MATLAB[98] code provided by collaborators at the University of Copenhagen.[99] Eckart tunneling previously shown good agreement with (more expensive and accurate) multidimensional small curvature tunneling calculations for hydrogen shift reactions. [100,101]

4.7.2 Variational transition state theory (VTST)

Variational transition state theory (VTST) is an extension and refinement of the TST that provides a more accurate description of chemical reaction rate. VTST is particularly useful for chemical systems where the TST assumptions may not hold. In order to understand the difference between the VTST and TST, let us make an analogy to hikers navigating through a rugged mountain pass to reach the other side. The highest point of the pass is analogous to the saddle point. The rugged terrain, steep slopes, and high pass (activation energy) make the journey challenging. In the classical approach (like TST), it is assumed that all hikers always try to climb over the pass along the minimum-energy path (MEP). However, not all of them perfectly follow the minimum energy path and they might veer off the path. After their journey, the hikers complain that the TS is not an accurate representation of a point where flux across the PES is minimized. Instead, they approach a generalized TS (GTS) by drawing a magenta line (dividing surface) in Figure 4. The actual GTS is the point of minimal flux across the potential energy surface (minimized rate for the reaction). In VTST, the hikers are not bound to strictly follow the MEP path. Instead, they can explore multiple different routes. As the hikers explores these diverse routes, some pathways may prove more intricate and time-consuming than a direct climb over the pass. Therefore, the average journey time of all hikers who take different routes may be slower, when this more thorough exploration of the complex landscape is accounted for.

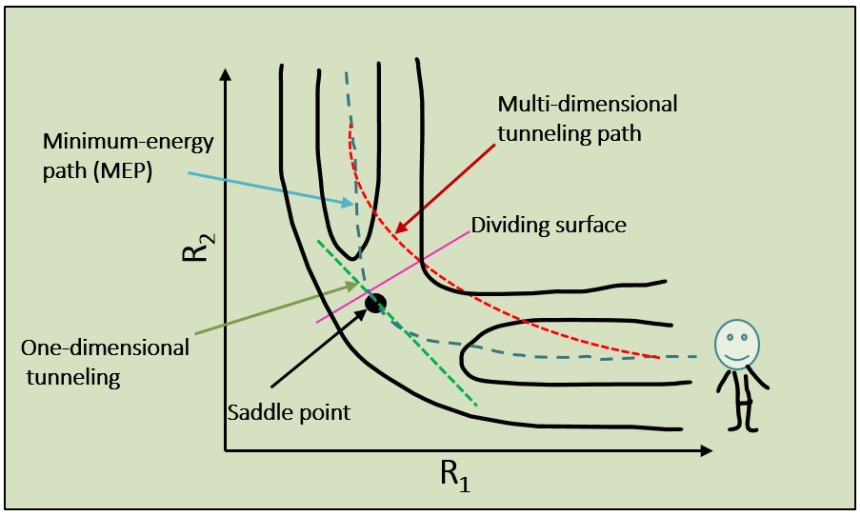


Figure 4: Schematic diagram of a multifaced dividing surface. The reaction coordinates R_1 denotes the O-H bond which is forming (represented along the X-axis), while R_2 denotes the C-H which is breaking (represented along the Y-axis). For more discussion and details, see reviews by Truhlar, *et al.*[95,96,102]

Using this analogy and figure 4, we can consider the ^{3}(RO...OR') → ROH + ^{3}R'-$_{H=}$OH-shift reaction as an example. Here, the H atom is transferring inside a cluster to another alkoxy oxygen, forming an alcohol and a carbonyl product. The hiker in the figure embodies the reactant molecules, represented by a low value of the R_2 co-ordinate, in this case the length of the C-H bond being broken in the reaction. As the hikers try to navigate through the mountain pass, R_2 increases, while R_1 (the length of the O-H bond being formed) decreases. The saddle point corresponds to an intermediate value of both co-ordinates (where one bond is being broken, and the other formed). The key features of variational transition state theory is that the dividing surface is optimized to obtain the lowest rate, which then represents the "best" rate for the system in question. This form of VTST is called the reaction path variational theory (RP-VTST) and is used when there is specific saddle point on a chemical reaction.[96,103,104] Like TST rates, VTST rates can be corrected for tunneling using multiple approaches: one-dimensional tunneling, where tunneling corrections are computed along the minimum energy path, and multidimensional tunneling, where multiple different tunneling routes are explored, as illustrated in

the figure. We used canonical variational transition state theory (CVT) with multidimensional tunneling in the temperature range of 0–600 K.[96] CVT uses the dividing surface concept explained above, providing more accurate rate coefficient as compared to TST. We applied both one-dimensional Winger tunneling (blue line) and multidimensional tunneling (red line) to cross the minimum energy path (magenta line).The multidimensional tunneling approximations considered were zero-curvature tunneling (ZCT) and small-curvature tunneling (SCT).[96] The main reason for using this RP-VTST was to double-check the accuracy of the "conventional" TST rate for the H-shift reaction (i.e., assessing the magnitude of the variational correction to this rate).

Another form of VTST, called variable reaction coordinate (VRC)-VTST, deals with reactions having loose transition states, where there is no well-defined barrier on the potential energy surface.[105–107] For VRC-VTST, the main problem is that there is no simple way to get a reaction path (as there is no saddle point), and thus no easy way to get a dividing surface either. Thus, VRC-VTST uses a complicated and computationally costly pivot point approach[108] to obtain a multidimensional dividing surface (which again is then variationally minimized). However, due to the way VTST codes are constructed, the dissociation rate cannot be calculated directly, but instead one need to compute the association rate, and then use detailed balance to obtain the dissociation rate.

In the detailed balance approach, the dissociation rate (γ; note that in cluster dynamics studies this is usually called the evaporation rate) can be estimated as:[109]

$$\gamma = \beta \, \frac{p_{\mathrm{ref}}}{k_{\mathrm{B}}T} \, \exp\left(\frac{\Delta G}{k_{\mathrm{B}}T}\right) \tag{41}$$

Where γ is the rate coefficient for the dissociation reaction ^{3}(RO…OR') → RO + R'O, β is the association rate coefficient for the reverse reaction ^{3}RO + R'O → ^{3}(RO…OR'), ΔG is the difference in Gibbs free energies of ^{3}(RO…OR') and RO + R'O, p_{ref} is the reference pressure at which ΔG is calculated (typically 1 atm), k_{B} is the Boltzmann constant, and T is the absolute temperature.[109]

4.7.3 ISC rate constant calculation

In non-relativistic quantum mechanics, the total angular momentum ($\vec{J}$) is equal to the sum of the orbital angular momentum ($\vec{L}$) and the spin angular momentum ($\vec{S}$) as follows:

$$\vec{J} = \vec{L} + \vec{S} \tag{42}$$

The spin angular momentum, $\vec{S}$, of a system comprising N electrons can take the following values:

$$\vec{S} = \left|\sum_i^N \vec{s}_i\right| = \hbar\left[\frac{N}{2}, \left|\frac{N}{2} - 1\right|, \left|\frac{N}{2} - 2\right|, \ldots\right] \tag{43}$$

where, $\vec{s}_i$ represents the spin angular momentum of the i-th electron.

The total (electronic) orbital angular momentum can similarly be written as

$$\vec{L} = \sum_i^N \left|\vec{l}_i\right| \tag{44}$$

where $\vec{l}_i$ represents the orbital angular momentum of the i-th electron.

If the spin multiplicity is conserved in a chemical reaction, $\Delta\vec{S} = 0$. The reaction is then called spin allowed - otherwise it is spin forbidden. Most molecules are singlets in their ground state ($\sum S = 0$), but the oxygen molecule is an exception, as its ground state is a triplet ($\sum S = 1$). This has consequences for the formation of ROOR' accretion products from RO_4R' tetroxides on a singlet surface. For example, the two RO radicals formed in the tetroxide dissociation must be coupled as a triplet to keep the overall spin multiplicity as a singlet, since the O_2 is formed in its triplet ground state. Thus, ROOR formation is spin forbidden, but can happen via intersystem crossings (ISC) of the ^{3}(RO...OR') to the singlet surface. In an ISC, the spin is exchanged through a coupling (spin-orbit coupling, SOC) with the orbital angular momentum, such that $\Delta\vec{S} = -\Delta\vec{L}$ and $\Delta\vec{J} = 0$. Thus, a strong spin-orbit

coupling makes it more probable for a ISC reaction to happen. Computational models for computing spin-orbit couplings are inherently relativistic.

ISC rate calculations typically invoke adiabatic and Franck Condon (FC) approximations.[110] The adiabatic approximation simplifies the quantum mechanical description of a system by assuming that the nuclear and electronic motions occur independently, while the FC approximation considers electronic transitions in molecules as instantaneous, neglecting nuclear motion during the transition process.[110] Using these approximations, the ISC rate can be calculated using the following formula.[110–112]

$$k_{ISC} = 1.6 \times 10^9 \langle i|\hat{H}_{SO}|f\rangle^2 \, FC \tag{45}$$

where $\langle i|\hat{H}_{SO}|f\rangle^2$ (in cm^{-1}) is the matrix element of the spin-orbit coupling interaction operator, $\hat{H}_{SO}$, between the initial and final electronic states i and f, and the FC is the Franck Condon factor. The FC factors describes the strength of the coupling between the electronic and vibrational modes of a molecule.[110,111] They also describe how well the vibrational modes of the molecule are matched, capturing the degree of overlap of vibrational wave functions. The larger the FC factors, the stronger the overlap between the vibrational states of the initial and final electronic states, indicating a greater degree of correspondence in vibrational modes.[110,111] In the work by Valiev *et al.* (Paper **IV**),[51] vibrational frequencies for the other electronic states were explicitly calculated, but in the thesis papers (**I, II** and **III**), the FC was computed based on an approximation using the energy gap between the states to estimate the FC. If the ground state of studied system is (as in our case) a triplet, the ISC process can be considered to be thermal, since it happens from the ground T_1 state to S_1 or to higher singlet states. Then the final ISC rate constant can be calculated using the following formula:

$$k'_{ISC} = k_{ISC} . e^{(-\frac{E_{if}}{k_B T})} \tag{46}$$

where k_B is the Boltzmann constant and E_{if} is the energy gap between the two electronic states. The ISC rate thus in principle depends on three main factors: E_{if},

$\langle i|\hat{H}_{SO}|f\rangle^2$, and FC (though as mentioned above, FC is in our case estimated using E_{if}). All these three parameters require knowledge of the energies of both singlet and triplet states. To obtain these, we need to use multireference methods (the spin-orbit coupling also requires consideration of relativistic effects). In this work, we used the Pauli-Breit spin orbital coupling interaction operator.[113] Note that the Pauli-Breit spin-orbital coupling interaction operator consists of two both one-electron two-electron terms. In our SOCME calculation for alkoxy clusters, the one-electron term is dominant.[51] Also, the $k_{ISC}(T_1 \rightarrow S_1)$ is completely determined by the SOCME ($\langle i|\hat{H}_{SO}|f\rangle^2$) due to these states being quasi degenerate (very small energy gap). The Bolzmann ($e^{(-\frac{E_{if}}{k_B T})}$) and FC factors were always unity for the $T_1 \rightarrow S_1$ transition, as the energy gaps were between 10 and 1000 cm^{-1}. Similarly, both factors were substantially larger for $T_1 \rightarrow S_2$ compared to $T_1 \rightarrow S_3$ and $T_1 \rightarrow S_4$, though still less than one.[51] The $k'_{ISC}(T_1 \rightarrow S_{2,3,4})$ rates thus depended on both the SOCME and the energy gap (via both the Bolzmann and the FC factors).

In the thesis (Paper **I-III),** and the work by Valiev *et al.* (Paper **IV**),[51] and Peräkylä *et al.* (Paper **V**),[35] the SOC between the states T_1-T_4 and S_1-S_4 were calculated at the CASSCF level of theory but with the XMC-QDPT2 energies as the zero-order energies within perturbation theory using proper active space.[110] Calculating the SOC at the XMC-QDPT2 level of theory is possible but too expensive, and also not really needed as the accuracy of the CASSCF SOCME calculation is already within one order of magnitude. We employed several sets of active space configurations to check which sets of orbitals contributing to the state averaging. These configurations included (6,4), (10,8) and (10,10) active spaces, where the first number in the bracket is the number of electrons and the second is the number of molecular orbitals. The rationale for selecting the active space for our studied cluster was to select the main orbitals that give the largest contributions (configuration interaction weights of more than 0.2) to the relevant low-lying electronic states (T_1...T_4 and S_1...S_4). For instance, the (6,4) active space (six electrons in four orbitals) represents good compromise between the computational cost and accuracy which is sufficient to describe the states of interest for our alkoxy-alkoxy clusters.

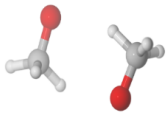
5 Results and discussion

Here, we first present all the energetically low-lying structures (global minima) for relevant model ^{3}(RO…OR') clusters studied in this thesis work (and in related articles where I was a co-author). Then we present reaction rates calculated for the three following reactions: dissociation to RO + R'O, "spin-flip" (intersystem crossing) leading to the formation of ROOR' dimer, and a hydrogen shift leading to the formation of an alcohol and a ketone/aldehyde. In work by Valiev *et al.* (Paper **IV**),[51] we studied the full pathway for ROOR' production for CH_3OO + CH_3OO as a model system, and calculated ISC rates for some small ^{3}(RO…OR') complexes. In Paper **I**, we studied all three product channels for a relatively small set of R and R'. Paper **II** then focuses especially on the ISCs for a limited set of very large and complex R and R'. Finally, paper **III** focuses especially on the H-shift, for a set of mostly medium-size R (with more systematic variation of the substituents). The work by Peräkylä *et al.*[35] (Paper **IV**) presents an additional reaction mechanism and product channel for ^{3}(RO…OR') complexes.

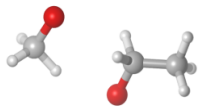
5.1 Configurational sampling and structures of the triplet ^{3}(RO…OR') cluster

We used the Jammy Key for configurational sampling (JKCS) approach[79] to generate ^{3}(RO…OR') cluster structures in our studies in papers **I, II, III** (and other works by *Valiev et al.*,[51] and Peräkylä *et al.*[35]). This approach is generally good enough for systematic conformer studies of atmospheric clusters. However, we found that the default settings did not actually work for our open-shell systems, as many unwanted reactions happened during the xTB optimization. Therefore, we performed several actions to get rid from the unwanted reactions, as described in

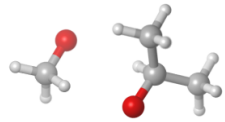
section 4.5.4. All the ^{3}(RO…OR') clusters from papers **I, II** and **III** are collected in **Figure 6**.



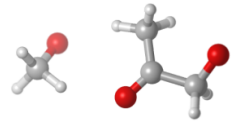
MeO…OMe



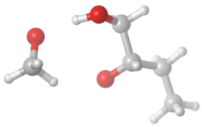
MeO…OEt



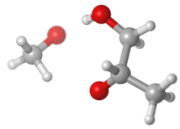
MeO…O*i*Pr



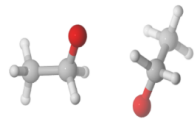
MeO…OAce



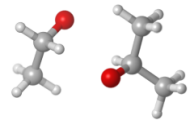
MeO…O-BuOH



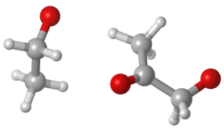
MeO…O*i*PrOH



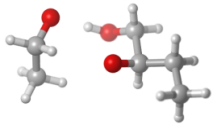
EtO…OEt



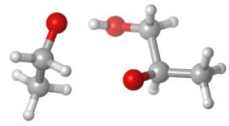
EtO…O*i*Pr



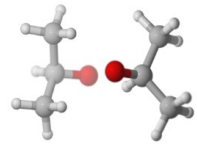
EtO…OAce



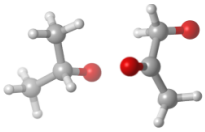
EtO…O-BuOH



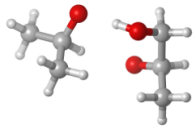
EtO…O*i*PrOH



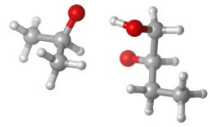
*i*PrO…O*i*Pr



*i*PrO…OAce



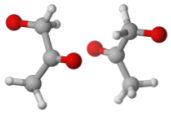
*i*PrO…OPrOH



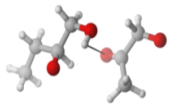
*i*PrO…O-BuOH



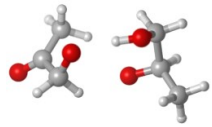
*t*BuO…O*t*Bu



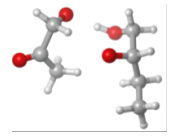
AceO…OAce



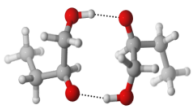
AceO…O-BuOH



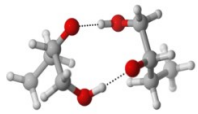
AceO…O*i*PrOH



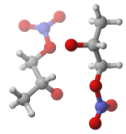
AceO…O-BuOH



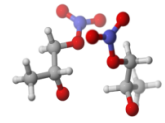
R,S-BuOH-O…O-BuOH



R,R-BuOH-O…O-BuOH



R,S-PrNO$_3$-O…O-PrNO$_3$



R,R-PrNO$_3$-O…O-PrNO$_3$

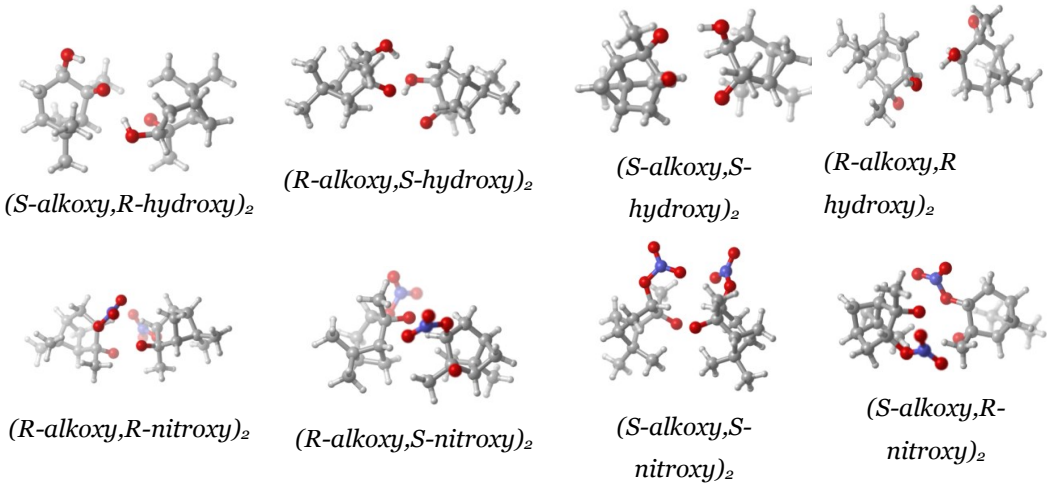


Figure 6. Optimized structures, at the ωb97X-D/aug-cc-pVTZ level, and at ωb97X-D/6-31++G** (given in italics) level for the of the ^{3}(RO…OR') clusters studied in this thesis work. The dotted lines denote H-bonds. Color coding: gray = C, white = H, red = O, blue = N.

5.2 Stability of the ^{3}(RO…OR') cluster with respect to RO + R'O

In **Table 1** we present the energetics for the reaction ^{3}(RO…OR') => RO + R'O for all clusters calculated in Papers **I** and **III** at the ωb97X-D/aug-cc-pVTZ level, and in Paper **II** at the ωb97X-D/6-31++G** level. Here, CCSD(T) results are omitted. For larger clusters we do not have them due to computational reasons, while for the smaller clusters the CCSD(T) results were internally inconsistent (not just inconsistent with DFT), as discussed further below (and in the papers). To evaluate the possible magnitude of basis-set effects on the energetics, we recomputed the MeO…OMe cluster using the smaller basis set (6-31++G**). The obtained electronic and Gibbs free energies for the dissociation reaction (using the sign convention in Table 1) were 2.86 kJmol^{-1} and 0.28 kJmol^{-1} higher, respectively, than those computed with the larger aug-cc-pVTZ basis. In other words, the cluster was more stable (stronger binding) with the smaller basis set, which implies the presence of

some basis-set superposition error, i.e. too strong binding with small basis sets due to the monomers "borrowing" each other's basis functions. However, 2.86 kJmol^{-1} (and especially 0.28 kJmol^{-1}) are fairly small values, given the likely overall errors in treating weakly bound triplet clusters of two radicals.

Table 1: Electronic energies (in kJmol^{-1}, not including zero-point energy), Gibbs free energy (at 298 K and 1 atm reference pressure), and radical distances for the reaction route ^{3}(RO…OR') ⇒ RO + R'O, computed at ωb97X-D/aug-cc-pVTZ level, and at the ωb97X-D/6-31++G** (*given in italics*) levels of theory.

^{3}(RO…OR') Cluster	ΔE_{DFT} (kJmol^{-1})	ΔG_{DFT} (kJmol^{-1})	O⋯O distance in (Å)
MeO…OMe	+13.89	-22.50	3.49
MeO…OEt	+19.12	-16.36	3.49
MeO…O*i*Pr	+15.48	-18.79	3.47
MeO…OAce	+13.97	-21.25	6.14
MeO…O-BuOH	+29.25	-17.28	3.35
MeO…O*i*PrOH	+32.93	-13.43	3.35
EtO…OEt	+23.46	-8.06	3.52
EtO…O*i*Pr	+21.84	-17.45	3.59
EtO…OAce	+25.02	-11.84	4.21
EtO…O-BuOH	+36.07	-10.38	3.27
EtO…O*i*PrOH	+39.71	-5.86	3.27
*i*PrO…O*i*Pr	+19.65	-20.89	3.68
*i*PrO…OAce	+24.39	-17.20	4.16
*i*PrO…O*i*PrOH	+37.57	-8.87	3.31
*i*PrO⋯O-BuOH	+34.85	-13.18	3.33
*t*BuO…O*t*Bu	+19.03	-24.92	3.69
AceO…OAce	+28.68	-12.94	6.47
AceO…*i*PrOH	+40.12	-10.63	3.42
AceO…O-BuOH	+41.42	-11.42	3.43
AceO…O-BuOH	+33.60	-11.38	6.16

R,S-BuOH-O…O-BuOH	+33.57	-23.85	4.29
R,R-BuOH-O…O-BuOH	+56.15	-1.46	3.46
R,S-$PrNO_3$-O…O-$PrNO_3$	+26.48	-20.57	3.89
R,R-$PrNO_3$-O…O-$PrNO_3$	+28.48	-13.75	3.94
α-pinene, (S-alkoxy,R-hydroxy)₂	*+38.24*	*-13.05*	*3.31*
α-pinene, (R-alkoxy,S-hydroxy)₂	*+38.07*	*-12.38*	*3.35*
α-pinene, (S-alkoxy,S-hydroxy)₂	*+40.84*	*-9.29*	*4.13*
α-pinene, (R-alkoxy,R-hydroxy)₂	*+50.04*	*4.06*	*3.50*
α-pinene, (R-alkoxy,R-nitroxy)₂	*+44.81*	*-7.91*	*3.19*
α-pinene, (R-alkoxy,S-nitroxy)₂	*+42.59*	*-6.82*	*5.49*
α-pinene, (S-alkoxy,S-nitroxy)₂	*+45.31*	*-10.92*	*3.48*
α-pinene, (S-alkoxy,R-nitroxy)₂	*+88.45*	*+40.00*	*4.32*

We note that the sign convention in Table 1 differs from that used e.g., in studies of atmospheric molecular clustering, in which energies are typically given for the cluster formation reaction rather than for dissociation. The reason for our choice is that the ^{3}(RO…OR′) ⇒ RO + R′O reaction is the one actually occurring in the atmosphere. The reverse reaction of RO + R'O colliding to form a cluster never happens due to the low concentration of alkoxy radicals. Positive values in Table 1 thus imply that the cluster is below the separated radicals in energy (or free energy), while negative values imply the opposite. The main motivation for computing these energetics was to see how they depend on the of the R/R' group's size and character, as well as the bonding pattern of the cluster. In Paper **I**, we considered MeO…OMe, EtO…OEt, *t*BuO…O*t*Bu, *i*PrO…O*i*Pr, AceO…OBuOH, BuOH-O…O-BuOH and $PrNO_3$-O…O-$PrNO_3$ clusters. We found that all of clusters were below the isolated RO + R'O molecules with respect to the electronic energies (+18.9 to +56.2 kJmol^{-1}at the DFT level) but slightly above them in free energies (-24.92 to -1.5 kJmol^{-1}). This demonstrates that all the clusters are relatively weakly bonded. We also observed that the presence of a large number of individually weak CH…HC and CH…O interactions slightly increases the binding energies. For example, the MeO…OMe cluster was the weakest bound, but the addition of alkyl groups (EtO…OEt, *t*BuO…O*t*Bu, *i*PrO…O*i*Pr) increases the binding energies by a few

kJmol^{-1}. As expected, the presence of H-bonds (AceO…OBuOH, BuOH-O…O-BuOH) stabilized the clusters. On the other hand, aldehyde and nitroxy group had very little effect on the on the ^{3}(RO…OR') stability. Most importantly, we found that stereochemistry had significant effect on the cluster binding energies. For example, the difference in DFT electronic binding energies was 20.9 kJmol^{-1} between the R,R and R,S diastereomers of BuOH-O…O-BuOH. We also noticed some inconsistencies between DFT and CCSD(T)-F12 energies in Paper **I**. We attempted many different approaches to resolve the problem, for example, orbital rotation at the HF stage, but all the attempts were unsuccessful. We also tested different approaches to check whether our clusters possess multireference character or not. However, it turned out that all of them had single reference character – likely because a triplet cluster of two separated radicals is well described by a single triplet reference. Thus, we concluded that we don't know the exact reason for the DFT-CCSD(T) discrepancies, and also note that we did not find any literature reference values for such exotic triplet systems.

In Paper **II**, we generated ^{3}(RO…OR') clusters for RO derived from the major RO_2 formed in the OH and NO_3 – initiated oxidation α-pinene. We considered four stereoisomers in each system due to the presence of two stereocenters. We denoted them as α-pinene, 1) S-alkoxy, R-hydroxy/nitroxy, 2) R-alkoxy, S-hydroxy/nitroxy, 3) S-alkoxy, S-hydroxy/nitroxy, 4) R-alkoxy, R-hydroxy/nitroxy. The binding energies of the lowest-energy conformers found for each stereoisomer pair are given in Table 1. We found that with one exception, all the clusters were bound by between 37.7 and 50.2 kJmol^{-1} in electronic energies, and about 4.2 to -12.6 12 kJmol^{-1} in Gibbs free energies. Again, this implies moderately weak bonding. Also, a similar trend as in Pape**r I** was observed, with the presence of H-bonding leading to more exergonic cluster formation for the OH - containing clusters compared to the smaller clusters in the other papers. However, the ONO_2 containing clusters do not have any H-bonds, and still have fairly similar binding. There can be few reasons behind this strong binding. First, organic molecules with ONO_2 groups typically have high dipole moments due to the electronegativity differences between oxygen (O) and nitrogen (N) on one hand, and hydrogen (H) and carbon (C) on the other. ONO_2 groups can even increase the polarization of adjacent functional groups in

the molecules, further increasing the dipole moments.[114] In clusters of molecules containing ONO_2 groups, these dipoles can interact with each other through dipole-dipole interactions. These interactions are typically attractive, and contribute to cluster stability. Second, ONO_2 groups can exhibit resonance structures, where electrons are delocalized across the oxygen and nitrogen atoms. This delocalization can stabilize the ONO_2 group, and could even extend to neighboring molecules within the cluster, enhancing overall cluster stability.

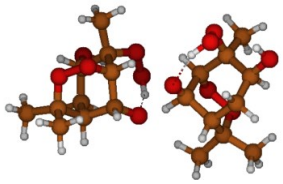
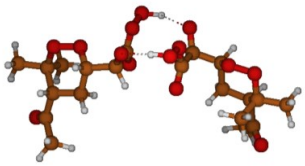
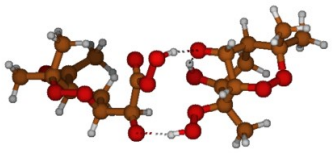
Stereoselectivity also played a significant role in cluster binding energies. The α-pinene,(S-alkoxy,R-nitroxy)$_2$ cluster is an outlier, with considerably stronger bonding than any of the other clusters. As seen from Figure 5, this cluster has a direct interaction between the alkoxy radical on one monomer, and the nitroxy group of the other monomer. Despite extensive testing, we were unable to assess whether the anomalously strong binding of the α-pinene,(S-alkoxy,R-nitroxy)$_2$ system represents an example of genuine strong stereoselectivity in cluster formation, or whether it is a computational artefact.

In Paper **III**, we carried out conformer sampling for a new set of ^{3}(RO···OR') systems with R/ R' = MeO, EtO, iPrO, AceO, HO-PrO and HO-BuO, varying the substituents on both R and R' (where R ≠ R'). We found binding energies between +13.8 to +41.4 kJmol^{-1} in electronic energies, and between +21.3 to -5.9 kJmol^{-1} in Gibbs energies. This was in line with the results of Paper **I and II.** The dominant factor determining the cluster binding energy was again H-bonding. In addition, we also found a common trend in all three papers that adding alkyl groups tends (with some exceptions) to slightly increase binding energies. This is due to the increasing number of weak CH···HC and CH···O interactions. We also observed that the interactions between ketone group and the H-bonded to alkoxyl carbon (α-oxyl C) are substantially weaker than those between the ketone and H atoms bonded to alkyl carbons (β-oxyl C). Unlike in Paper **I**, we found considerably less inconsistency between DFT and F12 binding energies.

The most important trend we observed in our studies that with the increasing size and complexity of the R and R', the binding energies keep going up. To confirm this

trend found in our published work, we decided to generate some further atmospherically relevant cluster and to support the thesis conclusion. We generated three representative clusters ^{3}(RO···OR'), derived from two highly oxidized molecule (HOM)-RO_2 proposed in the literature. $C_{10}H_{15}O_7$ corresponds to the RO derived from the ring-broken O_8-RO_2 from ozone-initiated α-pinene oxidation proposed by Iyer *et al.*[115] while $C_{10}H_{17}O_6$ corresponds to the RO derived from a O_7-RO_2 from OH-initiated α-pinene oxidation proposed by Berndt *et al.*[116]

Table 2: Relative electronic energies (in kJmol^{-1}, not including zero-point energy) and Gibbs free energy (at 298 K and 1 atm reference pressure) for the reaction route ^{3}(RO…OR') => RO + R'O for some larger highly oxygenated molecular (HOM) clusters, computed at the ωb97X-D/6-31+G* level of theory.

^{3}(RO…OR')	ΔE in kJmol^{-1}	ΔG in kJmol^{-1}
 $C_{10}H_{17}O_6$…$C_{10}H_{17}O_6$	+35.53	-14.28
 $C_{10}H_{15}O_7$…$C_{10}H_{15}O_7$	+64.71	+1.51
 $C_{10}H_{17}O_6$…$C_{10}H_{15}O_7$	+56.44	-5.04

We used the same approach of conformer sampling as discussed earlier, though with a slightly smaller basis set, as the system were very large. We found (see Table 2) the same trend as in Paper **I, II** and **III**, with the largest cluster ($C_{10}H_{15}O_7$...$C_{10}H_{15}O_7$) having the highest binding energy, followed by $C_{10}H_{15}O_7$...$C_{10}H_{17}O_6$. The reason for this ordering is the hydrogen bonding (O...H) between two separate RO moieties in these two clusters, and the lack of H-bonding in $C_{10}H_{17}O_6$...$C_{10}H_{17}O_6$. The latter instead forms intramolecular H-bonds, making the cluster less stable. We also note that the $C_{10}H_{15}O_7$...$C_{10}H_{15}O_7$ is more strongly bound in electronic energy than any of the less functionalized systems studied in Papers **I**, **II** and **III** (excluding the anomalous α-pinene, (S-alkoxy,R-nitroxy)$_2$ case).These additional calculations further support our findings that as the size and complexity of the R and R' increases, the binding energies keep going up.

5.3 VTST rate for the RO + R'O ⇒ ^{3}(RO…OR') reaction

In paper **I**, we tried to estimate the dissociation rate for a selected number of ^{3}(RO...OR') systems using variational transition state theory. However, the currently available VTST models were only able to calculate the reaction rate for the RO + R'O collision/association reaction. The actual dissociation rate could thus only be obtained using the detailed balance equation in Eq 41.[109] This gave unphysically large dissociation rates especially in some cases when CCSD(T)-F12 binding energies were used. XMC-QDPT2 binding energies led to more reasonable (though still very fast) rates in the 1.0×10^{10} s^{-1} range for the larger model systems. Surprisingly, for a few systems, the rate predicted using Eq 41 exceeds the typical frequency range of molecular vibrations, especially the intramolecular vibrational frequencies corresponding to the break-up of a cluster, by several orders of magnitude. This is physically impossible, as reactions cannot proceed faster than molecular vibrations. We also note that according to experimental evidence, the dissociation rate should not be orders of magnitude faster than the H-shift rate (as the two channels are roughly competitive for most small R).

A recent work from our group by Lauri Franzon[117] demonstrated a simple trajectory simulation method, which provides reliable (or at least physically consistent) estimates of ^{3}(RO…OR') dissociation rates. The calculated rates for systems larger than ^{3}(MeO…OMe) ranged from 1 to 10^9 s^{-1}, with more complex systems generally having lower rates. This model considers the relative motion of the two molecules rather than the statistical distribution of energy within the complex. The method requires only three parameters as an input: the bimolecular binding energy calculated from the quantum chemical method, together with the intermolecular equilibrium distance and moments of inertia of the complex. This is a good and cheap model for cases where the traditional detailed balance approach fails to determine dissociation rates.

5.4 H-shift rate calculation

H-shift transition states were computed for the systems studied in Papers **I** and **III**. Some representative systems with their H-shift rates are given in **Figure 7**. (The tertiary systems studied in Paper **II** cannot undergo the type of H-shifts studied here.) The rate constants were also calculated using RP-VTST (Reaction Path Variational Transition State Theory) in Paper **I** for four small subsets of the systems. Based on the calculations, the VTST rates were lower than the TST ones (but by less than this factor of 3 at 298 K). The calculated tunneling factor was between 1 and 6, but in over 50 % cases it was close to 1. This suggests that tunneling was not a dominant factor for our studied systems. This it is not very surprising, as our calculated barriers were fairly low. It also indicates that the reaction is proceeding primarily through the classical pathway, overcoming the energy barrier by thermal activation. However, for few clusters, the higher tunneling factor suggest that quantum mechanical tunneling might also be important, with the system having a notable probability of passing through the energy barrier even when its classical energy is lower than the barrier height.

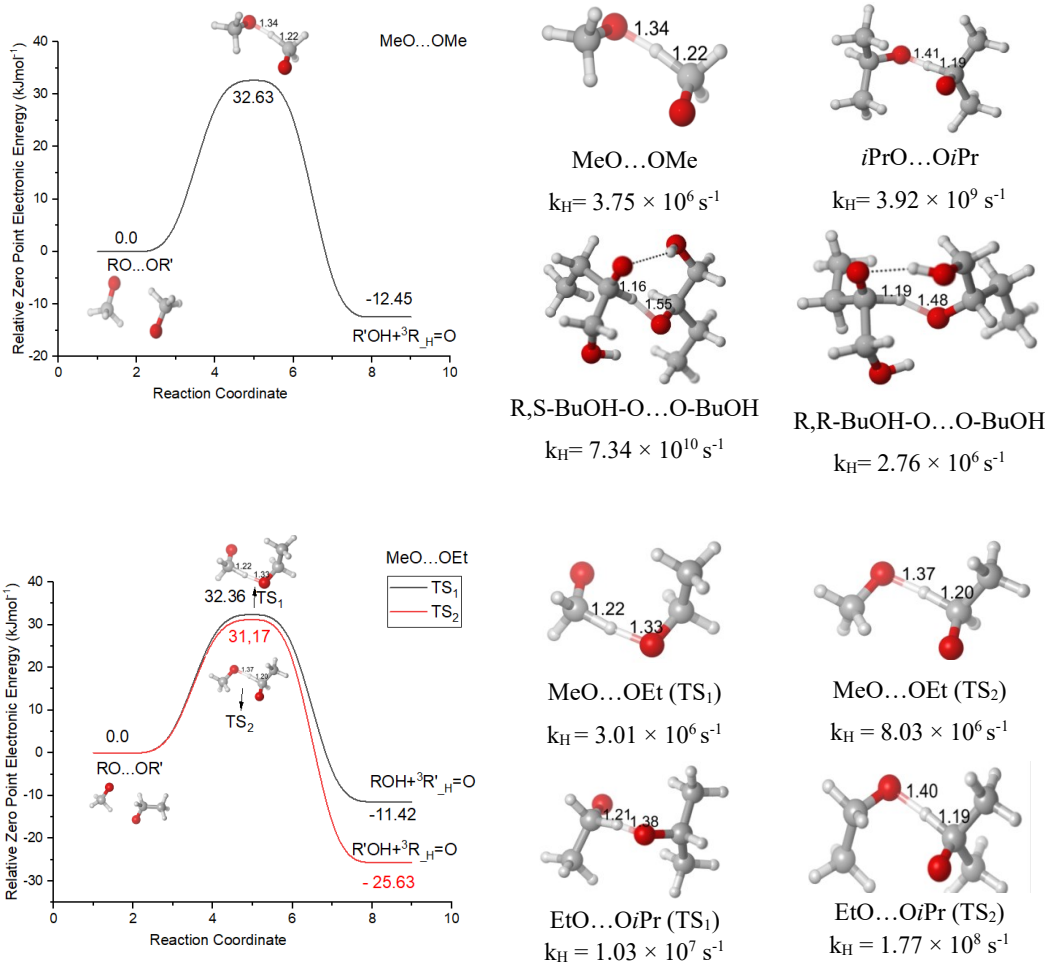


Figure 7. Transition state geometry and potential energy surface for some of representative ^{3}(RO…OR') systems, optimized at the ωb97X-D/aug-cc-pVTZ level. Here the distances between C-H and O-H bonds that are breaking and forming are given in angstrom (Å). Color coding: gray = C, white = H, red =O, blue =N.

In Paper **I,** we discussed the intermolecular hydrogen shift rate calculations for set of homoclusters ^{3}(RO…OR') where R=R'. In Paper **III**, we investigated the H-shift rate constant for a set of heteroclusters ^{3}(RO…OR') where R≠R'. We classified RO as donors and acceptors depending on which RO is donating or accepting a H atom during the H-shift reaction. In both papers, we tried to estimate the reaction rate for a set of ^{3}(RO…OR') complexes at atmospherically relevant conditions. We

described the barrier heights and relative energy between reactants and products of a studied system at different levels of theory. We also showed how substituents (R and R' groups) influenced the barrier heights and reaction rate for a set of ^{3}(RO…OR') complexes. We found that triplet carbonyl (R'$_{-H}$=O) and singlet alcohol (ROH) products were more favorable than singlet carbonyl (R'$_{-H}$=O) and triplet alcohol (ROH) for most of the studied systems. However, for the largest systems in Paper **I** (both diastereomers of BuOH-O…O-BuOH and $PrNO_3$-O…O-$PrNO_3$), the lowest-energy combination of products corresponded to singlet carbonyl and triplet alcohol products due to the presence of additional functional groups on the alcohol products. A similar anomaly was found for the ^{3}(AceO…OBuOH) cluster in Paper **III,** where the ^{1}R$_{-H}$=O + ^{3}R'OH pair was 1.3 kJmol^{-1} lower in energy than ^{3}R$_{-H}$=O + ^{1}R'OH. Overall, the reaction exothermicity (or endothermicity) was explained by the relative energy needed to form one of the two products in the triplet excited state. In Paper **III**, the reaction energies leading to the formation of ^{3}R$_{-H}$=O + ^{1}R'OH were categorized into three groups. The first group included reactions where the (triplet) carbonyl product was formaldehyde (HCHO), which was the least favorable, with reaction energies close to zero. The second group included reactions where the (triplet) carbonyl product is methylglyoxal, which was the most favorable with reaction energies ranging from -79.5 to -108.8 kJmol^{-1}. The third group includes all remaining reactions, which have energies between -12.6 and -33.5 kJmol^{-1}.

In both Paper **I** and **III**, the transition states of the system were found to be between 10.9 and 40.2 kJmol^{-1} above the reactants at the DFT level and between 11.3 and 59.4 kJmol^{-1} at the CCSD(T)-F12 level (hence abbreviated F12), with numbers containing zero-point vibrational energy corrections. In Paper **III**, we found that that the variation in barrier heights was small compared to the systems studied in Paper **I**. These barriers were roughly comparable to those found in the literature for fast unimolecular alkoxy radical H-shifts,[118] as well as to the predictions of Lee *et al.* [33] H-shift barriers in the ^{3}(RO…OR') clusters of of *n*BuO and *s*BuO alkoxy radicals. The DFT and F12 barrier heights were mostly consistent with each other, with the F12 barrier heights typically 8-12 kJmol^{-1} higher than the DFT barrier heights. However, in some cases, the F12 barrier was significantly higher than the DFT barrier due to the HF convergence issues, for example, the difference was 39

kJmol^{-1} for TS_1 of *i*PrO…O-BuOH. We caution that the F12 barriers in these cases are likely to be highly unreliable. All the H-shift rate constants were high, ranging from 1×10^6 s^{-1} to 9.9×10^{10} s^{-1} at the DFT level and 6 s^{-1} to 7.3×10^{10} s^{-1} after F12 energy correction. We also found (in Paper **I**) that stereoisomers and functional groups have a modest effect on the H-shift rate.

Nevertheless, we found that the barrier height was affected by three important factors: the weakening of the C-H bond strength in the donor RO due to interactions with adjacent functional groups, the overall binding energy of the cluster, and the thermodynamic stability of the reaction products. In order the understand the variation in barrier heights we tested various different approaches in Paper **III**. First, we compared our calculated barrier height with the reference structure activity relations ship (SAR) values calculated by Vereecken *et al.* [114] and found that the correlation was almost nonexistent. In the SAR, two main factors determine the rate: the "span", which refers to the number of atoms between the shifting hydrogen (H) atom and the accepting oxygen (O) atom in the transition state, and the substituents on the carbon atom from which the H atom is abstracted (donor). Our reactions lack a "span", as the donor and acceptor are on different molecules (not covalently connected). Second, we checked the barrier heights dependence on the cluster binding energies.

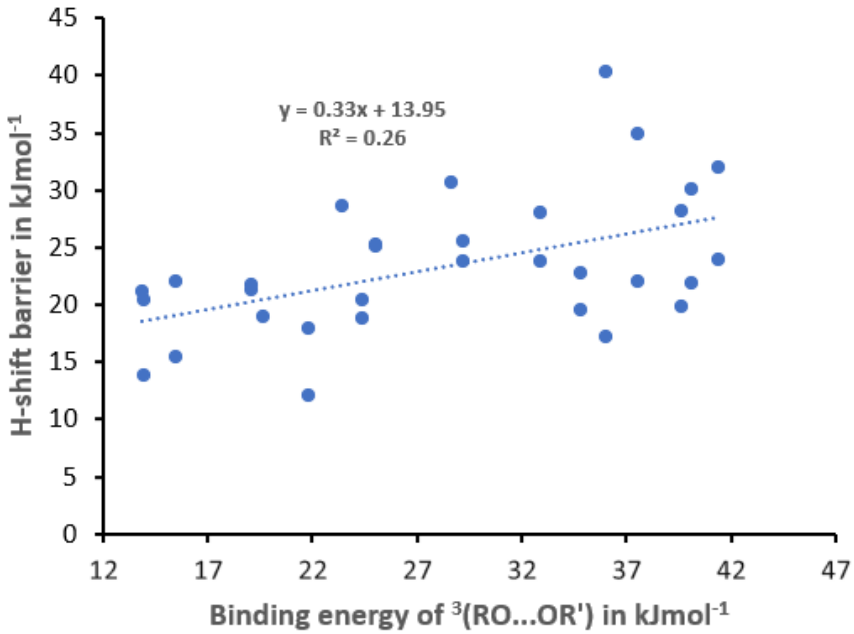


Figure 8. H-shift barrier height vs the binding energy of the ^{3}(RO…OR') complex, at the UωB97X-D/aug-cc-pVTZ level.

We found that there was a weak but non-negligible positive correlation between binding energies and barrier heights (**Figure 8**). For example, an increase in binding energy by 4.2 kJmol^{-1} will, on average, increase the H-shift barrier by 1.3 kJmol^{-1}. As the R and R' become more complex and functionalized, the binding energy keeps going up, and thus the H-shift rate goes down. This can be thought to play a role analogous to the "span" in the unimolecular case, i.e., strongly bonded clusters mean we have a higher energy penalty for the distortion of the cluster from the minimum geometry to the TS (just like the small-span unimolecular H-shifts have a large penalty).

5.5 Inter system crossing rate constant

The third channel studied in the thesis is an intersystem crossing (ISC) of ^{3}(RO…OR') complexes to the singlet surface, allowing for recombination to ROOR'. We first optimized all the (RO…OR') complexes on the singlet surface to check whether the complex eventually recombines to form singlet ROOR'. We found that 75 % of the triplet cluster recombine to form singlet ROOR'. Table 2 shows the values of the rate constant for intersystem crossing (k_{ISC}) for different ^{3}(RO…OR') complexes calculated in Papers **I, II, III** and **IV**[51]. Here, k_{ISC} refers to the transition from the triplet state to the singlet state, which can lead to the formation of ROOR'.

Table 3: ISC rate constants for our studied systems at 298 K, based on energies of triplet and singlet states computed using XMC-QDPT2(6,4)/6-311++G(d,p) and matrix elements of the spin-orbit coupling interaction (SOCME) between T_1-T_4 and S_1-S_4 states computed at the CASSCF(6,4)/6-311++G(d,p) level.

^{3}(RO···OR') cluster	**Σk_{ISC} (s^{-1})**
MeO…OMe	5.0×10^{12}
EtO…OEt	3.5×10^{9}
*i*PrO…O*i*Pr	8.0×10^{9}
*t*BuO…O*t*Bu	2.1×10^{9}
AceO…OAce	4.0×10^{8}
AceO…OBuOH	9.0×10^{8}
R,S-BuOH-O…O-BuOH	4.0×10^{3}
R,R-BuOH-O…O-BuOH	2.0×10^{9}
R,S-$PrNO_3$-O…O-$PrNO_3$	1.2×10^{9}
R,R-$PrNO_3$-O…O-$PrNO_3$	4.6×10^{9}
MeO…OEt	2.3×10^{9}
MeO…O*i*Pr	2.1×10^{9}
MeO…Oace	2.3×10^{9}

MeO…OBuOH	7.9×10^9
MeO…O*i*PrOH	7.9×10^9
EtO…O*i*Pr	4.1×10^9
EtO…OAce	1.5×10^9
EtO…OBuOH	1.4×10^9
EtO…O*i*PrOH	1.3×10^9
*i*PrO…OAce	9.0×10^8
*i*PrO…O*i*PrOH	3.0×10^9
*i*PrO…OBuOH	6.3×10^9
AceO…O*i*PrOH	1.0×10^8
AceO…OBuOH	1.1×10^9
α-pinene, (*S*-alkoxy,*R*-hydroxy)$_2$	1.2×10^{10}
α-pinene, (*R*-alkoxy,*S*-hydroxy)$_2$	5.7×10^7
α-pinene, (*S*-alkoxy,*S*-hydroxy)$_2$	9.4×10^9
α-pinene, (*R*-alkoxy,*R*-hydroxy)$_2$	1.5×10^{10}
α-pinene, (*R*-alkoxy,*R*-nitroxy)$_2$	1.6×10^{10}
α-pinene, (*R*-alkoxy,*S*-nitroxy)$_2$	1.0×10^6
α-pinene, (*S*-alkoxy,*S*-nitroxy)$_2$	5.7×10^{10}
α-pinene, (*S*-alkoxy,*R*-nitroxy)$_2$	2.1×10^8

From Table 3, we can see that the ISC rates varied between 4.0×10^3 s^{-1} and 5.0×10^{12} s^{-1}. However, most of the other systems have k_{ISC} values ranging from 10^8 to 10^{10} s^{-1}. The MeO…OMe and R,S-BuOH-O…O-BuOH systems from Paper **I** are the outliers with extreme ISC rates. The rest of the data suggest that almost all the studied systems have reasonably fast ISC rates.

We also found in Papers **I-III** and the work by Valiev *et al.* (in Paper **IV)** [51] and Peräkylä *et al.* (Paper **V**)[35] that the ISC rates strongly depend on the conformation of the ^{3}(RO···OR') clusters, and they also displayed stereoselectivity, with the ISC rates for different stereoisomer pairs in some cases varying by over 4 orders of magnitude. The overall ISC rate was dominated by the ISC between T_1 and S_1 for most studied systems, except for the cases, where the SOCME between T_1 and S_1 is zero or small. The variation in the ISC rate between T_1 and S_1 is driven almost

exclusively by variations in the SOCME (rather than the energy gap), which, in turn, is related to the orientation of the C-O bonds relative to each other. However, ISCs to the excited singlet state S_2 can also have high rates despite moderate energy gaps (on the order of 1500-2500 cm^{-1}). We found that the ISC rates for the studied systems are similar to the H-shift rates, suggesting that both channels are competitive.

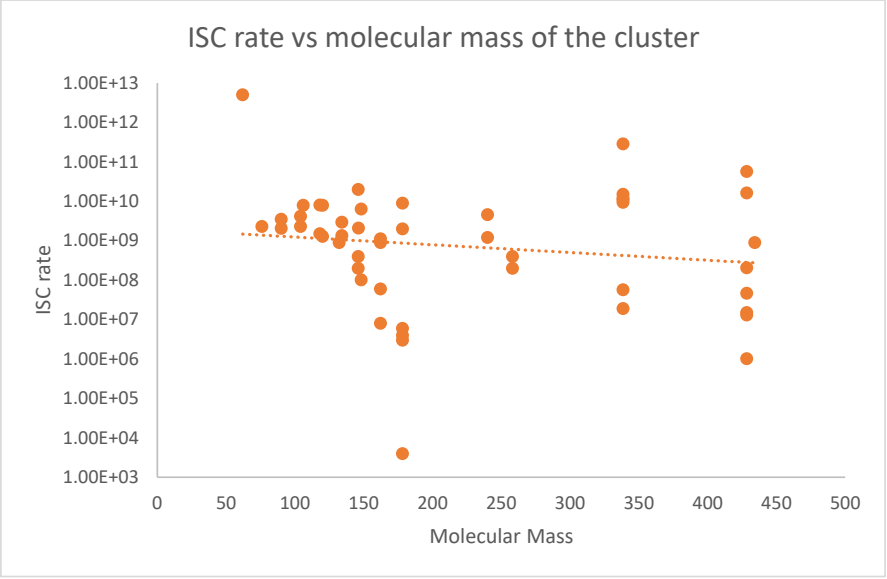


Figure 9: ISC rate vs. molecular mass for all the molecular ^{3}(RO…OR') clusters studied in this thesis.

Finally, we wanted to see if there is an overall trend in the ISC rates with respect to the system sizes. We plotted the ISC rate constants versus the molecular masses of all the studied systems, even including the local minima conformers calculated in Papers **I-III, IV**[51] and **V**[35] (see Figure 9). The rates are plotted on a logarithmic scale, and the fitted regression is thus an exponential trend. We found that there is little correlation between the ISC rate and the system size (molecular mass). Interestingly, for the smaller molecular masses, the ISC rates are clustered together, albeit with few outliers (e.g. MeO…OMe and R,S-BuOH-O…O-BuOH). As the molecular masses increases, the spread of the values become larger and larger.

Overall, the ISC rate doesn't seem to depend on the system size (or the functional groups), as there was very little correlation between them.

6 Conclusions and outlook

Paper **I**: In this work, we calculated the energetics and reaction rates of ^{3}(RO…OR') complexes that are the key intermediates in the RO_2 + R'O_2 reactions relevant to atmospheric chemistry. Our extensive computational study confirmed the hypothesis (proposed in earlier work[51]) that the reaction rates for RO_2 + R'O_2 are controlled by the formation and decomposition of the initial RO_4R' tetroxide intermediate. Furthermore, the dissociation reaction of the ^{3}(RO…OR') complex was predicted to be extremely fast. The study also indicated that functionalization affected the intermolecular H-shift rates less than the corresponding unimolecular H-shift rates. The study further confirmed and extended the prediction that intersystem crossings allowing the recombination of ROOR's accretion products are extremely fast and competitive in atmospheric gas phase condition. However, our quantitative rate estimations were inconsistent with experimental constraints, as the dissociation rates predicted for all but the largest studied systems imply a near-unity branching ratio for the RO + R'O channel, in contradiction with measurements. This study also provided possible explanations for this discrepancy, including that the binding energies predicted for the ^{3}(RO…OR') complexes may be too low, or that the RO + R'O association rates might also be too high. The study also suggested that successful modeling of the reaction channels will require methods going beyond statistical rate theories, and may require direct dynamic simulations, which will face significant challenges in accounting for the complicated electronic structure of the system and the calculation of intersystem crossing rates on-the-fly.

Paper **II:** In this work, we calculated the intersystem crossings (ISC) rate for ^{3}(RO…OR') complexes formed in the self-reactions of a set of RO_2 stereoisomers produced in the OH- and NO_3-initiated oxidation of α-pinene. We found that the

computed ISC rates were very fast, indicating that the proposed mechanism can explain accretion product formation observed in α-pinene oxidation products. The variation of calculated ISC rates between stereoisomers suggested that accretion product formation may be strongly stereoselective. Overall, the study provided insights into the formation of large ROOR' accretion products in atmospheric conditions.

Paper **III**: In the last thesis paper, we investigated substituents effects on the H-shift reaction channel that leads to the formation of carbonyl and alcohol products (which are much more volatile than the ROOR' accretion products). The reaction of interest involved an intermolecular H-shift inside a triplet intermediate complex ^{3}(RO…OR') of two alkoxy radicals. As compared to papers **I** and **II**, this study also involved a larger variety of different systems. We obtained both minimum-energy geometries of ^{3}(RO…OR') and H-shift transition states for a total of 14 systems, each with two possible H-shift reaction pathways. The calculated rate coefficients showed surprisingly small variation, with most H-shift rates being higher than 10^8 s^{-1} but lower than 10^9 s^{-1}. The effect of functional groups on the H-shift rates was a bit difficult to explain. We proposed a hypothesis that involves a competition between three different effects: 1) the weakening of the breaking C-H bond due to interactions with adjacent functional groups, 2) the overall binding energy of the ^{3}(RO…OR'), which tends to increase the barrier height, and 3) the thermodynamic stability of the reaction products. Intersystem crossing (ISC) rates were also computed for the same ^{3}(RO…OR') systems, and found to be of as the same order of magnitude as the H-shift rates, suggesting that the two studied channels should be competitive for small and medium-sized systems. For larger and more complex systems, we predicted that the increase in ^{3}(RO…OR') binding energy will tend to raise H-shift barriers and ultimately render the alcohol + carbonyl pathway uncompetitive. Thus, we have explained why accretion product formation becomes extremely efficient for complex RO_2.

From Papers **I-III**, we can also make a grand conclusion that our cost-effective systematic conformer search approach performs is reasonably good at finding minima of ^{3}(RO…OR') complex clusters (and also ^{3}(RO…R) clusters in work by

Valiev *et al.*[35]) on the triplet potential energy surface. One of the interesting findings of the work is that as the R and R' become more complex and functionalized, the binding energy of the complex keeps going up, and the H-shift rate and the dissociation rate keep going down, while the inter-system-crossing (ISC) does not change very much (with the exception of some outliers). Note that this trend holds even for the detailed balance computed dissociation rates, despite the actual values likely being substantially overestimated. Our overall results confirm the proposed novel RO_2 + RO_2 reaction mechanism for accretion product formation, which certainly contributes to organic aerosol formation.

Future outlook: Despite having many successes in this thesis work, we are still missing some crucial results. First, the branching ratio between the three channels studied in this work is still missing. We managed to produce some qualitative results, but due to the uses of different levels of theory, and various substantial error sources, actual branching ratios cannot be obtained. However, if only the three previously known RO_2 + $R'O_2$ channels were active, we would not actually need detailed branching ratios. The work in this thesis as well as Lauri Franzon's[117] dissociation model has shown that (and why) we can safely assume the branching ratio for accretion products is close to 1 for large enough systems. However, work by Peräkylä *et al.*[35] (Paper **V**) shows that there are many accretion product channels, and estimating the branching between these still requires ISC calculations. This branching may be important in the atmosphere, since esters and peroxides have very different chemical and thermal behavior. Specifically, peroxides might break up much quicker, and thus not be so good at making SOA. ISC calculations, while doable for large ^{3}(RO…OR') systems as shown in this thesis, require the use of very expensive multireference methods. In this thesis work, we have calculated ISC rates for 46 different clusters over several years. This is not a huge number. Therefore, faster methods are needed, since there are so many possible relevant RO…OR', RO…R' etc. clusters for which to calculate them.
Thus, we need more advanced machine learning, deep learning and probabilistic models to predict ISC rates with less computational expense, and also an automated method to generate relevant RO…OR', RO…R' etc. cluster structures efficiently. Existing models such as GECKO-A,[119] a generator for explicit chemistry and kinetics

of organic compounds in the atmosphere, need to be updated for generating atmospherically relevant (and HOM-relevant, *i.e.* large) RO…OR', RO…R' and other similar clusters from lists of all possible peroxy radicals. We aim to develop a new "GECKO plus accretion reactions" type code, which can give us structures (e.g. SMILES strings) for all possible ^{3}(RO…OR') formed in RO_2 + $R'O_2$ reactions. Then we need a fast machine learning - based model to evaluate the ISC rates for the corresponding ^{3}(RO…OR') clusters. For this, we need to generate data sets using multireference methods to predict the ISC rates for some subset of the systems generated by this approach.

However, the approach proposed above would still be useless (for machine-learning ISC rates) without the conformational sampling codes developed in this thesis, which allow us to get reasonable representative ^{3}(RO…OR') geometries after e.g. a modified GECKO-A code gives us a SMILES string for the RO. However, some further work is likely needed to develop our sampling algorithms for cases where the also monomers have a huge number of conformers.

7 References

(1) Hallquist, M. *et al.*, The Formation, Properties and Impact of Secondary Organic Aerosol: Current and Emerging Issues. *Atmos. Chem. Phys.* **2009**, *9*, 5155–5236.

(2) Brauer, M *et al.*, A. Ambient Air Pollution Exposure Estimation for the Global Burden of Disease. *Environ Sci Technol*, **2016**, *50*, 79–88.

(3) Loeb, N. G.; Kato, S. Top-of-Atmosphere Direct Radiative Effect of Aerosols over the Tropical Oceans from the Clouds and the Earth's Radiant Energy System (CERES) Satellite Instrument. *J Clim*, **2002**, *15*, 1474–1484.

(4) Rosenfeld, D. *et al.*, Global Observations of Aerosol-Cloud-Precipitation-Climate Interactions. *Reviews of Geophysics,* **2014**, *52*, 750–808.

(5) Merikanto, J. *et al.*, Impact of Nucleation on Global CCN. *Atmos Chem Phys*, **2009**, *9*, 8601–8616.

(6) Jimenez, J. L. *et al.*, Evolution of Organic Aerosols in the Atmosphere. *Science,* **2009**, *326*, 1525–1529.

(7) Kirkby, J. *et al.*, Ion-Induced Nucleation of Pure Biogenic Particles. *Nature,* **2016**, *533*, 521–526.

(8) Zhang, X. *et al.*, Influence of Vapor Wall Loss in Laboratory Chambers on Yields of Secondary Organic Aerosol. *Proceedings of the National Academy of Sciences,* **2014**, *111*, 5802–5807.

(9) Tröstl, J. *et al.*, The Role of Low-Volatility Organic Compounds in Initial Particle Growth in the Atmosphere. *Nature* **2016**, *533*, 527–531.

(10) Glowacki, D. R.; Pilling, M. J. Unimolecular Reactions of Peroxy Radicals in Atmospheric Chemistry and Combustion. *ChemPhysChem*, **2010**, *11*, 3836–3843.

(11) Orlando, J. J.; Tyndall, G. S. Laboratory Studies of Organic Peroxy Radical Chemistry: An

Overview with Emphasis on Recent Issues of Atmospheric Significance. *Chem Soc Rev,* **2012**, *41*, 6294–6317.

(12) Ehn, M. *et al.,* A Large Source of Low-Volatility Secondary Organic Aerosol. *Nature,* **2014**, *506*, 476–479.

(13) Savee, J. D. *et al.*, Direct Observation and Kinetics of a Hydroperoxyalkyl Radical (QOOH). *Science,* **2015**, *347*, 643–646.

(14) Riipinen, I. *et al.*, Organic Condensation: A Vital Link Connecting Aerosol Formation to Cloud Condensation Nuclei (CCN) Concentrations. *Atmos Chem Phys,* **2011**, *11*, 3865–3878.

(15) Criegee, R. Mechanism of Ozonolysis. *Angewandte Chemie International Edition in English,* **1975**, *14*, 745–752.

(16) Vereecken, L.; Novelli, A.; Taraborrelli, D. Unimolecular Decay Strongly Limits the Atmospheric Impact of Criegee Intermediates. *Physical Chemistry Chemical Physics,* **2017**, *19*, 31599–31612.

(17) Long, B.; Bao, J. L.; Truhlar, D. G. Atmospheric Chemistry of Criegee Intermediates: Unimolecular Reactions and Reactions with Water. *J Am Chem Soc,* **2016**, *138*, 14409–14422.

(18) Vereecken, L. Lifting the Veil on an Old Mystery. *Science,* **2013**, *340*, 154–155.

(19) Taatjes, C. A.; Shallcross, D. E.; Percival, C. J. Research Frontiers in the Chemistry of Criegee Intermediates and Tropospheric Ozonolysis. *Phys. Chem. Chem. Phys.,* **2014**, *16*, 1704–1718.

(20) Bianchi, F. *et al.*, Highly Oxygenated Organic Molecules (HOM) from Gas-Phase Autoxidation Involving Peroxy Radicals: A Key Contributor to Atmospheric Aerosol. *Chem Rev,* **2019**, *119*, 3472–3509.

(21) Teng, A. P.; Crounse, J. D.; Wennberg, P. O. Isoprene Peroxy Radical Dynamics. *J Am Chem Soc*, **2017**, *139*, 5367–5377.

(22) Praske, E. *et al.*, Atmospheric Autoxidation Is Increasingly Important in Urban and Suburban North America. *Proceedings of the National Academy of Sciences,* **2018**, *115*, 64–69.

(23) Xu, L. *et al.*, Unimolecular Reactions of Peroxy Radicals Formed in the Oxidation of α-Pinene and β-Pinene by Hydroxyl Radicals. *J Phys Chem A,* **2019**, *123*, 1661–1674.

(24) Berndt, T. *et al.*, A. Accretion Product Formation from Self- and Cross-Reactions of RO_2 Radicals in the Atmosphere. *Angewandte Chemie - International Edition,* **2018**, *57*, 3820–3824.

(25) Kalberer, M. *et al.*, Identification of Polymers as Major Components of Atmospheric Organic Aerosols. *Science,* **2004**, *303*, 1659–1662.

(26) Nguyen, T. B. *et al.*, S. A. Effect of Humidity on the Composition of Isoprene Photooxidation Secondary Organic Aerosol. *Atmos Chem Phys,* **2011**, *11*, 6931–6944.

(27) Berndt, T. *et al.*, Hydroxyl Radical-Induced Formation of Highly Oxidized Organic Compounds. *Nat Commun,* **2016**, *7*, 13677.

(28) Berndt, T. *et al.*, Gas-Phase Ozonolysis of Cycloalkenes: Formation of Highly Oxidized RO_2 Radicals and Their Reactions with NO, NO_2, SO_2, and Other RO_2 Radicals. *J Phys Chem A,* **2015**, *119*, 10336–10348.

(29) Noell, A. C. *et al.*, Near-Infrared Kinetic Spectroscopy of the HO_2 and $C_2H_5O_2$ Self-Reactions and Cross Reactions. *J Phys Chem A,* **2010**, *114*, 6983–6995.

(30) Zhang, P. *et al.,* Theoretical Study on the Mechanism and Kinetics for the Self-Reaction of $C_2H_5O_2$ Radicals. *J Phys Chem A,* **2012**, *116*, 4610–4620.

(31) Salo, V.-T. *et al.*, Gas-Phase Peroxyl Radical Recombination Reactions: A Computational Study of Formation and Decomposition of Tetroxides. *J Phys Chem A,* **2022**, *126*, 4046–4056.

(32) Ghigo, G. *et al.*, Combustion and Atmospheric Oxidation of Hydrocarbons: Theoretical Study of the Methyl Peroxyl Self-Reaction. *Journal of Chemical Physics,* **2003**, *118*, 10575–10583.

(33) Lee, R. *et al.*, Why Are: Sec -Alkylperoxyl Bimolecular Self-Reactions Orders of Magnitude Faster than the Analogous Reactions of Tert -Alkylperoxyls ? The Unanticipated Role of CH Hydrogen Bond Donation. *Phys. Chem. Chem. Phys.,* **2016**, *18*, 23673–23679.

(34) Vereecken, L. Peeters, J. Decomposition of Substituted Alkoxy Radicals–Part I: A Generalized Structure-Activity Relationship for Reaction Barrier Heights. *Phys. Chem. Chem. Phys.* **2009**, *11*, 9062-9074.

(35) Peräkylä, O. *et al.*, Large Gas-Phase Source of Esters and Other Accretion Products in the Atmosphere. *J Am Chem Soc,* **2023**, *145*, 7780–7790.

(36) Berndt, T. *et al.*, Accretion Product Formation from Ozonolysis and OH Radical Reaction of α-Pinene: Mechanistic Insight and the Influence of Isoprene and Ethylene. *Environ Sci Technol,* **2018**, *52*, 11069–11077.

(37) Dirac, P., The Quantum Theory of the Electron. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character,* **1928**, *117*, 610–624.

(38) Schrödinger, E. An Undulatory Theory of the Mechanics of Atoms and Molecules. *Physical Review*, **1926**, *28*, 1049–1070.

(39) Born, M.; Oppenheimer, R. Zur Quantentheorie Der Molekeln. *Ann Phys,* **1927**, *389*, 457–484.

(40) Hartree, D. R. The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part II. Some Results and Discussion. *Mathematical Proceedings of the Cambridge Philosophical Society*, **1928**, *24*, 111–132.

(41) Hartree, D. R. The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods. *Mathematical Proceedings of the Cambridge Philosophical Society*, **1928**, *24*, 89–110.

(42) Fock, V. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Zeitschrift für Physik,* **1930**, *61*, 126–148.

(43) Roos, B. O. et al., *Multiconfigurational Quantum Chemistry*. Hoboken, New Jersey: John Wiley & Sons, 2016

(44) Bartlett, R. J. Coupled-Cluster Approach to Molecular Structure and Spectra: A Step toward Predictive Quantum Chemistry. *J Phys Chem*, **1989**, *93*, 1697–1708.

(45) Van Voorhis, T.; Head-Gordon, M. Benchmark Variational Coupled Cluster Doubles Results. *J Chem Phys*, **2000**, *113*, 8873–8879.

(46) Knizia, G.; Adler, T. B.; Werner, H.-J. Simplified CCSD(T)-F12 Methods: Theory and Benchmarks. *J Chem Phys*, **2009**, *130*, 054104.

(47) Adler, T. B.; Knizia, G.; Werner, H.-J. A Simple and Efficient CCSD(T)-F12 Approximation. *J Chem Phys,* **2007**, *127*, 221106.

(48) Finley, J.; Malmqvist, P.-Å.; Roos, B. O.; Serrano-Andrés, L. The Multi-State CASPT2 Method. *Chem Phys Lett*, **1998**, *288*, 299–306.

(49) Granovsky, A. A. Extended Multi-Configuration Quasi-Degenerate Perturbation Theory: The New Approach to Multi-State Multi-Reference Perturbation Theory. *J Chem Phys,* **2011**, *134*, 214113.

(50) Nakano, H. Quasidegenerate Perturbation Theory with Multiconfigurational Self-Consistent-Field Reference Functions. *J Chem Phys,* **1993**, *99*, 7983–7992.

(51) Valiev, R. R. *et al.*, Intersystem Crossings Drive Atmospheric Gas-Phase Dimer Formation. *J Phys Chem A,* **2019**, *123*, 6596–6604.

(52) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Physical Review*, **1964**, *13*, B864–B871.

(53) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review*, **1965**, *140*, A1133–A1138.

(54) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J Chem Phys,* **1993**, *98*, 5648–5652.

(55) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys Rev A (Coll Park),* **1988**, *38*, 3098–3100.

(56) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys Rev B*, **1988**, *37*, 785–789.

(57) Perdew, J. P. Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas. *Phys Rev B,* **1986**, *33*, 8822–8824.

(58) Chai, J.-D.; Head-Gordon, M. Long-Range Corrected Hybrid Density Functionals with Damped Atom–Atom Dispersion Corrections. *Phys. Chem. Chem. Phys.,* **2008**, *10*, 6615–6620.

(59) Myllys, N.; Elm, J.; Kurtén, T. Density Functional Theory Basis Set Convergence of Sulfuric Acid-Containing Molecular Clusters. *Comput Theor Chem,* **2016**, *1098*, 1–12.

(60) Myllys, N. *et al.*, Coupled Cluster Evaluation of the Stability of Atmospheric Acid–Base Clusters with up to 10 Molecules. *J Phys Chem A,* **2016**, *120*, 621–630.

(61) Elm, J.; Bilde, M.; Mikkelsen, K. V. Assessment of Density Functional Theory in Predicting Structures and Free Energies of Reaction of Atmospheric Prenucleation Clusters. *J Chem Theory Comput*, **2012**, *8*, 2071–2077.

(62) Zerner, M. C. Semiempirical Molecular Orbital Methods. *Reviews in computational chemistry*, **1991**, *2*, 313–365.

(63) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-XTB–An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J Chem Theory Comput*, **2019**, *15*, 1652–1671.

(64) Bannwarth, C. *et al*., Extended tight-binding Quantum Chemistry Methods. *WIREs Computational Molecular Science,* **2021**, *11*.

(65) Lennard-Jones, J. E. Cohesion. *Proceedings of the Physical Society,* **1931**, *43*, 461–482.

(66) Glendening, E. D.; Reed, A. E.; Carpenter, J. E.; Weinhold, F. **NBO**, *Version 3.1*. Gaussian Inc. 2003.

(67) C. J. Cramer. *Essentials of Computational Chemistry - Theories and Models*, *2nd ed*.; John Wiley and Sons; **2004**.

(68) Friedman, P. A. R. *Molecular Quantum Mechanics*, *4th ed*.; Oxford University Press, **2005**.

(69) Huzinaga, S.; McWilliams, D.; Domsky, B. Approximate Atomic Wavefunctions. *J Chem Phys,* **1971**, *54*, 2283–2284.

(70) Boys, S. F.; Bernardi, F. The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors. *Mol Phys,* **1970**, *19*, 553–566.

(71) Simon, S.; Duran, M.; Dannenberg, J. J. How Does Basis Set Superposition Error Change the Potential Surfaces for Hydrogen-Bonded Dimers? *J Chem Phys,* **1996**, *105*, 11024–11031.

(72) Foresman, J. B.; Frisch, A. E. Exploring Chemistry with Electronic Structure Methods, *3rd ed*. Wallingford, CT, USA: Gaussian. Inc 2015.

(73) Atkins, P. W.; De Paula, J. *Atkins' Physical Chemistry*; Oxford university press, 2014.

(74) Atkins, P. W.; Friedman, R. S. *Molecular Quantum Mechanics*; Oxford university press, 2011.

(75) Ribeiro, R. F. *et al.*, Use of Solution-Phase Vibrational Frequencies in Continuum Models for the Free Energy of Solvation. *J Phys Chem B,* **2011**, *115*, 14556–14562.

(76) Grimme, S. Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory. *Chemistry - A European Journal,* **2012**, *18*, 9955–9964.

(77) Spartan 20, Wavefunction, Inc.: Irvine, C.A, 2020.

(78) Spartan 16, Wavefunction, Inc.: Irvine, C.A, 2016.

(79) Kubečka, J. *et al.*, Configurational Sampling of Noncovalent (Atmospheric) Molecular Clusters: Sulfuric Acid and Guanidine. *J Phys Chem A,* **2019**, *123*, 6022–6033.

(80) Zhang, J.; Dolg, M. Global Optimization of Clusters of Rigid Molecules Using the Artificial Bee Colony Algorithm. *Phys. Chem. Chem. Phys.,* **2016**, *18*, 3003–3010.

(81) Zhang, J.; Dolg, M. ABCluster: The Artificial Bee Colony Algorithm for Cluster Global Optimization. *Phys. Chem. Chem. Phys.,* **2015**, *17*, 24173–24181.

(82) Yu, W.; He, X. *et al.* Extension of the CHARMM General Force Field to Sulfonyl-Containing Compounds and Its Utility in Biomolecular Simulations. *J Comput Chem,* **2012**, *33*, 2451–2468.

(83) Vanommeslaeghe, K. *et al.*, CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J Comput Chem,* **2009**, 31, 671–690.

(84) Johnston, R. L. Evolving Better Nanoparticles: Genetic Algorithms for Optimising Cluster Geometries. *Dalton Transactions,* **2003**, 22, 4193–4207.

(85) Karaboga, D.; Basturk, B. On the Performance of Artificial Bee Colony (ABC) Algorithm. *Appl Soft Comput*, **2008**, *8*, 687–697.

(86) Dieterich, J. M.; Hartke, B. OGOLEM: Global Cluster Structure Optimisation for Arbitrary Mixtures of Flexible Molecules. A Multiscaling, Object-Oriented Approach. *Mol Phys,* **2010**, *108*, 279–291.

(87) Kanters, R. P. F.; Donald, K. J. cluster: Searching for Unique Low Energy Minima of Structures Using a Novel Implementation of a Genetic Algorithm. *J Chem Theory Comput,* **2014**, *10*, 5729–5737.

(88) Zhang, J.; Dolg, M. Global Optimization of Clusters of Rigid Molecules Using the Artificial Bee Colony Algorithm. *Phys. Chem. Chem. Phys.*, **2016**, *18*, 3003–3010.

(89) Luchini, G. *et al.*, GoodVibes: Automated Thermochemistry for Heterogeneous Computational Chemistry Data. *F1000Res,* **2020**, *9*, 291.

(90) Temelso, B. *et al.*, ArbAlign: A Tool for Optimal Alignment of Arbitrarily Ordered Isomers Using the Kuhn–Munkres Algorithm. *J Chem Inf Model* **2017**, *57*, 1045–1054.

(91) Walker, M. W.; Shao, L.; Volz, R. A. Estimating 3-D Location Parameters Using Dual Number Quaternions. *CVGIP: Image Understanding,* **1991**, *54*, 358–367.

(92) Kabsch, W. A Solution for the Best Rotation to Relate Two Sets of Vectors. *Acta Crystallographica Section A,* **1976**, *32*, 922–923.

(93) Eyring, H. The Activated Complex in Chemical Reactions. *J Chem Phys,* **1935**, *3*, 107–115.

(94) Evans, M. G.; Polanyi, M. Some Applications of the Transition State Method to the Calculation of Reaction Velocities, Especially in Solution. *Transactions of the Faraday Society,* **1935**, *31*, 875–894.

(95) Fernandez-Ramos, A. *et al.,* Variational Transition State Theory with Multidimensional Tunneling. *Reviews in Computational Chemistry*, **2007**, *23*, 125–232.

(96) Bao, J. L.; Truhlar, D. G. Variational Transition State Theory: Theoretical Framework and Recent Developments. *Chem Soc Rev,* **2017**, *46*, 7548–7596.

(97) Eckart, C. The Penetration of a Potential Barrier by Electrons. *Phys. Rev.* **1930**, *35*, 1303–1309.

(98) MathWorks. MATLAB R2018b. Natick, Massachusetts, United States. 2018.

(99) Møller, K. H. *et al.*, Cost-Effective Implementation of Multiconformer Transition State Theory for Peroxy Radical Hydrogen Shift Reactions. *J Phys Chem A,* **2016**, *120*, 10072–10087.

(100) Zhang, F.; Dibble, T. S. Impact of Tunneling on Hydrogen-Migration of the n-Propylperoxy Radical. *Phys. Chem. Chem. Phys.,* **2011**, *13*, 17969–17977.

(101) Skodje, R. T.; Truhlar, D. G.; Garrett, B. C. A General Small-Curvature Approximation for

Transition-State-Theory Transmission Coefficients. *Journal of Physical Chemistry,* **1981**, *85*, 3019–3023.

(102) Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. Current Status of Transition-State Theory. *J Phys Chem*, **1996**, *100*, 12771–12800.

(103) Garrett, B. C.; Truhlar, D. G. Generalized Transition State Theory. Classical Mechanical Theory and Applications to Collinear Reactions of Hydrogen Molecules. *Journal of Physical Chemistry,* **1979**, *83*, 1052–1079.

(104) Gao, J.; Truhlar, D. G. Quantum Mechanical Methods for Enzyme Kinetics. *Annu Rev Phys Chem,* **2002**, *53*, 467–505.

(105) Pilling, M. J. Radical-Radical Reactions. *Annual Review of Physical Chemistry*, **1996**, 81–108.

(106) Naroznik, M.; Niedzielski, J. Recombination of Tert-Butyl Radicals: The Role of Weak van Der Waals Interactions. *Theor Chem Acc,* **2002**, *108*, 103–112.

(107) Limitations of Variational Transition State Theory for Barrierless Radical–Radical Recombination Reactions. *Zeitschrift für Physikalische Chemie*. **2004**, *218*, 457–468.

(108) Georgievskii, Y.; Klippenstein, S. J. Variable Reaction Coordinate Transition State Theory: Analytic Results and Application to the $C_2H_3+H\rightarrow C_2H_4$ reaction. *Journal of Chemical Physics,* **2003**, *118*, 5442–5455.

(109) Myllys, N. *et al*., Guanidine: A Highly Efficient Stabilizer in Atmospheric New-Particle Formation. *J Phys Chem A*, **2018**, *122*, 4717–4729.

(110) Valiev, R. R. *et al*., First-Principles Method for Calculating the Rate Constants of Internal-Conversion and Intersystem-Crossing Transitions. *Phys. Chem. Chem. Phys.*, **2018**, *20*, 6121–6133.

(111) Plotnikov, V. G. Regularities of the Processes of Radiation less Conversion in Polyatomic Molecules. *Int J Quantum Chem,* **1979**, *16*, 527–541.

(112) Bixon, M.; Jortner, J. Intramolecular Radiationless Transitions. *J Chem Phys,* **1968**, *48*, 715–726.

(113) A. Szabo and N. S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications: New York, 1996.

(114) Iyer, S. *et al.*, Modeling the Detection of Organic and Inorganic Compounds Using Iodide-Based Chemical Ionization. *J Phys Chem A*, **2016**, *120*, 576–587.

(115) Iyer, S. *et al.,* Molecular Mechanism for Rapid Autoxidation in α-Pinene Ozonolysis. *Nat Commun*, **2021**, *12*, 878.

(116) Berndt, T. *et al.*, Hydroxyl Radical-Induced Formation of Highly Oxidized Organic Compounds. *Nat Commun*, **2016**, *7*, 13677.

(117) Franzon, L. Simple Physical Model for the Estimation of Irreversible Dissociation Rates for Bimolecular Complexes. *J Phys Chem A* **2023**, *127*, 5956–5966.

(118) Vereecken, L.; Peeters, J. A Structure-Activity Relationship for the Rate Coefficient of H-Migration in Substituted Alkoxy Radicals. *Phys. Chem. Chem. Phys.* **2010**, *12*, 12608-12620.

(119) Aumont, B.; Szopa, S.; Madronich, S. Modelling the Evolution of Organic Carbon during Its Gas-Phase Tropospheric Oxidation: Development of an Explicit Model Based on a Self-Generating Approach. *Atmos Chem Phys*, **2005**, *5*, 2497–2517.