



LARGE AND MULTI SCALE MECHANISTIC MODELING OF DIELS-ALDER REACTIONS

A thesis submitted in fulfillment of the requirements for the degree

of

MASTER OF SCIENCE IN CHEMISTRY

of

RHODES UNIVERSITY, SOUTH AFRICA

Department of Chemistry / Faculty of Science

By

Bienfait Kabuyaya Isamura

BSc(Hons)

21B8638

ORCID iD: 0000-0003-3278-3284

Supervisor: Prof. Kevin Lobb

2021

Table of contents

Table of contents	2
Summary.....	4
Abbreviations	7
Acknowledgment.....	9
Dedication	10
INTRODUCTION.....	12
Chapter 1: Conceptual background.....	18
1. Overview	19
2. The Diels-Alder reaction	20
2.1. The mechanism.....	20
2.2. Applications	25
2.3. Diels-Alder reaction of the [60] fullerene.....	27
3. Cheminformatics and computational methods	31
3.1. Molecular representation: SMILES strings	31
3.2. Chemical databases: ZINC	34
3.3. Computational methods of calculation.....	36
3.4. Hybrid methods: the ONIOM approach.....	46
3.5. Transition state prediction	48
4. Reaction force analysis (RFA)	49
4.1. Standard RF theory	49
4.2. Atomic resolution of energy derivatives.....	50
Chapter 2: The AMADAR program	53
0. Summary.....	54
1. Introduction.....	55
2. IT requirements	56
3. The Algorithm	57
4. Configuration files	66
5. Performance and case study.....	74
6. Concluding remarks	78
Chapter 3: On the (a)synchronicity of Diels-Alder reactions	80
0. Summary.....	81
1. Introduction.....	82
2. Methodology	83

3. The dataset.....	84
4. Reaction force analysis	87
5. Natural population analysis	94
6. Atomic decomposition of energy derivatives	100
7. Concluding remarks	105
Chapter 4: The Diels-Alder reaction of C ₆₀ fullerene	108
0. Summary.....	109
1. Introduction.....	110
2. Methodology	110
3. The ONIOM models	112
4. Accuracy and performance of ONIOM models	114
5. C ₆₀ DA reaction energetics and structural data	120
6. Concluding remarks	130
CONCLUSION AND PERSPECTIVES.....	132
Appendix.....	135
References.....	138

Summary

The [4+2] cycloaddition reaction between conjugated dienes and substituted alkenes is known as the Diels-Alder (DA) reaction, in honor of two German chemists, Otto Diels and Kurt Alder, who first reported this marvelous chemical transformation. The DA reaction is one of the most popular reactions in organic chemistry, allowing for the regio- and stereospecific establishment of six-membered rings with up to four stereogenic centers. This pericyclic reaction has found many applications in areas as diverse as natural products chemistry, polymer chemistry, and agrochemistry.

Over the past decades, the mechanism of the Diels-Alder (DA) reaction has been the subject of numerous studies, dealing with questions as diverse as the mechanistic pathway, the synchronicity, the use of catalysts, the effect of solvents and salts, etc. On the other hand, as an example, fullerenes (and particularly [60] fullerene) have been found to act as good dienophiles in DA reactions to the extent that many functionalized fullerenes with interesting applications are still synthesized by reacting C₆₀ with dienes. However, despite the very abundant literature about the mechanism of the DA reaction, some pertinent questions have been still pending, including, without being restricted to, the prediction of transition state (TS) geometries and the modeling of DA reactions involving large systems, such as those of C₆₀ fullerene.

It must be emphasized that TSs are not easy to predict and the main reason is that many existing algorithms require that the search is initiated from a good starting point (guess TS), which must be very similar to the actual TS. This problem is even more difficult when many TSs are to be located as may be the case in large-scale studies. Moreover, due to the large size of the C₆₀ molecule, the usage of accurate high-level computational methods in the investigation of its reactivity towards dienes is computationally costly, implying the need to find the best threshold between accuracy and computational cost.

Therefore, the present study was carried out to contribute to solving the problems of large-scale prediction of DA transition state geometries and the multi-scale modeling of C₆₀ fullerene DA reactions.

To address the first problem (large-scale prediction of TSs), we have developed a python program named “AMADAR”, which predicts an unlimited number of DA transition states, using only the SMILES strings of the cycloadducts. AMADAR is customizable and allows for the description of intramolecular DA reactions as well as systems resulting in competing paths.

In addition, The AMADAR tool contains two separate modules that perform reaction force analyses and atomic decomposition of energy derivatives from the predicted Intrinsic Reaction Coordinates (IRC) paths.

The performance of AMADAR was assessed using 2000 DA cycloadducts and showed a success rate of $\sim 95\%$. Most of the errors were due to basis set inconsistencies or convergence issues that we are still working on. Furthermore, a set of 150 IRC paths generated by the AMADAR program were analyzed to get insight into the (a)synchronicity of DA reactions. This investigation confirmed that the reaction force constant κ (second derivatives of the system energy with respect to the reaction coordinate) was a good indicator of synchronicity in DA reactions. A close inspection of the profile of κ has enabled us to propose an alternative classification of DA reactions based on their synchronicity degree, in terms of (quasi)-synchronous, moderate asynchronous, asynchronous, and likely two-steps DA reactions. Natural population analyses seemed to indicate that the global maximum of the reaction force constant could be identified with the formation of all the bonds in the reaction site. Finally, the atomic resolution of energy derivatives suggested that the mechanism of the DA reaction involves two inner elementary processes associated with the formation of each C-C bond. A striking mechanistic difference between synchronous and asynchronous DA reactions emerging from this study is that, in asynchronous reactions, the driving and retarding forces are mainly caused by the fast and slow-forming bonds (elementary process) respectively, while in the case of synchronous ones both elementary processes retard and drive the process concomitantly and equivalently.

Regarding the DA reaction of C_{60} fullerene that was considered to illustrate the problem of multiscale modeling, we have constructed 12 ONIOM2 and 10 ONIOM3 models combining five semi-empirical methods (AM1, PM3, PM3MM, PDDG, PM6) and the LDA(SVWN) functional in conjunction with the B3LYP/6-31G(d) level. Then, their accuracy and efficiency were assessed in comparison with the pure B3LYP/6-31G(d) level considering first the DA reaction between C_{60} and cyclopentadiene whose experimental data are available. Further, different DFT functionals were employed in place of the B3LYP functional to describe the higher-layer of the best ONIOM partition, and the results obtained were compared to experimental data. At this step, the ONIOM2(M06-2X/6-31 G(d): SVWN/STO-3G) model, where the higher layer encompasses the diene and pyracyclene portion of C_{60} , was found to provide the best tradeoff between accuracy and cost, with respect to experimental data. This model showed errors lower than 2.6 and 2.0 kcal/mol for the estimation of the activation and

reaction enthalpies respectively. We have also demonstrated, by comparing several ONIOM2(DFT/6-31G(d): SVWN/STO-3G) models, the importance of dispersion corrections in the accurate estimation of reaction and activation energies.

Finally, we have considered a set of 21 dienes, including anthracene, 1,3-butadiene, 1,3-cyclopentadiene, furan, thiophene, selenothiophene, pyrrole and their mono-cyano and hydroxyl derivatives to get insight into the DA reaction of C_{60} using the best ONIOM2(M06-2X/6-31 G(d): SVWN/STO-3G) model. For a given diene and its derivatives, the analysis of frontier molecular orbitals provides a consistent explanation for the substituent effect on the activation barrier. It revealed that electron-donating (withdrawing) groups such as -OH (-CN) cut down on the activation barrier of the reaction by lowering (extending) of the $HOMO_{diene} - LUMO_{C_{60}}$ gap and consequently enhancing (weakening) the interaction between the two reactants. Further, the decomposition of the activation energy into the strain and interaction components suggested that, for a given diene, electron-donating groups (here -OH) diminish the height of the activation barrier not only by favoring the attractive interaction between the diene and C_{60} , but also by reducing the strain energy of the system; the opposite effect is observed for electron-withdrawing groups (here -CN). In contrast with some previous findings on typical DA reactions, we could not infer any general rule applicable to the entire dataset for the prediction of activation energies because the latter does not correlate well with either of the TS polarity, electrophilicity of the diene, or the reaction energy.

Abbreviations

AM1: Austin Model 1

AMADAR: Automated workflow for Mechanistic Analysis of Diels-Alder Reactions

CASSCF: Complete Active Space Self-Consistent Field

CC: Coupled Cluster

CNDO: Complete Neglect of Differential Overlap

DA: Diels-Alder

DBOC: Diagonal Born-Oppenheimer Correction

DFT: Density Functional Theory

GGA: Generalized Gradient Approximation

GSM: Growing String Method

GTO: Gaussian-Type Orbital

HOMO: Highest Occupied Molecular Orbital

HOMA: Harmonic Oscillator Model of Aromaticity

INDO: Intermediate Neglect of Differential Overlap

IRC: Intrinsic Reaction Coordinate

L(S)DA: Localized (Spin) Density Approximation

LUMO: Lowest Unoccupied Molecular Orbital

MO: Molecular Orbital

NDDO: Neglect of Diatomic Differential Overlap

NEB: Nudged-Elastic Band

NICS: Nucleus Independent Chemical Shift

NOMO: Nuclear Orbital plus Molecular Orbital

NPA: Natural Population Analysis

PDI: Paradelocalization Index

PMn: Parametric Method n

QM: Quantum Mechanics

QM/MM: Quantum Mechanics / Molecular Mechanics

RFA: Reaction Force Analysis

SMILES: Simplified Molecular Input Line Entry System

STO: Slater-Type Orbital

SW: Surface Walking

TS: Transition State

TST: Transition State Theory

UFF: Universal Force Field

ZDO: Zero Differential Overlap

Acknowledgment

I would like to acknowledge all the people (and institutions) whose support, hands, words, minds and/or love made possible the achievement of this project.

I am grateful to my supervisor, Prof Kevin Lobb, who guided me all along this hectic journey of trials and errors. I could not expect more from him. May God reward you.

This work would never have started without the financial support of the BEBUC scholarship system, which granted me a Master's bursary covering all my living expenses and academic fees. I am also thankful to the Department of Chemistry, which gave me a comfortable place to learn and grow in science.

My gratitude is also addressed to my parents and siblings, who did not stop supporting me. On a special tone, I say thank you to my mother who kept telling me that I have seeds of greatness and that I will certainly make it.

I cannot forget the invaluable help and assistance I got from my lab mates, namely Washington Dendera, Yolanda Novokoza, Sophakama Zabo and Mofeli Leoma. I appreciated your company, and, overall, the encouraging vibes we used to share when the laboratory atmosphere smelled depression and discouragement.

To my friends, Grady Mukubwa, Justin Safari, Choonzo Chiyumba, Yolande Openda, Urbain Ndagano, Temlandvo Matshidiso, and Diana Kalume, thank you so much for your time and love. You have been such a blessing for me and made my stay in Grahamstown so adorable. I will always keep positive memories of any second we spent together.

Thank you very much to any other person who, implicitly or explicitly, contributed to the accomplishment of the present work.

Bienfait K. Isamura

Dedication

I dedicate this thesis to my parents, Angelus Paluku Isamura and Marie Stella Masika. May this scientific work be a reward for your inestimable efforts invested in my education and the inspiring models of hard work you are and will remain to me.

“Theories are like mathematical formulas. As the formulas help us to solve the sums,
similarly, the theories help us to establish our points in research”

Ziaul Haque

INTRODUCTION

For centuries, chemistry remained a totally experimental field. The observation of plenty of natural phenomena and the characterization of a variety of substances allowed experimental chemists to establish numerous empirical laws that are still in use to make predictions and explain new findings [1]. More interestingly, with technological progress, sophisticated instruments such as spectrometers, microscopes, etc. have been built up and these materials are used routinely in chemical laboratories to probe the reality from the very bottom of the matter [2].

However, one of the main drawbacks of laboratory experiments is that they are very time-consuming and error-prone, to the extent that several trials are often needed before any successful endeavor. In addition, some phenomena, although described and admitted theoretically, cannot be observed in the laboratory. This is the case of transition states, which can't be observed due to their transient nature and great instability [3].

Since the advent of powerful computers, computational chemistry has emerged as a complementary tool, helping chemists to circumvent some of the limitations of experiments. Computational chemistry allows for the (*in silico*) simulations of chemical processes and properties, including those which are difficult or even impossible to investigate experimentally. It is often used to confirm and rationalize experimental findings, make predictions of new materials or properties not yet measured about known substances [4]. In the past years, molecular modeling has got so much success, and has become a real research domain.

Despite the great assistance that computational chemistry is offering to chemists to tackle many chemistry-related problems, especially those for which experimentation alone is not sufficient or feasible, numerous challenges still need to be addressed concerning the cost and the efficiency of computational methods used for modeling molecular systems. Among these are the prediction of transition state geometries and the description of large systems.

On one hand, locating TS geometries on the potential energy surface of chemical reactions is one of the most challenging tasks in computational chemistry, often requiring expert-based knowledge and permanent human intervention. The main reason is that most existing codes require that the search is initiated from a good starting (guess) geometry that is structurally very similar to the actual TS [5]. Unfortunately, this guess structure is not easy to get, and the problem even worsens when many TSs are to be predicted.

On the other hand, despite being known to provide accurate results, approximate quantum mechanics methods are not suitable when it comes to treating large systems. Furthermore, force

field-based methods, which are commonly used to describe large molecular systems, do not offer quantitatively reliable data in some cases (and these methods certainly cannot handle bond-forming or breaking events). Therefore, there is the need to find the best trade-off between accuracy and computational cost in capable methods. Hybrid methods such as the ONIOM approach [6] seem to be the alternative solution as they allow for the usage of two or even three different methods to treat different parts of molecular systems. Unfortunately, it is not always easy to partition the system at hand and decide which method to use for each part to ensure the best compromise between accuracy and cost.

The present work is an attempt to address these two problems of large-scale and multi-scale molecular modelling, restricting the investigation to the mechanism of the Diels-Alder (DA) reactions, and that of [60] fullerene in particular. Why the Diels-Alder reaction? Why especially [60] fullerene?

First, the DA reaction is one of the most important reactions in organic chemistry and a typical case study for undergraduate chemistry students. The relevance of the DA reaction lies in the many applications it has found in domains as diverse as the total synthesis of natural products[7], polymer chemistry [8] as well as the industrial preparation of agrochemicals and fragrances [9]. Second, fullerenes are known to undergo DA reactions with several dienes. However, due to their large size, modelling their reactivity at high accurate computational levels is very time-consuming. As such, it is necessary to find the best threshold between accuracy and cost. These two questions have motivated our present study.

To reach the overall goal of this project, i.e. contributing to solving the problem of largescale and multiscale modelling of DA reactions, six specific objectives were set:

1. Largescale modeling of DA reactions. In this context, the aims were to:
 - develop a program that can predict an unlimited number of DA transition state geometries and carry out, in a quasi-automated fashion, reaction path analyses;
 - test the performance of the program on a large number of systems;
- analyze the results obtained and identify common patterns in the mechanism of DA reactions, especially in terms of their (a)synchronicity. In this study, the synchronicity of DA reactions is defined as the difference between the lengths of the two nascent C-C bonds at the TS. Note however that several other definitions have been proposed to estimate the synchronicity of cycloaddition reactions, such as considering the relative

variation of Wiberg indices of all the bonds directly involved in the active site of the reaction[10][11].

2. Multiscale modelling of [60] fullerene DA reactions. In this context, the aims were to:

- design multi-layer ONIOM models for the description of [60] fullerene DA reactions;
- assess the performance and accuracy of each model in comparison with quantum mechanics calculations and experimental data;
- study more systems using the optimum hybrid model.

Apart from this introduction and the conclusion at the end, the present thesis encompasses four chapters:

Chapter 1: Conceptual Background

In this chapter, we present the concepts necessary for a better understanding of the findings reported in later chapters.

Chapter 2: The AMADAR program

This chapter is about a python program that we have written to respond to the problem of largescale prediction of DA reactions. This program has been named “AMADAR”, which stands for “Automated workflow for Mechanistic Analysis of Diels-Alder Reactions”. The AMADAR program is hosted in a GitHub repository, freely accessible and usable after minor customization ([CMCDD/AMADAR \(github.com\)](https://github.com/CMCDD/AMADAR)).

Chapter 3: On the (a)synchronicity of Diels-Alder reactions

In this chapter, the AMADAR program is applied on a large set of DA cycloadducts, and the findings are mainly discussed in terms of the (a)synchronicity degree of the corresponding DA reactions in the context of the reaction force theory and atomic resolution of energy derivatives.

Chapter 4: The Diels-Alder reaction of C₆₀ fullerene

We present in this chapter different ONIOM models and assess their performances and accuracy in the investigation of the DA reaction between [60] fullerene and cyclopentadiene. Then, we apply the best model on 21 DA reactions to understand the effect of substituent of the activation energy and assess the validity of the Bell-Evans-Polanyi principle in this special

case. This study involves the analysis of frontier-molecular orbital energies, charge transfer effects, and reaction path analysis.

“We think there is color, we think there is sweet, we think there is bitter, but in reality there are atoms and a void”

Democritus, C.460 – C 370 B.C

Chapter 1: Conceptual background

1. Overview

This chapter is a literature review, written to provide the reader with the concepts upon which the chapters of this thesis are built. The discussion below is divided into three sections:

- The Diels-Alder reaction:

In this first section, we have reviewed previous works on the mechanism of the Diels-Alder reaction and its applications. A special paragraph (1.3) has been devoted to discussing the Diels-Alder reaction of the [60] fullerene, which constitutes a special case requiring multiscale modeling.

- Cheminformatics and Computational methods

Therein, we discuss the concept of molecular representation and chemical databases, before providing an overview of the most common computational methods used to model chemical systems, their performance, advantages, drawbacks and expected accuracy.

- Reaction force analysis

The theoretical background of the reaction force theory is presented, as well as the atomic decomposition of energy derivatives. The two approaches, implemented in the AMADAR package, are used in our work to get insight into the mechanism of DA reactions.

2. The Diels-Alder reaction

The [4+2] cycloaddition reaction between a conjugated diene and a dienophile is known as the Diels-Alder (DA) reaction. Discovered in 1928 by Otto Diels and Kurt Alder [12], the Diels-Alder reaction is one of the most important ring-forming reactions available in organic chemistry. This pericyclic process allows for the synthesis of six-membered rings with up to four stereogenic centers [13]. A distinctive feature of the reaction is its high flexibility, i.e. the ease with which the practitioners can customize the diene and dienophile to reach their synthetic targets.

Over the past decades, several versions of the DA reaction (including, but not restricted to, intramolecular, hetero, and asymmetric variants) have been developed and employed to solve a variety of chemical puzzles. One of the key applications of these Diels-Alder reactions is their usage in the total synthesis of natural products [7][14]. In this section, we review some of the most striking mechanistic features and applications of this very important organic reaction. A special section is devoted to the Diels-Alder reaction of C₆₀ fullerene.

2.1. The mechanism

The mechanism of the Diels-Alder (DA) reaction has been at the core of numerous enthusiastic studies and hot debates during past decades, from both experimental [15] and theoretical views [16]. Reported studies have concerned the (most likely) mechanistic pathway[17][18], the usage of catalysts [19][20], the aromaticity of the TS [21], the (a)synchronicity [22], or the effect of solvents [23][24] and salts [25] on the rate and stereoselectivity of the reaction.

a) The mechanistic pathway

First and foremost, much time has been invested in determining the most likely mechanistic pathway that a specific DA reaction should follow, between concerted and stepwise routes. The concerted pathway assumes no intermediate between the reactant and products configurations. As such, the reacting system passes through a unique transition state before relaxing towards the cycloadduct state. On the other hand, the stepwise mechanism postulates the existence of a diradical or zwitterionic intermediate between the initial and final equilibrium states, which then results in a two-steps process [16]. Reported studies indicate that either of these two steps can dictate the rate of the reaction [26].

Theoretical studies based on the transition state theory have asserted the preference of the concerted pathway over the stepwise one in many DA reactions. For example, Houk and his coworkers investigated the feasibility of the two mechanisms in the case of the prototypical DA reaction between butadiene and ethylene. These authors found the energy barrier of the concerted transition state to be lower of 7.7 kcal/mol than that of the first step of the stepwise trajectory at the B3LYP/6-31G(d) level. This difference was reduced to 2.3 kcal/mol when energy corrections were inserted to account for the spin-contamination of the diradical transition structure[27]. Li et al revisited the same reaction at the CASSCF/6-31G(d) level and obtained a gap of 5.0 kcal/mol between the concerted TS and the first of the stepwise trajectory[28]. Similarly, James et al studied the mechanism of the DA reactions between acrolein and several 1,3-dienes using the B3LYP and M06-2X density functionals in conjunction with the 6-31G(d) basis set and found the concerted cycloaddition was to be favored over the diradical pathway of about 2.5 and 5.3 kcal/mol respectively for 2,3-dimethyl-1,3-butadiene and cyclopentadiene. However, they did notice that (*E*)-1,3-pentadiene had no preference for either of the two pathways, as it showed a small energy gap of only 1.2 kcal/mol between the TS geometry of the concerted route and that of the stepwise trajectory. This result was supported by experimental findings revealing that a mixture of (*E*)-1,3-pentadiene and acrolein lead both to the cycloaddition product and the copolymer[17].

Despite the panoply of studies supporting the concertedness of many DA reactions, some isolated cases have been reported, where the DA [4+2] cycloaddition occurs preferentially following the stepwise mechanism. For example, the DA reaction between 4-nitrobenzodifuroxan (NBDF) and the Danishefsky's diene (1-methoxy-3-trimethylsilyloxy-1,3-butadiene) was investigated experimentally and theoretically by Steglenko et al. These authors found this reaction to proceed through the intermediacy of a zwitterionic nitrosilate with two possible conformations (cis and trans)[29]. Furthermore, DFT calculations suggested that the conversion between the cis and trans conformations of this intermediate could take place either by returning to the pre-reaction complex or by rotation through a less stable intermediate. Similarly, Pham et al showed that the intramolecular DA reaction between tryptamine-derived Zincke aldehydes and strychnine proceeds via a stepwise pathway[30]. In addition, they particularly highlighted the crucial role of the potassium counterion in the Zincke aldehyde diene, which seemed to promote its pre-organization in an *s-cis* conformation suitable for bicyclization. In the same thinking track, Wakayama et al investigated the DA reaction between 1,3-butadiene and disilene using the CASSF (complete active space self-consistent

field) level of theory[31]. Their calculations predicted the activation energy of the concerted TS to be 9.0 kcal/mol higher than that of the first step of the stepwise route.

b) Synchronicity

Today, DA reactions are almost universally admitted to be concerted processes. However, being concerted does not imply that they are all synchronous; because, on a time-resolved scale, there is no guarantee that the two nascent C-C bonds will be formed concomitantly. According to Dewar, a synchronous multi-bond reaction is one in which the bonds are formed in unison[22]. The mechanistic relevance of assessing the synchronicity of DA reactions is probably grounded in the fact that the more asynchronous a reaction tends to be, the more likely it is to follow a stepwise mechanism and vice-versa[32]. From a computational view, the simplest and most used indicator of the DA degree of synchronicity consists in measuring the difference between the lengths of the two emerging C-C bonds at the TS. Recently, Merino et al showed that topological analyses could provide information regarding synchronicity that, often, is not reflected in the geometry of transition structures[33]. However, the former structural estimator is still the most used.

The widespread opinion was that unsymmetrically substituted dienes/dienophiles lead to unsymmetrical TS and thus asynchronous mechanisms[32]. However, Souza et al found that this was not correct on a time-resolved scale and provided strong evidence showing that even unsymmetrical antagonists could proceed via synchronous routes[34]. Such controversies illustrate the need to dig deeper into the essence of synchronicity in DA reactions.

Various studies have attempted to elucidate the origin of asynchronicity in DA reactions. Houk et al ascribed the asynchronicity in the DA reaction between 1,3-butadiene and acrolein to the gap in the molecular orbital coefficients of the two C atoms in the LUMO of the acrolein molecule[35]. The symmetric distribution of the electron-density excess reached in the electrophilic reactants through the charge transfer process was advanced by Domingo et al as the main factor controlling the synchronicity in polar DA reactions[36]. Moreover, considering a dataset of 20 DA reactions, Bachrach et al found that the symmetry of the DA transition state depends on the interplay of orbital interactions, aromaticity, which favors a C_s structure, and strain, which tends to break symmetry[37]. By comparing the strain energies measured at the TS of their reactions, these authors provided a mnemonic for predicting the symmetry (synchronicity) of the TS: if the strain energy is greater than 10 kcal/mol, the TS is likely to be asymmetric; otherwise, it will be symmetric. Another interesting study based on the reaction

force theory was reported by Yepes et al, who demonstrated that the second derivative of the reaction energy along the reaction path, known as the reaction force constant, constituted a good indicator of the synchronicity of DA reactions[38]. Recently, Bickelhaupt and his co-workers reported that the asynchronicity in DA reactions could be rationalized in terms of the Pauli repulsion between the occupied molecular orbitals of the antagonists[39].

We must admit that all of these approaches have contributed to piercing the mystery behind the origin of (a)synchronicity in DA reactions. However, some other pertinent questions still needed to be tackled, such as the role of each atom (especially those involved in the reactive site) in the overall (a)synchronicity of DA reactions. This fundamental question is one of the problems addressed in this thesis, based on the atomic resolution of energy derivatives as suggested by Ledrzejewski et al [40].

c) Aromaticity of DA transition states

Aromaticity is an essential concept in chemistry, employed to account for the unusual stability, reactivity, molecular structures, and other properties of many unsaturated organic compounds[41]. It must be emphasized that aromaticity does not imply lack of reactivity. For example, despite the greater aromaticity of the anthracene central ring (based on geometric and magnetic considerations), it is more prone towards DA cycloaddition than the less aromatic outer rings. This is known as the “anthracene problem”[42].

Several computational studies have investigated the aromaticity of DA reactions transition states using different tools[21][43]. Early on, it was assumed that thermally induced pericyclic reactions should occur through an aromatic TS whereas their photochemical counterparts proceeded preferentially via an antiaromatic TS[44]. Jiao and Schleyer provided evidence supporting this assumption by showing that TSs of thermally induced pericyclic reactions have not only delocalized structures and large resonance stabilizations, but also strongly enhanced magnetic susceptibilities and appreciable NICS (nucleus-independent chemical shifts) values arising from the diatropic ring currents[45]. Furthermore, Matito et al investigated the validity of different aromaticity descriptors for the prediction of the changes in aromaticity and planarity along the reaction path of the Diels–Alder reaction between ethene and 1,3-butadiene[46]. These authors showed that the results given by the NICS (Nuclear Independent Chemical Shift), at variance with those obtained by means of the HOMA (Harmonic Oscillator Model of Aromaticity) [47] for example, confirmed the existence of an aromatic transition state for this reaction.

While the NICS has been successfully used to assess the aromaticity of many cyclic organic structures[48] and transition structures[46], its application to the study of the aromaticity of DA TS does not escape from criticism[49]. To support this argument, Bachrach et al demonstrated that NICS values were not suitable to measure the aromatic stabilization of DA transition states as they may be contaminated by other factors, such as σ -electron density, non-planarity, variable bonding, etc. [37]. Many other alternative descriptors have been developed and used for that purpose. For instance, the paradelocalization index (PDI) [50] was used by Mojaheri et al to measure the aromaticity of 44 DA transition states. Their study revealed that in normal and neutral DA reactions, there is a gain in aromatic stabilization of the TS which contributes slightly to reduce the activation energy[51].

Studying the aromaticity of DA TS is still an open and hot question probably because of the versatile nature of the aromaticity itself, which is computationally estimated using several means, based on structural[47], energetic[52], magnetic[53] criteria, etc. For more details about aromaticity descriptors for DA transition states, the reader is referred to this review paper[54].

d) Catalysts, solvent, and salts effects on the mechanism of DA reactions

Many factors have been found to influence the rate and stereochemistry of DA reactions. Among these are principally catalysts and solvents. Both experimental[19] and theoretical[55] studies have shown that Lewis acids, such as atomic cations, accelerate the rate of DA reactions. The mechanism behind their catalytic effect is the enhancement of the interaction between the HOMO_{diene} and LUMO_{dienophile} orbitals by lowering of the LUMO energy of the dienophile[56]. Like atomic cations, positively charged molecular fragments such amidinium ions [57], iminium[58], tetraalkylammonium [59], ammonium and sulfonium salts[60] can also play the role of accelerator in DA reactions [25]. Today, Lewis catalysis in DA reactions is also explained in terms of Pauli-lowering catalysis. This theory is based on the idea that, by increasing the asynchronicity the Diels–Alder reaction, Lewis acids make the occupied π -orbital located on the C=C double bond of the dienophile asymmetric. This leads to an extra reduction of the destabilizing steric Pauli repulsion and the destabilizing activation strain [13].

Besides Lewis acids, many other catalytic techniques have shown promising results in the boosting of DA reactions. These include the utilization of electric fields (electrostatic catalysis)[20], biological materials (antibody catalysis)[61] or polymers[62]. Particularly, some enzymes (‘Diels-Alderase’) seem to play a crucial role in the feasibility (and stereospecificity) of DA reactions in biological mediums[63].

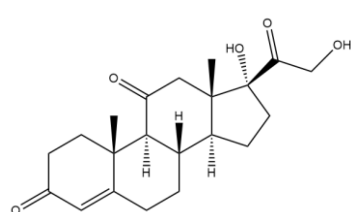
Furthermore, the effect of solvents and salts on DA reaction are also well documented. Cativiela et al experimentally studied the effect of fluorinated alcohols on the rate, *endo/exo*, regio- and diastereo-facial selectivities of several Diels–Alder reactions was [64]. The latter observed that the solvophobicity power (Sp) was the main factor influencing the reaction rate, although in some reactions hydrogen bond donating (HBD) ability may also play a role. Similarly, the use of ionic liquids has been proven to enhance the *endo* selectivity in the DA reaction between methyl acrylate and cyclopentadiene at room temperature[65][66]. A green chemistry aspect to highlight here is that water seems to provide a greater acceleration of DA reactions than ionic liquids at room temperature[67].

In computational chemistry, two main families of models are used to account for solvent effects, namely implicit and explicit models. Implicit approaches consist in embedding the solute in a cage surrounded by a continuum medium characterized by its relative permittivity. In explicit models, the simulated system is composed of the solute and a large number of solvent molecules. In terms of computing cost, implicit models are more affordable than their explicit analogs; though, explicit solvation gives rise to more reliable and insightful results than implicit models[68]. Coming back to DA reactions, semi-empirical AM1 and PM3 methods have been used to analyze solvent effects on the Diels-Alder reactions of cyclopentadiene and 2-methyl-1,3-butadiene (isoprene) with acrylonitrile, acrolein, methyl vinyl ketone, and methyl acrylate in the framework of the Self-Consistent Reaction Field (SCRF)[69] and supermolecule[70] models. Both SCRF and supermolecule models described correctly the effect of the solvent on *endo/exo* selectivities[71]. In the same track, Chandrasekhar conducted a QM/MM simulation in water for a set of DA reactions of cyclopentadiene (CPD) and found that free energies of activation were reduced on going from the gas phase to water by 1.5, 2.8, and 4.4 kcal/mol for the reactions with acrylonitrile, methyl vinyl ketone, and naphthoquinone, respectively[72]. More interestingly, Thomas et al proved that a water surface is less effective than bulk water for catalysis of DA reactions of CPD with 1,4-naphthoquinone, methyl vinyl ketone, and acrylonitrile[73].

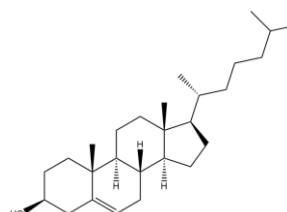
2.2. Applications

The DA reaction has found many applications in various domains, especially in the total synthesis of natural products [7], in polymer chemistry[8] as well as in the industrial preparation of agrochemicals[9].

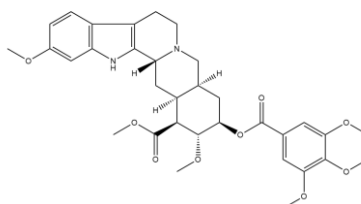
The first application of DA reactions in natural products chemistry was the synthesis of the steroids cortisone (Figure 1) and cholesterol by Woodward et al [74]. Since then, the DA reaction has been involved in many synthetic schemes. For instance, it was used by Danishefsky *et al* in the preparation of disodium prephanate, a biosynthetic precursor of the amino acids phenylalanine and tyrosine [75]. Corey and his coworkers also utilized a DA reaction in the early steps of the synthesis of prostaglandins F2 α and E2 to establish the relative stereochemistry of the three contiguous stereocenters on the prostaglandin cyclopentadiene core [76]. In the same optic, Wender *et al* reported the usage of a DA reaction in their synthesis of the alkaloid reserpine to set the cis-framework of its D and E rings [77]. Moreover, the total syntheses of vindoline and vindorosine, two alkaloids isolated from the Madagascan periwinkle in 1985, were also initiated by an imino DA reaction[78]. Many other examples can be found in the literature. For more applications and details, the reader is referred to the excellent literature review by Nicolau et al [7].



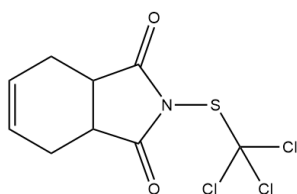
Cortisone



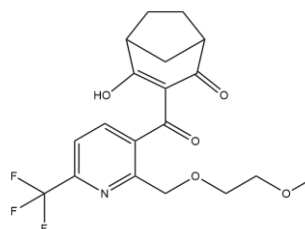
cholesterol



reserpine



Captan



Bicyclopyrone

Figure 1. 2D structures of some steroids (cortisone and cholesterol), alkaloid (reserpine) and agrochemicals (Captan and bicyclopyrone) synthesized using the Diels-Alder reaction.

In addition to their main utilization in the synthesis of biologically active compounds[79] and natural product chemistry [7] (mostly at reduced scale), DA reactions are also utilized in the industrial sector for the preparation of great amounts of chemicals, especially agrochemicals[9] and fragrances[80]. Captan for example (Figure 1) is a plant-growth regulator prepared from 1,3-butadiene in the presence of maleic anhydride followed by the treatment with ammonia and perchlorylmercaptan [9]. Bicyclopyrone [81] is a crop herbicide marketed by Syngenta and whose preparation involves a DA reaction to establish the 3-hydroxy-bicyclo[3.2.1]oct-3-en-2-one fragment. For a detailed review of industrial applications of the DA reactions, see [82] .

2.3. Diels-Alder reaction of the [60] fullerene

In 1985, Kroto and his co-workers reported the accidental preparation of a remarkably stable cluster consisting of 60 C atoms, to which they ascribed a truncated icosahedral geometry [83]. Five years later, the first synthesis of the [60]fullerene in macroscopic quantity was achieved by Krätschmer et al through an experiment using pure graphitic carbon soot as starting material [84]. Since then, a wide variety of chemical transformations have been used to produce a great diversity of fullerene derivatives [85][86]. Among them are cycloaddition reactions, especially the Diels-Alder reactions [87]. Due to the insolubility of fullerenes in most common solvents[88], their functionalization is imperative as it enhances their solubility [89][90]. It is worth noting that functionalized fullerenes have found widespread usage in chemistry [91], medicine [92][93], and material sciences [94][95].

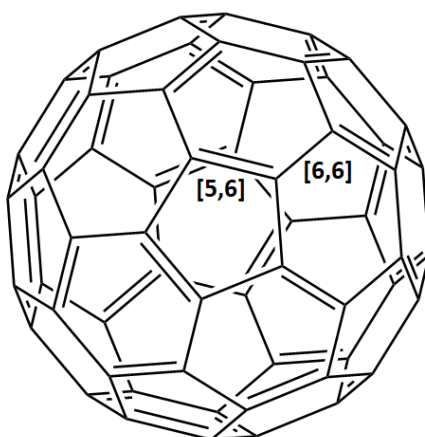


Figure 2. Simplified model of the [60] fullerene

[60] fullerene is the most prominent representative of the fullerene family (Figure 2). All the C atoms in C₆₀ are identical, and this high symmetry (I_h) is reflected by the unique signal at $\delta=143.2$ in its ¹³C NMR spectrum [96]. Despite the claimed aromaticity of the C₆₀ [97], which suggests some chemical stability, this molecule is very reactive. This observation meets the pertinent comment of Evans and Warhurst's saying that “*conjugated molecules show on the one hand enhanced thermochemical stability, while on the other hand they show in some of their reactions greater reactivity than do non-conjugated substances.*”[54].

The C₆₀ fullerene presents two types of C=C bonds, namely [5,6] and [6,6] bonds, whose lengths are 1.458 and 1.401 Å respectively [98]. The [6,6] bond is the most reactive and constitutes the preferred binding site of several dienes[99]. The functionalization of C₆₀ through DA reactions at the [6,6] bond can yield a large number of compounds including mono[100], di-[101], and polyfunctionalized products [102]. Meidine et al showed that C₆₀ reacts with up to six equivalents of cyclopentadiene, and the obtained adduct can be stabilized by hydrogenation of the pendant biclopentene moieties [103]. Krauter et al also obtained a sixfold DA adduct when reacting the buckminsterfullerene C₆₀ with 2,3-dimethyl-1,3-butadiene[104].

Furthermore, DA adducts of C₆₀ are known to be thermally unstable, and this often leads to the opening of the six-membered ring towards the initial C₆₀ and diene upon heating[105]. For instance, Wilson *et al* synthesized the mono-addition product of C₆₀ and Teflon-ponytailed cyclopentadiene, but any attempt to obtain a molecular ion during MALDI-TOF mass spectrometric analysis failed because the product underwent a retro-DA (at a temperature lower than 70°)[106]. Savora et al also observed that the DA monoadduct between anthracene and C₆₀ was unstable above room temperature and dissociated back to the reagents[107].

The mechanism of DA reactions involving C₆₀ fullerenes has been extensively investigated from a computational view. The first study in this sense was reported in 1994 by Sola et al, who employed the AM1 semi-empirical method to investigate the mechanism of the DA reaction between 2,3-dimethylidene-bicyclo [2.2.2]octane and C₆₀[108]. The authors considered the two possible paths, namely the reaction to the [5,6] and [6,6] bonds, and found the latter to be the most preferred route. The same authors also studied the regioselectivity of successive DA cycloaddition of 1,3-butadiene and C₆₀, and found that the enthalpy of activation slightly increased with successive attacks, from 16.2 kcal/mol in the formation of the monoadduct to 17.8 for the hexakisadduct[109]. The choice of the AM1 method in these two studies was probably driven by the need to reduce the computational cost. For the same

reason, Shang et al also used AM1 to investigate the ozonization of the [60]fullerene core, whose major product was predicted to be the C₆₀O [110].

Recently, more accurate studies have been carried out using DFT methods to describe the entire system. For example, Ravinder and Subramanian studied the DA reaction between C₆₀ and [3]dendralene using both B3LYP and B2PLYP-D functionals in conjunction with the 6-31G(d) basis set [111]. Their calculations suggested **1-2** additions to be thermodynamically and kinetically more favored than the **1-1** ones. The authors also noted the significant role played by dispersion effects in the stabilization of the transition states.

The improved reactivity of M@C₆₀⁺ (M=Li⁺, Na⁺, etc) against various dienes has also attracted so much attention. Ueno and his co-workers tried to explain the origin of the enhanced reactivity of Li@C₆₀⁺ towards 1,3-cyclohexadiene as compared to C₆₀ by using the hybrid meta-DFT functional M06-2X with the double zeta split valence 6-31G(d) basis set. Their results suggested that the Li⁺ ion catalyzed the DA reaction by lowering the energy of the LUMO (Lowest Unoccupied Molecular Orbital) of the C₆₀[112]. In the same reasoning, Cui et al showed that Na⁺ could be the best alkali metal cation to catalyze the DA reaction between M@C₆₀⁺ ions (M=Li, Na, K, Rb, and Cs) and cyclopentadiene[113]. The functionalization of C₆₀ with various 3,6-disubstituted tetrazines was investigated by Beheshtian et al at the B3LYP/6-31G(d) level. These authors found reaction energies between -1.26 and - 2.17 eV. Moreover, in contrast with the widely admitted electron-deficient nature of the C₆₀, their calculations indicated that C₆₀ could also behave as an electron-rich agent against some of the considered tetrazines[114].

However, regarding the size of the C₆₀ cage and the computing time required for quantum mechanics ab initio and DFT methods to converge on systems such large, a tradeoff between computational cost and accuracy is needed. This is where multiscale methods such as the hybrid ONIOM approach become very useful[115]. The ONIOM method allows for the combination of high-level (MP2, DFT, etc) and lower-level (UFF force-field, AM1, SVWN/STO-3G, etc) computational methods to describe different portions of the systems.

Osuna et al assessed the performance of the ONIOM method for the study of the chemical reactivity of fullerenes considering the DA reaction between C₆₀ and cyclopentadiene[116]. Four different partitions were tested, consisting of the cyclopentadiene molecule in conjunction with four other fragments (see Figure 2), namely, an ethylene molecule (Partition I, yellow), a naphthalene fragment (Partition II, green), a pyracylene molecule (Partition III, pink), and a

$C_{26}H_{12}$ buckybowl species[117] (Partition IV, violet). Their results suggested Partition III along with the ONIOM2(B3LYP/6-31G(d): SVWN/STO-3G) level to be good enough for getting excellent results in reference with the B3LYP/6-31G(d) level used on the entire system. In another study, the same authors showed that dispersion effects were essential to improve the accuracy of their previous ONIOM2 approach and get results closer to the experimental ones[118]. The ONIOM approach has also been applied to study other large C containing systems, such as nanotubes[119] and graphene sheets[120].

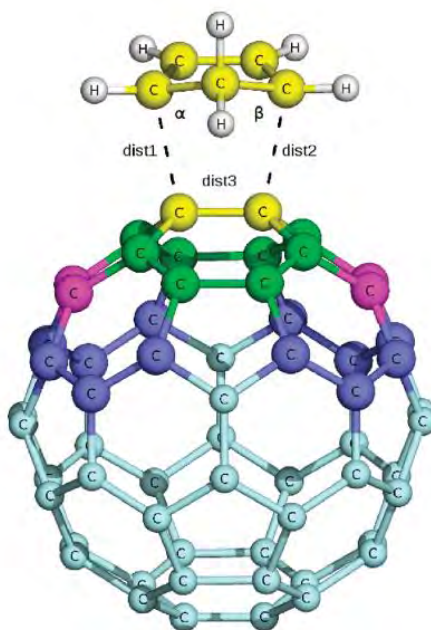


Figure 3. Example of possible partitioning models. Reprinted with permission from [116]. Copyright {2009} American Chemical Society

3. Cheminformatics and computational methods

Laboratory experiments are generally time-consuming and error-prone, to the extent that several trials are often needed before any successful endeavor. To get around this obstacle, molecular modeling has emerged as a complementary tool for the chemist, supplementing the common laboratory vessels and instruments. Molecular modeling has become a well-established research area during the last decade due to advances in computer hardware and software that have brought high-performance computing and graphics within the reach of most academic and industrial laboratories[121]. Molecular modeling allows for the (*in silico*) simulations of chemical processes and properties. It is often used to confirm and rationalize experimental findings[122], predict new materials[123], or properties[124] not yet measured about known substances.

However, to provide the most reliable results, *in silico* simulations require that the systems in the investigation are well represented (described) and that the appropriate tools are used to carry out the task. Two domains closely related molecular modeling, which should not be confused, are cheminformatics and computational chemistry.

Cheminformatics is the use of computer and informational technologies to collect, store, analyze and manipulate chemical data with the aim of getting relevant information from it[125]. Cheminformatics combines biological, chemical, pharmaceutical, and drug patient information to address large-scale data mining, curation, and visualization challenges[126]. On the other hand, computational chemistry is much about the implementation and application of methods that describe molecular systems and investigate their properties [127].

3.1. Molecular representation: SMILES strings

Before any simulation on a given molecular system, the latter must be represented. Chemical structures can first be understood as graphs, i.e., diagrams where vertices or nodes (atoms) are connected by edges (bonds).

Molecular graphs constitute the most primary way of collecting structural information about molecules. These 2D objects can include rings, branches, etc. Some structural features may also be inserted in terms of edges or nodes attributes. Therefore, it becomes possible for example to differentiate similar graphs such as CH₄ and CH₃F by labeling the nodes with atomic numbers or atomic charges. The connectivity between nodes is generally specified using

adjacency matrices [128], which implies that a certain order should be ascribed to different nodes. This order is determined using graph traversal algorithms [129], such as the depth-first, breadth-first, and random search techniques (Figure 1). Depending on the algorithm used by visualization tools to sort out the nodes, the graph returned may look different from tool to tool, while denoting the same molecule.

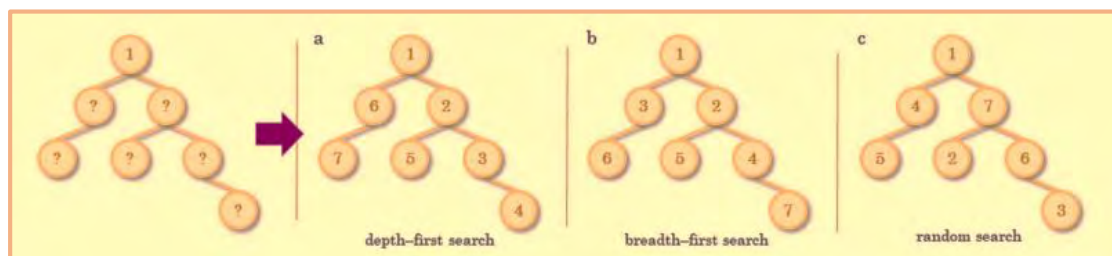


Figure 4. Graph traversal algorithms [130]

Despite some advantages of molecular graphs[131] like the fact that any subgraph is always chemically interpretable, they are not the best choice in most cases. Many types of molecules cannot be described by the graph model, such as structures containing any form of delocalized bonds[132], polycentric bonds (e.g. pure boron clusters)[133], ionic bonds[134], or metal-metal bonds[135]. For example, organometallic compounds (metallocenes, metal carbonyl complexes, etc.) are difficult to describe using molecular graphs because their bonding scheme cannot be explained by valence bond theory[136]. In other words, it would be difficult to describe the bonds using only pairwise relationships between atoms. Moreover, for molecules whose geometry is changing in 3D space, especially if these changes induce the breaking of bonds or the formation of new ones, graph models are meaningless. Finally, graph representations are not easy to search through and not compact at all, especially in terms of storage memory.

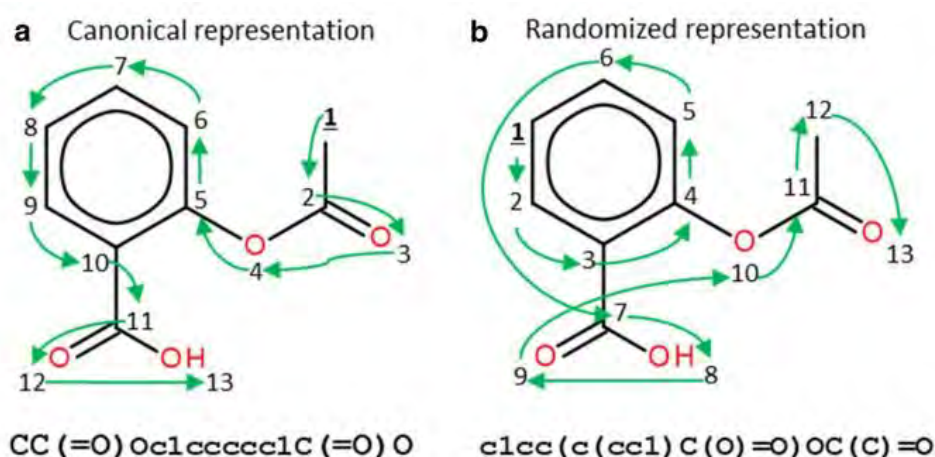


Figure 5. Canonical and randomized SMILES representation of alanine[137]

To overcome the memory problem of graph models and their reduced searchability, linear representations have been introduced and are now widely used in cheminformatics projects to store and manipulate structural data. One of the most common linear notations, used in this thesis to store a huge amount of structural information, is the “Simplified Molecular Input Line Entry System” or SMILES.

SMILES strings were developed in 1988 by David Weininger [138], and have been the most popular line notation ever since. The syntax or encoding rules of the SMILES notation were first incorporated in the Daylight Chemical Information System (<https://www.daylight.com/>), which is still maintaining it. Later on, the rules were implemented in other cheminformatics toolkits such as RDKit[139], openBabel[140], KNIME[141], etc.

The most basic SMILES rules one should know about concern how atoms, bonds, rings, and connectivity are specified. Atoms are denoted by their atomic symbols and placed one after the other to specify the connectivity. The notations -, =, #, and @ are used to define single, double, triple, and aromatic bonds respectively. When an atom is bound to more than two neighbors (branching), parentheses are used to indicate the branched atoms or groups. By default, H atoms are ignored as they may be filled in later to complete the valence of heavy atoms. Single bonds can also be ignored by putting adjacent atoms one next to the other. For example, the strings “C=CC”, “CCC(=O)C” and “C#CC(F)O” can be identified with propene, butan-2-one, and 1-fluoroprop-2-yn-1-ol respectively. In case the structure contains one or more rings, numbers are inserted in the SMILES string to indicate positions where the ring opens and closes. For example, the SMILES strings “CCCC” and “C1CCCC1” represent butane and cyclobutane respectively. Moreover, SMILES strings can also include stereochemical information. As such, the symbols “/” and “\” can be combined to distinguish *cis* and *trans* isomers.

Since SMILES rules are based on molecular graphs, they also suffer from the ordering issue of nodes (atoms). In the RDKit toolkit, for example, a depth-first search algorithm is used to sort atoms[139]. However, multiple atom numberings can be obtained for the same molecule when different initial nodes are chosen to traverse the underpinning node, leading to different SMILES. This situation can sometimes be beneficial as it may serve as a data augmentation means for neural network-based applications [142]. However, the ambiguity it brings is often avoided by generating a unique SMILES string called “canonical SMILES” (Figure 5). Several canonicalization approaches have been proposed at this end, such as the Schneider-Sayle-

Landrum algorithm found in the RDKit tool[143] or Boyle's method [144]. It is worth noting that SMILES line notations inherit some of the drawbacks of graph models, such as the inability to describe organometallic compounds and salts. Therefore, improving SMILES notations (JmolSMILES[145], HumanSMILES[146], deepSMILES[147], etc) is a hot and promising research domain where much is still to be done.

To conclude with this paragraph, we would like to highlight that, apart from SMILES and linear notations, many other molecular representations, and file formats have been developed and are currently utilized. For more details about molecular representations, the reader is referred to the review paper by David et al[130].

3.2. Chemical databases: ZINC

In the beginning, chemical databases (CDB) were probably developed for the sake of keeping track of the molecules synthesized around the world. However, the past 35 years have seen so much progress in the construction and usage of structural databanks. These have gone from being mere repositories of the compounds synthesized to being powerful research tools for discovering new lead compounds [148].

The sources of input structures in CDBs are very diversified, including, but not limited to, retrieval of compounds from catalogs of vendors, conformational searches, and combinatorial approaches [149]. Combinatorial libraries are often generated using various parallel synthetic schemes, where several reagents are mixed so that simultaneous reactions can be triggered and their products identified [150]. These combinatorial libraries can also be generated numerically. An example of such a tool is the DerivatizeMe program, written in our research group and which performs derivatization starting from a given parent molecule. DerivatizeMe has been used to expand the South African Natural Compounds Database (SANCDB)[151].

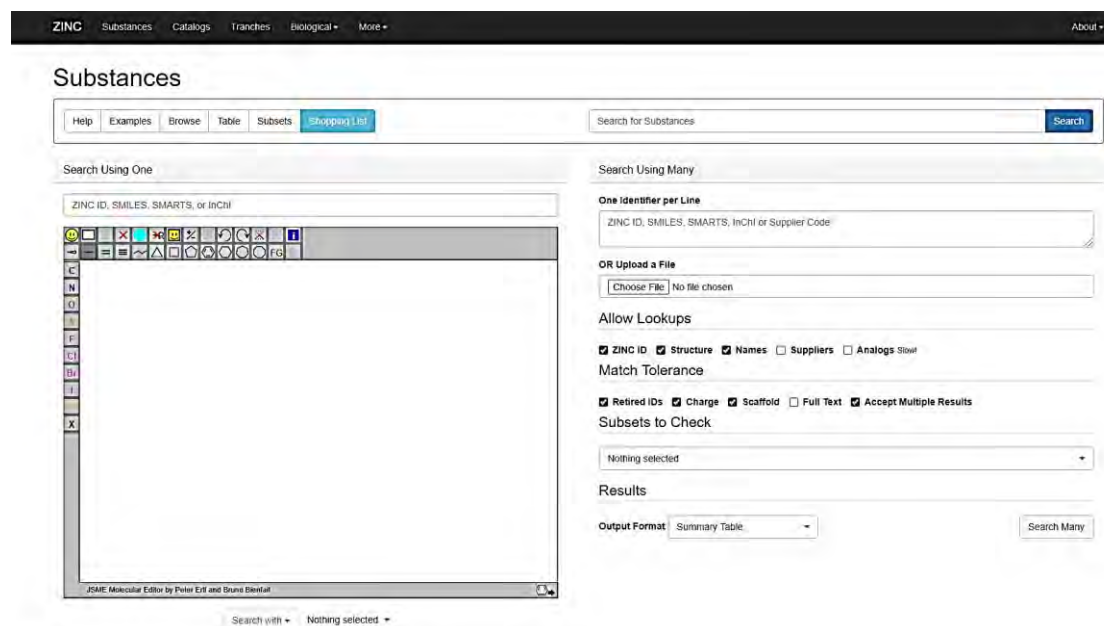


Figure 6. The ZINC website ([ZINC\(docking.org\)](http://ZINC(docking.org)))

Recent years have seen an unprecedented emergence of chemical databases, categorized into commercial (CSD [152]) and open-access ones (ChemSpider[153], LIGAND [154], NCI [155], etc.). On a special tone, the freely accessible ZINC chemical database stands out of the lot for its many interesting features. ZINC [156] is a relational database constructed with MySQL 4.0 [157] and the Perl DBI/DBD toolkit. It contains commercially available compounds, regularly collected from different vendor sources. At its creation, ZINC was populated with 727 824 compounds. Today, this number has reached millions of hits. Collected compounds are filtered before being integrated into the database. All compounds with one of the following features are filtered out: molecular weight greater than 700; calculated LogP greater than 6 and less than -4; the number of hydrogen-bond donors greater than 6; the number of hydrogen-bond acceptors greater than 11; and the number of rotatable bonds greater than 15. They also remove all molecules containing an atom other than H, C, N, O, F, S, P, Cl, Br, or I. Some exceptions are made, for example, to include some actual drugs that violate these constraints.

Molecules that pass the previous barrier must be prepared so that they are compatible with some internal pre-requisites. The preparation protocol includes many steps, such as the generation of isomeric SMILES (from the SDF files provided by the vendors) and their canonicalization, the generation of 3D conformers (using the OpenEye Omega program (<https://www.eyesopen.com/>)), the protonation, tautomerization, and calculation of several properties such as partial atomic charges (using the semi-empirical AMSOL program[158]),

LogP (using a fragment-based approach), etc. After molecules have been prepared, they are then annotated (labeled) and loaded in the database using a Perl script [159].

ZINC screening may be performed based on several criteria: ranges of molecular properties, unique serial code assigned to each compound, vendor name, substructure match, etc. Once screened, the results can be retrieved either by downloading individual molecules or the set of all molecules matched in SMILES, mol2, SDF, etc., or by downloading a table of molecular properties including purchasing information for analysis in a spreadsheet. In the context of this thesis, as will be seen later, compounds containing cyclohexene substructures were retrieved from the ZINC database in the form of SMILES strings. These molecules were used to test the performance of our “AMADAR” program, which is presented in Chapter 2.

3.3. Computational methods of calculation

Computational chemistry is emerging along with theoretical chemistry, with the efforts centered on solving chemistry-related problems by calculation and not on developing new theoretical methods. The past years have been marked by the development of a bewildering number of computational methods, corresponding to different levels of theory. Therefore, one of the greatest challenges in computational chemistry is selecting the suitable level of theory for a given problem and being able to assess the quality of the results.

Behind all computational methods, there is the fundamental question of how the system constituents are described and which equations are used to treat the dynamics of the system. Generally, the system can be described using either classical mechanics or quantum mechanics. This splits computational methods into force field (FF) and quantum mechanics (QM) methods. In their turn, QM methods are very diversified, encompassing wave function methods, semi-empirical methods, and density functional theory (DFT) methods.

In the following lines, we present the must-know of computational methods, stressing the approximation behind each family of techniques, the advantages, and the disadvantages. The mathematical background of each computational method is beyond the scope of this discussion. For further details, the reader is referred to the excellent book of Franck Jensen untitled “Introduction to Computational Chemistry” [160].

3.3.1. Force-field methods

Force-field (FF) methods constitute a family of computational methods that consider atoms as the “building blocks” of molecular systems. In other words, electrons are not considered, so that the bonding information should be provided explicitly. Moreover, the dynamic of the systems is treated classically using Newton’s second law. It is worth mentioning that molecules are described by a “ball and spring” model, where different atoms have different sizes and softness’s, while bonds have different lengths and stiffnesses [160]. FF methods are often referred to as “molecular mechanics methods”.

FF methods are grounded on the observation that different molecules tend to be composed of fragments or units that are structurally similar. For example, all C-H bonds lengths are roughly comprised between 1.06 and 1.10Å, while C-H stretch vibrations range from 2900 to 3300 cm⁻¹. However, a C atom forming only single bonds with its neighbors must behave differently than another with a double or triple bond. FF methods incorporate this aspect by defining “atom types”. Generally, atom types depend on the atomic number and the chemical bonding the atom is involved in. For example, in the MM2 force-field [161], there are 71 atom types (type 41 misses), with type 1 and 2 being sp³ and sp² hybridized C atoms respectively.

As shown in eq.1, the energy of the system in FF methods is split into several terms, each describing the energy needed to change the configuration of the system in a certain way:

$$E_{FF} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} + E_{cross} \quad (1)$$

In this equation, E_{str} is the energy required to stretch a bond between two atoms, E_{bend} accounts for the angle bending and E_{tors} is the torsional energy for rotating around a bond. These three energy functions are often written as Taylor-development series around some typical values. E_{vdw} is the Van der Waals term, which describes the interaction between non-charged atoms due to the uneven distribution of the electron density. The VDW interactions are commonly described using the Lennard-Jones potential [162] (or a modified version such as in Merck Molecular Force Field (MMFF) [163]). E_{el} is the electrostatic energy, which estimates the energy due to the interaction between localized (atomic) charges. The computation of this term requires that atoms are ascribed charges. These can be assigned by empirical rules, but are commonly assigned by fitting to the electrostatic potential obtained using electronic structure methods. Finally, the cross term (E_{cross}) is included in some FF to account for the coupling between bonded interactions. This term is based on the idea that any bond stretching induces some changes in the angles, and vice-versa. It must be noted that E_{str} , E_{bend} and E_{tors} are commonly named “bonded interactions”, while E_{vdw} and E_{el} are the “non-bonded interactions”.

All these terms depend on parameters, which are fitted either to experimental data or high-level computational calculations.

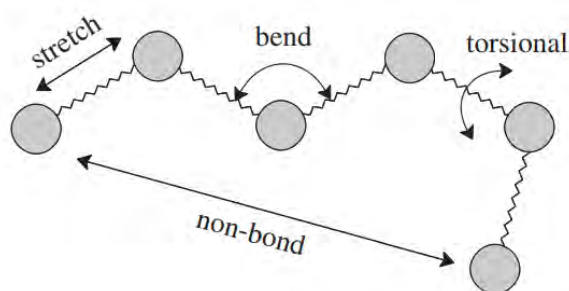


Figure 7. Illustration of Force-field terms [160]

All-atoms force fields (MM3 [164], CHARMM [165], AMBER [166], etc.) are generally used to predict geometries and relative energies. Geometries are known to be much easier to deal with than relative energies. In the past years, many force fields have been developed to describe more specific systems (such as metal complexes [167]) or improve on previous ones [168]. Today, there exist other more exotic families of force fields such as the universal, coarse-grained, and reactive force fields.

The Universal force field (UFF) is assumed to be applicable to almost all atoms of the periodic table, with a reduced number of parameters as compared to the common all-atoms force fields [169]. Unfortunately, this may result in slightly less accurate geometries and relative energies [170]. Coarse-grained force fields aim at increasing the computational efficiency by grouping atoms into units ('beads'). This allows simulations of larger systems and longer times. One of the biggest limitations of CG force fields is their incapacity to describe hydrogen bonds. Two examples of CG force fields are the UNRES [171] and MARTINI protein force fields [172]. Finally, reactive force fields contain no explicit bonding information since energy terms (functions) are defined in terms of interatomic distances. Reactive force fields (such as ReaxFF [173] for hydrocarbons) may be the best choice when bond breaking/formation processes are expected.

The main advantage of FF methods is the speed of calculations, which enables large systems to be treated. Especially when good parameters are available, it is possible to make excellent predictions of geometries and relative energies [160]. However, in general, force field methods provide poor quality results compared to QM methods. Furthermore, FF methods are almost

useless when it comes to understanding the chemical bonding or modeling quantum-related phenomena such as the tunneling effect or Pauli repulsion.

3.3.2. Quantum mechanics methods

In contrast with FF methods, QM methods treat electrons explicitly in the description of the system. Since electrons are very light particles, they can't be described by Newton's second law. Therefore, QM methods try to solve another equation, the Schrodinger equation, whose time-independent form is typed here below:

$$\hat{H}\psi = E\psi \quad (2) \text{ where } \hat{H} = -\frac{1}{2}\nabla^2 + V(R_i, r_j) \text{ in atomic units.}$$

Regarding its analytic form (second-order differential equation), eq.2 can admit an infinity of solutions, which are elements of the Hilbert space [174]. Each of these solutions ψ_i , called “wave function”, describes a quantum state of the system and corresponds to a given energy E_i . The lowest-energy state constitutes the ground state of the system, which is the most likely to observe at equilibrium. The difficulty is that equation 2 cannot be solved (analytically) for systems containing more than 2 electrons. Therefore, only approximate solutions can be deduced numerically. When these solutions are obtained without referring to experimental data, the methods are called “*ab initio*”. Otherwise, they are called “semi-empirical” methods.

The first step in the journey towards solving (numerically) eq.2 is the Born-Oppenheimer (BO) approximation [175]. The latter is based on the consideration that electrons are very light particles compared to nuclei. Keeping this in mind, it is obvious that electrons move so fast and readjust themselves several times before any noticeable change in the positions of nuclei. Therefore, it becomes possible to decouple the nuclei and electrons' motions. The consequence is that eq.2 can now be solved in two steps: first, an electronic solution is obtained by considering nuclei as fixed (their position are then parameters). The resulting potential energy surface (PES) is then used to solve the nuclear equation. It must be noted that the nuclear wave function contains information about vibrational and rotational energy levels.

The BO approximation is usually a good approximation for the most majority of systems and brings only tiny errors. However, when two or more solutions of eq2 for a system become very close energetically, the BO approximation breaks. To assess the error induced by the BO approximation, one can rely on the diagonal Born-Oppenheimer correction (DBOC), which is the second derivative of the electronic wave function with respect to the nuclear coordinates. Some methods have been developed that do not consider the BO approximation. This is the

case of the NOMO (Nuclear Orbital plus Molecular Orbital) method [176], which should give rise to an energy spectrum containing both nuclear (e.g. vibrations) and electronic states.

Despite the BO approximation, the Schrodinger equation for multi-electronic systems still cannot be solved analytically due to the inseparability of electron-electron interactions. The next approximation consists in assuming that the motion of a given electron is independent of the individual dynamics of all the other electrons. This approximation is the basis of the Hartree-Fock (HF) theory [177], which considers each electron as interacting with the average field of the rest.

As each electron is now independent of the others (HF theory), the Schrodinger equation can be solved for each electron and the resulting solution is a monoelectronic molecular orbital ϕ_i . Grossly speaking, the monoelectronic energy operator is called the “Fock” operator (in reality F is associated with the variation of the total energy, not the energy operator itself), in which the electron-electron interaction is accounted for by the “Coulomb” and “exchange” terms, noted J and K respectively. The Coulomb integral can be identified with the classical repulsion between two charge distributions described by ϕ_1^2 and ϕ_2^2 . The exchange integral has no classical counterpart and may be understood as the energy due to the permutation of two electrons with the same spin.

In line with the Hartree approximation [178], the total wave function should be expressed as the product of the monoelectronic molecular orbitals, i.e. $\psi = \phi_1 \phi_2 \dots \phi_N$, with N the number of electrons in the system. However, since electrons are indistinguishable fermionic particles (spin = $\frac{1}{2}$), ψ must be antisymmetric with respect to the permutation of any two electrons. Therefore, the monoelectronic solutions are always arranged into a Slater determinant (SD) [179] to ensure the antisymmetry constraint is fulfilled.

In addition, each ϕ_i is transformed into a spin-orbital (product of a spatial and spin function) to account for the spin of each electron. Finally, the best total wave function is obtained by applying the variational principle to the SD taken as a trial function and containing parameters that have to be readjusted by minimizing the energy of the system. The resulting system of equations is called HF equations which are solved iteratively (Self-Consistent Field procedure).

a) Basis sets (functions)

The usage of ab initio QM methods implies a supplementary approximation, consisting in using a set of known functions, named basis functions, to approximate molecular orbitals (MOs).

Normally, MOs should be written as an infinite expansion of such functions (complete basis set). Unfortunately, this is impossible in actual calculations.

There are two types of basis functions (basis set): Slater-type Orbitals (STO) and Gaussian-type Orbitals [180]. Owing to their analytical expressions, STOs ($\sim e^{-\zeta r}$) are known to behave better than GTOs ($\sim e^{-\zeta r^2}$) in the proximity of the nucleus where the latter present a zero slope [181]. However, GTOs are computationally less costly than STOs. The common compromise is then to combine several GTOs to reproduce the accuracy of STOs, and save computational power. In practice, a basis function is composed of several GTOs called primitives, which are combined or contracted using contraction coefficients. Therefore, there are two types of parameters to be determined (fitted): the ζ (Zeta) exponent and the contraction coefficients.

Different families of basis sets have been proposed to ameliorate the description of MOs. The minimum (Single Zeta) basis set [182] is the simplest, involving a unique basis function to describe each occupied atomic orbital. However, generally, more basis functions are used to represent occupied AOs, resulting in Double, Triple, Quadruple, etc, Zeta basis set. When the duplication, triplication, etc, is restricted only on valence electrons, we get a series of basis sets called split-valence basis sets [183]. This is the case of the Pople's basis sets like 6-31G which means that one contracted GTO (formed of 6 primitives GTOs) will be used to describe the core-shell(s), while each valence orbital is described as two GTOs, one formed of 3 primitives and a separate (primitive) GTO [184]. Note that the 3 primitives are contracted, and their contracted GTO is combined with another GTO using a fitted coefficient. Most often, other supplementary functions are added to the primary ones, such as polarization and diffuse functions.

Polarization functions are denoted by an asterisk in Pople's notation. Two asterisks mean that these functions will be added to the light (H and He) and heavy (C, N, O, etc.) atoms. For instance, 6-31G** is equivalent to 6-31G(d,p) and means that d-type functions must be added to heavy atoms and p-like ones to H atoms. Polarization functions add some flexibility in the system, allowing for a better representation of the asymmetry that is caused by chemical bonds[160]. On the other hand, diffuse functions are denoted by "+" in Pople's convention. Two "+" signs imply that they have to be added on both light and heavy atoms. These functions catch the "tail" portion of atomic orbitals. They are very crucial for the description of systems involving anions.

b) HF derived semi-empirical methods

The HF model must be considered as a branching point, from which other methods can be derived. For example, standard semi-empirical methods are derived from the HF method by neglecting, in the Fock matrix, all products of basis functions depending on the same electron coordinates, but centered on different atoms. This simplification is known as the Zero Differential Overlap approximation (ZDO) [185]. Applying the ZDO approximation on the Fock matrix implies that only one- and two-center integrals will remain. From there, further approximations are made, including the NDDO (Neglect of Diatomic Differential Overlap) [186], INDO (Intermediate Neglect of Differential Overlap), and CNDO (Complete Neglect of Differential Overlap) approximations [187]. The main difference between NDDO, INDO, and CNDO lies in the treatment of two-electron integrals: CNDO and INDO cut down these integrals to two parameters, while NDDO retains all of them [160].

To remediate the deficiencies due to these approximations, some parameters are inserted where integrals were removed. The resulting modified methods correspond to well-known semi-empirical methods. For example, the MINDO/1, MINDO/2, and MINDO/3 are deduced from the INDO approximation by introducing parameters that depend on diatomic variables. The two-atoms parameters used to establish the MINDO/3 are based on data of 10 atoms [188]. Three other semi-empirical methods are derived from the parametrization of the NDDO model, namely MNDO, AM1 (Austin Model 1) [189], and PM3 [190]. In opposition to the previous MINDOs methods, NDDO-derived methods consider atomic variables. AM1 was developed to correct a systematic error in the MNDO method, i.e. the misdescription of interatomic repulsion over 2-3Å. Similarly, PM3 is an improvement over AM1. Finally, in the aim to improve over the PM3 method, J.J. Steward developed PM6 and PM7 using new parametrization (covering around 70 elements) and changing the atomic variables by diatomic ones as in the MINDO schemes [191].

Normally, the accuracy of the aforementioned NDDO-derived SE methods in predicting quantities like heats of formation is expected to follow the trend $AM1 < PM3 < PM6 < PM7$. However, this is not always the case when other questions are addressed. In between the trend, other SE methods have been proposed, such as the PDDG/PM3 (Pairwise Distance Directed Gaussian PM3) method of Repasky et al, which was found to yield more accurate heats of formation (over AM1 and PM3) for 622 molecules containing C, H, N, and O [192].

Like FF methods, classical semi-empirical methods perform fairly well with systems for which there is enough experimental information [160]. Their main advantage over common FF methods is that they can describe bond breaking/formation processes. Since SE methods use a minimal basis set, they are not suitable for computing some properties such as the polarizability of molecules, which requires the use of diffuse functions on atoms. Finally, as SE methods are zero-dimensional, the only way to assess their accuracy is by comparing the results of different methods with experimental data.

c) Improving the HF method: Post-HF methods

HF method is known not to account for the (dynamic) correlation energy between electrons (a consequence of averaging the electron-electron interaction). More advanced methods try to include electron correlation either by introducing multideterminantal wave functions [193] or by considering that the real system is a deviation from the HF model so that the Hamiltonian can be written as the sum of the two terms: one denoting an ideal system where electron-electron interactions are described in an average way (HF model) and a perturbation term to correct the deviation from the real system. The latter family of post-HF approaches includes methods such as the Möller-Plesset Perturbation theory [194] (e.g. MP2 and MP4).

It is worth noting that post-HF methods can generate results that systematically tend towards the exact solution of the Schrodinger equation. Unfortunately, this comes with a cost in terms of computing time. As such, most of these methods are not suitable for systems with more than 20 atoms (unless one disposes of enough computing resources or power) [160].

d) Density functional theory

The main disadvantage of wave function methods is the computational cost, which increases drastically with the size of the system. To get around this limitation, Density Functional Theory (DFT) has been developed based on the Hohenberg-Kohn theorem [195], which states that the electronic energy of the ground state is determined totally by the electron density of the system. In DFT methods, the energy of the system is expressed as a function of the electron density, which in turn is the square of the wave function integrated over $N - 1$ electron. Therefore, the energy of the system appears to be a function of another function, i.e. a “functional” of the electron density.

The challenge in DFT methods is to find the best functional that connects the electron density to the energy of the system. Normally, the energy functional can be divided into three terms $T[\rho]$, $E_{\text{ne}}[\rho]$, and $E_{\text{ee}}[\rho]$, which denotes the kinetic energy, electron-nucleus attraction, and electron-electron repulsion respectively. Like in the HF method, the last term may be split into a Coulomb and an exchange term noted $J[\rho]$ and $K[\rho]$.

While $E_{\text{ne}}[\rho]$ and $J[\rho]$ have the same expressions as in the HF method, the problem is now how to obtain $T[\rho]$ and $K[\rho]$. Early studies proposed to consider a uniform gas model to derive directly their expressions without any reference to orbitals. Unfortunately, these orbital-free models (such as the Thomas-Fermi-Dirac (TFD) model [196]) did not perform well since they could not predict chemical bonding and subsequently no molecular structure. Despite further attempts to improve $J[\rho]$ and $K[\rho]$ by including higher-order derivatives of the electron density, the accuracy of these methods could not go beyond that of the HF method.

The main flaw in orbital-free models is the poor description of the kinetic energy. Modern DFT models are based on the Kohn-Sham formalism [197], which suggests that kinetic energy $T[\rho]$ can be derived from a set of auxiliary orbitals used to represent the electron density. Unfortunately, this brilliant trick has a computational price to be paid. Because, as orbitals are re-introduced, the complexity jumps from 3 to $3N$ variables.

The KS theory relies on the assumption that the kinetic energy can be split into two terms, one which can be calculated exactly ($T_{\text{s}}[\rho]$) and a smaller correction term known as exchange-correlation functional ($E_{\text{xc}}[\rho]$). The first term is computed by considering no interaction between electrons (such as in the HF model), while E_{xc} is now the only unknown. Therefore, different DFT methods are obtained as an attempt to find the E_{xc} term, which is often considered as the sum of the exchange E_{x} and correlation E_{c} terms.

Several approximations in the context of the KS DFT theory have been proposed. The simplest one is the local density approximation (LDA) [160], which assumes a uniform electron gas as in the electron-free TFD method. For open-shell systems, the LDA approximation is replaced by LSDA (S for spin) to account for the different α and β spin densities. A typical example of the LDA method is the X_{α} method [198] proposed by Slater in 1951, where the correlation energy is completely neglected so that $E_{\text{xc}} = E_{\text{x}}$. Two correlation functionals that can be used with Slater's exchange energy are the Vosko-Wilk-Nusair (VWN) [199] and Perdew-Wang (PW) [200] interpolation formula. The L(S)DA methods are known to provide results with

accuracy similar to those obtained by the wave function HF method, but with increased computational efficiency.

To improve over the LSDA approximation, the key idea is to derive new exchange-correlation functionals based on the assumption of a non-uniform electron gas. The Generalized Gradient Approximation (GGA) for example consists in introducing the first derivative of the electron density as a variable in the E_{xc} functional. Many exchange and correlation functionals have been derived in this way. For example, Becke's exchange energy proposed in 1988 (noted B or B88) [201] inserts a correction term in the LDA exchange energy, depending on $\Delta\rho$ and a β parameter (noted β) that is fitted on known data. A famous correlation functional in the GGA approximation is the Lee, Yang, and Parr functional (LYP) [202]. The LYP correlation functional is often combined with Becke's exchange term to give the BLYP exchange-correlation functional. Similarly, Perdew and his team developed several GGA exchange-correlation functionals, such as the PBE (Perdew-Burke-Ernzerhof) functional [203], which depends on three parameters (a, b and c) fitted on the Helium atomic data.

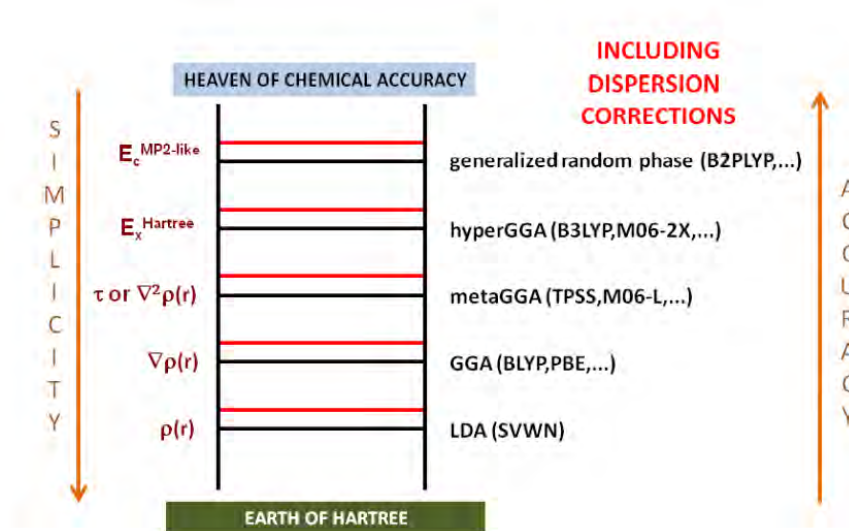


Figure 8. Jacob's ladder of DFT methods improvement [204]

More accurate approximations (functionals) beyond the GGA model have been developed. These include *meta-GGA methods* {such as the Tao-Perdew-Staroverov-Scuseria (TPSS) [205] and the Perdew-Kurth-Zupan-Blaha (PKZB) [206] exchange-correlation functionals} which include higher-order derivatives of ρ , *hybrid GGA-methods* {the famous B3LYP [207] exchange-correlation functional} including a portion of the HF exchange energy in the expression of $E_x[\rho]$, *dispersion-corrected methods* {such as the B3LYP-D3 [208] (with the Grimme's dispersion energy) and the ω B97XD [209] exchange-correlation functionals},

range-separated methods {such as the CAM-B3LYP [209] functional} which define an exchange switching whose expression switches between short- and long-range interactions (using, for example, an E_{xc} functional for short-range and the (exact) HF expression for long-range, and *double-hybrid methods* {such as B2PLYP [210] built on the BLYP functional} containing parametrized HF and MP2-like energy terms.

While new DFT functionals are still being developed, the main inconvenience of DFT methods is the impossibility to systematically improve the quality of results such as in wave function methods (MP4 is better than MP3, MP2, and MP1=HF). In addition, DFT methods are known to poorly describe energy gaps between states with different spins [160]. DFT methods are also accused of being prone to favoring electron delocalization in some cases. This is the case of radical cations where DFT methods are unable to localize the spin and charge, leading to unrealistic energy surfaces.

3.4. Hybrid methods: the ONIOM approach

Despite the recent advances in computational chemistry, which have been accelerated by the advent of powerful computers, it is still difficult and costly to accurately compute the properties of large and complex systems using QM methods. Colossal efforts have been invested in trying to come over this problem. Of the most promising solutions is the usage of hybrid methods, which allow for the application of different methods (from high-level accurate methods to FF ones) to different fragments on a system. The first ideas of combining different methods were reported by Honig and Karplus in 1971[211], who combined a Huckel model (to describe π electrons) and a force field for non-bonded Van der waals interactions to characterize the torsional potential energy surface of the retinal molecule. The methodology was extended, and formalized into the generic “QM/MM” scheme introduced by Warshel and Levitt [212].

In the QM/MM formalism, the system is divided into two parts: the “model” system which is treated with a QM method, and the environment which is described by a force field [213]. Between the two regions, there is a shared area (QM-MM boundary), whose description is the most challenging. The QM/MM scheme is an additive approach in that the total energy of the system is obtained as the sum of those of the QM, MM, and QM-MM parts. For more details on the applications of QM/MM methods, see [214].

$$E_{\text{QM/MM}} = E_{\text{QM}} + E_{\text{MM}} + E_{\text{QM-MM}} \quad (3)$$

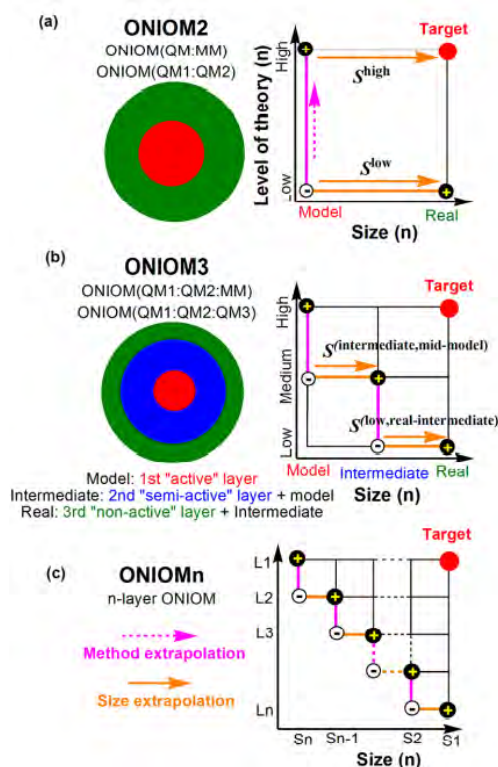


Figure 9. Schematic partitioning of the system using the ONIOM method [115].

Another hybrid method is the ONIOM approach. The first generation of ONIOM methods developed by Morokuma and coworkers [215] can be understood as the combination of two alternative schemes, namely the IMOMM (Integrated Molecular Orbital (MO) + Molecular Mechanics) [216] and IMOMO (Integrated Molecular Orbital + Molecular Orbital) techniques [217]. The IMOMM and IMOMO are based on partitioning the system into two layers (low and high) described either by QM and MM methods, or two QM methods of different accuracy. The total energy of the system in these approaches is calculated using the eq4, where high is the most accurate level and low is the least one.

$$E_{\text{IMOMM/IMOMO}} \approx E_{\text{ONIOM2}(\text{high}, \text{low})} = E_{\text{high}, \text{model}} + E_{\text{low}, \text{real}} - E_{\text{low}, \text{model}} \quad (4)$$

The error made by using the ONIOM2 method can be assessed by computing the difference between $E_{\text{ONIOM2}(\text{high}, \text{low})}$ and $E_{\text{high}, \text{real}}$. For example, for an INIOM2(B3LYP/6-31G(d): AMBER) model, the error could be calculated as the difference between the energy of the model and that obtained by applying the B3LYP/6-31G(d) on the entire system. It is recommended to evaluate this error to assess the reliability of a chosen ONIOM model [218].

A particular feature of the ONIOM method is that it is not limited to two layers. For example, the first three-layer ONIOM model (ONIOM3) was developed by Svensson et al. by combining

the extrapolative two-layer IMOMM and IMOMO schemes into an integrated approach using two QM methods and a force field method [6]. In their ONIOM3 model, the system was partitioned into the model, intermediate, and environment portions (layers), described by the high, medium, and lower levels of theory (eq5).

$$E_{\text{ONIOM3}(\text{high: medium: low})} = E_{\text{high, model}} + E_{\text{medium, intermediate}} - E_{\text{medium, model}} + E_{\text{low, real}} - E_{\text{low, intermediate}} \quad (5)$$

Similarly, the quality of an ONIOM3 model can be assessed by evaluating the energy difference between $E_{\text{ONIOM3}(\text{high: medium: low})}$ and $E_{\text{high, real}}$. The ONIOM method can be extended to higher-order models [219], but, in practice, the significance of using more than three layers may need to be examined carefully for any particular situation [115].

3.5. Transition state prediction

Computational chemistry has afforded the simulation of transient physical states or configurations such as reactions transition states. Topologically speaking, TSs are saddle points on the potential energy surface (PES) of chemical reactions [5]. They correspond to the highest energy point of the (lowest energy) path connecting the reactant(s) and product(s) [220]. The prediction of TSs is one of the most challenging tasks in computational chemistry. However, there exist two main families of methods (algorithms) used in the attempts to locate these saddle points: surface walking (SW) and interpolation algorithms [160].

Surface walking (SW) algorithms navigate on the PES, guided by the gradient and Hessian information [221]. By following the Hessian eigenvalues, SW algorithms can maximize the energy in one degree of freedom and minimize it in all the others [222]. These techniques only need the reactant structure, from which they walk throughout the PES until they reach the (assumed) product [223]. Unfortunately, it has been found that, when used alone, SW algorithms may perform poorly, especially for bimolecular reactions.

On the other hand, interpolation algorithms need both the reactant and product structures. Then, they construct a sequence of configurations (nodes) between the two extrema, before carrying out optimization calculations on the different structures [224]. In practice, an interpolation algorithm may be used to obtain a good estimate for the saddle point (guess) [225], which will then be refined using a surface walking technique such as the Newton-Raphson algorithm [5]. Interpolation algorithms are often called “double-ended” methods, in opposition to the single-ended SW algorithms. Many interpolation algorithms have been developed encompassing, but

not limited to, the nudged-elastic band (NEB) [226], the string method (SM) [227], and the growing string method (GSM) [228]. The discussion of these techniques will not be covered here.

4. Reaction force analysis (RFA)

4.1. Standard RF theory

The conversion of reactants into products during chemical reactions induces equivalent changes in the overall energy of the system [229]. As the reaction goes on, atoms are displaced towards or far from each other, and their motions are associated with a force acting on the whole system to bring it from the initial to the final state. Torro-Labbé and coworkers showed that this force F (reaction force) could be defined as the opposite of the system's potential energy $E(\xi)$ with respect to the reaction coordinate ξ [230]. Most often, the reaction path ξ is chosen along the Intrinsic Reaction Coordinate (IRC) path, which is the mass-weighted steepest-descent path connecting the reactants to the products through the TS [231].

$$F(\xi) = - \frac{dE(\xi)}{d\xi} \quad (6)$$

For a chemical process presenting a unique energy barrier, the profile of F splits the IRC path into three regions delimited by its global minimum α and maximum γ located at both sides of the TS [232]. The region between these extrema is known as the TS region and lies between the preparation phase, on its left, and the relaxation one situated on its right [233]. Previous studies have shown that the most important electronic changes take place in the TS region, while the other two regions are dominated by structural changes [234]. The fragmentation of the IRC path suggested by the reaction force allows for the partitioning of the activation energy in two terms, $E_{act,1}$ and $E_{act,2}$, defined in eqs 7 and 8. The first term accounts for the energy needed by the reacting system to overcome the structural resistance in the preparation phase, while the second term evaluates the energy consumed from α to the TS.

$$E_{act,1} = - \int_R^\alpha F(\xi) d\xi \quad (7); \quad E_{act,2} = - \int_\alpha^{TS} F(\xi) d\xi \quad (8)$$

In the same framework, Yepes et al introduced the second derivative of $E(\xi)$, known as the reaction force constant, and showed that it carried insightful information about the mechanism of chemical reactions [235].

$$\kappa(\xi) = \frac{d^2 E(\xi)}{d\xi^2} \quad (9)$$

The reaction force theory is one of the most powerful reactivity paradigms developed during the last two decades. This theory has been successfully applied to several types of reactions, such as the deamination of the cytosine molecule[236] or intramolecular proton transfer reactions[237]. In contrast with the transition state theory [238], which considers a fixed position on the IRC path as the (classical) transition state, the reaction force theory defines a whole region including the classical TS, where the most electronic changes of the reaction take place. This TS region concept is consistent with the TS spectroscopy of Zewai and Polanyi[239], which describes the TS region as a continuum of transient and unstable intermediate configurations of the reacting system.

4.2. Atomic resolution of energy derivatives

Although much can be learned from the profile of E , F , and κ as defined in Eqs 6 and 9 [240], the chemistry behind a reaction happens between the interacting atoms. Therefore, resolving the reaction force F and reaction force constants K in terms of atomic contributions should provide interesting information about the implication of atoms in various reaction stages.

Recently, Jędrzejewski et al [40] reported the mathematical basis for the atomic resolution of energy derivatives along the IRC path. These authors showed that the reaction force F and force constant κ can be decomposed into atomic contributions by introducing the Hellman-Feynman[241] (HF) forces acting on each nucleus in the standard definition of F and κ (eqs.10-11). Eq.10 shows that atomic contributions $F_A(\xi)$ are obtained as the scalar product of the HF force vector F_A and the deviation vector $\frac{\partial R_A}{\partial \xi}$, which denotes the change in the position of a given atom during the reaction.

$$F_\xi = -\frac{dE}{d\xi} = -\sum_{A \in M} \frac{\partial E}{\partial R_A} \frac{\partial R_A}{\partial \xi} = \sum_{A \in M} F_A \frac{\partial R_A}{\partial \xi} = \sum_A F_A(\xi) \quad (10)$$

$$\kappa_\xi = -\frac{dF_\xi}{d\xi} = -\sum_{A \in M} \frac{d}{d\xi} \left[F_A \frac{\partial R_A}{\partial \xi} \right] = \sum_A \kappa_A(\xi) \quad (11)$$

where the sums run over all the atoms in the molecule.

The foundations of this method rest on the work by Ordon and Komorowski[40] who

investigated the energy derivatives within the conceptual DFT framework [242] and the vibrational softening of molecules [243].

In the present thesis, we have implemented equations 6,9,10, and 11 in the AMADAR program (Chapter 2) for automating the analysis of reaction (IRC) paths. The code has been applied to a dataset of 150 IRC paths of DA reactions, and the results obtained are presented and discussed in Chapter 3.

“If you can’t do great things, do small things in a great way”

Napoleon Hill

Chapter 2: The AMADAR program

0. Summary

Predicting transition state geometries is one of the most challenging tasks in computational chemistry, which often requires expert-based knowledge and constant human intervention. In this chapter, we present technical details and preliminary results of a python-based tool (AMADAR) designed to generate (an unlimited number of) Diels-Alder (DA) transition state geometries (TS) and analyze IRC paths in a (quasi-) automated fashion. The TSs are predicted sequentially, from a constrained optimization in internal coordinates to the refinement at a quantum mechanical level using the single-ended Berny algorithm. Two special modules of the package are devoted to executing the numerical calculations concerning the reaction force analysis and the atomic resolution of energy derivatives. The AMADAR package does not need any training step as it is the case in deep learning-based approaches and uses only SMILES strings of the cycloadducts as inputs. Two key features of AMADAR are its high customizability and the ability to handle particular cases such as intramolecular reactions and situations resulting in competing paths. The performance of the protocol has been assessed using a dataset of 2000 likely DA cycloadducts retrieved from the ZINC database. The sequential location of the corresponding TSs was achieved with a success rate of 95%. The source code of the AMADAR tool is available on GitHub ([CMCDD/AMADAR \(github.com\)](https://github.com/CMCDD/AMADAR)) and can be used at the price of some minor customizations, mostly regarding the local working environment of the user.

1. Introduction

Many theories have been developed in the past to rationalize and predict the speed of chemical reactions. One of the simplest approaches is the Arrhenius relationship that connects the reaction rate to the reaction activation energy through a proportionality constant[244]. While his equation looks very handy, one should be very cautious before using it, especially when quantitatively accurate results are expected and when the study covers a long range of temperature since the activation energy may not remain constant [245]. In place of the Arrhenius equation which only relies on the classical molecular collision theory, a more accepted framework is the transition state theory (TST), combining classical, quantum, and statistical mechanics to derive an equation serving to predict reaction rates. In reality, there are many variants of the TST (conventional TST, variational TST, the RRKM theory, etc.) reposing on the same fundamental assumptions with some supplementary approximations [246].

In the framework of the TST, reactions are assumed to proceed through very unstable states called transition states (TSs) [5]. The relevance of the TS lies in the fact it is directly connected to the activation energy and subsequently the rate of chemical processes. The main challenge in computational studies based on the TST consists in predicting the TS because a good guess is always required to increase the chances of locating the saddle point. For reactions such as the Diels-Alder reaction, chemical intuition can be applied to build manually a good guess that will be refined using single-ended algorithms. On the other hand, a double-ended approach can also be considered by providing an interpolation algorithm with the geometries of the reactants and products. Most often, an interpolation algorithm is used to obtain a good estimate for the TS (guess), which will then be refined using a single-ended algorithm [225]. However, nothing guarantees whether the TS will be located or not. Especially in the former case, the problem generally worsens when we must predict several TSs because this implies that we should construct manually a great number of guesses, which is logically very time-consuming and error-prone.

Recently, deep learning (DL) approaches emerged as a potential solution to this problem. However, DL models require that huge amounts of good quality data are available for training before any prediction can be made [247]. Moreover, some DL pipelines must be fed with 3D geometries of both the reactants and products[248], which comes with a cost if their structures have to be first optimized.

In order to contribute to the challenge of predicting TS geometries and to the understanding of the mechanism of the DA reaction, we have built the AMADAR program (Automated workflow for Mechanistic Analysis of Diels-Alder Reactions). In comparison with DL approaches, the AMADAR tool does not require any training steps and uses only SMILES strings of the cycloadducts as inputs. It is designed to generate an unlimited number of Diels-Alder (DA) TS geometries, before carrying out subsequent analyses based on the intrinsic reaction coordinate (IRC) paths. Two of the key features of AMADAR are its ability to handle particular cases such as intramolecular reactions and situations resulting in competing paths; as well as its high customizability.

The source code of the AMADAR program is provided with appropriate documentation detailing the functioning of the program. The code is written in a user-friendly, efficient way, that should allow intermediate python programmers with some knowledge in computational chemistry to easily customize it. The AMADAR package is projected to be very beneficial to a large community of researchers working on the DA reaction [249] [250][251].

2. IT requirements

The usage of the AMADAR program on a computing platform (PC, cluster, etc) requires a python RDKit[139] virtual environment to be already installed on it. This is because the initial steps of the algorithm rely on the RDKit toolkit. RDKit is a cheminformatics package developed by Greg Landrum, and offering many facilities such as the generation of conformers, the modification of chemical structures, the computational of several molecular descriptors, etc.

Moreover, the user has to define in the main configuration file (da.ini) all the environment variables giving access to an electronic structure software which will have to perform some calculations in the guidance of the AMADAR code. The Gaussian [252] software has been considered as the default, but potentially several other programs could be used in place (for example, Psi4 [253], GAMESS [254], etc.) as long as they offer the possibility to carry out TS and IRC calculations. However, for this second condition methods that generate input files or analyse output files would have to be written in the context of the program for electronic structure calculations. Work is still in progress on how to integrate the GAMESS electronic structure package as a second default electronic structure package so that the user would then be able to choose between Gaussian and GAMESS.

3. The Algorithm

Before digging deeper into the technical details of the AMADAR program, we would like to give an overview of how the TSs are predicted. Briefly speaking, the location of Diels-Alder TSs is done sequentially. The first step consists in converting the SMILES string of the cycloadduct into a 3D conformer. Then, a constrained optimization in internal coordinates is performed on the latter structure and this results in a two-fragment structure which is here referred to as the pseudo-guess. This pseudo-guess structure is further refined at semi-empirical and quantum mechanics levels giving rise to the guess and TS respectively. Automatic checking is performed at this step to make sure the right stationary point has been located; otherwise, the type of error is reported (in the error.txt file) and the file is saved in the “TS/GTS/TS/ERROR_FILES” folder.

Let us now talk about the AMADAR algorithm. As shown in Figure 10, this encompasses 8 steps divided into three main phases: the preparation of 3D geometries, the electronic structure calculations, and the IRC path analysis. These steps are detailed hereafter.

Step 1: Initialization

The program is launched by running the `__init.py` script from a local RDKit environment. This script uploads the SMILES strings of the cycloadducts saved in a separate file called SMILES.txt file, packed from the top of the file to the bottom. This file should be saved in the same directory as the `__init.py` script; otherwise, its absolute path should be provided at the right place in the initialization script. The SMILES.txt file must not contain any blank at the top of the file. Only one string must be given per line. After uploading the contents of the SMILES.txt file, the `__init.py` script checks if the strings contained in the SMILES.txt file may be Diels-Alder cycloadducts.

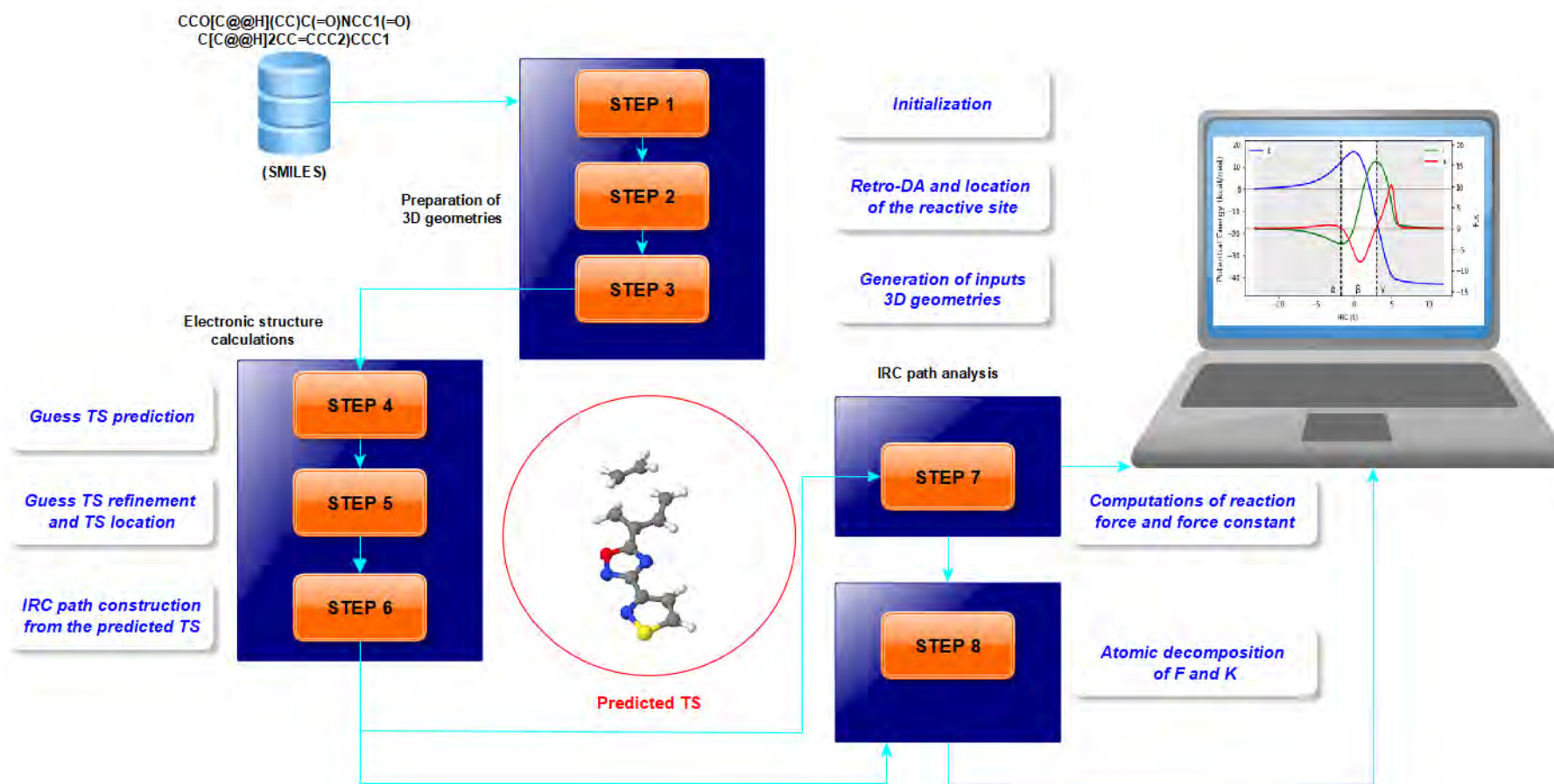


Figure 10. Simplified scheme of the AMADAR algorithm.

To check the validity of DA likely cycloadducts in the SMILES.txt file, their SMILES strings are converted into mol objects, which are then used to seek a cyclohexene substructure in the molecule. In case none of the strings provided in the SMILES.txt file is found to be a DA cycloadduct, or if there is a blank line at the top of this file, the execution of the code is terminated reporting the message:” **SMILES.txt file uploaded, but maybe empty or has a blank line at the top. Please check it. Process killed!!!!**. It is worth noting that the main configuration file (da.ini) contains flags that must be set according to the type of jobs the user would like to carry out. For example, the “*OVERWRITE_FLAG*” flag tells the code to remove all the generated folders during past executions and to create new ones. The next code snippet shows how the SMILES.txt file and SMILES strings are checked (lines 111 to 143 of the __init.py script).

```
smiles_path=os.path.join(cwd,'SMILES_temp.txt')
if os.path.isfile(smiles_path):
    with open(smiles_path,'r') as file:
        content=file.readlines()
    if len(content)!=0 and content[0]!='\n':
        print('SMILES strings uploaded!!!\n')
#####
#      Checking the SMILES strings to make sure they are real DA cycloadducts      #
#####
    cycloadducts_pos=[ ]
    not_cycloadducts_pos=[ ]
    for i in range(len(content)):
        if Chem.MolFromSmiles(content[i]).HasSubstructMatch(cyclo)==True:
            cycloadducts_pos.append(i)
        else:
            not_cycloadducts_pos.append(i)

    if len(cycloadducts_pos)>=1:
        time.sleep(2)
        print(' <<<<',len(cycloadducts_pos),' DA cycloadducts identified in the
              SMILES.txt file')
    path_reagents=os.path.join(cwd,'R')
```

```

path_cycloadd=os.path.join(cwd,'C')
path_pgts=os.path.join(cwd,'TS')
paths=[path_reagents,path_cycloadd, path_pgts]
for path in paths:
    if os.path.isdir(path):
        if OVERWRITE_FLAG==1:
            shutil.rmtree(path)
            os.mkdir(path)
        else:
            os.mkdir(path)

```

Step 2: Retro-DA and location of the reactive site

The SMILES strings approved in step 1 are converted one by one into *mol* objects. Then, each mol object is passed to the “Get_Atoms_of_Interest ()” method of the “Geom3D” module to locate and keep track of the reactive site (RS). A list of atomic indices (ListAtomsInt) defining the RS is returned. In this list, four atoms originate from the diene and the two others from the dienophile. In case the cycloadduct had more than one cyclohexene substructure, a 2×n shaped list is returned, with n the number of cyclohexene substructures. This suggests that the given cycloadduct is the product of more than one competing path, each of which being treated separately.

It must be noted that the “Geom3D” module depends on another module called “Retro-DA”, which contains the appropriate functions (methods) intended to locating atoms and bonds of interest. For example, the “process_retro_Diels_Alder ()” method is aimed at breaking the new C-C bonds of the cycloadduct and changing 3 single bonds into double bonds (and 1 double into a single-bonds) in the RS to obtain the structures of the initial reactants. This process is a retro-DA transformation that gives access to the diene and dienophile mol objects and SMILES strings. The following code shows how the reactive site is located.

```

def Get_Atoms_of_Interest(mol):
    system=mol
    paths=locate_bonds_of_interest(system)
    paths_atoms_of_interest=[]
    for path in paths:

```

```

double_bond=path[0][0]
set_of_four_atoms_indices=path[1]
atom_0=double_bond.GetBeginAtom()
atom_1=double_bond.GetEndAtom()
atom_2=system.GetAtomWithIdx(set_of_four_atoms_indices[2])
atom_3=system.GetAtomWithIdx(set_of_four_atoms_indices[3])
atom_4=system.GetAtomWithIdx(set_of_four_atoms_indices[1])
atom_5=system.GetAtomWithIdx(set_of_four_atoms_indices[0])
paths_atoms_of_interest.append([atom_0,atom_1,atom_2,atom_3,atom_4,atom_5])
return paths_atoms_of_interest

```

Step 3: Generation of input 3D geometries

At this step, the 3D geometry of the cycloadducts in Cartesian coordinates is obtained using a sequential procedure. First, the mol object is embedded in a 3D box and the returned conformer is optimized using the UFF force-field.

The choice of the UFF is here justified by both its ability to keep all the bonds intact and the fact that it was parametrized for a large number of elements from the periodic table [255]. In case of failure, the procedure is repeated, this time looking for more conformers (up to 60) and increasing the number of runs (up to 2000). The UFF optimization is carried out by setting the max number of iterations to 500. This value can be changed by the user. Note that there is no restriction about the chirality of C atoms in the cycloadduct so that the search can span a broader space. The selected conformer is used as input in the constrained optimization towards the pseudo-guess of the TS (step4). The code used to embed and optimize the cycloadduct's structure is given here below.

```

def Optimize_struct(mol):
    try:
        mol=Chem.AddHs(mol)
        embed=AllChem.EmbedMolecule(mol, maxAttempts=200)
        Error=False
        if embed==-1:

```

```

embed=AllChem.EmbedMolecule(mol,useRandomCoords=True,
maxAttempts=2000,enforceChirality=False)

if embed==-1:

    confs=rdDistGeom.EmbedMultipleConfs(mol,numConfs=20,
maxAttempts=2000,enforceChirality=False,useRandomCoords=True)

    if (len(confs)<1):

        confs=rdDistGeom.EmbedMultipleConfs(mol,numConfs=40,
maxAttempts=2000,enforceChirality=False,useRandomCoords=True)

        if (len(confs)<1):

            confs=rdDistGeom.EmbedMultipleConfs(mol,numConfs=60,
maxAttempts=2000,enforceChirality=False,useRandomCoords=True)

            if (len(confs)>=1):

                AllChem.UFFOptimizeMolecule(mol,confId=confs[random.randint(0,len
(confs)-1)],maxIters=500)

            else:

                Error=True

        else:

            AllChem.UFFOptimizeMolecule(mol,maxIters=500)

    except BaseException as e:

        print('Failed')
        print(e)

    if Error==True:

        print('Failed')
        print('Unable to embed molecule')

return mol

```

Step 4: Guess-TS prediction

The first 3 steps of the protocol are supposed to allow for the obtaining of a 3D geometry of the cycloadduct from its SMILES string and identifying the reactive site. The aim of this fourth step is to predict a guess for the TS. In comparison with interpolation techniques, which use the structures of both the reactants and products to deduce the most plausible TS guess, our approach does only need the cycloadduct geometry.

The construction of the guess TS is achieved in two steps: first, a constrained optimization in

internal coordinates coerces the cycloadducts to adopt a configuration where these pairs of terminal C atoms are 2.15Å apart. This separation distance can be changed by the user in the da.ini file. The positions of the terminal C atoms of the diene and dienophile moieties are retrieved from the ListAtomsInt (step 2). This optimization returns 16 successive configurations of the system, of which the highest energy structure (a 2 fragment structure) corresponds to the pseudo-guess TS. Then, the latter is isolated, cleaned up at the same level of theory using a single-ended Berny algorithm[252]. This leads to the guess-TS. The PM6 semi-empirical method has been found to perform well at this step (although the user is permitted to choose other methods).

Step 5: Guess TS refinement and TS location

For each system, a new TS calculation is performed to refine the previous guess structure at a user-defined quantum mechanics level of theory. This step is followed by a vibrational check to make sure the predicted stationary point is a real TS. The default level of theory for this task is B3LYP/6-31G(d) but this can be changed by the user. The choice B3LYP/6-31G(d) level was driven by the fact that it is the most used level in computational investigations of DA reactions, and is known to provide good TS geometries [17][34][256].

Step 6: IRC path construction from the predicted TS

Once the TS has been located, the IRC path can be determined. Details about the theory level or the number of IRC points are defined by the user in the da.ini configuration file. A module called *IRC* contains methods that extracts the TS geometry, and build the input file for IRC calculations. We have observed that at least 60 points per IRC direction from the TS are sufficient to obtain a good IRC path (for mid-size systems with heteroatoms) linking the reactants to the cycloadducts through the TS.

After IRC calculations, a separate script named *myIRCAalyzer.py* may be executed to perform reaction force analyses (RFA, step 7) and atomic decompositions of the reaction force and reaction force constant (step 8). For these analyses, only user-specified systems can be considered. Details about the system to analyse or the atoms to involve in the decomposition must be given in a separate configuration file named *analysis.ini*.

Step 7: Computation of reaction force (F) and reaction force constant (κ)

Two important quantities from the reaction force theory are the reaction force F and reaction force constant κ , which are defined using equations 6 and 9 (see chapter 1, paragraph

4.2), where E is the system's energy along the IRC path ξ . Torro-Labé and his coworkers have extensively demonstrated that F and κ can provide insights information about the mechanism of several reactions [257].

In the AMADAR package, there is integrated a module called *RFA* made of several methods that execute all the calculations related to the reaction force analysis and plot the results. F and κ are numerically calculated at each point of the IRC path using the finite difference approach. Technically, the average value of the forward and the backward derivatives at each given point is used as a good estimation of the exact derivative, except for the first and last points.

Step 8: Atomic decomposition of F and κ

As demonstrated by Jędrzejewski et al[40], the reaction force F and force constant κ can be decomposed into atomic contributions by introducing the Hellman-Feynman[241] forces acting on each nucleus in the standard definition of F and κ_ξ (see eqs.10-11, chapter 1, paragraph 4.2).

Table 1. Description of some modules of the AMADAR package

Modules	Description	Dependencies	classes (methods)	Length
IRC	Builds input files for IRC calculations and analyzes outputs files	-	2 (51)	2 127 lines
RFA	Carries out numerical derivations from IRC paths data; and returns 2D plots of E, F, and κ along the IRC path.	-	5 (48)	1 780 lines
TS	Prepares input files for the refinement of guess TS, and analyzes the outputs.	-	1 (36)	1 770 lines
Geom_3D	Converts <i>mol</i> objects into 3D geometries, and prepares input files for electronic structure calculations	Retro_DA	0 (5)	335 lines
Retro_DA	Realizes the retro-DA transformation of the cycloadducts (CA), and identifies the reactive site for each pathway to the (CA).	-	0 (16)	416 lines
RFD	Carries out the atomic (fragment) decomposition of the reaction force (constant)	IRC	2(20)	525 lines
Guess	Analyzes outputs of the redundant coordinate optimization (pseudo-guesses) and builds inputs for the guess TS	-	2 (27)	817 lines

The AMADAR package contains the “*RFD*” module, which executes the atomic decomposition of F and κ_{ξ} using IRC paths and information about the indexes of atoms to consider in the decomposition. Two flavours of methods for plotting the results have been implemented, allowing the rendering of contributions in terms of atomic or fragment contributions. In the latter case, the individual contributions of atoms forming the fragment are summed up to obtain the fragment contribution. To perform these analyses, the “*RFD_FLAG*” flag in the *da.ini* file must be set to 1 before running the *myIRCAalyzer.py* script. Details about the different configuration files are given in the next section. It must be noted that Hellman-Feynman (HF) forces are extracted for Gaussian output files, forming a series of $n \times 3$ matrices, with n the number of atoms in the molecule. The number of matrices is dictated by the length of the IRC path, each point being associated with a given geometry and a matrix of HL forces.

4. Configuration files

Two configuration files must be edited by the user to guide the code during the execution. The user has to provide details related to the levels of theory, the flow of execution, etc.

The first configuration is named *da.ini*. This file must be edited prior to any generation of transition states. The *da.ini* file has two sections, namely [**job_details**] and [**flags**]. These sections work like standard python dictionaries, with keywords linked to values. The [**job_details**] section contains 9 keywords. For example, the *CALC_LEVELS_RC* keyword tells the levels of theory to be used in the optimization of reactants and cycloadducts. This keyword must have two values separated by a semi-colon. Users are advised to choose a semi-empirical and QM levels as the 1st and 2nd levels respectively. The aim is to ensure a soft improvement sequence from the UFF conformer to a QM predicted geometry. The *NBR_PATHS* keyword indicates the nature of IRC path to be constructed. This tells the code either to construct the two directions from the TS in the same file or to generate two different files for each direction away from the TS. The latter option has not been activated yet, but will be very soon since the rest of analyses are still based on a unique IRC file containing both directions.

Table 2. First configuration file: da.ini

Section	Keyword	Description	Usage
[job_details]	NPROCshared	number of processors to be used for the jobs	Can have any integer value according to the computer platform
	D_SPLITTING	Final distance (in Å) between the two fragments during the constrained optimization.	Default value = 2.15Å. Should be chosen based on chemical intuition
	CALC_LEVELS_RC	Levels of theory to be used in the optimization of reactants and cycloadducts. Two levels must be given, separated by a semi-colon.	The defaults are PM6 and b3lyp/6-31G(d).
	CALC_LEVELS_TS	Levels of theory to be used in the optimization of reactants and cycloadducts. Three levels must be given, separated by a semi-colon.	The first level is used in the constrained optimization of the cycloadduct towards the pseudo-guess TS. The second level is for the refinement of the pseudo-guess into a guess TS. It is better using the same level as the 1st. The third level is needed for the final refinement of the guess-TS to the targeted QM accuracy. The defaults are PM6, PM6 and b3lyp/6-31G(d).

CALC_LEVEL_IRC	The level of theory to be used in IRC calculations.	Default is B3LYP/6-31G(d). This should be the same as the one used to refine the guess into the likely TS.
NBR_IRC_POINT_PP	Maximum number of points to use in the construction of the IRC path.	Default = 60 is good starting choice for mid-sized molecules, but may be increased or decreased depending on the system.
IRC_STEP_SIZE	Step-size to consider in the determination of the IRC path.	The default is 8, corresponding to $0.08 \text{ (amu)}^{0.5} \text{ Bohr}$
NBR_PATHS	Nature of IRC path to be constructed.	This keyword can have two states: 1 in case we want to build a unique IRC file or 2 if we want to have 2 separate input files.
TS_ID_numbers	ID numbers of the reactions for which for which the IRC path has to be generated. These values correspond to the position (line) of the SMILES string in the SMILES.txt file.	These must be integers values separated by commas. This flag can also be set to 'ALL', which means all the TS predicted will be involved in IRC calculations.

[flags]	SCRATCH	Tells the initialization code whether it should overwrite the existing files in the R, C and TS folders or not.	This flag can only have 1 and 0 values, which tell the code to overwrite or not the existing folders. In case 0 is chosen, then there must already exist files in the R, C and TS folders.
	RC_FLAG	Tells the initialization code whether it should optimize the reactants and cycloadducts geometries or not.	This flag can only have 0 and 1 values. Value 0 urge the code not to optimize the reactants and cycloadducts.
	TS_FLAG	Tells the initialization code whether it should optimize the TS or not.	This flag can only have 0 and 1 values. Value 1 tells the code to optimize the TSs, whereas 0 says the opposite
	IRC_FLAG	Tells the initialization code whether it should run IRC calculations or not.	This flag can only have 0 and 1 values. Value 1 tells the code to run IRC calculations, whereas 0 says the opposite. This will work if and only if there are TS already predicted and saved in the ...TS/GTS/TS folder.
	IRC_GEOMS_CONSTR	Tells the code to generate geometries of the system along the IRC path.	This will work only if there are IRC paths already constructed and saved in the IRC folder. Outputs are saved in the "paths" folder.

RFA_FLAG	Authorizes or blocks the code to run the reaction force analysis.	This flag can only have 0 and 1 values. Value 1 is for authorizing, 0 for blocking.
RFD_FLAG	Authorizes or blocks the atomic resolution of energy derivatives (along the IRC path) to be performed.	This flag can only have 0 and 1 values. Value 1 is for authorizing, 0 for blocking.
WBOA_FLAG	Controls the execution of Wiberg bond order analysis based on IRC geometries. It will work only if IRC natural population calculations have already been performed on the IRC geometries. These are saved in the "paths" folder.	This flag can only have 0 and 1 values. Value 1 enables the Wiberg Bond Order analysis to be carried out, while 0 is for the opposite.

Table 3. Second configuration file: analysis.ini

Section	Keyword	Description	Usage
[RFA]	Unq_RFA	This keyword indicates the list of ID numbers of IRC files to include in the reaction force analysis. Values must be integers, separated by commas.	In case all the files are to be considered, the value "-1" must be used. The value 0 means that no file is to be analysed.
	Multiple_RF	It indicates the list of ID numbers of IRC files to include in the multiple RFA in order to superimpose their reaction force curves. Values must be integers, separated by commas.	The value "-1" is not allowed as the number of overlayable systems must stay limited for efficiency. The value 0 disables the analysis
	Multiple_RE	It indicates the list of ID numbers of IRC files to include in the multiple RFA in order to superimpose their potential energy curves. Values must be integers, separated by commas.	The value "-1" is not allowed as the number of overlayable systems must stay limited for efficiency. The value 0 disables the analysis.
	Multiple_RFC	This keyword indicates the list of ID numbers of IRC files to include in the multiple RFA in order to superimpose their reaction force constant curves. Values must be integers, separated by commas.	The value "-1" is not allowed as the number of overlayable systems must stay limited for efficiency. The value 0 disables the analysis.

[RFD]	JOB_ID	This keyword indicates the ID number of IRC file to consider for the atomic decomposition of energy derivatives.	The value 0 disables the analysis.
	ATOMS	This keyword gives the list of atomic indexes to involve in the decomposition. Values must be integers, separated by commas.	The value 0 disables the analysis.
	FRAG	This keyword is for the list of fragments to involve in the decomposition.	Fragments are separated by semi-colons, while indexes of atoms from the same fragment are separated by commas. Make sure all the values are integers. The value 0 disables this analysis.
	FRAG_NAMES	Names of the fragments whose indexes of constituting atoms were given in the FRAG keyword.	Names are strings and must be given in the same order as indicated in the FRAG keyword. This section can be left blank or put any value when FRAG=0 because no fragment decomposition will be made.

[IRC_PATHS]	IRC_ID	ID number of the IRC file for which the geometries have to be extracted and run (single point calculation, with pop analysis)	ID numbers must be integers.
	LEVEL	Level of theory to use for the previous job	Default =B3LYP/6-31G(d)
	WBOA_ID	ID number of the system (IRC file) for which a Wiberg Bond order analysis should be done based on the output of the previous step.	This will work if single point calculations for the geometries extracted from the IRC path of the system of interest was be run and saved in the "path" folder. The value 0 blocks this module to be executed.

The **[flags]** section contains flags that authorize or forbid the code to run specific jobs. All of these flags are binary, i.e. they can only have two values: 0 and 1. The value 1 stands for authorizing, while 0 is for blocking the module or task. For example, the “RFA_FLAG” flag tells the code to run reaction force analysis. Note that these flags have priority on the keywords of the analysis.ini configuration file.

The analysis.ini file controls all the analyses based on IRC results. Information given in this file will only count if the flags in the da.ini file authorize the corresponding analyses to be performed. This file has three sections, namely **[RFA]**, **[RFD]** and **[IRC_PATHS]**. Details about the meaning of each keyword of these sections and their usage are given in Table 3.

The default settings of the da.ini and analysis.ini configuration files are provided in the Appendix 1 and 2.

5. Performance and case study

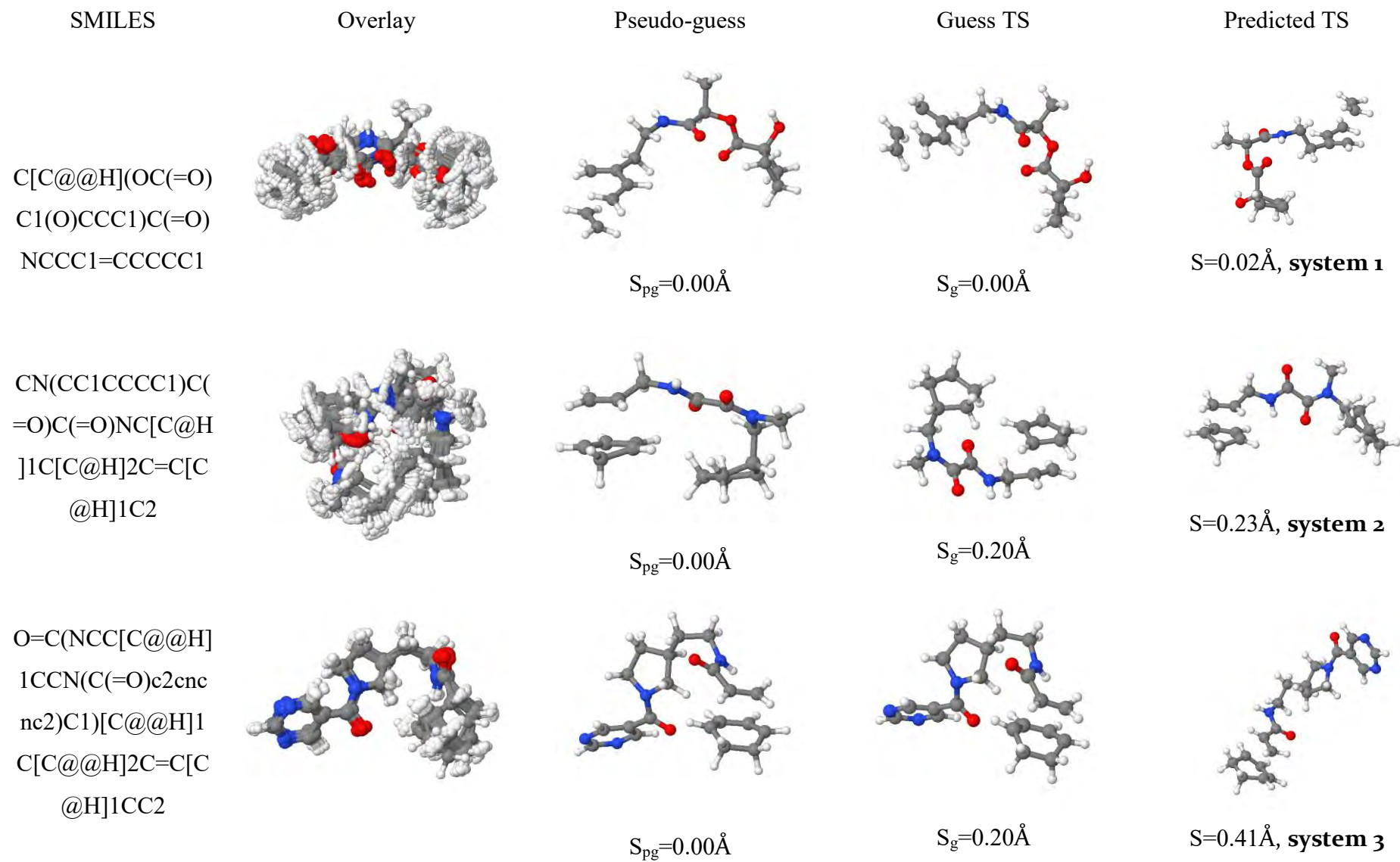
To assess the performance of the AMADAR program in predicting DA transition state geometries, 2000 cycloadduct SMILES strings were extracted from the ZINC database and pumped into the pipeline, keeping the default settings of the program. In reality, these compounds were randomly sampled from an original dataset of 5.7 million compounds containing at least one cyclohexene substructure.

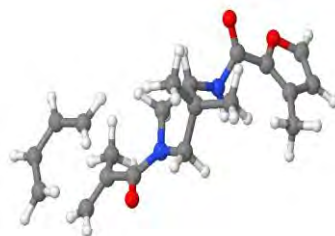
As explained in the section 3 (The Algorithm), each SMILES string had to be embedded into a 3D box and minimum-optimized using the universal force-field (UFF) [255]. Then, the returned conformer was subjected to a constrained optimization in internal coordinates, coercing the system to adopt a new configuration in which the terminal C atoms of diene and dienophile moieties were placed 2.15 Å apart (pseudo-guess TS). This two-fragments structure was then sequentially refined using the single-ended Berny algorithm at the semi-empirical PM6 and quantum mechanics b3lyp/6-31G(d) levels.


The analysis of outputs showed that about 1910 were successfully converted into the corresponding TSs, representing 95% of success. Each of these structures was confirmed to possess a unique imaginary frequency lower than 200 i. Basis set inconsistencies and unmet convergence criteria were the main causes of errors. Basis set errors were returned for all the systems containing the iodine atom, which cannot be described by the 6-31G(d) basis set. Errors in convergence criteria were observed for large systems and resulted in the abortion of


the calculation. A separate module is being developed to address these issues, and three others, in a systematic way and will be integrated into the next release of the package.

Figure 11 illustrates the results obtained for a very reduced set of four systems. For each system, this figure gives the pseudo-guess (PGTS), guess (GTS), predicted TSs geometries and the SMILE strings of the corresponding cycloadducts. These images are presented alongside the overlaying of the 16 configurations returned after the constrained optimization towards the PGTS. The predicted TSs cover a wide range of synchronicity indexes (0.00 to 1.07 Å), measured as the difference between the lengths of the two nascent C-C bonds. As revealed by the S_{pg} values, pseudo-guess geometries are symmetrical structures in which the diene and dienophile are placed at a user-defined distance from each other (default value = 2.15 Å). The refinement of the pseudo-guesses increases the asymmetry (asynchronicity) of the structure around the reactive site, from the pseudo-guess to the predicted TS. Obviously, the extend of this asymmetry depends on the nature of substituents inserted in the diene and dienophile. In the next chapter, we present in a more detailed fashion the results obtained upon the application of the AMADAR tool on the dataset of 2000 DA cycloadducts. These results are mainly discussed in terms of the (a)synchronicity of DA reactions.

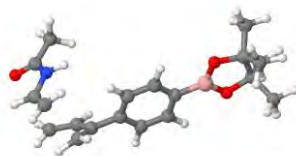










CC(=O)N[C@H]1C=C(c2ccc(B3OC(C)(C)C(C)(C)O3)cc2)CC1







77

6. Concluding remarks

In this chapter, we have presented the AMADAR program developed in our research group for the prediction of DA transition state geometries and the analysis of DA reaction paths. The performance of the program has also been discussed. The AMADAR program is expected to be very beneficial to a broad community of researchers working on the mechanism of DA reactions. Further works should integrate new modules in the package allowing for the investigation of hetero-Diels-Alder reaction as well as automating the investigation of stepwise IRC profiles in the case of very asynchronous reactions alongside with the concerted ones. In addition, implementation of modules that enable the use of GAMESS, as an alternative of Gaussian are planned. This work has led to a research article (short communication), already submitted in the *Journal of Cheminformatics* for publication.

“A valid scientific theory is predictive, verifiable, and replicable. To me, that’s beautiful.”

Dean Ornish

Chapter 3: On the (a)synchronicity of Diels-Alder reactions

0. Summary

The (a)synchronicity of Diels-Alder (DA) reactions is an open question that has gained the interest of many scientists and has generated many previous publications. In the present chapter, we report on some new insights into the origin and physical meaning of the (a)synchronicity in DA reactions, considering a large dataset of 150 systems. The results are discussed in the framework of the reaction force analysis, in conjunction with natural population calculations and atomic energy derivatives along the intrinsic reaction coordinate (IRC) path. The same methodologies were also utilized in previous works and were found to shed light on the mechanism of other chemical reactions. Our results suggest that the transition from concerted to stepwise DA processes may cautiously be located in a 0.10 Å window of synchronicity indexes, expanding from 0.90 to 1.00 Å. Further, the position of the global minimum of the reaction force constant with respect to the TS is proposed as being an alternative and quantifiable indicator of the (a)synchronicity in DA reactions. Finally, the atomic resolution of energy derivatives reveals that the mechanism of the DA reaction involves two inner elementary processes associated with the formation of the two C-C bonds. This decomposition went on to indicate that, in asynchronous reactions, the interplaying driving and retarding forces are mainly caused by the fast and slow-forming bonds (elementary process) respectively, while in the case of synchronous ones both elementary processes retard and drive the process concomitantly and equivalently.

1. Introduction

The mechanism of the Diels-Alder (DA) reaction has been the subject of numerous experimental [15] and theoretical studies [16]. Despite some isolated cases where the DA reaction has been found to follow a stepwise scheme [29], the debate around the mechanistic route of DA reactions has suggested that most DA reactions proceed through a concerted pathway [28].

While the concerted pathway is now widely viewed as being the putative mechanistic route of DA reactions, this does not imply that they are all synchronous on a time-resolved scale. Various studies have attempted to elucidate the origin of the (a)synchronicity in DA reactions using different approaches, including, but not restricted to, the frontier molecular orbital theory [35], molecular electron density theory [36] and activation strain model [39]. On a special note, the reaction force analysis has revealed that the second derivative of the system's energy with respect to the reaction coordinate (known as the reaction force constant) was a good indicator of the synchronicity of DA reactions [38].

It must be admitted that all of these studies have significantly contributed to piercing the mystery around the (a)synchronicity of DA reactions. However, a variety of pertinent questions and dark zones regarding the matter still remain. First, it must be noted that all of these studies have investigated specific or reduced-size samples of DA reactions, mostly similar, which obviously gives rise to concerns about the extrapolation of the findings. Second, to our knowledge, no one has tried to understand thoroughly the individual role(s) of each atom, especially those involved at the reactive site, in the overall (a)synchronicity of DA reactions. It is sad to note that this fundamental question cannot be answered even in the context of the standard Reaction Force Analysis (RFA), which only provides an explanation of the mechanism in terms of energy derivatives[257]. In the same context, the physical meaning of critical point in the profile of the reaction force constant as defined in the reaction force theory has not received much attention.

The question about the individual role of specific atoms or fragments in the mechanism of reactions can be addressed by decomposing energy derivatives along the reaction path into atomic contributions. In 2016, Ledrzejewski et al [40] demonstrated the feasibility of doing such a decomposition by referring to the Hellman-Feynman theorem, which defines the nature of individual forces acting on nuclei in molecular systems[258]. Their promising approach goes

beyond the standard reaction force analysis, bringing up the missing breaks for an accurate atomic resolution of the reactive force and reaction force constant. This has been implemented in the previous chapter in the AMADAR program, along with the standard reaction force analysis.

In this chapter, we report the results obtained by applying the AMADAR program on a large dataset of 2000 likely DA cycloadducts, and discuss them in terms of the (a)synchronicity of the corresponding DA reactions. To the best of our knowledge, this is the first time the atomic resolution of energy derivatives is applied on DA reactions, and the first time the reaction force analysis is extended to such a large dataset. In addition, natural populations calculations have been performed to understand the physical meaning of critical points in the profile of the reaction force constant.

2. Methodology

SMILES strings of 2000 potential Diels-Alder cycloadducts were retrieved from the ZINC database [156] (SMILES accessible from this GitHub repository ([BienfaitKI/MSc_Thesis github.com](https://github.com/BienfaitKI/MSc_Thesis))) and pushed into the AMADAR pipeline. These compounds were randomly sampled from an original dataset of 5.7 million compounds containing at least one cyclohexene substructure. Using the default settings of the AMADAR program, each SMILES string was embedded into a 3D box and minimum-optimized using the universal force-field (UFF) developed in Rappé's group[255]. Then, the returned conformer was subjected to a constrained optimization in internal coordinates, coercing the system to adopt a new configuration in which the terminal C atoms of diene and dienophile moieties were placed 2.15 Å apart (pseudo-guess TS). This two-fragments structure was then sequentially refined using the single-ended Berny algorithm[259] at the semi-empirical PM6[191] and quantum mechanics B3LYP/6-31G(d)[260][202] levels.

For each TS predicted, the synchronicity index *S* was estimated as the difference between the lengths of the two C-C bonds. Then, natural population [261] calculations were performed to evaluate the Wiberg bond orders of the two emerging C-C bonds as well as the charge transfer between the diene and dienophile in the TS geometry. The latter was identified with the absolute value of the global NPA charge on each fragment. It is worth noting that the B3LYP/6-31G(d) level of theory has been successfully utilized in previous studies to investigate the mechanism of the Diels-Alder reaction [17][34] [256].

Furthermore, 150 of the predicted TSs were used for determining the IRC paths with a step size of $0.8 \text{ (amu)}^{1/2} \text{ Bohr}$ at the B3LYP/ 6-31G(d) level. To obtain this sample, the TS dataset was first split into 11 groups of synchronicity degree, where each group corresponded to a 0.10 \AA gap in S values, starting at $S=0.00 \text{ \AA}$ to most asynchronous TS. Then, the 150 systems were selected using a python script that randomly collected up to 20 hits in each 0.10 \AA synchronicity index gap. The IRC paths constructed were first examined in the framework of the reaction force theory. Then, the reaction force and reaction force constants were decomposed in terms of atomic (fragment) contributions using the equations proposed by Ledrzejewski et al [40] and implemented in the AMADAR package. Finally, to assess the bond breaking/formation processes during the reaction, NBO calculations were performed at the b3lyp/6-31G(d) level on the geometries extracted from the IRC paths.

3. The dataset

Among the 2000 DA cycloadducts extracted from the ZINC database, about 1910 were successfully converted into the corresponding TSs, representing a success rate of 95%. Each of these structures was confirmed to possess a unique imaginary frequency lower than 200 i . The 3D geometries and the numbering used in this study to denote all the predicted TSs and corresponding reactions are accessible from the following GitHub repository ([BienfaitKI/MSc_Thesis \(github.com\)](https://github.com/BienfaitKI/MSc_Thesis)).

The retro-DA transformation of the 2000 cycloadducts led to 90 different dienes and 1140 dienophiles. 1,3-butadiene was the most frequent diene, involved in $\sim 55 \%$ of the DA reactions, followed by 1,3-cyclopentadiene and 1,3-cyclohexadiene that counted for 10%. The remaining dienes were substitutional derivatives of the previous. The ethylene molecule was the most popular dienophile, representing about 30 % of the entire population. The comprehensive list of reactants and cycloadducts SMILES is provided in a separate file here ([MSc_Thesis/Reactants and Cycloadducts.xlsx at main · BienfaitKI/MSc_Thesis \(github.com\)](https://github.com/BienfaitKI/MSc_Thesis/blob/main/MSc_Thesis/Reactants_and_Cycloadducts.xlsx)). The dataset of TSs generated covers a wide range of electronic and structural features. Its diversity was examined in terms of polarity and synchronicity (Figure 12).

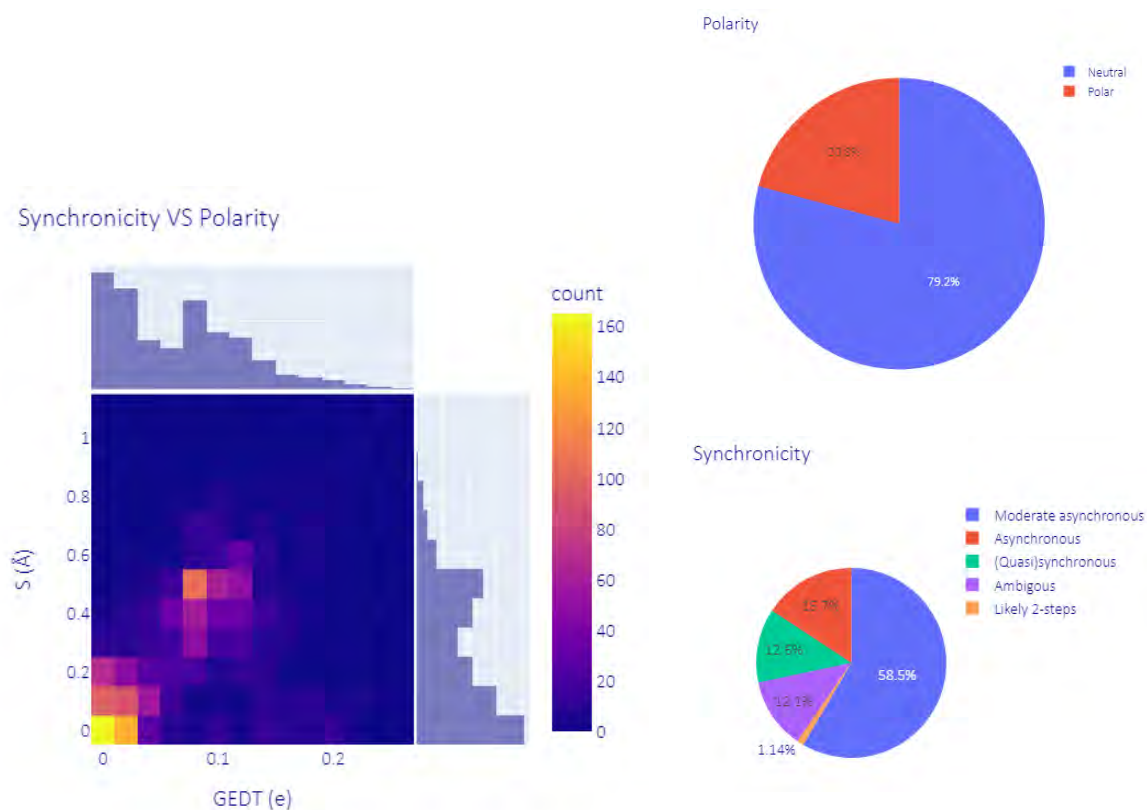


Figure 12. Synchronicity and polarity of the predicted TSs

The magnitude of the electron density exchanged between the diene and dienophile at the TS was computed to assess the polarity of each TS. This amount is known as the “global electron density transfer” (GEDT) in the framework of the molecular electron density theory[256]. Predicted GEDT values ranged between 0.00 and 0.26 e. According to the classification of DA reactions proposed here[262] based on GEDT values, our dataset encompasses 79.2 % of neutral DA reactions ($0 < \text{GEDT} < 0.15$ e) and 20.8 % of polar ones ($0.15 \text{ e} < \text{GEDT} < 0.45$ e). Furthermore, the sign of the global charge on each fragment revealed that the dienes were positively charged in 73 % of the predicted TSs, which is a typical feature of normal electron-demand DA reactions. In opposition to Domingo and his co-workers who reported a good correlation between the polarity and the activation energy ($R^2=0.99$) [40], no such correlation was observed in our dataset ($R^2 = 0.32$, Figure 13). This is probably due to the high heterogeneity of our dataset as compared to theirs, which consisted of similar DA reactions involving a unique diene (cyclopentadiene) against twelve substituted ethylene. Therefore, it

may be postulated that high correlations between GEDT values and activation energies should only be expected within groups of very similar DA reactions.

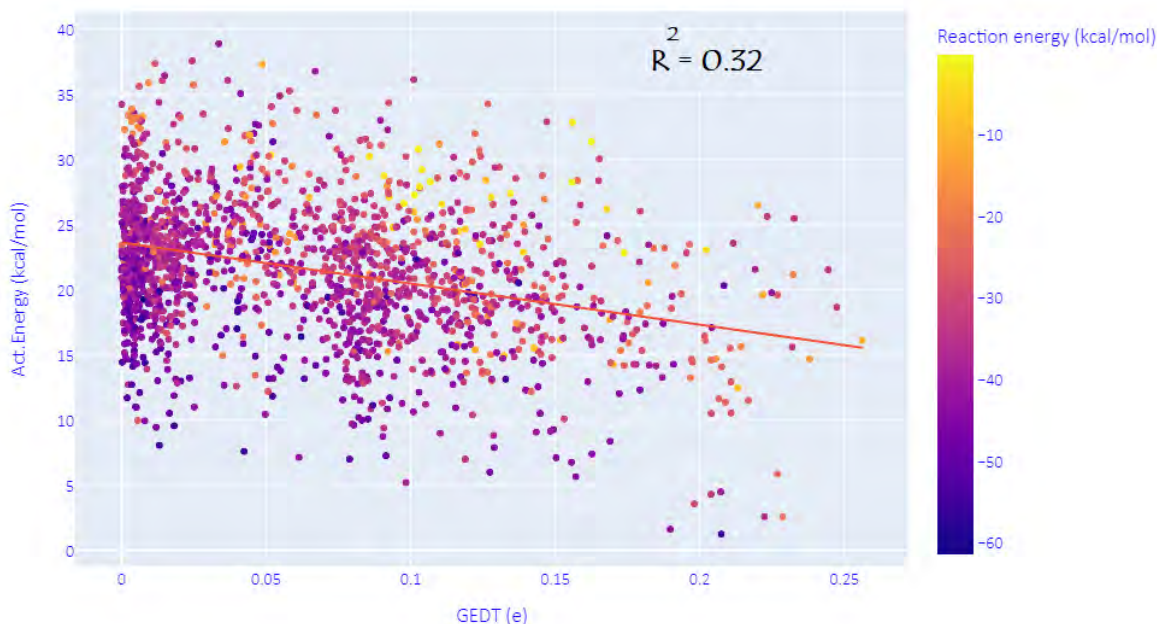


Figure 13. No linear relationship between the polarity and activation energy of DA reactions

We assessed the (a) synchronicity of the TSs (DA reactions) using the widely admitted structural index S , which is defined as the difference between the lengths of the two emerging C-C bonds at the TS. As shown in Figure 13, 58.5 % of the reactions considered in this study are moderate asynchronous processes, with S values comprised between 0.25 Å and 0.55 Å. The rest of systems comprise 12.6 % of (quasi)synchronous ($0.00 < S < 0.20$ Å), 15.7 % of asynchronous ($0.60 < S < 0.90$ Å) and 1.1 % of likely two-steps DA reactions ($S > 0.95$ Å). This alternative classification of DA reactions is here proposed based on the analysis of the reaction force constant profile (section 4.2), which revealed that, in each of these ranges of synchronicity index, the reactions present the same patterns on the profile of κ so that one may consider them as following similar mechanisms.

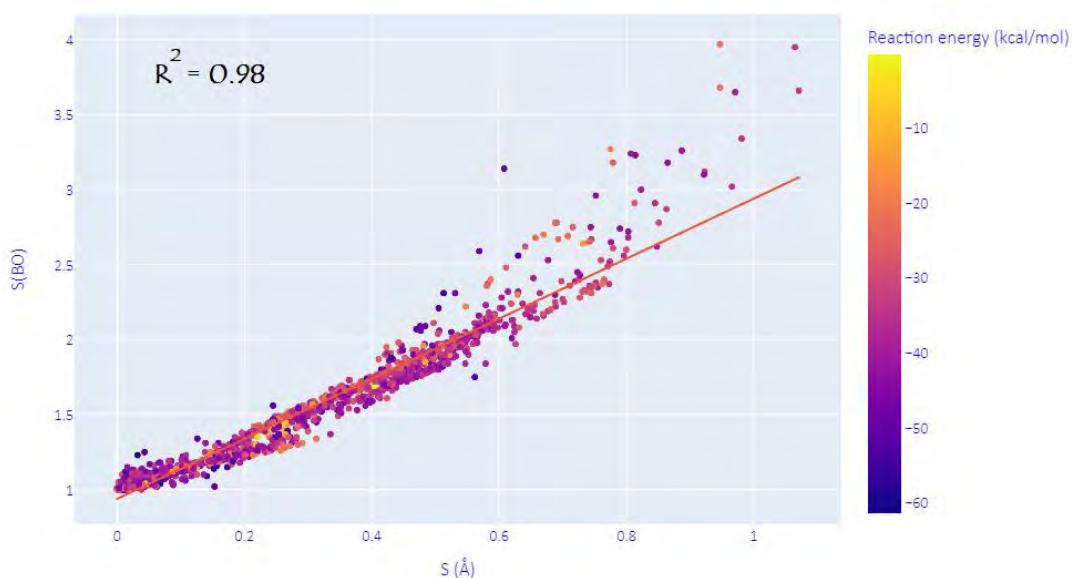


Figure 14. Correlation between S(BO) and S synchronicity indices.

In the aim of finding an alternative descriptor of (a)synchronicity in DA reactions, the Wiberg bond orders (WBO) of the two nascent C-C bonds were evaluated in all the TSs. Then, the natural logarithm of their ratio noted S(BO) was plotted against S (Figure 14), resulting in a correlation coefficient of $R^2 = 0.99$ between the two descriptors. This suggests that S(BO) and S are qualitatively equivalent and thus, can be used interchangeably to measure the (a)synchronicity of DA reactions. However, in the discussion below we have chosen to consider the commonly used index S for a better concord with previous studies.

4. Reaction force analysis

The reaction force theory is one of the most powerful reactivity paradigms developed during the last two decades. This theory has been successfully applied to several types of reactions, such as the deamination of the cytosine molecule[236] or intramolecular proton transfer reactions[237]. In contrast with the transition state theory[238], which considers a fixed position on the IRC path as the (classical) transition state, the reaction force theory defines a whole region including the classical TS, where the most electronic changes of the reaction take place. The concept of TS region is consistent with the TS spectroscopy of Zewai and Polanyi[239], which describes the TS region as a continuum of transient and unstable intermediate configurations of the reacting system.

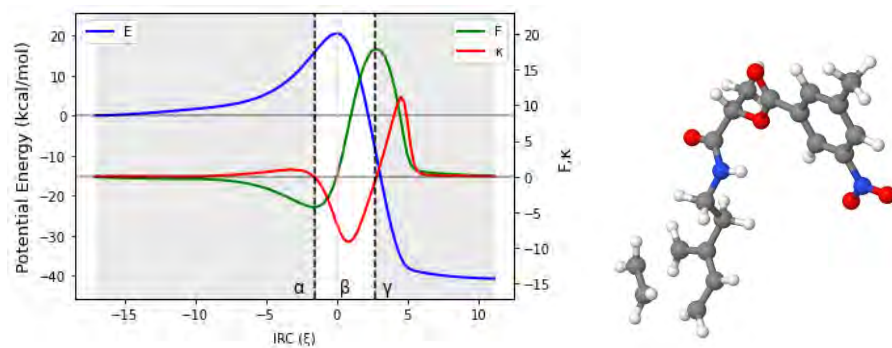
In the present study, the RFA was applied to 150 Diels-Alder reactions which were chosen as indicated in section 2 (paragraph 3). Figure 15 presents the profiles of the potential energy E , the reaction force F , and the reaction force constant κ of ten DA reactions among the 150 systems considered for IRC calculations to pave a wide range of synchronicity indexes. Be aware that qualitatively equivalent conclusions to those reported below were inferred from any set of reactions with similar degrees of synchronicity. For each reaction, the two components of the activation energy (see Chapter 1, section 4), $E_{\text{act},1}$ and $E_{\text{act},2}$, are given. The position of the classical TS is indicated by β , while the minimum and maximum of F are denoted by the Greek letters α and γ respectively.

Looking at $E_{\text{act},1}$ and $E_{\text{act},2}$, it comes out that the most important part of the activation energy is consumed in the preparation phase, taking on average about 75% of the energy required to bring the system up to its transition state. For example, the activation energy of the (quasi) synchronous **R73** DA reaction (20.6 kcal/mol) is predicted to be composed of 15.7 kcal/mol and 4.8 kcal/mol involved respectively in the preparation phase and the first part of the TS region at the B3LYP/6-31G(d) level. Similarly, the activation energy of the asynchronous **R1453** DA reaction is predicted to be 23.8 kcal/mol, with $E_{\text{act},1}$ and $E_{\text{act},2}$ accounting for 18.9 and 4.9 kcal/mol respectively. Mancera et al investigated the DA reactions between a 9,10-disubstituted chiral anthracene and a short series of dienophiles, and found that more than 70% of the activation energy was involved in the preparation phase [235]. While this feature is apparently common to all DA reactions, it should not be generalized. One exception is the $\text{S}_{\text{N}}2$ reaction between CH_3Cl and H_2O . This reaction requires approximately the same amount of energy during the preparation step ($\text{R} \rightarrow \alpha$) and the first part of the TS region, estimated at +27.0 and +25.0 kcal/mol respectively at b3lyp/6-31G(d) level[263]. It is worth mentioning that the profile of the potential energy is almost insensitive to the synchronicity of DA reactions. Only strongly asynchronous DA reactions ($S > 0.90\text{\AA}$) present a barely noticeable inflection point in the second portion of the TS region. As such, the potential energy curve cannot serve as an indicator for the synchronicity of DA reactions.

Figure 15 reveals that the reaction force F switches from negative (retarding) to positive (driving) values when the system crosses the classical TS. This indicates first the existence of a dominating retarding force in the preparation phase that opposes structural changes in the system[257]. But, as the system approaches the TS region, a positive driving force arises and foils the retarding one. The two forces acting on the system keep interplaying up to the TS, where their magnitude exactly counterbalance. Mechanically speaking, the TS is an

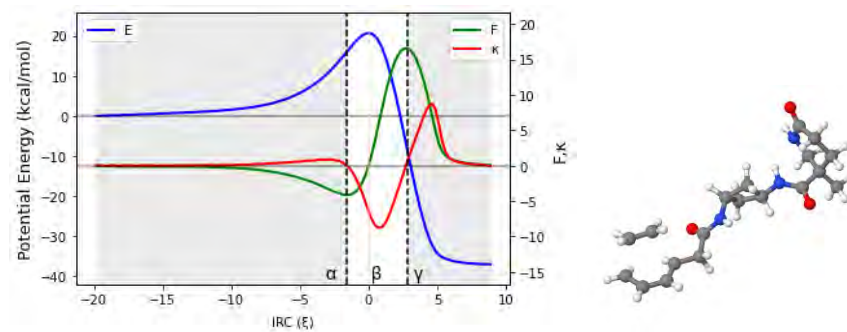
equilibrium point since the resultant of all forces acting on the system is null. After the TS, the driving force wins over the retarding one and drags the system to the product state.

In the framework of the activation strain model (ASM) developed by Bickelhaupt and his co-workers [264], the potential energy E at each point of the IRC path is decomposed into the strain and interaction components. By combining the ASM and the RFA, Politzer et al were able to quantify the retarding and driving forces that compose the total reaction force[265]. Their study showed that the retarding force observed in the preparation phase was the sum of positive strain and interaction (repulsive) forces. The two components were identified with the opposite gradient of the strain and interaction energy constituents of the system's potential energy. In this chapter (section 6), another picture of the same phenomenon is presented, where the reaction force appears to be mainly composed of two components associated with the formation of each C-C bond.



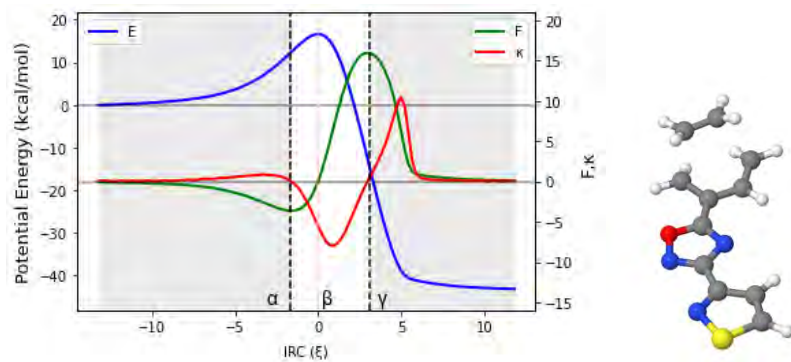
$$E_{\text{act}} = 20.58, E_{\text{act},1} = 15.73, E_{\text{act},2} = 4.84,$$

$$S = 0.00 \text{ \AA} \quad (\mathbf{R76})$$



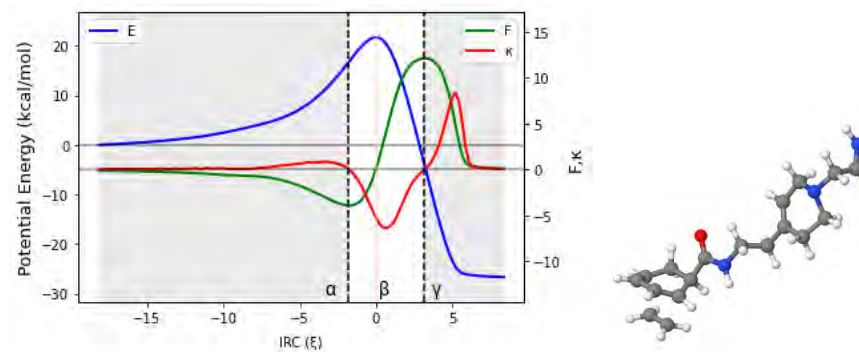
$$E_{\text{act}} = 20.70, E_{\text{act},1} = 15.79, E_{\text{act},2} = 4.91,$$

$$S = 0.12 \text{ \AA} \quad (\mathbf{R313})$$



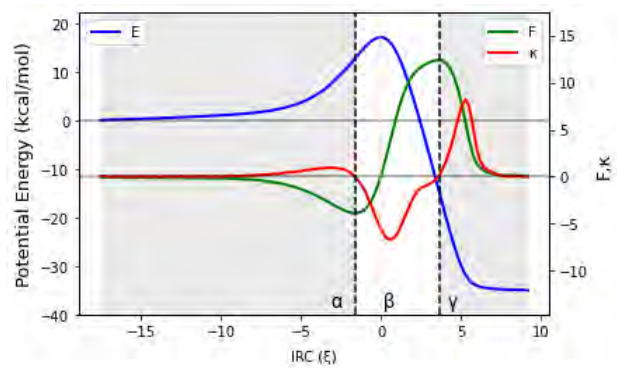
$$E_{\text{act}} = 16.64, E_{\text{act},1} = 12.30, E_{\text{act},2} = 4.34,$$

$$S = 0.25 \text{ \AA} \quad (\mathbf{R1336})$$



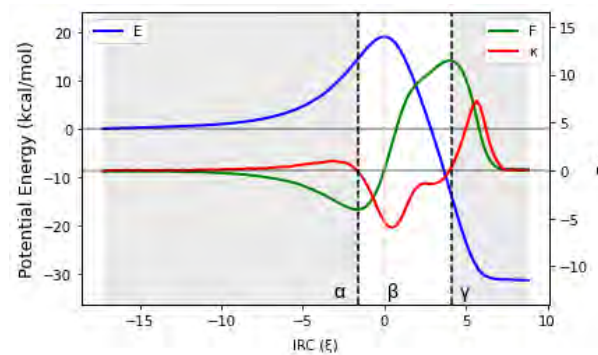
$$E_{\text{act}} = 21.69, E_{\text{act},1} = 16.56, E_{\text{act},2} = 5.13,$$

$$S = 0.37 \text{ \AA} \quad (\mathbf{R359})$$



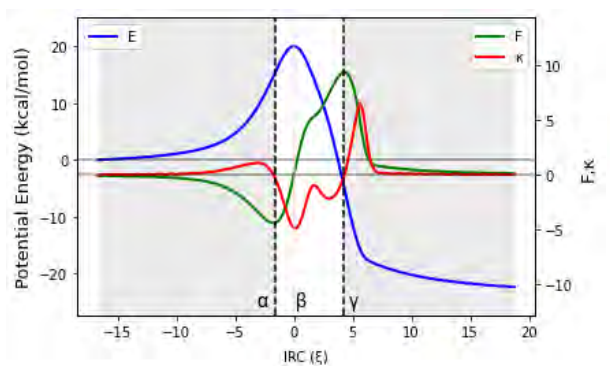
$$E_{\text{act}} = 17.09, E_{\text{act},1} = 12.65, E_{\text{act},2} = 4.44$$

$$S = 0.44 \text{ \AA} \text{ (R279)}$$



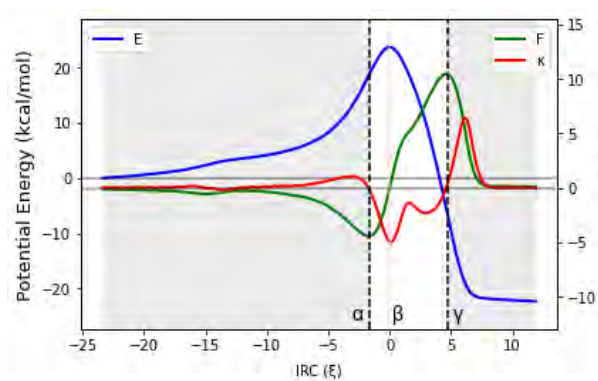
$$E_{\text{act}} = 19.06, E_{\text{act},1} = 14.64, E_{\text{act},2} = 4.42,$$

$$S = 0.54 \text{ \AA} \text{ (R793)}$$



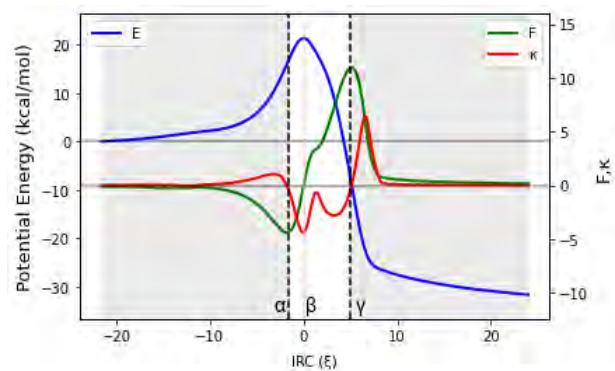
$$E_{\text{act}} = 20.06, E_{\text{act},1} = 15.23, E_{\text{act},2} = 4.83,$$

$$S = 0.65 \text{ \AA} \text{ (R318)}$$



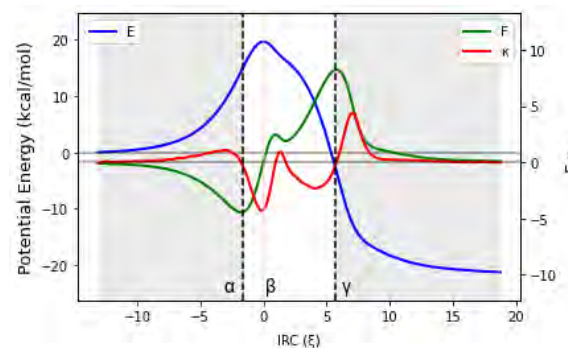
$$E_{\text{act}} = 23.73, E_{\text{act},1} = 18.87, E_{\text{act},2} = 4.86,$$

$$S = 0.76 \text{ \AA} \text{ (R1453)}$$



$$E_{\text{act}} = 21.38, E_{\text{act},1} = 16.48, E_{\text{act},2} = 4.9,$$

$$S = 0.82 \text{ \AA} \text{ (R727)}$$



$$E_{\text{act}} = 19.65, E_{\text{act},1} = 14.80, E_{\text{act},2} = 4.85,$$

$$S = 0.95 \text{ \AA} \text{ (R407)}$$

Figure 15. Reaction force analysis for an illustrative set of ten Diels-Alder reactions covering a wide range of synchronicity degrees (from $S = 0.00 \text{ \AA}$ to $S = 0.95 \text{ \AA}$). Each plot contains three graphs associated with the potential energy (E , in blue), the reaction force (F , in green), and the reaction force constant (κ , in red) of the considered reaction along the IRC path, as well as the image of the corresponding TS (color code: black=C, white=H, blue=N, red=O, green=F, yellow=S). Energies (E_{act} , $E_{\text{act},1}$ and $E_{\text{act},2}$) are given in kcal/mol, F in kcal/mol. ξ and κ in kcal/mol. ξ^2 .

Figure 15 reveals that the profile of the reaction force constant is the most affected by the changes in the degree of synchronicity, so that one may establish a direct correlation some patterns and the level of synchronicity. Therefore, the reaction force constant must be a good indicator of the (a)synchronicity in DA reactions. *Yepes et al* were the first to report this observation in their investigation of a series of seven DA reactions involving cyclopentadiene and cyanoethylenes[38]. Regarding the patterns appearing in the profile of κ , one can categorize DA reactions in four subsets including (quasi)synchronous ($0.00 < S \text{ (\AA)} < 0.20$), moderate asynchronous ($0.25 < S \text{ (\AA)} < 0.55$), fully asynchronous ($0.60 < S \text{ (\AA)} < 0.90$) and likely two-steps ($S \text{ (\AA)} \geq 0.95$) processes. Notice that these boundaries are not continuous. This is because the DA reactions within the gaps may be of either type depending on the structural and electronic features of each reaction. In every subgroup, the following features seem to be unanimous. In (quasi-)synchronous DA reactions, the reaction force constant κ has a unique minimum in the TS region. At the same time, F increases continuously in the same region, before crossing a sharp maximum. As the reaction becomes more and more asynchronous, some striking changes are observed on the graph of F and κ . For moderate asynchronicity ($0.25 < S \text{ (\AA)} < 0.55$), κ presents not only a minimum in the TS region but also an inflection point (shoulder) located after the minimum in the direction of the reaction coordinate. As we shall show in paragraph 6, this sudden appearance of an inflection point in the TS region of κ indicates the formation of the first C-C bonds. A slight change is also observed in the graph of F , which flattens around its maximum. For asynchronous DA reactions ($0.60 < S \text{ (\AA)} < 0.90$), the reaction force F shows an inflection point in the TS region, while the reaction force constant κ presents two minima and a maximum (between these minima) in the same region.

The sign of the reaction force constant over the TS region can be very informative. In the case of fully synchronous to asynchronous DA reactions, it remains negative all along the TS region. However, when the reaction becomes very asynchronous ($S > 0.95 \text{ \AA}$), the maximum of κ in the TS region is lifted to more and more positive values, and a new inflection point is (barely) noticeable in the TS region of E . A very drastic change is observed in the profile of the reaction force: the characteristic inflection point observed in the TS region of asynchronous reactions falls apart into a new local minimum and a local maximum. Similar observations were reported by *Politzer et al*[266] for a series of double proton transfer reactions. To the best of our knowledge, this is the first time this observation is reported in the case of DA reactions.

From a mechanistic view, the presence of a positive local maximum on the profile of κ in the TS region has been identified with the emergence of a metastable intermediate in the region situated between the new local extrema (maximum and minimum) of the reaction force [266]. To quickly assess this hypothesis, vibrational frequency calculations of all the geometries located in the second part of the TS region for the reactions **R313** ($S=0.12\text{\AA}$), **R279** ($S=0.44\text{\AA}$), **R727** ($S = 0.82\text{\AA}$), and **R404** ($S = 0.95\text{\AA}$) were carried out. This restricted analysis was intended to see if there was a link between the (a)synchronicity and the presence of a possible TS structure and subsequently a possible intermediate in the TS region. The results obtained revealed the existence of one TS among the structures in the previous region for the very asynchronous **R404**, and not in those of **R313**, **R279**, and **R727**. This finding is a prior argument upholding the assumption of a possible two-step mechanism for DA reactions with a synchronicity index higher than 0.95\AA . However, since the situation cannot be generalized for the moment due to the peculiarity of each system, it is more cautious to declare that the transition between a preferentially concerted mechanism and a two-steps pathway can be cautiously located between $S=0.90\text{\AA}$ and $S=1.00\text{\AA}$. Therefore, very asynchronous DA reactions ($S=0.95\text{\AA}$ and $S=1.00\text{\AA}$) are likely to proceed through a two-steps non-concerted pathway, with a diradical or zwitterionic intermediate. This conclusion is consistent with previous findings[31], but more interestingly, it adds a relevant detail regarding the range of synchronicity indexes where the transition from the concerted to the stepwise routes in homo-DA reactions may be observed.

5. Natural population analysis

Natural population calculations were performed on the geometries extracted from the IRC paths of 150 DA reactions. Figure 16 depicts the convention used herein to designate the six atoms and six bonds involved in the reactive site. In Figure 17, the Wiberg bond orders of the six bonds involved in the reactive site are plotted against the reaction coordinate for an illustrative set of ten DA reactions ($0 < S(\text{\AA}) < 1$). The position of global maximum of the reaction force constant, noted ω , is also indicated.

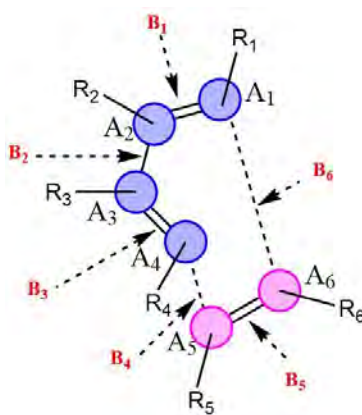


Figure 16. Reactive site of DA reactions: atoms and bonds from the diene (blue) and dienophile (pink)

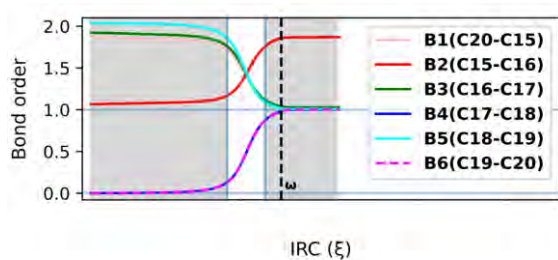
Figure 17 reveals that both the bond orders and atomic charges remain almost constant all over the preparation phase. This finding supports the insignificance of electronic changes during this step and confirms previous findings by Burda et al[267]. Furthermore, the constancy of atomic charges in the reactive site during the preparation phase suggests that the overall interaction between the dienes and dienophiles are mostly of electrostatic nature since very little charge transfers are experienced between the antagonists. This agrees well with Politzer et al who found that the diene and dienophile experienced mostly a purely repulsive electrostatic interaction in the preparation phase due to the interaction between their π clouds[265].

The six bonds in the reactive site begin to change considerably before entering the TS region. The most important changes in bond orders take place in the TS region, in agreement with the fact that this region is admitted to be dominated by electronic changes. However, it is only after crossing the second maximum of $\kappa(\omega)$ that all the bonds in the reactive site reach their bond order asymptotic values and can then be considered as completely formed. Since the reactive site is the part of the system that is the most affected by the transformation, we can postulate that ω represents the moment of the reaction when the system becomes a topological equivalent of the cycloadduct, as it now contains the same types and number of bonds.

The relative positions of the B4 and B6 curves constitute another indicator of the synchronicity of DA reactions. For a fully synchronous DA reaction, the two curves are perfectly overlaid, suggesting the two bonds to form at the same rate and time. At the position of the classical TS, the nascent C-C bonds are almost halfway in the process of their formation, with bond orders around 0.5. Therefore, one can assume that synchronous DA reactions proceed through cyclic

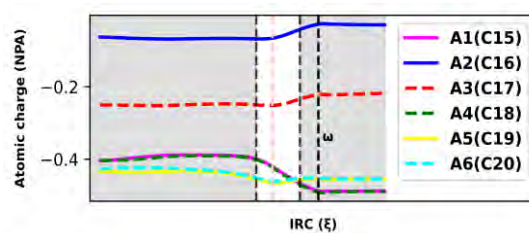
TSs where the terminal C atoms are non-covalently bound. For asynchronous processes, one of the two C-C bonds is delayed and this deviation is as much pronounced as the reaction is asynchronous. In addition, the first C-C bond is (almost) completely formed in the TS region, reaching a bond order > 0.95 . In very asynchronous processes such as **R404** and **R727**, we could not get fixed at first on whether the TSs were cyclic or not since the bond order of the slow-forming bond was still very small at the classical TS (< 0.25). However, a quick topological QTAIM analysis of the TS of **R404** and **R727** revealed the existence of bond critical points between the two pairs of terminal C atoms, which gainsaid the assumption of an acyclic structure. Similar findings were obtained by Hammoudan et al on a set of 3 asynchronous hetero-DA based on the QTAIM approach [268]. Therefore, based on the results we have got so far, there is no sufficient proof to deny the cyclicity of asynchronous DA TSs, at least for the systems we have investigated.

R76



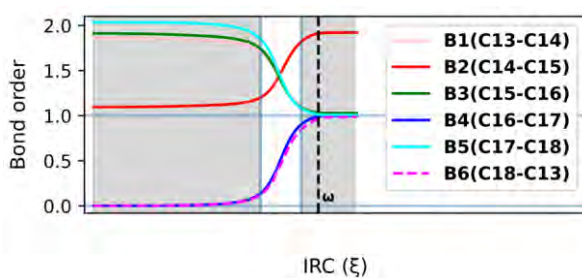
$S = 0.00 \text{ \AA}$, $S_{BO} = 1.01$, $CT = 0.001$, Flux = INVERSE

R76



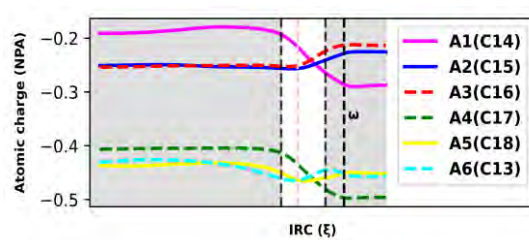
$S = 0.00 \text{ \AA}$, $S_{BO} = 1.01$, $CT = 0.001$, Flux = INVERSE

R313



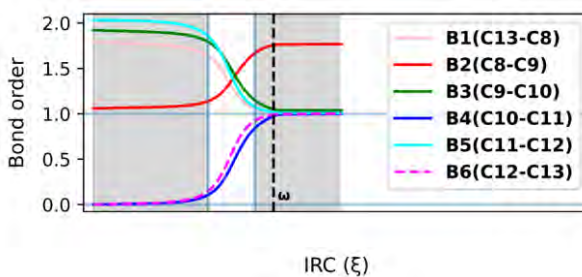
$S = 0.12 \text{ \AA}$, $S_{BO} = 1.14$, $CT = 0.002$, Flux = NORMAL

R313



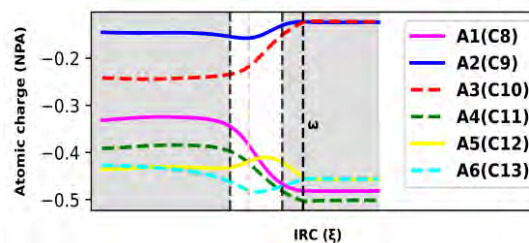
$S = 0.12 \text{ \AA}$, $S_{BO} = 1.14$, $CT = 0.002$, Flux = NORMAL

R1336



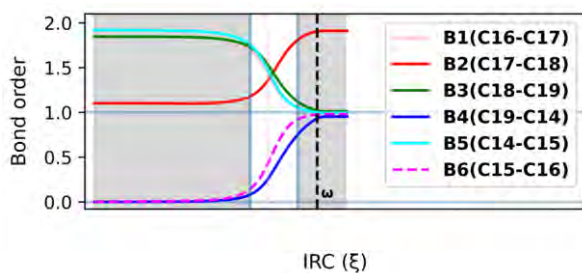
$S = 0.25 \text{ \AA}$, $S_{BO} = 1.37$, $CT = 0.05$, Flux = NORMAL

R1336



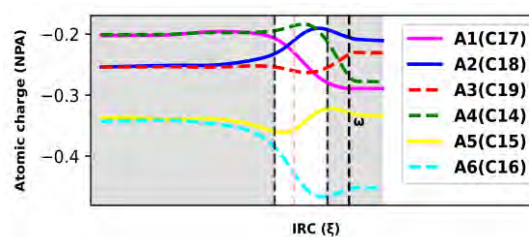
$S = 0.25 \text{ \AA}$, $S_{BO} = 1.37$, $CT = 0.05$, Flux = NORMAL

R359

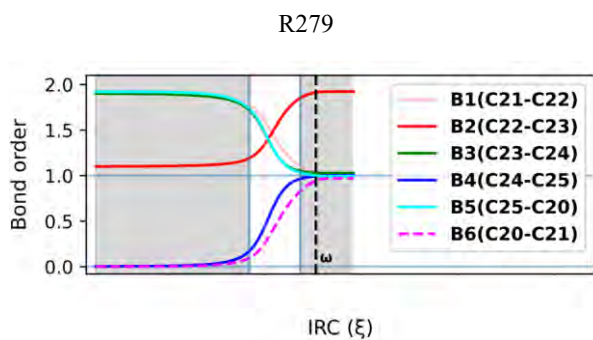


$S = 0.37 \text{ \AA}$, $S_{BO} = 1.68$, $CT = 0.11$, Flux = NORMAL

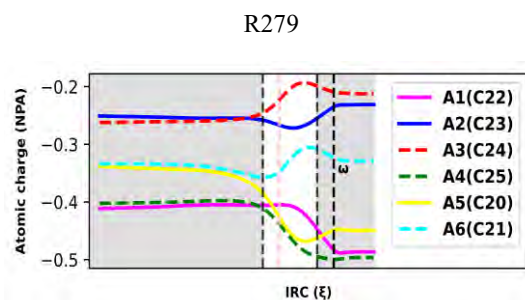
R359



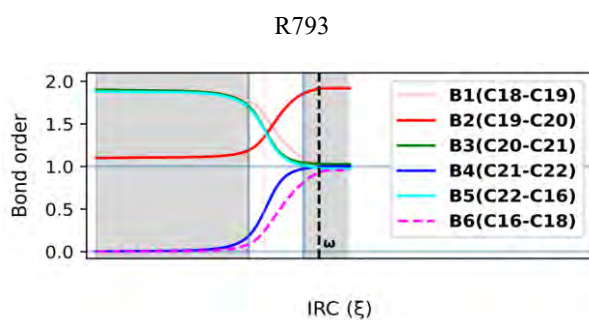
$S = 0.37 \text{ \AA}$, $S_{BO} = 1.68$, $CT = 0.11$, Flux = NORMAL



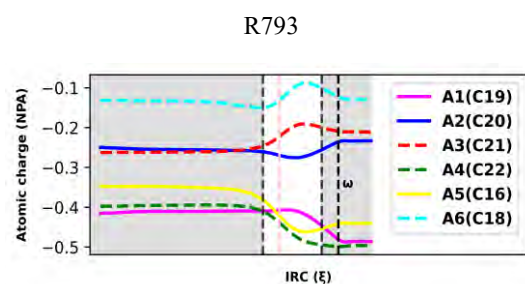
$S = 0.44 \text{ \AA}$, $S_{BO} = 1.71$, $CT = 0.08$, Flux = NORMAL



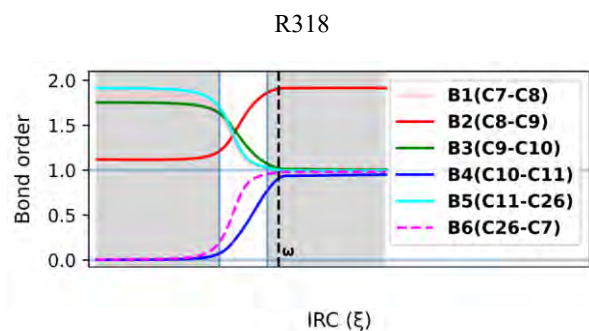
$S = 0.44 \text{ \AA}$, $S_{BO} = 1.71$, $CT = 0.08$, Flux = NORMAL



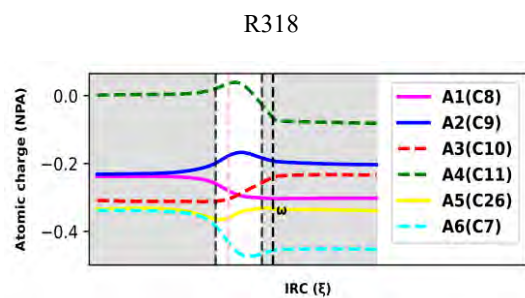
$S = 0.54 \text{ \AA}$, $S_{BO} = 1.97$, $CT = 0.08$, Flux = NORMAL



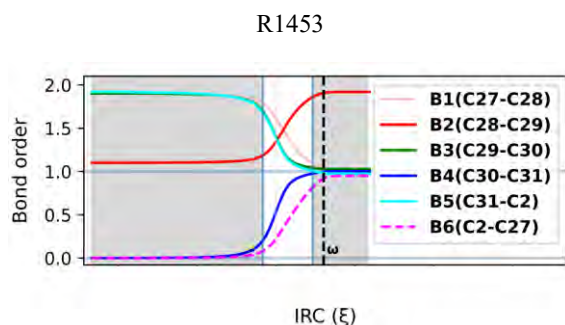
$S = 0.54 \text{ \AA}$, $S_{BO} = 1.97$, $CT = 0.08$, Flux = NORMAL



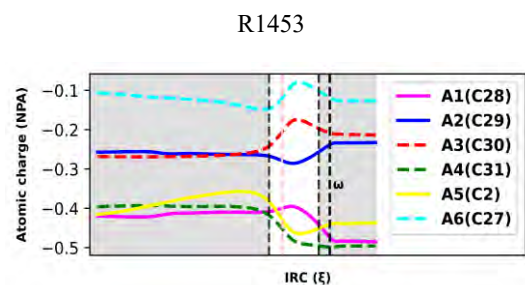
$S = 0.65 \text{ \AA}$, $S_{BO} = 2.62$, $CT = 0.18$, Flux = INVERSE



$S = 0.65 \text{ \AA}$, $S_{BO} = 2.62$, $CT = 0.18$, Flux =
INVERSE



$S = 0.76 \text{ \AA}$, $S_{BO} = 2.62$, $CT = 0.18$, Flux = INVERSE



$S = 0.76 \text{ \AA}$, $S_{BO} = 2.62$, $CT = 0.18$, Flux =
INVERSE

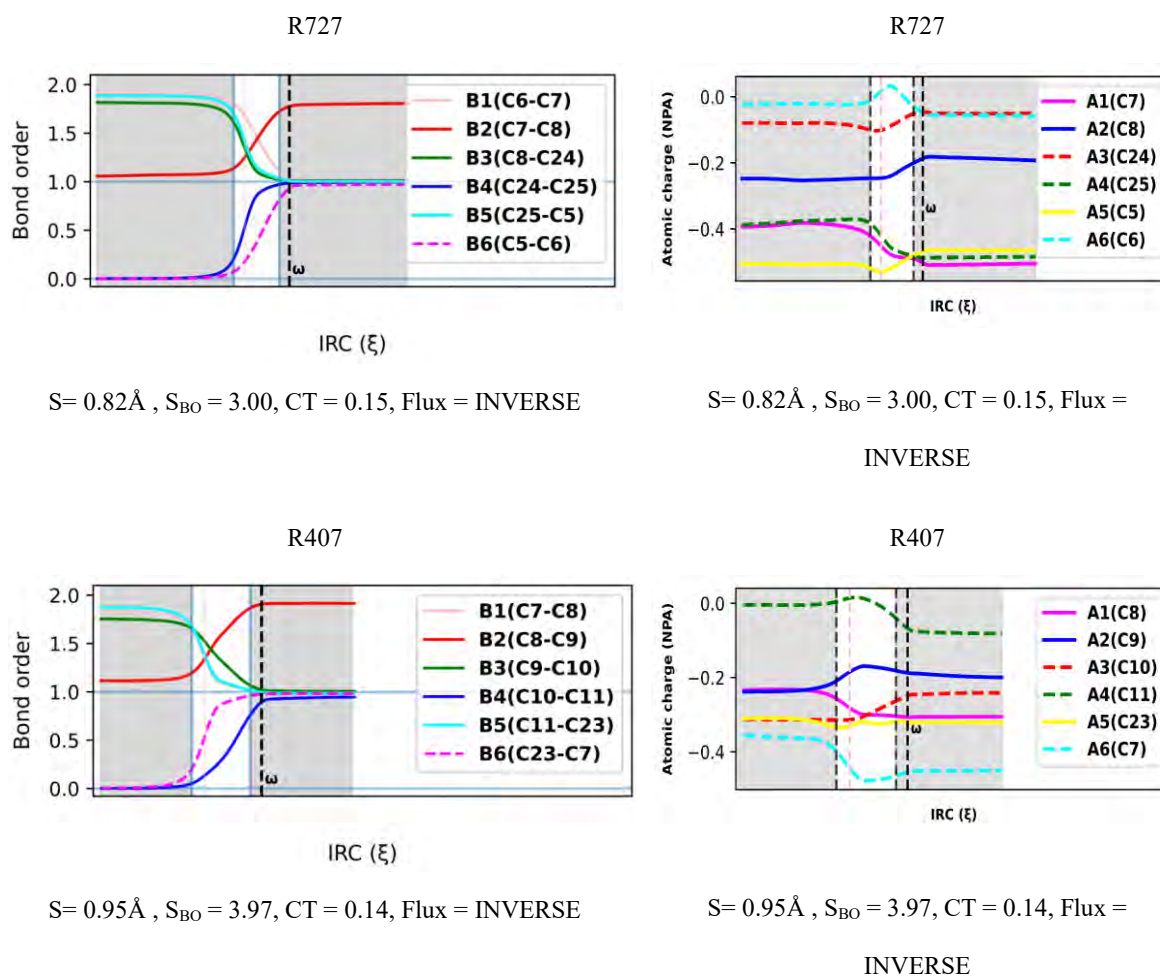


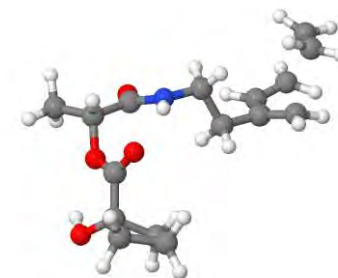
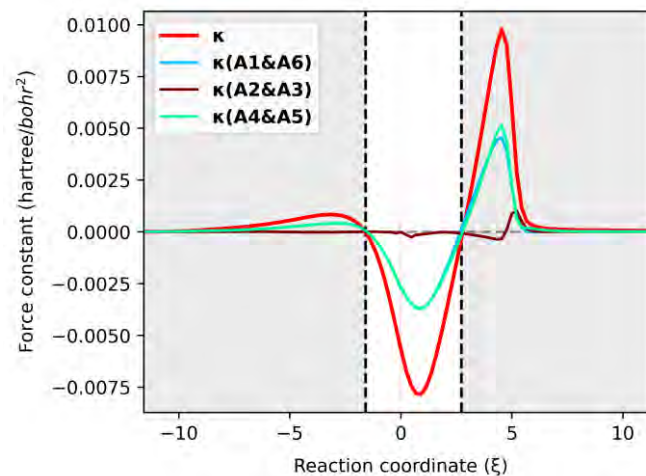
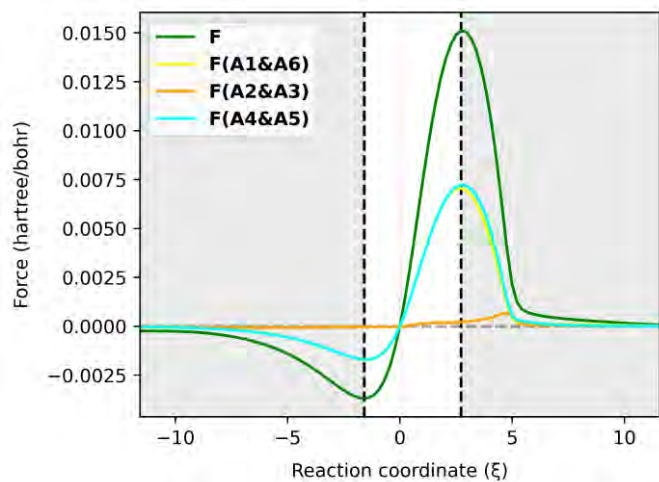
Figure 17. Bond order analysis of an illustrative set of ten Diels-Alder reactions covering a wide range of synchronicity degree (from $S = 0.00 \text{ \AA}$ to $S = 0.95 \text{ \AA}$). Wiberg bond orders of the reactive site bonds are plotted along the IRC path. The position of the global maximum of the reaction force constant is denoted by w . The synchronicity indices, the charge transfer at the TS and the sense of the electron density flux are given in their respective units.

The process of bond breaking/formation during DA reactions has also been assessed using other tools in previous studies. The bond evolution theory (BET) has probably been one of the most promising approaches for this purpose. It is a combination of the ELF topological analysis[269] and the catastrophe theory[270]. Berski et al applied the BET to the normal-electron demand DA reaction between 1,3-butadiene (BD) and acrolein (Acr), and the inverse electron-demand one between 2,4-pentadienal and methyl vinyl ether. These authors found that the entire process could be split into 11 or 10 catastrophe sequences respectively, including the

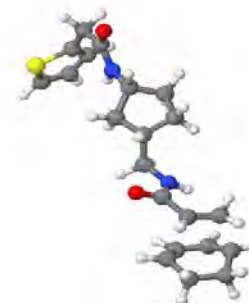
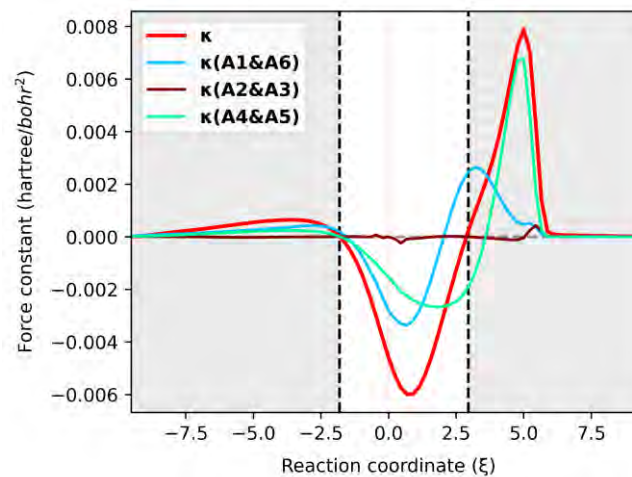
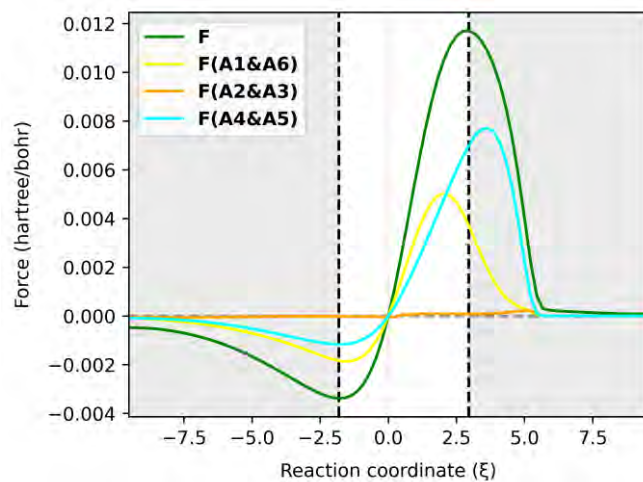
cuspid and fold types of catastrophes. Cusp catastrophes were associated with the reduction of the three C=C bonds into C-C bonds, as well as with the formation of the two new C-C bonds. Fold catastrophes resulted from the concentration of the electron density on terminal C atoms, which preceded the formation of new C-C bonds [251]. Moreover, using the ELF topological analysis, Domingo et al demonstrated that the formation of the new C-C bonds in polar DA reactions was induced by the merging of two monosynaptic basins (pseudo-radical) located on terminal C atoms into a unique bisynaptic basin[36].

6. Atomic decomposition of energy derivatives

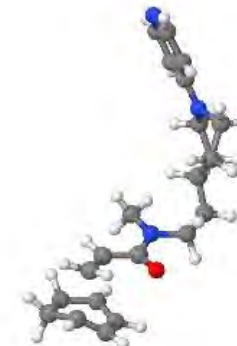
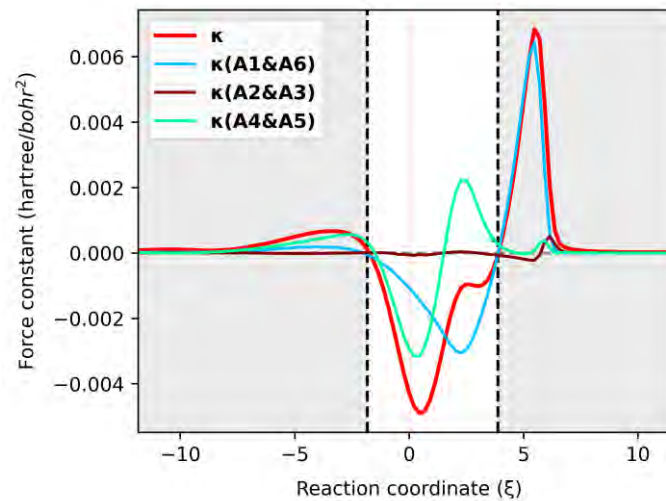
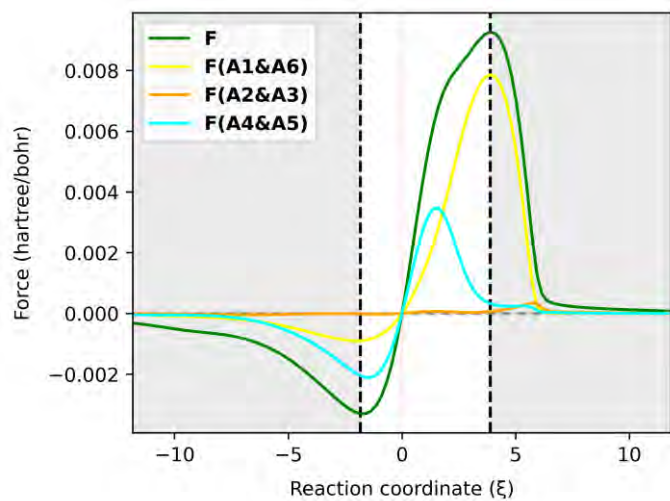
The reaction force F and reaction force constant κ were resolved in terms of atomic contributions using the approach proposed here[40] and implemented in our AMADAR program. For an illustrate set of six DA reactions covering a range of synchronicity from 0.02 to 1.02 Å, Figure 18 depicts these grouped contributions for the six atoms involved in the reactive site in terms of 3 pairs. It should be noted that qualitatively equivalent conclusions to those reported below were deduced by picking up any random set of other DA reactions.



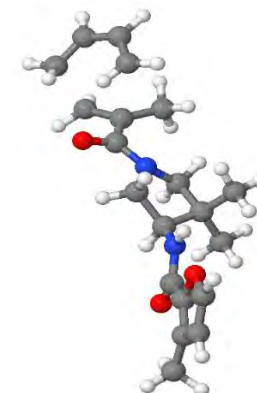
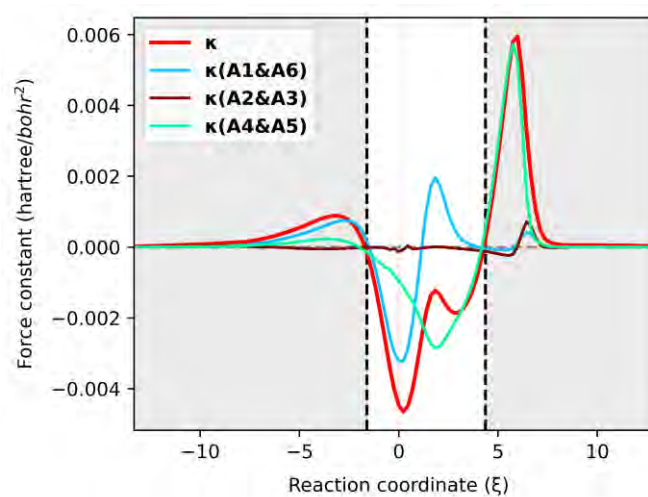
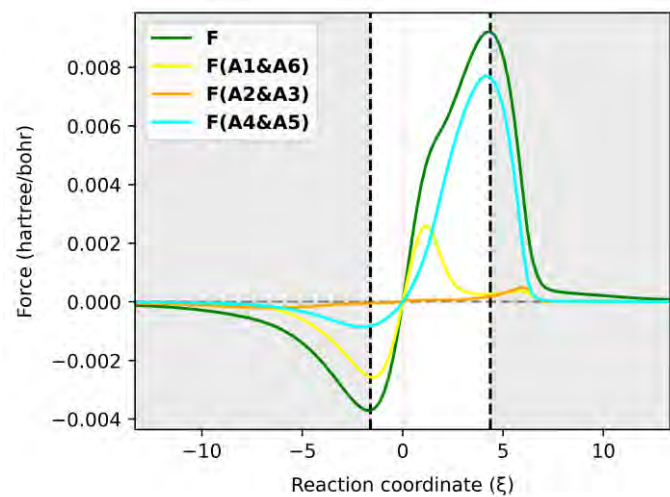
$S=0.02 \text{ \AA}$, $\kappa_{min}(\xi)=0.908$, **R1310**



$S=0.24 \text{ \AA}$, $\kappa_{min}(\xi)=0.682$, **R218**



$S=0.43 \text{ \AA}$, $\kappa_{min}(\xi)=0.456$, **R585**



$S=0.67 \text{ \AA}$, $\kappa_{min}(\xi)=0.230$, **R732**

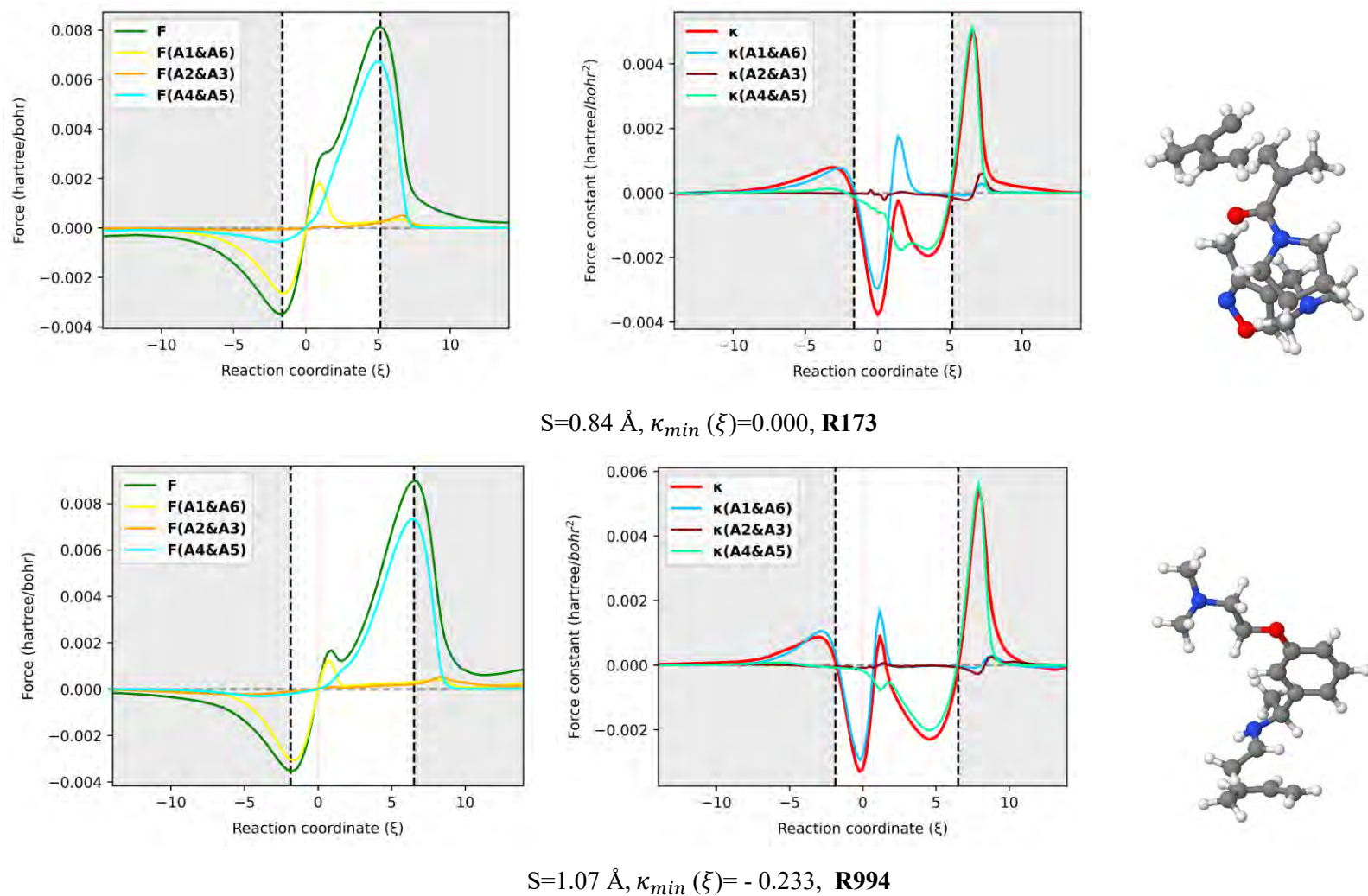


Figure 18. Atomic (fragment) resolution of the reaction force F and reaction force constant κ for both atoms in the reactive site and pair of atoms (A1&A6, A2&A3 and A4&A5) in the case of ten DA reactions with synchronicity indexes between 0.02 \AA to $S = 1.07 \text{ \AA}$

Let us first examine the atomic resolution of the reaction force F . Figure 18 shows that the terminal C atoms (A1, A4, A5, and A6) bring the most important contributions to F , while the two others (A2 and A3) can barely be noticed. This observation can be rationalized by the fact that A1, A4, A5, and A6 are directly involved in the interaction between the diene and dienophile. As such, they must experience the greatest electronic effects as the reaction proceeds, which result in considerable electrostatic interactions and Feynman forces[271]. Let us recall that the atomic contributions to the reaction force, noted $F_A(\xi)$ for a given atom A, is obtained as the scalar product of the Hellman-Feynman force F_A acting on the corresponding nucleus and the vectorial change of the nuclear position $\frac{\partial R_A}{\partial \xi}$. Thus, since all the atoms in the reactive site move considerably during the process, the only reason for the insignificant contributions of A2 and A3 must be related to small Hellman-Feynman forces acting on their nucleus.

The most striking feature from Figure 18 is that the overall profile of the reaction force F is dictated by two dominant contributions stemming from the two pairs of interacting atoms. This finding suggests that the mechanism of these DA reactions can be understood in terms of two inner elementary processes identified with the formation of the two C-C bonds. Each of the two processes has its proper critical points and it is only their combination that gives rise to the observed profile. One can follow these individual critical points to get insight into the formation of the two C-C bonds. In this optic, it is clear from Figure 18 that both processes have their global minima matching with that of the reaction force F , with a magnitude that depends on the degree of synchronicity of the DA reaction. For synchronous reactions, the two minima are situated about halfway of the reaction force over the preparation phase. This suggests that the two nascent C-C bonds oppose equally to the structural change taking place in the reacting system before entering the TS region. As the DA reaction gets more asynchronous, the main retarding force comes from the fast-forming bond, while the second bond brings the most driving force towards the formation of the product. Therefore, one mechanistic difference between synchronous and asynchronous DA reactions seems to be that, in asynchronous reactions, the two opposing effects that vanish at the classical TS are mainly caused by two separate elementary processes, while in the case of synchronous ones both elementary processes retard and drive the process concomitantly and equivalently. Nevertheless, it is worth noting that, for asynchronous DA reactions, the steep increase of F in the surrounding of the classical TS is still due to the fast-forming bond, whose positive effect is more pronounced in

that area than that of the second bond, before decaying quickly and leaving the control to the second bond that drives the process just after.

Turning now on the reactive force constant, Figure 18 confirms and even strengthens the results reported in section 5 based on the Wiberg bond order analysis. Similarly to κ_{max} , the position of the $\kappa_{A1,A6}$ and $\kappa_{A4,A5}$ global maxima should indicate when each of the C-C bonds is formed. Then, some striking features can be inferred from Figure 18. First, as the DA reaction becomes asynchronous, $\kappa_{A1,A6}$ maximum shifts to the left and its magnitude decreases, while $\kappa_{A4,A5}$ maximum tends to fit κ_{max} . The typical inflection point and local maximum on the κ profile in the TS region of (moderate) asynchronous processes correspond to the global maximum of $\kappa_{A1,A6}$ and indicate therefore the formation of the fast-forming bond. This finding is in line with the analysis of bond orders reported above, which suggested that the first bond is formed more and more around the TS when the reaction becomes asynchronous. Moreover, the position of κ_{min} with regards to the classical TS changes with the synchronicity of DA reactions. For synchronous reactions, κ_{min} is right-deviated as compared to the TS. Similar deviations were reported in the reaction between HF and CO, where κ_{min} was located at the left of TS ($\xi = -0.22$)[40]. This deviation was found to be beyond any kind of statistical error of the method and may suggest an effect of the nuclear interaction on κ_{ξ} .

It must be emphasized that the magnitude of the deviation between κ_{min} and TS decreases as the reaction becomes asynchronous. More interestingly, for likely two-step DA reactions ($S > 0.95\text{\AA}$), κ_{min} is located at the left of the classical TS such as in the case of **R994** ($S=1.07\text{\AA}$) where it is found at $\xi = -0.23$. For example, it is valued to $\xi = 0.90$ in **R1310**, $\xi = 0.682$ in **R218**, $\xi = 0.45$ in **R585**, $\xi = 0.23$ in **R732**, and $\xi = 0.00$ in **R173** whose synchronicity indexes are 0.02, 0.24, 0.43, 0.67 and 0.84 \AA respectively. Therefore, the relative position of κ_{min} with respect to the TS seems to be both an indicator of the degree of synchronicity in DA reactions and a potential descriptor of the transition between likely concerted and stepwise mechanisms. However, there is not sufficient evidence to prove the last assumption at present. Anyway, in contrast with the profile of κ , the deviation between κ_{min} and the TS can be quantified. As such, it becomes possible to compare two reactions from the same synchronicity category, something which is very hard to do by only looking at the profile of their reaction force constant κ .

7. Concluding remarks

In this chapter, we have tried to address the perpetual question of (a)synchronicity in DA reactions using the reaction force theory and the atomic resolution of energy derivatives. It has been found that the transition between preferentially concerted to stepwise processes can be located in the range of synchronicity indexes from 0.90 to 1.00Å. Furthermore, the global maximum of the reaction force constant may indicate the formation of the slow-forming C-C bond. The atomic resolution of energy derivatives suggested that the mechanism of the DA reaction involves two inner elementary processes associated with the formation of the C-C bonds. The position of the global minimum of the reaction force constant with respect to the TS is shown to correlate with the synchronicity of DA reactions. Further studies are being conducted to understand common features in the reaction electron flux along the IRC path of this huge dataset. This work has produced another research paper, which will soon be submitted for publication.

“The greatest obstacle to discovery is not ignorance – it is the illusion of knowledge”

Daniel J. Boorstin

Chapter 4: The Diels-Alder reaction of C₆₀ fullerene

0. Summary

The Diels-Alder reaction between C_{60} and a series of dienes has been studied using the ONIOM approach. In all, 22 ONIOM models were constructed and their accuracy discussed. We have demonstrated, by comparing several ONIOM2(DFT/6-31G(d): SVWN/STO-3G) models, the importance of dispersion corrections in the accurate estimation of reaction and activation energies. Our findings suggest that the ONIOM2(M06-2X/6-31G(d): SVWN/STO-3G) model, where the higher layer encompasses the diene and pyracyclene portion of C_{60} , provides the best compromise in the estimation of both the activation and reaction enthalpies, showing errors lower than 2.6 and 2.0 kcal/mol respectively. The effect of substituents (-CN and -OH) on the activation energy of C_{60} DA reactions is partially explained in the framework of the frontier molecular orbital theory and activation strain model.

1. Introduction

As mentioned previously, C_{60} can undergo Diels-Alder (DA) reactions with a variety of dienes. This transformation leads to the formation of diverse functionalized C_{60} fullerenes, some of which having found applications. Normally, as it is the case for small to mid-size systems, one should gain invaluable insight into the mechanism of C_{60} DA reactions using high level computational methods. However, the big size of the C_{60} dienophile restricts the usage of quantum mechanics methods, and so does it for any attempt of extending the investigation to many C_{60} DA reactions.

In this chapter, we report on the usage of the ONIOM approach for the investigation of DA reactions involving [60] fullerene as the dienophile. This approach is found to reduce considerably the computational cost associated with the prediction of DA transition states involving the C_{60} Buckminster fullerene. It also allows for the expansion of the study on several other systems. In short, we have discussed the accuracy and cost of 22 ONIOM models with respect to the B3LYP/6-31G(d) level, as well as the ability of different DFT functionals to reproduce the experimental data of the DA reaction between C_{60} and cyclopentadiene. The best performing DFT functional (M06-2X) and ONIOM partition are used to investigate the DA reactions between C_{60} and 21 dienes. The results are analyzed in terms of reaction activation energy and its strain and interaction components, frontier molecular orbitals of the diene, $HOMO_{diene}-LUMO_{C60}$ gaps, polarity and synchronicity of the TSs. We also insist of the substituent effect on the symmetry and polarity of the TS.

2. Methodology

The present study was carried out in two main steps. First, the DA reaction between C_{60} and cyclopentadiene (CPD), whose experimental data are available, was considered to assess the accuracy (and performance) of different ONIOM models in reproducing B3LYP/6-31G(d) energetic data. The choice of B3LYP/6-31G(d) level as the reference was driven by the fact that it has been the most used in the investigation of DA reactions [114] and it is also known to provide good quality TS geometries [17][34]. Figure 19 shows selected geometrical parameters of the TS geometry for the DA reaction between C_{60} and CPD. These parameters were evaluated for each ONIOM model in order to discuss their ability to reproduce the TS structure as compared to the B3LYP/6-31G(d) reference. Then, having identified the best model, the latter was employed to test the accuracy of DFT functional with respect to the

experimental data. For this purpose, a single-point frequency calculation was performed on the optimized stationary points obtained using the best ONIOM model, where the higher-layer is described by various DFT functional.

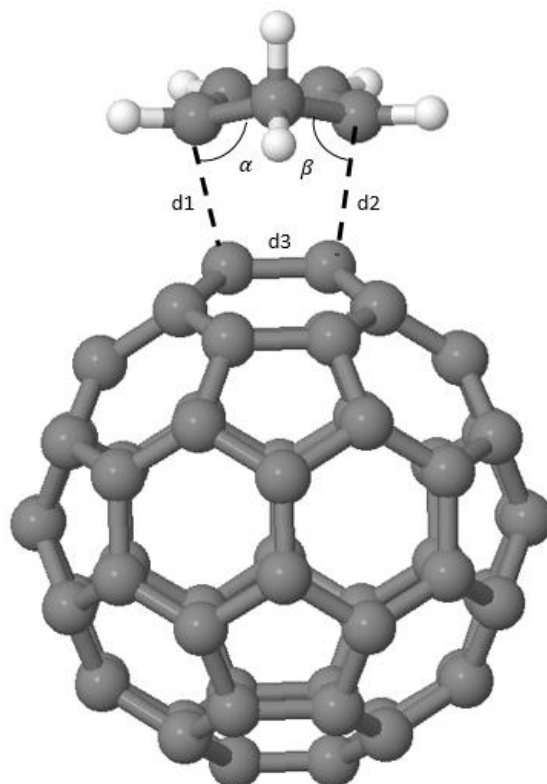


Figure 19. Geometrical parameters of interest for the description of the TS of the DA reaction between C₆₀ and CPD

Further, 21 dienes were considered to get insight into the DA reaction of C₆₀ fullerene using the best level of theory (DFT functional and ONIOM partition) in terms of accuracy and computational cost among all those considered in this study. For each reaction, structural, electronic and energetic data were computed, discussed and the results obtained are here reported.

We estimated the polarity of TSs as the charge transfer (noted q) between the two reactants at the TS. To obtain this charge transfer, a single-point calculation coupled with a Natural Population Analysis (NPA) was carried out on the optimized geometry of the TS, using the best functional to describe the entire system. Moreover, the propensity of each diene to retain its electron density was evaluated using the electrophilicity index proposed by Parr and coworkers [Ref].

M. Bickelhaupt and coworkers introduced the activation strain model, which suggests that the activation energy of a given chemical reaction can be decomposed into the strain ($\Delta E_s^\#$) and interaction ($\Delta E_i^\#$) components. These values were examined to understand grossly the origin of the activation energy of each reaction. Note that the interaction energy was estimated using the following equation, where $E^\#$ (TS), $E^\#$ (C₆₀) and $E^\#$ (diene) are respectively the total electronic energy of the TS, and that of C₆₀ and the diene in the distorted geometry in the TS. The strain energy ($\Delta E_s^\#$) was deduced as the difference between the activation energy and the interaction energy. All of these calculations were performed using the Gaussian 09 package[252].

$$(\Delta E_i^\#) = E^\# (\text{TS}) - [E^\# (\text{C}_{60}) + E^\# (\text{diene})] \quad (12)$$

$$(\Delta E_s^\#) = (\Delta E^\#) - (\Delta E_i^\#) = \Delta E_s^\#(\text{diene}) + \Delta E_s^\# (\text{C}_{60}) \quad (13)$$

Note that $\Delta E_s^\#(\text{diene})$ and $\Delta E_s^\#(\text{C}_{60})$ are computed as the difference in electronic energy between the distorted geometry of the reactant at the TS and the minimum one.

3. The ONIOM models

Regarding the size of the C₆₀ buckminsterfullerene and the inherent computational cost associated with the location of TSs, it looked compulsory to find first the appropriate method that gives the best threshold between accuracy and performance. For this purpose, 22 ONIOM models (12 ONIOM2 and 10 ONIOM3 models) were built by combining five semi-empirical methods (PM6, PDDG, AM1, PM3, PM3MM) and the LSDA(SVWN)/STO-3G level in conjunction with the B3LYP/6-31G(d) level (Figure 20). This calibration study was restricted to the reaction between C₆₀ and cyclopentadiene (CPD) because its experimental data are available. In fact, Pang and Wilson [99] reported an activation energy of 6.9 kcal/mol for this cycloaddition reaction at 293 K. Giovani et al found a value of 26.7 ± 2.2 kcal/mol for the retro-DA transformation of the C₆₀-CPD cycloadduct. By combining these two values, one can estimate the reaction energy at 19.8 ± 2.2 kcal/mol. These data are used later on to compare the accuracy of several DFT functionals in predicting activation and reaction energies of DA reaction of C₆₀.

In ONIOM2 models, two different partitions were considered, where either of the naphthalene or pyracyclene fragment of C₆₀ was involved in the high-layer with the cyclopentadiene molecule. ONIOM3 models have three partitions: the higher, medium and lower layers. The higher layer was described with the B3LYP/6-31G(d) level, the medium layer with the

SVWN/STO-3G and the lower-layer with each of the five semi-empirical methods. The higher-layer includes either the naphthalene or pyracyclene moieties of C₆₀, while the medium one involves respectively the pyracyclene or a cupped buckybowl fragment. It must be noted that we could not restrict the medium partition to the buckybowl shape since this was found to violate ONIOM partition rules by keeping two atoms in the lower-layer with no connection with the medium one. The results obtained using these models were compared to those predicted by the B3LYP/6-31G(d) level applied on the entire system.

The choice of these partition schemes was mainly guided by previous findings, especially the study by Osuna et al on the performance of the ONIOM method for the study of chemical reactivity in fullerenes [116]. These authors modelled the DA reaction between C₆₀ and CPD by constructing 8 ONIOM2 models, where AM1 semi-empirical method and the LSDA(SVWN)/STO-3G were combined with the B3LYP/6-31G(d) level. Four different partitions were established, including the [6,6] double bond, the naphthalene, pyracyclene, or buckybowl fragments of C₆₀ respectively in the higher-layer alongside the CPD. The lower layer was described using either AM1 or SVWN/STO-3G.

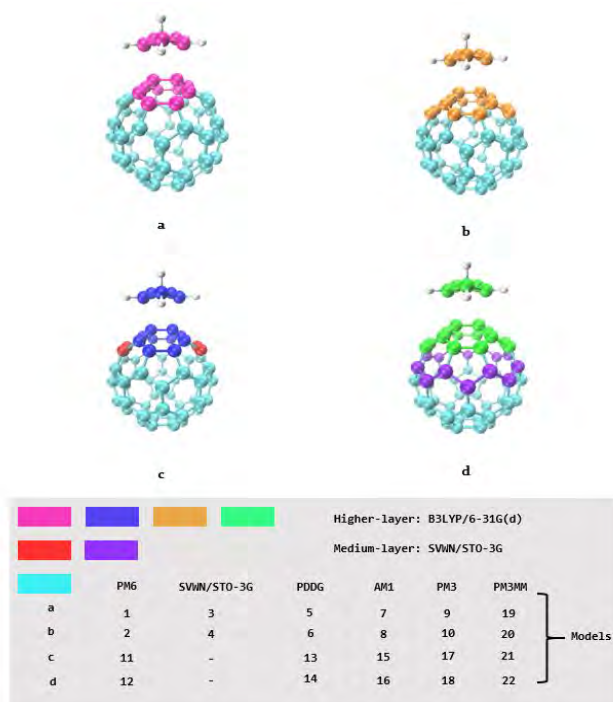


Figure 20. ONIOM models (1-22) considered to describe the C₆₀-CPD system. Given a model number (1-22), the low layer for ONIOM may be determined from the individual column headings (PM6, SVWN/STO-3G, PDDG, AM1, PM3, PM3MM). The specific decomposition (a, b, c, d) used for the model is read from the row titles. From this the colours on the structures may be used to determine the full ONIOM model used.

4. Accuracy and performance of ONIOM models

For each of the 22 ONIOM models, we optimized the corresponding reactants, product and TS towards their appropriate stationary points. All the minima were successfully located as they showed no imaginary frequency. We were unable to locate the TSs using models **5**, **7**, and **12**. In the case of models **12 (5)**, none of the convergence criteria (only one convergence criterion) was met and the wave function seemed to oscillate after long running times. These models were then discarded. In line with Osuna et al [115], two of the four convergence criteria (the maximum force and its RMS) were met when applying model **7**. Therefore, the final structure was assumed to be a fairly good estimate for the actual TS.

Table 4 collects the activation energy and enthalpies (ΔE^\ddagger and ΔH^\ddagger in kcal/mol), reaction energies and enthalpies (ΔE and ΔH in kcal/mol) and HOMO-LUMO gaps of C_{60} computed with the different models and at the B3LYP/6-31G(d) level. Deviation from the reference and the time elapsed to locate the TS (CC in minutes) are also included in the Table. Table 4 indicates that the computational cost is significantly reduced when ONIOM models are used in place of the B3LYP/6-31G(d) level. With the ONIOM models the location of the TS using one CPU takes between 20 and 140 minutes instead of more than 5200 minutes at the B3LYP/6-31G(d) level. In other words, the convergence towards the TS is 40 to 260 faster when ONIOM models used employed. However, being faster does not mean that the results are accurate.

Looking at the activation and reaction energies in Table 4, it appears that ONIOM2 models perform better than their ONIOM3 analogs. In addition, partitions involving the pyracyclene and cupped buckybowl fragments respectively in the higher and medium levels of ONIOM3 models tend to overestimate the activation energy and underestimate the reaction energy. Four models perform well with respect to the reference, namely models **4**, **7**, **9** and **19**. Particularly, the model **4**, i.e. ONIOM2(B3LYP/6-31G(d): SVWN/STO-3G), yields the best results with deviations of 0.4 and 0.6 kcal/mol respectively for the activation and reaction energy. Similar deviations are observed in terms of activation and reaction enthalpies, estimated at 0.6 and 0.7 kcal/mol. Models **7**, **9** and **19** also give good results with deviation lower than 1 kcal/mol. However, only model **4** seems to reproduce fairly well the HOMO-LUMO gap of the C_{60} fullerene, with the smallest deviation of -0.17 kcal/mol. It is worth noting that the location of the TS using the ONIOM2(B3LYP/6-31G(d): SVWN/STO-3G) model (**4**) was 37 times faster than using uniquely the B3LYP/6-31G(d) level.

Table 1. Performance of the ONIOM models with respect to the B3LYP/6-31G(d) level

Models	ΔE^\ddagger	ΔE	$\Delta\Delta E^\ddagger$	$\Delta\Delta E$	HL	ΔHL	ΔH^\ddagger	ΔH	ΔH^\ddagger	$\Delta\Delta H$	CC
Model 1	19.6	-4.2	2.2	5.1	7.40	4.64	19.7	-2.3	1.8	4.3	48
Model 2	18.5	-9.9	1.1	-0.6	6.41	3.65	18.7	-8.1	0.8	-1.5	23
Model 3	16.0	-12.4	-1.4	-3.1	2.59	-0.17	16.6	-9.5	-1.3	-2.9	94
Model 4	17.8	-8.7	0.4	0.6	2.59	-0.17	18.5	-5.9	0.6	0.7	137
Model 5	-	-	-	-	-	-	-	-	-	-	-
Model 6	17.1	-13.0	-0.3	-3.7	7.26	4.50	17.3	-10.9	-0.6	-4.3	68
Model 7	16.9	-9.3	-0.5	0.0	7.30	4.54	17.1	-7.4	-0.8	-0.8	52
Model 8	16.8	-12.3	-0.6	-3.0	7.30	4.54	17.1	-10.3	-0.8	-3.7	70
Model 9	18.3	-8.4	0.9	0.9	7.27	4.51	18.7	-6.4	0.8	0.2	35
Model 10	18.0	-11.9	0.6	-2.6	7.27	4.51	18.3	-9.8	0.4	-3.2	68
Model 11	16.9	-13.7	-0.5	-4.4	7.40	4.64	17.1	-11.9	-0.8	-5.3	59

Model 12	-	-	-	-	-	-	-	-	-	-	-
Model 13	15.6	-16.8	-1.8	-7.5	7.26	4.50	15.8	-14.8	-2.1	-8.2	59
Model 14	20.3	-4.0	2.9	5.3	7.26	4.50	19.2	-3.6	1.3	3.0	137
Model 15	15.3	-16.1	-2.1	-6.8	7.30	4.54	15.6	-14.1	-2.3	-7.5	-
Model 16	19.9	-4.4	2.5	4.9	7.30	4.54	14.5	-8.6	-3.4	-2.0	84
Model 17	16.4	-15.9	-1.0	-6.6	7.26	4.50	16.8	-13.8	-1.1	-7.2	42
Model 18	20.5	-3.8	3.1	5.5	7.26	4.50	19.6	-3.3	1.7	3.3	134
Model 19	18.3	-8.4	0.9	0.9	7.26	4.50	18.7	-6.4	0.8	0.2	30
Model 20	18.0	-11.9	0.6	-2.6	6.24	3.48	18.3	-9.8	0.4	-3.2	63
Model 21	16.4	-15.9	-1.0	-6.6	7.26	4.50	16.8	-13.8	-1.1	-7.2	37
Model 22	20.5	-3.8	3.1	5.5	7.26	4.50	19.6	-3.3	1.7	3.3	85
B3LYP	17.4	-9.3	0.0	0.0	2.76	0.00	17.9	-6.6	0.0	0.0	5216

Table 5 presents selected structural features of the TS predicted using the different models and the B3LYP/6-31G(d) reference level with no symmetry constraint. Owing to the symmetry of C₆₀ and CPD, we expected a synchronous TS for the DA reaction between C₆₀ and CPD. This assumption was confirmed at the B3LYP/6-31G(d) level, which predicted the two C-C bonds to be 2.203 Å long at the TS. This was also qualitatively confirmed for about half of the ONIOM models, which predicted synchronous TSs, even though they did not reproduce the individual C-C bond lengths. ONIOM models involving the PM6 method gave rise to the most distorted TSs, with C-C bonds length differences of 0.689 and 0.341 Å respectively for models **1** and **2**.

To understand the origin of these large deviations, we tried to locate the same TS using the PM6 method on the entire system without any restriction on the symmetry. This calculation revealed a very asynchronous TS, with C-C bonds of 1.862 and 3.020 Å, in clear contradiction with the B3LYP level. Therefore, one can confirm that the asymmetry observed in the TS predicted with the ONIOM2(B3LYP/6-31G(d): PM6) models is caused uniquely by the PM6 method, which is apparently unable to describe correctly the high curvature of the C₆₀ molecule. Moreover, the TS geometry predicted using the PM6 level does not reflect a concerted mechanistic pathway. This hypothesis is inferred from the visual analysis of the imaginary vibrational mode. Normally, it should be moving both pairs, instead it displaces only one of the two pair of interacting atoms and keep the other one intact. This suggests the likeliness of seeing an intermediate being formed upon formation of the C_{CPD}–C_{C60} bond lying in the axis of the vibrational mode.

Table 5 shows that model **4** gives the most similar TS to the B3LYP/6-31G(d) level, with C-C bonds long of 2.197 Å and identical α and β angles of 90.3°. This finding supports the energetic data reported in Table 4, which suggested this model to present the smallest deviation with respect to the reference level. Although models **7**, **9** and **19** showed fairly good energetic data and less costly convergence towards the corresponding TSs, they fail to predict the expected synchronous TSs. Therefore, the ONIOM2 (B3LYP/6-31G(d): SVWN/STO-3G) model **4** appears to be the most accurate and time-saving level for the study of Diels-Alder reactions involving C₆₀ fullerene as compared to B3LYP/6-31G(d).

However, in line with the literature[115], the B3LYP functional, either full or in model **4**, does not lead to accurate activation and reaction enthalpies in comparison with the experimental values reported by Pang et Wilson[98], and Giovani et al[271]. A more accurate model was then needed to remediate this problem.

Table 5. Comparative structural description of transition states using different ONIOM models and the B3LYP/6-31G(d) level.

	d1	d2	S	d3	α	β
B3LYP	2.203	2.203	0.000	1.460	90.2	90.2
Model1	1.984	2.673	0.689	1.481	98.6	80.0
Model2	2.437	2.096	0.341	1.465	85.0	93.9
Model3	2.219	2.219	0.000	1.455	90.0	90.0
Model4	2.197	2.197	0.000	1.462	90.3	90.3
Model5	-	-	-	-	-	-
Model6	2.266	2.266	0.000	1.462	89.4	89.4
Model7	2.092	2.502	0.410	1.474	94.8	83.7
Model8	2.265	2.265	0.000	1.462	89.4	89.4
Model9	2.163	2.356	0.193	1.480	92.2	87.0
Model 10	2.259	2.259	0.000	1.461	89.5	89.5
Model 11	2.084	2.535	0.451	1.456	95.0	83.1
Model 12	-	-	-	-	-	-
Model 13	2.289	2.289	0.000	1.453	89.2	89.2
Model 14	2.165	2.166	0.001	1.473	90.8	90.8
Model 15	2.301	2.273	0.028	1.453	88.8	89.6
Model 16	2.168	2.168	0.000	1.472	90.8	90.8
Model 17	2.281	2.281	0.000	1.452	89.2	89.2
Model 18	2.163	2.163	0.000	1.473	90.0	90.9
Model 19	2.163	2.356	0.193	1.480	92.2	87.0
Model 20	2.259	2.259	0.000	1.461	89.5	89.5
Model 21	2.281	2.281	0.000	1.452	89.2	89.2
Model 22	2.163	2.163	0.000	1.473	90.9	90.9

Therefore, single-point frequency calculations were performed from the geometry obtained with the best ONIOM2(DFT/6-31G(d): SVWN/STO-3G) model **4**, but this time using each of the following DFT functionals to describe the higher-layer: M06-2X, CAM-B3LYP, TPSSH, PBE and ω B97XD. The results are shown in Table 6.

Table 6. Comparison between different DFT functionals. Values in parentheses are corrected with the ZPE. All the values are given in kcal/mol.

Method	ΔH^\ddagger	ΔH	$\Delta\Delta H^\ddagger$	$\Delta\Delta H$
experimental	6.9	-19.8	0.0	0.0
B3LYP	18.5 (17.8)	-5.5 (-2.1)	11.6 (10.9)	14.3 (17.7)
CAM-B3LYP	18.2 (19.2)	-13.7 (-9.9)	11.3 (12.3)	6.1 (9.9)
PBE	7.7 (8.3)	-14.6 (-11.3)	0.8 (1.4)	5.2 (8.5)
TPSSH	11.0 (8.9)	-10.2 (-9.7)	4.1 (2.0)	9.6 (10.1)
M06-2X	8.6 (9.5)	-21.8 (-18.3)	1.7 (2.6)	-2.0 (1.5)
wB97XD	9.5 (10.4)	-24.2 (-20.3)	2.6 (3.5)	-4.4 (3.5)

The reaction and activation enthalpies (obtained using different functionals that describe the higher-layer of the ONIOM2(DFT/6-31G(d): SVWN/STO-3G) model 4) were calculated with and then without correction from the Zero-Point-Energy (ZPE). Table 6 reveals that the B3LYP functional gives the worst results among all the DFT functionals. Furthermore, in contrast with the known tendency of the B3LYP functional to underestimate activation energies [272], our results suggest instead an overestimation of the activation energy for the DA reaction between C₆₀ and CPD, with an error between 10.9 and 11.6 kcal/mol with respect to the experimental data. Osuna et al suggested that this discrepancy could be attributed to the incapacity of the functional to describe the steep curvature of the C₆₀ molecule [115]. Poater et al showed that activation energies computed using the B3LYP functional could be improved by increased the value of the a_0 parameter appearing in the expression of the exchange energy from 0.1 to 0.9 [273]. However, we did not attempt that here because it was not within the scope of this study. Moreover, the range-separated CAM-B3LYP functional seems to improve slightly over B3LYP the estimation of the reaction enthalpy showing an error of 6 to 10 kcal/mol. However, it performs worse for the prediction of activation energies. One way to address this issue could be by adjusting the separation distance in the damping function used to shift between short and

long range interactions in the calculation of the exchange energies. However, such an attempt is beyond the scope of this work.

The GGA functional of Perdew-Burke-Ernzerhof (PBE) performs better than B3LYP for the estimation of both the reaction and activation enthalpies, as it predicts more accurate values as compared to experimental data. While it gives the most accurate activation enthalpy (7.7 to 8.3 kcal/mol) with an error lower than 1.5 kcal/mol, its evaluation of the reaction enthalpy is still good enough showing deviation of the order of 5.2 to 8.5 kcal/mol.

The hybrid meta-GGA M06-2X and dispersion corrected ω B97XD functionals give the best compromise between activation and reaction enthalpies. The former presents errors lower than 2.6 and 2.0 kcal/mol for the estimation of the activation and reaction enthalpy respectively, while the latter shows slightly higher errors lower than 4.5 and 3.5 kcal/mol. The common factor between these functionals is that they both include dispersion corrections in the estimation of the total electronic energy. In the ω B97XD functional, dispersion effects are computed empirically using the Grimme's function [208], while in M06-2X dispersion effects are accounted for both by compensation through the double (54%) exchange HF energy and with a Grimme's function [274]. This finding highlights the prime role played by dispersion effects in the description of the DA reaction of C₆₀ fullerene in general. Osuna et al came to the same conclusion in one of their previous studies [117].

5. C₆₀ DA reaction energetics and structural data

Figure 21 displays the optimized geometries of all the dienes considered as model counterparts to get insight into the DA reaction of the C₆₀ fullerene. These include anthracene, 1,3-butadiene, cyclopentadiene, thiophene, oxothiophene, selenothiophene pyrrole, and their mono- cyano and hydroxyl derivatives. Note that the substituent is fixed at one of the two indistinguishable terminal C atoms of the diene, which is engaged into a one-to-one interaction with one [6,6] C atom of C₆₀. Table 8 collects the reaction energy and free energy (ΔE , ΔG in kcal/mol), activation energy and enthalpy (ΔE^\ddagger in kcal/mol), interaction (ΔE_i^\ddagger) and strain (ΔE_s^\ddagger) energies of the system at the TS, the diene's (ΔE_{sd}^\ddagger in kcal/mol) and C₆₀'s (ΔE_{sC60}^\ddagger in kcal/mol) contributions to the strain energy as well as the percentage of interaction energy in the total activation energy (I_{TS} in %). The synchronicity index (S in Å), TS polarity (q in e), the electrophilicity index (ω in eV) and HOMO_{diene}-LUMO_{C60} gap (HL in eV) of the diene are given in Table 8. Reactions are numbered following the corresponding dienes, i.e. reaction 1 for example is the DA reaction between C₆₀ and the diene **1** as shown in Figure 21.

To begin with, it is worth noting that (almost) all the expected stationary points were successfully located. For example, TSs were obtained with the correct unique imaginary frequency, associated to the vibrational mode moving the two pairs of interacting atoms forward and backward. Reactants and the cycloadducts were real minima on the PES, even though for 1,3-butadiene (**2**) and its cyano (**9**) and hydroxyl (**16**) derivatives, the gauche conformers were returned upon optimization. Baraban et al recently reported the experimental evidence supporting the existence of a stable gauche conformer for 1,3-butadiene with a torsional angle of 34° [275]. However, 1,3-butadiene is known to have a trans global minima and the isomerization pathway between the trans and gauche conformers has already been studied [276]. Guided by these previous findings, the total electronic energy of the trans conformer of the dienes **2**, **9** and **16** was computed at the same level of theory. This led to the trans global minima, and allowed for the estimation of the conformational barriers at the M06-2X/6-31G(d), which were found to be 2.5, 2.4, 2.0 and kcal/mol for **2**, **9** and **16** respectively. The first value is close to the experimental rotational barrier of 2.8 kcal/mol [277].

Results in Table 7 suggest that 1-hydroxycyclopentadiene (diene **17**) and 1-cyanothiophene (CTP, diene **12**) respectively react the fastest and slowest with C_{60} , showing activation energies of 2.7 kcal/mol and 24.5 kcal/mol. This observation is in line with the strain and interaction energy components of the activation energy. These values indicate that the attractive interaction between 1-hydroxycyclopentadiene and C_{60} , estimated to -17.0 kcal/mol, almost counteracts the repulsive strain energy of 19.7 kcal/mol needed to distort the reactants up to the TS. On the contrary, the system C_{60} – CTP requires more energy (36.9 kcal/mol) to reach the transition state, and this results in a higher activation barrier.

The Bell-Evans-Polanyi principle states that, for a series of similar reactions, the activation energies and reaction enthalpies should be linearly correlated [278]. However, in the present case, no significant correlation was observed between the activation and reaction energies for the entire dataset. This indicates that the DA reactions of C_{60} fullerene [HOMO-LUMO gap = 2.16 eV at the ONIOM2(M06-2X/6-31G(d): SVWN/STO-3G) level] do not respond to the Bell-Evans-Polanyi (BEP) principle ($R^2=0.49$, See Appendix 3). This finding is consistent with Pla et al [279], who considered a series of fullerenes (C_{68} , C_{70} , C_{76} , C_{78} , C_{80} and C_{84}) against 1,3-butadiene, and found that only fullerenes with large HOMO-LUMO gaps such as C_{68} (4.0 eV, $R^2=0.98$) tend to obey the BEP principle, while those with small gaps did not show significant correlations (example C_{70} , HOMO-LUMO gap = 2.22 eV and $R^2=0.35$) at the M06-2X/6-31G(d,p) level. According to these authors, the low correlation observed for some

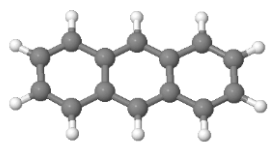
fullerenes can be explained by the change in the electronic structure of the corresponding cycloadducts due to the mixing of the ground state with excited states close in energy. Many other physico-chemical transformations do not satisfy the BEP principle. These comprise for example some enzymatic reactions manifesting what is known as negative catalysis[280] as well as oxide ion migration processes[280].

In general, electron-withdrawing groups (here -CN) increase the activation energy (and enthalpy), while electron-donating ones (here -OH) results in lower activation barriers. For example, the DA reaction between C₆₀ and furan (**Fur**, diene **5**) is predicted with an activation enthalpy of 22.0 kcal/mol, which turns out to be higher by 24.8 kcal/mol and lower by 17.4 kcal/mol respectively for the cyano (diene **12**) and hydroxyl (diene **19**) derivatives. This observation can be first rationalized by looking at the HOMO orbital of the diene. Table 5 shows that the cyano group, which depletes the electron density of the diene, decreases the energy of the HOMO and then enhances the HOMO_{diene}-LUMO_{C60} gap. Therefore, the insertion of the CN group weakens the interaction between the two antagonists because the higher the HOMO_{diene}-LUMO_{C60} gap, the lower the overlap between atomic orbitals of terminal C. The opposite effect is observed for the hydroxyl group. For example, the HOMO_{diene}-LUMO_{C60} gap is evaluated to 5.46 eV for furan, and 6.05 and 4.77 eV respectively for its cyano and hydroxyl derivatives. These values are consistent with the previous interaction energies.

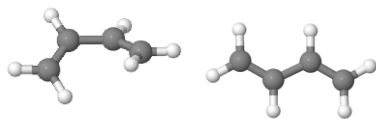
Table 8 reveals that the insertion of a -OH (-CN) group in a given diene increases (diminishes) the polarity of the TS as a consequence of the lowering of (increase of) the electrophilicity of the substituted diene with respect to the parent one. For example, the computed electrophilicity index ω of diene **6** (selenothiophene = STP) at the M062X/6-31G(d) level is 0.74 eV, while that of its -CN (diene **13**) and -OH (diene **20**) derivatives are 1.36 and 0.56 eV respectively. These values suggest that diene **13** tends to retain its electron density more than diene **6** and **20** respectively. This is asserted by the electron density transferred to C₆₀ at the TS (polarity), which are predicted to be 0.217, 0.116 and 0.307 e respectively for dienes **6**, **13** and **20** at the ONIOM2(M06-2X/6-31G(d): SVWN/STO-3G) level. Domingo et al studied the DA reaction between CPD and a series of cyano-ethylenes, and found that the polarity of the TS was strongly correlated with the activation energy[280]. Their finding also applies here and can also serve to understand the substituent effect in our case. Indeed, looking at Table 5, one can notice that, for a given diene (**1-7**), the more polar (less electrophilic) the TS (the diene) is, the lower the activation energy will be. As such, it is clear that by decreasing the TS polarity,

electron withdrawing groups such as $-\text{CN}$ should give rise to a higher activation energy as compared to the unsubstituted diene, while the opposite is true for the $-\text{OH}$ group.

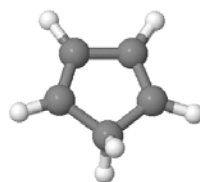
In the previous paragraphs, the effect of the $-\text{CN}$ and $-\text{OH}$ substituents on the activation energy was explained based on the analysis of frontier molecular orbitals, polarity of the TS and electrophilicity of the dienes. Another equivalent picture is suggested by the interaction and strain energies computed at the TS. In agreement with all the previous explanations, Table 7 confirms that the interaction between the dienes and C_{60} tends to be more attractive (repulsive) when an H atom of the terminal C of the diene is substituted with the $-\text{OH}$ ($-\text{CN}$) group. In addition, the strain or distortion energy seems to play a prominent role. Because, not only the interaction between the two antagonists is affected, but it is also clear from Table 7 that inserting the $-\text{OH}$ ($-\text{CN}$) group tends to decrease (increase) the amount of energy necessary to distort the reactants up to their geometry at the TS. Therefore, the overall diminution (increase) of the activation energy upon the insertion of the $-\text{OH}$ ($-\text{CN}$) group should be understood as resulting from the interplay between two opposite effects, the distortion of and interaction between the diene and C_{60} . Once again, this rule is restricted to a given diene and its derivatives, and does not hold on the entire dataset.



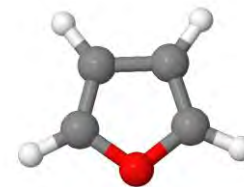
1



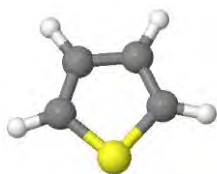
2



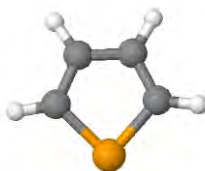
3



4



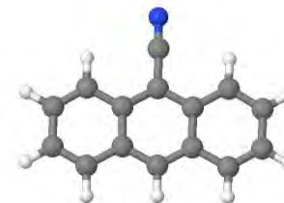
5



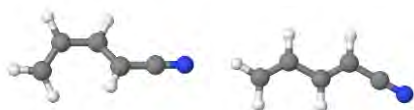
6



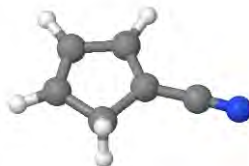
7



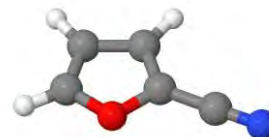
8



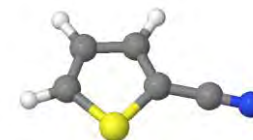
9



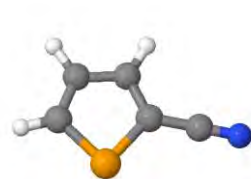
10



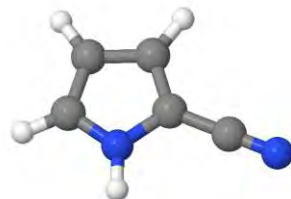
11



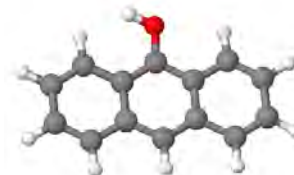
12



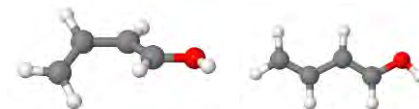
13



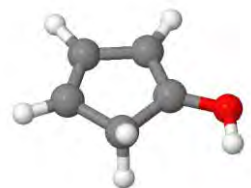
14



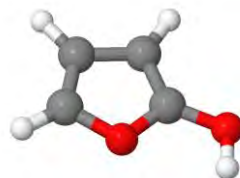
15



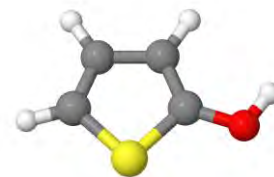
16



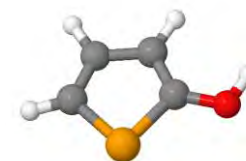
17



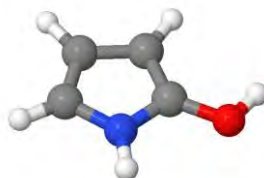
18



19



20



21

Figure 21. Optimized geometries of 21 dienes. Coloring code: blue (N), red (O), yellow (S), dark orange (Se), gray (C) and white (H).

Table 7. Structural, electronic and energetic data of 21 DA reactions between C60 and a series of 21 dienes.

Reaction	$\Delta E^\#$	ΔE	$\Delta E_i^\#$	$\Delta E_s^\#$	$\Delta H^\#$	$\Delta E_{sd}^\#$	$\Delta E_{sC60}^\#$	I_{Ts}
1	11.1	-20.3	-15.8	26.9	11.5	19.8	7.1	37.0
2	13.3	-33.1	-9.8	23.1	11.5	17.8	5.3	29.8
3	7.7	-24.8	-12.5	20.3	8.3	14.8	5.5	38.2
4	15.1	-6.2	-14.8	29.9	15.4	21.2	8.7	33.1
5	22.0	-4.0	-13.3	35.3	22.0	25.9	9.4	27.4
6	16.5	-11.9	-17.5	34.0	16.7	24.9	9.1	34.0
7	15.9	1.8	-18.0	33.9	16.3	25.8	8.2	34.7
8	13.0	-14.8	-13.7	26.7	13.3	21.2	5.5	34.0
9	14.8	-26.2	-12.9	27.8	13.1	19.2	8.6	31.8
10	10.7	-19.2	-11.7	22.3	11.2	16.8	5.5	34.3
11	18.7	-1.4	-13.3	32.1	18.8	23.5	8.6	29.4
12	24.5	0.7	-11.5	36.0	24.8	27.0	8.9	24.3

13	19.2	-7.1	-15.6	34.8	19.0	25.8	9.0	31.0
14	22.0	8.2	-18.1	40.0	22.1	28.8	11.2	31.1
15	5.9	-19.4	-18.8	24.7	6.2	18.0	6.7	43.2
16	8.1	-29.1	-14.1	22.2	6.8	17.1	5.2	38.8
17	2.7	-24.4	-17.0	19.7	3.2	14.4	5.2	46.4
18	12.6	-10.5	-18.1	30.7	12.9	21.8	8.9	37.1
19	17.1	-7.9	-16.5	33.7	17.4	24.1	9.6	32.9
20	12.1	-14.8	-20.1	32.2	12.3	22.8	9.4	38.5
21	10.9	-0.9	-29.4	40.3	11.2	26.3	14.0	42.2

Table 8. Synchronicity index (S in Å), polarity (q in e), frontier molecular orbitals (H_{diene} and L_{diene} in Hartree), electrophilicity (ω in eV) and HOMO_{diene} – LUMO_{C60} gap (H L in eV)

Reaction	S	q	H_{diene}	L_{diene}	ω	$H_d - L_{C60}$
1	0.000	0.174	-0.23695	-0.03100	1.19	4.12
2	0.000	0.166	-0.28697	0.02065	0.78	5.48
3	0.000	0.209	-0.26622	0.03018	0.64	4.92
4	0.000	0.265	-0.27735	0.06527	0.45	5.22
5	0.000	0.22	-0.28615	0.03195	0.69	5.46
6	0.000	0.217	-0.28477	0.02525	0.74	5.42
7	0.000	0.314	-0.25408	0.09525	0.25	4.59
8	0.216	0.079	-0.25464	-0.05828	1.70	4.60
9	0.229	0.083	-0.30987	-0.03410	1.46	6.11
10	0.170	0.117	-0.29208	-0.02385	1.27	5.62
11	0.153	0.164	-0.3039	-0.00461	1.08	5.95
12	0.268	0.117	-0.30788	-0.02344	1.31	6.05
13	0.269	0.116	-0.30651	-0.02770	1.36	6.02
14	0.179	0.234	-0.28331	0.01839	0.79	5.39
15	0.245	0.233	-0.22652	-0.02753	1.10	3.84
16	0.228	0.255	-0.25897	0.03600	0.57	4.72

17	0.209	0.276	-0.24134	0.04017	0.49	4.24
18	0.363	0.363	-0.25314	0.07708	0.32	4.56
19	0.410	0.305	-0.26088	0.04429	0.52	4.77
20	0.418	0.307	-0.26022	0.03843	0.56	4.76
21	0.572	0.472	-0.23415	0.10782	0.16	4.05

The decomposition of the strain energy in two terms arising from the diene and C₆₀, namely $\Delta E_{sd}^{\#}$ and $\Delta E_{sC60}^{\#}$, suggests that the diene always requires more energy than C₆₀ to get distorted into the strained geometry at the TS enhancing the interaction between the antagonists. Quantitatively speaking, Table 7 suggests that the dienes accounts for 65 to 80% of the strain energy. As a result, $\Delta E_{sd}^{\#}$ should dictate, with few exceptions, the trend in $\Delta E_s^{\#}$. In other words, increasing the strain in the diene should result in a very strained TS and vice-versa. For example, $\Delta E_{sd}^{\#}$ of 16.8, 23.5, 27.0, 25.8 and 28.8 kcal/mol are predicted for the series of 5-membered rings cyanurated dienes **10-14**, which correlate well with the ΔE_s values of 22.3, 32.1, 36.0, 34.8 and 40.0 kcal/mol. The plotting of $\Delta E_s^{\#}$ against $\Delta E_{sd}^{\#}$ supports this observation as it shows a correlation coefficient of 0.95 (See Appendix 4). Note that, for a given series of dienes and derivatives (1-7, 8-14 and 15-21), cyclopentadiene containing structures are the least strained, while pyrrole including ones show the highest strain.

As could be expected, unsubstituted dienes lead to symmetrical TSs. The insertion of –CN or –OH group at one of the terminal C atoms (engaged in the interaction) results in an asynchronous TS structure. For a given diene framework, the induced asynchronicity is generally more pronounced in the hydroxylated derivative than in the cyanurated one. This observation suggests that the –OH group provokes a larger gap between the atomic charges of the two terminal C atoms that –CN does, and causes them to interact noticeably differently with the identical [6,6] C atoms of C₆₀.

6. Concluding remarks

In this chapter, we have assessed the accuracy of different ONIOM models and DFT functionals for the description of the DA reactions of C₆₀ fullerene. Our results suggest that the ONIOM2(M06-2X/6-31G(d): SVWN/STO-3G) model, with the higher-layer including both the diene and the pyracyclene fragment of C₆₀, gives the best trade between accuracy and computational cost in the estimation of both the activation and reaction enthalpies. We have demonstrated the role of dispersion effects in the description of DA reactions involving C₆₀ fullerene, especially in the prediction of reaction energies. The analysis of frontier molecular orbitals suggests that electron donating (withdrawing) groups such as -OH (-CN) cut down on the activation barrier of the reaction due to the lowering (extending) of the HOMO_{diene} – LUMO_{C60} gap and enhancement (weakening) of the interaction between the two reactants. The decomposition of the activation energy into the strain and interaction components indicates that, for a given diene, electron donating groups (here -OH) diminish the height of the activation barrier not only by favoring the attractive interaction between the diene and C₆₀, but also by reducing the strain energy of the system; the opposite effect is observed for electron withdrawing groups (here -CN). A third paper that presents and discusses the findings reported in this chapter is under consideration in *ChemPhysChem* (revisions step).

“In so far as quantum mechanics is correct, chemical questions are problems in applied mathematics”

Eyring, Walter and Kimball

CONCLUSION AND PERSPECTIVES

Throughout this thesis, we have tried to address the problem of large and multi scale mechanistic modelling of Diels-Alder (DA) reactions. Due to the huge size of the dataset generated and analysed, chapters 2-4 were written in such a way that gives an overview of the main findings. We have placed most of the output files and some plots in the following GitHub repository, just in case the reader would like to have a look at them ([BienfaitKI/MSc_Thesis \(github.com\)](https://github.com/BienfaitKI/MSc_Thesis)).

The first problem of large-scale modelling of DA reactions has been tackled by designing and implementing a python-based package (AMADAR) that predicts an unlimited number of DA transition state geometries and carries out subsequent reaction path analyses, provided the SMILES string of the cycloadduct. The AMADAR tool is hosted on GitHub (link provided in chapter 2) and is expected to be helpful to a broad community of researchers working on the mechanism of DA reactions. The application of the AMADAR tool on a set of 2000 cycloadducts has provided insights into the origin and physical meaning of the (a)synchronicity in DA reactions.

On the other hand, we have shown that the ONIOM approach could be used to reduce the computational cost associated with the mechanistic modelling of DA reactions involving the Buckminster fullerene C₆₀. Our results suggest that the ONIOM2(M06-2X/6-31G(d):SVWN/STO-3G) model, where the higher-layer includes the diene and the pyracyclene portion of C₆₀, provides the best trade-off between accuracy and computing cost.

Some further questions related to the present work still need to be responded. These include:

1. The extension of the AMADAR program: this work has already been started. It aims at integrating new functionalities in the package. For example, a separate module called *Retro-HeteroDA* is being developed. This will allow for the analysis of aza and oxo-DA reactions. A standalone copy of this module, not yet tested and not connected with the rest of the package, is currently hosted in a GitHub repository ([BienfaitKI/MSc_Thesis \(github.com\)](https://github.com/BienfaitKI/MSc_Thesis)). Moreover, we would like to integrate another module called *ASM* (hosted in the same previous repository) to the rest of the AMADAR package. This module is designed to automate the analysis of IRC paths in the framework of the activation strain model developed by M. Bickelhaupt and coworkers.
2. Apply our ONIOM2(M06-2X/6-31G(d):SVWN/STO-3G) to a large dataset of DA reactions involving C₆₀ and other members of the fullerene family. This task will

obviously demand a certain degree of automation, which we hope to achieve using python scripting.

3. Finally, we would like to evaluate the possibility of extending the same automation methodology implemented within the AMADAR tool on other types of chemical reactions, beginning with other forms of cycloaddition up to carbocation and sigmatropic rearrangements processes.

Appendix

Appendix 1 : Main configuration file (da.ini)

[gaussian_environment]

GAUSS_SCR = /scr

GAUSS_SCRDIR = /scr

G09BASIS = /opt/chemistry/gaussian/g09.E/g09/basis

GAUSS_ARCHDIR = /opt/chemistry/gaussian/g09.E/arch

GAUSS_BSDDIR = /opt/chemistry/gaussian/g09.E/bsd

GAUSS_EXEDIR =

/opt/chemistry/gaussian/g09.E/bsd:/opt/chemistry/gaussian/g09.E/local:/opt/chemistry/gaussian/g09.E/extras:/opt/chemistry/gaussian/g09.E

GAUSS_LEXEDIR = /opt/chemistry/gaussian/g09.E/linda-exe

LD_LIBRARY_PATH =

/opt/chemistry/gaussian/g09.E/bsd:/opt/chemistry/gaussian/g09.E/local:/opt/chemistry/gaussian/g09.E/extras:/opt/chemistry/gaussian/g09.E

PATH =

/opt/chemistry/gaussian/g09.E/extras:/opt/chemistry/gaussian/g09.E/local:/opt/chemistry/gaussian/g09.E:/opt/chemistry/gaussian/g09.E/bsd

[gamess_environment]

PATH = /opt/chemistry/gamess

version = 00

np = 1

[prog]

engine = gamess

[job_details]

NPROCShared = 4

D_SPLITTING = 0.24

CALC_LEVELS_RC = pm6; b3lyp/6-31G(d)

CALC_LEVELS_TS = pm6; pm6; b3lyp/6-31G(d)

CALC_LEVEL_IRC = b3lyp/6-31G(d)

NBR_IRC_POINT_PP = 60

IRC_STEP_SIZE = 8

NBR_PATHS=1

TS_ID_numbers = 1,2,3,4

[flags]

SCRATCH = 1

RC_FLAG = 0

TS_FLAG = 1

IRC_FLAG = 0

IRC_GEOMS_CONSTR = 0

RFA_FLAG = 0

RFD_FLAG = 0

WBOA_FLAG = 0
OVERWRITE_FLAG = 0

Appendix 2 : Second configuration file (analysis.ini)

[RFA]

Unq_RFA = 1
Multiple_RF = 0
Multiple_RE = 0
Multiple_RFC = 0

[RFD]

JOB_ID = 585
ATOMS = 0
FRAG = 18,19;20,21
FRAG_NAMES = C18&C19, C20&21

[IRC_PATHS]

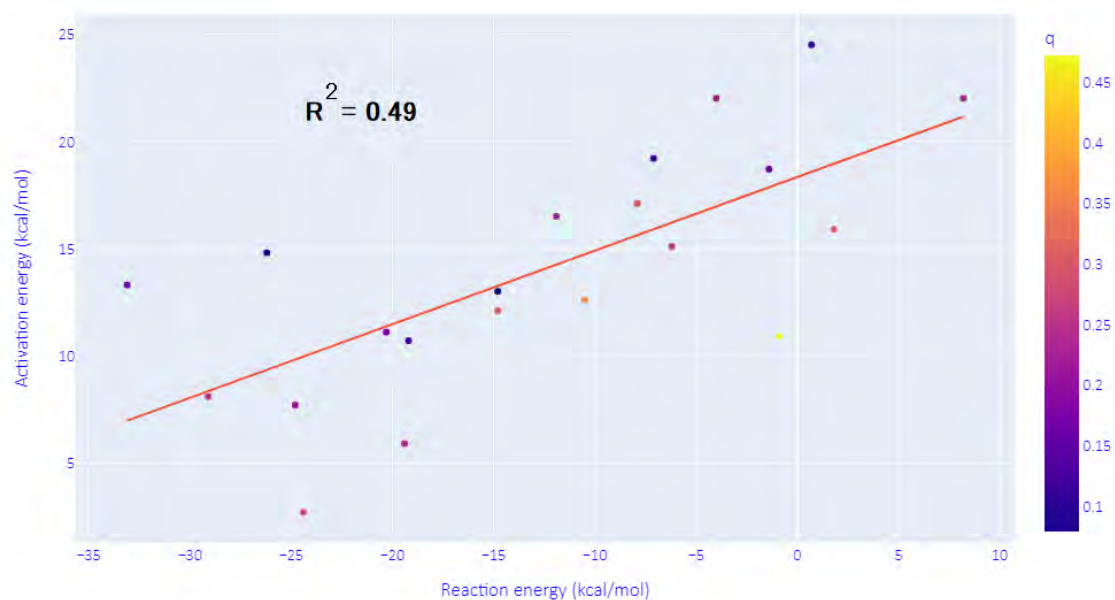
IRC_ID= 1
LEVEL = b3lyp/6-31G(d)
WBOA_ID= 1

[NPA]

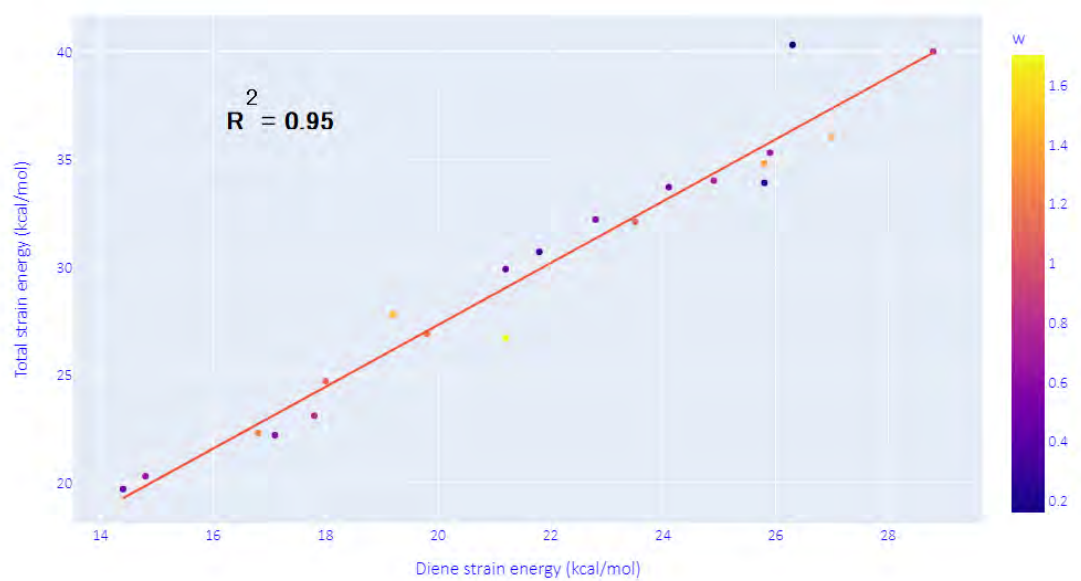
[REF]

[ASM]

Appendix 3: Correlation between the activation and reaction energies in the Diels-Alder reaction of C60 fullerene with 21 different dienes. The color scale denotes the polarity of the TS.



Appendix 4: Correlation between the total strain energy and the contribution from the diene in the Diels-Alder reaction of C60 fullerene with 21 different dienes. The color scale denotes the degree of electrophilicity of the diene.



References

1. Feifer N (1966) The relationship between Avogadro's principle and the law of Gay-Lussac. *J Chem Educ* 43:411–412. <https://doi.org/10.1021/ed043p411>
2. Schwartz JC, Senko MW, Syka JEP (2002) A two-dimensional quadrupole ion trap mass spectrometer. *J Am Soc Mass Spectrom* 13:659–669. [https://doi.org/10.1016/S1044-0305\(02\)00384-7](https://doi.org/10.1016/S1044-0305(02)00384-7)
3. Simons J (1993) Finding transition states when second-order Jahn–Teller instability occurs. *Int J Quantum Chem* 48:211–218. <https://doi.org/10.1002/qua.560480306>
4. Lewars EG (2016) Computational chemistry: Introduction to the theory and applications of molecular and quantum mechanics: Third Edition 2016
5. Cerjan CJ, Miller WH (1981) On finding transition states. *J Chem Phys* 75:2800–2801. <https://doi.org/10.1063/1.442352>
6. Svensson M, Humbel S, Froese RDJ, et al (1996) ONIOM: A multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and Pt(P(t-Bu)₃)₂ + H₂ oxidative addition. *J Phys Chem* 100:19357–19363. <https://doi.org/10.1021/jp962071j>
7. Nicolaou KC, Snyder SA, Montagnon T, Vassilikogiannakis G (2002) The Diels-Alder reaction in total synthesis. *Angew Chemie - Int Ed* 41:1668–1698. [https://doi.org/10.1002/1521-3773\(20020517\)41:10<1668::AID-ANIE1668>3.0.CO;2-Z](https://doi.org/10.1002/1521-3773(20020517)41:10<1668::AID-ANIE1668>3.0.CO;2-Z)
8. Briou B, Améduri B, Boutevin B (2021) Trends in the Diels–Alder reaction in polymer chemistry. *Chem Soc Rev*. <https://doi.org/10.1039/d0cs01382j>
9. Kittleson AR (1953) Preparation and Some Properties of N - Trichloromethylthiotetrahydrophthalimide. *Agric food Chem* 1:677–679
10. Cossío FP, Morao I, Jiao H, Von Ragué Schleyer P (1999) In-plane aromaticity in 1,3-dipolar cycloadditions. Solvent effects, selectivity, and nucleus-independent chemical shifts. *J Am Chem Soc* 121:6737–6746. <https://doi.org/10.1021/ja9831397>
11. Novikov AS, Kuznetsov ML (2012) Theoretical study of Re(IV) and Ru(II) bis-isocyanide complexes and their reactivity in cycloaddition reactions with nitrones. *Inorganica Chim Acta* 380:78–89. <https://doi.org/10.1016/j.ica.2011.08.016>
12. Diels O, Alder K (1928) Synthesen in der hydroaromatischen Reihe. *Justus Liebigs Ann Chem* 460:98–122. <https://doi.org/10.1002/jlac.19355190112>
13. Martin JG, Hill RK (1961) Stereochemistry of the Diels-Alder reaction. *Chem Rev*

14. Kinsman AC, Kerr MA (2003) The Total Synthesis of (+)-Hapalindole Q by an Organomediated Diels-Alder Reaction. *J Am Chem Soc* 125:14120–14125. <https://doi.org/10.1021/ja036191y>
15. Dewar MJS, Pierini AB (1984) Mechanism of the Diels-Alder Reaction. Studies of the Addition of Maleic Anhydride to Furan and Methylfurans. *J Am Chem Soc* 106:203–208. <https://doi.org/10.1021/ja00313a041>
16. Sakai S (2003) Theoretical analysis of concerted and stepwise mechanisms of the hetero-Diels-Alder reaction of butadiene with formaldehyde and thioformaldehyde. *J Mol Struct THEOCHEM* 630:177–185. [https://doi.org/10.1016/S0166-1280\(03\)00153-2](https://doi.org/10.1016/S0166-1280(03)00153-2)
17. James NC, Um JM, Padias AB, et al (2013) Computational investigation of the competition between the concerted diels-alder reaction and formation of diradicals in reactions of acrylonitrile with nonpolar dienes. *J Org Chem* 78:6582–6592. <https://doi.org/10.1021/jo400900x>
18. Chronakis N, Orfanopoulos M (2001) Concerted Diels-Alder reaction between anthracene and C60. *Tetrahedron Lett* 42:1201–1204. [https://doi.org/10.1016/S0040-4039\(00\)02211-5](https://doi.org/10.1016/S0040-4039(00)02211-5)
19. Otto S, Bertoncin F, Engberts JBFN (1996) Lewis acid catalysis of a Diels-Alder reaction in water. *J Am Chem Soc* 118:7702–7707. <https://doi.org/10.1021/ja960318k>
20. Aragonès AC, Haworth NL, Darwish N, et al (2016) Electrostatic catalysis of a Diels-Alder reaction. *Nature* 531:88–91. <https://doi.org/10.1038/nature16989>
21. Bini R, Chiappe C, Mestre VL, et al (2009) A theoretical study of the solvent effect on Diels-Alder reaction in room temperature ionic liquids using a supermolecular approach. *Theor Chem Acc* 123:347–352. <https://doi.org/10.1007/s00214-009-0525-0>
22. Cativiela C, García JI, Mayoral JA, Salvatella L (1996) Modelling of solvent effects on the diels-alder reaction. *Chem Soc Rev* 25:209–218. <https://doi.org/10.1039/CS9962500209>
23. Kumar A (2001) Salt effects on Diels-Alder reaction kinetics. *Chem Rev* 101:1–19. <https://doi.org/10.1021/cr990410+>
24. Houk KN, Gonzalez J, Li Y (1995) Pericyclic reaction transition states: passions and punctilios, 1935–1995. *AccChemRes* 28:81–90
25. Goldstein E, Beno B, Houk KN (1996) Density functional theory prediction of the relative energies and isotope effects for the concerted and stepwise mechanisms of the

- Diels-Alder reaction of butadiene and ethylene. *J Am Chem Soc* 118:6036–6043.
<https://doi.org/10.1021/ja9601494>
26. Li Y, Houk KN (1993) Diels-Alder Dimerization of 1,3-Butadiene: An ab Initio CASSCF Study of the Concerted and Stepwise Mechanisms and Butadiene–Ethylene Revisited. *J Am Chem Soc* 115:7478–7485. <https://doi.org/10.1021/ja00069a055>
 27. Steglenko D V., Kletsky ME, Kurbatov S V., et al (2011) The stepwise Diels-Alder reaction of 4-nitrobenzodifuroxan with Danishefsky's diene. *Chem - A Eur J* 17:7592–7604. <https://doi.org/10.1002/chem.201003695>
 28. Pham H V., Martin DBC, Vanderwal CD, Houk KN (2012) The intramolecular Diels-Alder reaction of tryptamine-derived Zincke aldehydes is a stepwise process. *Chem Sci* 3:1650–1655. <https://doi.org/10.1039/c2sc01072k>
 29. Wakayama H, Sakai S (2007) Theoretical analysis of concerted and stepwise mechanisms of Diels-Alder reactions of butadiene with silaethylene and dislene. *J Phys Chem A* 111:13575–13582. <https://doi.org/10.1021/jp074391g>
 30. Dewar MJS (1984) Multibond Reactions Cannot Normally Be Synchronous. *J Am Chem Soc* 106:209–219. <https://doi.org/10.1021/ja00313a042>
 31. Nendel M, Tolbert LM, Herring LE, et al (1999) Strained allenes as dienophiles in the Diels-Alder reaction: An experimental and computational study. *J Org Chem* 64:976–983. <https://doi.org/10.1021/jo982091c>
 32. Merino P, Chiacchio MA, Legnani L, et al (2017) Introducing topology to assess the synchronicity of organic reactions. Dual reactivity of oximes with alkenes as a case study. *Org Chem Front* 4:1541–1554. <https://doi.org/10.1039/c7qo00233e>
 33. De Souza MAF, Ventura E, Do Monte SA, et al (2016) Revisiting the concept of the (a)synchronicity of diels-alder reactions based on the dynamics of quasiclassical trajectories. *J Comput Chem* 37:701–711. <https://doi.org/10.1002/jcc.24245>
 34. Loncharich RJ, Brown FK, Houk KN (1989) Transition Structures of the Diels-Alder Reaction of Butadiene with Acrolein. *J Org Chem* 54:1129–1134.
<https://doi.org/10.1021/jo00266a026>
 35. Domingo LR, Pérez P, Sáez JA (2012) Origin of the synchronicity in bond formation in polar Diels-Alder reactions: An ELF analysis of the reaction between cyclopentadiene and tetracyanoethylene. *Org Biomol Chem* 10:3841–3851.
<https://doi.org/10.1039/c2ob25152c>
 36. Bachrach SM, White PB (2007) Towards assessing the aromaticity of the Diels-Alder transition state. *J Mol Struct THEOCHEM* 819:72–78.

- <https://doi.org/10.1016/j.theochem.2007.05.026>
37. Yepes D, Donoso-Tauda O, Pérez P, et al (2013) The reaction force constant as an indicator of synchronicity/ nonsynchronicity in [4+2] cycloaddition processes. *Phys Chem Chem Phys* 15:7311–7320. <https://doi.org/10.1039/c3cp44197k>
 38. Vermeeren P, Hamlin TA, Bickelhaupt FM (2021) Origin of Asynchronicity in Diels-Alder Reactions. *Phys Chem Chem Phys*. <https://doi.org/10.1039/d1cp02456f>
 39. Jedrzejewski M, Ordon P, Komorowski L (2016) Atomic Resolution for the Energy Derivatives on the Reaction Path. *J Phys Chem A* 120:3780–3787. <https://doi.org/10.1021/acs.jpca.6b03408>
 40. Schleyer PVR, Wu JI, Cossío FP, Fernández I (2014) Aromaticity in transition structures. *Chem Soc Rev* 43:4909–4921. <https://doi.org/10.1039/c4cs00012a>
 41. Von Ragué Schleyer P, Manoharan M, Jiao H, Stahl F (2001) The acenes: Is there a relationship between aromatic stabilization and reactivity? *Org Lett* 3:3643–3646. <https://doi.org/10.1021/ol016553b>
 42. Jun-ichi Aihara (1978) Aromaticity based theory of pericyclic reactions. *Bull Chem Soc Jpn* 51:1788–1792. <https://doi.org/10.1246/bcsj.51.1788>
 43. Manoharan M, Venuvanalingam P (1998) Gain or loss of aromaticity in Diels-Alder transition states and adducts: A theoretical investigation. *J Phys Org Chem* 11:133–140. [https://doi.org/10.1002/\(SICI\)1099-1395\(199802\)11:2<133::AID-POC981>3.0.CO;2-R](https://doi.org/10.1002/(SICI)1099-1395(199802)11:2<133::AID-POC981>3.0.CO;2-R)
 44. Dewar M (1971) Aromaticity and pericyclic reactions. *Angew Chemie - Int Ed* 17:761–776. <https://doi.org/10.1002/anie.197107611>
 45. Jiao H, Von Schleyer PR (1998) Aromaticity of pericyclic reaction transition structures: Magnetic evidence. *J Phys Org Chem* 11:655–662. [https://doi.org/10.1002/\(sici\)1099-1395\(199808/09\)11:8/9<655::aid-poc66>3.0.co;2-u](https://doi.org/10.1002/(sici)1099-1395(199808/09)11:8/9<655::aid-poc66>3.0.co;2-u)
 46. Matito E, Poater J, Duran M, Solà M (2005) An analysis of the changes in aromaticity and planarity along the reaction path of the simplest Diels-Alder reaction. Exploring the validity of different indicators of aromaticity. *J Mol Struct THEOCHEM* 727:165–171. <https://doi.org/10.1016/j.theochem.2005.02.020>
 47. Dobrowolski JC (2019) Three Queries about the HOMA Index. *ACS Omega* 4:18699–18710. <https://doi.org/10.1021/acsomega.9b02628>
 48. Chen Z, Wannere CS, Corminboeuf C, et al (2005) Nucleus-independent chemical shifts (NICS) as an aromaticity criterion. *Chem Rev* 105:3842–3888. <https://doi.org/10.1021/cr030088+>

49. Feixas F, Matito E, Poater J, Sola M (2009) On the performance of some aromaticity indices: a critical assessment using a test set. *J Comput Chem* 29:1543–1554. <https://doi.org/10.1002/jcc.20914>
50. Poater J, Fradera X, Duran M, Solà M (2003) The delocalization index as an electronic aromaticity criterion: Application to a series of planar polycyclic aromatic hydrocarbons. *Chem - A Eur J* 9:400–406. <https://doi.org/10.1002/chem.200390041>
51. Mohajeri A, Shahamirian M (2010) Theoretical study of Diels-Alder reaction: Role of substituent in regioselectivity and aromaticity. *J Iran Chem Soc* 7:554–563. <https://doi.org/10.1007/BF03246043>
52. Von Ragué Schleyer P, Puhlhofer F (2002) Recommendations for the evaluation of aromatic stabilization energies. *Org Lett* 4:2873–2876. <https://doi.org/10.1021/ol0261332>
53. Fowler PW, Steiner E (1997) Ring Currents and Aromaticity of Monocyclic π - Electron Systems C₆H₆, B₃N₃H₆, B₃O₃H₃, .. 5639:1409–1413
54. Evans MG, Warhurst E (1938) The activation energy of diene association reactions. *Trans Faraday Soc* 614–624
55. Vermeeren P, Brinkhuis F, Hamlin TA, Bickelhaupt FM (2020) How Alkali Cations Catalyze Aromatic Diels-Alder Reactions. *Chem - An Asian J* 15:1167–1174. <https://doi.org/10.1002/asia.202000009>
56. Vermeeren P, Hamlin TA, Fernández I, Bickelhaupt FM (2020) How Lewis Acids Catalyze Diels–Alder Reactions. *Angew Chemie - Int Ed* 59:6201–6206. <https://doi.org/10.1002/anie.201914582>
57. Schuster T, Kurz M, Göbel MW (2000) Catalysis of a Diels-Alder reaction by amidinium ions. *J Org Chem* 65:1697–1701. <https://doi.org/10.1021/jo991372x>
58. Larsen SD, Grieco PA (1985) Aza Diels-Alder Reactions in Aqueous Solution: Cyclocondensation of Dienes with Simple Iminium Salts Generated under Mannich Conditions. *J Am Chem Soc* 107:1768–1769. <https://doi.org/10.1002/asia.201600781>
59. Kumatabara Y, Kaneko S, Nakata S, et al (2016) Hydrogen-Bonding Catalysis of Tetraalkylammonium Salts in an Aza-Diels–Alder Reaction. *Chem - An Asian J* 11:2126–2129. <https://doi.org/10.1002/asia.201600781>
60. He CQ, Lam CC, Yu P, et al (2020) Catalytic Effects of Ammonium and Sulfonium Salts and External Electric Fields on Aza-Diels-Alder Reactions. *J Org Chem* 85:2618–2625. <https://doi.org/10.1021/acs.joc.9b03446>
61. Hilvert D, Hill KW, Nared KD, Auditor MM (1989) Antibody Catalysis of a Diels-

- Alder Reaction. *J Am Chem Soc* 92:61–92:62. <https://doi.org/10.1038/nature16989>
62. Kirsch N, Hedin-Dahlström J, Henschel H, et al (2009) Molecularly imprinted polymer catalysis of a Diels-Alder reaction. *J Mol Catal B Enzym* 58:110–117. <https://doi.org/10.1016/j.molcatb.2008.12.002>
 63. Oikawa H, Tokiwano T (2004) Enzymatic catalysis of the Diels-Alder reaction in the biosynthesis of natural products. *Nat Prod Rep* 21:321–352. <https://doi.org/10.1039/b305068h>
 64. Cativiela C, García JI, Gil J, et al (1997) Solvent effects on Diels-Alder reactions. the use of aqueous mixtures of fluorinated alcohols and the study of reactions of acrylonitrile. *J Chem Soc Perkin Trans 2* 653–660. <https://doi.org/10.1039/a601717g>
 65. Fischer T, Sethi A, Welton T, Woolf J (1999) Diels-Alder Reactions in Room-Temperature Ionic Liquids. *Chem Commun* 40:793–796
 66. Lee CW (1999) Diels-Alder reactions in chloroaluminate ionic liquids: Acceleration and selectivity enhancement. *Tetrahedron Lett* 40:2461–2464. [https://doi.org/10.1016/S0040-4039\(99\)00249-X](https://doi.org/10.1016/S0040-4039(99)00249-X)
 67. Tiwari S, Kumar A (2006) Diels-Alder reactions are faster in water than in ionic liquids at room temperature. *Angew Chemie - Int Ed* 45:4824–4825. <https://doi.org/10.1002/anie.200600426>
 68. Tapia O, Goscinski O (1975) Self-consistent reaction field theory of solvent effects. *Mol Phys* 29:1653–1661. <https://doi.org/10.1080/00268977500101461>
 69. Peeters A, Van Alsenoy C, Zhang ML, Van Doren VE (2000) On the use of supermolecule model for calculation of Young's modulus of crystalline polymers. *Int J Quantum Chem* 80:425–431. [https://doi.org/10.1002/1097-461X\(2000\)80:3<425::AID-QUA12>3.0.CO;2-0](https://doi.org/10.1002/1097-461X(2000)80:3<425::AID-QUA12>3.0.CO;2-0)
 70. Cativiela C, Dillet V, García JI, et al (1995) Solvent effects on Diels-Alder reactions. A semi-empirical study. *J Mol Struct THEOCHEM* 331:37–50. [https://doi.org/10.1016/0166-1280\(94\)03795-M](https://doi.org/10.1016/0166-1280(94)03795-M)
 71. Chandrasekhar J, Shariffskul S, Jorgensen WL (2002) QM/MM simulations for Diels-Alder reactions in water: Contribution of enhanced hydrogen bonding at the transition state to the solvent effect. *J Phys Chem B* 106:8078–8085. <https://doi.org/10.1021/jp020326p>
 72. Thomas LL, Tirado-Rives J, Jorgensen WL (2010) Quantum mechanical/molecular mechanical modeling finds diels alder reactions are accelerated less on the surface of water than in water. *J Am Chem Soc* 132:3097–3104.

- <https://doi.org/10.1021/ja909740y>
73. Woodward BRB, Sondheimer F, Taub D, et al (1952) Journal of the Chemical Society. J Am Chem Soc 74:4223–4250. <https://doi.org/10.1021/ja01137a001>
 74. Danishefsky S, Hirama M, Fritsch N, Clardy J (1979) Synthesis of Disodium Prephenate and Disodium Epiprephenate. Stereochemistry of Prephenic Acid and an Observation on the Base-Catalyzed Rearrangement of Prephenic Acid to /7-Hydroxyphenyllactic Acid. J American Chem Soc 101:7013–7018. <https://doi.org/10.1021/ja00517a039>
 75. Corey EJ, Weinshenker NM, Schaaf TK, Huber W (1969) Stereo-Controlled Synthesis of Prostaglandins F2a and E2. J American Chem Soc 91:5675–5677
 76. Wender PA, Schaus JM, White AW (1980) General Methodology for cis-Hydroisoquinoline Synthesis: Synthesis of Reserpine. J Am Chem Soc 102:6157–6159. <https://doi.org/10.1021/ja00539a038>
 77. Andriamialisoa RZ, Langlois N, Langlois Y (1985) A New Efficient Total Synthesis of Vindorosine and Vindoline. J Org Chem 50:961–967. <https://doi.org/10.1021/jo00207a011>
 78. Eschenbrenner-Lux V, Kumar K, Waldmann H (2014) The Asymmetric Hetero-Diels-Alder Reaction in the Syntheses of Biologically Relevant Compounds. Angew Chemie - Int Ed 53:11146–11157. <https://doi.org/10.1002/anie.201404094>
 79. Stepanyuk A, Kirschning A (2019) Synthetic terpenoids in the world of fragrances: Iso E Super® is the showcase. Beilstein J Org Chem 15:2590–2602. <https://doi.org/10.3762/bjoc.15.252>
 80. Chen Y, Hu C, Doohan D (2018) Safety of Bicyclopyrone on Several Vegetable Crops and Efficacy of Weed Control. Weed Technol 32:498–505. <https://doi.org/10.1017/wet.2018.26>
 81. Funel JA, Abele S (2013) Industrial applications of the Diels-Alder reaction. Angew Chemie - Int Ed 52:3822–3863. <https://doi.org/10.1002/anie.201201636>
 82. Kroto HW, Heath JR, O'Brien SC, et al (1985) C60: buckminsterfullerene. Nature 318:162–163
 83. Krätschmer W, Lamb LD, Fostiropoulos K, Huffman DR (1990) Solid C60: a new form of carbon. Nature 347:354–358. <https://doi.org/10.1038/347354a0>
 84. Hirsch A (2006) Functionalization of fullerenes and carbon nanotubes. Phys Status Solidi Basic Res 243:3209–3212. <https://doi.org/10.1002/pssb.200669191>
 85. Hirsch A (1995) Addition reactions in buckminsterfullerene C60. Synthesis (Stuttg)

- 8:895–913. <https://doi.org/10.1055/s-1995-4046>
86. Rubin Y, Khan S, Freedberg DI, Yeretdzian C (1993) Synthesis and X-ray Structure of a Diels-Alder Adduct of C₆₀. *J Am Chem Soc* 115:344–345.
<https://doi.org/10.1021/ja00054a049>
 87. An YZ, Ellis GA, Viado AL, Rubin Y (1995) A Methodology for the Reversible Solubilization of Fullerenes. *J Org Chem* 60:6353–6361.
<https://doi.org/10.1021/jo00125a022>
 88. Nebhani L, Barner-Kowollik C (2010) Functionalization of fullerenes with cyclopentadienyl and anthracenyl capped polymeric building blocks via Diels-Alder chemistry. *Macromol Rapid Commun* 31:1298–1305.
<https://doi.org/10.1002/marc.201000142>
 89. Nakamura E, Isobe H (2003) Functionalized Fullerenes in Water. The First 10 Years of Their Chemistry, Biology, and Nanoscience. *Acc Chem Res* 36:807–815.
<https://doi.org/10.1021/ar030027y>
 90. McEwen CN, McKay RG, Larsen BS (1992) C₆₀ as a radical sponge. *J Am Chem Soc* 114:4412–4414
 91. Nakamura E, Isobe H, Tomita N, et al (2000) Functionalized fullerene as an artificial vector for transfection. *Angew Chemie - Int Ed* 39:4254–4257.
[https://doi.org/10.1002/1521-3773\(20001201\)39:23<4254::AID-ANIE4254>3.0.CO;2-O](https://doi.org/10.1002/1521-3773(20001201)39:23<4254::AID-ANIE4254>3.0.CO;2-O)
 92. Bakry R, Vallant RM, Najam-ul-Haq M, et al (2007) Medicinal applications of fullerenes. *Int J Nanomedicine* 2:639–649. [https://doi.org/10.1016/S0081-1947\(08\)60578-0](https://doi.org/10.1016/S0081-1947(08)60578-0)
 93. Wudl F (2002) Fullerene materials. *J Mater Chem* 12:1959–1963.
<https://doi.org/10.1039/b201196d>
 94. Jia L, Chen M, Yang S (2020) Functionalization of fullerene materials toward applications in perovskite solar cells. *Mater Chem Front* 4:2256–2282.
<https://doi.org/10.1039/d0qm00295j>
 95. Johnson RD, Meijer G, Bethune DS (1990) C₆₀ Has Icosahedral Symmetry. *J Am Chem Soc* 112:8983–8984. <https://doi.org/10.1021/ja00180a055>
 96. Marek T, Ciesielski A (1995) Local aromatic character of C₆₀ and C₇₀ and their derivatives. *J Chem Inf Comput Sci* 35:1001–1003
 97. Hedberg K, Hedberg L, Bethune DS, et al (1991) Bond Lengths in Free Molecules of Buckminsterfullerene, C₆₀ from Gas-Phase Electron Diffraction. *Science* (80-)

- 254:410–412. <https://doi.org/10.1126/science.254.5030.410>
98. Pang LSK, Wilson MA (1993) Reactions of C₆₀ and C₇₀ with Cyclopentadiene. *J Phys Chem* 97:6761–6763
 99. Ohno M, Azuma T, Kojima S, et al (1996) An efficient functionalization of [60]fullerene. Diels-Alder reaction using 1,3-butadienes substituted with electron-withdrawing and electron-donating (silyloxy) groups. *Tetrahedron* 52:4983–4994. [https://doi.org/10.1016/0040-4020\(96\)00106-8](https://doi.org/10.1016/0040-4020(96)00106-8)
 100. Kräutler B, Müller T, Maynollo J, et al (1996) A Topochemically Controlled, Regiospecific Fullerene Bisfunctionalization. *Angew Chemie (International Ed English)* 35:1204–1206. <https://doi.org/10.1002/anie.199612041>
 101. Hirsch A, Lamparth I, Grösser T, Karfunkel HR (1994) Regiochemistry of Multiple Additions to the Fullerene Core: Synthesis of a Tb-Symmetric Hexakisadduct of C₆₀ with Bis(ethoxycarbonyl)methylene. *J Am Chem Soc* 116:9385–9386. <https://doi.org/10.1021/ja00099a088>
 102. Meidine MF, Roers R, Langley GJ, et al (1993) Formation and stabilisation of the hexa-cyclopentadiene adduct of C₆₀. *J Chem Soc Chem Commun* 1342–1344. <https://doi.org/10.1039/C39930001342>
 103. Kräutler B, Maynollo J (1995) A Highly Symmetric Sixfold Cycloaddition Product of Fullerene Photoswitchable Phase Separation in Hydrophobically Modified Poly (acrylamide) / Surfactant Systems **. *Angew Chem Int Ed* 34:87–88
 104. Wang GW, Saunders M, Cross RJ (2001) Reversible Diels - Alder addition to fullerenes: A study of equilibria using ³He NMR spectroscopy. *J Am Chem Soc* 123:256–259. <https://doi.org/10.1021/ja001346c>
 105. Wilson SR, Yurchenko ME, Schuster DI, et al (2000) Synthesis of fluorous fullerene adducts: Reversible solubilization of fullerenes in perfluorinated solvents. *J Org Chem* 65:2619–2623. <https://doi.org/10.1021/jo991008y>
 106. Sarova GH, Berberan-Santos MN (2004) Kinetics of the Diels-Alder reaction between C₆₀ and acenes. *Chem Phys Lett* 397:402–407. <https://doi.org/10.1016/j.cplett.2004.09.005>
 107. Solà M, Mestres J, Martí J, Duran M (1994) An AM1 study of the reactivity of buckminsterfullerene (C₆₀) in a Diels-Alder model reaction. *Chem Phys Lett* 231:325–330. [https://doi.org/10.1016/0009-2614\(94\)01249-0](https://doi.org/10.1016/0009-2614(94)01249-0)
 108. Solà M, Duran M, Mestres J (1996) Theoretical study of the regioselectivity of successive 1,3-butadiene Diels-Alder cycloadditions to C₆₀. *J Am Chem Soc*

- 118:8920–8924. <https://doi.org/10.1021/ja9614631>
109. Shang Z, Pan Y, Cai Z, et al (2000) An AM1 study of the reaction of ozone with C₆₀. *J Phys Chem A* 104:1915–1919. <https://doi.org/10.1021/jp9934937>
 110. Ravinder P, Subramanian V (2012) Density functional theory studies on the Diels-Alder reaction of [3]dendralene with C₆₀: An attractive approach for functionalization of fullerene. *Theor Chem Acc* 131:1–11. <https://doi.org/10.1007/s00214-012-1128-8>
 111. Ueno H, Kawakami H, Nakagawa K, et al (2014) Kinetic Study of the Diels–Alder Reaction of Li⁺@C₆₀ with cyclohexadiene: greatly increased reaction rate by encapsulation of Li⁺. *J Am Chem Soc* 136:11162–11167. <https://doi.org/10.1021/ja505952y>
 112. Cui CX, Liu YJ (2015) Role of encapsulated metal cation in the reactivity and regioselectivity of the C₆₀ Diels-Alder reaction. *J Phys Chem A* 119:3098–3106. <https://doi.org/10.1021/acs.jpca.5b00194>
 113. Beheshtian J, Peyghan AA, Bagheri Z (2012) Theoretical investigation of C₆₀ fullerene functionalization with tetrazine. *Comput Theor Chem* 992:164–167. <https://doi.org/10.1016/j.comptc.2012.05.039>
 114. Chung LW, Sameera WMC, Ramozzi R, et al (2015) The ONIOM Method and Its Applications. *Chem Rev* 115:5678–5796. <https://doi.org/10.1021/cr5004419>
 115. Osuna S, Morera J, Cases M, et al (2009) Diels - Alder Reaction between Cyclopentadiene and C₆₀ : An Analysis of the Performance of the ONIOM Method for the Study of Chemical Reactivity in Fullerenes and Nanotubes. *J Phys Chem A* 113:9721–9726. <https://doi.org/10.1021/jp904294y>
 116. Scott LT (1996) Fragments of fullerenes: Novel syntheses, structures and reactions. *Pure Appl Chem* 68:291–300. <https://doi.org/10.1351/pac199668020291>
 117. Osuna S, Swart M, Solà M (2011) Dispersion corrections essential for the study of chemical reactivity in fullerenes. *J Phys Chem A* 115:3491–3496. <https://doi.org/10.1021/jp1091575>
 118. Osuna S, Houk KN (2009) Cycloaddition reactions of butadiene and 1,3-dipoles to curved arenes, fullerenes, and nanotubes: Theoretical evaluation of the role of distortion energies on activation barriers. *Chem - A Eur J* 15:13219–13231. <https://doi.org/10.1002/chem.200901761>
 119. Cao Y, Osuna S, Liang Y, et al (2013) Diels-Alder reactions of graphene: Computational predictions of products and sites of reaction. *J Am Chem Soc* 135:17643–17649. <https://doi.org/10.1021/ja410225u>

120. Cohen NC, Blaney JM, Humblet C, et al (1990) MEDICINAL. *J Med Chem* 33:883–894
121. Cruz J, Martínez R, Genesca J, García-Ochoa E (2004) Experimental and theoretical study of 1-(2-ethylamino)-2-methylimidazoline as an inhibitor of carbon steel corrosion in acid media. *J Electroanal Chem* 566:111–121.
<https://doi.org/10.1016/j.jelechem.2003.11.018>
122. Dagdelen J, Montoya J, De Jong M, Persson K (2017) Computational prediction of new auxetic materials. *Nat Commun* 8:1–8. <https://doi.org/10.1038/s41467-017-00399-6>
123. Neese F (2009) Prediction of molecular properties and molecular spectroscopy with density functional theory: From fundamental theory to exchange-coupling. *Coord Chem Rev* 253:526–563. <https://doi.org/10.1016/j.ccr.2008.05.014>
124. Prakash N, Gareja DA (2010) Cheminformatics. *J Proteomics Bioinforma* 3:249–252. <https://doi.org/10.4172/jpb.1000147>
125. Wegner JK, Sterling A, Guha R, et al (2012) Cheminformatics. *Commun ACM* 55:65–75. <https://doi.org/10.1145/2366316.2366334>
126. Fernández I, Cossío FP (2014) Applied computational chemistry. *Chem Soc Rev* 43:4906–4908. <https://doi.org/10.1039/c4cs90040e>
127. Singh H, Sharma R (2012) Role of Adjacency Matrix & Adjacency List in Graph Theory. *Int J Comput Technol* 3:179–183. <https://doi.org/10.24297/ijct.v3i1c.2775>
128. Merrill D, Garland M, Grimshaw A (2015) High-performance and scalable GPU graph traversal. *ACM Trans Parallel Comput* 1:. <https://doi.org/10.1145/2717511>
129. David L, Thakkar A, Mercado R, Engkvist O (2020) Molecular representations in AI - driven drug discovery : a review and practical guide. *J Cheminform* 12:1–22. <https://doi.org/10.1186/s13321-020-00460-5>
130. Gutman I, Rücker C, Rücker G (2001) On Walks in Molecular Graphs. *J Chem Inf Comput Sci* 41:739–745. <https://doi.org/10.1021/ci000149u>
131. Velasquez JD, Echeverría J, Alvarez S (2021) Delocalized Bonding in Li₂X₂ Rings: Probing the Limits of the Covalent and Ionic Bonding Models. *Inorg Chem* 60:345–356. <https://doi.org/10.1021/acs.inorgchem.0c02972>
132. Zubarev DY, Boldyrev AI (2008) Developing paradigms of chemical bonding: Adaptive natural density partitioning. *Phys Chem Chem Phys* 10:5207–5217. <https://doi.org/10.1039/b804083d>
133. Kang S, Yang K, White SR, Sottos NR (2017) Silicon Composite Electrodes with

- Dynamic Ionic Bonding. *Adv Energy Mater* 7:1–7.
<https://doi.org/10.1002/aenm.201700045>
134. Powers IG, Uyeda C (2017) Metal-Metal Bonds in Catalysis. *ACS Catal* 7:936–958.
<https://doi.org/10.1021/acscatal.6b02692>
 135. Klein DJ, Trinajstić N (1990) Valence-bond theory and chemical structure. *J Chem Educ* 67:633–637. <https://doi.org/10.1021/ed067p633>
 136. David L, Arús-Pous J, Karlsson J, et al (2019) Applications of deep-learning in exploiting large-scale and heterogeneous compound data in industrial pharmaceutical research. *Front Pharmacol* 10:1–16. <https://doi.org/10.3389/fphar.2019.01303>
 137. Weininger D (1988) SMILES, a Chemical Language and Information System. 1 Introduction to methodology and Encoding Rules. *J Chem Inf Comput Sci* 28:31–36.
<https://doi.org/10.1021/ci00057a005>
 138. Landrum G (2011) RDKit : A software suite for cheminformatics , computational chemistry , and predictive modeling. Components
 139. O’Boyle NM, Banck M, James CA, et al (2011) Open Babel : an open chemical toolbox. *J Cheminform* 3:1–14. <https://doi.org/10.1186/1758-2946-3-33>
 140. Beisen S, Meinel T, Wiswedel B, et al (2013) KNIME-CDK : Workflow-driven cheminformatics. 2–5
 141. Bjerrum EJ (2017) SMILES Enumeration as Data Augmentation for Neural Network Modeling of Molecules
 142. Schneider N, Sayle RA, Landrum GA (2015) Get Your Atoms in Order-An Open-Source Implementation of a Novel and Robust Molecular Canonicalization Algorithm. *J Chem Inf Model* 55:2111–2120. <https://doi.org/10.1021/acs.jcim.5b00543>
 143. O’Boyle NM (2012) Towards a Universal SMILES representation - A standard method to generate canonical SMILES based on the InChI. *J Cheminform* 4:1–14.
<https://doi.org/10.1186/1758-2946-4-22>
 144. Hanson RM (2016) Jmol SMILES and Jmol SMARTS: Specifications and applications. *J Cheminform* 8:1–20. <https://doi.org/10.1186/s13321-016-0160-4>
 145. Garay-Ruiz D, Bo C (2021) Human-Readable SMILES: Translating Cheminformatics to Chemistry
 146. O’Boyle NM, Dalke A (2018) DeepSMILES: An adaptation of SMILES for use in machine-learning of chemical structures. *ChemRxiv* 1–9.
<https://doi.org/10.26434/chemrxiv.7097960>
 147. Miller MA (2002) Chemical database techniques in drug discovery. *Nat Rev Drug*

- Discov 1:220–227. <https://doi.org/10.1038/nrd745>
148. Edwards P (2002) Combinatorial chemistry. *Drug Discov Today* 7:383–384. [https://doi.org/10.1016/S1359-6446\(02\)02181-5](https://doi.org/10.1016/S1359-6446(02)02181-5)
 149. Pottorf RS, Chadha NK, Katkevics M, et al (2003) Parallel synthesis of benzoxazoles via microwave-assisted dielectric heating. *Tetrahedron Lett* 44:175–178. [https://doi.org/10.1016/S0040-4039\(02\)02495-4](https://doi.org/10.1016/S0040-4039(02)02495-4)
 150. Diallo BN, Glenister M, Musyoka TM, et al (2021) SANCDB: an update on South African natural compounds and their readily available analogs. *J Cheminform* 13:1–15. <https://doi.org/10.1186/s13321-021-00514-2>
 151. Groom CR, Bruno IJ, Lightfoot MP, Ward SC (2016) The Cambridge structural database. *Acta Crystallogr Sect B Struct Sci Cryst Eng Mater* 72:171–179. <https://doi.org/10.1107/S2052520616003954>
 152. Pence y HE, Pence y HE (2010) ChemSpider: An Online Chemical Information Resource. *Chem Educ today* 87:1123–1124. <https://doi.org/10.1021/ed100697w>
 153. Goto S, Okuno Y, Hattori M, et al (2002) LIGAND: Database of chemical compounds and reactions in biological pathways. *Nucleic Acids Res* 30:402–404. <https://doi.org/10.1093/nar/30.1.402>
 154. Voigt JH, Bienfait B, Wang S, Nicklaus MC (2001) Comparison of the NCI Open Database with Seven Large Chemical Structural Databases. *J Chem Inf Comput Sci* 41:702–712. <https://doi.org/10.1021/ci000150t>
 155. Irwin JJ, Shoichet BK (2005) ZINC - A free database of commercially available compounds for virtual screening. *J Chem Inf Model* 45:177–182. <https://doi.org/10.1021/ci049714+>
 156. Bannon R, Chin A, Kassam F, Roszko A (2004) Conceptual Architecture
 157. Cramer CJ, Truhlar DG (1992) AM1-SM2 and PM3-SM3 parameterized SCF solvation models for free energies in aqueous solution. *J Comput Aided Mol Des* 6:629–666. <https://doi.org/10.1007/BF00126219>
 158. Sharp O (1994) The Perl Programming Language. *Dr Dobb's J Softw Tools* 19:1–70
 159. Jensen F (2017) *Introduction to Computational Chemistry Programs*, 3rd ed. John Wiley & Sons, Ltd, Chichester
 160. Frierson MR, Imam MR, Allinger NL, Zalkow VB (1988) The MM2 Force Field for Silanes and Polysilanes. *J Org Chem* 53:5248–5258. <https://doi.org/10.1021/jo00257a010>
 161. Wang X, Ramírez-Hinestrosa S, Dobnikar J, Frenkel D (2020) The Lennard-Jones

- potential: When (not) to use it. *Phys Chem Chem Phys* 22:10624–10633.
<https://doi.org/10.1039/c9cp05445f>
162. Jász Á, Rák Á, Ladjánszki I, Cserey G (2019) Optimized GPU implementation of Merck Molecular Force Field and Universal Force Field. *J Mol Struct* 1188:227–233.
<https://doi.org/10.1016/j.molstruc.2019.04.007>
 163. Lii J -H, Allinger NL (1991) The MM3 force field for amides, polypeptides and proteins. *J Comput Chem* 12:186–199. <https://doi.org/10.1002/jcc.540120208>
 164. MacKerell AD, Banavali N, Foloppe N (2000) Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* 56:257–265.
[https://doi.org/10.1002/1097-0282\(2000\)56:4<257::AID-BIP10029>3.0.CO;2-W](https://doi.org/10.1002/1097-0282(2000)56:4<257::AID-BIP10029>3.0.CO;2-W)
 165. Wang J, Wolf RM, Caldwell JW, et al (2004) Development and testing of a general Amber force field. *J Comput Chem* 25:1157–1174. <https://doi.org/10.1002/jcc.20035>
 166. Norrby PO, Brandt P (2001) Deriving force field parameters for coordination complexes. *Coord Chem Rev* 212:79–109. [https://doi.org/10.1016/s0010-8545\(00\)00296-4](https://doi.org/10.1016/s0010-8545(00)00296-4)
 167. Klauda JB, Monje V, Kim T, Im W (2012) Improving the CHARMM force field for polyunsaturated fatty acid chains. *J Phys Chem B* 116:9424–9431.
<https://doi.org/10.1021/jp304056p>
 168. Artemova S, Jaillet L, Redon S (2016) Automatic molecular structure perception for the universal force field. *J Comput Chem* 37:1191–1205.
<https://doi.org/10.1002/jcc.24309>
 169. Casewit CJ, Colwell KS, Rappé AK (1992) Application of a Universal Force Field to Organic Molecules. *J Am Chem Soc* 114:10036–1046.
<https://doi.org/10.1038/220833b0>
 170. Liwo A, He Y, Scheraga HA (2011) Coarse-grained force field: General folding theory. *Phys Chem Chem Phys* 13:16890–16901. <https://doi.org/10.1039/c1cp20752k>
 171. Marrink SJ, Risselada HJ, Yefimov S, et al (2007) The MARTINI force field: Coarse grained model for biomolecular simulations. *J Phys Chem B* 111:7812–7824.
<https://doi.org/10.1021/jp071097f>
 172. Van Duin ACT, Dasgupta S, Lorant F, Goddard WA (2001) ReaxFF: A reactive force field for hydrocarbons. *J Phys Chem A* 105:9396–9409.
<https://doi.org/10.1021/jp004368u>
 173. Mackey GW (1957) Quantum Mechanics and Hilbert Space. *Am Math Mon* 64:45–57.
<https://doi.org/10.1080/00029890.1957.11989120>

174. Woolley RG, Sutcliffe BT (1977) Molecular structure and the born-Oppenheimer approximation. *Chem Phys Lett* 45:393–398. [https://doi.org/10.1016/0009-2614\(77\)80298-4](https://doi.org/10.1016/0009-2614(77)80298-4)
175. Nakai H (2007) Nuclear Orbital Plus Molecular Orbital Theory: Simultaneous Determination of Nuclear and Electronic Wave Functions Without Born–Oppenheimer Approximation. *Int J Quantum Chem* 107:2849–2869. <https://doi.org/10.1002/qua>
176. Lieb EH, Simon B, U EM (1977) The Hartree-Fock Theory for Coulomb Systems. *Commun Math Phys* 194:185–194
177. Hartree DR (1928) The Wave Mechanics of an Atom with a Non-Coulomb Central Field Part I Theory and Methods. *Math Proc Cambridge Philos Soc* 24:89–110. <https://doi.org/10.1017/S0305004100011919>
178. Matsen FA, Cantu AA (1969) Spin-Free Quantum Chemistry. VII. The Slater Determinant. *J Phys Chem* 2488–2494
179. Stewart RF (1970) Small Gaussian Expansions of Slater-Type Orbitals. *J Chem Phys* 52:425–431. <https://doi.org/10.1063/1.1672702>
180. Magalhães AL (2014) Gaussian-type orbitals versus slater-type orbitals: A comparison. *J Chem Educ* 91:2124–2127. <https://doi.org/10.1021/ed500437a>
181. Pitzer RM (1967) Optimized molecular orbital wavefunctions for methane constructed from a minimum basis set. *J Chem Phys* 46:4871–4875. <https://doi.org/10.1063/1.1840649>
182. Weigend F, Ahlrichs R (2005) Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys Chem Chem Phys* 7:3297–3305. <https://doi.org/10.1039/b508541a>
183. Rassolov VA, Ratner MA, Pople JA, et al (2001) 6-31G* basis set for third-row atoms. *J Comput Chem* 22:976–984. <https://doi.org/10.1002/jcc.1058>
184. Fischer-Hjalmars I (1965) Deduction of the zero differential overlap approximation from an orthogonal atomic orbital basis. *J Chem Phys* 42:1962–1972. <https://doi.org/10.1063/1.1696232>
185. Gonzalez-Lafont A, Truong TN, Truhlar DG (1991) Direct dynamics calculations with neglect of diatomic differential overlap molecular orbital theory with specific reaction parameters. *J Phys Chem* 95:4618–4627. <https://doi.org/10.1021/j100165a009>
186. Birner P, Hofmann H -J (1982) Critical examination of approximate LCAOMO methods. I. Reasons for the failure of the CNDO and INDO methods in theoretical

- conformational analysis of conjugated compounds. *Int J Quantum Chem* 21:833–843.
<https://doi.org/10.1002/qua.560210508>
187. Lewis DFV (1986) MINDO/3: a review of the literature. *Chem Rev* 86:1111–1123
 188. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) AM1: A New General Purpose Quantum Mechanical Molecular Model. *J Am Chem Soc* 107:3902–3909.
<https://doi.org/10.1021/ja00299a024>
 189. Stewart JJP (1989) Optimization of parameters for semiempirical methods II. Applications. *J Comput Chem* 10:221–264. <https://doi.org/10.1002/jcc.540100209>
 190. Stewart JJP (2007) Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J Mol Model* 13:1173–1213. <https://doi.org/10.1007/s00894-007-0233-4>
 191. Repasky MP, Chandrasekhar J, Jorgensen WL (2002) PDDG/PM3 and PDDG/MNDO: Improved semiempirical methods. *J Comput Chem* 23:1601–1622.
<https://doi.org/10.1002/jcc.10162>
 192. Shavitt I (1998) The history and evolution of configuration interaction. *Mol Phys* 94:3–17. <https://doi.org/10.1080/002689798168303>
 193. Christiansen O (2003) Møller-Plesset perturbation theory for vibrational wave functions. *J Chem Phys* 119:5773–5781. <https://doi.org/10.1063/1.1601593>
 194. Honenberg P, W.Kohn (1964) Inhomogeneous Electron gas. *Phys Rev* 136:864–870.
<https://doi.org/10.1007/BF01198136>
 195. Dirac PAM (1930) Note on Exchange Phenomena in the Thomas Atom. *Math Proc Cambridge Philos Soc* 26:376–385. <https://doi.org/10.1017/S0305004100016108>
 196. Kohn W, Sham LJ (1965) Self-Consistent Equations Including Exchange and Correlation Effects. *Phys Rev* 140:1133–1138.
<https://doi.org/10.1103/PhysRev.140.A1133>
 197. Slater JC (1951) A simplification of the Hartree-Fock method. *Phys Rev* 81:385–390.
<https://doi.org/10.1103/PhysRev.81.385>
 198. Vosko SH, Wilk L, Nusair M (1980) Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can J Phys* 58:1200–1211. <https://doi.org/10.1139/p80-159>
 199. Perdew JP, Wang Y (1992) Accurate and simple analytic representation of the electron-gas correlation energy. *Phys Rev B* 98:244–249.
<https://doi.org/10.1103/PhysRevB.98.079904>
 200. A.D.Becke (1988) Density-fnctional exchange-energy approximation with correct

- asymptotic behavior. *Phys Rev* 38:3098–3100.
<https://doi.org/10.1103/physreva.38.3098>
201. Lee C, Yang W, Parr RG (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B* 37:785–789.
<https://doi.org/10.1103/physrevb.37.785>
 202. Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. *Phys Rev Lett* 77:3865–3868. <https://doi.org/10.1103/PhysRevLett.77.3865>
 203. Gomes JRB, Fajín JLC, Cordeiro MNDS, et al (2013) Density functional treatment of interactions and chemical reactions at interfaces. *Density Funct Theory Princ Appl Anal* 1–58
 204. Tao J, Perdew JP, Staroverov VN, Scuseria GE (2003) Climbing the density functional ladder: Nonempirical meta-generalized gradient approximation designed for molecules and solids. *Phys Rev Lett* 91:.
<https://doi.org/10.1103/PhysRevLett.91.146401>
 205. Avellaneda F, Bustacara C, Garzón JP, González E (2006) Implementation of a molecular simulator based on a MultiAgent system. *Proc - 2006 IEEE/WIC/ACM Int Conf Intell Agent Technol (IAT 2006 Main Conf Proceedings)*, IAT'06 82:117–120.
<https://doi.org/10.1109/IAT.2006.80>
 206. Becke AD (1993) A new mixing of Hartree-Fock and local density-functional theories. *J Chem Phys* 98:1372–1377. <https://doi.org/10.1063/1.464304>
 207. Darvish Ganji M, Hosseini-Khah SM, Amini-Tabar Z (2015) Theoretical insight into hydrogen adsorption onto graphene: A first-principles B3LYP-D3 study. *Phys Chem Chem Phys* 17:2504–2511. <https://doi.org/10.1039/c4cp04399e>
 208. Chai J Da, Head-Gordon M (2008) Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys Chem Chem Phys* 10:6615–6620. <https://doi.org/10.1039/b810189b>
 209. Bousquet D, Brémond E, Sancho-García JC, et al (2013) Is there still room for parameter free double hybrids? Performances of PBE0-DH and B2PLYP over extended benchmark sets. *J Chem Theory Comput* 9:3444–3452.
<https://doi.org/10.1021/ct400358f>
 210. Honig B, Karplus M (1971) Implication of torsional potential of retinal isomers for visual excitation. *Nature* 229:558–560. <https://doi.org/10.1038/229558a0>
 211. Warshel A, Levitt M (1976) Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J*

- Mol Biol 103:227–249. [https://doi.org/10.1016/0022-2836\(76\)90311-9](https://doi.org/10.1016/0022-2836(76)90311-9)
212. Wallrapp FH, Guallar V (2011) Mixed quantum mechanics and molecular mechanics methods: Looking inside proteins. *Wiley Interdiscip Rev Comput Mol Sci* 1:315–322. <https://doi.org/10.1002/wcms.27>
213. Senn HM, Thiel W (2009) QM/MM methods for biomolecular systems. *Angew Chemie - Int Ed* 48:1198–1229. <https://doi.org/10.1002/anie.200802019>
214. Morokuma K, Wang Q, Vreven T (2006) Performance evaluation of the three-layer ONIOM method: Case study for a zwitterionic peptide. *J Chem Theory Comput* 2:1317–1324. <https://doi.org/10.1021/ct600135b>
215. Maseras F, Morokuma K (1995) IMOMM: A new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J Comput Chem* 16:1170–1179. <https://doi.org/10.1002/jcc.540160911>
216. Humbel S, Sieber S, Morokuma K (1996) The IMOMO method: Integration of different levels of molecular orbital approximations for geometry optimization of large systems: Test for n-butane conformation and SN2 reaction: $\text{RCl} + \text{Cl}^-$. *J Chem Phys* 105:1959–1967. <https://doi.org/10.1063/1.472065>
217. Vreven T, Morokuma K (2000) On the Application of the IMOMO (Integrated Molecular Orbital + Molecular Orbital) Method. *J Comput Chem* 21:1419–1432. [https://doi.org/10.1002/1096-987X\(200012\)21:16<1419::AID-JCC1>3.0.CO;2-C](https://doi.org/10.1002/1096-987X(200012)21:16<1419::AID-JCC1>3.0.CO;2-C)
218. Sokkar P, Boulanger E, Thiel W, Sanchez-Garcia E (2015) Hybrid quantum mechanics/molecular mechanics/coarse grained modeling: A triple-resolution approach for biomolecular systems. *J Chem Theory Comput* 11:1809–1818. <https://doi.org/10.1021/ct500956u>
219. Banerjee A, Adams N, Simons J, Shepard R (1985) Search for stationary points on surfaces. *J Phys Chem* 89:52–57. <https://doi.org/10.1021/j100247a015>
220. Baker J (1986) An algorithm for the location of transition states. *J Comput Chem* 7:385–395. <https://doi.org/10.1002/jcc.540070402>
221. Shang C, Liu ZP (2012) Constrained broyden dimer method with bias potential for exploring potential energy surface of multistep reaction process. *J Chem Theory Comput* 8:2215–2222. <https://doi.org/10.1021/ct300250h>
222. Henkelman G, Jónsson H (2000) Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J Chem Phys* 113:9978–9985. <https://doi.org/10.1063/1.1323224>
223. Maragakis P, Andreev SA, Brumer Y, et al (2002) Adaptive nudged elastic band

- approach for transition state calculation. *J Chem Phys* 117:4651–4658.
<https://doi.org/10.1063/1.1495401>
224. Henkelman G (2000) A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J Chem Phys* 113:9901–9904.
<https://doi.org/10.1063/1.1329672>
 225. Dittrich R, Schrefl T, Suess D, et al (2002) A path method for finding energy barriers and minimum energy paths in complex micromagnetic systems. *J Magn Magn Mater* 250:12–19. [https://doi.org/10.1016/s0304-8853\(02\)00388-8](https://doi.org/10.1016/s0304-8853(02)00388-8)
 226. Pan AC, Sezer D, Roux B (2008) Finding transition pathways using the string method with swarms of trajectories. *J Phys Chem B* 112:3432–3440.
<https://doi.org/10.1021/jp0777059>
 227. Quapp W (2005) A growing string method for the reaction pathway defined by a Newton trajectory. *J Chem Phys* 122:1–11. <https://doi.org/10.1063/1.1885467>
 228. Rincón E, Jaque P, Toro-Labbé A (2006) Reaction force analysis of the effect of Mg(II) on the 1,3 intramolecular hydrogen transfer in thymine. *J Phys Chem A* 110:9478–9485. <https://doi.org/10.1021/jp062870u>
 229. Toro-Labbé A (1999) Characterization of chemical reactions from the profiles of energy, chemical potential, and hardness. *J Phys Chem A* 103:4398–4403.
<https://doi.org/10.1021/jp984187g>
 230. Fukui K (1981) The Path of Chemical Reactions —The IRC Approach. *Acc Chem Res* 14:365–368. <https://doi.org/10.1021/cen-v046n004.p005>
 231. Inostroza-Rivera R, Herrera B, Toro-Labbé A (2014) Using the reaction force and the reaction electronic flux on the proton transfer of formamide derived systems. *Phys Chem Chem Phys* 16:14489–14495. <https://doi.org/10.1039/c3cp55159h>
 232. Hargis JC, Vöhringer-Martinez E, Woodcock HL, et al (2011) Characterizing the mechanism of the double proton transfer in the formamide dimer. *J Phys Chem A* 115:2650–2657. <https://doi.org/10.1021/jp111834v>
 233. Herrera B, Toro-Labbé A (2007) The role of reaction force and chemical potential in characterizing the mechanism of double proton transfer in the adenine-uracil complex. *J Phys Chem A* 111:5921–5926. <https://doi.org/10.1021/jp065951z>
 234. Yepes D, Murray JS, Pérez P, et al (2014) Complementarity of reaction force and electron localization function analyses of asynchronicity in bond formation in Diels-Alder reactions. *Phys Chem Chem Phys* 16:6726–6734.
<https://doi.org/10.1039/c3cp54766c>

235. Labet V, Morell C, Toro-labbe A (2008) Theoretical Study of Cytosine Deamination from the Perspective of the Reaction Force Analysis. *J Phys Chem* 112:11487–11494. <https://doi.org/10.1021/jp8059097>
236. Toro-Labbé A, Gutiérrez-Oliva S, Concha MC, et al (2004) Analysis of two intramolecular proton transfer processes in terms of the reaction force. *J Chem Phys* 121:4570–4576. <https://doi.org/10.1063/1.1777216>
237. Truhlar DG, Garrett BC, V BB, et al (1996) Current Status of Transition-State Theory. *J Phys Chem* 100:12771–12800. <https://doi.org/10.1021/jp953748q>
238. Zewail AH (2000) Femtochemistry: Atomic-scale dynamics of the chemical bond. *J Phys Chem A* 104:5660–5694. <https://doi.org/10.1021/jp001460h>
239. Yepes D, Murray JS, Politzer P, Jaque P (2012) The reaction force constant: An indicator of the synchronicity in double proton transfer reactions. *Phys Chem Chem Phys* 14:11125–11134. <https://doi.org/10.1039/c2cp41064h>
240. Politzer P, Murray JS (2018) The Hellmann-Feynman theorem: a perspective. *J Mol Model* 24:1–7. <https://doi.org/10.1007/s00894-018-3784-7>
241. Geerlings P, Fias S, Boisdenghien Z, De Proft F (2014) Conceptual DFT: Chemistry from the linear response function. *Chem Soc Rev* 43:4989–5008. <https://doi.org/10.1039/c3cs60456j>
242. Komorowski L, Ordon P (2001) Vibrational softening of diatomic molecules. *Theor Chem Acc* 105:338–344. <https://doi.org/10.1007/s002140000244>
243. Hansen DA, Traut DE (1989) The kinetics of leaching rock ilmenite with hydrofluoric acid. *Jom* 41:34–36. <https://doi.org/10.1007/BF03220221>
244. Peleg M, Normand MD, Corradini MG (2012) The Arrhenius equation revisited. *Crit Rev Food Sci Nutr* 52:830–851. <https://doi.org/10.1080/10408398.2012.667460>
245. Truhlar DG, Garrett BC, V BB, et al (1996) Current Status of Transition-State Theory. *J Phys Chem* 100:12771–12800
246. Pattanaik L, Ingraham JB, Grambow CA, Green WH (2020) Generating transition states of isomerization reactions with deep learning. *Phys Chem Chem Phys* 22:23618–23626. <https://doi.org/10.1039/d0cp04670a>
247. Makoś MZ, Verma N, Larson EC, et al (2021) Generative adversarial networks for transition state geometry prediction. *J Chem Phys* 155:. <https://doi.org/10.1063/5.0055094>
248. Hernández Mancera JP, Núñez-Zarur F, Gutiérrez-Oliva S, et al (2020) Diels-Alder reaction mechanisms of substituted chiral anthracene: A theoretical study based on the

- reaction force and reaction electronic flux. *J Comput Chem* 41:2022–2032.
<https://doi.org/10.1002/jcc.26360>
249. Bickelhaupt FM, Houk KN (2017) Analyzing Reaction Rates with the Distortion/Interaction-Activation Strain Model. *Angew Chemie - Int Ed* 56:10070–10086. <https://doi.org/10.1002/anie.201701486>
 250. Berski S, Andrés J, Silvi B, Domingo LR (2006) New findings on the Diels - Alder reactions. An analysis based on the bonding evolution theory. *J Phys Chem A* 110:13939–13947. <https://doi.org/10.1021/jp068071t>
 251. Frisch MJ, Trucks GW, Schlegel HB, et al (2009) Gaussian 09 (C.01) inc. Wallingford CT
 252. Turney JM, Simmonett AC, Parrish RM, et al (2012) Psi4: An open-source ab initio electronic structure program. *Wiley Interdiscip Rev Comput Mol Sci* 2:556–565. <https://doi.org/10.1002/wcms.93>
 253. Gordon MS, Schmidt MW (2005) Advances in electronic structure theory: Gamess a decade later. *Theory Appl Comput Chem* 1185–1189. <https://doi.org/10.1016/B978-044451719-7/50084-6>
 254. Casewit CJ, Colwell KS, Rappé AK (1992) Application of a Universal Force Field to Organic Molecules. *J Am Chem Soc* 114:10035–10046. <https://doi.org/10.1038/220833b0>
 255. Domingo LR (2016) Molecular electron density theory: A modern view of reactivity in organic chemistry. *Molecules* 21:1–15. <https://doi.org/10.3390/molecules21101319>
 256. Toro-Labbé A, Gutiérrez-Oliva S, Murray JS, Politzer P (2009) The reaction force and the transition region of a reaction. *J Mol Model* 15:707–710. <https://doi.org/10.1007/s00894-008-0431-8>
 257. Di Ventra M, Pantelides ST (2000) Hellmann-Feynman theorem and the definition of forces in quantum time-dependent and transport problems. *Phys Rev B - Condens Matter Mater Phys* 61:16207–16212. <https://doi.org/10.1103/PhysRevB.61.16207>
 258. Suleimanov Y V., Green WH (2015) Automated Discovery of Elementary Chemical Reaction Steps Using Freezing String and Berny Optimization Methods. *J Chem Theory Comput* 11:4248–4259. <https://doi.org/10.1021/acs.jctc.5b00407>
 259. Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. *J Chem Phys* 98:5648–5652. <https://doi.org/10.1063/1.464913>
 260. Weinhold F, Landis CR, Glendening ED (2016) What is NBO analysis and how is it useful? *Int Rev Phys Chem* 35:399–440.

<https://doi.org/10.1080/0144235X.2016.1192262>

261. Domingo LR, Sáez JA (2009) Understanding the mechanism of polar Diels-Alder reactions. *Org Biomol Chem* 7:3576–3583. <https://doi.org/10.1039/b909611f>
262. Politzer P, Burda J V., Concha MC, et al (2006) Analysis of the reaction force for a gas phase S_N2 process: CH₃Cl + H₂O → CH₃OH + HCl. *J Phys Chem A* 110:756–761. <https://doi.org/10.1021/jp0582080>
263. Van Zeist WJ, Bickelhaupt FM (2010) The activation strain model of chemical reactivity. *Org Biomol Chem* 8:3118–3127. <https://doi.org/10.1039/b926828f>
264. Politzer P, Murray JS, Yepes D, Jaque P (2014) Driving and retarding forces in a chemical reaction. *J Mol Model* 20:1–6. <https://doi.org/10.1007/s00894-014-2351-0>
265. Politzer P, Murray JS, Jaque P (2013) Perspectives on the reaction force constant. *J Mol Model* 19:4111–4118. <https://doi.org/10.1007/s00894-012-1713-8>
266. Burda J V., Murray JS, Toro-Labbé A, et al (2009) Reaction force analysis of solvent effects in the addition of HCl to propene. *J Phys Chem A* 113:6500–6503. <https://doi.org/10.1021/jp9025927>
267. Hammoudan I, Chtita S, Riffi-Temsamani D (2020) QTAIM and IRC studies for the evaluation of activation energy on the C=P, C=N and C=O Diels-Alder reaction. *Heliyon* 6:3–7. <https://doi.org/10.1016/j.heliyon.2020.e04655>
268. Becke AD, Edgecombe KE (1990) A simple measure of electron localization in atomic and molecular systems. *J Chem Phys* 92:5397–5403. <https://doi.org/10.1063/1.458517>
269. Zeeman EC (1976) Catastrophe theory. *Sci Am* 234:65–83. <https://doi.org/10.1038/371683a0>
270. Dominikowska J, Jabłoński M, Palusiak M (2016) Feynman force components: Basis for a solution to the covalent: Vs. ionic dilemma. *Phys Chem Chem Phys* 18:25022–25026. <https://doi.org/10.1039/c6cp03774g>
271. Giovane LM, Barco JW, Yadav T, et al (1993) Kinetic stability of the C60-cyclopentadiene Diels-Alder adduct. *J Phys Chem* 97:8560–8561. <https://doi.org/10.1021/j100135a004>
272. Bento AP, Solà M, Bickelhaupt FM (2005) Ab initio and DFT benchmark study for nucleophilic substitution at carbon (S_N2@C) and silicon (S_N2@Si). *J Comput Chem* 26:1497–1504. <https://doi.org/10.1002/jcc.20261>
273. Poater J, Solà M, Duran M, Robles J (2002) Analysis of the effect of changing the a₀ parameter of the Becke3-LYP hybrid functional on the transition state geometries and energy barriers in a series of prototypical reactions. *Phys Chem Chem Phys* 4:722–

731. <https://doi.org/10.1039/b108910m>
274. Walker M, Harvey AJA, Sen A, Dessent CEH (2013) Performance of M06, M06-2X, and M06-HF density functionals for conformationally flexible anionic clusters: M06 functionals perform better than B3LYP for a model system with dispersion and ionic hydrogen-bonding interactions. *J Phys Chem A* 117:12590–12600. <https://doi.org/10.1021/jp408166m>
 275. Baraban JH, Martin-Drumel M, Changala PB, et al (2018) The Molecular Structure of gauche -1,3-Butadiene: Experimental Establishment of Non-planarity . *Angew Chemie* 130:1839–1843. <https://doi.org/10.1002/ange.201709966>
 276. Sancho-García JC, Pérez-Jiménez AJ, Moscardó F (2001) Description of C(sp²)-C(sp²) rotation in butadiene by density functionals. *J Phys Chem A* 105:11541–11548. <https://doi.org/10.1021/jp0120615>
 277. Engeln R, Consalvo D, Reuss J (1992) Evidence for a gauche minor conformer of 1,3-butadiene. *Chem Phys* 160:427–433. [https://doi.org/10.1016/0301-0104\(92\)80010-S](https://doi.org/10.1016/0301-0104(92)80010-S)
 278. Wubbels GG (2015) The Bell-Evans-Polanyi Principle and the regioselectivity of electrophilic aromatic substitution reactions. *Tetrahedron Lett* 56:1716–1719. <https://doi.org/10.1016/j.tetlet.2015.02.070>
 279. Pla P, Wang Y, Alcamí M (2020) When is the Bell-Evans-Polanyi principle fulfilled in Diels-Alder reactions of fullerenes? *Phys Chem Chem Phys* 22:8846–8852. <https://doi.org/10.1039/c9cp06977a>
 280. de Visser SP, Lin YT, Ali HS, et al (2021) Negative catalysis / non-Bell-Evans-Polanyi reactivity by metalloenzymes: Examples from mononuclear heme and non-heme iron oxygenases. *Coord Chem Rev* 439:213914. <https://doi.org/10.1016/j.ccr.2021.213914>