Local Orbital Analysis of Electronic Correlation in Molecular Systems

Dissertation
for the award of the degree
"Doctor rerum naturalium" (Dr.rer.nat.)
of the Georg-August-Universität Göttingen

within the doctoral program Chemistry
of the Georg-August University School of Science (GAUSS)

submitted by
Xaiza Aniban
from Kalilangan, Bukidnon

Göttingen, 2022

Thesis Committee

Prof. Dr. Ricardo A. Mata
Institute for Physical Chemistry, Georg-August-University Göttingen

Prof. Dr. Inke Siewert
Institute for Inorganic Chemistry, Georg-August-University Göttingen

Prof. Dr. Julia Contreras-García
CNRS, Laboratoire de Chimie Théorique, LCT, Sorbonne Université, Paris, France

Members of the Examination Board

Reviewer:
Prof. Dr. Ricardo A. Mata
Institute for Physical Chemistry, Georg-August-University Göttingen

Second Reviewer:
Prof. Dr. Inke Siewert
Institute for Inorganic Chemistry, Georg-August-University Göttingen

Further members of the Examination Board:

Prof. Dr. Manuel Alcarazo
Institute for Organic and Biomolecular Chemistry, Georg-August-University Göttingen

Prof. Dr. Anna Krawczuk
Institute for Inorganic Chemistry, Georg-August-University Göttingen

Prof. Dr. Daniel Obenchain
Institute for Physical Chemistry, Georg-August-University Göttingen

Prof. Dr. Martin Suhm
Institute for Physical Chemistry, Georg-August-University Göttingen

Date of the oral examination: 13.01.2023

Acknowledgement

I am immensely thankful to Prof. Dr. Ricardo Mata for granting me the opportunity to pursue my doctoral degree in his group. His warm welcome, unwavering support from the beginning till the end, patience and for all the opportunities he paved the way for me have been pivotal for my growth during this exciting phase, both scientifically and personally. I always thought quantum chemistry is very complex and difficult to understand, but under his guidance, I realized that tinkering around the complexity of the field can also be a lot of fun. I especially thank him for patiently walking me through the field and for illustrating to me, in a very creative and fun way, the different quantum chemical concepts. With his guidance, I fulfilled what I wanted in this PhD – understand the intricacies of quantum chemistry and have a lot of fun while doing so!

I extend my gratitude to my thesis advisory committee, Prof. Dr. Inke Siewert and Prof. Dr. Julia Contreras-García. Their support was crucial during this phase, most especially when the pandemic hit. I am also always thankful for the fun discussions during the TAC meetings and outside discussions during the collaborations on different projects. A special thanks to Prof. Dr. Martin Suhm, who is always very welcoming and very generous with imparting his expertise to me and my collaborators on several projects involving benchmarking. Furthermore, I would also like to thank Prof. Dr. Manuel Alcarazo for the very interesting project on helicenes. Also my gratitude to Jun. Prof. Anna Krawczuk and Jun. Prof. Daniel Obenchain for graciously accepting to be a part of my examination board.

Of course, I'd like to thank my collaborators for all the discussions and fun projects I was involved in. Thanks to Beppo Hartwig for the adventure of discovering the impact of dispersion in several diol systems, as well as to Robert Medel for the interesting quest of uncovering the delicate balances in ketone-alcohol systems.

The fascinating projects I pursued during this PhD journey were made achievable due to the invaluable support from funding institutions. I am profoundly grateful to Deutsche Forschungsgemeinschaft (DFG) for their financial support through the RTG 2455 Benchmarking Experiments for Numerical Quantum Chemistry (BENCh). The allure of this specific research area is precisely why I chose the University of Göttingen for my studies. The RTG has brought together a lot of remarkable people from all over the world, and the BENCh members have provided an invaluable support network, fostering enriching discussions and enjoyable gatherings throughout my academic pursuit.

My journey through the PhD was nothing short of incredible, largely due to the amazing members of the Computational Chemistry and Biochemistry group. Thanks to Dr. Jon Uranga and Dr. Benjamin Schröder for patiently helping me navigate (especially when I started) the technical aspects as well as generously generously sharing their knowledge in any scientific query I ask them. A special thanks to Jon for helping me push through my thesis submission, most especially the last-minute manuscript edits. I am also thankful to Dr. Rainer Oswald for helping me get through all the technical difficulties in my calculations and programs. My heartfelt thanks to Lukas Hasecke, who has been there since I started my PhD in 2019, and was instrumental in making my stay in the group truly memorable. Moreover, my heartfelt thanks to Tlektes Zhanabekova, Sophia Bazzi, Anton Römer, Jin Ye, Johannes Kircher, Jun. Prof. Jonny Proppe and Björn Hein-Janke for the engaging discussions, and fun get-together. I must not overlook expressing my gratitude to those who streamlined administrative aspects, ensuring a smoother academic journey - Karin Schoppe, Petra Lawecki and Martina Plaettner.

Big thanks to Dilshana Shanavas Rasheeda and Martin Liebetrau for their unwavering support, remarkable friendship, fruitful scientific (and non-scientific) discussions, and coding challenges. My stay in the Göttingen has been one of a kind because of you both. Especially to Dilshana for all the pep-talks during the ups and downs all throughout. Also thanks to Marius Herbold, Jonas Finkler, Tsz Wai Ko (Kenko) and Ting Yi Chen (Kenny), for all the fun before, during, and after corona. Special mention to Katharina Meyer, for the warm welcome and kind advice how to navigate around.

In addition, I would like to wholeheartedly thank my parents, Leah and Simeon Aniban, for backing me up all the way. They always believed in me no matter what, and were always ready to welcome me with open arms (and loads of good food), no matter how things would have went with my endeavors. Finally, I have reached the pinnacle of my academic journey (*Ma, Pa, doctor na jud ko!*).

Lastly, I would like to thank my loving boyfriend, Paul Czapp, who has been an incredible partner throughout. Not only did he believed in me and my abilities, but he had my back through the highs and lows of this degree. In moments where I felt like giving up, he patiently encouraged me. Most importantly, he made sure I went to the library to write my manuscript. So supportive, isn't he? ;)

My PhD time in Göttingen has been incredibly enriching and exceptionally memorable, and I owe it all to the amazing individuals I mentioned. Once again, a big THANK YOU to each and every one of you for making this journey truly unforgettable!

To God be the glory!

Abstract

Chemical problems nowadays are explored not only with experiments but also with accompanying theoretical studies. The advancement in computer hardware and improved accuracy of quantum chemical algorithms has even facilitated the exploration of chemical problems that are not yet feasible with current experimental facilities. Most interpretations of the theoretical results are based on the calculated numbers, which presents a challenging task. In order to alleviate the difficulty, local descriptors can facilitate the interpretation of different chemical phenomena. This can be very helpful to analyze the intricate energy balance of chemical systems and can efficiently assist in the design of target experiments/molecules. Local correlation serves as an excellent means of providing such descriptors. Furthermore, local correlation methods are highly cost-effective as these efficiently lower the computational cost without sacrificing the accuracy. Importantly, local correlation makes use of localized orbitals bridging classical concepts with quantum chemical results.

In this thesis, several tools were developed using local correlation to provide new approaches to solve diverse chemical problems. Dispersion-free gradient LMP2 was developed to investigate the impact of chirality recognition in diols. This tool provided a clear picture of the impact of dispersion in chirality recognition in both structures and energies of the dimers. On the other hand, the PNO-SCS-LMP2, in tandem with the extended Kitaura-Morokuma (KM) energy decomposition scheme, has helped unlock the delicate energy balance between the dimers of fenchone and fenchol systems. This combination of energy decomposition analysis showed that most of the *SR* systems exhibit a stronger electrostatics than their some of their *RR* counterparts. For the most stable structures (indicated by -I), the dispersion contribution is higher in the *RR*-I than in *SR*-I. This is affirmed by the dispersion interaction density (DID) visualization done on both species. Furthermore, the inversion mechanism of expanded helicenes was also evaluated, and it was observed that the in-fjord substitution, provided that the substituent interacts with the arm, helps in stabilizing the reactant thereby raising the energy barrier of the inversion process. DID visualization also affirmed the presence of moieties interacting due to dispersion.

It was also of interest to examine forces, specifically dispersion in this case, as an overlap between molecules. This gives a different perspective on the interacting fragments and provides a directionality of the force upon interaction. In this thesis, this approach is

called o-DID. Different toy systems were evaluated and revealed the capabilities of the method. This is well complemented by pair orbital analysis, which helps to quantify the extent of dispersion interaction at an orbital level.

Another tool implemented herein is dedicated to aid chemists in pinpointing the specific orbital that needs a multireference treatment. Not all systems can be adequately described by a single determinant. Hence, if some important configurations are neglected, the system's description is already on the wrong footing. To remedy this, the extent of multireference character must be assessed. In this research, the D_2 diagnostic from the OI-MP2 formulation was used to examine the multireference character in an orbital level. It is a targeted approach which can, within the capabilities of D_2 from OI-MP2, examine a potentially problematic orbital.

Overall, a multitude of tools have been developed to enhance the proficiency of chemists in chemical analysis. The diverse applications of local orbital analysis, as presented, offer a refined, targeted approach for both qualitative and quantitative analyses across a spectrum of complex chemical problems.

Contents

Chapter 1

Introduction

The rapid advancement of computer hardware in the last few decades has also brought with it a revolutionary development in the field of chemistry. In the domain of electronic structure calculations, this progress has not only enabled the modeling of increasingly larger systems but also ushered in a plethora of highly accurate computational methods for simulating complex chemical systems. This facilitated a change in the way chemists approach their investigations. Over the years, theoretical calculations became a significant tool to predict chemical properties, assisting in rationalizing the obtained experimental results. At the same time, theoretical chemistry can save a huge effort of experimental trial-and-error, as seen in drug design and material development, for example. A step further than that is conducting pure theoretical studies, exploring chemical phenomena which are often challenging to capture in the experimental bench.

Quantum mechanical calculations have been extremely successful in describing the electronic structure of molecules, particularly isolated molecular systems. Their precision has, in some cases, come very close to high resolution spectroscopic results. The methods developed have not been only tested against the high-level methods available but have also been rigorously compared with existing experimental results, for example the Gn[1–5] and Wn[6–8] methods. This has been a significant step towards a closer collaboration between theory and experiment. However, upon looking closer at these benchmarking practices, more often both parties rely on numbers (e.g., enthalpies of formation, rate constants, atomization energies, bond distances, to mention a few). In fact, a lot of chemical interpretations are based on the numbers obtained from both ends. While this is not necessarily incorrect and serves as a critical link between the theory and experiment, drawing chemical interpretations from a single quantity can be quite challenging.

Despite the advances from both ends of theory and experiment, chemists continue to seek descriptors that will supplement the numerical results with a chemical picture. This proves to be helpful in instances where there is little to no change in a measurable quantity but also has an important effect in terms of structure. In fact, this becomes increasingly relevant for systems with non-covalent interactions. Since these systems are not covalently bonded, a significant change in their structures is not necessarily reflected in, say for example, their relative energies. Thus, experimental approaches to probe their structures, e.g., Fourier transform infrared (FTIR) jet spectroscopy and microwave spectroscopy, are important. Such experiments can be well supplemented by theory, but a deeper analysis regarding the delicate interplay of different forces responsible for the interaction needs more than the typical geometry optimizations.

Local correlation is a powerful tool to extract *ab initio* properties from a molecular system. Not only are these methods very cost-effective without compromising chemical accuracy, but the use of localized molecular orbitals effectively bridges the gap between quantum chemical results and classical chemical concepts. This facilitates the development of a chemical picture that aids chemists in rationalizing results. In this thesis, we utilize local correlation to offer a set of tools for chemists to investigate various chemical phenomena, with a particular emphasis on systems with noncovalent interactions. In the next chapters, these tools are discussed in depth, accompanied by insights into their applications.

In Chapter 3, we examined the significance of dispersion in chirality recognition and molecular inversions. This chapter aims to demonstrate that examining both the energy and the molecular structure of compounds is very crucial for understanding chemical phenomena, particularly chirality recognition in this case. A dispersion-free gradient local 2nd-order Møller Plesset (LMP2) method was implemented to see how dispersion affects chirality recognition. Additionally, local correlation methods were used to isolate and quantify the extent of the influence of this subtle force. Furthermore, local orbital analysis was utilized to identify specific molecular moieties responsible for the dispersion interaction.

After establishing the importance of structure and how it affects the energy as well as the chemical property of the system, Chapter 4 discusses a newly developed tool. This tool allows chemists to connect quantum mechanical results and chemical intuition. Specifically, this tool is exclusive for isolating and viewing overlaps in systems resulting from dispersion interactions. This comes with a quantitative analysis which equips the user with a targeted approach in pinpointing which specific orbital interacts significantly due to dispersion.

In benchmarking, the goal is to attain accurate values for justifiable reasons, rather than relying on error cancellation by chance. However, this is not possible when the calculation starts on an incorrect foundation. In reality, single-reference methods encounter significant limitations if there are one or more relevant electron configurations that need to be considered in the calculation. In Chapter 5, a collection of tools and approaches using local correlation is presented to guide the user on whether a single-reference method is appropriate or if a multireference method is necessary. More importantly, this chapter introduces a targeted approach to diagnose 'problematic' orbital(s) using local orbital analysis.

Chapter 2

Theory Background

2.1 Schrödinger Equation and the Hartree-Fock Approximation

The mathematical description of the behavior of electrons, and thus of chemistry, is all reflected in quantum mechanics (QM). In principle, a property of an individual atom or molecule can be predicted by QM exactly, but in practice, due to their mathematical complexity, QM equations have only been solved exactly for one-electron systems. To remedy this, a multitude of approximations has been developed to approximate the solution for multi-electron problems, some of which will be introduced in the later part of the chapter.

Schrödinger and Heisenberg formulated QM in equivalent ways, but in this chapter, the Schrödinger form will be presented as it is the most familiar to chemists. The time-independent Schrödinger equation is expressed as follows:

$$\hat{H}\Psi = E\Psi. \tag{2.1}$$

In the eigenvalue equation above, the wavefunction Ψ is a function of electrons and nuclear positions, where the description of the electron is a wave (as the name 'wavefunction' implies). The energy E is the eigenvalue which connects to physical observables. The $\hat{H}$ is the Hamiltonian operator, which will consist of a kinetic energy part ($\hat{T}$) and the potential energy part ($\hat{V}$), generally expressed as follows (in atomic units):

$$\begin{aligned}\hat{H} &= \hat{T}_n + \hat{T}_e + \hat{V}_{ne} + \hat{V}_{ee} + \hat{V}_{nn} \\ &= -\sum_{A}^{M} \frac{1}{2m_A}\nabla_A^2 - \sum_{i}^{N} \frac{1}{2}\nabla_i^2 - \sum_{i}^{N}\sum_{A}^{M} \frac{Z_A}{|\mathbf{R}_i - \mathbf{R}_A|} + \sum_{i>j}^{N} \frac{1}{r_{ij}} + \sum_{A>B}^{M} \frac{Z_A Z_B}{R_{AB}},\end{aligned} \tag{2.2}$$

where n are the nuclei, e are electrons, M is the number of nuclei, N is the number of electrons, A, B are nuclei indices and i, j are electron indices. Charges are represented by Z while masses are indicated by m. The first two terms in the equation above represent the nuclear and electronic kinetic energy, respectively. On the other hand, the last three terms are the potential energy between particles.

Note that the above formulation is the time-independent, nonrelativistic Schrödinger equation. As shown in Equation 2.2, the nuclear and electronic motions are separated.

Such decoupling of electronic and nuclear motions is possible because the electrons, due to their very small mass relative to the nuclei, can move much faster. This fundamental assumption is called the Born-Oppenheimer approximation. By virtue of this assumption, the last term in Equation 2.2, $\hat{V}_{nn}$, is constant.

To complete the description of the Equation 2.2, the ∇_i^2 is the Laplacian operator acting on (for example) particle i, as shown below:

$$\nabla_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}. \tag{2.3}$$

Once the wavefunction is established, molecular properties can be determined. This is done by taking the expectation value (denoted with angled brackets) of the operator for a specific property. Energy, for example, is the expectation value of the Hamiltonian operator (for a normalized function) given by

$$\langle E \rangle = \int \Psi^* \hat{H} \Psi. \tag{2.4}$$

The type of QM approximations discussed in this book is the *ab initio* (Latin for ‘from the beginning’) methods. This set of approximations is applied to computations which are derived directly from theoretical principles and do not include experimental data. The Hartree-Fock (HF) approximation is the most common type of *ab initio* calculation, usually serving as the foundational calculation for post-HF methods to account for electron correlation. In this method, the primary approximation is the mean field approximation, which means electron-electron repulsion is treated in an ‘average’ manner. In other words, each electron is considered to be moving in an electrostatic field of nuclei and the average field of the rest ($N-1$, where N is the total number of electrons) of the electrons.

Taking into account the Pauli principle, an HF wavefunction can be built from molecular orbitals using an anti-symmetrized product represented by a Slater determinant, as shown below:

$$\Psi^{\mathrm{HF}}(\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n}) = \frac{1}{\sqrt{N!}} \begin{bmatrix} \psi_1(\mathbf{x_1}) & \psi_2(\mathbf{x_1}) & \dots & \psi_N(\mathbf{x_1}) \\ \psi_1(\mathbf{x_2}) & \psi_2(\mathbf{x_2}) & \dots & \psi_N(\mathbf{x_2}) \\ \vdots & \vdots & \dots & \vdots \\ \psi_1(\mathbf{x_N}) & \psi_2(\mathbf{x_N}) & \dots & \psi_N(\mathbf{x_N}) \end{bmatrix}, \tag{2.5}$$

where each function used in the determinant is a spin orbital consisting of a spatial and spin component ($\mathbf{x_i} = \{\mathbf{r_i}, \mathbf{s_i}\}$). In *closed-shell* representation where the number of α and β spin electrons are the same, the spin orbitals can be integrated out and the same spatial orbitals $\{\phi_i\}$ can be used for both sets. Note also that the wavefunction is reflected in a single Slater determinant, earning the classification of a single-reference method.

Using Dirac's *bra-ket* notation and with the help of Slater-Condon rules, the HF energy can be obtained in a similar manner as in Equation 2.4:

$$\begin{aligned} E^{\mathrm{HF}} &= \langle \Psi_{\mathrm{HF}} | \hat{H} | \Psi_{\mathrm{HF}} \rangle \\ &= 2\sum_{i=1}^{N/2} \langle i | \hat{h} | i \rangle + \frac{1}{2}\sum_{i=1}^{N/2}\sum_{j=1}^{N/2} [2(ii|jj) - (ij|ji)], \end{aligned} \tag{2.6}$$

where the $\hat{h}$ is the one-electron operator, $\hat{T}_e + \hat{V}_{ne}$ from Equation 2.2. The first term in the above equation (Eq. 2.6) is the one-electron integral, while the second term are composed of two-electron integrals, where $(ii|jj)$ is the Coulomb expression (denoted by J) while $(ij|ji)$ represents the exchange term (denoted by K). On one hand, the one-electron integral illustrates the energy of the electron due to its motion ($\hat{T}_e$) and the attraction of the nuclear core $\hat{V}_{ne}$. On the other hand, the second sum represents the Coulombic repulsions resulting from the interaction of the electron with all the other electrons within the atom or molecule, plus exchange non-classical terms.[9]

Hartree-Fock is a variational method, which means that the approximate energies calculated are all equal to or greater than the exact energy. This can be used to improve the form of the wavefunction, and the lower the energy obtained, the closer it is to the 'correct' description of the system. To minimize the energy, a gradient can be built from Equation 2.6 under the constraint that the orbitals are orthonormal. This can be done by using a Lagrange functional, which can be built in the form:

$$\langle \partial i | \hat{h} | i \rangle + \sum_{j}^{N/2} [2(\partial ii|jj) - (\partial ij|ji)] - \sum_{j}^{N/2} \varepsilon_{ji} \langle \partial i | j \rangle = 0. \tag{2.7}$$

This can be simplified by defining the Fock operator,

$$\hat{f}\left|i\right\rangle = \sum_{j}^{n/2} \varepsilon_{ji}\left|j\right\rangle , \tag{2.8}$$

where $\hat{f}$ is

$$\begin{aligned}\hat{f}(1) &= \hat{h}(1) + \sum_{j=1}^{N/2}[2\hat{j}_j(i) - \hat{k}_j(i)] \\ &= \hat{h}(i) + \hat{g}(i).\end{aligned} \tag{2.9}$$

The operator $\hat{j}_j(i)$ is the Coulomb operator and the $\hat{k}_j(i)$ is the exchange. It is important to note that the Fock operator is a one-electron operator.

Essentially, HF theory provides a stepping stone on the way to the exact solution of the Schrödinger equation. This approach provides a very well-defined energy, one which can be converged in the limit of an infinite basis set. Unfortunately, this does not mean that it does not suffer from certain chemical limitations. The pinnacle of its limitation is the one-electron nature of the Fock operators. The only 'correlation' accounted for so far in Equation 2.6 is the exchange. All the rest is ignored. Of course, it is an interesting point to make, just how important other types of correlation are for other molecular properties? HF energy commonly recovers 99% of the total energy, but the remaining 1% which describes the instantaneous correlation as electrons move with respect to each other, is an important factor to take into account to when describing different chemical phenomena. Dispersion, for example, does not exist at all in the HF-approximation and in the recent years, this force has been proven to be a relevant driving force in a number of chemical problems - a point that will be emphasized in the next chapters. This, and all the other wavefunction-based methods discussed in the succeeding sections, belong to the field of wavefunction theory (WFT).

Note that the HF method is based on several constraints. First, the Born-Oppenheimer approximation, which decouples nuclear and electronic motion. Second, the mean field approximation, which assumes that electronic repulsion is due to the movement of every electron in an average electrostatic field. Third is the single-reference approximation, where it is assumed that a single determinant is enough to describe the electronic configuration of the system.

2.2 Electron Correlation

In HF theory, the interaction of two electrons can be seen in the second term of Equation 2.6, which contains the Coulomb (J) and exchange (K) terms. In this treatment, the handling of electron-electron repulsion is via electron movement in a smeared-out, average electrostatic field due to the rest ($N - 1$) of other electrons. This also means that the probability of an electron having a particular set of spatial coordinates at some moments does not depend on the coordinates of the other electrons in the same moment. However, this model is flawed. In reality, every electron moves under the influence of repulsion due to the presence of other electrons (i.e., their movement is 'correlated' with one another), not of an average electrostatic field. As a consequence, in real atom or molecule, the electrons are better at dodging each other. Such behaviour means that the electronic energy in reality is lower (more negative) than the one predicted by the HF calculation. Even with the biggest basis set approaching the HF limit, the HF method still overestimates the electron-electron repulsion because of its failure to treat electron correlation properly.

Correlation energy is a measure of the extent to which any *ab initio* calculation fails to deal with electron correlation perfectly. Löwdin, in a canonical exposition, provided the definition *"... The correlation energy for a certain state with respect to a specified Hamiltonian is the difference between the exact eigenvalue of the Hamiltonian and its expectation value in the HF approximation for the state under consideration."*[10] In other words, the correlation energy is the difference between the restricted HF (RHF) energy and the energy calculated by some 'perfect' quantum mechanical procedure with the same basis set. Mathematically, this is expressed as:

$$E_{\text{corr}} = E_{\text{exact}} - E_{\text{RHF}}. \tag{2.10}$$

From the definition above, the correlation energy is always negative since $\mathrm{E}_{\text{exact}}$ is more negative than E_{RHF}. Essentially, the correlation energy is the energy that the HF procedure failed to account for. Of course, this is under the assumption that relativistic effects and other effects like spin-orbit coupling are negligible (same Hamiltonian).

2.2.1 Dynamic and Static Correlation

In terms of the 'permanence' of electrons avoiding each other, electron correlation can also be classified as *static* and *dynamic*. When electrons 'instantly' avoid each other, such as in the case of those occupying the same spatial orbital, that contributes to the *dynamic* *correlation*. If the avoidance of electrons is in a more 'permanent' basis, such as those with different spatial orbitals, that is referred to as the *static correlation*. The latter becomes more relevant if there are other configurations close in energy to the ground state, thus sometimes called *near-degeneracy effect*. Classic example is the H_2 molecule under an RHF description. At the equilibrium bond distance, the correlation is mainly dynamic. However, at its dissociation limit, H_2 is dominated by static correlation. At this point, two relevant configurations must be considered so as to properly calculate the energy of the system.

In some chemical cases, like stretching of H_2, unrestricted Hartree-Fock procedure (UHF) can partially model static correlation by relaxing the constraint that the α and β orbitals are equivalent, yielding a lower total energy.[11,12] In case of O_3 at its equilibrium geometry, UHF recovers 10% of the correlation energy.

Since UHF can capture part of the static correlation, this led Pople to propose an alternative definition of the correlation energy:[13]

$$E_{\text{corr}} = E_{\text{exact}} - E_{\text{HF}}, \tag{2.11}$$

where $\mathrm{E_{HF}}$ is the lowest energy of *any* single determinant wavefunction. Pople correlation energies may be more robust than the Löwdin ones despite avoiding anomalous behaviour in multireference systems.[14]

Many workers have found it useful to subdivide the electron correlation into dynamic and static.[10]

$$E_{\text{corr}} = E_{\text{dyn}} + E_{\text{stat}}. \tag{2.12}$$

The accuracy of electronic structure calculations relies heavily on the proper description of electron correlation. In this section, a thorough discussion about the differences in both types of correlation, and how they are accounted for, are presented.

Dynamic Correlation

Strictly speaking, HF method allows a certain degree of electron correlation - i.e., two electrons of the same spin cannot be in the same place at the same time (Pauli's exclusion principle). Since there is zero probability at any moment that two electrons of the same spin are at the same point in space, given that the wavefunction is continuous, there should be a smooth decrease of the probability of finding that at the same space versus their separation.

The inability of the HF wavefunction to model interelectronic cusps[15] and dispersion interactions[16, 17] is due to dynamic correlation.[18] Recall that under this assumption (as discussed earlier), the electrons do not see other electrons individually but just as an average electrostatic field. This type of correlation is regarded as a universal contribution. Simply put, this type of correlation arises from the need for mixing the Hartree-Fock state with higher-order excited states and is a correction to the second constraint imposed on HF method, i.e., the mean field approximation. This means that there is a large number of excited state determinants with comparable, small occupations which are relevant to properly describe the molecular system.

If a quantitative prediction of experimental observables is desired, a significant consideration of dynamic correlation is required. Usually, this can be treated using post-HF methods with two dominant classes of methods: perturbative and coupled cluster (CC) approaches. These will be discussed further in succeeding sections.

Static Correlation

Systems with stretched bonds or multiple bonds, electronically excited or partially occupied degenerate orbitals have a considerable static correlation.[19] This arises from the near-degeneracies of the HF occupied and virtual orbitals. These types of systems are poorly described by a single Slater determinant. This shows that static correlation energy (also called non-dynamical, near-degeneracy, left-right, or first-order) is a system-dependent correlation effect. This is mainly due to the inaccuracy introduced by the third constraint to the HF method, i.e., the description of the wavefunction using a single determinant.

Hollet and Gil[18] classified static correlation into two flavors: Type A as *absolute near-degeneracy* and Type B as *relative near-degeneracy.* The distinction between these types lies in their susceptibility to different electronic structure methods. Specifically, Type A static correlation can be effectively captured by UHF, while Type B static correlation eludes such treatment. According to this classification, stretched H_2 molecule is dominated by Type A static correlation while Be-like ions with large nuclear charge (Z) are dominated by Type B static correlation.

In systems with 'quasidegenerate' electronic structures, several electronic configurations need a proper description. If a single determinant assumption is required, this can lead to significant errors. In such cases, not even a very good method to capture dynamic correlation can help. Multireference (MR) treatments will be necessary, for example full configuration interaction (FCI), multireference treatment of different flavors (e.g., MCSCF, MRCI, MRCC) and complete active space (CAS) approaches of different varieties (CASSCF or CASPT2).

2.2.2 Møller-Plesset Perturbation Theory

In perturbation methods, electron correlation is treated as perturbation or 'fluctuation' or 'small deformation' of the reference HF energy. The Hamiltonian is written as:

$$\hat{H} = \hat{H}^{(0)} + \lambda \widehat{H^{(1)}}, \tag{2.13}$$

where $\hat{H}^{(0)}$ is the unperturbed Hamiltonian, λ is the scaling parameter defining the strength of the perturbation operator $\widehat{H^{(1)}}$. The latter accounts for the electron-electron repulsion beyond mean field approximation.

There are two approaches to use the perturbation theory applied to chemical systems. One is *Brillouin–Wigner* perturbation theory which, unfortunately, is not size consistent and the *Rayleigh-Schrödinger perturbation theory*, which is size consistent. In the latter method, the total energy is computed in a step-wise manner and has a nice feature that the correction of a particular order is independent of the maximum order chosen. For a detailed derivation of both perturbation methods, see Reference 20.

Møller-Plesset perturbation theory[21] (MPPT) is a version of Rayleigh-Schrödinger perturbation theory where the electronic Hamiltonian is partitioned as in Equation 2.13 with:

$$\hat{H}^{(0)} = \sum_{i}^{N} \hat{f}(i), \text{and} \tag{2.14}$$

$$\hat{H}^{(1)} = \hat{H} - \hat{H}^{(0)}. \tag{2.15}$$

Note that $\hat{f}$ is the Fock operator which represents the zero-order operator and the $\hat{H}^{(1)}$ as the perturbation. The wavefunction and energy is expanded via Taylor series, as follows:

$$\begin{aligned} |\Psi\rangle &= \sum_{i=0} \lambda^i |\Psi^{(i)}\rangle \text{, and} \\ E &= \sum_{i=0} \lambda^i E^{(i)}. \end{aligned} \tag{2.16}$$

Inserting Equation 2.16 in the Schrödinger equation provided in Equation 2.1, the general expression for such expansion is the following:

$$(\hat{H}^{(0)} - E^{(0)}) |\Psi^{(0)}\rangle + \lambda[(\hat{H}^{(0)} - E^{(0)}) |\Psi^{(1)}\rangle + (\hat{H}^{(1)} - E^{(1)}) |\Psi^{(0)}\rangle] + ... = 0. \tag{2.17}$$

In this perturbation approach, one can carry out calculations through a given order of n, and is referred to as MPn approximations. Note that for each power of λ, the terms associated with it must be zero.

Applying the standard procedure of perturbation theory, the following breakdown of energies is obtained:

$$E_{\text{MP}}^{(0)} = \langle\Psi_{\text{HF}}| \hat{H}^{(0)} |\Psi_{\text{HF}}\rangle = \sum_i \varepsilon_i, \tag{2.18}$$

$$E_{\text{MP}}^{(1)} = \langle\Psi_{\text{HF}}| \hat{H}^{(1)} |\Psi_{\text{HF}}\rangle = \frac{1}{2} \sum_{ij} [2(ii|jj) - (ij|ji)], \tag{2.19}$$

$$E_{\text{MP}}^{(2)} = \sum_{\mu \neq 0} \frac{\left| \langle \Psi^{HF} \mid \hat{H}^{(1)} \mid \Psi^{(1)} \rangle \right|^2}{E^{(0)} - E^{(1)}} \tag{2.20}$$

where μ is the eigenstate index and $\mu \neq 0$ refers to determinants which are constructed from different spin orbitals than the ground state determinant, and $\Psi^{(1)}$ refers to the excited determinants. The Fock operator represents the zero-order operator, thus the zeroth-order energy is basically the sum of occupied orbital energies, as reflected in Equation 2.18. Equation 2.19 is the first-order energy, whose operator can be interpreted as the mean-field approximation.[22] The sum of Equations 2.18 and 2.19 is basically the Hartree Fock energy. This means that the inclusion of explicit electron-electron correlation via perturbation only starts to appear from the second-order energy reflected in Equation 2.20.

By virtue of Brillouin's theorem and orthonormality conditions between orbitals, there is no coupling between the reference wavefunction and singly excited determinants $|\Psi_i^a\rangle$. Double excitations $|\Psi_{ij}^{ab}\rangle$, on the other hand, contribute to the energy. By integrating out the spin component, the total MP electronic energy with 2nd order correction (MP2) can be expressed as:

$$\begin{aligned} E^{\mathrm{MP2}} &= E_{\mathrm{MP}}^{(0)} + E_{\mathrm{MP}}^{(1)} + E_{\mathrm{MP}}^{(2)} \\ &= E^{\mathrm{HF}} + \sum_{i<j} \sum_{a<b} \frac{|\,(ia\|jb)\,|^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}, \end{aligned} \tag{2.21}$$

where i, j represents occupied orbitals and a, b the virtual ones. The term $(ia\|jb)$ simply means $(ia \mid jb) - (ib \mid ja)$. As an extension of Rayleigh-Schrödinger perturbation theory, MPPT is size-extensive, which means it scales properly with the size of the system. This method, however, is non-variational since the wavefunction is not optimized by going for the lowest energy, therefore it does not provide an upper bound for the ground state energy of the Schrödinger equation. Despite that, in the scope of wavefunction theory approaches, MP2 is the 'go-to' method to integrate correlation with a good compromise in cost and accuracy. By all means, the order of correction can be increased (using MPn notation, e.g., MP3, MP4 ... MPn), but this does not guarantee convergence. In fact, several studies[22, 23] show that the energy oscillates and sometimes diverges with the inclusion of higher excitations as well as being affected by the use of augmented basis functions.

The inclusion of MPn algorithms in computer programs since the mid-1970s made the approach easily accessible to chemists. Rapid development and implementation greatly benefited from the competition of the Bartlett group in Florida and the Pople group in Pittsburgh.[24] Its availability made it easier for chemists to deploy chemical investigations beyond mean-field approximations. In practice, nowadays, higher order of MP methods (MP3 and beyond) are not often used due to steep scaling. However, MP2 is still very useful because even if it is less accurate than many DFT methods,[25] it still has certain advantages. For example, in cases of understanding dispersion forces or charge-transfer processes. In fact, in the last 30 years, MP methods have been greatly developed to reduce their scaling and increase their accuracy. A good overview of such developments can be read in the review of Cremer.[24]

Spin Component Scaled MP2 (SCS-MP2)

In the original MP2 ansatz, the $E^{(2)}$ in Equation 2.20 and 2.21 can be reformulated as:

$$E^{(2)} = E_{ss}^{(2)} + E_{os}^{(2)}, \tag{2.22}$$

where ss is the same-spin or parallel spin ($\alpha\alpha, \beta\beta$,'triplet') spin of electrons while os is the opposite spin or antiparallel-spin ($\alpha\beta$, 'singlet') electron pairs. Integrating out spin component, terms of Equation 2.22 can be expressed as:

$$\begin{aligned} E_{ss}^{(2)} &= \sum_{ij}\sum_{ab} \frac{(ia \mid jb)[(ia \mid jb) - (ib \mid ja)]}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}, \\ E_{os}^{(2)} &= \sum_{ij}\sum_{ab} \frac{(ia \mid jb)(ia \mid jb)}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}. \end{aligned} \tag{2.23}$$

Accordingly, MP2 is biased toward same spin (ss) excitations and correlations. This results in inaccurate quantification of several experimental observables (e.g., atomization energies, reaction energies, etc.). In response to this, Grimme[26] proposed in 2003 to scale spin components ss and os separately, yielding

$$E_{\text{SCS-MP2}}^{(2)} = \frac{1}{3}E_{ss}^{(2)} + \frac{6}{5}E_{os}^{(2)}. \tag{2.24}$$

In general, the separate scaling significantly improved the MP2 performance (see their original paper in Reference 26 for the mean absolute error values). SCS-MP2 is still a single-reference method, so in cases where static correlation is important, multiconfigurational treatment is still advised. This spin component scaling methods are also extended to SCS-MP3, and further chemical properties have been examined.[27–29]

2.2.3 Coupled-Cluster Theory

It is well known that one of the pitfalls of configuration interaction (CI) method is its lack of size consistency and its slow convergence toward the full CI (FCI) limit. To remedy this, the coupled-cluster (CC) model was introduced, a mathematically elegant technique to estimate electron correlation pioneered by Čížek in 1966.[30] This particular model represents a nonlinear but separable parametrization of the correlated electronic state. It is size consistent and size extensive, however, it is not variational.

The main assumption of CC theory is that the full-CI wavefunction (within the basis set approximation) can be described as:

$$|\Psi^{\mathrm{CC}}\rangle = e^{\hat{T}} |\Psi^{\mathrm{HF}}\rangle , \tag{2.25}$$

where

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + ...\hat{T}_N. \tag{2.26}$$

The subscript represents the excitation state while N is the total number of electrons. For example, $\hat{T}_1$ and $\hat{T}_2$ can be expressed as

$$\begin{aligned} \hat{T}_1 |\Psi^{\mathrm{HF}}\rangle &= \sum_i \sum_a t_i^a |\Psi_i^a\rangle , \text{and} \\ \hat{T}_2 |\Psi^{\mathrm{HF}}\rangle &= \sum_{i<j} \sum_{a<b} t_{ij}^{ab} |\Psi_{ij}^{ab}\rangle . \end{aligned} \tag{2.27}$$

Recall that i, j are occupied orbitals while a, b are virtual orbitals. The terms t_i^a and t_{ij}^{ab} are single and double excitation amplitudes, respectively, which are basically the coefficients in front of determinants.

The exponential excitation operator $e^{\hat{T}}$ from Equation 2.25 is defined by the Taylor expansion:

$$e^{\hat{T}} = 1 + \hat{T} + \frac{1}{2!}\hat{T}^2 + \frac{1}{3!}\hat{T}^3 + ..., \tag{2.28}$$

where $\hat{T}$ is expressed in Equation 2.26. Collecting the terms of the same excitation number, one can obtain:

$$|\Psi^{\mathrm{CC}}\rangle = [\underbrace{1 + \hat{T}_1}_{\text{singles}} + \underbrace{(\hat{T}_2 + \frac{1}{2}\hat{T}_1^2)}_{\text{doubles}} + \underbrace{(\hat{T}_3 + \hat{T}_1\hat{T}_2 + \frac{1}{6}\hat{T}_1^3)}_{\text{triples}} + ...] |\Psi^{\mathrm{HF}}\rangle . \tag{2.29}$$

From Equation 2.29, two types of excitations can be observed. The excitations $\hat{T}_1, \hat{T}_2, \hat{T}_3, ...$

are referred to as *connected excitations* and usually have large contributions. Via Slater-Condon rules, $\hat{T}_1$ has a small contribution because singles only come in the second order wavefunction, but $\hat{T}_2$ directly connects to the HF configuration and is the dominant contributor to the correlation energy. It is expected that $\hat{T}_3$ is the second largest contributor, and for well-behaved systems, the higher excitations have lesser contribution to the energy. On the other hand, excitation operators which are a product of two or more excitation operators (e.g., $\frac{1}{2}\hat{T}_1^2$, $\hat{T}_1\hat{T}_2$) are called *disconnected excitations.* Assuming the normal behaviour of the system, the disconnected excitation $\frac{1}{2}\hat{T}_1^2$ will have the third largest contribution. It is important to note that one of the advantages of coupled cluster theory is even if the wavefunction only has singly and doubly excited determinants, it is possible to account for some higher excitations via disconnected excitations. So even with CCSD (coupled cluster singles and doubles), one can obtain a large combination of higher excitations.

The next challenge is the determination of the cluster amplitudes for all the operators included in the specific CC model. The standard implementation follows the usual procedure of left-multiplying the Schrödinger equation by trial wavefunctions expressed as determinants from HF orbitals. This results in a set of coupled, non-linear equations in the amplitudes which must be solved iteratively. With the amplitudes available, the coupled cluster energy is calculated as:

$$E^{\mathrm{CC}} = \langle \Psi^{\mathrm{HF}} | \, \hat{H} \, | e^{\hat{T}} \Psi^{\mathrm{HF}} \rangle \, . \tag{2.30}$$

The scaling of CCSD is $\mathcal{O}(\mathcal{N}^6)$ while the inclusion of the full, connected triples $\hat{T}_3$ in CCSDT increases the scaling to $\mathcal{O}(\mathcal{N}^8)$, making the latter very computationally costly and intractable for all but small molecules. Several approaches were done to include connected triples without incurring much cost. One of the most robust methods was introduced by Raghavachari *et al.*[31] in 1989 and is called CCSD with perturbative triples, CCSD(T). This is currently regarded as the *gold standard* of single-reference calculations in quantum chemistry. The latter method has a scaling of $\mathcal{O}(\mathcal{N}^7)$.

In the past decades, notable developments were made in CC methods. This includes the use of localized orbitals exploiting the short-range character of the dynamical electron correlation, which decays as r_{ij}^{-6}. When local basis functions are used, the integrals over distant pairs (in two electron integrals) can be neglected, thereby lowering the computational cost.[32–37] On the other hand, standard CC methods are still plagued with problems presented by the electron-electron cusp which hampers the fast convergence of CC to the full basis set limit. This is caused by discontinuous higher derivatives of the wavefunction where the coordinates of the two particles coincide. This can be removed by including terms in the wavefunction that depend on the interelectronic distance r_{12}. Explicit inclusion of r_{12} or f_{12} ansatz can significantly improve convergence (see review of Ma and Werner[38] for an extensive list of references). PNO-LCCSD(T)-F12,[39,40] which utilizes explicit correlation using f_{12} ansatz, was used in several investigations in Chapter 3 for high level energy calculations.

2.2.4 Multireference Diagnostics

The methods described above – HF, MP2, CC – are generally suitable if a single determinant is enough to describe the system. Unfortunately, as discussed in Section 2.2.1, some systems need more than one determinant to properly get the qualitative description of the system, which in turn is crucial for any quantitative results to be calculated. A classic case is that of O_3. Looking at its enthalpy of formation ΔH_f°, a deviation of ~3 kcal mol^{-1} using CCSD(T)/CBS when compared to the experimental result, 34.10 ± 0.4 kcal mol^{-1}.[7,41] It obviously provides an insufficient description of O_3 at its molecular ground state.

Unfortunately, multireference methods have a steep scaling with respect to the system size. Large systems that can be treated quickly by single-reference methods may be computationally unviable with multireference approaches due to the inclusion of all chemically relevant electrons and orbitals into the active space. Although restricted active space (RAS), generalized active space (GAS) or occupation restricted multiple active space (ORMAS) can be utilized, the selection of active space typically relies on the user's chemical intuition which can significantly affect the predicted properties.

Before the deployment of the computationally expensive and cumbersome multireference calculations, several molecular diagnostic tools are available to preliminarily evaluate the need of an multireference treatment. In this section, the general idea of some diagnostic tools relevant to Chapter 5 are presented. The diagnostics discussed here are mainly those which are derived from the behaviour of single-reference methods.

C_0 and C_0^2 Diagnostics

A typical diagnostic tool is the absolute value of the leading coefficient C_0 or its weight C_0^2 in a CISD or CASSCF calculations.[42–45] This indicates the contribution of the dominant configuration state function to the wavefunction. Values of $C_0 \leq 0.95$ or $C_0^2 \leq 0.90$ is an indication of significant multireference character.[42] If the coefficient values are from the reference HF configuration in a CISD wavefunction, the diagnostic might be biased towards the HF determinant of a pathological system which can result in a misleadingly large coefficient for the HF determinant.[43] On the other hand, if the calculation is based on the CASSCF wavefunction, the weight of the principal configuration is only considered reliable if CASSCF is feasible and includes all the critical orbitals in the active space. The requirement of large active spaces can lead to computationally prohibitive calculation and small active spaces may lead to unreliable diagnoses.

Moreover, this diagnostic gets unreliable with increasing system size. In the case of a system of non-interacting water molecules C_0^2 approaches zero in the limit of infinite number of monomers.[22] However, this is a well known single-reference case. This means that a threshold limit for classifying multireference character based solely in C_0 is difficult to define.

Total Atomization Energy Diagnostic, $\%\mathrm{TAE_e}$

Total atomization energy (TAE) is calculated as the difference in the electronic energy (same basis set and level of theory) between the system and all its component dissociated atoms, shown below:

$$\mathrm{TAE} = \left(\sum_{i}^{N} E_i\right) - E_{\text{total}}. \tag{2.31}$$

Martin and co-workers,[8] when developing their W4 composite method, proposed that the percentage of the connected quadruple and quintuple excitations contribution $\%\mathrm{TAE_e[T_4 + T_5]}$ to the total atomization energy is a good *a posteriori* indicator of the extent of non-dynamical correlation. This is based on the performance of the $\%\mathrm{TAE_e[T_4 + T_5]}$ versus the largest double excitation amplitude and HOMO-LUMO occupation number.[8,46] However, there is a formidable cost to go to quadruple and quintuple excitations, and they found a strong correlation between $\%\mathrm{TAE[(T)]}$ and $\%\mathrm{TAE_e[T_4 + T_5]}$ ($R^2 = 0.941$, provided the removal of BeO and MgO). This led them to propose $\%\mathrm{TAE[(T)]}$ as an affordable *a priori* energy-based diagnostic to assess non-dynamical correlation of the system. Such diagnostic is defined as follows:

$$\%\mathrm{TAE_e[(T)]} = 100 \times \left(\frac{\%\mathrm{TAE_e[CCSD(T)]} - \%\mathrm{TAE_e[CCSD]}}{\%\mathrm{TAE_e[CCSD(T)]}}\right) \tag{2.32}$$

Using the diagnostic in Equation 2.32, the following threshold was proposed: %TAE$_{\mathrm{e}}$[CCSD(T)] less than 2% are appropriately described by single-reference methods, 2-5% indicates mild static correlation, 5-10% hints at moderate static correlation and if the value exceeds 10%, an multireference treatment is needed.[46] Even if this diagnostic has the capability to predict the contributions from T_4 and T_5 excitations and how important these higher excitations to the system in question, the neccessity to go to CCSD and CCSD(T) can be cost-prohibitive for larger molecules.

T_1 Diagnostic

One of the oldest and most commonly used diagnostics (even with large systems nowadays) is the T_1 diagnostic, proposed by Lee and Taylor in 1989.[43] In a different paper,[47] it was already observed that the Euclidean norm of t_1 gave large values for problematic systems and smaller values for the well behaved ones. Further studies regarding its reliability of prediction was then expanded by Lee and Taylor. This is based on the single excitation amplitudes t_1 from the CCSD wavefunction. Formally, the diagnostic is defined as follows:

$$\mathrm{T}_1 = \sqrt{\frac{\mathbf{t}_1 \cdot \mathbf{t}_1}{\mathrm{n}_{\mathrm{corr}}}}, \tag{2.33}$$

where $\mathbf{t}_1$ is the vector of t_1 and n_{corr} is the number of correlated electrons. This diagnostic basically quantifies the averaged value of the single excitation amplitudes. If the values of T_1 exceeds 0.02, this diagnostic indicates that a single-reference method may not be reliable for the system, and will certainly not yield accurate results. However, since T_1 diagnostic is an averaged value, the non-homogeneity of the t_1 vector is not reflected or possible singular t_2 amplitudes. This diagnostic has been extended to open shell CC theory in a consistent manner with the original definition to that of closed shell CCSD.[48]

T_2 Diagnostic

One significant limitation of T_1 diagnostic previously described is the need to access the single excitation amplitude from the CCSD wavefunction, which in 1990s, was still a formidable task. This prompted an extension to have the same diagnostic tool but from perturbation theory, proposed by Lee *et al.*[49] The t_1 amplitudes from a full coupled-cluster wavefunction may be thought of as a sum of the singles coefficients from the mth-order wavefunctions of the MBPT

$$t_1 = \sum_m u_1^{(m)} \lambda^m, \tag{2.34}$$

where λ is the perturbation parameter and m is the order. Based on Equation 2.33, the T_2 diagnostic can be defined as

$$\mathrm{T}_2 = \sqrt{\frac{\mathbf{u}_1^{(2)} \cdot \mathbf{u}_1^{(2)}}{\mathrm{n}_{\mathrm{corr}}}}. \tag{2.35}$$

When using canonical orbitals within the context of restricted MP and restricted open shell MP (ROMP), $\mathbf{u}_1^{(2)}$ can be obtained by

$$u_1^{(2)} \equiv u_i^a = \frac{1}{f_i - f_a} \left(\sum_{cdk} (ac|kd) t_{ik}^{cd} - \sum_{clk} (ck|il) t_{lk}^{ac} \right). \tag{2.36}$$

Recall that t_{ij}^{ab} are the double excitation coefficients from the first order wavefunction, while i, j and a, b are the occupied and unoccupied orbitals, respectively. Lastly, f is the Fock operator. A more detailed explanation as to the derivation of the single excitation amplitudes via MPBT can be found in the article of Lee.[49] For this specific diagnostic, the following guideline was established: $\mathrm{T}_2 \leq 0.012$ indicates an appropriate description by low orders of perturbation theory, if T_2 is between 0.012 and 0.015, care must be taken when analyzing results from low order MBPT and finally, if $\mathrm{T}_2 \geq 0.015$, low order MBPT methods will not likely yield accurate results, except in cases where there is significant error cancellation. One problem of T_2 is that it is system size dependent. T_2 values have been calculated for a series of molecules - HNC, CH_3NC, C_2H_5NC, C_3H_7NC - where H is basically replaced by larger alkyl groups. Results gave different values, i.e. 0.0146, 0.0122, 0.0108 and 0.0099, respectively.[50] The proposed threshold of the diagnostic suggests that MP2 should be reliable for the larger molecules in the series but probably not for the smallest one. This dependence makes it unsuitable for comparing molecules of different size.

A quick note on its computing time. The MP2 T_2 diagnostic, due to the contractions done in Equation 2.36 to obtain the single excitation amplitudes, requires more operations than the MP2 energy evaluation itself. However, it becomes free if the analytic gradient is computed.[49]

D_1 Diagnostic from MP2 and CCSD

To overcome problems with size inconsistency, Janssen and Nielsen[50] proposed a new diagnostic D_1 using single excitations (thus the subscript 1) arising from MP2 and CCSD wavefunctions. This is basically the same set of excitations as the previous ones, except that instead of using the Euclidean norm (or its equivalent Frobenius norm), it uses a 2-norm matrix. This means that instead of the averaged values of excitation amplitudes, utilizing the 2-norm matrix is closely related to the value of the largest single excitation amplitude. Simply put, D_1 is defined as follows:

$$\mathrm{D}_1 = \sqrt{\lambda_{max}(\mathbf{T}\mathbf{T}^{\mathrm{T}})}. \tag{2.37}$$

The elements of the matrix $\mathbf{T}$ are the single excitation amplitudes from MP2 and CCSD. These amplitudes are obtained the same way as T_1 and T_2 single excitation amplitudes described previously. The following cut-offs were proposed for the D_1 diagnostic: single-reference methods are generally applicable if $\mathrm{D}_1(\mathrm{MP2}) \leq 0.015$ and $\mathrm{D}_1(\mathrm{CCSD}) \leq 0.020$. MP2 and CCSD methods are inadequate if $\mathrm{D}_1(\mathrm{MP2}) \geq 0.040$ and $\mathrm{D}_1(\mathrm{CCSD}) \geq 0.050$. In between values of the defined thresholds indicate that MP2 and CCSD theory generally perform well, but caution must be taken for systems where factors other than orbital relaxation could affect the quality of the reference wavefunction. This was initially extended to open shell CCSD wavefunction by Leininger *et al.*[51] and then revised by Lee.[52]

D_1 diagnostics are invariant with respect to orbital rotations and are not size dependent (size-intensive). The latter property makes it appropriate for large systems and for system comparison with different sizes.

D_2 Diagnostic from MP2 and CCSD

There have been concerns regarding the reliability of single excitation amplitudes as a metric to detect the inadequacy of the HF reference determinant, especially when this is due to the low-lying double excited states. Thus, it has been suggested to utilize the largest double substitution amplitude as a diagnostic.[53] However, it has been demonstrated[50] that this diagnostic is dependent on the system size and lacks invariance to orbital rotations within each of the occupied and virtual spaces, making it inappropriate for systematic comparison of different molecules.

To address this issue, double excitation amplitudes from MP2 and CCSD were utilized in a formulation similar to D_1, i.e., using 2-norm matrix. The diagnostic value is obtained as follows:

$$\mathrm{D}_2 = \sqrt{\lambda_{max}(\mathbf{T}^{\mathrm{T}}\mathbf{T})}, \tag{2.38}$$

where the overall matrix can be built from the respective occupied $(\mathbf{T^o})^{\mathrm{T}}\mathbf{T^o}$ and virtual $(\mathbf{T^v})^{\mathrm{T}}\mathbf{T^v}$ matrices. These are then diagonalized, and the largest eigenvalue is selected. The elements of the matrices are given as

$$\begin{aligned}
[(\mathbf{T^o})^{\mathrm{T}}\mathbf{T^o}]_{i,j} &= \sum_{kab} T^{ik}_{ab} T^{kj}_{ab}, \\
[(\mathbf{T^v})^{\mathrm{T}}\mathbf{T^v}]_{a,b} &= \sum_{ijc} T^{ij}_{ac} T^{ij}_{cb}.
\end{aligned} \tag{2.39}$$

MP2 and CCSD are generally applicable if D_2(MP2) and D_2(CCSD) values are 0.15 or below. It is encouraged to go beyond MP2 if $\mathrm{D}_2(\mathrm{MP2}) > 0.17$ and beyond CCSD if $\mathrm{D}_2(\mathrm{CCSD}) > 0.18$. Values in between imply that both single-reference methods can be used with caution. These diagnostics are size-intensive, have the same invariance properties as the energy with respect to orbital rotations, and are easily computed since double excitation amplitudes are readily available.

2.3 Local Correlation Methods

The previous discussions in this chapter were mainly focused on methods which use canonical molecular orbitals (CMOs). CMOs consist of eigenvectors of the self-consistent field (SCF) Fock operator and are usually delocalized over large parts of a molecule. The use of such orbitals presents two challenges: 1) steep scaling and 2) chemical interpretation.

Conventional wavefunction methods, such as MP2 and CCSD(T) discussed in the previous sections, are still very costly when applied to large systems. Despite the advancement in computer hardware and efficient parallelization of the existing algorithms, the ‘scaling wall’ of these conventional methods still cannot be overcome. It is well known that the dynamic correlation is a short-range effect. The pair correlation energies decay at short range, i.e. r^{-6}, where r is the distance between two localized spin orbitals. The use of CMOs inhibits the conventional correlation approaches from benefiting from the electron locality. This also makes the ‘steep scaling’ unphysical. Another challenge presented by the use of CMOs is the chemical interpretation chemists can obtain after deploying quantum chemical

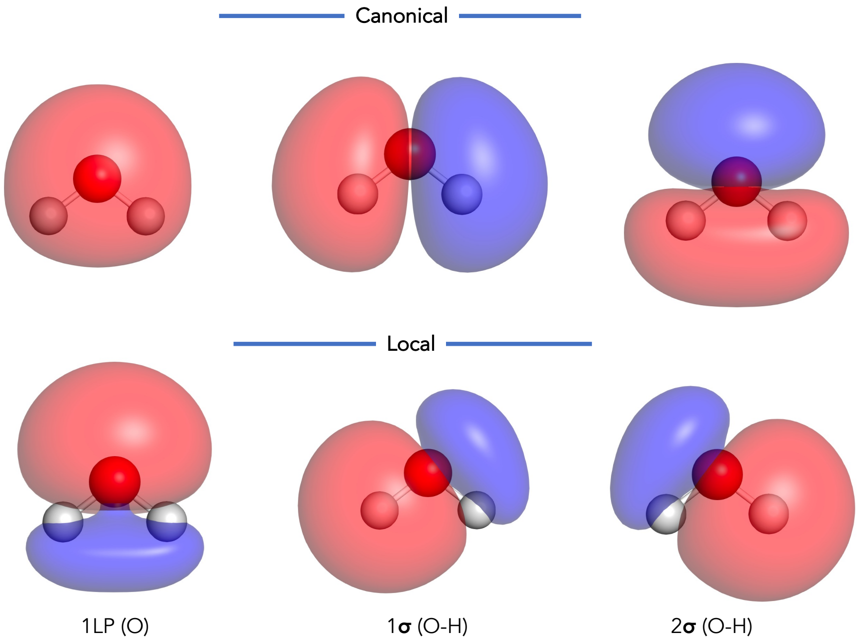


Figure 2.1: Selected occupied valence orbitals of water molecule. Canonical (top) and local (bottom) representations were obtained from the calculation at the LMP2/cc-pVDZ level using Molpro 2022.[54] LP refers to lone pair, while σ refers to the sigma bond indicated. Orbitals were visualized using PyMOL.[55]

calculations. The delocalized nature of CMOs does not often coincide with the molecular orbital pictures chemists have learned. See for example, in Figure 2.1 the top part are the valence occupied orbitals of a water molecule. These are selected molecular orbitals (MOs) from calculations conventional calculations, and as shown, the orbital density is scattered and difficult to associate with any physical meaning.

The challenges mentioned above can be addressed by the use of *local correlation methods*. These sets of methods are the traditional ones where the whole molecule is treated in one calculation, and various approximations that take advantage of the fast decay of the correlation energy are applied. The other approach is called *fragmentation methods*, in which the whole system is divided into smaller subsystems and these pieces are treated independently, mainly via conventional methods but local correlation methods are also possible. The latter approach is beyond the scope of this discussion, but further reading can be done in Reference 56.

In local correlation methods, the type of orbitals used are the localized molecular orbitals (LMOs). In contrast to CMOs, these are normally restricted to one or two atoms with a small intensity on the nearby atoms. In the case of π-delocalized systems, LMOs might have a significant intensity on three or more atoms. The use of LMOs addresses the difficulty of chemical interpretation which is inherent with the use of CMOs. Localizing molecular orbitals bridge the gap between chemical intuition and molecular wavefunctions, making it easier to connect with classical chemical concepts, for example in Figure 2.1. In this case, the lower orbitals can be viewed as Lewis representations of the two O−H bonds and one of the lone pairs of oxygen. From the practical point of view, LMOs significantly reduce the scaling of conventional correlated wavefunction methods due to a dramatic reduction of the virtual orbitals needed for the correlation of each electron pair, as well as the number of pairs overall computed.

The idea of local correlation is almost as old as the idea of electron correlation itself, and has been suggested even before HF was feasible for medium sized systems.[57,58] However, it took two decades to revive the idea through the pioneering works of Pulay and Saebø,[59–64] and in the last 25 years, enormous progress has been made towards the development of accurate local correlation methods. One of the ways to go about the scaling problem of the conventional methods is to take advantage of the short-range character of electron correlation. This can be done by the use of LMOs (discussed above) and by introducing local approximations. Two basic approximations are central to local correlation methods: first, the total correlation energy is the sum of pair energies where each pair describes the correlation of electrons in a pair of occupied LMOs, where each pair describes the correlation of the electrons in a pair of LMOs. A hierarchy is introduced depending on the magnitude of pair energies, and this is discussed further in Section 2.3.4, and depending on the method, approximations can be introduced for each class, ranging from full LCCSD for strong pairs to a non-iterative perturbation correction for distant pairs. Second is the *domain approximation*, which is applied to each individual pairs. A domain is a subset of local virtual orbitals which is spatially close to the LMO pair under consideration. With these local approximations in hand, the accuracy of the wavefunction-based correlation is preserved as much as possible. This approach also reduces the unfavorable scaling due to system size, ideally reaching linear scaling.

2.3.1 Localization of Orbitals

Pulay's main idea[59] regarding local description of electron correlation is to transform the occupied orbitals into local orbitals, as follows:

$$|i\rangle \equiv |\phi_i^{\mathrm{loc}}\rangle = \sum_k^{\mathrm{val}} |\phi_k^{\mathrm{can}}\rangle\, U_{ki}^{\mathrm{loc}}. \tag{2.40}$$

The equation above shows that unitary transformation (U) of CMOs $|\phi_k^{\mathrm{can}}\rangle$ results to LMOs $|i\rangle \equiv |\phi_i^{\mathrm{loc}}\rangle$. Superscript 'can' refers to CMOs, 'loc' to LMOs and 'val' to valence orbitals. Note that localization excludes uncorrelated core orbitals. This is because mixing of core and valence orbitals, even without any local approximation, would affect the correlation energy.

$$\begin{aligned} |i\rangle &= \sum_{\mu=1}^{N_{\mathrm{AO}}} |\mu\rangle\, L_{ui}, \\ \text{where} \quad L_{\mu i} &= \sum_k^{\mathrm{val}} C_{\mu k} U_{ki}^{\mathrm{loc}}. \end{aligned} \tag{2.41}$$

The localized valence orbitals $|i\rangle$ can also be expressed in AO basis $|\mu\rangle$ using a rectangular transformation matrix $\mathbf{L}$, whose dimension is given by $N_{\mathrm{AO}} \times N_{\mathrm{val}}$. The transformation matrix $\mathbf{L}$ is related to the canonical MO coefficient $\mathbf{C}$ by the unitary transformation matrix $\mathbf{U}^{\mathrm{loc}}$. In the following text, LMOs are indicated by indices i, j, k, l while Greek letters μ, ν, ρ, σ are used for AOs.

2.3.2 Localization Schemes

In general, local orbitals are obtained by maximization or minimization procedures of certain properties to determine the transformation matrix U_{ki}^{loc} from Equation 2.41. One of the most common localization schemes is the Foster-Boys (FB).[65,66] In this approach, the spatial extension of orbitals is minimized. Despite providing well-localized orbitals, the method suffers from the so-called 'banana bonds'.[67] This can be seen, for example, in the double bonds between carbons. The FB method generates two equal banana-shaped orbitals resulting from sp^2 hybridization of the carbons. Another approach to localization is the Edmiston-Ruedenberg (ER) method.[68,69] This approach is mainly based on the maximization of the self-repulsion energy, $\frac{1}{r_{12}}$. It is not used as extensively as others, probably due to the high scaling of 5th order compared to the 3rd order scaling of FB.

In 1989, Pipek and Mezey introduced a localization function that measures the number of atomic centers over which a molecule extends.[70,71] Pipek-Mezey (PM) scheme generally maximizes the sum of the squared Mulliken partial charges, q_{iA}

$$P^{\text{PM}} = \sum_{A}^{\text{atoms}} \sum_{i}^{\text{val}} q_{iA}^2, \tag{2.42}$$

where

$$q_{iA} = 2 \sum_{\mu \in A} L_{\mu i} \sum_{\nu} S_{\mu\nu} L_{\nu i}. \tag{2.43}$$

The first sum in Equation 2.43 indicates that it runs over all the basis functions μ centered on atom A. In the end, this minimizes the number of centers at which orbitals are localized.

In Mulliken population analysis, the overlap populations occur since the AO basis is not orthogonal. Thus, this overlap population is equally divided between different atomic centers. This ignores the possibility of both atoms having different electronegativity. This results in Mulliken's population displaying some unphysical behaviour which increases with the AO basis set size. This problem is aggravated when, on top of a large basis set, augmented functions are included.

The problematic definition of Mulliken partial charges in Equation 2.43 can be avoided by using intrinsic bond orbitals (IBOs).[72] To expand the occupied Hartree-Fock orbitals exactly, the molecule-intrinsic minimal basis of polarized orbitals (IAOs) $|\rho\rangle$ is created.[72–75] A suitable minimal basis of free-atom AO $|\tilde{\rho}\rangle$ generates the IAOs using projection operators, which are then symmetrically orthogonalized. The redefined partial charges are then

$$q_{iA} = 2 \sum_{\rho \in A} \langle i|\rho\rangle \langle \rho|i\rangle \,. \tag{2.44}$$

Equation 2.44 is used in PM localization shown in Equation 2.42.

2.3.3 Population Analysis

Chemical models are very important for aiding the understanding of various chemical phenomena. For example, in the context of reactivity, the concept of 'electron rich' and 'electron poor' regions is vital to predict possibilities of reaction or to rationalize the outcome of a synthetic route. From the quantum mechanical perspective, this can be aided by the mapping of atomic charges through population analysis.

Population analysis is a mathematical way of partitioning a wavefunction (or electron density) into charges on the nuclei, bond orders, or other related information. Since partial charges do not correspond to any unique physical properties, these are non-observables. In reality, there are no partial charges on each atom, just the fact that there is a positive nucleus and negative electrons. However, pinning down electron density and nuclear charges into partial charges per atom greatly helps chemists to understand the electron density distribution. Although an artificial construct, it is often useful for synthesis design and mechanism prediction.

Mulliken population analysis was partially discussed in the previous section (Section 2.3.2). In this approach, the electron population of an atom A is calculated as:

$$\rho_A = \sum_{\mu \in A}^{N_{\text{bas}}} \sum_{\nu}^{N_{\text{bas}}} D_{\mu\nu} S_{\mu\nu}. \tag{2.45}$$

The density matrix $D_{\mu\nu}$ can be obtained from

$$D_{\mu\nu} = \sum_{i}^{\text{occ}} n_i C_{\mu i} C_{\nu i}, \tag{2.46}$$

where n_i is the occupation number of the orbital. Note that the Mulliken charges in Equation 2.43 are related to the Mulliken population in Equation 2.45 by

$$\rho_A = \sum_i q_{iA}. \tag{2.47}$$

As mentioned, large basis sets with augmented functions can yield unreasonable results. Diffuse functions may describe adjacent atoms more than the atom which they are centered. For example, this approach tends to underestimate the charge separation in ionic bonded systems. Regardless of the deficiencies, this scheme is still popular because, as seen in the above equations, it is easy to implement.

2.3.4 Local MP2

In Section 2.3.1, the localization method was discussed, where it mainly focused on localizing occupied orbitals. The virtual space, on the other hand, is spanned by a set of nonorthogonal functions $\{\tilde{\phi}_r\}$. This can be done in several ways, e.g. projected atomic orbitals (PAOs),[59] pair natural orbitals (PNOs),[76] or orbital-specific virtuals (OSVs).[77] Using PAOs, the virtual space is obtained directly from AOs $\{\chi_\mu\}$. This can be done by projecting the occupied space as

$$|\tilde{\phi}_r\rangle = \left(1 - \sum_{i=1}^{m} |\phi_i\rangle \langle\phi_i|\right) |\chi_\mu\rangle = \sum_{\mu} |\chi_\mu\rangle P_{\mu r}, \tag{2.48}$$

in which the projection matrix can be obtained via

$$\mathbf{P} = \mathbf{1} - \mathbf{L}\mathbf{L}^\dagger\mathbf{S}. \tag{2.49}$$

The same notation is followed for the occupied LMOs $(i, j, ...)$ while notation $r, s, ...$ is assigned to projected AOs. Projected AO basis is indicated with tilde.

In the context of LMP2,[78] a subset $[i]$ (*orbital domain*) of the projected AOs $|\tilde{\phi}_r\rangle$ is assigned to each LMO $|\phi_i\rangle$. In the local basis, the first order wavefunction is

$$|\Psi^{(1)}\rangle = \frac{1}{2} \sum_{ij\in P} \sum_{rs\in[ij]} \tilde{T}^{ij}_{rs} |\Phi^{rs}_{ij}\rangle \text{ with } \tilde{T}^{ij}_{rs} = \tilde{T}^{ji}_{sr}. \tag{2.50}$$

From the equation above, P is the pair list. The pair domains $[ij]$ are defined as:

1. *Strong pairs* ($R \leq 1$ bohr) - treated with highest level, e.g. LCCSD.
2. *Weak pairs* ($1 < R \leq 8$ bohr) - treated (optionally) with LMP2.
3. *Distant pairs* ($8 < R \leq 15$ bohr) - treated (optionally) with LMP2.
4. *Very distant pairs* ($R \geq 15$ bohr) - can be entirely neglected.

In this context, however, strong and weak pairs are both treated by LMP2. Note that the number of projection functions $r, s \in [ij]$ for a given pair (ij) is independent of molecular size. R is the closest distance between atoms in the primary domains $[i]$ and $[j]$.

The amplitudes $\tilde{T}^{ij}$ are determined using iterative procedures since a local orbital basis does not diagonalize the zeroth-order Hamiltonian. The linear equations can be obtained by minimizing the Hylleraas functional:[61]

$$
\begin{aligned}
\tilde{R}^{ij}_{rs} &= \tilde{K}^{ij}_{rs} + \sum_{tu\in[ij]} \tilde{f}_{rt}\tilde{T}^{ij}_{tu}\tilde{S}_{us} + \sum_{tu\in[ij]} \tilde{S}_{rt}\tilde{T}^{ij}_{tu}\tilde{f}_{us} \\
&- \sum_{k} \left[\sum_{tu\in[kj]} \tilde{S}_{rt} f_{ik}\tilde{T}^{kj}_{tu}\tilde{S}_{us} + \sum_{tu\in[ki]} \tilde{S}_{rt} f_{kj}\tilde{T}^{ik}_{tu}\tilde{S}_{us} \right] \\
&= 0,
\end{aligned}
\tag{2.51}
$$

where $\tilde{S}$ and $\tilde{f}$ are the overlap and Fock terms in projected basis, respectively. These residuals $\tilde{R}^{ij}_{rs}$ must vanish to zero for $r, s \in [ij]$. The exchange term $\tilde{K}^{ij}_{rs}$ represents a small subset of the transformed two-electron integral, which is expressed as:

$$
\tilde{K}^{ij}_{rs} = (ri|sj) = \sum_{\mu\nu} P_{\mu r} P_{\nu s} \left[\sum_{\rho\sigma} L_{\rho i} L_{\sigma j} (\mu\rho|\nu\sigma) \right], \tag{2.52}
$$

where all r, s are close to either i or j. Once the amplitudes have converged, the second-order energy can be obtained by:

$$
E^{\text{LMP2}} = \sum_{ij\in P} \sum_{rs\in[ij]} (2\tilde{T}^{ij}_{rs} - \tilde{T}^{ij}_{sr})\tilde{K}^{ij}_{rs}. \tag{2.53}
$$

2.4 Energy Decomposition Analysis

The intermolecular forces governing the interaction of chemical systems, like in a dimer or in the bulk, represent a fascinating facet of chemistry and an intriguing area for chemical investigations. These interactions are ubiquitous, with their influence significantly affecting many fields, such as catalysis, biochemistry, material science and medicinal chemistry. However, gaining a deeper insight into the intricate interplay of these forces is a substantial challenge for theoretical chemistry. Despite the difficulty, several methods have been developed in the attempt to isolate and (approximately) quantify these forces from the interaction energy of the system. In this section, two schemes will be discussed, distinguishing between the decomposition of the HF energy or the correlation energy.

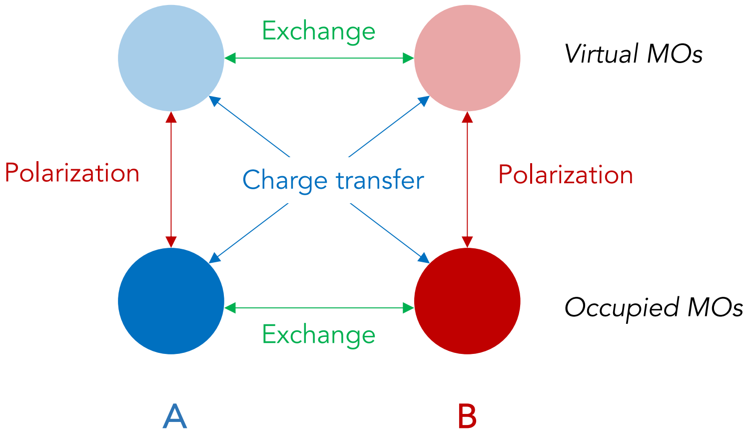


Figure 2.2: Energy decomposition of the HF interaction energy proposed by Kitaura and Morokuma.[79]

2.4.1 Kitaura-Morokuma Energy Decomposition Scheme

Consider a dimer system with closed-shell molecule A and B. The interaction energy of the dimer is defined as

$$E_{\text{int}} = E_{\text{AB}} - E_{\text{A}} - E_{\text{B}}, \tag{2.54}$$

where E_{A} and E_{B} are energies of molecules A and B computed at the dimer geometry. The individual energies should be calculated at a consistent level of theory and basis set. If calculations are done using the HF approximation, E_{int} is obtained by:

$$E_{\text{int}}^{\text{HF}} = \Delta E_{\text{AB}}^{\text{HF}} - \Delta E_{\text{A}}^{\text{HF}} - \Delta E_{\text{A}}^{\text{HF}}. \tag{2.55}$$

In 1976, Kitaura and Morokuma proposed a decomposition scheme within the Hartree-Fock approximation. In their scheme, the interaction energy $E_{\text{int}}^{\text{HF}}$ is divided into the following contributions:

$$E_{\text{int}}^{\text{HF}} = \Delta E_{\text{ele}} + \Delta E_{\text{pol}} + \Delta E_{\text{ex}} + \Delta E_{\text{ct}}. \tag{2.56}$$

The components in Equation 2.56, referred to here as KM-EDA, are defined as follows:

1. *Electrostatic* (ele) is the classical electrostatic interaction between occupied MOs. It does not cause any mixing of MOs, thus it does not appear in Figure 2.2.

2. *Polarization* (pol) is shown in red arrows in Figure 2.2. This type of interaction causes the mixing of occupied MOs (dark red/blue) with virtual MOs (light red/blue).

3. *Exchange* (ex) is the interaction between the occupied MOs of molecule A and B and virtual MOs of A and B (see green arrow in Figure 2.2). This interaction causes exchange and delocalization between molecules.

4. *Charge transfer* (ct) involves the mixing of occupied MOs of A with the virtual MOs of B, and vice versa, as indicated by the blue arrow in Figure 2.2. This type of interaction causes intermolecular delocalization.

The 'charge transfer' term in Equation 2.56 includes the rest of the contributions after isolating electrostatic, polarization, and exchange. This term can be further divided into 1) true charge transfer term, E_{ct}, 2) exchange-polarization term, E_{expl} and 3) coupling term E_{mix}, where

$$E_{\mathrm{mix}} = \Delta E - (\Delta E_{\mathrm{ele}} + \Delta_{\mathrm{pol}} + \Delta E_{\mathrm{ex}} + \Delta E_{\mathrm{ct}} + \Delta E_{\mathrm{expl}}). \tag{2.57}$$

In 2009, Su and Li[80] extended the described energy decomposition of the interaction energy from HF approximation. Their approach differs from the original KM-EDA in that it isolates the exchange and repulsion terms. The isolation of electrostatic, exchange, and repulsion terms was done according to Hayes and Stone's method.[81] This new implementation is formulated with spin orbitals that accommodate both closed and open-shell systems described by RHF, ROHF, and UHF. This enables the analysis of covalent bonds and intermolecular interactions.

Additionally, in this extension, the polarization energy is redefined as *orbital relaxation*, describing the transition from monomer HF spin orbitals to the supermolecule HF spin orbitals. This conceptually resembles the *electronic interaction energy* defined by Ziegler and Rauk[82] for the Hartree-Fock-Slater method.

To simplify things, assume that the dimer is composed of similar molecules, called A (i.e., A=B), and the dimer is called X. The total interaction energy of the dimer X composed of molecules A is described as

$$\Delta E^{\mathrm{HF}} = \langle \Psi_X | \hat{H}_X | \Psi_X \rangle - \sum_A \langle \Psi_A | \hat{H}_A | \Psi_A \rangle . \tag{2.58}$$

Ψ_X and Ψ_A are the HF wavefunctions for the dimer X and monomers A. The following equations will show how each individual contribution to the total interaction energy is calculated.

Before defining the terms of different energy contributions to the total interaction energy, three energy approximations are first defined. The approximate energy expression $E_X^{(1)}$ for a dimer X composed of monomers A is:

$$E_X^{(1)} = \sum_{i\in X}^{\alpha,\beta} h_i + \frac{1}{2}\sum_{i\in X}^{\alpha,\beta}\sum_{j\in X}^{\alpha,\beta} \langle ii|jj\rangle - \sum_A \left(\frac{1}{2}\sum_{i\in A}^{\alpha}\sum_{j\in A}^{\alpha} \langle ij|ij\rangle + \frac{1}{2}\sum_{i\in A}^{\beta}\sum_{j\in A}^{\beta} \langle ij|ij\rangle \right) + E_X^{\mathrm{nuc}}. \tag{2.59}$$

In this equation (Eq. 2.59) $h_i, \langle ii|jj\rangle$ and $\langle ij|ij\rangle$ represent one-electron and two-electron Coulomb and exchange integrals. E_{nuc} is the nuclear repulsion energy. The spin orbitals i and j are the variationally optimized HF orbitals that minimize the HF energy of each monomer and are orthonormal to each other within each monomer. They are not variationally optimized to minimize the dimer molecule HF energy and are not necessarily orthonormal to each other between the monomers. Note that this specific equation does not contain an exchange term between the monomers.

The approximate energy expression $E_X^{(2)}$ is provided by

$$E_X^{(2)} = \sum_{i\in X}^{\alpha,\beta} h_i + \frac{1}{2}\sum_{i\in X}^{\alpha,\beta}\sum_{j\in X}^{\alpha,\beta} \langle ii|jj\rangle - \frac{1}{2}\sum_{i\in X}^{\alpha}\sum_{j\in X}^{\alpha} \langle ij|ij\rangle - \frac{1}{2}\sum_{i\in X}^{\beta}\sum_{j\in X}^{\beta} \langle ij|ij\rangle + E_X^{\mathrm{nuc}}. \tag{2.60}$$

The spin orbitals i and j are the orthonormal HF spin orbitals of the monomers. Similar to Equation 2.59., they are not necessarily orthonormal to each other between monomers. Note that this energy approximation, unlike 2.59, already contains the exchange terms between monomers.

The last approximate energy is $E_X^{(3)}$, given by

$$
\begin{aligned}
E_X^{(3)} &= \sum_{i\in X}^{\alpha,\beta}\sum_{j\in X}^{\alpha,\beta} h_{ij}(S^{-1})_{ij} + \frac{1}{2}\sum_{i\in X}^{\alpha,\beta}\sum_{j\in X}^{\alpha,\beta}\sum_{k\in X}^{\alpha,\beta}\sum_{l\in X}^{\alpha,\beta}\langle ij|kl\rangle \\
&\times (S^{-1})_{ij}(S^{-1})_{kl} - \frac{1}{2}\sum_{i\in X}^{\alpha}\sum_{j\in X}^{\alpha}\sum_{k\in X}^{\alpha}\sum_{l\in X}^{\alpha}\langle ik|jl\rangle \\
&\times (S^{-1})_{ij}(S^{-1})_{kl} - \frac{1}{2}\sum_{i\in X}^{\beta}\sum_{j\in X}^{\beta}\sum_{k\in X}^{\beta}\sum_{l\in X}^{\beta}\langle ik|jl\rangle \\
&\times (S^{-1})_{ij}(S^{-1})_{kl} + E_X^{\text{nuc}}.
\end{aligned}
\tag{2.61}
$$

The equation above for the dimer X can be obtained if the monomer orbitals are used to form a single-determinant wavefunction. Orbital orthonormality is enforced by S^{-1}. The indices i, j, k and l are orthonormal HF spin orbitals of the monomers. Like the previous approximate energies, these may not necessarily be orthonormal between monomers. S is the overlap matrix or the monomer spin orbitals. Since orthonormality between monomers is not imposed, S and its inverse S^{-1} are not unit matrices.

From the approximate energies outlined above, the individual energy contributions to the total interaction energy can now be defined. The electrostatic interaction energy between the monomers A and dimer X is

$$
\begin{aligned}
\Delta E_{\text{ele}} &= E_X^{(1)} - \sum_A E_A^{\text{HF}} \\
&= \frac{1}{2}\sum_{i\in X}^{\alpha,\beta}\sum_{j\in X}^{\alpha,\beta}\langle ii|jj\rangle + E_X^{\text{nuc}} - \sum_A\left(\frac{1}{2}\sum_{i\in A}^{\alpha,\beta}\sum_{j\in A}^{\alpha,\beta}\langle ii|jj\rangle + E_A^{\text{nuc}}\right)
\end{aligned}
\tag{2.62}
$$

In RHF cases, Equation 2.62 is the same as in the original KM-EDA. This energy component is additive for a supermolecule consisting of many monomers.

The exchange energy is basically the difference of $E_X^{(2)}$ and $E_X^{(1)}$:

$$
\Delta E_{\text{ex}} = E_X^{(2)} - E_X^{(1)}. \tag{2.63}
$$

The exchange energy in Equation 2.63 is additive for a molecule composed of several monomers.

The repulsion contribution, on the other hand, can be obtained as:

$$\Delta E_{\text{rep}} = E_X^{(3)} - E_X^{(2)}. \tag{2.64}$$

In the case of RHF, $\Delta E_{\text{ex}} + \Delta E_{\text{rep}}$ is the same as the exchange repulsion term in KM-EDA formulation. Unlike the first two contributions defined above, the repulsion energy is not pairwise additive for a supermolecule consisting of many monomers. This is due to the enforced simultaneous orthonormalization of all the orbitals from all monomers by using the inverse of the supermolecule overlap matrix S, as reflected in Equation 2.61.

Lastly, the polarization term is defined as:

$$\Delta E_{\text{pol}} = E_X^{\text{HF}} - E_X^{(3)}, \tag{2.65}$$

where E_X^{HF} is the HF energy of the dimer X. For a supermolecule consisting of many monomers, ΔE_{pol} is not additive. This term is equivalent to the sum of polarization, charge transfer and the mixing term in the original formulation of KM-EDA.

Using the extended and modified version of Su and Li,[80] the total HF interaction energy ΔE^{HF} is decomposed into the following:

$$\Delta E^{\text{HF}} = \Delta E_{\text{ele}} + \Delta E_{\text{ex}} + \Delta E_{\text{rep}} + \Delta E_{\text{pol}}. \tag{2.66}$$

Note that the new decomposition scheme in Equation 2.66 will be referred to as **extended KM-EDA**. Any reference to extended KM-EDA in the succeeding chapters pertain to this definition.

2.4.2 PNO-LMP2 Energy Decomposition Scheme

In the framework of the energy decomposition scheme introduced by Su and Li,[80] the assessment of dispersion contributions to the overall interaction energy is accomplished through a supermolecular approach, coupled with the utilization of size-consistent correlation methods, such as MP2 and CCSD(T). This means they are treating the correlation energy as a pure dispersion contribution. However, that is not appropriate since correlation energy includes other forces on top of dispersion.

In the recent years, several approaches have been developed to dissect not only the HF total interaction energy but also the correlation energy. One of which is the perturbative approach where the Hamiltonian is partitioned into contributions of non-interacting fragments and a series of perturbating potentials.[83] One of its very well known methods is the Symmetry-Adapted Perturbation Theory (SAPT),[84] which is unfortunately very computationally demanding as its scaling is $\mathcal{O}(\mathcal{N}^7)$. Within SAPT, the total interaction energy is the sum of its *electrostatic, induction, dispersion* and *exchange-repulsion* contributions.

On a different front, Neese and colleagues have introduced the Local Energy Decomposition[85] (LED) scheme, designed to partition the Hartree-Fock and correlation energy into distinct components. This approach uses pairs of localized occupied orbitals and the contributions are grouped into various families, 1) *intrafragment excitations*, 2) *dispersion contributions* (composed of genuine dispersion and exchange dispersion) and 3) *charge transfer* contributions (two types of dynamic charge polarization and double dynamic charge polarization). Notably, LED has recently been extended to facilitate the analysis of interaction energies of open-shell molecular systems, utilizing the UHF-DLPNO-CCSD(T).[86]

Pair natural orbitals (PNOs) represent a cutting-edge in the local correlation methods. PNO-based coupled cluster methods are currently stand as the most accurate and efficient approaches in this domain.[38, 39, 87–91] It is noteworthy that the computational cost scales linearly with system size.[91–93] Their capability to treat big molecules at high accuracy makes them an interesting tool to explore the intricate interplay of intermolecular forces in a given system.

Within the Molpro quantum chemical package, a decomposition scheme using projected atomic orbital (PAO) LMP2 provides a way to breakdown the correlation energy into several components, i.e., *dispersion, exchange-dispersion* and *ionic* contributions. This is similar to the scheme presented by Schütz *et al.* in Figure 2.3. Recently, Wuttke and Mata modified the scheme by introducing the the charge transfer terms between distinct fragments, in contrast to the traditional bundling of all charge transfer components into ionic contributions. A detailed derivation of PNO-LMP2, as well as the technicalities of its implementation, is discussed Reference 94. In this section, the relevant equations towards the PNO-LMP2 decomposition scheme are highlighted.

Consider a dimer consisting of monomers A and B. At the correlation energy level, the dispersion and exchange-dispersion interactions of the orbitals i and j are located on different fragments A and B, respectively. In this context, LMOs i are excited to PNOs a, same as j to b. These contributions are expressed as follows:

$$\Delta E_{\text{disp}}(ij) = \sum_{ab}^{\text{PNOs}} \tilde{T}_{ab}^{ij} K_{ab}^{ij} P_a^A P_b^B, \text{where } i \in A, j \in B, \tag{2.67}$$

$$\Delta E_{\text{exdisp}}(ij) = \sum_{ab}^{\text{PNOs}} \tilde{T}_{ab}^{ij} K_{ab}^{ij} P_a^B P_b^A, \text{where } i \in A, j \in B. \tag{2.68}$$

From the equation above, P_a^A is the percent weight of PNO a on fragment A and $\tilde{T}_{ab}^{ij}$ is the contravariant amplitudes in the localized PNO basis. As for the charge transfer contributions, two types of contributions are present – between intermolecular and intramolecular orbital pairs. These contributions are calculated as:

$$\Delta E_{\text{CT}_{A\rightarrow B}}(ij) = \sum_{ab}^{\text{PNOs}} \tilde{T}_{ab}^{ij} K_{ab}^{ij} (P_a^B P_b^B + P_a^A P_b^B + P_a^B P_b^A), \text{where } i, j \in A \tag{2.69}$$

$$\Delta E_{\text{CT}_{A\rightarrow B}}(ij) = \sum_{ab}^{\text{PNOs}} \tilde{T}_{ab}^{ij} K_{ab}^{ij} P_a^B P_b^B, \text{where } i \in A, j \in B. \tag{2.70}$$

Finally, the intramolecular energy term, which is derived by the excitations from LMOs to PNOs located in the same fragment, is written as:

$$\Delta E_{\text{intra}}(ij) = \sum_{ab}^{\text{PNOs}} \tilde{T}_{ab}^{ij} K_{ab}^{ij} P_a^A P_b^A, \text{where } i, j \in A. \tag{2.71}$$

These contributions are printed in the output file of the calculations. PNO-MP2 as well as PNO-SCS-LMP2, in which the latter is used, are both available in the in-house program of the group and is implemented in Molpro 2018 developer's version.[95]

2.5 Local Orbital Analysis of Molecular Interactions

As mentioned in the earlier sections of this chapter, the use of LMOs instead of CMOs is beneficial in two ways – for chemical interpretation and computational speed. The former is an important point but is often not highlighted in quantum mechanical investigations.

The use of local orbital subspaces bridges the gap between the chemists' perceived chemical picture and the results derived from quantum mechanical calculations. In this section, the basics of local methods as well as the derivation yielding to the physical interpretation of the dispersion forces are discussed.

2.5.1 Basic Principles

For any correlated wavefunction-based methods, the energy expression can be written as the sum of the reference (E_{ref}) and correlation (E_{corr}) contributions:

$$E_{\text{tot}} = E_{\text{ref}} + E_{\text{corr}}, \tag{2.72}$$

where E$_{\text{ref}}$ is the HF energy.

For any correlated energy decomposition analysis (EDA) scheme, a useful starting point is the observation that the correlation energy of a closed-shell system can be written as:

$$E_{\text{corr}} = \sum_{i \geq j}^{\text{occ}} \sum_{ab}^{\text{virt}} (ia|jb)\tilde{\tau}_{ab}^{ij}, \tag{2.73}$$

where i, j are the occupied orbitals, a, b are virtual orbitals and $\tilde{\tau}_{ab}^{ij}$ are the contravariant amplitudes $\tilde{\tau}_{ab}^{ij} = \frac{4\tau_{ab}^{ij} - 2\tau_{ab}^{ij}}{1+\delta_{ij}}$. Note that in MP2, τ_{ab}^{ij} are the double excitation amplitudes T_{ab}^{ij}. Equation 2.73 can be expressed as the sum of double excitation contributions from pairs of occupied orbitals to the virtual orbitals. Therefore, E$_{\text{corr}}$ can be expressed as pair correlation energies, ϵ_{ij},

$$E_{\text{corr}} = \sum_{ij}^{\text{occ}} \epsilon_{ij}. \tag{2.74}$$

The correlation energy can be further divided into inter-fragment and intra-fragment contributions (using Equation 2.74) if each orbital is assigned to the fragment where it is dominantly localized. As initially suggested by Schütz,[96] if a local representation of the virtual orbitals is used, virtual orbitals can also be assigned to the fragments where they are predominantly localized. Looking at Equation 2.73, the double excitations can be regrouped into several classes corresponding to the different physical components of the interaction using the LMP2 method.

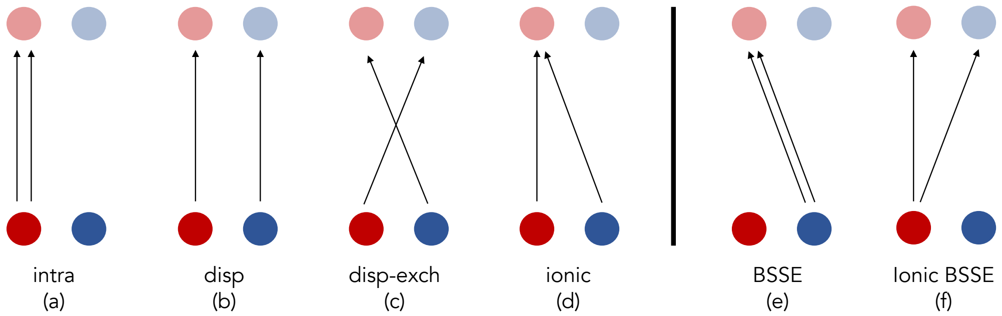


Figure 2.3: Schematic representation of different types of double excitation in the local correlation approach in the context of intermolecular interactions. Red represents fragment A and blue for fragment B. Dark colors (bottom part) depicts occupied orbitals while the lighter ones (upper part) are for virtual orbitals.

The decomposition of different double excitations in a local correlation approach is illustrated in Figure 2.3. In (a), the double substitution occurs within the same monomer and this corresponds to *intramolecular correlation effects*. If there is simultaneous single excitations on two different monomers to orbitals in the same domain, as in (b), this substitution class is responsible for the *dispersive coupling* between the individual monomers. On the other hand, cross excitations transferring one electron from an occupied LMO of monomer A (blue) to the correlating virtual space of monomer B (light red), and vice versa as shown in (c), this class refers to the *dispersion-exchange* effects. Lastly, we have three types of ionic substitutions which can be generally categorized as *charge transfer*. Excitation in (d) shows an ionic substitution transferring one electron from an occupied LMO of monomer B to the correlating virtual space of monomer A coupled with single excitation in monomer A. Excitations in (e) corresponds to the doubly ionic excitations from the occpied LMO of monomer B to the virtual space of monomer A. Lastly in (f), the type of ionic substitution is when the electrons are transferred from an occupied LMO of monomer A to the correlating virtual space of monomers A and B.

Per the construction of PAO-based local methods, (e) and (f) are automatically excluded.[96] Double excitation in (e) is mainly responsible for the BSSE in conventional calculations. This, of course, only considers the post-HF energy. The double excitations on (f), on the other hand, may contribute to some extent to the interaction energy, most especially when hydrogen bonds are involved.

Using the partition scheme described above, the correlation energy using a local wavefunction method (LMP2 in this case) is as follows:

$$E_{\mathrm{corr}} = \epsilon_{\mathrm{intra}} + \epsilon_{\mathrm{disp}} + \epsilon_{\mathrm{disp-exch}} + \epsilon_{\mathrm{ionic}}. \tag{2.75}$$

2.5.2 Dispersion Analysis Using Local Orbitals

The relevance of dispersion, most especially with an increasing number of electrons, is undeniable. This will be established in Chapter 3 and recent studies continuously prove its importance.[97–101] In local correlation methods, excitations are spatially resolved (see Figure 2.3), which allows one to distinguish London dispersion energy terms from other correlation contributions. Given a dimer system with monomer A and B, this arises from the excitation pairs which involves both monomers, but the ‘ownership’ of the electrons is untouched. Dispersion energy, as defined by each orbital pair, is as follows:

$$\epsilon_{ij}^{\mathrm{disp}} = \sum_{a\in A, b\in B} \frac{K_{ab}^{ij}(2K_{ab}^{ij} - K_{ab}^{ji})}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}, \tag{2.76}$$

with an excitation $\{ij\} \to \{ab\}$ where i, a is located in one monomer (say A) while j, b in a different monomer (say B). The monomers should be defined, but such approach is flexible enough that it can also handle intramolecular terms. Note that in both cases, it is necessary to define grouping criteria for different orbitals.[102, 103]

To build a visual representation of dispersion interactions, the pair energy terms $\epsilon_{ij}^{\mathrm{disp}}$ can be used. Considering fragment A, a matrix density D^{A} in the AO fragment is computed as

$$D_{\mu\nu}^{\mathrm{A}} = \sum_{i\in A} \frac{1}{N_i} \left(\epsilon_{ij}^{\mathrm{disp}}\right) P_{\mu\nu}^{i}, \tag{2.77}$$

where $P_{\mu\nu}^{i}$ corresponds to the closed-shell orbital density of orbital i. This density matrix is obtained as

$$P_{\mu\nu}^{i} = 2C_{\mu i}C_{\nu i}, \tag{2.78}$$

which is simply the AO density matrix for orbital ϕ_i.

In order to recover the dispersion interaction between fragment A and B, $\Gamma^{\mathrm{A}}(\mathrm{r})$ is integrated over the whole space, as in the following equation:

$$\Gamma^A(\mathbf{r}) = \sum_{\mu\nu} D^A_{\mu\nu} \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}). \tag{2.79}$$

This is interpreted as the dispersion interaction density (DID) in AO matrix representation. Such definition has already been used in previous works as well as by other groups.[104–106]

Chapter 3

Dispersion Forces in Diols, Ketone-Alcohols and Helicenes

Some parts of this chapter are reproduced from the results, discussion and conclusion sections of the following published access articles:

X. Aniban, B. Hartwig, A. Wuttke, and R.A. Mata. Dispersion Forces in Chirality Recognition – A Density Functional and Wave Function Theory Study of Diols. *Phys. Chem. Chem. Phys.*, 23(21):12093–12104, 2021

S. Suárez-Pantiga, P. Redero, X. Aniban, M. Simon, C. Golz, R.A. Mata, and M. Alcarazo. In-Fjord Substitution in Expanded Helicenes: Effects of the Insert on the Inversion Barrier and Helical Pitch. *Chem. Eur. J.*, 27(53):13358–13366, 2021

3.1 Introduction

The interaction between enantiomers and diastereomers is a determining factor for the structure of the chiral molecular aggregates and is tied to their function and corresponding molecular properties. One scenario involves chirality recognition, which is the ability of a chiral probe to distinguish between two enantiomers of different molecules. The interactions between these molecules subtly change once one of the systems is replaced by its mirror image, making the study of non-covalent interactions more interesting albeit challenging. Since chiral molecules constitute the building blocks of life, understanding the fundamentals of such phenomena provides a lot of insight as to how nature works. No wonder this has fascinated chemists and physicists since its discovery in the middle 19th century.[109]

The homochirality of life represents a very striking aspect of chirality recognition and remains one of the great, unanswered questions in evolutionary science.[110,111] Macrobiomolecules, for example, consist of units of the same chirality. In human beings the proteinogenic amino acids have the L configurations while bacterial cell walls contain their D enantiomers. The interaction of drugs with enzymes or protein receptors is also characterized by marked enantioselectivity.[112] That is why in the fields of synthesis, drug development and toxicology, chirality recognition plays a very important role.[113–117] The most prominent example involves the drug thalidomide, which contained a racemic mixture of the active ingredient. This latter was prescribed to (pregnant) women to alleviate morning sickness. Unfortunately, it was later realized that the (R)-enantiomer produced the desired result, but the (S)-enantiomer caused severe birth defects.[118–120] Thus, chirality recognition is a very important consideration in biochemistry[121] and organic synthesis.[122,123] These interactions also find their relevance in supramolecular chemistry[124–127] and chirality effects in molecular imprinting.[128]

Chirality recognition is thought to take place when weakly bound contact pairs form, which involve specific interactions, one of which is hydrogen bonding. Experimental methods have been developed to gauge these interactions, and indirect information can be obtained through optical spectroscopy,[129] NMR spectroscopy,[130] circular dichroism[131] and some thermodynamic measurements.[132] However, there remains a scarcity of direct characterization at the molecular level. This would be very helpful in acquiring an accurate understanding of forces at play in this phenomenon. Experimental techniques that do not involve perturbation brought by solvent are highly desirable. Chirality recognition in molecular encounters can then be effectively studied by gas-phase experiments on isolated neutral or ionic clusters. Linking supersonic expansions with high- or medium-resolution spectroscopic techniques allows for the formation of non-equivalent weakly bound diasterioisomer complexes and their spectroscopic characterization (see a review of methods in Ref. 133).

High-level quantum chemistry calculations thrive very well with experiments done in gas phase. The electronic, vibrational, rotational as well as photoionization spectroscopic data serve as excellent benchmarks for *ab initio* calculations. In a number of chirality recognition studies done using theory and experiment, DFT (Density Functional Theory) has always been the workhorse on the theory side.[133–135] This is understandable because WFT methods are usually more expensive. Conventional MP2, for example, scales to $\mathcal{O}(\mathcal{N})^5$ and the 'gold standard' CCSD(T) scales to $\mathcal{O}(\mathcal{N})^7$. However, developments in WFT led to local methods which, apart from making WFT methods more affordable, also provide a better picture of orbitals and orbital interactions which is highly favorable for chemists' interpretation.

In this chapter, several forms of chirality are explored. Molecules with stereogenic centers - a common form of molecular chirality and perhaps the most familiar to chemists (e.g. ketone-alcohol systems), are investigated. Several systems whose chirality is due to molecular asymmetry caused by steric hindrance (e.g., diols) are also explored. Lastly, molecules that are chiral due to internal steric strains are probed.

Throughout this chapter, we address a central question concerning subtle intermolecular forces responsible for chirality recognition. Specifically, we inquire about the role of dispersion in the chirality recognition of diols. Our investigation covers not only the energetic aspect but also the extent to which dispersion has affected the molecular structure. Additionally, the intriguing case of fenchone-fenchol systems is featured in this chapter. The different molecular systems are isoenergetic, which means that if chirality recognition is only viewed from the energy perspective, it does not exist on these systems. However, their structural aspect differ significantly and such difference is evident on spectroscopic data available. This prompts us to scrutinize the delicate balance of the various forces that dictate and stabilize these configurations. Lastly, we delve into the realm of expanded helicenes where the effect of substituent in the fjord of the helix was reviewed. Specifically, we investigated how the substituent affected the potential energy surface of the system, as well as how dispersion plays a role in shaping the helicenes particularly when in-fjord substitution and helix expansions were performed.

3.2 Definition of Dispersion-Free LMP2 Gradients

In order to properly describe non-covalent interactions, especially the weak forces that significantly contribute to the molecular stability of the system under scrutiny, a proper treatment of the electron correlation is necessary. The simplest useful post-HF method which can incorporate dynamic electron correlation is MP2. In MP2, unlike in HF, London dispersion forces are naturally accounted for. One drawback, though, is that it typically overestimates the non-covalent interaction energies for dispersively bound systems. Despite that, MP2 generally outperforms HF and exchange correlation functionals which have not been corrected for dispersion in the calculation of interaction energies.

Based on Equation 2.75 and Figure 2.3, the class of excitation responsible for dispersion contribution can be easily identified. In order to remove the dispersion gradients from (SCS)-LMP2 method, the following steps are taken:

1. An LMP2 energy calculation is carried out at each step. Using non-canonical orbitals, the amplitudes are obtained in an iterative fashion.
2. The amplitudes for pair excitations describing dispersion are set to zero at the end of the LMP2 iterations, i.e., T_{ab}^{ij} with $i, a \in A$ and $j, b \in B \neq A$.
3. Using the altered amplitudes matrix, the coupled perturbed localization equations are solved and the (SCS)-LMP2 gradients are computed.
4. If the Hess-matrix is computed, the same amplitudes matrix is used, consistent with the gradients.

The results for dispersion-free LMP2 and SCS-LMP2 gradients were confirmed by numerical gradients. All of these methods were integrated in the Molpro 2018.1[136] development version as an internal program for the research group.

3.3 Chirality Recognition in Diol Systems

As mentioned earlier, studying non-covalently bound systems without the perturbations brought by the solvent is of great interest because it can provide a more accurate understanding of how different forces are at play. Thus, benchmarking by gas phase experiments is highly favorable for quantum mechanical studies. In the case of studying chirality recognition via supersonic jet expansions, the use of jet cooling allows for low temperatures, which closely coincide with the commonly assumed 0 K for the quantum mechanical calculations. This allows an optimal collaboration between theory and experiment.

From the theory perspective, one of the aspects that is of interest is the relative stability of some local minima representing possible structures observed in the gas phase experiments. IR, Raman and microwave spectroscopy experiments are helpful in deducing these structures. In the case of neutral molecular aggregates, the choice of electronic structure method plays an important role because their relative stability is strongly influenced by van der Waals interactions.

3.3.1 The Interesting Case of Diols

When systems contain hydrogen bonding, their conformational stability is usually interpreted on the basis of charge transfer and electrostatic effects. While this is often correct, other intermolecular forces of attraction are also present, for example, dispersion.[a] Despite the latter being considered weak, its contribution usually increases with system size and can become an important driving force of some molecular properties, e.g., for molecular stability and reactivity. Dispersion has long been neglected in theoretical treatments because it is a weak force (which is only true for smaller systems) decaying at R^{-6} and percentage-wise. In general, it was thought to have a small effect on the total energy of the system. However, in recent years, it was shown that this force is not as negligible as the scientific community thought earlier. In some instances, it is a key force for stabilization[138,139] and is an important point to consider in interpreting reactivity.[140–146] With its demonstrated relevance, quantum chemical approaches have evolved to capture this, and for methods which naturally account for dispersion, prime importance is given to its quantitative approximation.

In studying chirality recognition in gas phase, the importance of hydrogen bonding as well as dispersion interactions has been explored extensively using infrared, Raman, microwave and mass spectroscopy.[135,147–156] Although most studies are concerned with intermolecular chiral interactions, intramolecular recognition can also occur. One example is the folding of *n*-alkanes where all the *gauche* angles in the kink of the most stable hairpin-like structure

---

[a] It is unfortunate that some literature interchange dispersion and van der Waals forces. Sometimes, even if the intermolecular force referred to is van der Waals, the community has an implicit assumption that this is dispersion. It is important to highlight here that van der Waals force is defined by IUPAC as *the attractive or repulsive forces between molecular entities (or between groups within the same molecular entity) other than those due to bond formation or to the electrostatic interaction of ions or of ionic groups with one another or with neutral molecules. The term includes: dipole–dipole, dipole-induced dipole and London (instantaneous induced dipole-induced dipole) forces. The term is sometimes used loosely for the totality of nonspecific attractive or repulsive intermolecular forces.*[137]

exhibit the same sign.[157] In reactions involving chiral molecules, dispersion is also vital, as in the case of catalytic asymmetric Diels-Alder reactions.[145] In this case, a chiral ion forms a pocket for the diene unit, and dispersion is responsible for stabilizing its enantioselectivity. Moreover, in host-guest interactions, dispersion is a relevant force responsible for the chiral separation in chromatographic methods.[158–161]

Chirality induced spin selectivity (CISS) is one interesting mechanism proposed to explain chirality recognition between chiral species. According to Kumar *et al.*,[162] spin polarization interaction is less repulsive for homochiral molecules than their heterochiral counterpart. This results in an enantiospecific overall interaction between each type of species. It should be noted, however, that this is a short-range interaction where a significant orbital overlap is important. From the theoretical end, it is not clear which level of quantum chemistry is appropriate to capture CISS accurately. Also, it is of interest to know if dispersion can at all drive chirality recognition.

In the quest to understand better the phenomena of chirality recognition, the aggregation of axially chiral 1,2-diols via conventional hydrogen bonds are particularly interesting. Such molecular species offer a multitude of directional intermolecular hydrogen bond patterns. This has interesting consequences in solids because the simplest 1,2-diol, i.e., ethanediol, crystallizes separately depending on the handedness and extended hydrogen-bonded networks.[163] This occurs despite the fact that this molecular species is only transiently chiral in the liquid phase.

In this section, several species of diols were investigated as shown in Figure 3.1. It is important to note that two systems, i.e., ethanediol (EDO) and pinacol, exhibit transient chirality while cyclohexanediol (CHexDO) has permanent chiral properties. Note that the first two are examples of chirality cases whose molecular asymmetry is due to steric hindrances (in this case, depending on the phase). Cyclohexanediol, on the other hand, is an example of a chiral system with stereogenic centers.

3.3.2 DFT and Dispersion Corrections

In standard DFT methods, energies are approximated on the basis of local quantities of the local electron density or reduced density gradient in GGA functionals.[164] However, London dispersion is an interaction which originates from non-overlapping densities which cannot be accounted for in classical DFT functionals. The failure of DFT to describe dispersion interactions has long been documented since at least 25 years ago.[165–167] The success of some functionals (most notably B3LYP/6-31G(d)) was due to error compensation

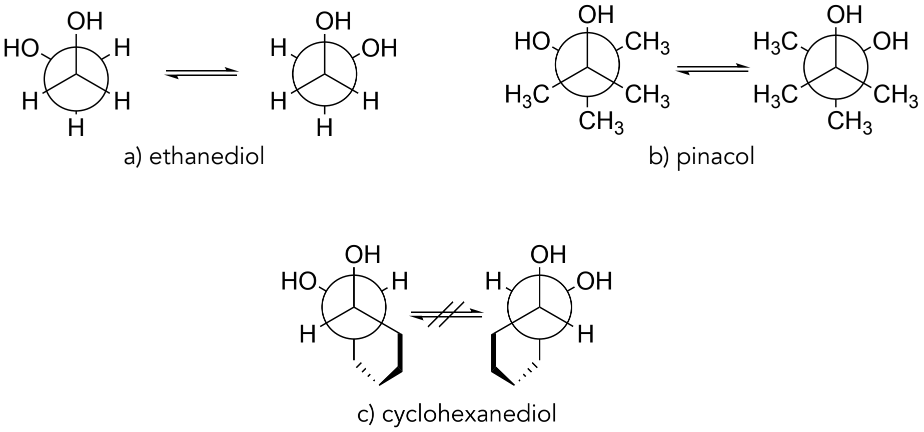


Figure 3.1: Molecular systems studied here, represented using their respective Newman projection. The *gauche* + and - conformers are portrayed for a) ethanediol (EDO) and b) pinacol, showing the transient nature of these systems. On the other hand, the permanent chirality of c) cyclohexanediol (CHexDO) are depicted between its *R,R* and *S,S* conformers.

because of lack of complete description of dispersion and pronounced BSSE.[168] To remedy this problem, several dispersion corrections devised by Grimme were incorporated. This involves a 'correction' potential function that is added to the exchange-correlation functional of choice.[169–171] However, this does not only capture dispersion but corrects other shortcomings of the functionals as well.

Note that DFT does not completely lack dispersion. The functionals themselves have already some dispersion incorporated albeit incomplete. The addition of the empirical dispersion corrections gives a better description of this force in chemical systems. However, in investigations that require a total neglect of dispersion, simply removing D-terms will not necessarily eliminate it. One reason is that density functionals already account for dispersion-like contributions at short distances. Also, the extent of empirical correction is not clear in terms of how it corrects for other shortcomings like correlation and exchange parts of the functional. In hybrid functionals, for example, the degree to which the exact HF exchange is mixed introduces another uncertainty.

3.3.3 Local Orbital Scheme in Chirality Recognition

WFT, unlike DFT, captures dispersion not as a separate term but as a part of the total energy. To isolate dispersion contributions and non-covalent interactions in general, a multitude of approaches are available. Some of these methods have been discussed in

Section 2.4.2, such as SAPT. An energy decomposition scheme based on local correlation methods can naturally describe long-range correlation effects like London dispersion. This approach has been described in detail in Section 2.5.1. If one wants to go to coupled cluster level, there is also local variants of the method to compute non-covalent interactions. PNO-LCCSD(T)-F12[40] and DLPNO-CCSD(T) LED[85, 104, 106] are representative methods. The inclusion of orbital relaxation terms and the incorporation of the triples in the CCSD(T)-based energy decomposition schemes make it problematic; thus triples are commonly excluded in the analysis.[85]

In the context of local orbital approaches, excitation terms contributing to dispersion can be pinpointed and isolated. This section delves into an examination of whether dispersion effects play a role in chirality recognition for some selected diols shown in Figure 3.1. To address this, dispersion contributions for both DFT and local WFT were activated and deactivated to see how the energy and structure of the systems were affected. Since most works quantify dispersion by examining energy differences in fixed structures, a different approach was taken by also carefully analyzing the changes in structures – something that is often neglected. Two important structural parameters were analyzed: intermolecular hydrogen (OH⋯O) bonding and the distance between the center of masses of each monomer in the dimer, R(CM-CM). These are important considerations in benchmarking, especially when experimental techniques such as high resolution infrared and microwave spectroscopies are involved. In the energy aspect, the range of relative stabilities and more importantly, the energy gap between heterochiral-homochiral (het-hom) species were carefully analyzed. To provide further insights, DID visualization[103, 172] was used to delineate specific moieties within the system that interact due to dispersion effects.

3.3.4 Computional Methods

DFT calculations

Conformational sampling was done using the CREST (Conformer-Rotamer Ensemble Sampling Tool)[173, 174] program with the semi-empirical GFN2-xtb (Geometries Frequencies Noncovalent Interactions version 2- extended Tight Binding)[175, 176] method. Non-covalent interaction (keyword: NCI) mode was specified in the said approach. To ensure that no conformers have been overlooked, multiple sampling runs, as well as manual cross checking were done.

The large number of conformers generated by CREST were pre-optimized with B97-3c/madef2-TZVP.[177–179] Relevant structures were then optimized using BP86,[180–182] PBE,[183] PBE0[184,185] and B3LYP[180,186,187] using the same basis set. This basis set was chosen for these set of structures because it is cost-effective. It was also observed that when using the mentioned basis set with B3LYP, it can get aVQZ level of results in terms of relative energies but in a very fast computation time (approximately 90 times faster). Note that density fitting (RI-J) was used for PBE and BP86.[188] The optimization was followed by frequency calculation within the double harmonic approximation. All these calculations were run using ORCA (version 4.2.1).[189,190] Dispersion effects were incorporated using D3(BJ,abc).[170,191] DFT calculations 'without' dispersion were done by removing the D3(BJ,abc) keyword.

WFT Calculations

Relevant structures found in the DFT calculations above were used for the WFT calculations. Optimization and frequency calculations of the molecular systems were done using PAO-based LMP2[64,192] and spin component scaled LMP2 (SCS-LMP2)[26] methods with the aug-cc-pVTZ, H=cc-pVTZ basis set.[193,194] For the localization part, Pipek-Mezey localization scheme was used.[70,71] Density fitting was also utilized for both the reference wavefunction and (SCS)-LMP2 calculations. All these calculations were conducted using Molpro 2018.1 commercial version.[136]

For WFT calculations without dispersion, the same specifications as in the calculations above were employed, except that the local methods were modified as described in Section 3.2 in an in-house program. Calculations were carried out using the Molpro 2018.1 development version.[136] For the DID calculations, Molpro 2019.2 was used.[195]

Unless otherwise stated, all presented energies are zero-point corrected.

Naming Scheme for the Molecular Species

To identify each species, the nomenclature used was the same as in Ref. 135.

- It distinguishes between dimers of the same chirality (homochiral = hom) and different chirality (heterochiral = het).
- The different conformers are characterized by the amount of intermolecular hydrogen bonds and how the terminal hydrogen bond is oriented (only the intermolecular hydrogen bonds), thus the number after het or hom.

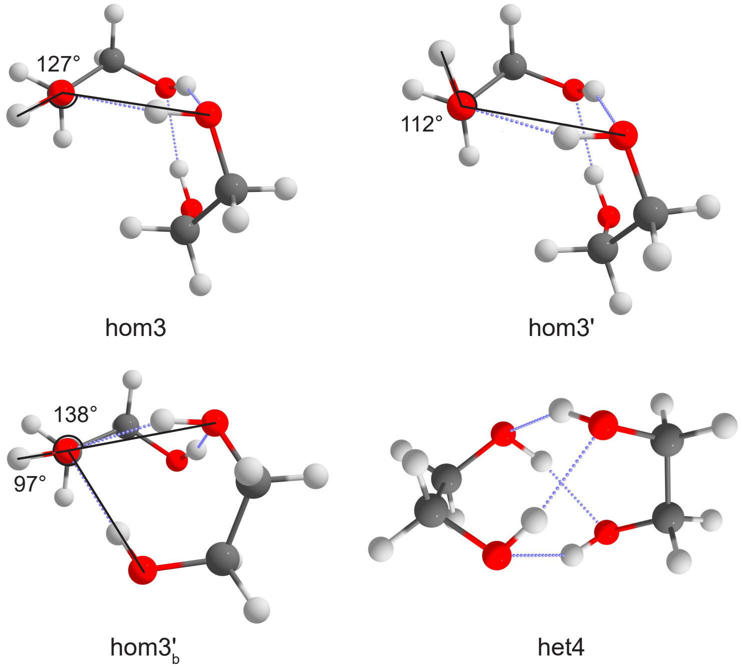


Figure 3.2: Sample illustration of O-OH angle and how different specifiers are determined. A structure with four intermolecular hydrogen bonds is labeled as **het4** and does not require additional specification because no terminal nor dangling OH groups exist. On the other hand, a structure with three intermolecular hydrogen bonds with bifurcation is labeled as **hom3$_b$'**.

- As shown in Figure 3.2, an apostrophe (') sign is added if the terminal O−OH angle is below 120°, as shown in the figures labeled as hom3 and hom3'.

- Each label must be unique, so in cases where any label reoccurs, an **'a'** is added. This also represents its own structural motif, where the C−C backbone is oriented parallel whereas it is more orthogonal otherwise.

- For bifurcated hydrogen bond arrangements, a subscript **'b'** is added. As shown in Figure 3.2, the bifurcated arrangement makes it necessary to account for two different O-OH angles which overall results in hom3$_b$' as its label. On the other hand, het4 does not have terminal/dangling OH groups, thus no more specification is needed.

- All conformers with the same name of the different compounds share close structural similarity.

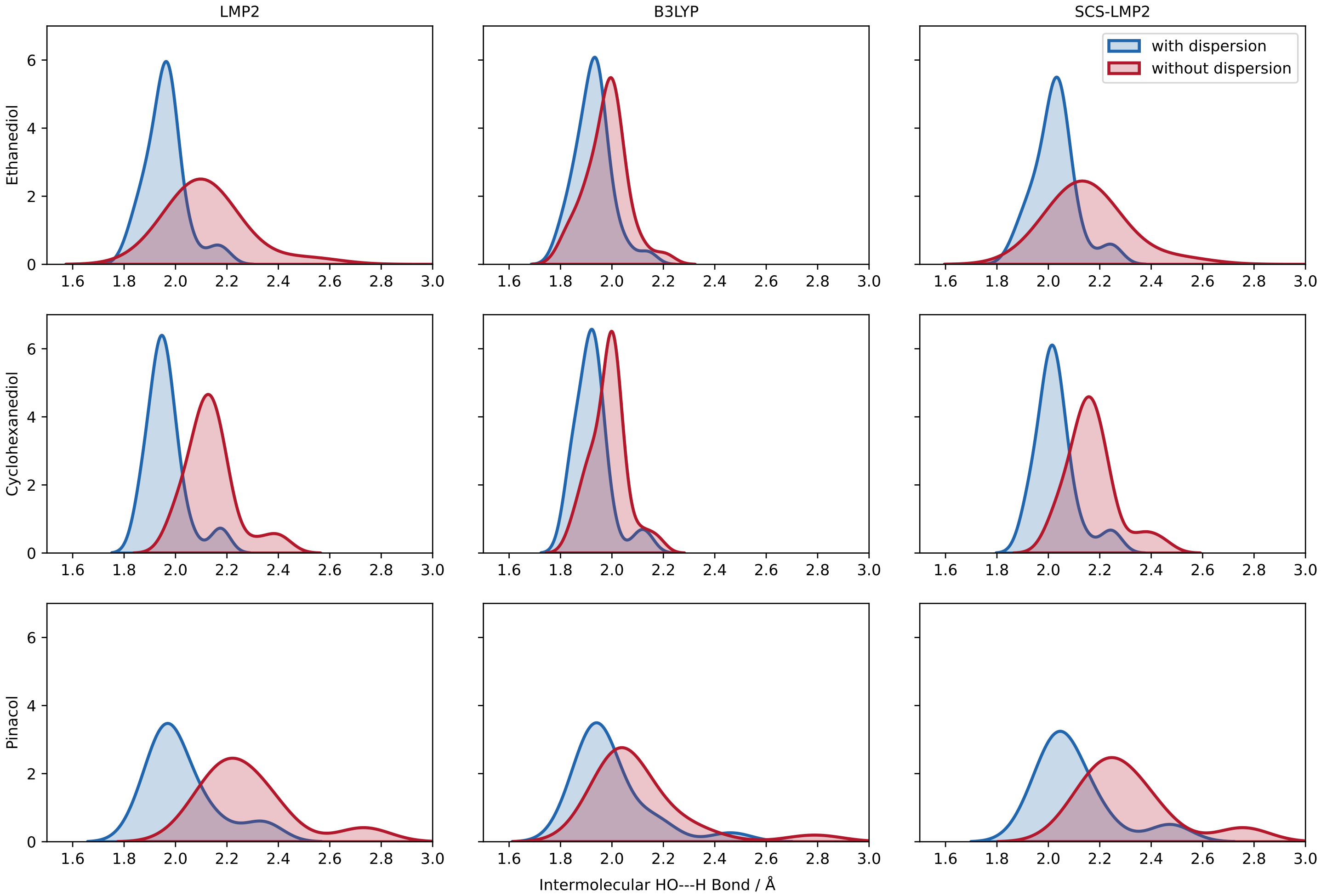


Figure 3.3: The behaviour of the intermolecular hydrogen bond (HO···H) distances was mapped using kernel density estimate (KDE) plot. Three molecular systems (top to bottom) were evaluated at different levels of theory (left to right). The density is taken from every HO···H distance, thereby showing the frequency and distribution. To ensure comparability between different approaches, the y-axis is normalized. See Appendix A.1, Figure A.1 for more details.

3.3.5 Is Dispersion a Driving Force of Chirality Recognition?

Structural Changes

For systems with hydrogen bonding, the intermolecular hydrogen bond distances are very important to examine because the simulated vibrational frequency calculations are highly dependent on the structure. Thus, care must be taken in benchmarking signature peaks. In this section, geometry optimizations, with and without dispersion, were done on the structures of different diol systems. The impact of dispersion on the different structures of the selected diol systems is presented.

As shown in Figure 3.3, the density distribution of the HO···H bond distances was noted in the presence of dispersion and once it was removed. Uniform scaling was ensured to make sure that comparisons among different systems can be made. A 0.20 Å and 0.17 Å average HO···H bond shift was observed in EDO for LMP2 and SCS-LMP2 results, respectively. B3LYP results showed a lesser degree of shift, i.e., 0.05 Å. The same was observed for CHexDO. The mean values of the intermolecular hydrogen shift were 0.18 Å (LMP2), 0.14 Å (SCS-LMP2) and 0.06 Å (B3LYP). For pinacol systems, there was an increase in the mean values, notably from B3LYP which gave 0.11 Å shift. The WFT methods also showed higher shifts for this system, i.e., 0.27 Å for LMP2 and 0.22 Å for SCS-LMP2.

Based on the values discussed above, WFT methods indicated a more pronounced shift in HO···H bonding once dispersion was removed. There was also a broadening in the distribution alongside with the shift. DFT methods, however, had a very slight shift relative to their WFT counterparts on the same system once D3 correction was not included in the geometry optimization. Structurally, longer hydrogen bonds were observed once dispersion was neglected, resulting in weaker hydrogen bonds. The larger impact in (SCS)-LMP2 is easy to explain. With uncorrected DFT, dispersion is still included in the short range, that is why the impact is lesser. This is a common issue in interpretations of the impact of dispersion based solely on the D3 correction. Overall, the changes in intermolecular hydrogen bonding (which can be seen from the density shifts) as well as the broadening of the density peaks reflects the scattering of distances after dispersion was removed - a clear indication on how dispersion strongly affects the structure during geometry optimization.

Another structural aspect worthy of analysis is the distance between the center of masses of the two monomers in the dimer. This is important because any changes to the structure, either the monomers are closer or farther from each other, can affect interpretations in microwave spectroscopy. This change, denoted as ΔR(CM-CM), indicates whether or not

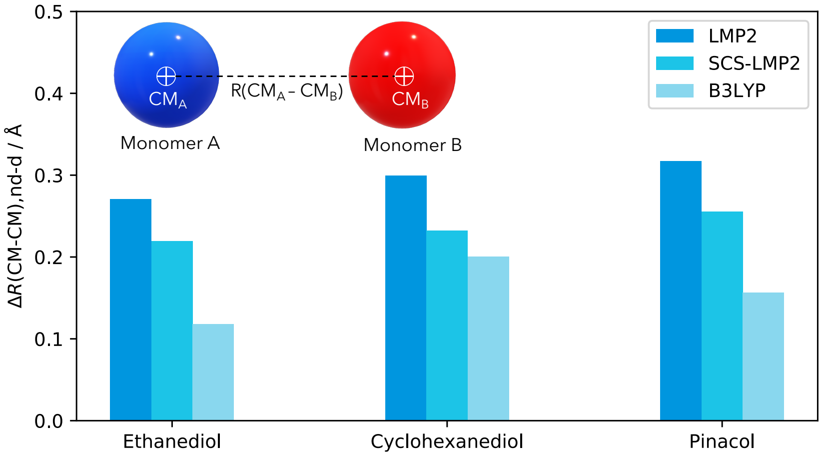


Figure 3.4: Differences in the center of mass (CM), ΔR(CM-CM), of each monomer in the dimer. Positive values indicate the increase in the distances of the monomers once dispersion was not included.

the monomers move towards or far from each other. We now observe how average ΔR(CM-CM) changes once dispersion was removed.

Upon analyzing the trend in R(CM-CM), it was observed that it has a similar behavior with the HO···H bond mapping. In EDO, for example, the removal of dispersion caused the monomers to move away from each other by 0.22 Å (SCS-LMP2) and 0.27 Å (LMP2). Removal of D3 in B3LYP showed only a slight increase in separation, i.e., 0.12 Å. Pinacol shows a similar pattern. Then again, B3LYP only resulted in an average of 0.16 Å separation, while the ΔR(CM-CM) doubled for LMP2, at 0.32 Å. SCS-LMP2 also resulted in a wide separation of 0.26 Å. In a slightly contrasting picture, CHexDO has a similar trend to others but the differences between the methods are not as much as the EDO and pinacol. LMP2 indicated an average separation of 0.30 Å while B3LYP has 0.20 Å. SCS-LMP2 has closer results to B3LYP, with only 0.23 Å increase in the distances of the monomers once dispersion was removed. The explanation regarding this particular structural aspect is similar to the intermolecular hydrogen bond distances.

Energy Differences: Het–Hom Gaps

Chirality recognition is often discussed in terms of relative energies, which determine the populations of different conformers. As shown in Figure 3.5, one of the noticeable effects of dispersion is in the energy range of the relative stability of every molecular system being examined, as shown in the energy range of the red and blue dots. These dots represent

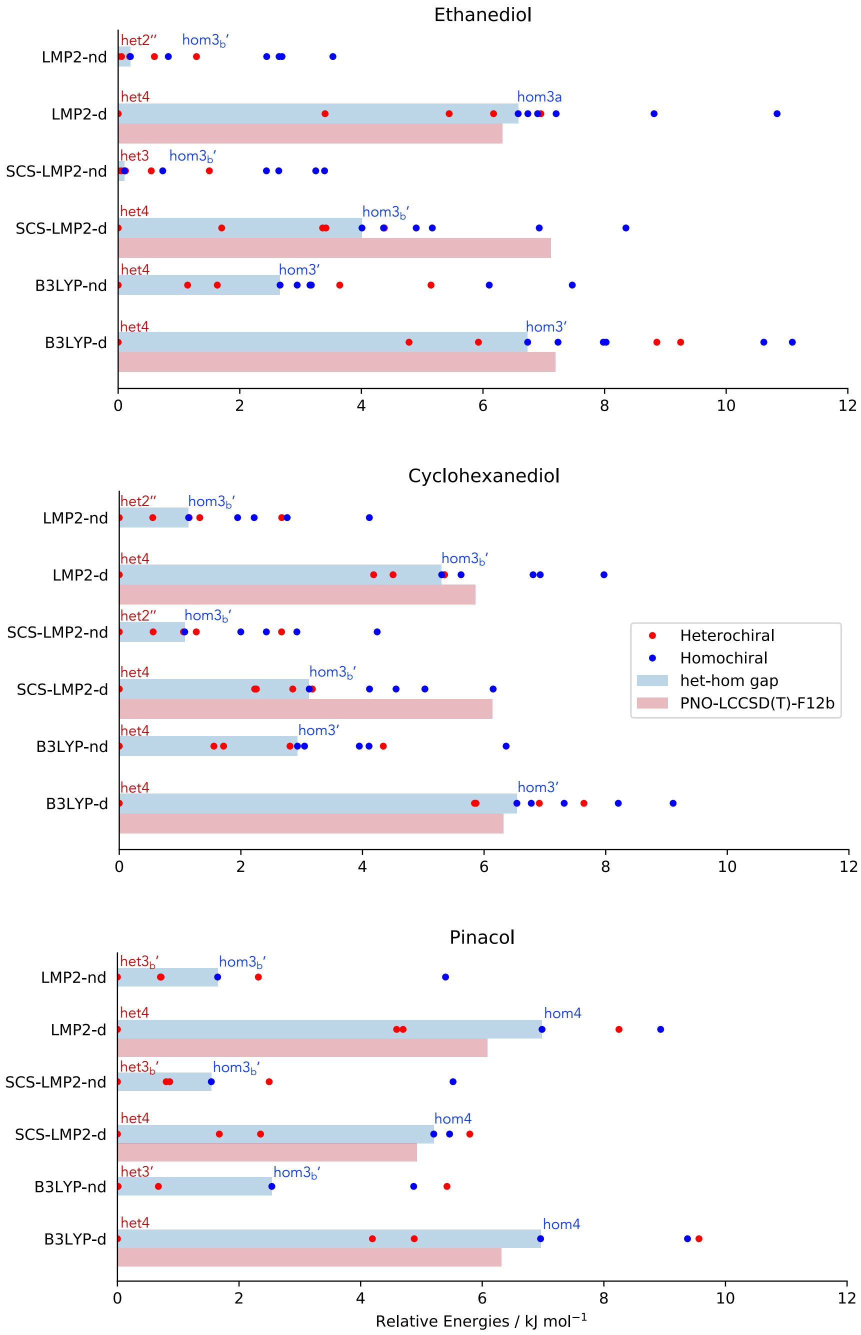


Figure 3.5: Chirality recognition in diols, as reflected by their heterochiral-homochiral (het-hom) gaps. Relative stabilities of all relevant species (red dots: heterochiral, blue dots: homochiral) were plotted and are indicated by the length of blue bar, which reflects the het-hom gap of every species. To benchmark with a higher level calculation, PNO-LCCSD(T)-F12/cc-pVTZ-F12[40,196] calculations were executed for the structures responsible for the het-hom gap, corrected with the ZPVE values of their respective optimization methods.

the optimized structure using three different methods, each with and without dispersion. In the case of EDO, a range of 8.4–11.0 kJ mol^{-1} decreased to 3.4–6.1 kJ mol^{-1} once dispersion was removed. For CHexDO, a change from 6.1–9.1 kJ mol^{-1} to 4.1–6.4 kJ mol^{-1} was observed, while 5.5–9.6 kJ mol^{-1} for pinacol was reduced to at most 5.5 kJ mol^{-1} when optimization was done without dispersion.

The decrease in the energy range of the relative stability of the molecular species has also affected the heterochiral-homochiral (het-hom) gap. This gap represents the energy difference between the most stable heterochiral and homochiral structure and it is the most important value to consider regarding chirality recognition. A significant decrease in this gap, when comparing methods with and without dispersion, indicates that dispersion is a driving force of chirality recognition. As shown in Figure 3.5, the het-hom gap is represented by the blue bar and is calculated using both DFT and WFT.

Looking at the blue bars of EDO, a dramatic decrease in the het-hom gap was observed using LMP2 with no dispersion, leading to isoenergetic hetero- and homochiral structures (6.6 down to 0.2 kJ mol^{-1}), while the het-hom gap of SCS-LMP2 went down to 0.1 kJ mol^{-1} from 4.0 kJ mol^{-1}, a 3.9 kJ mol^{-1} difference. When B3LYP was used, the removal of D3 correction decreased the het-hom gap to more than half, i.e., 6.7 kJ mol^{-1} was reduced to 2.7 kJ mol^{-1}. For the CHexDO, the LMP2 gap was lowered by 4.2 kJ mol^{-1} while SCS-LMP2 gap was reduced by approximately 1/3 of its gap with dispersion (3.1 to 1.1 kJ mol^{-1}). B3LYP gap decreased more than half, i.e., 3.6 kJ mol^{-1}. Finally, in the case of pinacol, the het-hom gaps were decreased by 5.3 kJ mol^{-1}, 3.7 kJ mol^{-1} and 4.4 kJ mol^{-1} for LMP2, SCS-LMP2 and B3LYP methods, respectively. Evidently, dispersion in these three systems is a determining component in chirality recognition.

It is interesting to note that the structures responsible for the het-hom gap change depending on 1) the choice of method and 2) the presence of dispersion (see Figure 3.5). What is constant is that for structure optimizations with dispersion, het4 is always the most stable structure. For homochiral species, however, it varies depending on the method and the system. For calculations with dispersion, hom3 is the most stable structure for EDO and CHexDO while it is hom4 for pinacol. Unfortunately, hom4 for EDO and CHexDO were not included in this specific set of comparison because the removal of dispersion for both systems causes the structure to converge into something else. This inhibits a proper comparison on the effect of dispersion.

Calculations of PNO-LCCSD(T)-F12/cc-pVTZ-F12[40,196] energies provide an overview of how B3LYP, LMP2 and SCS-LMP2 energies fare in terms of predicting the het-

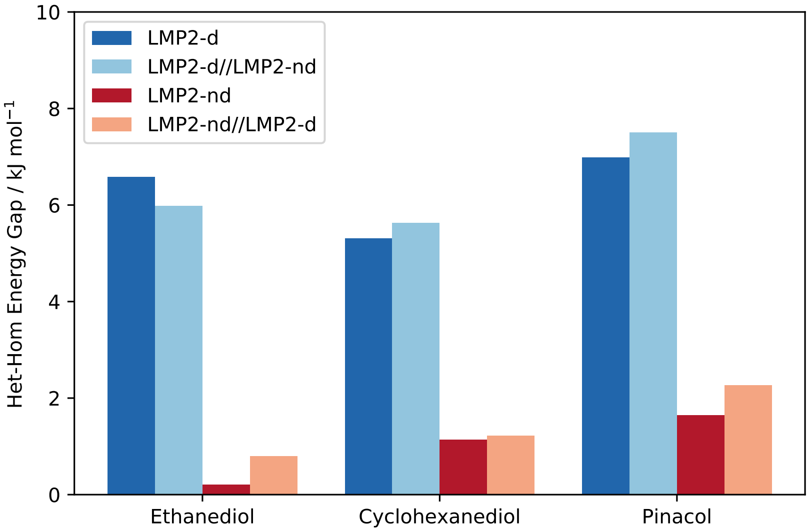


Figure 3.6: Het-hom gaps of different systems evaluated to determine if the chirality recognition is due to electron correlation or a structural effect. Different LMP2 approaches were used. Structural effects were reflected in bars colored dark blue to light blue as well as dark red to light red. Any change from blue to red indicates an electron correlation effect. When dispersion is added, **-d** is shown. Otherwise, **-nd** is written.

hom gaps. Note that this is only possible on structures optimized with dispersion, as an in-house program for PNO-LCCSD(T)-F12 without dispersion is not available. For example, for the pink bar in B3LYP in pinacol, het4 and hom4 structures were picked and had their electronic energies recalculated using the PNO-LCCSD(T)-F12 level of theory. Note that ZPVEs used were those of the structures at which they were optimized. Going back to pinacol, the trend was clear - all the WFT and DFT methods employed slightly overestimated the het-hom gap compared to the values given by PNO-LCCSD(T)-F12. A mixed behavior was observed for EDO and CHexDO where some gaps were overestimated while some were underestimated. For both systems, SCS-LMP2 significantly underestimated the het-hom gap when compared to the coupled cluster energy (approximately 3 kJ mol^{-1}).

Since the het-hom gap in Figure 3.5 is a result of optimization with and without dispersion, is the het-hom gap then a result of a structural effect or solely an electron correlation effect? To answer this question, LMP2 optimized structures responsible for the het-hom gaps for each species were subjected to a series of single point energy calculations. Results are shown in Figure 3.6. These are all ZPVE-corrected, using the ZPVE of the optimization procedure. Blue bars show the final total electronic energy with dispersion while red bars represent gaps for total electronic energy without dispersion. The shift from dark blue to light blue as well as dark red to light red reflects the structural effect. Any shift from blue to red indicates the correlation effect.

As shown in Figure 3.6, there is an insignificant structural effect on the het-hom gap, as reflected in the small energy change between dark and light colors. Note that dark blue represents LMP2 with dispersion, light blue is a single point energy calculation with dispersion on top of structures optimized with dispersion. Same goes with red bars. These results support the common approach to quantify dispersion effects – taking fixed structures and observing solely the energy differences. This happens despite the significant structural changes observed earlier.

Which Moiety Contributes the Most?

The previous discussions have clearly shown the impact of dispersion in chirality recognition, both in energies and structures of the conformations investigated. It is then of great interest to know which specific part of the dimer has a larger dispersion contribution. To help map the source, DID visualization[103] was employed. Density contribution to the interaction was calculated using PAO-based SCS-LMP2. In this section, only the structures responsible for the het-hom gaps were examined.

In Figure 3.7, the most stable heterochiral and homochiral conformations for the three systems are shown, except for pinacol where hom3_b' was chosen to be consistent with the other two molecular systems. A consistent range of density (3.0×10^{-10} to $1.5\ ea_0^{-3}$) was chosen for the visualization to ensure comparability. The red color in the map indicates a strong interaction while blue indicates a weak interaction. The DID plot shows that the OH moieties of the monomers contribute the most in terms of dispersion. More specifically, more pronounced dispersion contributions are coming from the oxygen atom, and an observable dispersion density on the hydrogen atom. This is due to the increased polarizability of the oxygen when involved in a hydrogen bond.

The het4 systems in Figure 3.7 showed larger dispersion interaction density compared to its homochiral counterpart. This means that dispersion, as a stabilizing force, is stronger in these set of conformations. Looking closely, this is because of the nature of the structure itself. The het4 structures have four hydrogen bond networks while hom3_b' family only has three. When comparing the DIDs, it becomes obvious that the OH groups are strongly interacting in het4 while in hom3_b', the density in the OHs are not as strong as in their het4 counterparts. Furthermore, the hydrogen in the free OH group in hom3_b' does not

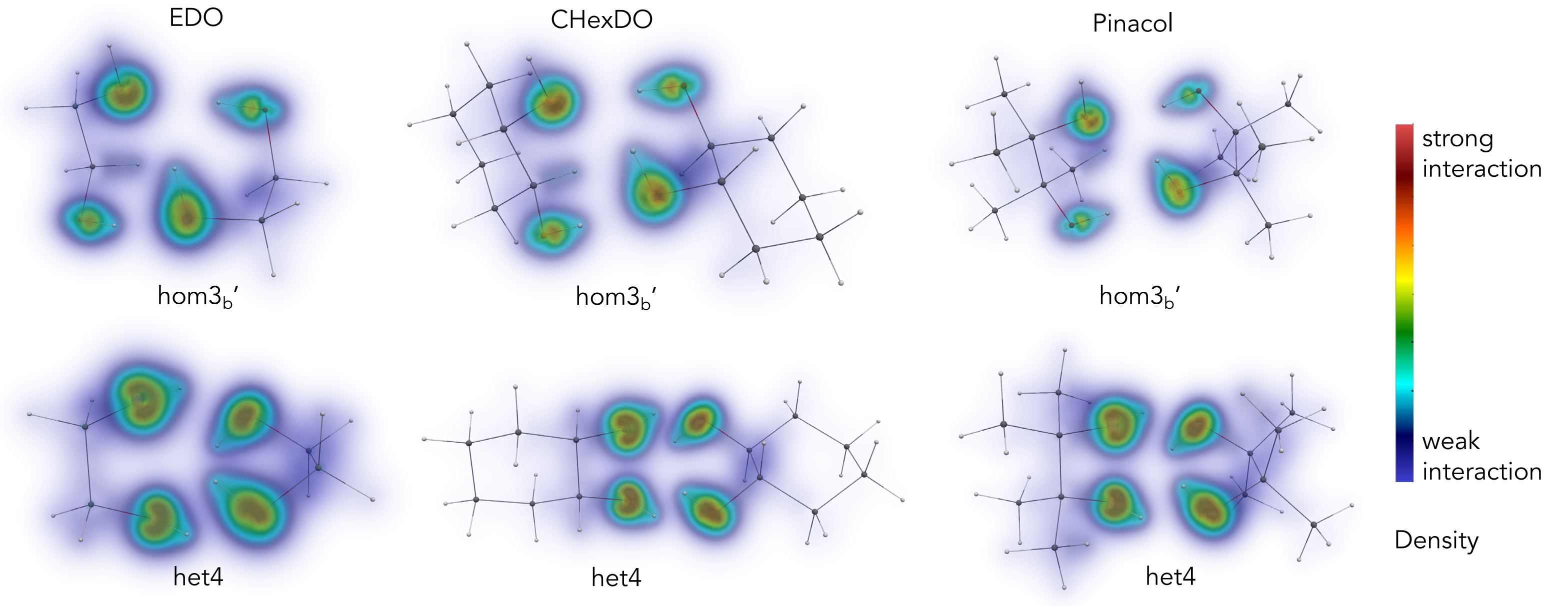


Figure 3.7: Mapping of interacting fragments due to dispersion using Dispersion Interaction Density (DID) visualization.[103] These structures are the ones responsible for the het-hom gap in each species, except for hom3$_b$' of pinacol. It was chosen anyway to allow a consistent comparison across different systems. Results from SCS-LMP2/aug-cc-pVTZ,H=cc-pVTZ optimization. Density range: 3.0×10^{-10} to $1.5ea_0^{-3}$ using ParaView 5.4.0.[197]

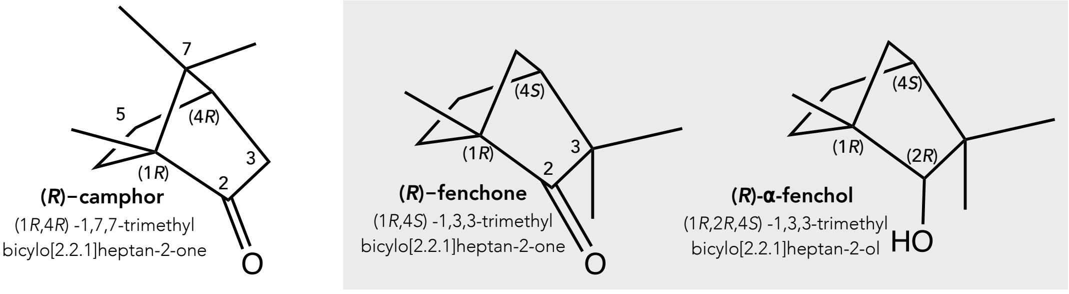


Figure 3.8: Structural formulas of camphor, fenchone and α-fenchol. The molecules in the box are considered in this study.

significantly contribute to the DID. It is notable that the density is symmetric among het4 structures due to its S4 symmetry (see Appendix A.1, Figure A.2). Since the intermolecular hydrogen bonds make the distances of oxygen atoms from different monomers closer, this leads to stronger dispersion interactions.

3.4 Energy Balances in Ketone-Alcohol Systems

As mentioned, one aspect of chirality with which chemists are very familiar is the presence of stereogenic centers. Since this type of chirality is permanent, it is interesting to see how different structural handedness affects the relative stability between homochiral and heterochiral species, revealing information about their chirality recognition. For largely rigid molecules with stereogenic centers, chirality recognition can be used in benchmark studies for theoretical methods, because one molecule is just merely mirrored and not a significant amount of deformation energy is necessary. This also means that shortcomings of the methods, e.g., not including anharmonicity, should mostly cancel.

In a recent study by Schnell and co-workers[198] regarding chirality recognition of *(R)*- and *(S)*-camphor with *(R)*-α-fenchol, they uncovered the delicate balance of intermolecular forces responsible for the stabilization of SR and RR species. They found that larger London dispersion forces are present in the SR isomer than the RR and pointed to it as a reason of the SR's more compact structure and larger rotational constants. On the other hand, the RR isomer features a slightly larger electrostatic and inductive contributions, which are correlated to its OH bonding strength. The latter is evident in the OH stretching fundamental wavenumber observed in vibrational spectroscopy, thereby allowing the discrimination between the two isomers.

Slightly changing the structure may or may not have a big effect on the chirality recognition of the system. In this section, instead of camphor as in the previous study,[198] fenchone was utilized. Fenchone and camphor are bicyclic monoterpenes. Their constitutional isomers differ in the position of their two geminal methyl groups. As shown in Figure 3.8, the geminal methyl group is located at C7 for camphor while it is at C3 for fenchone. Essentially, for fenchone and α-fenchol, they just differ in the substituent, one being alcohol and the other being ketone.

In this section, chirality recognition of *(S)*- and *(R)*-fenchone with *(R)*-α-fenchol was investigated. Specifically, the relative energies of the SR and RR configurations were given special attention and were benchmarked with the observed results of vibrational and microwave spectroscopy. A subtle balance of the forces keeping the RR and SR species was also explored using energy decomposition methods and was used to rationalize the fundamental OH stretching wavenumbers. Lastly, using DID visualization, the parts of the molecule interacting due to dispersion were mapped.

3.4.1 Computational Methods

To search for possible conformations of the SR and RR systems, a combination of manual conformational search, CREST[173,174] (version 2.11.1) based on GFN2-xtb[175,176] (version 6.4.0), and AIMD (ab initio molecular dynamics) was utilized. ORCA[189,190] 4.0.1 was used to carry out AIMD calculations, at the HF-3c[199] level of theory. The Berendsen thermostat at 298K and 450K were utilized. With the available conformational samples, geometry optimizations were done using B3LYP/may-cc-pVTZ[180,186,187,200,201] using the Gaussian 16[202] quantum chemical package. Grimme's dispersion correction with Becke-Johnson damping (D3BJ) was incorporated.[170,191] Unique species within 2 kJ mol^{-1} relative energies were picked for further analysis. To better account for dynamic correlation, high level PNO-LCCSD(T)-F12/cc-pVTZ-F12[40,196] calculations were done on top of the B3LYP structures. The ZPVEs of the B3LYP calculations were used along with the electronic energies calculated from PNO-LCCSD(T)-F12 level of theory.

Since it is of great interest to uncover the forces that govern the interactions of the molecular aggregates and have a more grounded view of the dimerization energies, energy decomposition analyses have been carried out. However, the interpretation of these forces may vary, depending on the decomposition scheme employed toward the total (interaction) energy of the system. The Kitaura-Morokuma (KM) scheme was utilized to decompose the reference HF energy.[79] The original form, however, is prone to numerical instability and basis set sensitivity, which mainly comes from the antisymmetric intermediate wavefunction. Thus, we used the KM method as extended by Su and Li,[80] which is referred here as extended KM. In this scheme, the supermolecular approach is used to calculate the interaction energy, and the latter is decomposed into four components: *electrostatic*, *exchange*, *repulsion* and *polarization*. For the dynamic correlation contribution to the interaction energy, this is decomposed using local correlation analysis. Such scheme was first proposed by Schütz *et al.*,[96] which initially used projected atomic orbitals (PAOs). This has been recently extended to pair natural orbital-based methods in the Molpro program package. One notable consideration when using PNOs as virtual spaces is their lack of intrinsic locality like PAOs. This can result to instabilities since localization of the pair domains is required. To overcome this, bundling some terms is a possible solution.[104] The procedure can be made robust by the use of Boys localization[65] and Løwdin partial charges. Thus, the original interaction classes of Schütz are still used here.[96] In this approach, the correlation part of the interaction energy is divided into three contrubutions: *dispersion* (which contains dispersion and exchange dispersion), *ionic* and *intramolecular contributions*. The final two terms are bundled into *non-dispersion* contributions.

For both energy decomposition schemes (extended KM and PNO-SCS-LMP2, see Section 2.4), an in-house code implemented in Molpro 2018.1 developer's version[136] was utilized. All calculations pertaining to energy decomposition use cc-pVTZ[194] basis set. In order to visualize which part of the fenchone and α-fenchol were interacting due to dispersion, DID visualization[103] was done on most stable structures of SR and RR systems using Molpro 2021.3 commercial version.[54]

3.4.2 Relative Stability of Dimers

For molecular species within the window of 1 kJ mol^{-1} from the most stable structure, the four most stable conformers of RR and SR species were selected for further analysis, as shown in Figure 3.9. For the SR conformers, B3LYP and PNO-LCCSD(T)-F12 energies predicted the same species, SR-I, as the global minimum. However, in the case of RR conformers, B3LYP predicted RR-IV, which is 0.31 kJ mol^{-1} lower than RR-I. The PNO-LCCSD(T)-F12 results, however, indicated otherwise, pointing to RR-I as the global minimum. Note that the energy differences of these species are very small, even below the chemical accuracy of quantum mechanical methods. Thus, vibrational and rotational spectroscopies are of great help in determining which among these structures are observed in experiment.

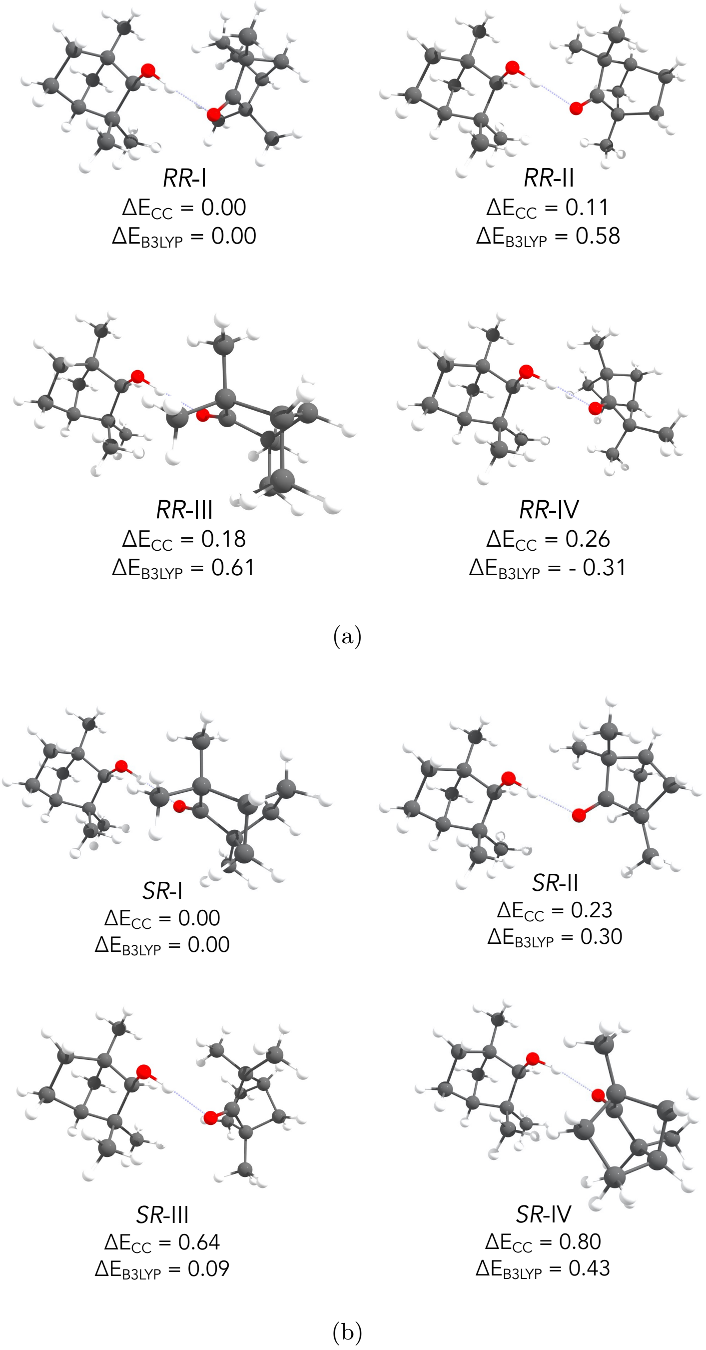


Figure 3.9: Most stable conformations of a) *RR* and b) *SR* species within 1 kJ mol^{-1} energy window. Relative energies are in kJ mol^{-1} and are ZPVE corrected. ΔE_{CC} is a high-level energy correction from PNO-CCSD(T)-F12/cc-pVTZ-F12 using Molpro 2019.1. ZPVEs for the CC energies are obtained from the B3LYP/may-cc-pVTZ calculations. Note that as much as possible, the projection of the (R)-α-fenchol remains constant so differences on how the different fenchone conformations rearrange themselves are noticeable.

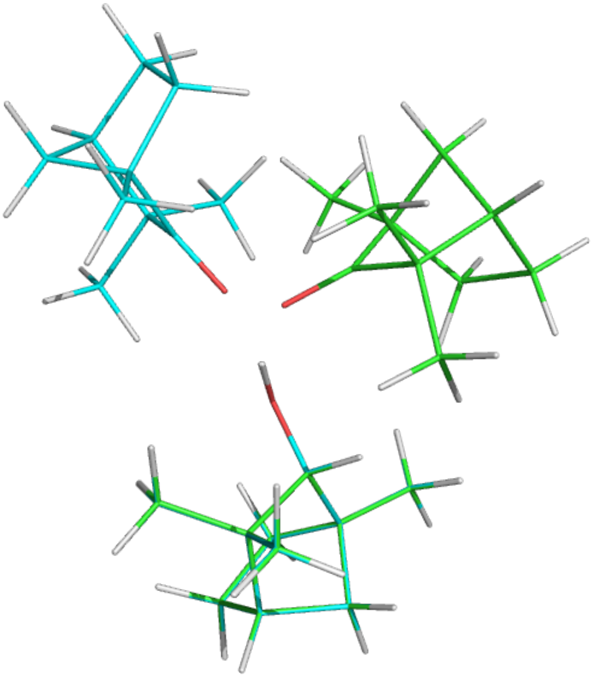


Figure 3.10: Superimposed structures of *RR*-I (green) and *SR*-I (blue). The two conformations were fitted via the fenchol molecule (bottom structure) because this is the common structure of both systems. The pair fitting was done via PyMOL.[55]

As demonstrated in Section 3.3, chirality recognition manifests through various molecular features, one of which is the energy difference between homochiral and heterochiral species. In the case of the fenchone-fenchol systems, the relative energy between *RR*-I and *SR*-I is essentially zero based on the calculations done. One cannot discriminate between the most stable complexes on the basis of energetics. On the other hand, vibrational spectroscopy (FTIR jet spectroscopy) and microwave spectroscopy results provide tangible evidence for the discernible existence of *RR*-I and *SR*-I. This evidence is elucidated by notable discrepancies in the observed band shifts and rotational constants.

According to the results of FTIR jet spectroscopy (see Appendix A.1, Figure A.3), assigned spectra for *RR* and *SR* species are on the opposite ends of the shown spectrum, i.e., *RR* -I with 3526 cm^{-1} while *SR*-I with 3470 cm^{-1}. These are associated with the OH stretching fundamentals of hydroxy groups acting as a donor for hydrogen bonds. A 56 cm^{-1} difference in the experimental results is a clear indication that these belong to different structures with different OH stretching fundamentals, a manifestation of unique structural characteristics of the system.

Experimental rotational constants for the heterochiral and homochiral species are available (see Appendix A.1, Table A.6 for the data). The experimental results match well with the predicted rotational constants of the global minima of every category, i.e., *RR*-I and *SR*-I. The experimental rotational constants corresponding to *RR*-I and *SR*-I are not very similar. Looking at the structures, this is not surprising. Shown in Figure 3.10 are the superimposed structures via the fenchol molecule. The spatial distribution of the systems shown is very different, i.e., at different structural motifs relative to the reference fenchol system.

These specific systems indicate that the usual approach of benchmarking, i.e., looking only at the energy observables like relative energies or dimerization energies, can lead to wrong interpretations. Structures and structural effects are equally important to look at, especially when there are existing high-end experimental techniques that can help elucidate how the structures of these systems look like. This further illustrates the prime importance of complementary approaches with theory and experiment in disentangling chemical phenomena, for example in explaining delicate balances of isoenergetic systems.

3.4.3 Analysis of Interaction Energies

Table 3.1: Local orbital analysis of the interaction energy of *SR* and *RR* fenchol-fenchone systems. ΔE_{elec}, ΔE_{exch}, ΔE_{rep} and ΔE_{pol} were calculated using extended KM scheme while the rest were calculated using PNO-SCS-LMP2. The interaction energy, ΔE_{int}, was calculated using the supermolecular approach at the same level of theory PNO-SCS-LMP2. The basis set used for all calculations was cc-pVTZ. All values in kJ mol^{-1}.

	ΔE(HF)				ΔE(PNO-SCS-LMP2)		E_{int}
Isomer	ΔE_{elec}	ΔE_{exch}	ΔE_{rep}	ΔE_{pol}	ΔE_{disp}	$\Delta E_{\text{non-disp}}$	
Fenchone-fenchol							
SR-I	−54.2	−75.6	135.2	−23.4	−17.7	−0.1	−35.8
SR-II	−50.5	−70.2	125.1	−22.0	−17.9	0.0	−35.5
SR-III	−44.8	−70.5	123.5	−19.3	−21.4	−0.6	−33.1
SR-IV	−48.9	−72.9	129.1	−20.9	−19.3	−0.3	−33.2
RR-I	−44.3	−72.2	126.2	−19.6	−22.5	−1.3	−33.7
RR-II	−49.0	−71.2	125.9	−21.2	−19.5	−0.3	−35.1
RR-III	−54.1	−75.1	134.4	−23.2	−17.7	0.1	−35.6
RR-IV	−45.4	−74.0	129.1	−19.3	−22.3	1.2	−33.2

Despite the *SR*-I and *RR*-I systems being isoenergetic, spectroscopic data (FTIR and MW) revealed the presence of both species and that they are distinguishable. It is then interesting to uncover which force(s) dominates one over the other, thereby leading to the observed spectra and constants. Table 3.1 shows energy decomposition analysis using the

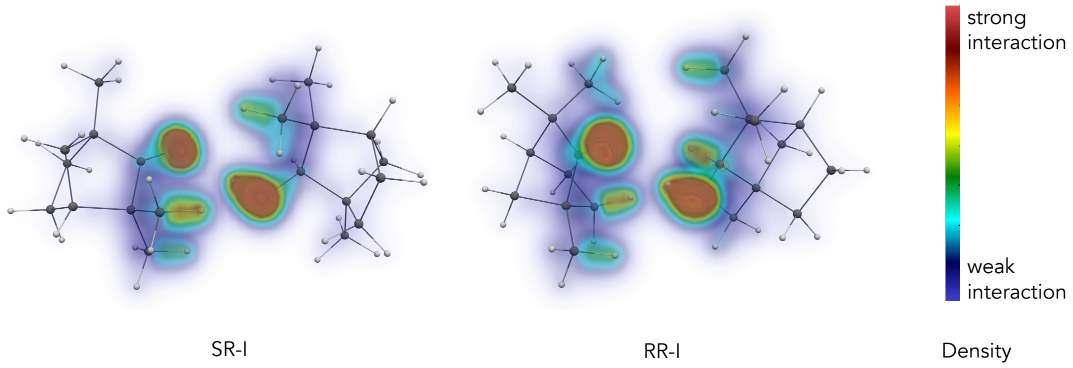


Figure 3.11: Density map of dispersion interaction using DID visualization.[103] Conformations considered are only *SR*-I and *RR*-I since they are the most stable structures of both systems. Calculations were done using SCS-LMP2/aug-cc-pVTZ,H=cc-pVTZ. Density range: 2.5^{-4} to 0.5 using ParaView 5.4.0.[197]

extended KM and PNO-SCS-LMP2 schemes. Note that the sum of individual contributions corresponds to the total interaction energy, E_{int}. As described in the methods, the ionic and intramolecular contributions are added together as non-dispersion contributions. In case one is curious of the non-bundled values, see Table A.7 in Appendix A.

Looking at the values in Table 3.1, higher (more negative) electrostatic and exchange contributions were calculated for *SR*-I than for that of *RR*-I. This can be the reason why the signature OH-stretching of *SR*-I is at lower wavenumbers (3470 cm^{-1}) than that of *RR*-I (3526 cm^{-1}). The 56 cm^{-1} experimental shift is large enough to distinguish the two bands from each other. This is nicely supported by theory, which indicated a shift difference of 67 cm^{-1} in the double harmonic approximation, or 64 cm^{-1} after rescaling (scaling factor of 0.9613 was used).

On the other hand, Table 3.1 shows that the dispersion contribution is higher in *RR*-I than in *SR*-1 by about 5 kJ mol^{-1}, as provided by the PNO-SCS-LMP2 results. Looking at their structures in Figure 3.10, there is more contact surface of the *RR*-I that can be a source of dispersion interaction. This is confirmed by the DID visualization, which shows the relevant parts of the monomers are interacting via London dispersion as they form the dimer. As shown in Figure 3.11, *RR*-I has more regions of space interacting due to dispersion (more blobs lighting up) compared to *SR*-I. Based on the *RR*-I structure, on top of the OH groups nearby each other, some −CH moieties are also nearer in proximity, thus contributing to the dispersion interaction. Note that dispersion is a function of the number of electrons near each other which can interact due to instantaneous dipole, as well as the polarizability of the moieties available.

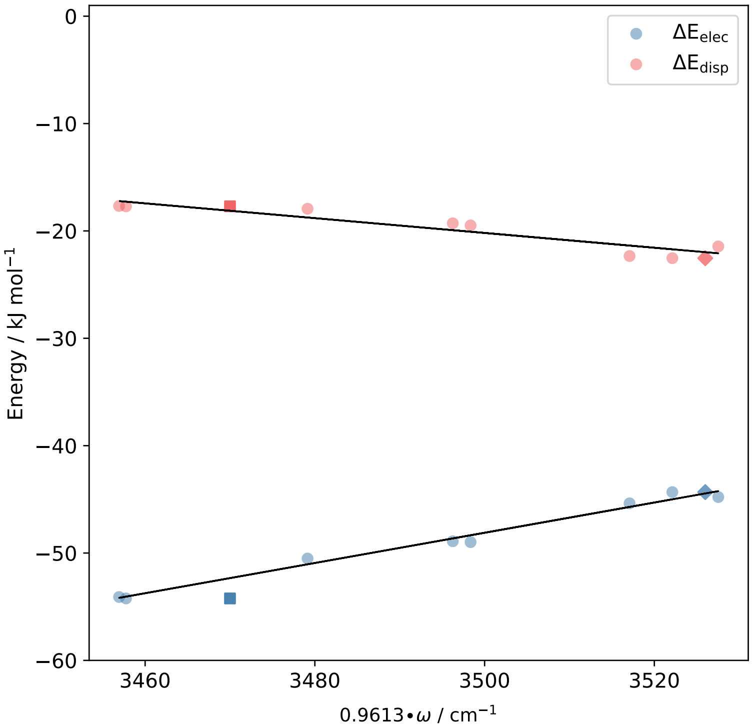


Figure 3.12: Behaviour of selected intermolecular energies versus the computed scaled harmonic frequencies. ΔE_{elec} (light blue) and ΔE_{disp} (light red) as a function of the experimental (■ (*SR*-I) and ◆ (*RR*-I)) or scaled harmonic (●) OH stretching wavenumber.

Overall, the total interaction energy of the dimers does not correlate with their relative stability. This set of dimers exhibits a small energy range concerning their relative stability, effectively making them isoenergetic. Moreover, their respective interaction energies display only marginal differences. For instance, in the *RR* category, the maximum difference of E_{int} is 2.4 kJ mol^{-1}, while it is 2.7 kJ mol^{-1} for the *SR* counterpart. It also does not necessarily reflect that the global minimum has the strongest interaction (more negative interaction energy). This is true for the *SR*-I but not for the *RR*-I. It is important to note that these differences fall below the chemical accuracy of the methods used. On the other hand, these dimers are structurally very different, particularly in the spatial arrangement of *S*- or *R*-fenchone with respect to fenchol. The decomposition of their interaction energies revealed a more notable difference in contributions, highlighting that there is a delicate interplay of different components in the interaction energy.

The energy decomposition analysis in Table 3.1 can also provide insights on the behaviour of selected non-covalent force in relation to the observable indicators of hydrogen bond strength. Figure 3.12 illustrates the relationship regarding the dynamics of electrostatics and dispersion forces with respect to both experimentally observed and computationally determined scaled stretching wavenumber. As shown in the graph, the light red dots indicate a negative correlation between dispersion forces and increasing scaled frequencies. It displays a substantial correlation with a correlation coefficient of -0.93. In stark contrast, the same graph shows that the electrostatic contributions have a positive correlation with the increasing trends in experimental and calculated harmonic O−H stretching. In this regard, a 0.99 correlation coefficient indicates a robust correlation between these parameters.

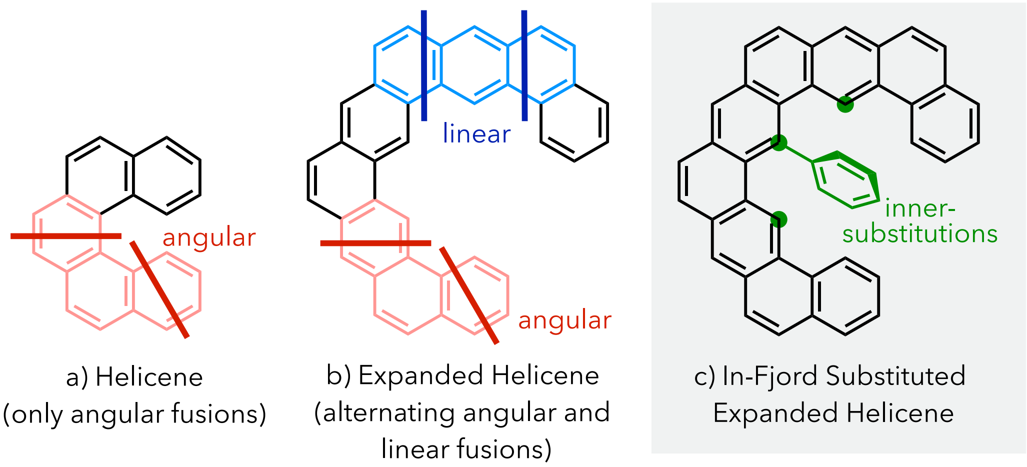
The fenchol O-H stretch frequency in the complexes is found to be red-shifted in good correlation with the electrostatic component of the interaction. This relates well to the common assumption that hydrogen-bonds with larger electrostatic contributions are, in fact stronger, implying a weaker O−H bond (lower force constant). In contrast, there is an inverse relationship between the hyrodgen bond strength and dispersion contributions among these systems.

3.5 Inversion Mechanisms in Substituted Helicenes

In the preceding two sections, we probed into the intricacies of chirality recognition with a focus on the dynamic interplay within molecular aggregates displaying either transient or permanent chirality. In this section, the main point of discussion revolves around the configurational stability of the systems and its influence on the enantiomerization from one chiral form to the other. As we transition from our prior discussion of dimers, we now direct our attention to an interesting system: the expanded helicene molecule. This large molecular structure derives its chirality from a unique combination of chiral molecular strain and the presence of axial chirality.

Helicenes and their related compounds represent a distinctive class of polyaromatic hydrocarbons (PAHs) that have garnered considerable scientific interest. Unlike other PAHs like acenes and nanographenes, helicenes are characterized by unique properties:

chirality originating from their non-planar, helical structure. The handedness of the helicity is a consequence of the presence of axial chirality. Consequently, the clockwise and counterclockwise helices are non-superposable. This family of compounds is of great interest because of their unique properties they demonstrate in areas such as molecular recognition[203,204] molecular machines[205] and asymmetric catalysis,[206,207] to mention a few.



a) Helicene (only angular fusions)

b) Expanded Helicene (alternating angular and linear fusions)

c) In-Fjord Substituted Expanded Helicene

Figure 3.13: Geometric and structural features of the different families of helicenes. Classical helicene is shown in a), while expanded helicenes with alternating angular and linear fusions are depicted in b). The helicenes discussed in this section are a form of expanded helicene with in-fjord substituent, as presented in c).

The skeleton of helicene is typically composed of *ortho*-fused rings, as shown in Figure 3.13a. The helical structure is induced due to the steric hindrance between its terminal rings, making its backbone twist in opposite directions. Typically, helicenes are denoted as $[n]$helicene, where n is the number of rings. According to IUPAC, only the structures with $n > 4$ are helicenes.[137] The unsubstituted [4]helicene, due to its poor configurational stability, has never been resolved into enantiomers.[208] [5]helicene displays partial configurational stability, making it a borderline case. This means that it can be resolved into enantiomers but then racemizes within a couple of days at room temperature. Configurationally stable $[n]$helicenes with $\mathrm{n} \geq 6$ can be resolved at room temperature, and $\Delta G^{\ddagger}(T)$ for inversion increases with increasing number of *ortho*-fused rings.

Configurational stability is one of the most important stereodynamic feature of helicenes, and is usually quantified in terms of the Gibbs activation energy, $\Delta G^{\ddagger}(T)$. This value is a metric if the compound can or cannot be resolved into enantiomers under ambient conditions. The higher $\Delta G^{\ddagger}(T)$, the better access there is to the chiral functional materials from helicenes. This makes the mechanism of helicene inversion an interesting feature. In 1970s, Martin *et al.*[209] proposed three hypotheses for the helicene interconversion: 1) bond breaking at the inner helix, 2) intramolecular double Diels-Alder reaction and 3) *'conformational pathway'*. The first one was easily ruled out because this would lead to a diradical species in which the C−C bond breaking would require a huge energy. The second one, involves an intermediate which, in principle, can be detected by ^{1}H-NMR spectroscopy. However, this intermediate was not observed in the spectrum so this hypothesis was discarded. Recent studies have shown that intramolecular Diels-Alder reactions were attainable but unfavorable.[210–212] The last hypothesis, i.e., via conformational pathway, is the most reasonable possibility arguing that helicenes are more flexible than previously believed. This hypothesis justified the low potential barriers observed until nonahelicenes since necessary molecular deformations are spread over a large number of bonds. This justification was supported by several theoretical simulations over the years, as discussed below.

The inversion of [*n*]helicenes were revisited by Grimme and Peyerminhoff[208] in 1994 for $n = 3 - 6, 8$. They used semiempirical AM1, *ab initio* SCF and density functional theory calculations using BLYP functional (basis set: SV+d+p).[213] In their study, they affirmed the flexibility of the fused benzene rings. The older notion about fused benzene rings is that they are rigid and would prefer a planar structure. They also validated the increasing barrier of helicenes with increasing rings until $n = 6$. Beyond that point, an increase led to a plateau of energy. Unfortunately, their predictions led to large overestimation using *ab initio* SCF method, with $10 - 14$ kcal mol^{-1} error for $n > 5$. However, BLYP calculations were able to approximate the barrier height of inversion with an accuracy of ± 1 kcal mol^{-1}. It is important to note that the activation barriers referred to here are activation enthalpies, $\Delta H^{\ddagger}$.

Janke *et al.*,[214] in 1996, reevaluated the inversion barriers of [*n*]helicenes with $n = 5 - 9$. They used several computational approaches but this discussion only highlights their results using B3LYP/3-21G//B3LYP/3-21G results (other approaches are semiempirical methods). Their reference experimental results showed a plateau of activation barrier ($\Delta H^{\ddagger}$) from $n = 7 - 9$. The increase of barrier with increasing *n* was supported by the calculations, although within maximum deviation of approximately 3 kcal mol^{-1}. Of particular interest is the inversion path of most unsubstituted helicenes. For most cases, it follows a one-step mechanism with C_2 ground state and C_s transition state. However, in the case of [9]helicene, it follows a three-step mechanism as well as an intermediate

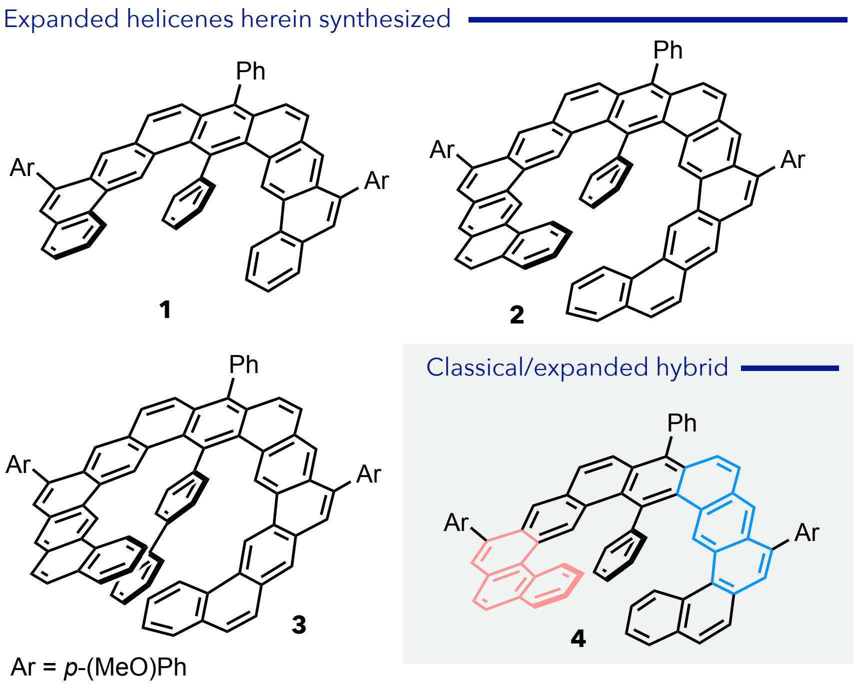


Figure 3.14: Helicoidal structures synthesized in the work of Suárez-Pantiga *et al.*[108]

structure near the C_s transition state in the second step. Furthermore, this C_s transition state in the second step has a lower energy than the other two transition state structures found.

The most recent theoretical explorations in helicenes were done by Barroso *et al.*[215] in 2017. In this study, they scrutinized the inversion mechanism of $[n]$helicenes with $n = 4 - 24$. For this analysis, they used PBE0/def2-TZVP//PBE0/6-31G(d), where dispersion effects were accounted by using Grimme's D3[170] correction. They validated that for $n = 4 - 7$ the inversion mechanism follows a single-step pathway but with $n \geq 8$, the process involves multi-step pathways with $2n - 14$ intermediates. These computations showed that the inversion mechanism is more complex than previously conceived, particularly when more rings are involved. More importantly, their results underscored the delicate effect of the steric hindrance and the number of π interactions as crucial factors influencing the changes in inversion barrier heights.

The helicenes with *ortho*-fused benzene rings possess a fixed diameter of the inner and outer helix. In contrast, expanded helicenes, as depicted in Figure 3.13b, exhibit an alternating linear and angular connectivity. One significant distinction between expanded helicenes and the 'orthodox' ones is the helix diameter. The order of *ortho* and *meta*-fused rings dictates the diameter of the helix. Unfortunately, with the expanded helix diameter comes the lower configurational stabilities. Up to date, no configurationally stable helicene which follows the alternating pattern in 3.13b has been reported, although the ones which follow angular-angular-linear fusions are available.[216–221] The synthesis of helicene-type compounds with larger radius, featuring sufficiently high barriers to be resolved into enantiomers, remains a formidable challenge.

Recently, Suarez-Pantiga *et al.*[108] synthesized a series of expanded helicenes of different sizes and shapes. Unlike ordinary unsubstituted expanded helicenes in Figure 3.13b, they incorporated phenyl and bi-phenyl substituents at the deepest part of their fjord, as shown in Figure 3.13c. The products (see Figure 3.14) were accessed via sequential Au-catalyzed hydroarylation of appropriately designed diynes. Most experimental work done, such as synthesis, x-ray crystal structure as well as photophysical characterization, can be found in the published article.

In this section, the helical inversion processes were assessed using state-of-the-art quantum chemical methods. Specifically, the potential energy surface during inversion was evaluated, and density mapping of London dispersion was conducted to elucidate how certain in-fjord substituents enhance the system's stability in comparison to non-substituted helical structure.

3.5.1 Computational methods

Density functional theory (DFT) was employed to explore the conformational dynamics of the synthesized compounds. While the compounds presented in this study were all substituted, the conformational dynamics of their unsubstituted counterparts, specifically those with no substitutions in their fjord region, were also examined to gain insights into the impact of inner substitution. Substituents on the outer rim were removed to expedite computation.

The structures were optimized at the PBEh3c[222] level of theory, and the energies were subsequently recomputed using B2PLYP,[223] with both calculations conducted using the def2-TZVP[178] basis set and default fitting basis. The inclusion of D3 corrections with Becke-Johnson damping was integral to these computations.[170,191] To construct the inversion profile, nudged elastic band (NEB) calculations were carried out where the energy estimates for the barrier was provided by the climbing image (CI). In cases where the inversion mechanism passed through a C_s-symmetric intermediate, the energy profile was only calculated in one part of the mechanism, specifically from the optimized ground state geometry to transition state structure and, subsequently symmetrized. The energy barriers presented in the results are determined based on the electronic energy difference between the global structure minimum and the NEB CI. Convergence criteria were set for the maximum component of the atomic force acting on the CI, with the tolerance of 5×10^{-4}, ensuring for a tight convergence close to the maximum. Note that the deviations from the optimized transition states were found to be below 2 kJ mol^{-1} (for example, $\mathbf{[1]_{Ph}}$ and $\mathbf{[4]_{H}}$). All these computational procedures were carried out using the Orca 4.2.1 program package.[189,190]

For the calculation part, the structures are represented by the number of the the compounds as indicated in Figure 3.14 enclosed in brackets, with the inner substituent indicated as a subscript. Intermediates are designated as **INT** while transition state structures are labeled as **TS**. Unless specified otherwise, the structure represents a reactant state structure. The energies reported are in electronic energies, unless specifically noted otherwise. The difference between the latter values and Gibbs free energy barriers are minimal, mainly because no bond breaking and bond formation happened which would significantly impact other energetic contributions.

In order to understand the trends of some mechanisms, dispersion interaction density (DID)[103] analyses of the structures were done. The energies were computed at the SCS-LMP2/cc-pVTZ[26,194] level with the Molpro 2020.1 program package.[224]

3.5.2 Inversion Mechanism

From the synthesized helicenes in the work shown in Figure 3.14, this section gives a detailed discussion of the inversion mechanism of every compound. Analyses are mainly focused on the configurations of reactant and transition state structures as well as the inversion mechanism and the barrier associated with it. Note that substituents outside the expanded helicenes were replaced by H atoms to reduce computational costs.

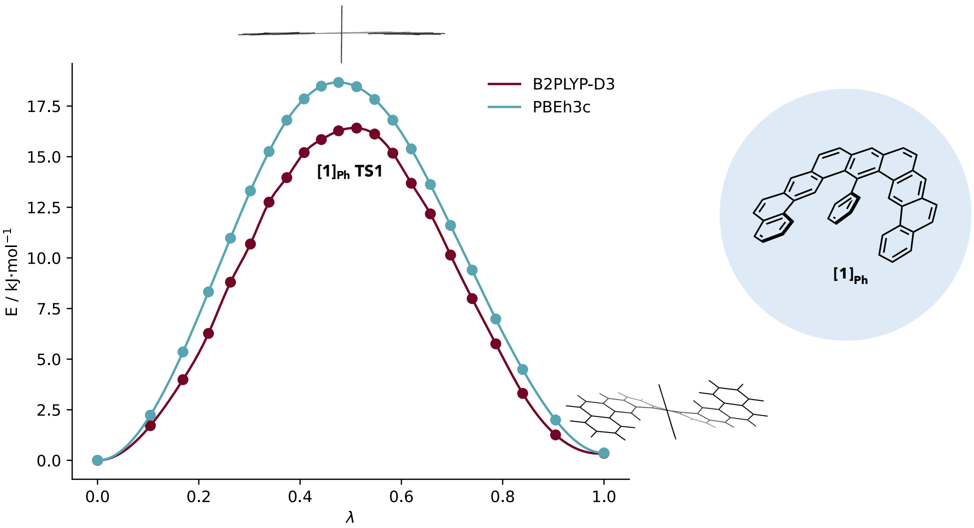


Figure 3.15: Enantiomerization profiles of Compound **1**, labeled as $\mathbf{[1]_{Ph}}$. Energy profiles were provided using NEB calculations. The independent coordinate λ is the normalized abelian distance between atoms from reactant to transition state.

Compound **1**, $\mathbf{[1]_{Ph}}$, is the 'shortest' expanded helicene examined in this chapter. A phenyl group is inserted in-fjord position, and the inversion profile is shown in Figure 3.15. The inversion process is characterized by a one-step mechanism involving a C_2 symmetric reactant state and a C_s symmetric transition state, wherein the phenyl group is perfectly perpendicular to the benzene ring to which it is attached to. Note that its mechanism is similar to those of $[n]$helicenes where $n \leq 7$.[215] The symmetry of the reactant and transition states also resembles that of the 'orthodox' helicenes. Despite its considerable number of rings in the helix (9 rings), the alternating linear and angular fusions enable this compound to possess a sufficiently large helix diameter, which allows its inversion to proceed seamlessly, even with the in-fjord phenyl substituent. Indeed, the barrier (calculated in the NEB profile) is quite small, only 15.9 kJ mol^{-1} (free barrier for activation of 19.1 kJ mol^{-1}).

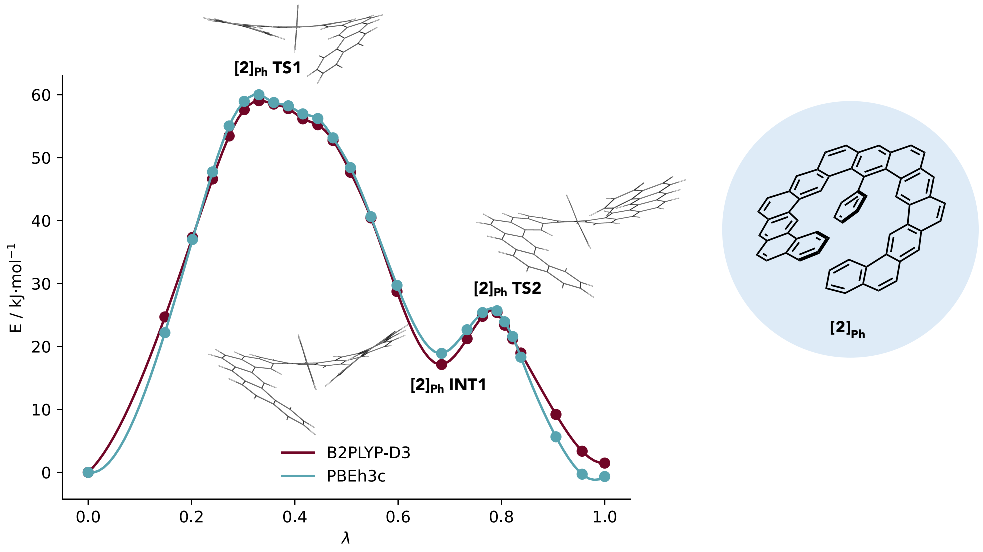


Figure 3.16: Enantiomerization profiles of Compound **2**. Energy profiles were provided using NEB calculations. The independent coordinate λ is the normalized abelian distance between atoms from reactant to transition state.

Moving to Compound **2**, here labeled as $\mathbf{[2]_{Ph}}$, this features a longer expanded helicene arm compared to $\mathbf{[1]_{Ph}}$, yet still bearing one phenyl substituent in-fjord. Its enantiomerization process is presented in Figure 3.16. A glance at its profile already reveals that this is not the typical inversion profile observed in helicenes, including other expanded helicenes. Firstly, the enantionerization profile does not exhibit symmetry. In this two-step mechanism, the initial step involves a relatively significant barrier while the second step is merely a small elevation. The transition states lack definite symmetry either, and only the reactant enantiomer adhere to the more ‘typical’ C_2 symmetry. The barrier has been computed to be $\Delta E^{\ddagger} = 59.0$ kJ mol^{-1}, closely resembling the barrier of its unsubstituted counterpart (i.e. no phenyl ring in the middle), which is $\Delta G^{\ddagger} = 54.3$ kJ mol^{-1}.[217] This implies that the addition of phenyl ring does not significantly affect the height of inversion barrier, which can be attributed to the diameter of the ring. Despite the extension of the arm length compared to $\mathbf{[1]_{Ph}}$, the sequence of angular and linear fusion allows for an increase of the diameter of the helix, where the phenyl ring in the fjord position has minimal interaction with the arms. This interaction, or the lack thereof, is a primary factor influencing the heights of the inversion barriers. In terms of configurational stability, the inversion barrier remains quite low, making the conversion between chiralities relatively facile.

If a single phenyl substituent does not have any effect in the inversion barrier due to its

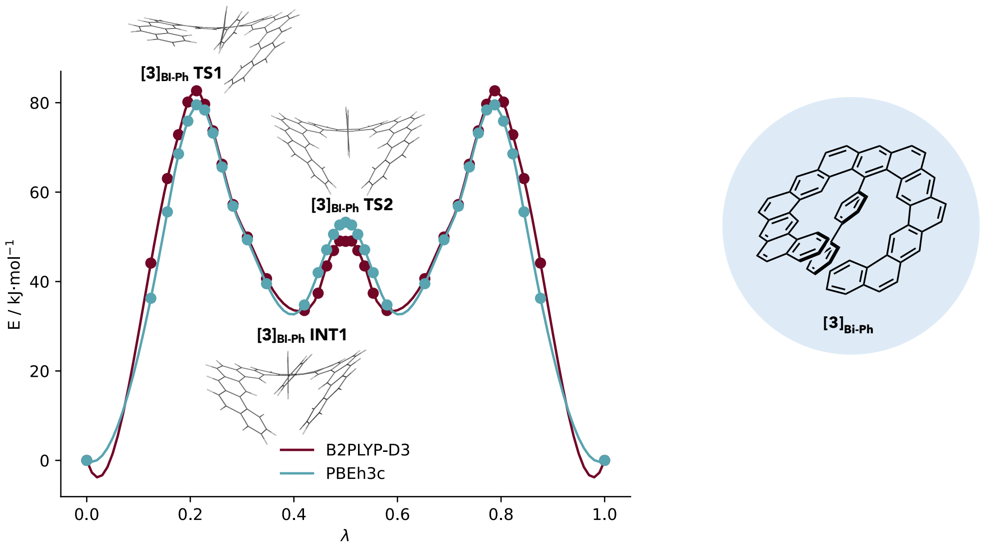
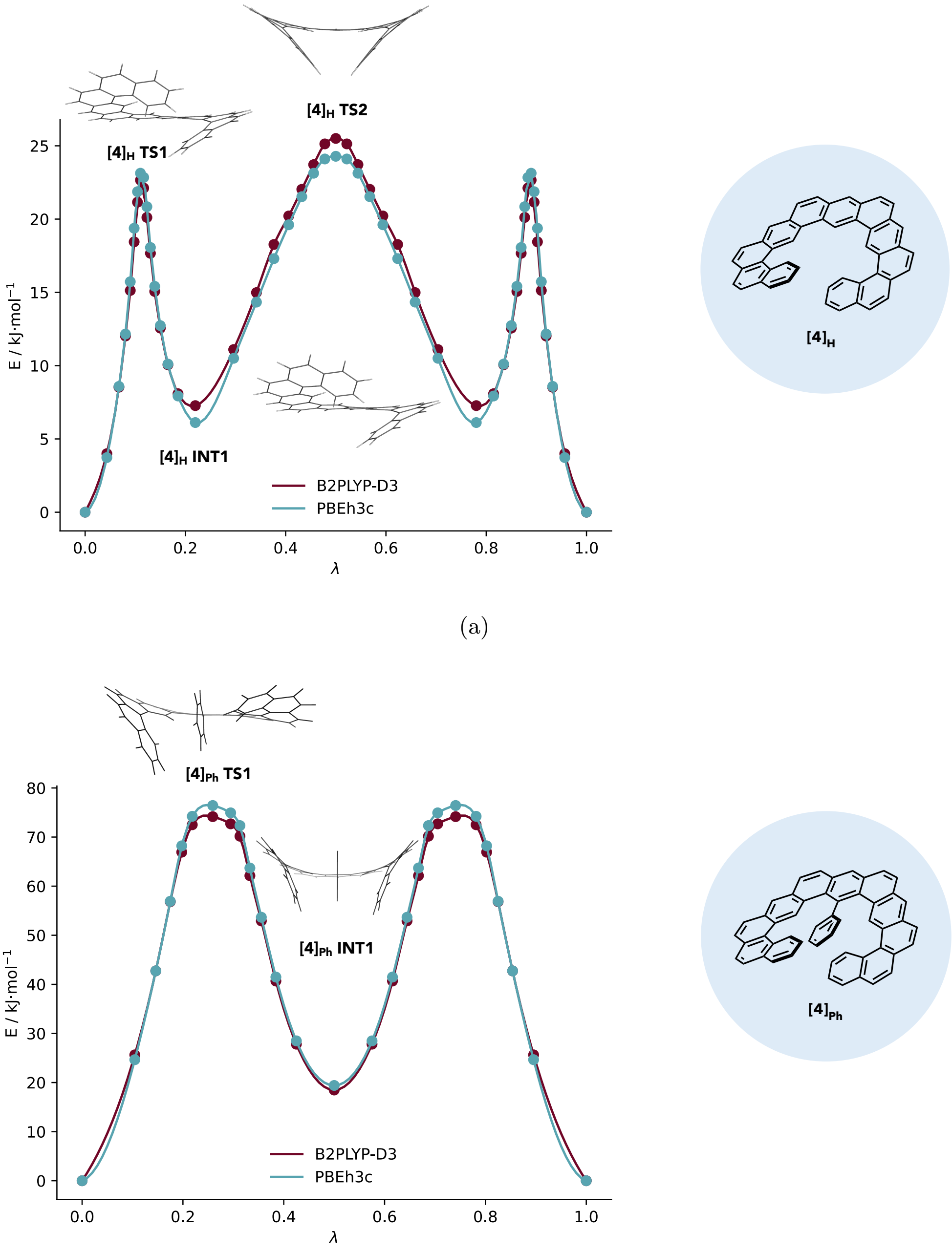


Figure 3.17: Enantiomerization profiles of Compound **3**. Energy profiles were provided using NEB calculations. The independent coordinate λ is the normalized abelian distance between atoms from reactant to transition state.

diameter size, what if there is a substituent in the middle which can interact with the arms? This is precisely what **[3]$_{\textbf{Bi-Ph}}$** looks like. In terms of the helix structure, it is identical to Compound **2**. The only difference lies in the substituent in the middle, where, instead of a single phenyl ring, a biphenyl ring is incorporated. A brief examination of its enantiomerization process in Figure 3.17 reveals a symmetric, three-step mechanism. The profile closely resembles that of the nonahelicene, which was simulated by Janke *et al.*[214]As expected, the reactant maintains a C_2 symmetry. **TS1** (which is the mirror image of **TS3**) lacks a specific symmetry, but one arm is visibly reaches a planar position while the other arm remains in its position. **TS1** also displays the highest barrier, i.e., $\Delta E^{\ddagger} =$ 82.7 kJ mol^{-1}, around 20 kJ mol^{-1} increase to that of **[2]$_{\textbf{Ph}}$**. This can be considered an intermediate case, not extremely low but not exceptionally high either, making it challenging to resolve the two enantiomers. Two factors contribute to this situation: 1) the π interaction of the end phenyl ring in the biphenyl substitutent, which interacts with the arm starting when the inversion is complete, and 2) the need for a more abrupt rotation of the internal substituent compared to **[2]$_{\textbf{Ph}}$** to reduce steric hindrance. In the second step, this is where the C_s symmetric transition state is observed. The barrier from **[3]$_{\textbf{Bi-Ph}}$ INT1** to **[3]$_{\textbf{Bi-Ph}}$ TS2** is very low. The overall profile of the inversion shows that the arms operate independently upon inversion, one after the other.



(a)

(b)

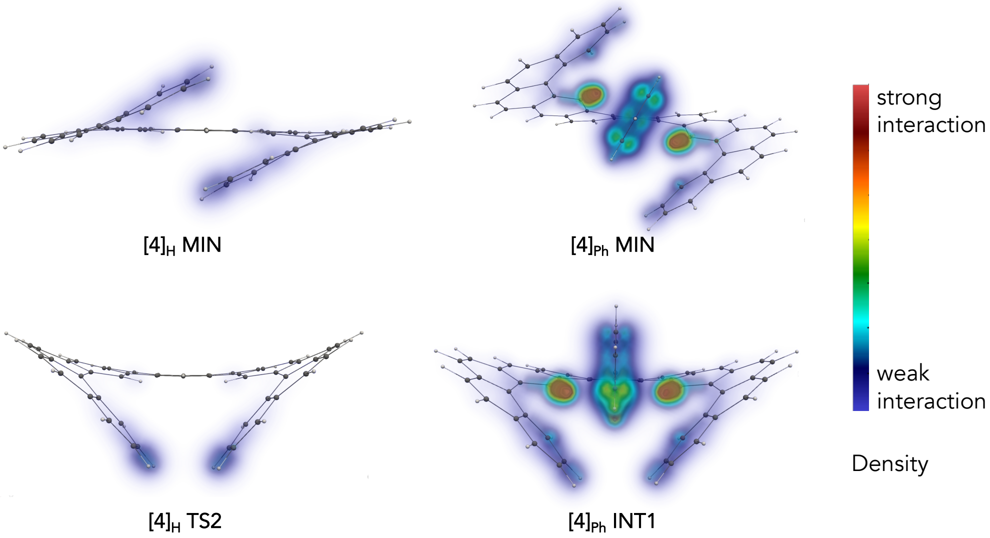
Figure 3.18: Enantiomerization profiles of Compound **4**, (a) without substituent and (b) with phenyl substituent. Energy profiles were provided using NEB calculations. The independent coordinate λ is the normalized abelian distance between atoms from reactant to transition state.

A shorter arm with fewer linearly fused rings as the previous molecules essentially combines the characteristics of classical and expanded helicenes. Compound **4**, both with and without in-fjord phenyl substituent, is an example of this hybrid. Due to its structural configuration, Compound **4** (in general) is more rigid than Compounds **1-3**. To gain a better understanding of its behavior, the inversion processes for both **[4]$_{\mathbf{Ph}}$** and **[4]$_{\mathbf{H}}$** were investigated.

The inversion mechanism of **[4]$_{\mathbf{H}}$** is straightforward, as shown in Figure 3.18a, essentially following a 3-step mechanism. Initially, one of the arms undergoes inversion via **[4]$_{\mathbf{H}}$ TS1**, resulting in an intermediate state where the inverted arm points towards the opposing arm, which remains stationary at this stage. This transition continues to a C$_s$-symmetric **[4]$_{\mathbf{H}}$ TS2** (transition state) exhibiting a barrier similar to **[4]$_{\mathbf{H}}$ TS1**. The second TS shows two arms facing each other, pointing in the same directions. A comparable process involving the other arm completes the inversion. The calculated inversion barrier for this process was calculated to be $\Delta E^{\ddagger}$ = 25.5 kJ mol^{-1}, determined by the barrier height from **[4]$_{\mathbf{H}}$ TS2**.

Intriguingly, the in-fjord insertion of phenyl substituent significantly altered the shape of isomerization profile, shifting it from a 3-step mechanism to a 2-step mechanism. It is worth noting that in Figure 3.18a, three transition states were identified. However, in this case, only two symmetric but relatively high barriers were observed ($\Delta E^{\ddagger}$ = 74.2 kJ mol^{-1}). The transition state **[4]$_{\mathbf{Ph}}$ TS1** is similar to **[4]$_{\mathbf{H}}$ TS1**, where one arm moves to a semi-planar position relative to the backbone where the substituent is attached. However, the C$_s$-symmetric structure with two arms facing each other is no longer the highest transition state as in **[4]$_{\mathbf{H}}$ TS2**. Instead, this becomes a minimum, leading to an intermediate **[4]$_{\mathbf{Ph}}$ INT1**. This might seem counterintuitive because the Ph-substituent would be expected to increase the steric strain, leading to a more pronounced deformation the the helical framework. However, other factors come into play. To provide a more comprehensive insight of what is happening, it is important to note that the Ph-substituent introduces a penalty when one arm points towards the center of the helix. The barrier heights between **[4]$_{\mathbf{H}}$ TS1** and **[4]$_{\mathbf{Ph}}$ TS1** is 22.7 versus 74.2 kJ mol^{-1}, respectively, representing a significant difference of 51.1 kJ mol^{-1}. Such comparison is fair given the structural similarities. This difference can be attributed to the increased steric clash in **[4]**$_{\mathrm{Ph}}$.

Taking a closer look, it becomes evident that the Ph-substituent in the C$_s$-symmetric structure actually serves as a slightly stabilizing element. The **[4]$_{\mathbf{Ph}}$ INT1** is only 18.5 kJ mol^{-1} above the energy minimum, while the same structural motif in **[4]$_{\mathbf{H}}$** is



$[4]_H$ MIN $[4]_{Ph}$ MIN

$[4]_H$ TS2 $[4]_{Ph}$ INT1

Figure 3.19: Map of dispersion interaction density for **[4]**H and **[4]**$_{\text{Ph}}$, depicting the interactions between the arms and the inner substituent in the latter case. Visualization was done using ParaView 5.4.0.[197]

25.5 kJ mol^{-1} above the starting configuration, and is in fact a transition state. The reduced energy difference in the former structure is a manifestation of the attractive London dispersion forces that come into play between the terminal parts of the arm and the central phenyl group. Despite the presence of a phenyl substituent in a helix with a smaller diameter compared to Compounds **2** and **3**, the inversion energy is relatively low, not permitting the resolution into its respective enantiomers.

3.5.3 Dispersion's Role in Stabilizing Intermediates

Mapping out the density resulting from dispersion interactions proves invaluable in elucidating the specific atom groups where such a force is in effect. Quantified at the SCS-LMP2 level, the DIDs of **[4]$_{\mathbf{H}}$** were evaluated in comparison with its substituted counterpart, **[4]$_{\mathbf{Ph}}$**. For the latter, the interaction with the arms and the phenyl group in the fjord position was analyzed. As seen in Figure 3.19, the presence of the substituent leads to a radical change in the density. Calculations revealed that for **[4]$_{\mathbf{Ph}}$ INT1**, the dispersion contribution amounts to 29 kJ mol $^{-1}$, while **[4]$_{\mathbf{H}}$ TS2** only exhibits 3 kJ mol $^{-1}$. Thus, the London forces due to the presence of phenyl insertion effectively compensates the energetic penalty of **[4]$_{\mathbf{Ph}}$ INT1**, mainly because of the more pronounced deformation of the polyaromatic skeleton. Such deformation favors the electronic interaction via London dispersion. This attractive force is also significantly operating in **[4]$_{\mathbf{Ph}}$ TS1**. Despite the inversion barrier height of 74.2 kJ mol^{-1} in **[4]$_{\mathbf{Ph}}$**, the latter is significantly lowered by dispersion since this interaction is partially present between the phenyl center and the terminal end of one arm of the helicene.

The inversion mechanism of helicenes is a fascinating subject, especially when in-fjord substituents are involved. Unlike the typical helicenes, several factors are crucial in the inversion of expanded helicenes, as elaborated in this section. Firstly, the sequence of ring fusion is very important. The series of angular and linear fusion is relevant to the diameter of the helix. The bigger the diameter, the less likely the arms would interact, the lower the inversion barrier will be. Secondly, the presence of in-fjord substituents is important, but their impact is only observed when these substituents can interact with the helicene arms. This interaction is evident in the mapped inversion barrier of helicenes [**2**]$_{\mathbf{Ph}}$ and [**3**]$_{\mathbf{Bi-Ph}}$, as illustrated in Figures 3.16 and 3.17, respectively.

In scenarios where interactions occur between helicene arms (e.g., Compounds **2, 3** and **4** with Ph-substituent), steric hindrance is a major factor resulting in a higher inversion barrier. However, one should not neglect the stabilizing force of dispersion stemming from π-π interactions between benzene rings. The latter stabilizes the reactant state, needing more energy to move towards the transition state. The stronger the dispersion forces come into play, the more difficult it is to transition from the reactant structure to the transition state, leading to high energy barriers.

In the context of molecular design, several recommendations emerge. One is that depending on the sequence of ring fusion, the resulting arm length which allows the arm interactions will more likely lead to a higher inversion barrier due to steric hindrace. In-fjord substitution is also a clever way of stabilizing the reactant, provided that it interacts with the arms, enhancing the reactant's stability and influencing the conversion between isomers. One can take advantage of adding substituents which will facilitate an increase of dispersion interaction since it was shown that this force is vital in the inversion process. In this manner, the inversion barrier will be high, leading to an expanded helicene which can be resolved experimentally.

Chapter 4

A New Perspective on Dispersion Interaction Density: oDID

4.1 Introduction

Based on the findings presented in Chapter 3, it is imperative to examine carefully not only the energies but also give due consideration to structural attributes. This is particularly significant, given that high-end tools for structure elucidation are readily available and quite effective for theoretical benchmarking. For some systems, the changes in energy effectively imply changes in structures as well. However, utmost care must be taken because in cases of isoenergetic species where energy differences are little to none, the structural motifs can be starkly different. Most importantly, when making conclusions regarding a certain phenomenon, e.g., chirality, it is highly advised to look not only on the energy observables but also at structural changes which may not be obvious with energy differences.

Particularly prominent in non-covalently bound systems is the ease of changing structural motifs without much change in energy. This is seen in the results in Section 3.4, and is observed in some other dimer systems.[198,225] This is because, unlike in covalent bonding, there are more subtle effects in non-covalent interactions which affects the electronic structure of the interacting species only to a slight extent. That means the dominant intermolecular force in one conformer might be lower (or higher), and it is compensated by a different force in its other conformer. The net result leads to a close to zero-sum, as reflected in their relative stabilities of these conformers.

Leveraging advanced spectroscopic techniques like FTIR jet spectroscopy and microwave spectroscopy, one can discern these subtle structural differences, which may not be readily apparent in terms of their energy characteristics. More importantly, such results are amenable to quantum chemical benchmarking. Given the recent advances in theoretical frameworks, interaction energies are not only captured accurately but can also be decomposed into several categories providing tangible physical interpretation. Of course, since these distinct intermolecular forces are not experimentally measurable, theoretical approaches can provide an approximate overview of the different intermolecular forces at play.

Unfortunately, describing the interactions in non-covalently bound systems is challenging. This is mainly because the forces in these interactions are non-observable terms so there is no experiment that can track their behaviour which will serve as a benchmark for the models developed. Amidst the difficulty, there are several approaches to assess them. Some methods like SAPT,[83] and the Interacting Quantum Atoms (IQA)[226] framework proceed by partitioning the Hamiltonian and the resulting integrals. The development of real space approaches were done which enabled the revision of some classical concepts. Such approaches are based on a chemically grounded scalar or vector field. Several examples include restricted partitioning,[227] Quantum Theory of Atoms in Molecules (QTAIM),[228] Electron Localization Function (ELF)[229] and NonCovalent Interaction (NCI) index.[230]

This chapter is dedicated to the development of a method for isolating and analyzing the London dispersion forces between interacting fragments. Qualitatively, the method allows the visualization of dispersion as a contact surface between the fragments, providing a visual guide to interpret the spatial domains of dispersion effects. Quantitatively, this method is an excellent complement using the already available orbital-specific approximation of the pair interaction forces. These approximations are crucial for determining the scope and magnitude of dispersion interactions at the orbital level.

4.2 Dispersion-Weighted Orbital Overlap

When it comes to orbital information, local correlation analysis has the upper hand, mainly because electrons are localized on a specific orbital. Its advantage not only lies in the reduction of computational effort compared to calculations which use canonical orbitals, local orbital subspaces are also able to isolate different physical forces at play in a non-covalently bound system. This affords chemists a deeper insight on what is happening. Using local correlation methods, a space-resolved analysis is proposed to carry out orbital decomposition of non-covalent interactions, with a specific focus on London dispersion.

4.2.1 Overlap of Dispersion Interaction Densities

The local orbital representation can effectively highlight the dispersion densities in each monomer which interacts productively, and this has been described thoroughly in Section 2.5.2. Such a depiction is closer to the London picture of dispersion. This means that most of the time, the electron clouds maintain their form while also interacting with the space separation between them. NCI index,[230] on the other hand, describes a volume with subtle changes in the density to represent a contact - closer to the Feynman interpretation of dispersion. To come closer to the latter description, an overlap of DID was devised, which we will refer to as **o-DID**.

In this work, we aim to develop an extension to current local correlation methods that provides a parallel approach to visualizing non-covalent interaction akin to NCI index. Since NCI index represents a contact between systems analyzed, to align more closely with this concept, it is necessary for o-DID to define interaction zones between monomers. This can be achieved by displaying the regions of space where the dispersion interaction density of different monomers overlap. To this end, we employ the following expression:

$$\Gamma^{\text{o-DID}}(\mathbf{r}) = \sum_{i \in A} \sum_{j \in B} \rho_i(\mathbf{r}) \rho_j(\mathbf{r}) \; e_{ij}^{\text{disp}}. \tag{4.1}$$

The equation above simply means that the orbital density contribution from monomer A ($\rho_i(\mathbf{r})$)is multiplied by the orbital density contribution from monomer B ($\rho_(\mathbf{r})$). This product is then further multiplied by the dispersion energy associated with that specific pair of orbitals (e_{ij}^{disp}). The summation of all the individual orbital contributions, as they interact via dispersion, represents the dispersion contact between the dimer. The computation of each orbital density contribution is facilitated by the following equation:

$$\rho_i(\mathbf{r}) = 2 \sum_{\mu\nu} C_{\mu i} C_{\nu i} \chi_\mu(\mathbf{r}) \chi_\nu(\mathbf{r}). \tag{4.2}$$

4.2.2 Removal of Orthogonalization Tails

Several localization approaches have been mentioned in Section 2.3.2. In the development of our methodology, we mainly employ Pipek-Mezey localization scheme.[70,71] This scheme involves a minimization procedure done on a functional correlated to the Mulliken population analysis. However, even with alternative localization procedures (e.g., Foster-Boys[65,66] and Edmiston-Ruedenberg[68,69]) centered on small subunits, the residual presence of orthogonalization tails persists. The coefficient values for these tails are very low, and these are usually beyond the localization region. In this chapter, one important objective of using LMOs is that it connects the quantum chemical results with classical chemical concepts, for example looking at specific orbitals or interacting orbitals. Unfortunately, the presence of orthogonalization tails poses difficulty in bridging these concepts. Truncation of these tails, however, can lead to errors which are not negligible ($\sim$ 1.5 eV).[231,232]

For the purpose of obtaining a meaningful picture of the dispersion density overlap, these tails must be eliminated. Several *a priori* strategies have been devised for this purpose. Notable among them are the concept of absolutely localized molecular orbitals (ALMOs)[233] and the development of extremely localized molecular orbitals (ELMOs),[234,235] both originated from the idea proposed by Stoll *et al.*[236] in 1980. On the other hand, *a posteriori* method like preliminary deorthogonalization of LMOs before truncation does not show an

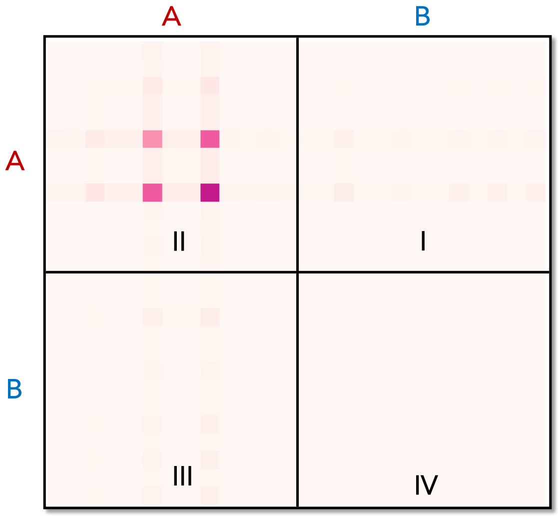


Figure 4.1: Anatomy of the single orbital density matrix, $D^i_{\mu\nu}$. Quadrants II and IV represent the MO centers exclusively for that fragment. Quadrants I and III, on the other hand, represent the part where MOs are partially present in both fragments A and B. Orthogonalization tails are present in quadrants (I, III and IV) if the fragment considered is A.

increase in localization nor lead to disappearance of tails. This is well demonstrated by Sundberg *et al.*.[237]

Where do these tails come from? When examining Equation 4.2, this can be simplified further into

$$D^i_{\mu\nu} = 2C_{\mu i}C_{\nu i}. \tag{4.3}$$

The resulting density matrix (see Figure 4.1) is partitioned into four quadrants. Quadrant I: $\mu \in A, \nu \in B$, quadrant II: $\mu, \nu \in A$, quadrant III: $\mu \in B, \nu \in A$ and quadrant IV: $\mu, \nu \in B$. Assuming MOs are centered in fragment A, there are still values of the density showing up in quadrants I, III, and IV. These are usually very low values and are a result of orthogonalization procedure, thus called orthogonalization tails. These tails are necessary for the orthonormality condition of the SCF procedure. For the purpose of our analysis, the aforementioned quadrants have to be 'cleaned-up'.

ALMOs is one of the approaches to remove these tails. This is based on the expansion of molecular orbitals (MOs) in local subsets of atomic orbitals (AOs). The only computational disadvantage is its non-orthogonality. It was first based on the work of Stoll *et al.*[236] where they generalized the SCF equations for the non-orthogonal ALMOs. Subsequently, Nagata *et al.*[238] proposed that such schemes will be referred to as locally projected self-consistent field for molecular interactions (LP SCF MI), or simply SCF MI. Khaliullin *et al.*[233] later demonstrated that SCF MI equations can replace diagonalization with a procedure that has a better scaling properties and significantly reduces computational time.

In a nutshell, nuclei and electrons of the entire system are divided into fragments. Note that these fragments are non-overlapping subsets, and each fragment must contain a specific number of electrons. This is partitioned naturally and does not rely on any cut-off threshold or bonds. Upon division, the AOs which are localized in the atoms are also partitioned into subsets $\{|\chi_{\mu i}\rangle\}$, where the first index denotes the number of subsets and the second number is the number of basis function within the given subset. The (occupied) LMOs can be labeled in a similar way as the basis functions, $\{|\phi_{\mu i}\rangle\}$ where μ now denotes the localization center and k are the different orbitals that can be attributed to the same center. This one to one correspondence naturally leads to the following approximations for the (occupied) LMOs:[236]

$$|\phi_{\gamma k}\rangle = C_{\mu i,\gamma k}\,|\chi_{\mu i}\rangle \tag{4.4}$$

where $C_{\mu i,\gamma k} = 0$ if $\gamma \neq \mu$. Under these conditions, MOs are localized on fragments in the same way AOs are localized on atoms. Dealing with the non-orthogonality of these MOs is discussed elsewhere (see References 233, 236).

The approach developed here employs the same concept of constraints. However, instead of implementing these constraints while in the SCF procedure, these are introduced once the final density matrix has been generated. These constraints can be applied to either the final coefficient matrix $C_{\mu i}$ or density matrix $D^i_{\mu\nu}$. In this work, the modification was made to the density matrix, which after the imposition of constraints, is now denoted as $\widetilde{D}^i_{\mu\nu}$:

$$\widetilde{D}^i_{\mu\nu} = \begin{cases} D^i_{\mu\nu}, & \text{if } \mu,\nu \in A \\ 0, & \text{if } \mu \in B \cup \nu \in B \end{cases} \tag{4.5}$$

This simply means that values of the new density matrix $\tilde{D}^i_{\mu\nu}$ are retained from the final $D^i_{\mu\nu}$ if AOs μ and ν belong to the fragment where they are localized. If not, these values are set to zero. An alternative way that gives the same result is if the conditions are imposed on the final $C_{\mu i}$ matrix. In this scenario, the new coefficient matrix $\tilde{C}_{\mu i}$ is defined as:

$$\tilde{C}_{\mu i} = \begin{cases} C^{i}_{\mu i}, & \text{if } \mu \in A \\ 0, & \text{if } \mu \in B. \end{cases} \tag{4.6}$$

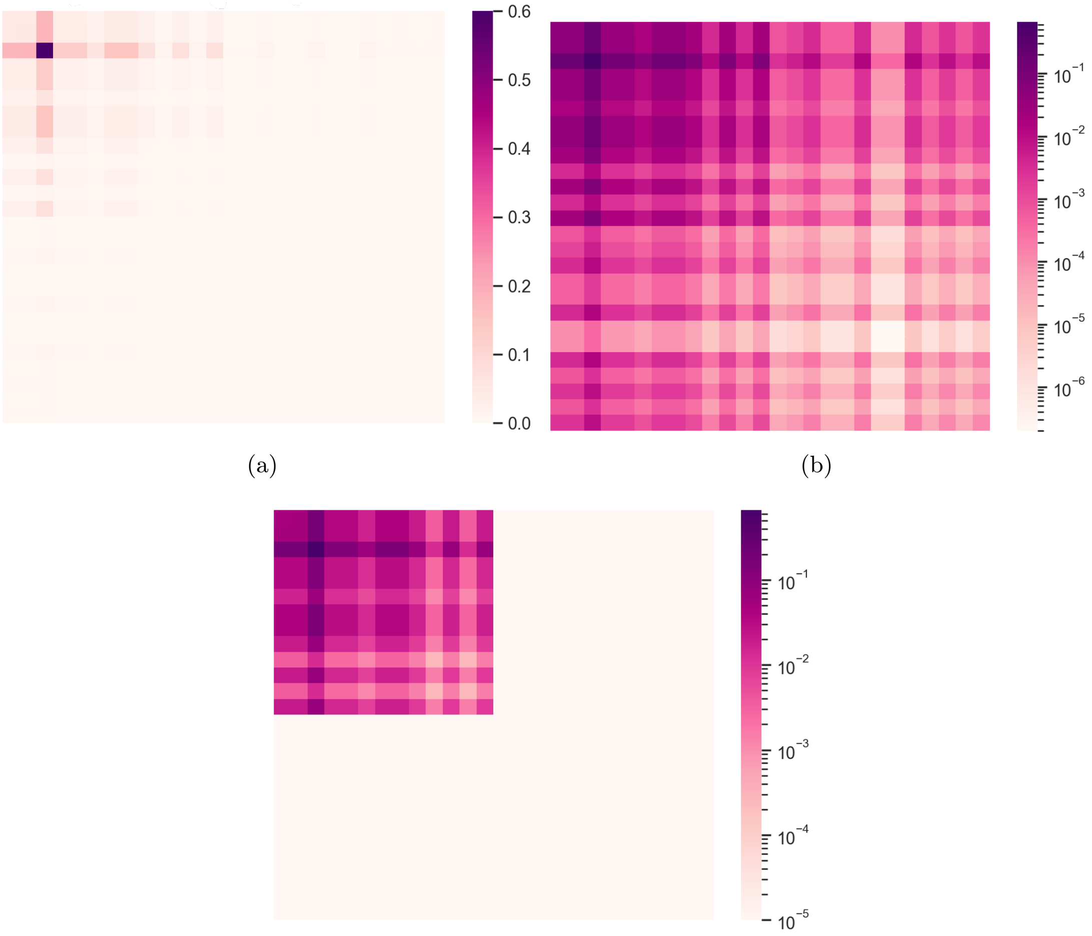
Analysis on Single Orbital Density Matrices

To investigate the impact of removing the orthogonalization tails through constraints imposed in Equation 4.5, a toy system of water dimer (non-optimized, geometry available at Appendix B) was used. Given that the matrix can become considerably large with larger basis set, for the purpose of matrix analysis, a single point calculation for the water dimer was done at the LMP2/3-21G level. All the occupied valence orbitals of water were included in the examination (total of eight orbitals).

For simple illustration, one orbital was selected and is shown in Figure 4.2. In Figure 4.2a, a linear scale is provided. Unfortunately, due to small density values, the orthogonalization tails are not readily apparent. When changing the scaling into logarithmic scale in Figure 4.2b, the tails are now more visible. However, caution should be taken when interpreting the log scale. In this case, a single change of color means 10 times more (or less) than the reference value. By defining the new density matrix $\widetilde{D}^{i}_{\mu\nu}$, the orthogonaliztion tails are effectively 'cleaned up' (see Figure 4.2c). Note that the constraints to generate the new density matrix $\widetilde{D}^{i}_{\mu\nu}$ is only applied on the last density matrix of the SCF iteration using LMOs.

o-DID in Parallel Benzene Dimer

The gravity of the orthogonalization tails problem becomes more visible in bigger systems, for example, the benzene dimer. The optimized structure of a parallel benzene dimer, as provided in Reference 240, was used and the algorithm for o-DID implemented in Molpro 2018 developer's version[136] was utilized. As illustrated in the results presented in Figure 4.3a, the dispersion density is not just an overlap between the dimer, but are also present in the each benzene unit. This complicates the interpretation that we seek to derive from o-DID. By applying the constraints as expressed in Equation 4.5, the tails are effectively eliminated, as demonstrated in Figure 4.3b. This is consistent with the desired description outlined in Equation 4.1.



(a)

(b)

(c)

Figure 4.2: Analysis of the single orbital density matrix of a water dimer at the LMP2/3-21G level of theory using the Pipek-Mezey localization scheme. A water dimer was used just to see how the density looks like before and after removing tails. Single orbital density matrix of the water dimer,$D^i_{\mu\nu}$, in a) on a linear scale and b) on a logarithmic scale. After setting constraints, $\widetilde{D}^i_{\mu\nu}$ in c) reflects the single orbital density without orthogonalization tails.

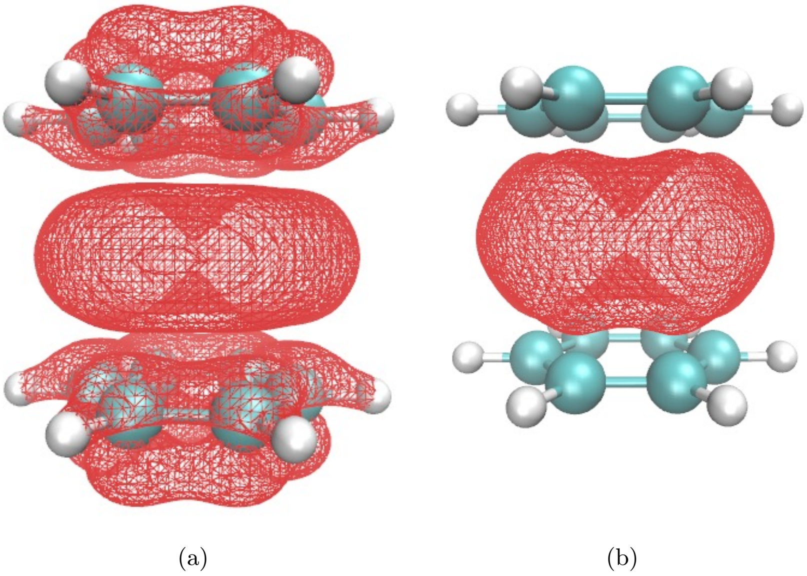


Figure 4.3: o-DID images of parallel benzene a) with and b) without orthogonalization tails. Calculations were done at the SCS-LMP2/cc-pVDZ level using a Molpro 2018 developer's version.[95] Visualized using VMD software version 1.9.4[239] with an isovalue of 0.15.

4.3 Model Systems and Computational Methods

The removal of orthogonalization tails gives a clearer view of the dispersion overlap between systems. To further explore the capabilities of o-DID, four dimer systems were employed as model systems, as shown in Figure 4.3. Two of these, argon dimer (Ar···Ar) and methane dimer ($CH_4 \cdots CH_4$), are characterized by dispersion-dominated interactions. To introduce diversity from the polar end, water dimer ($H_2O \cdots H_2O$) was utilized, known for its strong electrostatic and polarization interactions. The benzene-acetylene dimer ($C_6H_6 \cdots C_2H_2$) was selected as an intermediate system between the two interaction regimes.

Optimization of the dimer systems were done at SCS-MP2/aug-cc-pVTZ,H=cc-pVTZ using Molpro 2021.2.[54] The coordinates are available in Appendix B. The optimized structures were used as reference structures to generate the o-DID. Single point calculations at the SCS-LMP2 level, using the same basis set, were carried out using Molpro 2018 developer's version,[95] which integrated the o-DID algorithm. Absolute localization of the orbitals was achieved using the Pipek-Mezey scheme.[70,71] In the following sections, the density overlap due to dispersion is shown. This is followed by an in-depth analysis of the

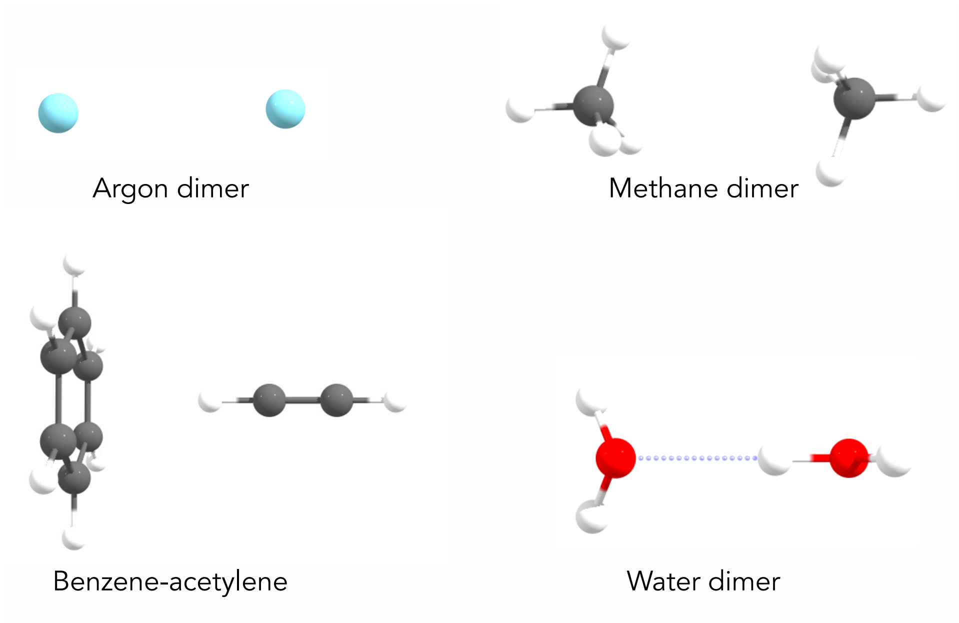


Figure 4.4: Selected model systems to evaluate the o-DID performance.

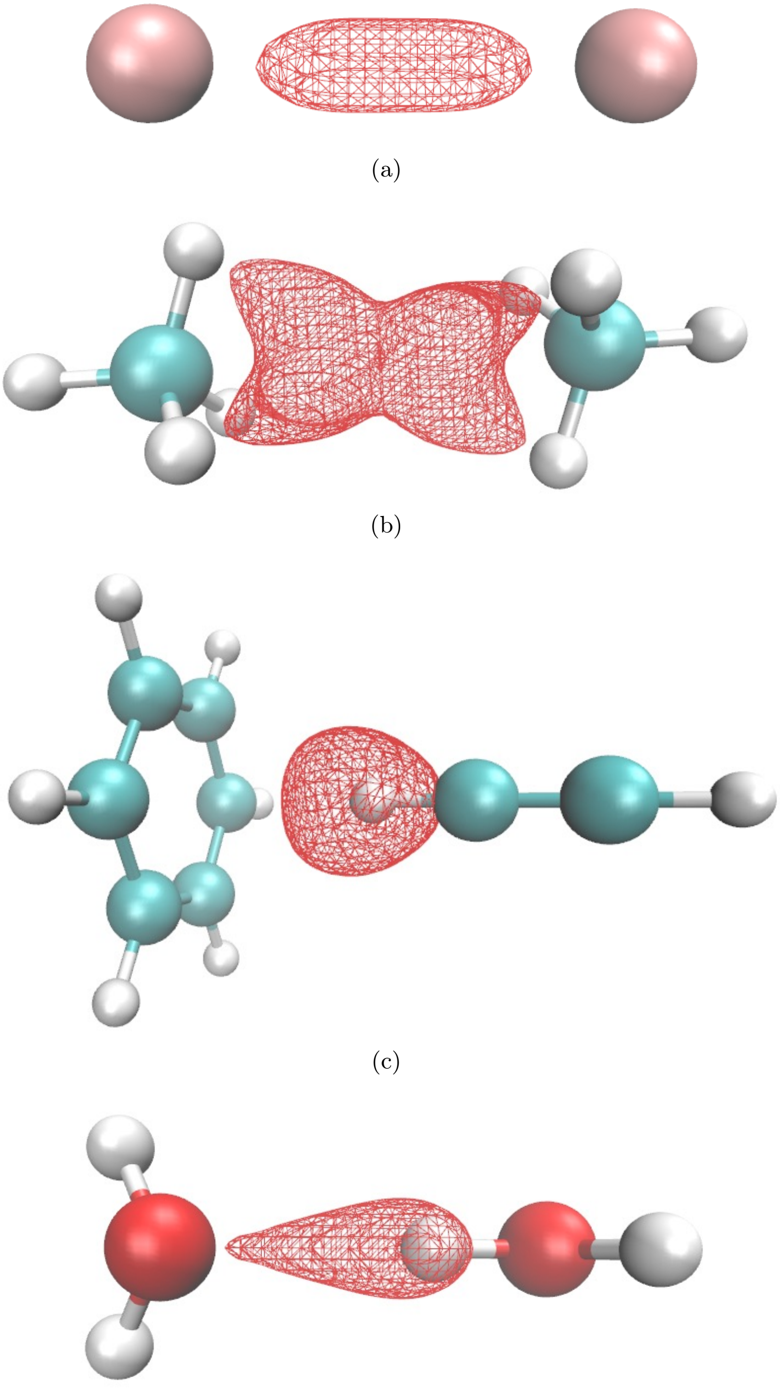
interacting orbital pairs.

4.4 Contact Surfaces

Using the toy systems (Figure 4.4) examined in this chapter, it is intriguing to explore the overlap of dispersion interaction density between the dimers/systems. In this case, the dispersion force is isolated from other forces. A brief scan of the overlapping densities in Figure 4.5 already reveals that the shape is significantly influenced by the interacting system. Additionally, it is noteworthy to observe the directionality of the contact surfaces.

Looking at the argon dimer in Figure 4.5a, the overlapping density between the two argon atoms resembles a prolate ellipsoid (with the major axis passing through both argon atoms). It is a symmetric density situated between the two atoms, indicating that the orbitals interacting from both directions contribute equally.

Another dispersion-dominated system presented here is the methane dimer, shown in Figure 4.5b. The overlapping density prominently elucidates the C−H interactions within the dimer. Since the methane geometry used here is in a *gauche* conformation, one C−H arm from one methane does not directly face the other C−H arm from the other methane, resulting in the slight decrease of density in the middle. Nevertheless, this still shows the



(a)

(b)

(c)

(d)

Figure 4.5: Total densities generated by o-DID method at the SCS-LMP2/aug-cc-pVTZ, H=cc-pVTZ. Molecular systems and their respective isovalues in parenthesis are as follows: a) argon dimer (0.2), b) methane dimer (0.11), c) benzene–acetylene (7.0) and d) water dimer (52). Visualization was done using VMD software version 1.9.4.[239]

'pyramidal' density coming from both methane, showing a uniform density contribution from all the C−H arms in the middle of two molecules. The shape of the density suggests that the majority of contribution originates from the C−H arms facing the other molecule. The trailing C−H arm does not appear to significantly influence the shape of the overall contact shape.

Moving on to the intermediate system, which lies in the spectrum between dispersion-dominated and electrostatic-dominated, we probe the benzene-acetylene dimer, as depicted in Figure 4.5c. Similar to the previously discussed systems, this T-shaped configuration retains a degree of symmetry. The density overlap due to dispersion exhibits a pear-shape distribution, with a semi-flat base near the benzene ring. This itself is a consequence of contributions from multiple orbitals originating from the benzene ring, more likely the π-orbitals. As we move towards the acetylene end, the density gets more pointed,influenced by the shape of the molecule and the proximity of the contributing orbitals. In contrast to the first two systems discussed earlier, the contact density does not center between the systems. Instead, it is skewed more towards the acetylene molecule. This observation suggests that the density contributing to the contact surface is not only due to C−H interaction with the π orbitals in the benzene ring system but also from the other orbitals in the C−C bonds of acetylene which are significantly overlapping with the π orbitals of benzene. Consequently, the total overlap is drawn closer to the acetylene.

Finally, we examine the contact density of the eletrostatic and polarization-dominated water dimer system, as shown in Figure 4.5d. The shape of the overlapping density is interesting in itself, resembling the lone pair of the oxygen atom from the water molecule on the left, engaging with the O−H orbital of the water molecule on the right. Furthermore, the density highlights that the dispersion interaction in this system is due to the O−H orbital interaction with the lone pairs of the oxygen atom of the other water molecule. The bigger blob of the density engulfing the O−H arm of the water molecule is an indication that the majority of the dispersion interaction emanates from this region of the molecule, as it interacts with several orbitals from the other water molecule.

4.5 Pair Orbital Interactions

To complement the qualitative information regarding dispersion overlap in Section 4.4, this section employs an additional method – pair orbital interaction analysis. This type of analysis offers quantitative information regarding the magnitude of the orbital interaction due to dispersion. These pair energies serve as the foundation for the DID and o-DID methods and are inherently available in any EDA calculation within the Molpro quantum chemistry package.

Recall that for each pair of orbitals, the interaction arising from dispersion is computed, and the sum of these interactions constitutes the total dispersion energy of the system under scrutiny. The information for the pair interaction due to dispersion is readily accessible, given the calculations conducted in the preceding sections were done. In this section, these values were analyzed to get a more vivid picture of what contributes to the interactions observed in the previous discussion (Section 4.4). A general note: to make the pair interactions easier to visualize, orbital densities are chosen as a representation.

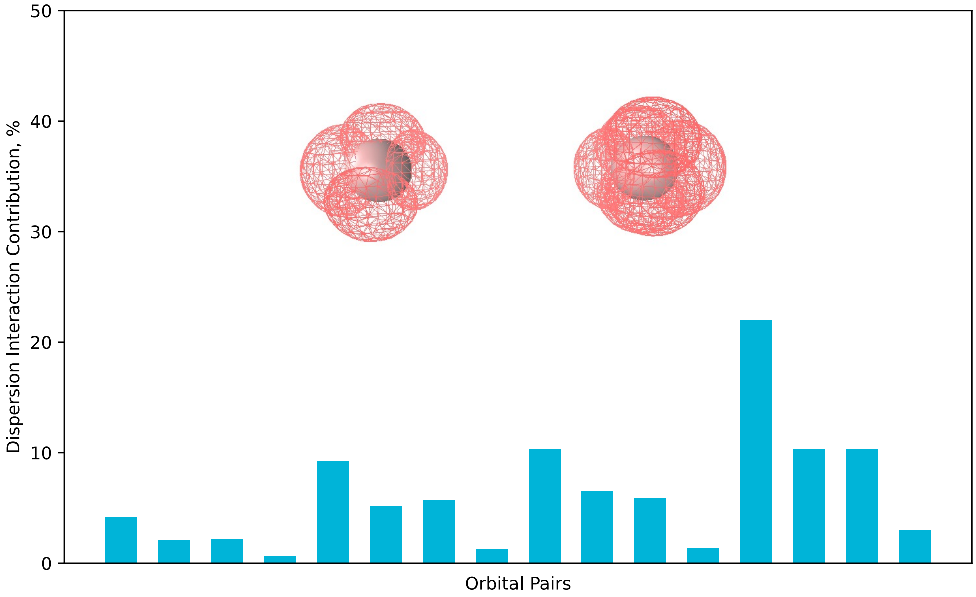


Figure 4.6: Pair orbital analysis of the Argon dimer. The inset illustrates the different active valence orbital densities of the argon atoms. Isovalue used for the orbital is 0.05. The total dispersion interaction amounts to -1.2 kJ mol^{-1}.

The uniform ellipsoidal shape of the total dispersion density in the argon dimer is the consequence of contributions from multiple orbital pairs, elucidated in Figure 4.6. The inset picture depicts different localized orbital densities of Argon atoms. These are supposed to be the *sp* hybridized orbitals which serve as the lone pairs of each argon atoms. However, note that in cases of atomic systems, the orbital assignment is arbitrary. So instead of focusing on individual orbital shapes, it is better to focus on the totality of the interactions. In this analysis, there is a distinct peak which stands out among the orbital pair contributions (around 22%). This phenomenon is due to the head-on dispersion overlap between this specific orbital pair from each argon atom. Overall, each pair contribution leads to a symmetric overlap dispersion density between these atoms. For details regarding the different orbital pair contributions, please refer to Appendix B, Table B.1.

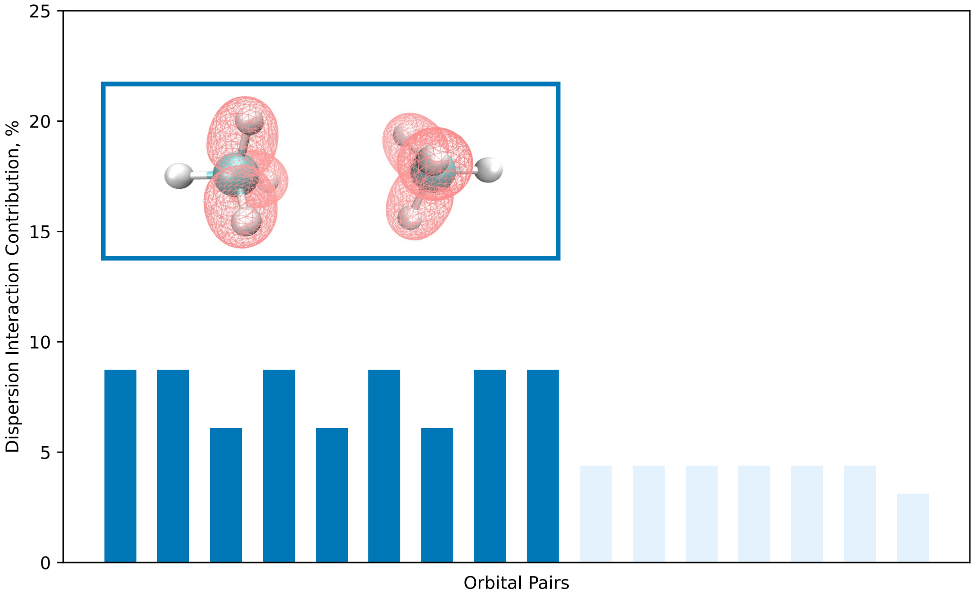


Figure 4.7: Pair orbital analysis of the methane dimer. The inset picture of the dimer system shows the different active valence orbital densities of the methane molecules. Isovalue used for the orbital is 0.07. The total dispersion interaction amounts to -2.0 kJ mol^{-1}.

The pair orbital analysis of the methane dimer in Figure 4.7 serves as an insightful illustration of how dispersion manifests as a function of distance and orientation. In this discussion, the focus is on the set of interactions responsible for o-DID pictures. In the inset of Figure 4.7, the dark blue bars represent the interactions from facing C−H orbitals between the methane dimer (refer to the picture inset). The overall $\sigma_{\mathrm{C-H}}$ interactions contribute around 71% to the total dispersion interaction density. All the high dark blue bars are due to the interaction of one C−H orbital with its adjacent C−H orbitals in the other molecule, while the shorter dark blue bars denote the interactions between one C−H orbital and the opposite C−H orbitals in the adjacent molecule. Specific values for each orbital interaction are reflected in the Appendix B, Table B.2. This substantial contribution from these orbital pairs is attributed to their close proximity, rendering them favorable for dispersion interactions. Conversely, the light blue bars on the right side of the graph represent the interactions between the trailing C−H orbital and the all other orbital of the other CH_4 molecule. Noticeably, these pairs constitute the remaining 29%

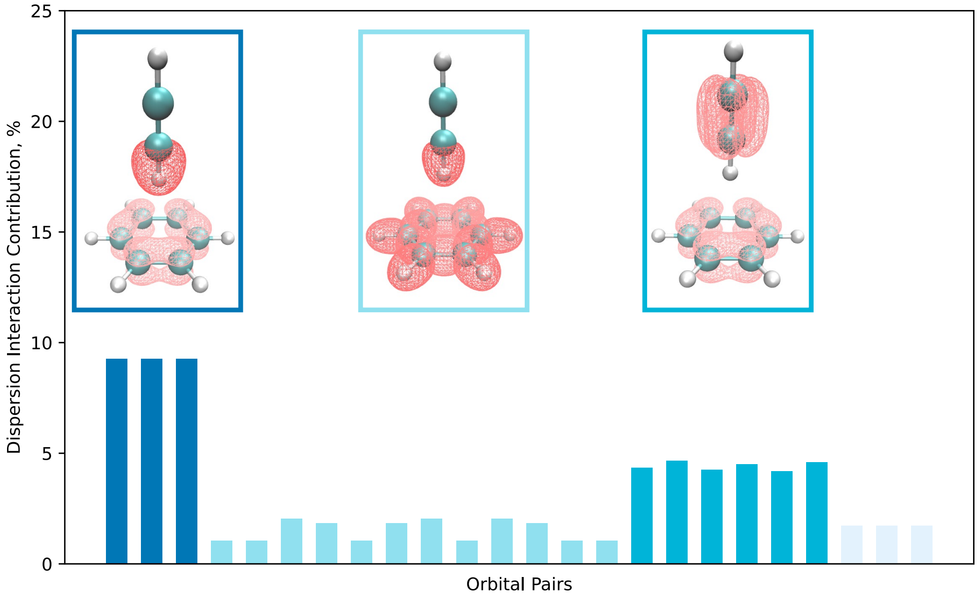


Figure 4.8: Pair orbital analysis of the benzene-acetylene system. The inset pictures show the different active valence orbital densities available for interaction. Isovalue used for the orbital is 0.04. The total dispersion interaction amounts to -8.8 kJ mol^{-1}. Pair interaction values below 1% are not included for ease of analysis and visualization.

of the dispersion contribution (approximately 3-4% each pair), with the tailing orbital pair having the least contribution. This discrepancy is visibly apparent due to the greater distance and less favorable orientation of these trailing orbitals.

The detailed orbital interaction of the intermediate system, benzene-acetylene, is shown in Figure 4.8. In a similar style as that of the methane dimer, the color gradient is employed to indicate the percentage of each orbital contribution. Simply put, the higher pair orbital contribution has a darker color, and it fades with the value of contribution. In this analysis, the set of interactions is classified according to the type of orbital interaction, as reflected by the several pictures directly above the pair orbitals representing each.

As indicated by the pair orbital analysis, the most significant contributor to the dispersion interaction occurs between the π orbitals of benzene ($\pi_{C_6H_6}$) and the sigma orbital of nearest C−H of acetylene ($\sigma_{C_2H_2}$), as shown in the dark blue bar on the leftmost part of the graph. Among these three orbital pairs, this already constitutes 28% of the dispersion contribution (with 9.3% per orbital pair). Among the orbitals of benzene, the π ring system is the most polarizable, thus easier to interact with the C−H sigma bond. It is noteworthy that, in comparison to the sigma orbitals of benzene, the π-orbitals of the ring exhibit a more favorable orientation (in terms of formation of instantaneous dipole moments) for interaction and are closer in electronic distances. All these factors collectively contribute to a stronger interaction between the orbitals.

Another significant contributor, albeit not chemically intuitive, is the π-π interactions between C_6H_6 and C_2H_2, as reflected in the blue bar on the right part of the graph. This specific interaction is intriguing because these π orbitals are already far from each other, yet the overall contribution still amounts to 27%, almost as high as the $\sigma_{C_2H_2} - \pi_{C_6H_6}$ interaction on the left. This can again be attributed to the polarizable nature of π-orbitals. These orbitals lie outside of the internuclear axis and on average, less tightly held because of its farther distance from the nuclei. Such polarizability favors the emergence of temporary dipoles resulting in significant dispersion interactions. Although each orbital pair is contributes around 4.7% each, this is a good illustration of not undermining the cumulative effect of the weaker π-π interactions. This substantial interaction explains the o-DID shape; not positioned in the middle of the system, but rather skewed toward C_2H_2.

Lastly, one might think that σ-σ interaction does not have a significant contribution to dispersion interaction because of the nature of the orbital itself - electrons are tightly bound to the nucleus compared to a π orbital. That is indeed the case, as seen in the light blue bars in the middle part of Figure 4.8. Each pair contributes around 1-2%. However, since there are several pairs responsible for this set of interaction, the total contribution is still high, i.e., 18%, to the dispersion interaction. All other orbital pairs which are below 1% were not included as these also do not represent an intuitive chemical picture. Further details on the pair orbital interactions, including those below 1% contributions, are available on Appendix B, Table B.3.

The uniform contribution from the π and σ orbitals of the benzene ring towards all other active orbitals of acetylene is the reason why the dispersion overlap shown in Figure 4.5c is somehow flat near the benzene ring. On the other hand, the slight decrease in the volume of the overlap density in the acetylene part is mainly due to the more constrained orbitals of the acetylene interacting with the orbitals of benzene.

From the polar end of the toy systems, the detailed orbital interaction of the water dimer is shown in Figure 4.9. The lone pair shaped dispersion density overlap of the molecule shown in Figure 4.5d was mainly contributed by the O$-$H orbital in the upper water molecule (see inset picture with dark blue frame) pointing towards the available lone pair of oxygen of the lower water molecule. Specifically, these interactions are between the σ-orbital and p-orbital which accounts for 24% of the dispersion interaction. This is followed by the same

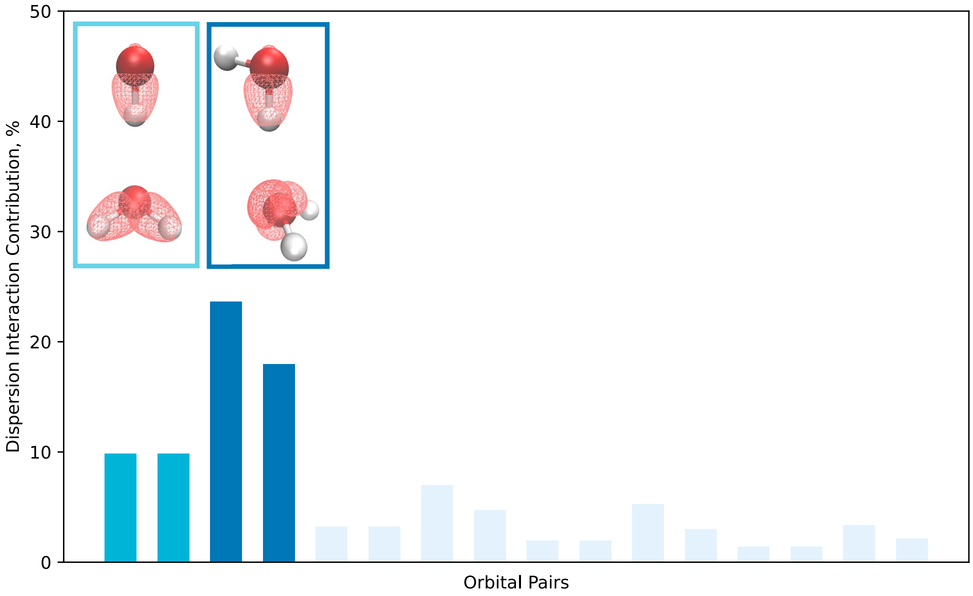


Figure 4.9: Pair orbital analysis of the water dimer. Inset pictures show the active valence orbital densities interacting due to dispersion. Isovalue used for orbitals is 0.15. The total dispersion interaction amounts to -3.5 kJ mol^{-1}.

$\sigma_{\mathrm{O-H}}$ orbital, but this time interacting with the hybrid orbital (sp orbital) of the other molecule, which contributes 18% of the interaction due to dispersion. The aforementioned interactions (dark blue bars) account for a total of 42% dispersion contribution and can be easily rationalized because of the polarizability of the lone pairs as well as their distance and orientation towards the O−H orbital. These factors have been discussed in detail above. What is also interesting is the kind of interaction which one might not put so much attention, but in the overall picture, is still a significant contributor, the $\sigma_{\mathrm{O-H}}$ orbital pair interactions between the two H_2O molecules shown in the left side of the graph in light blue bars. Normally, σ bonds have the electrons tightly held in space between the nuclei involved in a bond. But in the case of water molecules, the high electronegativity of oxygen relative to hydrogen effectively pulls the electron density towards it, resulting in an electron-rich oxygen atom which has a higher probability of forming dipole moments with the other monomer. In this specific case, this set of orbitals accounts for 20% of the total dispersion, making the overall interaction of the $\sigma_{\mathrm{O-H}}$ of the top water molecule with all the active valence orbitals of the bottom water molecule about 62%. The rest of the interacting orbitals contribute between 1 to 7%, as reflected in pale blue bars from the middle to right side of the graph. Since it is the $\sigma_{\mathrm{O-H}}$ of the top water molecule that interacts most significantly, thus the bigger blob of o-DID in this region as shown in Figure 4.5. Details of the pair orbital analysis for the water dimer is shown in Appendix B, Table B.4.

One may observe differences in the orbital shapes of each monomer in some dimers consisting of the same molecules. For instance, consider the argon dimer in Figure 4.6. Here, one argon atom (lower part) features perfectly localized valence orbitals, one *s*-orbital and three *p*-orbitals, while the other argon atom (upper part) has two *p*-orbitals and two *sp*-orbitals. A similar case can be observed in the water dimer presented Figure 4.9. For both oxygen atoms, instead of having two *p*-orbitals or two *sp*-orbitals to act as lone pairs, localization resulted in one *p*-orbital and one *sp*-orbital as lone pair per oxygen atom. These variations in orbital configurations are a direct consequence of the localization approach used, in this case, which is the Pipek-Mezey scheme. It is important to acknowledge that different localization schemes may yield slightly different orbital pictures. This can affect, to a minor extent, the pair orbital interactions. However, it is imperative to emphasize that these differences do not significantly impact the overall dispersion density overlap, and the overall dispersion interaction energy remains. Ultimately, different localization schemes can yield slightly different chemical pictures, and it is up to the user to decide which one is more appropriate to use depending on the system under investigation.

Chapter 5

Local Orbital Multireference Diagnostics

5.1 Introduction

Chapter 3 has shown the relevance of benchmarking, not only in terms of energy considerations but also in terms of its structural aspects. This evaluation is supported by several advanced spectroscopic techniques. In Chapter 4, a novel method for the visualization of dispersion interactions has been developed, elucidating them as an overlap between interacting molecular fragments. This innovative approach serves as a valuable tool for identifying the specific regions within a given system that are involved in productive dispersion interactions. Furthermore, this method also allows for the quantification of dispersion at the orbital level, thereby facilitating the identification of strongly interacting orbitals influenced by this intermolecular force. However, all these studies are based on the premise that a single-reference determinant is appropriate to represent the wavefunction describing our systems and that the dynamic correlation is the main aspect of interest. This can then be effectively captured by high-level calculations such as CC and its different variants.

Unfortunately, not all chemical systems can be described by single-reference methods. In some instances, the use of single-reference methods for systems with multireference character can lead to a discrepancy between theoretical and experimental results. This can provide an opportunity to consider the suitability of the method. Unfortunately, there are cases where theoretical predictions fit well with experiments due to error compensation, resulting in accurate numbers obtained for the wrong reasons. The fatal error starts showing up when exploring different chemical properties, e.g., relative energies, barrier heights, excitation energies, atomization energies and dissociation energies, to mention a few.

Multireference character is prevalent in many chemical systems but it is not merely defined by the molecular species under focus. A change in the geometry can significantly alter the wavefunction character of the system. A typical example is H_2 molecule. At equilibrium geometry, a single-reference wavefunction adequately captures the chemical properties of the molecule. Stretch it towards dissociation and another wavefunction is necessary to account for its ionic character, thereby requiring multiple wavefunctions to accurately calculate its energy. The complexity is magnified when dealing with multiple bonds, such as in stretched N_2. Multireference character is not limited to stretched molecules; it is also present in ground-state systems like ozone (O_3)[7,41] and the small but terrible C_2[241–243] molecule. This problem becomes more pronounced when dealing with low lying electronic states are present, as is often the case of metals. In light of these considerations, it is highly desirable to incorporate multireference diagnostic testing as a routine step before doing quantum mechanical calculations.

Recent developments in the multireference diagnostic tools have equipped chemists with the capability to assess which kind of methods are appropriate for their systems. These tools span from a range of the very affordable T_1 and D_1 diagnostics to computationally expensive and/or technically difficult diagnostics like assessing the natural orbital occupation number (NOONs) of a DMRG calculation,[244] $\%TAE_e[(T)]$[46] and the orbital-entanglement based multi-configurational diagnostic, $Z_{s(1)}$,[245] to name a few. If a routine check is desired, the affordable ones are more attractive tools because they are automatically generated upon calculations. However, the sensitivity of such tools is not guaranteed.

One interesting case on multireference character assessment pertains to non-heme iron complexes. T_1, D_1, %TAE[T] and $|t_2, max|$ all suggest a single-reference character for the systems investigated.[246] However, the NOONs of the most important MOs reveal that triplet states have single-reference character, whereas quintet states exhibit significant multireference character. This is particularly relevant in light of the discussion around their postulated two-state reactivity. In a system where experiment predicts triplet as a ground state while theoretical results suggest otherwise (i.e., quintet is a low-lying state),[247] it is necessary to achieve a proper description of the wavefunction to settle the dispute. Unfortunately, existing diagnostics are also in disagreement, despite the exhaustive computational efforts to determine which CC 'flavor' can best describe the system. Theoretical calculations still exhibit a 5 kcal mol^{-1} error when compared to experimental data available.[246]

Such instances are sufficient reasons to test the waters before embarking on expensive explorations. It may be necessary to find a better diagnostic tool or, perhaps, to discover more effective ways to utilize the existing ones. The results of both approaches will greatly assist in warning chemists to the potential high degree of multireference character within the molecular systems they investigate.

In this chapter, second excitation amplitudes of orbital invariant (OI) MP2 were utilized to examine the multireference character of small molecular systems. This approach corresponds to the application of the D_2(MP2) diagnostic developed by Nielsen and Janssen,[248] but our version utilizes the excitation amplitudes obtained from localized orbitals using the orbital-invariant version of MP2. This will be noted as **D_2(OI-MP2)**. We demonstrate that by analyzing the D_2(OI-MP2) values of every orbital, it becomes possible to identify the specific orbital(s) with a higher degree of multireference character. This information will aid users in selecting which orbitals to include if they intend to proceed to multireference treatments, such as complete active space (CAS).

5.2 The Orbital-Invariant MP2 Residual

One desirable quality of a molecular diagnostic tool is its orbital invariance. This means that the diagnostic tool should be ideally independent of the type of orbitals used in the calculations. For instance, when local methods are employed due to their affordable computational costs as well as the more chemically intuitive picture they provide to chemists, the value of the molecular diagnostic should not change compared to when canonical orbitals are used.

OI-MP2 is a reformulation of the MP2 theory (based on the Hylleraas functional) in terms of localized internal orbitals and the atomic orbitals in the virtual space.[61] Despite the use of local orbitals in the active (valence) space, the results are strictly equivalent to the canonical results under the condition that no further approximations are introduced. This type of formulation is helpful when it is desired to interpret pair correlation energies with clear chemical meaning (e.g., energy decomposition analyses). This is particularly difficult to do with canonical orbitals. There are several additional advantages of the use of local orbitals, but it is beyond the scope of this Section. Nonetheless, in this context, it is helpful to have a diagnostic which is independent of the choice of orbitals used.

The first-order perturbed wavefunction, $|\Psi^{(1)}\rangle$, consists of doubly-excited configurations which are used as a linear combination. This is written as: e

$$|\Psi^{(1)}\rangle = \frac{1}{2}\sum_{ij}\sum_{ab} T^{ij}_{ab} \, |\Phi^{ab}_{ij}\rangle \,, \tag{5.1}$$

where i, j are occupied and a, b are virtual orbitals. In this context, the occupied orbitals are only correlated valence orbitals. The expansion coefficients T^{ij}_{ab} are referred to as double excitation amplitudes. Note that:

$$T^{ij}_{ab} = T^{ji}_{ba}. \tag{5.2}$$

The doubly excited configurations, $|\Phi^{ab}_{ij}\rangle$, can be obtained by applying spin-summed excitation operators $\hat{E}^{a}_{i}$ to the reference Hartree Fock wavefunction. These spin-adapted excitation operators denote the excitation of an electron from an occupied orbital to a virtual orbital. Since we deal with doubly excited determinants, this leads to

$$|\Phi_{ij}^{ab}\rangle = \hat{E}_i^a \hat{E}_j^b |\Psi^{(0)}\rangle , \tag{5.3}$$

where $|\Psi^{(0)}\rangle$ is the reference wavefunction.

One caveat of these doubly excited configurations $|\Phi_{ij}^{ab}\rangle$ is that they are pairwise non-orthogonal for ij and ji. One solution is to use contravariant configurations and amplitudes

$$\widetilde{\Phi}_{ij}^{ab} = \frac{1}{6}(2\,|\Phi_{ij}^{ab}\rangle + |\Phi_{ij}^{ba}\rangle), \text{and} \tag{5.4}$$

$$\widetilde{T}_{ab}^{ij} = 2T_{ab}^{ij} - T_{ba}^{ij}. \tag{5.5}$$

The contravariant configurations have the following properties:

$$\begin{aligned} \langle\widetilde{\Phi}_{ij}^{ab}|\Psi^{(1)}\rangle &= T_{ab}^{ij}, \\ \langle\widetilde{\Phi}_{ij}^{ab}|\,\hat{H}\,|\Psi^{(0)}\rangle &= K_{ab}^{ij} \end{aligned} \tag{5.6}$$

Note that the Hylleraas functional, when minimized, can result to MP2 energy in the form

$$E^{(2)} = 2\,\langle\Psi^{(1)}|\,\hat{H}\,|\Psi^{(0)}\rangle + \langle\Psi^{(1)}|\,\hat{H}^{(0)} - E^{(0)}\,|\Psi^{(1)}\rangle\,. \tag{5.7}$$

When contravariant configurations are used, this results in the following second-order correlation energy:

$$\begin{aligned} E^{(2)} &= \sum_{ij}\sum_{ab}\langle\widetilde{\Phi}_{ij}^{ab}|\,\hat{H}\,|\Psi^{(0)}\rangle\,\widetilde{T}_{ab}^{ij} \\ &= \sum_{ij}\sum_{ab} K_{ab}^{ij}\widetilde{T}_{ab}^{ij} \end{aligned} \tag{5.8}$$

This then leads to the following first-order amplitude equations

$$R_{ab}^{ij} = \langle\widetilde{\Phi}_{ij}^{ab}|\,\hat{H}\,|\Psi^{(0)}\rangle + \langle\widetilde{\Phi}_{ij}^{ab}|\,\hat{H}^{(0)} - E^{(0)}\,|\Psi^{(1)}\rangle \overset{!}{=} 0. \tag{5.9}$$

The terms R^{ij}_{ab} are called residuals, which are equal to zero for the optimized amplitudes. The Hylleraas functional can now be written as

$$E^{(2)} = \sum_{ij} \sum_{ab} (K^{ij}_{ab} + R^{ij}_{ab}) \widetilde{T}^{ij}_{ab}. \tag{5.10}$$

As mentioned, the residuals R^{ij}_{ab} vanish when amplitudes are optimized, then $E^{(2)}$ becomes equal to the Equation 5.8.

The orbital-invariant form of the MP2 residuals is written as

$$R^{ij}_{ab} = K^{ij}_{ab} + \sum_{c} (f_{ac} T^{ij}_{cb} + T^{ij}_{ac} f_{cb}) - \sum_{k} (f_{ik} T^{kj}_{ab} + T^{ik}_{ab} f_{kj}), \tag{5.11}$$

where the quantities f_{rs} are matrix elements of the Fock matrix which in case of the canonical case is diagonal. If canonical orbitals are used, the residuals R^{ij}_{ab} are equal to zero, which then results in

$$T^{ij}_{ab} = \frac{K^{ij}_{ab}}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b}. \tag{5.12}$$

Recall that T^{ij}_{ab} are the MP2 amplitudes and the $\varepsilon_a, \varepsilon_b$ are the orbital energies of the virtual space while $\varepsilon_i, \varepsilon_j$ are orbital energies of the occupied space.

However, in cases where the virtual orbitals remain canonical while occupied orbitals are local (and such is the case in this chapter), the residuals can be written as

$$R^{ij}_{ab} = K^{ij}_{ab} + \sum_{c} (f_{ac} T^{ij}_{cb} + T^{ij}_{ac} f_{cb}) - \sum_{k} (f_{ik} T^{kj}_{ab} + T^{ik}_{ab} f_{kj}) + G^{ij}_{ab} + G^{ji}_{ba}, \tag{5.13}$$

where

$$G^{ij}_{ab} = -\sum_{k \neq i} f_{ik} T^{kj}_{ab}. \tag{5.14}$$

Finally, by applying the first-order perturbation theory, the amplitudes update by solving the linear amplitude equations iteratively. This can be done by computing the residuals in each iteration, as in the following equation:

$$\Delta T_{ab}^{ij} = \frac{R_{ab}^{ij}}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b}. \tag{5.15}$$

5.3 D_2 Diagnostic from OI-MP2

In Section 2.2.4, several diagnostic tools have been discussed in detail. In this section, the D_2 diagnostic is utilized. However, rather than resorting to conventional canonical orbitals, we adopt the orbital invariant MP2 discussed in Section 5.2. This formulation works by localizing the valence occupied orbital and the rest remains as canonical. The resulting amplitudes are used in the same as D_2(MP2) by Nielsen and Janssen,[248] in which the 2-norm matrix is invoked to address the system size dependence of previous diagnostics (i.e. T_1 and T_2).

There has been some skepticism regarding the use of double excitation amplitudes as a 'measure' of the quality of the wavefunction, mainly because of the use of canonical orbitals. Canonical orbitals, by nature, tend to be delocalized. In instances where molecular systems grow in size, the dominant double excitations may be 'hidden' primarily due to several double excitations with smaller coefficients. This has been a problem for M_2 diagnostics (for both MP2 and CCSD variants) which effectively lack size-intensivity.[50] Thus, a diagnostic tool using localized orbitals is an interesting alternative.

The D_2(OI-MP2) diagnostic can be simply calculated using the following expression:

$$\mathrm{D_2(OI-MP2)} = \sqrt{\lambda_{\max}(\mathbf{T}^{\mathrm{T}}\mathbf{T})} \tag{5.16}$$

where $\mathbf{T}^{\mathrm{T}}\mathbf{T}$ is composed of matrices $(\mathbf{T}^{\mathbf{o}})^{\mathrm{T}}\mathbf{T}^{\mathbf{o}}$ from the occupied space and $(\mathbf{T}^{\mathbf{v}})^{\mathrm{T}}\mathbf{T}^{\mathbf{v}}$ from the virtual space. These matrices are diagonalized separately and the largest value is chosen. The elements of the matrices are given as

$$\begin{aligned} [(\mathbf{T}^{\mathbf{o}})^{\mathrm{T}}\mathbf{T}^{\mathbf{o}}]_{i,j} &= \sum_{kab} T_{ab}^{ik} T_{ab}^{jk}, \\ [(\mathbf{T}^{\mathbf{v}})^{\mathrm{T}}\mathbf{T}^{\mathbf{v}}]_{a,b} &= \sum_{ijc} T_{ac}^{ij} T_{bc}^{ij}. \end{aligned} \tag{5.17}$$

The routine to calculate the double excitation amplitudes using OI-MP2 (see Section 5.2) and the corresponding diagnostic derived from it were implemented with Numpy interface in Psi4, an open-source quantum chemistry package.[249,250] Unless otherwise noted, a frozen-core approximation was used.

5.4 Evaluation of D_2(OI-MP2)

Unlike the D_2(MP2) developed by Nielsen and Janssen,[248] the D_2(OI-MP2) makes use of locally correlated valence orbitals. Localization was done using the Pipek-Mezey scheme.[70,71] Several examinations on its properties were conducted to demonstrate the capabilities of this new diagnostic tool.

To investigate the performance of the D_2(OI-MP2) diagnostic, we compared its calculated values to the literature values of D_2(MP2)[248] using several small molecular systems. The same geometries as those in their MP2 calculations were used, and the cc-pVTZ basis set was employed to ensure comparability with the reference values.

Table 5.1 shows the D_2 diagnostics from MP2 and OI-MP2. As shown in the table, the maximum difference among the molecular systems examined is 0.00007 from the N_2 molecule. Even on well known difficult systems like C_2 and O_3, the D_2(OI-MP2) diagnostic values match well with the reference values, with 0.00005 and 0.00006 difference for their D_2(MP2) and D_2(OI-MP2), respectively. These results show that even with the use of localized orbitals for the correlated valence electrons, the maximum double excitation amplitudes captured using canonical or localized orbitals are very similar. Thus, the threshold for D_2(MP2) will be adapted for the D_2 diagnostic values using OI-MP2.

5.5 Local Orbital Analysis Using D_2(OI-MP2)

If, indeed, one would like to use *a priori* approach to examine the extent of multireference character of a molecular system, one usually launches CCSD calculations where T1 or D1 diagnostic values are printed in the output file. These values are automatically provided when programs like Molpro or Orca are used. Note that only the ‘affordable’ and readily available diagnostic tools are being referred to here. Typically, the whole system is examined. Depending on the type of diagnostic tool used, either first excitation amplitudes or second excitation amplitudes are being picked up and mathematically processed. Depending on the predefined threshold values published in the literature is the interpretation of the possible multireference character.

In this section, a more targeted approach is presented on how to measure the multireference character of system. Unlike the conventional approach, our method operates by capturing

Table 5.1: Comparison of D_2(MP2)[248] and D_2(OI-MP2) diagnostics with the cc-pVTZ basis set.

Molecular system	D_2(MP2)	D_2(OI-MP2)
BCl	0.15038	0.15038
BF	0.14339	0.14339
C_2	0.26411	0.26406
C_2H_2	0.16498	0.16499
C_3	0.19579	0.19579
CH_2NH	0.16578	0.16577
CH_2O	0.16182	0.16180
CH_4	0.11542	0.11541
Cl_2	0.15253	0.15251
Cl_2O	0.16269	0.16269
ClOH	0.15542	0.15542
CO	0.14638	0.14632
CO_2	0.15078	0.15075
F_2	0.18164	0.18166
H_2O	0.11618	0.11619
HCl	0.12399	0.12397
HCN	0.16775	0.16778
HF	0.11003	0.11003
N_2	0.16665	0.16672
NH_3	0.11843	0.11843
O_3	0.26972	0.26978
P_2	0.19442	0.19436
SiH_4	0.11533	0.11532

the double excitation amplitudes of each valence orbital and calculating the D_2(OI-MP2) diagnostic value for each orbital. This technique not only provides a broader overview of the behaviour of each orbital in terms of excitation amplitudes but also assists in situations where one would specifically need a multireference calculation. This provides insights into which specific orbitals are more likely to be relevant to the active space.

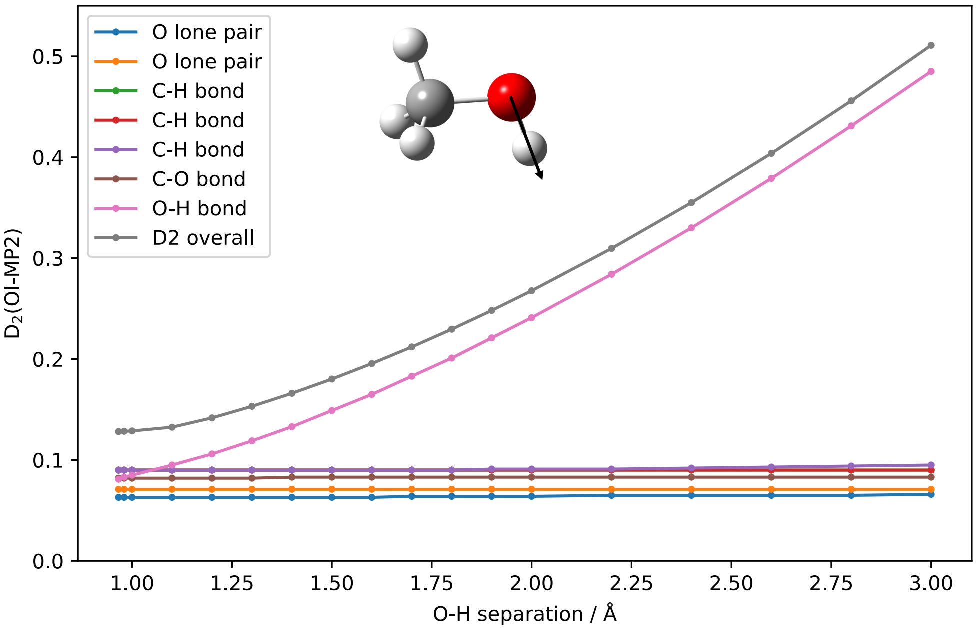
This targeted approach works by defining the centers of the active valence orbital using Mulliken population. A function for this purpose is integrated in the in-house OI-MP2 routine developed for this study. Note that since the excitation amplitude is orbital-specific, diagnostic thresholds commonly used are not necessarily applicable.

When stretching the molecule, as mentioned, it is one method to increase the static correlation of the system, hence increasing its multireference character. Such a case is observed in the process of, for example, proton transfer. In this section, we explored the behavior of two molecular systems to investigate how the multireference character of each orbital evolves during stretching towards the point where (approximately) dissociation occurs and to identify which orbitals do not present issues. Two systems are investigated, namely CH_3OH and NCOH. These are optimized at the MP2/cc-pVTZ level of theory using Gaussian 16 program.[202] Using their respective equilibrium geometries, the behaviour of the orbital-specific D_2(OI-MP2) diagnostic was evaluated as the O−H bond is stretched. The results are shown in Figure 5.1, and optimized structures can be found in Appendix C, Section C.1.

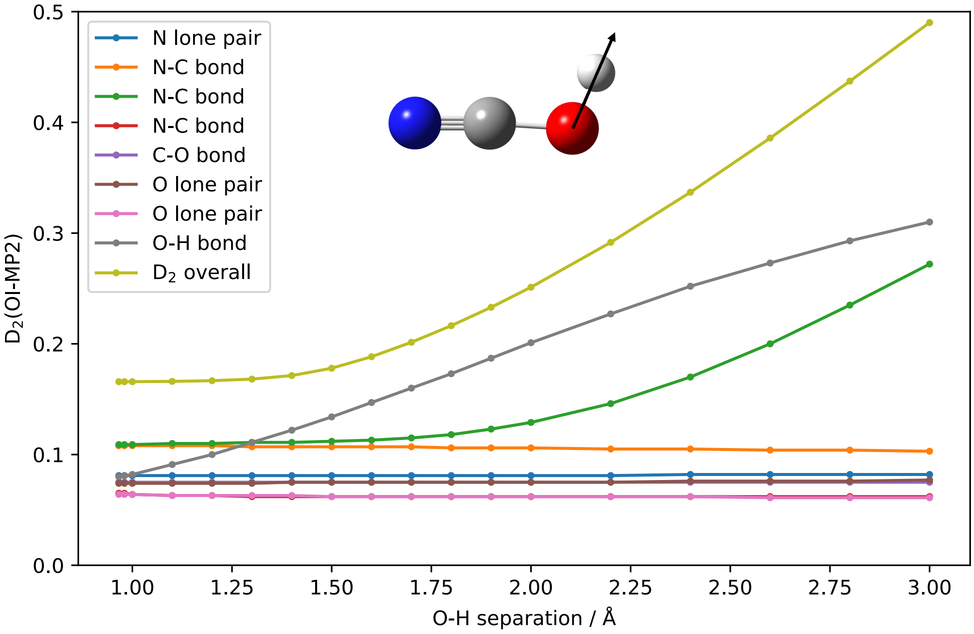
Upon analyzing the first system, CH_3OH, as shown in Figure 5.1a, all the σ-bonds (C−H and C−O) as well as the lone pair of oxygen remains unaffected while O−H is being stretched. As anticipated, the multireference character of the O−H σ-bond steadily increases as its length is extended. Intuitively, the behaviour of this specific orbital also mirrors the overall behaviour of the D_2(OI-MP2) diagnostic.

Another system examined is the NCOH molecule, as depicted in Figure 5.1b. Similar to the previous analysis, the O−H bond was stretched, while all other orbitals remained unchanged. Unquestionably, this specific bond exhibits an increase in multireference character as it elongates, as indicated by D_2(OI-MP2) orbital diagnostic. This reflects the same behaviour observed in the CH_3OH molecule. Interestingly, the O−H bond is not the only bond where an increase in static correlation is noticed. Surprisingly, one of the N−C bonds is also impacted, resulting in an increase in the static correlation, as highlighted by the D_2(OI-MP2) orbital values. What is more striking is that, unlike in CH_3OH where the O−H D_2(OI-MP2) diagnostic values reflects the overall D_2(OI-MP2) diagnostic behavior, in the case of NCOH, one of the N−C bonds takes the central role.

Based on the systems investigated above, the use of local orbital analysis to assess the degree of static correlation is a useful targeted approach to analyze the multireference character of the system. This orbital-based analysis of static correlation not only provides a more comprehensive overview of the behaviour of each valence orbital of the system in question, but can also aids in identifying orbitals which may not be obviously contributing to the multireference character of the system. The latter is particularly helpful to the chemists as it allows a logical identification of which orbitals to incorporate in the active



(a)



(b)

Figure 5.1: Orbital-specific D_2(OI-MP2) diagnostic calculation for a) CH_3OH molecule and b) NCOH molecule upon O−H stretching (indicated by the black arrow). Overall D_2(OI-MP2) of the system is also included.

space when deploying multireference calculations.

5.6 Molecular Diagnostic Inconsistencies

To ensure an accurate description of the molecular system in a particular calculation, it is advisable to look at the existing parameters that signal the presence of other significant configurations. Usually, one employs molecular diagnostics that are readily available in program packages that are computationally affordable. This, in itself, is not a problem. However, the scientific community has not reached a consensus on which diagnostic tool to use, and the choice may vary depending on the molecular system under investigation. Moreover, diagnostic thresholds are being adjusted based on the specific nature of the molecular system. For example, D_1(CCSD) signals a multireference character in main group atoms once value is >0.05. However, when systems involve 3d transition metals, Jiang *et al.*[251] suggested a value of 0.15 to indicate multireference character. Conversely, for systems involving 4d transition metals, Wang *et al.*[252] proposed a threshold of 0.12. These adjustments have not been limited to for D_1(CCSD) but also for T_1 as well as %TAE.

Before accepting or disputing the proposed adjustments, it is noteworthy to revisit to the original data set on which diagnostics based on single and double excitation amplitudes were initially developed. Historically, these diagnostics were established for small, well-behaved main group species. However, a critical question arises: Are these diagnostic tools consistently reflective of the multireference signal they give on different molecular systems? In this section, a comprehensive analysis delves into how different diagnostic tools 'diagnose' diverse molecular systems. This endeavor aims to provide a context to the reliability of the existing diagnostics and their consistency of assessing the molecular characteristics of the system.

The original data set of Janssen and Nielsen[50, 248] comprised 29-34 small closed-shell molecules examined. From their calculations, eight molecular systems were selected and closely inspected in this section. Specifically, the T_1, T_2, D_1(MP2), D_1(CCSD), D_2(MP2) and D_2(CCSD) diagnostic values were used in comparison. The bar charts follow a specific coloring scheme, akin to a traffic light. Green signals a single-reference character, typically indicating the suitability of single-reference calculations. Yellow represents intermediate diagnostic values, suggesting the need for caution in systems where other factors beyond orbital relaxation can significantly affect the quality of the reference wavefunction. Lastly, red signals a multireference character, indicating that single-reference methods are totally

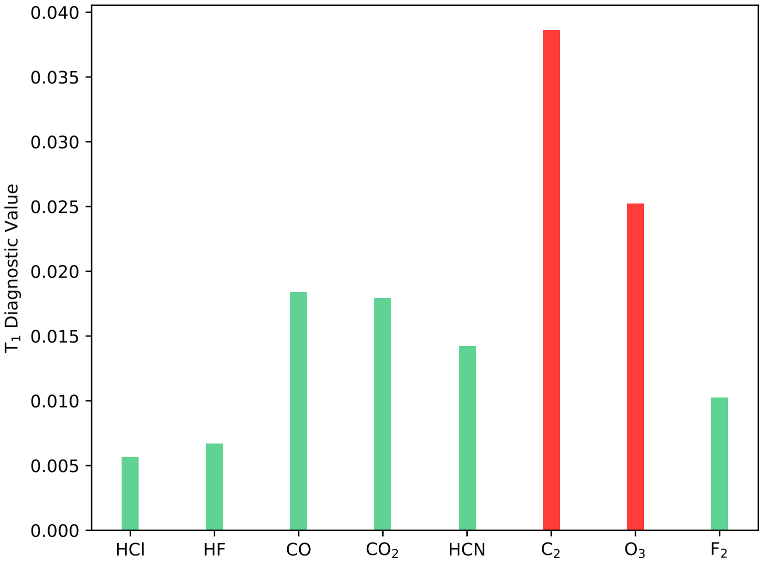
Table 5.2: Threshold values of several molecular diagnostics, where x is the calculated diagnostic value for a molecular system. Adjusted thresholds for 3d and 4d transition metals are not included in this table.

Diagnostic tool	Single-reference	Intermediate	Multireference
T_1[43,49]	≤ 0.02	–	> 0.02
T_2[49]	≤ 0.012	$0.012 < x < 0.015$	≥ 0.015
D_1(MP2)[50]	≤ 0.015	$0.015 < x < 0.040$	≥ 0.040
D_1(CCSD)[50]	≤ 0.02	$0.02 < x < 0.05$	≥ 0.05
D_2(MP2)[248]	≤ 0.15	$0.15 < x \leq 0.17$	> 0.17
D_2(CCSD)[248]	≤ 0.15	$0.15 < x \leq 0.18$	> 0.18

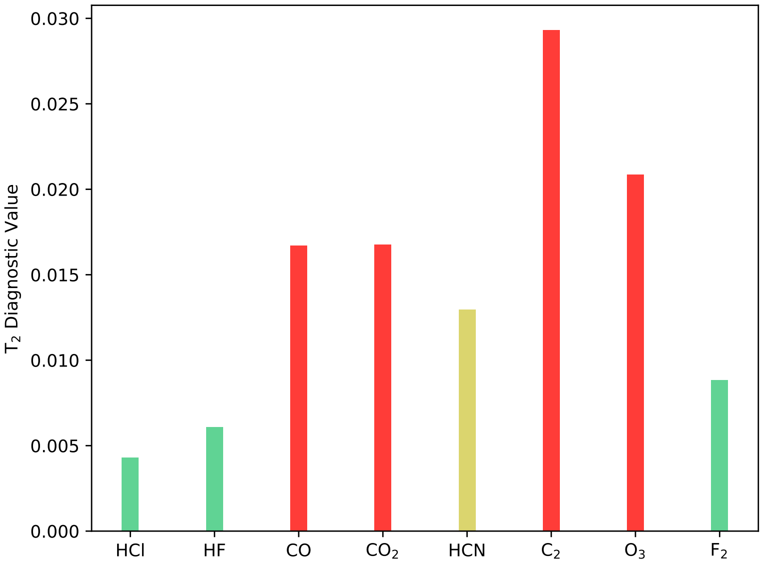
inadequate for the system.

For the selected molecules, Figure 5.2 reveals the approximations of T_1 and T_2 diagnostics (thresholds are summarized in Table 5.2). Recall from Section 2.2.4 that T_1 utilizes single-substitution amplitudes in the CCSD wavefunction, while T_2 employs single-excitation amplitudes from the MP2 wavefunction. When comparing both graphs in Figure 5.2a and b, a change in color of every bar from top (a) to bottom (b) signifies a disagreement between the diagnostic assessments. In this case, the T_2 diagnostic identifies CO and CO_2 as possessing multireference character, while HCN gives an intermediate value, suggesting that care must be taken when using single-reference methods - contradicting T_1 diagnosis. For the rest of the molecules, especially with well-known difficult systems like C_2 and O_3, both diagnostic tools yield identical predictions. It is worth noting that the only difference between the two diagnostics is the source of the single-excitation amplitude is from (Equation 2.33 for T_1 and Equation 2.35 for T_2), yet these differences manifest in their results.

Turning to D_1 diagnostics, the multireference character of the same molecules was analyzed with D_1 calculations, as reflected in Figure 5.3. This particular diagnostic tool uses a 2-norm matrix and relies on the single-excitation amplitudes of the MP2 and CCSD wavefunction, selecting a maximum value of the resulting vector. The same color scheme is applied; a change in bar color from top (a) to bottom (b) indicates a discrepancy between the diagnostics. In most cases, there is an agreement on both diagnostics, with one notable exception being CO_2. D_1(MP2) suggests a multireference character, while D_1(CCSD) implies that single-reference may still be applicable with caution. From their results, utilizing a 2-norm matrix appears to reduce the dependence of the diagnostic results on



(a)



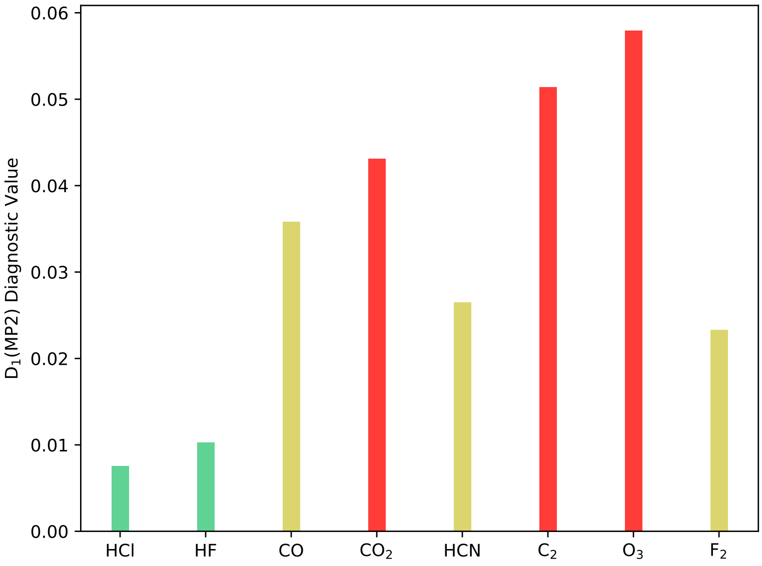
(b)

Figure 5.2: Comparison of T_1 and T_2 diagnostics.[50]

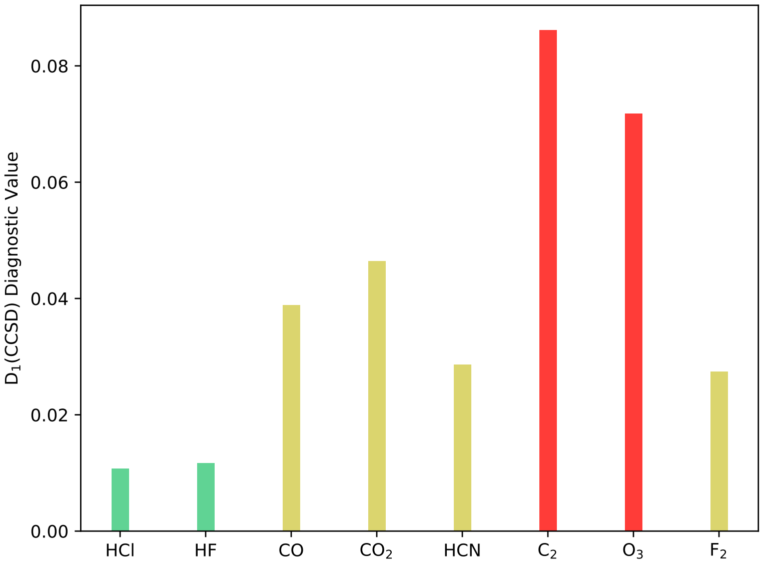
the source of single excitation amplitudes. Additionally, it is interesting to observe how the D_1 results compare to T_1 and T_2. There is an obvious disagreement between D_1 diagnostic and T_1 and T_2 diagnostics in classifying F_2 molecule, with D_1 indicating that the system in the yellow zone (intermediate value) while T_1 and T_2 place it in the green zone (single-reference). These diagnostics also differ in their assessment of whether or not CO_2 exhibits multireference character. Nevertheless, these diagnostics are consistently categorize HCl and HF as single-reference, while identifying the well known problematic systems O_3 and C_2 in the red zone (multireference).

In their subsequent paper, Nielsen and Janssen[248] also used the same data set when they developed their D_2(MP2) and D_2(CCSD) diagnostics. In this iteration, the 2-norm matrix of double-excitation amplitudes was utilized, selectively extracting the maximum value of the resulting vector (see Equation 2.38). The same set of molecules, characterized by unaltered geometries, were used as data set. The D_2 diagnostic values are shown in Figure 5.4. At first glance, one might think that finally there is more agreement in the predictions than the previous diagnostics, except of course with that of CO. Although D_2(MP2) and D_2(CCSD) agrees that single-reference methods are sufficient for most of the systems examined, D_2(CCSD) issues a warning that CO molecule might still have problems with single-reference methods (thus in the yellow zone). What is striking though is how the D_2 diagnostics (both MP2 and CCSD) classifies the F_2 molecule. While all the other diagnostics predict F_2 to be a well-behaved system where single-reference methods pose no issues, D_2(MP2) and D_2(CCSD) says otherwise. The results reflected in Figure 5.4 indicates that the choice of second excitation amplitudes, either from MP2 or CCSD wavefunction, gives the same prediction of the molecular character of the systems. However, this choice does not necessarily coincide with the results of other diagnostics, i.e., compared to T_1, T_2, D_1(MP2) and D_1(CCSD).

The results above show that, despite the existence of several affordable diagnostic tools, their diagnoses of the molecular character of some systems remain inconclusive. The different results for certain systems add to the confusion rather than offering a clear guidance on which methods can be used with confidence. Is the Euclidean norm adequate, or does the 2-norm matrix offer better results in terms of the values obtained? Is single excitation amplitude sufficient to indicate a possibility of multireference character, or is double excitation amplitude necessary? Can the arbitrary thresholds assigned be generally applied, or is there a need for further adjustments based on specific systems? The disparities in these diagnoses and the question regarding the reliability of the established thresholds leave the end users in a state of uncertainty.

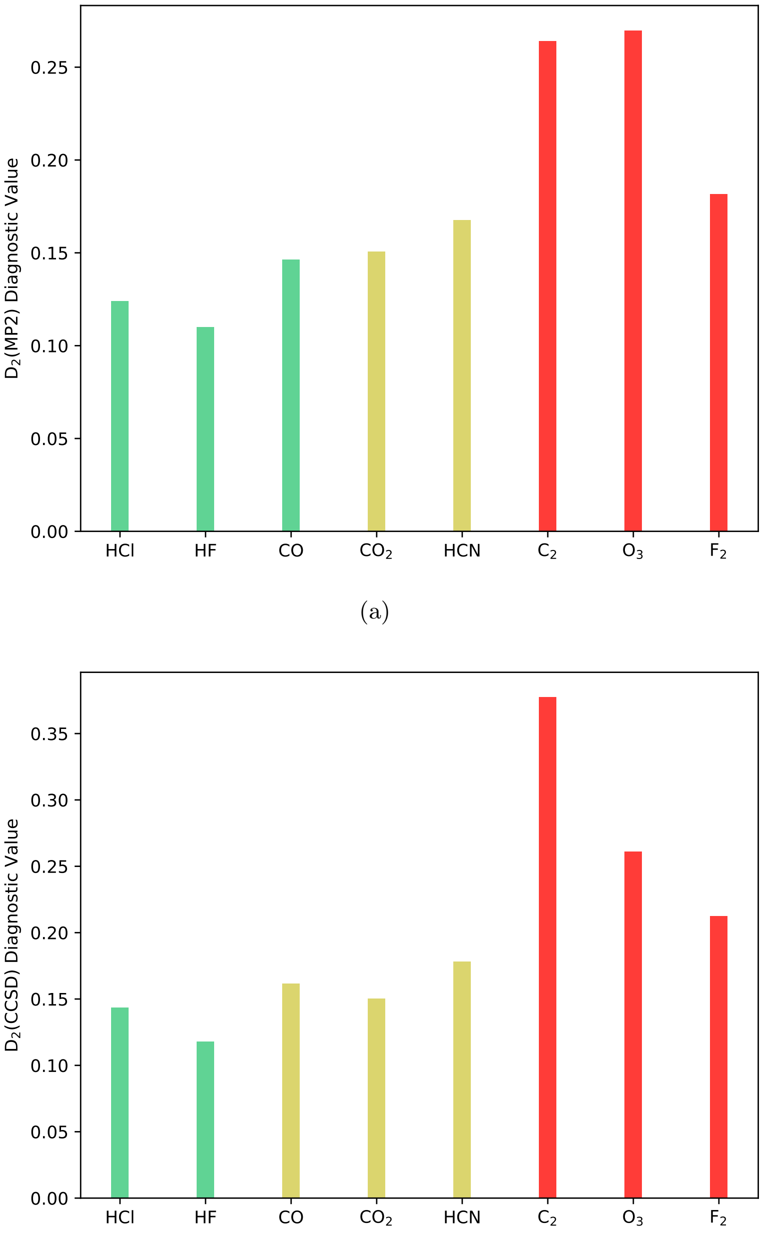


(a)



(b)

Figure 5.3: Comparison of the different D_1 diagnostics.[50, 248]



(a)

(b)

Figure 5.4: Comparison of the different D_2 diagnostics.[248]

5.7 The Case of Differential Diagnostics

One of the outstanding issues in current molecular diagnostics is the reliability of their thresholds. As mentioned, the diagnostic thresholds were established using small molecules from the main group. Specifically, the thresholds for T_1 were calibrated against the leading coefficient, C_0, of the CISD wavefunction.[43] On the other hand, the thresholds of T_2 were mainly based on the performance of T_1 on several molecules, specifically tailored to gauge the reliability of perturbation theory to approximate chemical properties.[49] D_1 (MP2 and CCSD) thresholds were established by looking at the error of equilibrium bond distances and harmonic vibrational frequencies of MP2 and CCSD methods compared to CCSD(T) values.[50] Similarly, the thresholds of D_2 (MP2 and CCSD) were determined, albeit this time using 34 molecular systems from the main group, an expansion from the 29 used for D_1 diagnostic.[248]

In 2012, Jiang *et al.*[251] introduced a comprehensive recalibration of the threshold values for T_1 and D_1 diagnostics when applied to systems containing transition metal species. Their in-depth analysis of the ccCA-TM/11all dataset[253] indicated that the historical criterion $T_1 < 0.02$ and $D_1 < 0.05$, originally designed for main group atoms, lacked practicality in the context of transition metal containing species. Through a meticulous examination involving linear fitting of T_1 and D_1 diagnostic for 225 molecules in the ccCA-TM/11all data set, they recommended a new cut-off of $T_1 < 0.05$ and $D_1 < 0.15$ for a reliable single-reference calculations. Furthermore, they proposed using the tandem of T_1, D_1 and $|\%\mathrm{TAE}| < 10$ as a combination to properly identify pathological species. The recalibration of the criteria was extended to 4d transition metals. Wang *et al.*[252] examined 118 4d transition metal-containing molecules, and established new threshold values based on the T_1, D_1 and $|\%\mathrm{TAE}|$ diagnostics. They propose that although previous adjustment for 3d transition metal are still applicable for 4d transition metals, their newly proposed threshold values are more fine-tuned for 4d metals to enhance precision in identifying the reliability of single-reference methods for these molecular species.

However, in 2017, Aoto *et al.*[254] contested the proposal of modifying the threshold of diagnostics based on their location in the periodic table. They examined the reliability of the adjusted threshold in reference to the $\Delta\mathrm{MR}$ value of 60 diatomic molecules containing transition metals. This $\Delta\mathrm{MR}$ represents the disparity between a single-reference and multireference treatments using icMRCCSD(T)/awCVTZ for an array of several molecular properties, including dissociation energy $\Delta\mathrm{MR}_{\mathrm{D_e}}$, harmonic frequency $\Delta\mathrm{MR}_{\omega_e}$, and bond distances $\Delta\mathrm{MR}_{\mathrm{R_e}}$. To consolidate these values it into a single dimensionless value, the they used the following criteria:

$$\Delta\mathrm{MR}_{\max} = \left\{ \frac{|\Delta\mathrm{MR}_{\mathrm{D_e}}|}{1\,\mathrm{kcal\,mol^{-1}}}, \frac{|\Delta\mathrm{MR}_{\mathrm{R_e}}|}{1\,\mathrm{pm}}, \frac{|\Delta\mathrm{MR}_{\omega_\mathrm{e}}|}{15\,\mathrm{cm^{-1}}} \right\} \tag{5.18}$$

where the denominators are the established benchmarks for chemical accuracy across various molecular properties. For example, if the difference of single-reference and multireference treatment for dissociation energy does not exceed 1 kcal mol^{-1}, that means single-reference method is an appropriate treatment for the system.

Their analysis revealed that the new threshold proposed by Jiang[251] is, in fact, not reliable for T_1 and D_1 diagnostics. Their results indicated many of false negatives, resulting in significant misdiagnosis of the molecular properties. This means that the diagnostic tool gives false confidence in the appropriateness of single-reference methods, whereas they actually result in a significant error compared to the established chemical accuracy of the calculated molecular properties. They concluded that the traditional criteria of $\mathrm{T}_1 \leq 0.02$ and $\mathrm{D}_1 \leq 0.05$ render these diagnostics highly reliable but significantly less efficient.

Most of the values of the multireference diagnostics are based on the behaviour of a single species when, in reality, the observables that often concern chemists are based on properties which often involve several molecules. Common benchmark values include atomization energies, dissociation energies, enthalpies of formation/reaction, interaction energies, high spin–low spin energy states, and excitation energies, among others. For example, in a typical reaction mechanism, usually it is the starting ground state geometry of molecule that is examined. Depending on the values calculated by the diagnostic for the ground state system is then the type of calculation deployed to all of the species, regardless of their electronic state. In other words, while most of the molecular properties desired to benchmark involves a differential approach (e.g., relative energies), the diagnostic measure is just often conducted on one participating species and a uniform interpretation is assumed for all other participating species. What if, all along, the approach of using the existing diagnostics is not the most appropriate?

In Figure 5.5, a subset of 36 out of 60 of Aoto's dataset[254] of transition metals diatomics was analyzed in a form of a *confusion matrix*. Specifically, the dissociation energy, as a benchmark molecular property is utilized. In this specific analysis, the common approach of using molecular diagnostics was used. For example, in a dissociation reaction $\mathrm{CuF} \longrightarrow \mathrm{Cu}^+ + \mathrm{F}^-$, the D_2 diagnostic is only used for the CuF species. The horizontal line indicates the maximum error (absolute deviation) of the calculated dissociation energy from single-

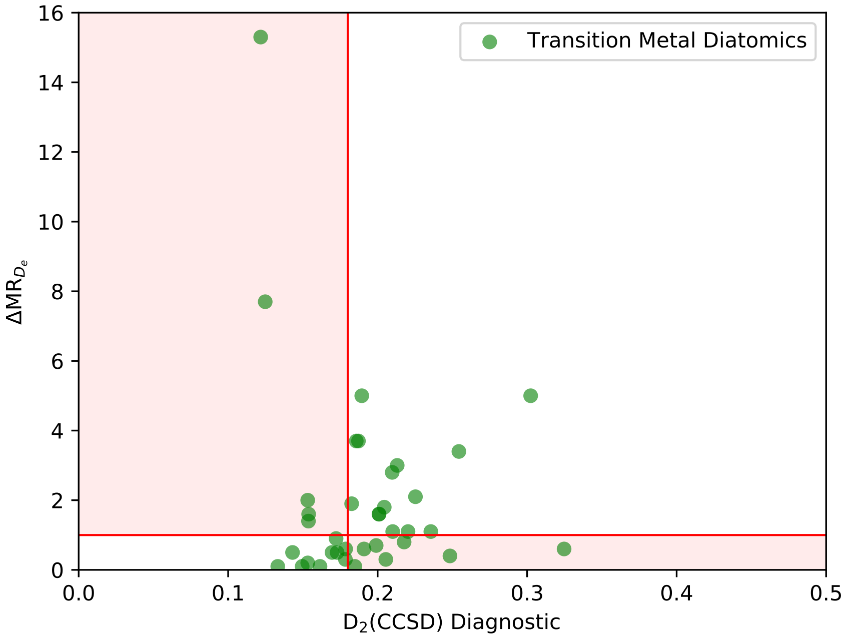


Figure 5.5: Performance of D_2(CCSD) diagnostic on selected transition metal complexes. D_2(CCSD) values were computed using Molpro 2020.2 developer's version.[224]

reference and multireference treatments, set at the chemical accuracy of 1 kcal mol^{-1}. The vertical bar, on the other hand, indicates the D_2(CCSD) threshold as established by Nielsen and Janssen[248] based on the original formulation of this diagnostic (see the threshold in Table 5.2). However, instead of accounting for any intermediate values, the threshold in this analysis is considered at 0.18, and anything exceeding that requires a multireference treatment. Points falling within the white region of the confusion matrix indicate that the diagnostic tool makes a proper prediction as to the reliability of single-reference method or if the multireference treatment is necessary based on the calculated ΔMR of the dissociation energy. Species which fall on the upper left red region contain false negative predictions, while those which fall on the lower right red region are the false positives. Among the two regions, the false negatives are the most critical results. This region means that the diagnostic tool gives the user false confidence that the systems are classified as single-reference appropriate systems, whereas in reality, the systems need multireference treatment. As shown in Figure 5.5, using the traditional approach, there are still a number of species that fall into the false positives and false-negative regions, which means that the D_2 diagnostic as well as its thresholds do not necessarily ensure the indication of the need to use multireference treatment. Worst is, it signals some species which have a very high values of $\Delta\text{MR}_{\text{D}_\text{e}}$ as safe for the use of a single-reference treatment.

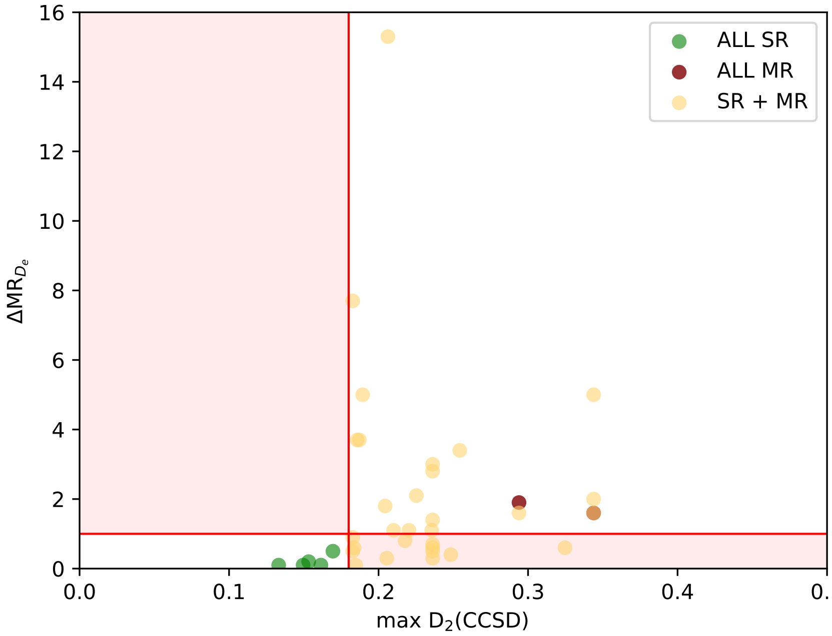


Figure 5.6: Performance of D_2(CCSD) diagnostics to selected transition metal complexes if the whole reaction is examined. Diagnostic values were computed using Molpro 2020.2 developer's version.[224]

In this section, an alternative approach to the evaluation of molecular diagnostics is proposed. Staying within the context of a dissociation reaction, e.g., $\text{CuF} \longrightarrow \text{Cu}^+ + \text{F}^-$, we dapart from the conventional practice of assessing the diagnostic value of individual species, like CuF. Instead, we analyze all the participating species are tested using D_2(CCSD). The maximum value of D_2(CCSD) among all the species serves as a parameter to decide whether or not single-reference treatment is suitable. Since only the maximum value of the diagnostic results is taken per reaction, the D_2(CCSD) thresholds still hold.

The results of this approach are shown in Figure 5.6. The same value of $\Delta\text{MR}_{\text{D}_\text{e}}$ is used, along with the D_2(CCSD) thresholds employed. To provide a clearer picture of the chemical species involved in each reaction, these are color-coded according to the diagnostic values of all the chemical species in the reaction. Green dots signify that the transition metals diatomic as well as its dissociated species are considered as well behaved system according to the D_2(CCSD) standards, indicating that single-reference methods are suitable. Yellow dots represent cases where at least one species in the reaction exhibits a D_2(CCSD) > 0.18. Lastly, red dots indicate that all species involved are characterized as having multireference character according to D_2(CCSD) standards. Upon analyzing Figure 5.6, it becomes evident that this approach effectively shifts all the species from the false negative region (upper left corner) to the white region, indicating a proper diagnosis of D_2(CCSD). Although the false positives are still present, the fatal misdiagnoses in the false-negative region have been eliminated.

While this type of analysis currently pertains to a single molecular property, it serves as a proof of concept for an alternative approach to the utilization of molecular diagnostics. Future work can explore the application of the same approach to other molecular properties such as excitation states (HS–LS) and atomization energies, to mention a few.

Chapter 6

Summary and Outlook

This thesis primarily advocates giving equal attention to both the energy and the structural features of chemical systems under investigation. For some systems, advanced spectroscopic techniques are available to probe their structure, and this can be well complemented by state-of-the-art quantum chemical methods. Nevertheless, achieving a comprehensive chemical representation remains a challenging endeavor in many cases. Local correlation methods provide not only accurate results at a modest computational cost but also bridge the gap between quantum chemical results and classical chemical concepts. Thus, we presented several tools and approaches to explore both aspects in rationalizing several chemical phenomena.

One of the main focuses of this investigation is chirality recognition discussed in Chapter 3. This research delves into various types of chiral systems, including transiently chiral molecules, permanently chiral molecules, and molecules with chirality arising from its molecular strain and the presence of axial chirality. For the first two types of chiral sytems, EDO, CHexDO and pinacol were examined, with a specific emphasis on the impact of dispersion on their chirality recognition. Specifically, the energy and structural effects due to dispersion were explored. To accomplish this, dispersion-free LMP2 gradients for the WFT part was developed, and removal of D3 corrections for the DFT part was used. The results demonstrated a significant influence on the structure of molecular aggregates when dispersion was neglected, evidently on the intermolecular HO···H bonds of the diols as well as the differences in the center of masses, ΔR(CM-CM). In terms of the differential chiral binding of these molecular systems, it was also found that dispersion is a relevant driving force. A significant decrease in the het-hom gap of the systems was observed once dispersion was eliminated. These results were more obvious with WFT results, although DFT methods still reflected these observations.

In systems with stereogenic centers (i.e., permanent chirality), the fenchone-fenchol systems were explored. These systems are pretty rigid, so there is not much deformation from its isolated structure and when it interacts in a dimer form. They are interesting because the *RR* and *SR* species are isoenergetic, so if chirality recognition is judged based on its het-hom gap, in this case it is non-existent. Although no experiment were done to conclude the chirality recognition in these systems, initial results from FTIR and microwave spectroscopy show that the stable conformers are very different in the structural aspect. To unravel the delicate balance of intermolecular forces at play, a tandem of extended KM and PNO-SCS-LMP2 energy decomposition schemes were employed. The most stable *SR*, *SR*-I, exhibits a higher electrostatic and exchange contribution than its most stable *RR* counterpart, *RR*-I. This can be the cause of OH-stretching at a lower wavenumber of *SR*-I relative to *RR*-I, thus resulting to a distinguishable peaks in the IR spectrum (exp.: 56 cm^{-1}, theory: 64 cm^{-1} after rescaling). In contrast, the *RR*-I has more dispersion contribution than *SR*-I. There is more surface of interaction of the monomers of the former, facilitating dispersion interactions. This observation is substantiated by DID visualization.

The third system examined are expanded helicenes, whose chirality is due to molecular strain and the presence of axial chirality. In this case, the inversion mechanism holds a great significance as it constitutes the key to resolving into its different enantiomers. The systems investigated were synthesized by Suárez-Pantiga *et al.*[108] Regarding the theoretical work, NEB calculations were done to explore the PES of the inversion mechanism. Several critical factors were observed to play a crucial role in the inversion barrier. One factor is the diameter of the arms, which is influenced by the manner and series of ring fusions (linear and angular modes). Expanded helicenes with longer arms shows a mechanism with higher inversion barrier. Equally important factor is the presence of substituent in the in-fjord position. When the substituent can interact with the arms, e.g., in the case of Compound **3** and **4** shown in Section 3.5, it significantly raises the inversion barrier to a significant extent, compared to the helicene lacking a substituent. Apart from steric hindrance which significantly affect the inversion barrier, both the interaction of arms with each other, as well as the interaction of the in-fjord substituent with the arms, are also affected by dispersion. The π-π interactions facilitate dispersion interactions, and the greater the available surface areas for these interactions, the more stable the reactant state, and thus the greater the inversion barrier. As for the mode of inversion itself, it strongly depends on the type of helicenes and the presence of substituents. Knowing which factors can effectively affect the inversion mechanism, these concepts can be used to further improve molecular designs of expanded helicenes.

In all the systems mentioned above, dispersion interactions play a crucial role. It is one thing to see how they affect the calculated values, it is another to discern the origin of dispersion forces. Through local correlation, DID visualization[103] was possible and it stirred us to a good direction in rationalizing the results.

Chapter 3 hopefully convinced the reader about the relevance of exploring both energies and structures when investigating a chemical phenomena. Results from the previous chapter also underscore the importance of considering dispersion forces, particularly when noncovalent interactions are involved. Chapter 4 presented a new way of looking at dispersion, where instead of focusing on its source, it is viewed from the perspective of an overlap. This chapter introduced a new approach, referred to as o-DID, which employs the overlap of dispersion densities to elucidate noncovalent interaction resulting from this force. This approach highlights the directionality of the overlap, as exemplified in Figure 4.5, serving as a valuable guide to identify regions of high dispersion density. This method is complemented by pair orbital analysis, which can pinpoint which specific orbital(s) with the strongest dispersion interaction, offering enhanced guidance for rationalizing diverse chemical phenomena.

In launching calculations, we hope to get the right numbers for the right reasons. To facilitate this, the foundation of the calculation should be right from the onset. Unfortunately, some systems are not adequately described by a single electronic configuration. There are systems, such as C_2, O_3, and metal-containing compounds, which often requires more than one relevant single electronic configuration that needs to be treated, indicating multireference character. Various diagnostic tools are available to detect the inadequacy of single-reference methods, but these diagnostics are inconsistent with each other or prove too costly and complicated for use as a preliminary test.

In Chapter 5, we employed OI-MP2 as source of double excitation amplitudes to calculate the D_2 diagnostics for several molecules. We also developed a targeted approach capable of identifying any problematic orbital that may contribute to multireference character. This strategy proves valuable in defining active spaces or deciding which orbitals to include in multireference treatment. This chapter also explored the potential of utilizing the existing diagnostics in a different manner, for example differential approach. For instance, in dissiocation reactions $AB \longrightarrow A + B$, diagnostic analysis is often performed on the molecule AB alone. We proposed to use the diagnostic in all species, taking the maximum value as a metric whether or not multireference calculations are necessary. This procedure, at least in the case of dissociation reactions, have shown to remove false negative results, i.e., a diagnosis saying there is no need for an multireference treatment where in reality, the system has to be treated with multireference calculations. However, this work is not yet complete as additional datasets for such a methodology are still necessary, and other types of currently available diagnostic tools need further examination.

Regarding multireference diagnostics using local orbitals, it remains a task for future work to establish their advantages over existing methods. The creation of a more comprehensive database is also imperative to properly establish a consistent metric for evaluating the current and future multireference diagnostic tools. This can involve the establishment of database containing information about excitation states, dissociation energies, and bond dissociation curves, to mention a few. Such a varied dataset could provide unique insights into the nature of dynamic and static correlation, a pivotal question central to the development of wave function methods.

In general, this research provides chemists with relevant tools and approaches to interpret energy results, which, in turn, are reflected in the molecular structures, specifically through the use of local correlation methods. The dispersion-free gradient LMP2 offers an effective means of eliminating dispersion, facilitating the examination of the extent of dispersion contribution in the systems under study. This thesis also demonstrates the use of extended KM and PNO-SCS-LMP2 as to how they can disentangle the delicate interplay of forces reflected in the balance of energy, particularly in the context of chiral recognition. Of course, this approach is applicable in a broad chemical problems involving the examination of intermolecular forces of attraction. The relevance of dispersion has been established in several chemical problems, thus this thesis also presented a new method to view such a force, specifically through the use of o-DID. This method provides chemists with a novel perspective of the dispersion as an overlap, which features directionality. It can be effectively complemented with pair orbital analysis to assess the extent of orbital interactions between non-interacting fragments. Lastly, this research introduces a new aspect on addressing multireference problems, one approach is via local orbital analysis to evaluate multireference character at an orbital level, and the other is using the diagnostics in a differential manner.

Bibliography

[1] J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari, and L. A. Curtiss. Gaussian-1 Theory: A General Procedure for Prediction of Molecular Energies. *J. Chem. Phys.*, 90(10):5622–5629, 1989.

[2] L. A. Curtiss, C. Jones, G. W. Trucks, K. Raghavachari, and J. A. Pople. Gaussian-1 Theory of Molecular Energies for Second-Row Compounds. *J. Chem. Phys.*, 93(4):2537–2545, 1990.

[3] L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople. Gaussian-2 Theory for Molecular Energies of First- and Second-Row Compounds. *J. Chem. Phys.*, 94(11):7221–7230, 1991.

[4] L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov, and J. A. Pople. Gaussian-3 (G3) Theory for Molecules Containing First- and Second-Row Atoms. *J. Chem. Phys.*, 109(18):7764–7776, 1998.

[5] L. A. Curtiss, P. C. Redfern, and K. Raghavachari. Gaussian-4 Theory. *J. Chem. Phys.*, 126(8):084108, 2007.

[6] S. Parthiban and J. M. L. Martin. Assessment of W1 and W2 Theories for the Computation of Electron Affinities, Ionization Potentials, Heats of Formation, and Proton Affinities. *J. Chem. Phys.*, 114(14):6014–6029, 2001.

[7] A. D. Boese, M. Oren, O. Atasoylu, J. M. L. Martin, M. Kállay, and J. Gauss. W3 theory: Robust Computational Thermochemistry in the kJ/mol Accuracy Range. *J. Chem. Phys.*, 120(9):4129–4141, 2004.

[8] A. Karton, E. Rabinovich, J. M. L. Martin, and B. Ruscic. W4 Theory for Computational Thermochemistry: In Pursuit of Confident Sub-kJ/mol Predictions. *J. Chem. Phys.*, 125(14):144108, 2006.

[9] E. Lewars. Computational Chemistry. *Introduction to the Theory and Applications of Molecular and Quantum Mechanics*, 2003.

[10] P.-O. Löwdin. Correlation Problem in Many-Electron Quantum Mechanics I. Review of Different Approaches and Discussion of Some Current Ideas. *Adv. Chem. Phys.*, pages 207–322, 1958.

[11] J. A. Pople and R. K. Nesbet. Self-Consistent Orbitals for Radicals. *J. Chem. Phys.*, 22(3):571–572, 1954.

[12] R. Seeger and J. A. Pople. Self-Consistent Molecular Orbital Methods. XVIII. Constraints and Stability in Hartree-Fock Theory. *J. Chem. Phys.*, 66(7):3045–3050, 1977.

[13] J. A. Pople and J. S. Binkley. Correlation Energies for $\mathrm{AH_n}$ Molecules and Cations. *Mol. Phys.*, 29(2):599–611, 1975.

[14] D. P. O'eill and P. M. W. Gill. Benchmark Correlation Energies for Small Molecules. *Mol. Phys.*, 103(6-8):763–766, 2005.

[15] T. Kato. On the Eigenfunctions of Many-Particle Systems in Quantum Mechanics. *Commun. Pure Appl. Math.*, 10(2):151–177, 1957.

[16] F. London. Zur theorie und systematik der molekularkräfte. *Zeitschrift für Physik*, 63(3):245–279, 1930.

[17] J. H. Van Lenthe, J. G. C. M. Duijneveldt-Van Devan Rijdt, and F. B. Van Duijneveldt. Weakly Bonded Systems. *Adv. Chem. Phys.*, 69:521–566, 1987.

[18] J. W. Hollett and P. M. W. Gill. The Two Faces of Static Correlation. *J. Chem. Phys.*, 134(11):114111, 2011.

[19] D. G. Truhlar. Valence Bond Theory for Chemical Dynamics. *J. Comput. Chem.*, 28(1):73–86, 2007.

[20] L. Piela. *Ideas of Quantum Chemistry*, volume 2, chapter 2. Elsevier, 3rd edition, 2020.

[21] C. Møller and M. S. Plesset. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.*, 46(7):618, 1934.

[22] T. Helgaker, P. Jørgensen, and J. Olsen. *Molecular Electronic-Structure Theory*, chapter 14. John Wiley & Sons, 2012.

[23] J. Olsen, O. Christiansen, H. Koch, and P. Jørgensen. Surprising Cases of Divergent Behavior in Møller-Plesset Perturbation Theory. *J. Chem. Phys.*, 105(12):5082–5090, 1996.

[24] D. Cremer. Møller-Plesset Perturbation Theory: From Small Molecule Methods to Methods for Thousands of Atoms. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 1(4):509–530, 2011.

[25] C. J. Cramer. *Essentials of Computational Chemistry: Theories and Models.* John Wiley & Sons, 2013.

[26] S. Grimme. Improved Second-Order Møller–Plesset Perturbation Theory by Separate Scaling of Parallel- and Antiparallel-Spin Pair Correlation Energies. *J. Chem. Phys.*, 118(20):9095–9102, 2003.

[27] S. Grimme. Improved Third-Order Møller-Plesset Perturbation Theory. *J. Comput. Chem.*, 24(13):1529–1537, 2003.

[28] M. Gerenkamp and S. Grimme. Spin-Component Scaled Second-Order Møller-Plesset Perturbation Theory for the Calculation of Molecular Geometries and Harmonic Vibrational Frequencies. *Chem. Phys. Lett.*, 392(1-3):229–235, 2004.

[29] T. P. M. Goumans, A. W. Ehlers, K. Lammertsma, E.-U. Würthwein, and S. Grimme. Improved Reaction and Activation Energies of [4+2] Cycloadditions, [3,3] Sigmatropic Rearrangements, and Electrocyclizations with the Spin-Component-Scaled MP2 Method. *Chem. Eur. J.*, 10(24):6468–6475, 2004.

[30] J. Čížek. On the Correlation Problem in Atomic and Molecular Systems: Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *J. Chem. Phys.*, 45(11):4256–4266, 1966.

[31] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon. A Fifth-Order Perturbation Comparison of Electron Correlation Theories. *Chem. Phys. Lett.*, 157(6):479–483, 1989.

[32] C. Hampel and H.-J. Werner. Local Treatment of Electron Correlation in Coupled Cluster Theory. *J. Chem. Phys.*, 104(16):6286–6297, 1996.

[33] M. Schütz and H.-J. Werner. Local Perturbative Triples Correction (T) with Linear Cost Scaling. *Chem. Phys. Lett.*, 318(4-5):370–378, 2000.

[34] M. Schütz. Low-Order Scaling Local Electron Correlation Methods. III. Linear Scaling Local Perturbative Triples Correction (T). *J. Chem. Phys.*, 113(22):9986–10001, 2000.

[35] M. Schütz and H.-J. Werner. Low-Order Scaling Local Electron Correlation Methods. IV. Linear Scaling Local Coupled-Cluster (LCCSD). *J. Chem. Phys.*, 114(2):661–681, 2001.

[36] M. Schütz. Low-Order Scaling Local Electron Correlation Methods. V. Connected Triples Beyond (T): Linear Scaling Local CCSDT-1b. *J. Chem. Phys.*, 116(20):8772–8785, 2002.

[37] M. Schütz. A New, Fast, Semi-Direct Implementation of Linear Scaling Local Coupled Cluster Theory. *Phys. Chem. Chem. Phys.*, 4(16):3941–3947, 2002.

[38] Q. Ma and H.-J. Werner. Explicitly Correlated Local Coupled-Cluster Methods Using Pair Natural Orbitals. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 8(6):e1371, 2018.

[39] Q. Ma and H.-J. Werner. Scalable Electron Correlation Methods. 5. Parallel Perturbative Triples Correction for Explicitly Correlated Local Coupled Cluster with Pair Natural Orbitals. *J. Chem. Theory Comput.*, 14(1):198–215, 2018.

[40] Q. Ma and H.-J. Werner. Accurate Intermolecular Interaction Energies Using Explicitly Correlated Local Coupled Cluster Methods [PNO-LCCSD(T)-F12]. *J. Chem. Theory Comput.*, 15(2):1044–1052, 2019.

[41] N. H. Morgon. Enthalpies of Formation of Phosphorus and Oxygen Compounds Determined by the Correlation Consistent Composite Approach. *Int. J. Quantum Chem.*, 112(19):3256–3260, 2012.

[42] S. R. Langhoff and E. R. Davidson. Configuration Interaction Calculations on the Nitrogen Molecule. *Int. J. Quantum Chem.*, 8(1):61–72, 1974.

[43] T. J. Lee and P. R. Taylor. A Diagnostic for Determining the Quality of Single-Reference Electron Correlation Methods. *Int. J. Quantum Chem.*, 36(S23):199–207, 1989.

[44] J. S. Sears and C. D. Sherrill. Assessing the Performance of Density Functional Theory for the Electronic Structure of Metal-Salens: The d^2-Metals. *J. Phys. Chem. A*, 112(29):6741–6752, 2008.

[45] J. S. Sears and C. D. Sherrill. Assessing the Performance of Density Functional Theory for the Electronic Structure of Metal-Salens: The $3d^0$-Metals. *J. Phys. Chem. A*, 112(15):3466–3477, 2008.

[46] A. Karton, S. Daon, and J. M. L. Martin. W4-11: A High-Confidence Benchmark Dataset for Computational Thermochemistry Derived from First-Principles W4 Data. *Chem. Phys. Lett.*, 510(4-6):165–178, 2011.

[47] T. J. Lee, J. E. Rice, G. E. Scuseria, and H. F. Schaefer. Theoretical Investigations of Molecules Composed Only of Fluorine, Oxygen, and Nitrogen: Determination of the Equilibrium Structures of FOOF, $(NO)_2$, and FNNF, and the Transition State Structure for FNNF *cis-trans* Isomerization. *Theor. Chim. Acta*, 75(2):81–98, 1989.

[48] D. Jayatilaka and T. J. Lee. Open-Shell Coupled-Cluster Theory. *J. Chem. Phys.*, 98(12):9734–9747, 1993.

[49] T. J. Lee, M. Head-Gordon, and A. P. Rendell. Investigation of a Diagnostic for Perturbation Theory: Comparison to the T_1 Diagnostic of Coupled-Cluster Theory. *Chem. Phys. Lett.*, 243(5-6):402–408, 1995.

[50] C. L. Janssen and I. M. B. Nielsen. New Diagnostics for Coupled-Cluster and Møller–Plesset Perturbation Theory. *Chem. Phys. Lett.*, 290(4-6):423–430, 1998.

[51] M. L. Leininger, I. M. B. Nielsen, T. D. Crawford, and C. L. Janssen. A New Diagnostic for Open-Shell Coupled-Cluster Theory. *Chem. Phys. Lett.*, 328(4-6):431–436, 2000.

[52] T. J. Lee. Comparison of the T_1 and D_1 Diagnostics for Electronic Structure Theory: A New Definition for the Open-Shell D_1 Diagnostic. *Chem. Phys. Lett.*, 372(3-4):362–367, 2003.

[53] J. D. Watts, M. Urban, and R. J. Bartlett. Accurate Electrical and Spectroscopic Properties of $X^1\Sigma^+$ BeO from Coupled-Cluster Methods. *Theor. Chim. Acta*, 90(5):341–355, 1995.

[54] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, et al. MOLPRO Version 2021.3: A Package of *Ab Initio* Programs, 2021. see http://www.molpro.net.

[55] Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. November 2015.

[56] M. S. Gordon. *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems.* John Wiley & Sons, 2017.

[57] O. Sinanoglu. Many-Electron Theory of Atoms, Molecules and Their Interactions. *Adv. Chem. Phys.*, 6, 1964.

[58] R. K. Nesbet. Electronic Correlation in Atoms and Molecules. *Adv. Chem. Phys.*, pages 321–363, 1965.

[59] P. Pulay. Localizability of Dynamic Electron Correlation. *Chem. Phys. Lett.*, 100(2):151–154, 1983.

[60] S. Sæbø and P. Pulay. Local Configuration Interaction: An Efficient Approach for Larger Molecules. *Chem. Phys. Lett.*, 113(1):13–18, 1985.

[61] P. Pulay and S. Saebø. Orbital-Invariant Formulation and Second-Order Gradient Evaluation in Møller-Plesset Perturbation Theory. *Theor. Chim. Acta*, 69(5):357–368, 1986.

[62] S. Sæbø and P. Pulay. Fourth-order Møller-Plessett Perturbation Theory in the Local Correlation Treatment. I. Method. *J. Chem. Phys.*, 86(2):914–922, 1987.

[63] S. Saebo and P. Pulay. The local correlation treatment. II. Implementation and tests. *J. Chem. Phys.*, 88(3):1884–1890, 1988.

[64] S. Sæbø, W. Tong, and P. Pulay. Efficient Elimination of Basis Set Superposition Errors by the Local Correlation Method: Accurate *Ab Initio* Studies of the Water Dimer. *J. Chem. Phys.*, 98(3):2170–2175, 1993.

[65] J. M. Foster and S. F. Boys. Canonical Configurational Interaction Procedure. *Rev. Mod. Phys.*, 32(2):300, 1960.

[66] S. F. Boys. Construction of Some Molecular Orbitals to be Approximately Invariant for Changes from One Molecule to Another. *Rev. Mod. Phys.*, 32(2):296, 1960.

[67] J. W. Boughton and P. Pulay. Comparison of the Boys and Pipek-Mezey Localizations in the Local Correlation Approach and Automatic Virtual Basis Selection. *J. Comput. Chem.*, 14(6):736–740, 1993.

[68] C. Edmiston and K. Ruedenberg. Localized Atomic and Molecular Orbitals. *Rev. Mod. Phys.*, 35(3):457, 1963.

[69] C. Edmiston and K. Ruedenberg. Localized Atomic and Molecular Orbitals. II. *J. Chem. Phys.*, 43(10):S97–S116, 1965.

[70] J. Pipek and P. G. Mezey. Dependence of MO Shapes on a Continuous Measure of Delocalization. *Int. J. Quantum Chem.*, 34(S22):1–13, 1988.

[71] J. Pipek and P. G. Mezey. A Fast Intrinsic Localization Procedure Applicable for *Ab Initio* and Semiempirical Linear Combination of Atomic Orbital Wave Functions. *J. Chem. Phys.*, 90(9):4916–4926, 1989.

[72] G. Knizia. Intrinsic Atomic Orbitals: An Unbiased Bridge Between Quantum Theory and Chemical Concepts. *J. Chem. Theory Comput.*, 9(11):4834–4843, 2013.

[73] W. C. Lu, C. Z. Wang, M. W. Schmidt, L. Bytautas, K. M. Ho, and K. Ruedenberg. Molecule Intrinsic Minimal Basis Sets. I. Exact Resolution of *Ab Initio* Optimized Molecular Orbitals in Terms of Deformed Atomic Minimal-Basis Orbitals. *J. Chem. Phys.*, 120(6):2629–2637, 2004.

[74] A. C. West, M. W. Schmidt, M. S. Gordon, and K. Ruedenberg. A Comprehensive Analysis of Molecule-Intrinsic Quasi-Atomic, Bonding, and Correlating Orbitals. I. Hartree-Fock Wave Functions. *J. Chem. Phys.*, 139(23):234107, 2013.

[75] T. Janowski. Near Equivalence of Intrinsic Atomic Orbitals and Quasiatomic Orbitals. *J. Chem. Theory Comput.*, 10(8):3085–3091, 2014.

[76] F. Neese, A. Hansen, and D. G. Liakos. Efficient and Accurate Approximations to the Local Coupled Cluster Singles Doubles Method Using a Truncated Pair Natural Orbital Basis. *J. Chem. Phys.*, 131(6):064103, 2009.

[77] J. Yang, G. K. L. Chan, F. R. Manby, M. Schütz, and H.-J. Werner. The Orbital-Specific-Virtual Local Coupled Cluster Singles and Doubles Method. *J. Chem. Phys.*, 136(14):144105, 2012.

[78] M. Schütz, G. Hetzer, and H.-J. Werner. Low-Order Scaling Local Electron Correlation Methods. I. Linear Scaling Local MP2. *J. Chem. Phys.*, 111(13):5691–5705, 1999.

[79] K. Kitaura and K. Morokuma. A New Energy Decomposition Scheme for Molecular Interactions Within the Hartree-Fock Approximation. *Int. J. Quantum Chem.*, 10(2):325–340, March 1976.

[80] P. Su and H. Li. Energy Decomposition Analysis of Covalent Bonds and Intermolecular Interactions. *J. Chem. Phys.*, 131(1):014102, July 2009.

[81] I. C. Hayes and A. J. Stone. An Intermolecular Perturbation Theory for the Region of Moderate Overlap. *Mol. Phys.*, 53(1):83–105, 1984.

[82] T. Ziegler and A. Rauk. Carbon Monoxide, Carbon Monosulfide, Molecular Nitrogen, Phosphorus Trifluoride, and Methyl Isocyanide as σ Donors and π Acceptors: A Theoretical Study by the Hartree-Fock-Slater Transition-State Method. *Inorganic Chemistry*, 18(7):1755–1759, 1979.

[83] E. G. Hohenstein and C. D. Sherrill. Wavefunction Methods for Noncovalent Interactions. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2(2):304–326, 2012.

[84] B. Jeziorski, R. Moszynski, and K. Szalewicz. Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes. *Chem. Rev.*, 94(7):1887–1930, 1994.

[85] W. B. Schneider, G. Bistoni, M. Sparta, M. Saitow, C. Riplinger, A. A. Auer, and F. Neese. Decomposition of Intermolecular Interaction Energies within the Local Pair Natural Orbital Coupled Cluster Framework. *J. Chem. Theory Comput.*, 12(10):4778–4792, 2016. PMID: 27564403.

[86] A. Altun, M. Saitow, F. Neese, and G. Bistoni. Local Energy Decomposition of Open-Shell Molecular Systems in the Domain-Based Local Pair Natural Orbital Coupled Cluster Framework. *J. Chem. Theory Comput.*, 15(3):1616–1632, 2019.

[87] Q. Ma, M. Schwilk, C. Köppl, and H.-J. Werner. Scalable Electron Correlation Methods. 4. Parallel Explicitly Correlated Local Coupled Cluster with Pair Natural Orbitals (PNO-LCCSD-F12). *J. Chem. Theory Comput.*, 13(10):4871–4896, 2017.

[88] P. Pinski, C. Riplinger, E. F. Valeev, and F. Neese. Sparse Maps - A Systematic Infrastructure for Reduced-Scaling Electronic Structure Methods. I. An Efficient and Simple Linear Scaling Local MP2 Method That Uses an Intermediate Basis of Pair Natural Orbitals. *J. Chem. Phys.*, 143(3):034108, 2015.

[89] F. Pavošević, C. Peng, P. Pinski, C. Riplinger, F. Neese, and E. F. Valeev. SparseMaps - A Systematic Infrastructure for Reduced Scaling Electronic Structure Methods. V. Linear Scaling Explicitly Correlated Coupled-Cluster Method with Pair Natural Orbitals. *J. Chem. Phys.*, 146(17):174108, 2017.

[90] Y. Guo, C. Riplinger, U. Becker, D. G. Liakos, Y. Minenkov, L. Cavallo, and F. Neese. Communication: An Improved Linear Scaling Perturbative Triples Correction for the Domain Based Local Pair-Natural Orbital Based Singles and Doubles Coupled Cluster Method [DLPNO-CCSD(T)]. *J. Chem. Phys.*, 148(1):011101, 2018.

[91] M. Schwilk, Q. Ma, C. Köppl, and H.-J. Werner. Scalable Electron Correlation Methods. 3. Efficient and Accurate Parallel Local Coupled Cluster with Pair Natural Orbitals (PNO-LCCSD). *J. Chem. Theory Comput.*, 13(8):3650–3675, 2017.

[92] C. Riplinger and F. Neese. An Efficient and Near Linear Scaling Pair Natural Orbital Based Local Coupled Cluster Method. *J. Chem. Phys.*, 138(3), 2013.

[93] H.-J. Werner, G. Knizia, C. Krause, M. Schwilk, and M. Dornbach. Scalable Electron Correlation Methods I.: PNO-LMP2 with Linear Scaling in the Molecular Size and Near-Inverse-Linear Scaling in the Number of Processors. *J. Chem. Theory Comput.*, 11(2):484–507, 2015.

[94] A. Wuttke. *Computational Study of Dispersion Interactions Through Local Orbital Analysis.* PhD thesis, Niedersächsische Staats-und Universitätsbibliothek Göttingen, 2019.

[95] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, et al. MOLPRO Version 2018: A Package of *Ab Initio* Programs, 2018. see http://www.molpro.net.

[96] M. Schütz, G. Rauhut, and H.-J. Werner. Local Treatment of Electron Correlation in Molecular Clusters: Structures and Stabilities of $(H_2O)_n$, $n = 2 - 4$. *J. Phys. Chem. A*, 102(29):5997–6003, 1998.

[97] D. J. Liptrot and P. P. Power. London Dispersion Forces in Sterically Crowded Inorganic and Organometallic Molecules. *Nat. Rev. Chem. Chemistry*, 1(1):1–12, 2017.

[98] C. R. Stennett, M. Bursch, J. C. Fettinger, S. Grimme, and P. P. Power. Designing a Solution-Stable Distannene: The Decisive Role of London Dispersion Effects in the Structure and Properties of {Sn (C6H2-2, 4, 6-Cy3) 2} 2 (Cy= Cyclohexyl). *J. Am. Chem. Soc.*, 143(51):21478–21483, 2021.

[99] We. Zou, J. C. Fettinger, P. Vasko, and P. P. Power. The Unusual Structural Behavior of Heteroleptic Aryl Copper (I) Thiolato Molecules: Cis vs Trans Structures and London Dispersion Effects. *Organometallics*, 41(6):794–801, 2022.

[100] J. Gramüller, M. Franta, and R. M. Gschwind. Tilting the Balance: London Dispersion Systematically Enhances Enantioselectivities in Brønsted Acid Catalyzed Transfer Hydrogenation of Imines. *J. Am. Chem. Soc.*, 2022.

[101] K. L. Mears and P. P. Power. Beyond Steric Crowding: Dispersion Energy Donor Effects in Large Hydrocarbon Ligands. *Acc. Chem. Res.*, 55(9):1337–1348, 2022.

[102] S. Riedel, P. Pyykkö, R. A. Mata, and H.-J. Werner. Comparative Calculations for the A-Frame Molecules $[S(MPH_3)_2]$ (M = Cu, Ag, Au) at Levels up to CCSD(T). *Chem. Phys. Lett.*, 405(1-3):148–152, 2005.

[103] A. Wuttke and R. A. Mata. Visualizing Dispersion Interactions Through the Use of Local Orbital Spaces. *J. Comput. Chem.*, 38(1):15–23, 2017.

[104] A. Altun, F. Neese, and G. Bistoni. Effect of Electron Correlation on Intermolecular Interactions: A Pair Natural Orbitals Coupled Cluster-Based Local Energy Decomposition Study. *J. Chem. Theory Comput.*, 15(1):215–228, 2018.

[105] K. Dzialkowski, A. Gehlhaar, C. Wölper, A. A. Auer, and S. Schulz. Structure and Reactivity of 1, 8-Bis (naphthalenediyl) dipnictanes. *Organometallics*, 38(15):2927–2942, 2019.

[106] G. Bistoni. Finding Chemical Concepts in the Hilbert Space: Coupled Cluster Analyses of Noncovalent Interactions. *WIREs Comput. Mol. Sci.*, 10(3):e1442, 2020.

[107] X. Aniban, B. Hartwig, A. Wuttke, and R.A. Mata. Dispersion Forces in Chirality Recognition – A Density Functional and Wave Function Theory Study of Diols. *Phys. Chem. Chem. Phys.*, 23(21):12093–12104, 2021.

[108] S. Suárez-Pantiga, P. Redero, X. Aniban, M. Simon, C. Golz, R.A. Mata, and M. Alcarazo. In-Fjord Substitution in Expanded Helicenes: Effects of the Insert on the Inversion Barrier and Helical Pitch. *Chem. Eur. J.*, 27(53):13358–13366, 2021.

[109] J. Gal. The Discovery of Biological Enantioselectivity: Louis Pasteur and the Fermentation of Tartaric Acid, 1857—A Review and Analysis 150 Years Later. *Chirality: The Pharmacological, Biological, and Chemical Consequences of Molecular Asymmetry*, 20(1):5–19, 2008.

[110] J. Jacques. *The Molecule and Its Double.* McGraw-Hill Companies, 1993.

[111] J.S. Siegel. Homochiral Imperative of Molecular Evolution. *Chirality: The Pharmacological, Biological, and Chemical Consequences of Molecular Asymmetry*, 10(1-2):24–27, 1998.

[112] R. Mehvar and D. R. Brocks. Stereospecific Pharmacokinetics and Pharmacodynamics of Beta-Adrenergic Blockers in Humans. 2001.

[113] P. Cintas. *Biochirality.* Springer Berlin Heidelberg, 2013.

[114] F. Carey. *Advanced Organic Chemistry.* Springer, New York, NY, 2007.

[115] E. Francotte and W. Lindner. *Chirality in Drug Research.* Wiley-VCH Verlag GmbH & Co. KGaA, sep 2006.

[116] B. Testa. Mechanisms of Chiral Recognition in Xenobiotic Metabolism and Drug-Receptor Interactions. *Chirality*, 1(1):7–9, 1989.

[117] S. W. Smith. Chiral Toxicology: It's the Same Thing... Only Different. *Toxicol. Sci.*, 110(1):4–30, 2009.

[118] W. Lenz. A Short History of Thalidomide Embryopathy. *Teratology*, 38(3):203–215, 1988.

[119] W. Heger, H. J. Schmahl, S. Klug, A. Felies, H. Nau, H. J. Merker, and D. Neubert. Embryotoxic Effects of Thalidomide Derivatives in the Non-Human Primate Callithrix jacchus. IV. Teratogenicity of µg/kg Doses of the EM12 Enantiomers. *Teratog. Carcinog. Mutagen.*, 14(3):115–122, 1994.

[120] T. Eriksson, S. Björkman, B. Roth, and P. Höglund. Intravenous Formulations of the Enantiomers of Thalidomide: Pharmacokinetic and Initial Pharmacodynamic Characterization in Man. *J. Pharm. Pharmacol.*, 52(7):807–817, 2000.

[121] A. Pietropaolo. *Chirality in Biochemistry: A Computational Approach for Investigating Biomolecule Conformations*, chapter 12, pages 291–312. John Wiley & Sons, Ltd, 2010.

[122] X. Han, C. Yuan, B. Hou, L. Liu, H. Li, Y. Liu, and Y. Cui. Chiral Covalent Organic Frameworks: Design, Synthesis, and Property. *Chem. Soc. Rev.*, 49(17):6248–6272, 2020.

[123] M. Hu, Y. X. Yuan, W. Wang, D. M. Li, H. C. Zhang, B. X. Wu, M. Liu, and Y. S. Zheng. Chiral Recognition and Enantiomer Excess Determination Based on Emission Wavelength Change of AIEgen Rotor. *Nat. Commun.*, 11(1):1–10, 2020.

[124] L. Song, H. Wang, and Y. Yang. Experimental and Theoretical Study on the Inclusion Phenomena of o-Methoxyphenol with alpha-and beta-Cyclodextrin. *Acta Chim. Sin.*, 65(16):1593, 2007.

[125] R. Breslow and S. D. Dong. Biomimetic Reactions Catalyzed by Cyclodextrins and Their Derivatives. *Chem. Rev.*, 98(5):1997–2012, 1998.

[126] S. H. Gellman. Introduction: Molecular Recognition. *Chem. Rev.*, 97(5):1231–1232, 1997.

[127] W. Y. Lee, C. H. Park, and S. Kim. A Novel Star-Shaped Crown Ether: Spontaneous Isomerization of a Macrocyclic Polyketone to a Spirobicyclic Polyketal. *J. Am. Chem. Soc.*, 115(3):1184–1185, 1993.

[128] M. Kempe and K. Mosbach. Molecular Imprinting Used for Chiral Separations. *J. Chromatogr. A*, 694(1):3–13, 1995.

[129] L. Pu. Fluorescence of Organic Molecules in Chiral Recognition. *Chem. Rev.*, 104(3):1687–1716, 2004.

[130] K. B. Lipkowitz, S. Raghothama, and J. A. Yang. Enantioselective binding of tryptophan by α-cyclodextrin. *J. Am. Chem. Soc.*, 114(5):1554–1562, 1992.

[131] N. Berova, K. Nakanishi, and R. W. Woody. *Circular Dichroism: Principles and Applications.* John Wiley & Sons, 2000.

[132] M. Rekharsky and Y. Inoue. Chiral Recognition Thermodynamics of β-Cyclodextrin: The Thermodynamic Origin of Enantioselectivity and the Enthalpy-Entropy Compensation Effect. *J. Am. Chem. Soc.*, 122(18):4418–4435, 2000.

[133] A. Zehnacker and M. A. Suhm. Chirality Recognition Between Neutral Molecules in the Gas Phase. *AAngew. Chem. Int. Ed.*, 47(37):6970–6992, 2008.

[134] T. Greber, Ž. Šljivančanin, R. Schillinger, J. Wider, and B. Hammer. Chiral Recognition of Organic Molecules by Atomic Kinks on Surfaces. *Phys. Rev. Lett.*, 96(5):056103, 2006.

[135] B. Hartwig, M. Lange, A. Poblotzki, R. Medel, A. Zehnacker, and M. A. Suhm. The Reduced Cohesion of Homoconfigurational 1,2-Diols. *Phys. Chem. Chem. Phys.*, 22:1122–1136, 2020.

[136] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, et al. MOLPRO Version 2018.1: A Package of *Ab Initio* Programs, 2018. see http://www.molpro.net.

[137] S. J. Chalk. The IUPAC Gold Book website. 2019.

[138] J. Echeverría, G. Aullón, D. Danovich, S. Shaik, and S. Alvarez. Dihydrogen Contacts in Alkanes: Subtle but Not Faint. *Nat. Chem.*, 3(4):323–330, 2011.

[139] P. R. Schreiner, L. V. Chernish, P. A. Gunchenko, E. Y. Tikhonchuk, H. Hausmann, M. Serafin, S. Schlecht, Jeremy E. P. Dahl, R. M.K. Carlson, and A. A. Fokin. Overcoming Lability of Extremely Long Alkane Carbon-Carbon Bonds Through Dispersion Forces. *Nature*, 477(7364):308–311, 2011.

[140] M. S. G. Ahlquist and P. O. Norrby. Dispersion and Back-Donation Gives Tetracoordinate [Pd(PPh_3)$_4$]. *Angew. Chem. - Int. Ed.*, 50(49):11794–11797, 2011.

[141] I. Fenández López, M. Solà i Puig, and F. M. Bickelhaupt. Why Do Cycloaddition Reactions Involving C_{60} Prefer [6,6] Over [5,6] Bonds? *Chem. Eur. J.*, 19(23):7416–7422, 2013.

[142] E. Lyngvi, I. A Sanhueza, and F. Schoenebeck. Dispersion Makes the Difference: Bisligated Transition States Found for the Oxidative Addition of $Pd(PtBu_3)_2$ to $Ar{-}OSO_2R$ and Dispersion-Controlled Chemoselectivity in Reactions with $Pd[P(iPr)(tBu_2)]_2$. *Organometallics*, 34(5):805–812, 2015.

[143] T. H. Meyer, W. Liu, M. Feldt, A. Wuttke, R. A. Mata, and L. Ackermann. Manganese(I)-Catalyzed Dispersion-Enabled C−H/C−C Activation. *Chem. Eur. J.*, 23(23):5443–5447, 2017.

[144] E. Detmar, V. Müller, D. Zell, L. Ackermann, and M. Breugst. Cobalt-Catalyzed C−H Cyanations: Insights into the Reaction Mechanism and the Role of London Dispersion. *Beilstein J. Org. Chem.*, 14(1):1537–1545, 2018.

[145] D. Yepes, F. Neese, B. List, and G. Bistoni. Unveiling the Delicate Balance of Steric and Dispersion Interactions in Organocatalysis Using High-Level Computational Methods. *J. Am. Chem. Soc.*, 142(7):3613–3625, 2020. PMID: 31984734.

[146] C. Eschmann, L. Song, and P. R. Schreiner. London Dispersion Interactions Rather than Steric Hindrance Determine the Enantioselectivity of the Corey–Bakshi–Shibata Reduction. *Angew. Chem. - Int. Ed.*, 60(9):4823–4832, 2021.

[147] F. J. Winkler, R. Medina, J. Winkler, and H. Krause. Mass Spectral and Semi-Empirical Studies on Chiral Discrimination in Gaseous Aggregation Products of Protonated Dialkyl Tartrates. *J. Chromatogr. A*, 666(1):549 – 556, 1994. Chiral Separations Fundamental Aspects and Applications.

[148] N. Seurre, K. Le Barbu-Debus, F. Lahmani, A. Zehnacker, N. Borho, and M. A. Suhm. Chiral Recognition Between Lactic Acid Derivatives and an Aromatic Alcohol in a

Supersonic Expansion: Electronic and Vibrational Spectroscopy. *Phys. Chem. Chem. Phys.*, 8:1007–1016, 2006.

[149] A. Mahjoub, A. Chakraborty, V. Lepere, K. Le Barbu-Debus, N. Guchhait, and A. Zehnacker. Chirality-Dependent Hydrogen Bond Direction in Jet-Cooled (S)-1,2,3,4-Tetrahydro-3-isoquinoline Methanol (THIQM): IR-Ion Dip Vibrational Spectroscopy of the Neutral and the Ion. *Phys. Chem. Chem. Phys.*, 11:5160–5169, 2009.

[150] D. Scuderi, K. Le Barbu-Debus, and A. Zehnacker. The Role of Weak Hydrogen Bonds in Chiral Recognition. *Phys. Chem. Chem. Phys.*, 13:17916–17929, 2011.

[151] J. Altnöder, A. Bouchet, J. J. Lee, K. E. Otto, M. A. Suhm, and A. Zehnacker-Rentien. Chirality-Dependent Balance Between Hydrogen Bonding and London Dispersion in Isolated (±)-1-Indanol Clusters. *Phys. Chem. Chem. Phys.*, 15:10167–10180, 2013.

[152] K. Schwing and M. Gerhards. Investigations on Isolated Peptides by Combined IR/UV Spectroscopy in a Molecular Beam – Structure, Aggregation, Solvation, and Molecular Recognition. *Int. Rev. Phys. Chem.*, 35(4):569–677, 2016.

[153] F. Kollipost, K. E. Otto, and M. A. Suhm. A Symmetric Recognition Motif between Vicinal Diols: The Fourfold Grip in Ethylene Glycol Dimer. *Angew. Chem. - Int. Ed.*, 55(14):4591–4595, 2016.

[154] J. Klyne, A. Bouchet, S. Ishiuchi, M. Fujii, M. Schneider, C. Baldauf, and O. Dopfer. Probing Chirality Recognition of Protonated Glutamic Acid Dimers by Gas-Phase Vibrational Spectroscopy and First-Principles Simulations. *Phys. Chem. Chem. Phys.*, 20:28452–28464, 2018.

[155] F. Xie, M. Fusè, A. S. Hazrah, W. Jäger, V. Barone, and Y. Xu. Discovering the Elusive Global Minimum in a Ternary Chiral Cluster: Rotational Spectra of Propylene Oxide Trimer. *Angew. Chem. - Int. Ed.*, 59(50):22427–22430, 2020.

[156] K. Le Barbu-Debus, D. Scuderi, V. Lepère, and A. Zehnacker. Homochiral vs. Heterochiral Sodium Core Dimers of Tartaric Acid Esters: A Mass Spectrometry and Vibrational Spectroscopy Study. *J. Mol. Struct.*, 1205:127583, 2020.

[157] N. O. B. Lüttschwager, T. N. Wassermann, R.A. Mata, and M. A. Suhm. The Last Globally Stable Extended Alkane. *Angew. Chem. - Int. Ed.*, 52(1):463–466, 2013.

[158] K. B. Lipkowitz, B. Coner, M. A. Peterson, and A. Morreale. Enantioselective Binding in Gas Chromatography: A Computational Study of Chiral Selection by Permethyl-β-Cyclodextrin. *J. Phys. Org. Chem.*, 10(5):311–322, 1997.

[159] K. B. Lipkowitz, R. Coner, M. A. Peterson, A. Morreale, and J. Shackelford. The Principle of Maximum Chiral Discrimination: Chiral Recognition in Permethyl-β-cyclodextrin. *J. Org. Chem.*, 63(3):732–745, 1998. PMID: 11672068.

[160] V. Schurig. Separation of Enantiomers by Gas Chromatography. *J. Chromatogr. A*, 906(1):275 – 299, 2001. Chiral Separations.

[161] C. Cagliero, B. Sgorbini, C. Cordero, E. Liberto, P. Rubiolo, and C. Bicchi. Enantioselective Gas Chromatography with Derivatized Cyclodextrins in the Flavour and Fragrance Field. *Isr. J. Chem.*, 56(11-12):925–939, 2016.

[162] A. Kumar, E. Capua, M. K. Kesharwani, J. M. L. Martin, E. Sitbon, D. H. Waldeck, and R. Naaman. Chirality-Induced Spin Polarization Places Symmetry Constraints on Biomolecular Interactions. *PNAS*, 114(10):2474–2478, 2017.

[163] A. D. Fortes and E. Suard. Crystal Structures of Ethylene Glycol and Ethylene Glycol Monohydrate. *J. Chem. Phys.*, 135(23):234501, 2011.

[164] J. P. Wagner and P. R. Schreiner. London Dispersion in Molecular Chemistry—Reconsidering Steric Effects. *Angew. Chem. - Int. Ed.*, 54(42):12274–12296, 2015.

[165] S. Kristyán and P. Pulay. Can (Semi) Local Density Functional Theory Account for the London Dispersion Forces? *Chem. Phys. Lett.*, 229(3):175–180, 1994.

[166] J. Pérez-Jordá and A. D. Becke. A Density-Functional Study of van der Waals Forces: Rare Gas Diatomics. *Chem. Phys. Lett.*, 233(1-2):134–137, 1995.

[167] P. Hobza, J. Šponer, and T. Reschel. Density Functional Theory and Molecular Clusters. *J. Comput. Chem.*, 16(11):1315–1325, 1995.

[168] H. Kruse, L. Goerigk, and S. Grimme. Why the Standard B3LYP/6-31G* Model Chemistry Should Not Be Used in DFT Calculations of Molecular Thermochemistry: Understanding and Correcting the Problem. *J. Org. Chem.*, 77(23):10824–10834, 2012.

[169] S. Grimme. Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.*, 27(15):1787–1799, 2006.

[170] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg. A Consistent and Accurate *Ab Initio* Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.*, 132(15):154104, 2010.

[171] E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth, and S. Grimme. A Generally Applicable Atomic-Charge Dependent London Dispersion Correction. *J. Chem. Phys.*, 150(15):154122, 2019.

[172] A. Wuttke, M. Feldt, and R. A. Mata. All That Binds Is Not Gold—The Relative Weight of Aurophilic Interactions in Complex Formation. *J. Phys. Chem. A*, 122(34):6918–6925, 2018. PMID: 30088931.

[173] S. Grimme. Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *J. Chem. Theory Comput.*, 15(5):2847–2862, 2019.

[174] P. Pracht, F. Bohle, and S. Grimme. Automated Exploration of the Low-Energy Chemical Space with Fast Quantum Chemical Methods. *Phys. Chem. Chem. Phys.*, 22:7169–7192, 2020.

[175] S. Grimme, C. Bannwarth, and P. Shushkov. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1–86). *J. Chem. Theory Comput.*, 13(5):1989–2009, 2017.

[176] C. Bannwarth, S. Ehlert, and S. Grimme. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.*, 15(3):1652–1671, 2019.

[177] J. G. Brandenburg, C. Bannwarth, A. Hansen, and S. Grimme. B97-3c: A Revised Low-Cost Variant of the B97-D Density Functional Method. *J. Chem. Phys.*, 148(6):064104, 2018.

[178] F. Weigend and R. Ahlrichs. 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(18):3297, 2005.

[179] J. Zheng, X. Xu, and D. G. Truhlar. Minimally Augmented Karlsruhe Basis Sets. *Theor. Chem. Acc.*, 128(3):295–305, 2011.

[180] A. D. Becke. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A*, 38:3098–3100, 1988.

[181] J. P. Perdew. Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas. *Phys. Rev. B*, 33:8822–8824, Jun 1986.

[182] J. P. Perdew. Erratum: Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas. *Phys. Rev. B*, 34:7406–7406, Nov 1986.

[183] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.*, 77:3865–3868, Oct 1996.

[184] C. Adamo and V. Barone. Toward Reliable Density Functional Methods Without Adjustable Parameters: The PBE0 Model. *J. Chem. Phys.*, 110(13):6158–6170, 1999.

[185] M. Ernzerhof and G. E. Scuseria. Assessment of the Perdew–Burke–Ernzerhof Exchange-Correlation Functional. *J. Chem. Phys.*, 110(11):5029–5036, 1999.

[186] C. Lee, W. Yang, and R. G. Parr. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B*, 37:785–789, 1988.

[187] B. Miehlich, A. Savin, H. Stoll, and H. Preuss. Results Obtained with the Correlation Energy Density Functionals of Becke and Lee, Yang, and Parr. *Chem. Phys. Lett.*, 157(3):200 – 206, 1989.

[188] F. Weigend. Accurate Coulomb-Fitting Basis Sets for H to Rn. *Phys. Chem. Chem. Phys.*, 8(9):1057, 2006.

[189] F. Neese. The ORCA Program System. *WIREs Comput. Mol. Sci.*, 2(1):73–78, 2012.

[190] F. Neese. Software Update: The ORCA Program System, Version 4.0. *WIREs Comput. Mol. Sci.*, 8(1):e1327, 2017.

[191] S. Grimme, S. Ehrlich, and L. Goerigk. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.*, 32(7):1456–1465, 2011.

[192] S. Sæbø and P. Pulay. Local Treatment of Electron Correlation. *Annu. Rev. Phys. Chem.*, 44(1):213–236, 1993.

[193] R. A. Kendall, T. H. Dunning, and R. J. Harrison. Revisiting Electron Affinities of First-Row Atoms: Systematic Basis Sets and Wave Functions. *J. Chem. Phys.*, 96(9):6796–6806, 1992.

[194] T. H. Dunning. Gaussian Basis Sets for Use in Correlated Molecular Calculations: I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.*, 90(2):1007–1023, 1989.

[195] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, et al. MOLPRO Version 2019.2: A Package of *Ab Initio* Programs, 2019. see http://www.molpro.net.

[196] K. A. Peterson, T. B. Adler, and H.-J. Werner. Systematically Convergent Basis Sets for Explicitly Correlated Wavefunctions: The Atoms H, He, B−Ne, and Al−Ar. *J. Chem. Phys.*, 128(8):084102, 2008.

[197] Utkarsh Ayachit. *The ParaView Guide: A Parallel Visualization Application.* Kitware, Inc., 2015.

[198] M. M. Quesada-Moreno, M. Fatima, R. Medel, C. Pérez, and M. Schnell. Sniffing out Camphor: The Fine Balance Between Hydrogen Bonding and London Dispersion in the Chirality Recognition with α-Fenchol. *Phys. Chem. Chem. Phys.*, 24(21):12849–12859, 2022.

[199] R. Sure and S. Grimme. Corrected Small Basis Set Hartree-Fock Method for Large Systems. *J. Comput. Chem.*, 34(19):1672–1685, 2013.

[200] E. Papajak and D. G. Truhlar. Convergent Partially Augmented Basis Sets for Post-Hartree-Fock Calculations of Molecular Properties and Reaction Barrier Heights. *J. Chem. Theory Comput.*, 7(1):10–18, January 2011.

[201] E. Papajak, J. Zheng, X. Xu, H. R. Leverentz, and D. G. Truhlar. Perspectives on Basis Sets Beauty: Seasonal Plantings of Diffuse Basis Functions. *J. Chem. Theory Comput.*, 7(10):3027–3034, 2011.

[202] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V.

Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. Gaussian 16 Revision A.03, 2016. Gaussian Inc. Wallingford CT.

[203] M. Fang, L.and Li, W. B. Lin, Y. Shen, and C. F. Chen. Synthesis, Structures, and Photophysical Properties of Optically Stable 1,16-Diphenyl-3,14-Diaryl-Substituted Tetrahydrobenzo[5]helicenediol Derivatives: Enantioselective Recognition Toward Tryptophan Methyl Esters. *J. Org. Chem.*, 82(14):7402–7409, 2017.

[204] A. Mairena, J. I. Mendieta, O. Stetsovych, A. Terfort, I. G. Stará, I. Starỳ, P. Jelínek, and K. H. Ernst. Heterochiral Recognition Among Functionalized Heptahelicenes on Noble Metal Surfaces. *Chem. Commun.*, 55(71):10595–10598, 2019.

[205] S. Kassem, T. van Leeuwen, A. S. Lubbe, M. R. Wilson, B. L. Feringa, and D. A. Leigh. Artificial Molecular Motors. *Chem. Soc. Rev.*, 46(9):2592–2621, 2017.

[206] K. Yavari, P. Aillard, Y. Zhang, F. Nuter, P. Retailleau, A. Voituriez, and A. Marinetti. Helicenes with Embedded Phosphole Units in Enantioselective Gold Catalysis. *Angew. Chem.*, 126(3):880–884, 2014.

[207] P. Aillard, A. Voituriez, and A. Marinetti. Helicene-Like Chiral Auxiliaries in Asymmetric Catalysis. *Dalton Transactions*, 43(41):15263–15278, 2014.

[208] S. Grimme and S. D. Peyerimhoff. Theoretical Study of the Structures and Racemization Barriers of [n] Helicenes (n = 3–6, 8). *Chem. Phys.*, 204(2-3):411–417, 1996.

[209] R. H. Martin and M. J. Marchant. Thermal Racemization of Hepta-, Octa-, and Nonahelicene: Kinetic Results, Reaction Path, and Experimental Proofs that the Racemization of Hexa- and Heptahelicene Does Not Involve an Intramolecular Double Diels-Alder Reaction. *Tetrahedron*, 30(2):347–349, 1974.

[210] M. J. Fuchter, M. Weimar, X. Yang, D. K. Judge, and A. J. P. White. An Unusual Oxidative Rearrangement of [7]-Helicene. *Tetrahedron Lett.*, 53(9):1108–1111, 2012.

[211] R. J. F. Berger, M. J. Fuchter, I. Krossing, H. S. Rzepa, J. Schaefer, and H. Scherer. Gold(I)-Mediated Rearrangement of [7]-Helicene to Give a Benzo[cd]pyrenium Cation Embedded in a Chiral Framework. *Chem. Comm.*, 50(40):5251–5253, 2014.

[212] O. Stetsovych, M. Švec, J. Vacek, J. V. Chocholoušová, A. Jančařík, J. Rybáček, K. Kosmider, I. G. Stará, P. Jelínek, and I. Starỳ. From Helical to Planar Chirality by On-Surface Chemistry. *Nat. Chem.*, 9(3):213–218, 2017.

[213] A. Schäfer, H. Horn, and R. Ahlrichs. Fully Optimized Contracted Gaussian Basis Sets for Atoms Li to Kr. *J. Chem. Phys.*, 97(4):2571–2577, 1992.

[214] R. H. Janke, G. Haufe, E.-U. Würthwein, and J. H. Borkent. Racemization Barriers of Helicenes: A Computational Study. *J. Am. Chem. Soc.*, 118(25):6031–6035, 1996.

[215] J. Barroso, J. L. Cabellos, S. Pan, F. Murillo, X. Zarate, M. A. Fernandez-Herrera, and G. Merino. Revisiting the Racemization Mechanism of Helicenes. *Chem. Comm.*, 54(2):188–191, 2018.

[216] G. R. Kiel, S. C. Patel, P. W. Smith, D. S. Levine, and T. D. Tilley. Expanded Helicenes: A General Synthetic Strategy and Remarkable Supramolecular and Solid-State Behavior. *J. Am. Chem. Soc.*, 139(51):18456–18459, 2017.

[217] Y. Nakakuki, T. Hirose, and K. Matsuda. Synthesis of a Helical Analogue of Kekulene: A Flexible π-Expanded Helicene with Large Helical Diameter Acting as a Soft Molecular Spring. *J. Am. Chem. Soc.*, 140(45):15461–15469, 2018.

[218] G. R. Kiel, K. L. Bay, A. E. Samkian, N. J. Schuster, J. B. Lin, R. C. Handford, C. Nuckolls, K. N. Houk, and T. D. Tilley. Expanded Helicenes as Synthons for Chiral Macrocyclic Nanocarbons. *J. Am. Chem. Soc.*, 142(25):11084–11091, 2020.

[219] F. Chen, W. Gu, A. Saeki, M. Melle-Franco, and A. Mateo-Alonso. A Sterically Congested Nitrogenated Benzodipentaphene with a Double π-Expanded Helicene Structure. *Organic letters*, 22(9):3706–3711, 2020.

[220] K. Fujise, E. Tsurumaki, G. Fukuhara, N. Hara, Y. Imai, and S. Toyota. Multiple Fused Anthracenes as Helical Polycyclic Aromatic Hydrocarbon Motif for Chiroptical Performance Enhancement. *Chem. Asian J.*, 15(16):2456–2461, 2020.

[221] K. Fujise, E. Tsurumaki, K. Wakamatsu, and S. Toyota. Construction of Helical Structures with Multiple Fused Anthracenes: Structures and Properties of Long Expanded Helicenes. *Chem. Eur. J.*, 27(14):4548–4552, 2021.

[222] S. Grimme, J. G. Brandenburg, C. Bannwarth, and A. Hansen. Consistent Structures and Interactions by Density Functional Theory with Small Atomic Orbital Basis Sets. *J. Chem. Phys.*, 143(5):054107, 2015.

[223] S. Grimme. Semiempirical Hybrid Density Functional with Perturbative Second-Order Correlation. *J. Chem. Phys.*, 124(3):034108, 2006.

[224] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, et al. MOLPRO Version 2020.1: A Package of *Ab Initio* Programs, 2020. see http://www.molpro.net.

[225] R. A. Mata and M. A. Suhm. Benchmarking Quantum Chemical Methods: Are We Heading in the Right Direction? *Angew. Chem. Int. Ed.*, 56(37):11011–11018, 2017.

[226] M. A. Blanco, A. Martín Pendás, and E. Francisco. Interacting Quantum Atoms: A Correlated Energy Decomposition Scheme Based on the Quantum Theory of Atoms in Molecules. *J. Chem. Theory Comput.*, 1(6):1096–1109, 2005.

[227] M. Kohout. A Measure of Electron Localizability. *Int. J. Quantum Chem.*, 97(1):651–658, 2004.

[228] R. F. W. Bader. Atoms in Molecules. *Acc. Chem. Res.*, 18(1):9–15, 1985.

[229] A. D. Becke and K. E. Edgecombe. A Simple Measure of Electron Localization in Atomic and Molecular Systems. *J. Chem. Phys.*, 92(9):5397–5403, 1990.

[230] E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen, and W. Yang. Revealing Noncovalent Interactions. *J. Am. Chem. Soc.*, 132(18):6498–6506, 2010.

[231] M. D. Newton, E. Switkes, and W. N. L. Localized Bonds in SCF Wavefunctions for Polyatomic Molecules. III C−H and C−C Bonds. *J. Chem. Phys.*, 53(7):2645–2657, 1970.

[232] M. Levy, W. J. Stevens, H. Shull, and S. Hagstrom. Transferability of Electron Pairs Between H_2O and H_2O_2. *J. Chem. Phys.*, 61(5):1844–1856, 1974.

[233] R. Z. Khaliullin, M. Head-Gordon, and A. T. Bell. An Efficient Self-Consistent Field Method for Large Systems of Weakly Interacting Components. *J. Chem. Phys.*, 124(20):204105, 2006.

[234] M. Sironi and A. Famulari. An Orthogonal Approach to Determine Extremely Localised Molecular Orbitals. *Theo. Chem. Acc.*, 103(5):417–422, 2000.

[235] M. Sironi, A. Genoni, M. Civera, S. Pieraccini, and M. Ghitti. Extremely Localized Molecular Orbitals: Theory and Applications. *Theo. Chem. Acc.*, 117(5):685–698, 2007.

[236] H. Stoll, G. Wagenblast, and H. Preuβ. On the Use of Local Basis Sets for Localized Molecular Orbitals. *Theor. Chim. Acta*, 57(2):169–178, 1980.

[237] K. R. Sundberg, J. Bicerano, and W. N. Lipscomb. Variationally Deorthogonalized Localized Molecular Orbitals. *J. Chem. Phys.*, 71(4):1515–1524, 1979.

[238] T. Nagata, O. Takahashi, K. Saito, and S. Iwata. Basis Set Superposition Error-Free Self-Consistent Field Method for Molecular Interaction in Multi-Component Systems: Projection Operator Formalism. *J. Chem. Phys.*, 115(8):3553–3560, 2001.

[239] W. Humphrey, A. Dalke, and K. Schulten. VMD – Visual Molecular Dynamics. *J. Mol. Graph.*, 14:33–38, 1996. see http://www.ks.uiuc.edu/Research/vmd/.

[240] M. Pitonak, P. Neogrády, J. Rezac, P. Jurecka, M. Urban, and P. Hobza. Benzene Dimer: High-Level Wave Function and Density Functional Theory Calculations. *J. Chem. Theory Comput.*, 4(11):1829–1834, 2008.

[241] E. A. Ballik and D. A. Ramsay. Ground State of the C_2 Molecule. *J. Chem. Phys.*, 31(4):1128–1128, 1959.

[242] P. F. Fougère and R. K. Nesbet. Electronic Structure of C_2. *J. Chem. Phys.*, 44(1):285–298, 1966.

[243] J. D. Watts and R. J. Bartlett. Coupled-Cluster Calculations on the C_2 Molecule and the ${C_2}^+$ and ${C_2}^-$ Molecular Ions. *J. Chem. Phys.*, 96(8):6073–6084, 1992.

[244] K. Pierloot, Q. M. Phung, and A. Domingo. Spin State Energetics in First-Row Transition Metal Complexes: Contribution of (3s3p) Correlation and Its Description by Second-Order Perturbation Theory. *J. Chem. Theory Comput.*, 13(2):537–553, 2017.

[245] C. J. Stein and M. Reiher. Measuring Multi-configurational Character by Orbital Entanglement. *Mol. Phys.*, 115(17-18):2110–2119, 2017.

[246] M. Feldt, Q. M. Phung, K. Pierloot, R. A. Mata, and J. N. Harvey. Limits of Coupled-Cluster Calculations for Non-Heme Iron Complexes. *J. Chem. Theory Comput.*, 15(2):922–937, 2019.

[247] L. Que Jr. The Road to Non-Heme Oxoferryls and Beyond. *Acc. Chem. Res.*, 40(7):493–500, 2007.

[248] I. M. B. Nielsen and C. L. Janssen. Double-Substitution-Based Diagnostics for Coupled-Cluster and Møller–Plesset Perturbation Theory. *Chem. Phys. Lett.*, 310(5-6):568–576, 1999.

[249] R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince III, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, et al. Psi4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability. *J. Chem. Theory Comput.*, 13(7):3185–3197, 2017.

[250] D. G. A. Smith, L. A. Burns, D. A. Sirianni, D. R. Nascimento, A. Kumar, A. M. James, J. B. Schriber, T. Zhang, B. Zhang, A. S. Abbott, et al. Psi4NumPy: An Interactive Quantum Chemistry Programming Environment for Reference Implementations and Rapid Development. *J. Chem. Theory Comput.*, 14(7):3504–3511, 2018.

[251] W. Jiang, N. J. DeYonker, and A. K. Wilson. Multireference Character for 3d Transition-Metal-Containing Molecules. *J. Chem. Theory Comput.*, 8(2):460–468, 2012.

[252] J. Wang, S. Manivasagam, and A. K. Wilson. Multireference Character for 4d Transition Metal-Containing Molecules. *J. Chem. Theory Comput.*, 11(12):5865–5872, 2015.

[253] W. Jiang, N. J. DeYonker, J. J. Determan, and A. K. Wilson. Toward Accurate Theoretical Thermochemistry of First Row Transition Metal Complexes. *J. Phys. Chem. A*, 116(2):870–885, 2012.

[254] Y. A. Aoto, A. P. de Lima Batista, A. Kohn, and A. G. S. de Oliveira-Filho. How to Arrive at Accurate Benchmark Values for Transition Metal Compounds: Computation or Experiment? *J. Chem. Theory Comput.*, 13(11):5291–5316, 2017.

Appendix A

Supplemental Materials for Chapter 3

A.1 Energy and Structural Analysis of Diols

Table A.1: Relative energies of different conformations of EDO, CHexDO and pinacol. Optimization and frequency calculations were done using (SCS)LMP2/aug-cc-pVTZ, H=cc-pVTZ and B3LYP-D3(BJ,abc)/ma-def2-TZVP. With dispersion is indicated by *disp* while without dispersion considerations were labeled as *nodisp*. PNO-F12 in the table refers to the PNO-LCCSD(T)-F12/cc-pVTZ level of theory.

	SCSLMP2		LMP2		B3LYP	
EDO	disp	nodisp	disp	nodisp	disp	nodisp
het2”	4.38	0.12	6.95	0.00	9.25	5.14
het3	1.70	0.00	3.40	0.06	4.78	1.14
het3$_b$’	3.42	0.05	6.17	0.19	8.86	3.64
het3’	3.36	0.54	5.44	0.60	5.93	1.63
het4	0.00	1.50	0.00	1.29	0.00	0.00
hom2”	6.93	3.25	8.81	2.65	10.62	7.47
hom3	4.90	2.64	6.90	2.69	7.23	3.15
hom3$_b$’	4.01	0.10	6.74	0.20	8.02	3.18
hom3’	5.17	2.44	7.20	2.44	6.73	2.66
hom3a	4.36	0.73	6.58	0.83	7.98	2.94
hom3a’	8.35	3.39	10.84	3.53	11.09	6.10
PNO-F12	**7.12**		**6.32**		**7.19**	
CHexDO						
het2”	2.23	0.00	4.50	0.00	6.91	4.34
het3	2.25	1.06	4.18	1.15	5.84	1.72
het3$_b$’	2.85	0.56	5.35	0.55	7.64	2.81
het3’	3.18	1.27	5.35	1.32	5.87	1.56
het4	0.00	2.67	0.00	2.67	0.00	0.00
hom2”	4.12	2.42	5.62	1.95	7.32	6.36
hom3	6.15	4.24	7.97	4.11	8.21	4.11
hom3’	5.03	2.92	6.81	2.76	6.54	2.93
hom3a	4.55	2.00	6.92	2.22	9.11	3.95
hom3$_b$’	3.12	1.08	5.30	1.14	6.78	3.05
PNO-F12	**6.14**		**5.86**		**6.32**	
Pinacol						
het2”	5.79	0.86	8.25	0.72	9.56	5.42
het3$_b$’	1.68	0.00	4.59	0.00	4.88	0.01
het3’	2.35	0.80	4.70	0.71	4.19	0.00
het4	0.00	2.50	0.00	2.32	0.00	0.67
hom3$_b$’	5.46	1.54	8.93	1.65	9.38	2.54
hom4	5.20	5.52	6.98	5.40	6.96	4.87
PNO-F12	**4.93**		**6.08**		**6.31**	

Table A.2: Intermolecular H-bonding (HO···H) of EDO systems at different levels of theory. Units in Å.

LMP2/aug-cc-pVTZ, H=cc-pVTZ

Molecular System	with dispersion				without dispersion			
het2”	1.983	1.983			2.133	2.133		
het3$_b$’	2.188	2.031	1.892		2.532	2.112	2.000	
het3	1.905	1.998	1.899		2.045	2.152	2.063	
het3’	1.928	1.976	1.920		2.085	2.131	2.100	
het4	1.959	1.959	1.959	1.959	2.111	2.111	2.111	2.111
hom2	1.926	1.973			2.081	2.145		
hom3a	1.841	1.972	2.065		1.956	2.215	2.244	
hom3a’	1.822	1.961	2.021		2.034	3.390	2.046	
hom3$_b$’	2.145	1.983	1.886		2.491	2.083	2.006	
hom3	1.847	1.980	1.959		1.956	2.154	2.136	
hom3’	1.876	1.971	1.961		2.004	2.159	2.135	

SCS-LMP2/aug-cc-pVTZ, H=cc-pVTZ

Molecular System	with dispersion				without dispersion			
het2”	2.052	2.052			2.178	2.178		
het3$_b$’	2.255	2.083	1.944		2.510	2.152	2.030	
het3	1.968	2.059	1.964		2.079	2.181	2.096	
het3’	1.997	2.038	1.992		2.119	2.161	2.133	
het4	2.024	2.024	2.024	2.024	2.144	2.144	2.144	2.144
hom2”	1.992	2.051			2.116	2.196		
hom3a	1.897	2.051	2.140		1.987	2.250	2.288	
hom3a’	1.866	2.075	2.102		2.063	3.474	2.073	
hom3$_b$’	2.232	2.031	1.937		2.499	2.112	2.035	
hom3	1.900	2.044	2.026		1.986	2.183	2.169	
hom3’	1.936	2.039	2.025		2.035	2.188	2.166	

B3LYP-D3(BJ,abc)/ma-def2-TZVP

Molecular System	with dispersion				without dispersion			
het2”	1.939	1.939			1.980	1.980		
het3$_b$’	2.141	1.988	1.876		2.203	2.046	1.915	
het3	1.881	1.973	1.865		1.925	2.008	1.911	
het3’	1.885	1.943	1.877		1.913	2.000	1.935	
het4	1.951	1.951	1.951	1.951	1.997	1.997	1.997	1.997
hom2”	1.898	1.932			1.957	1.977		
hom3a	1.810	1.926	2.018		1.832	1.998	2.091	
hom3a’	1.792	1.916	2.001		1.811	2.018	2.044	
hom3$_b$’	2.062	1.956	1.864		2.114	2.024	1.909	
hom3	1.828	1.940	1.916		1.852	2.015	1.987	
hom3’	1.841	1.935	1.921		1.870	2.019	1.990	

Table A.3: Intermolecular H-bonding (HO···H) of CHexDO systems at different levels of theory. Units in Å.

LMP2/aug-cc-pVTZ, H=cc-pVTZ

Molecular System	with dispersion				without dispersion			
het2”	2.002	2.002			2.175	2.175		
het3$_b$’	2.176	2.002	1.873		2.378	2.131	2.007	
het3	1.910	1.988	1.898		2.089	2.139	2.077	
het3’	1.960	1.956	1.921		2.140	2.122	2.106	
het4	1.939	1.939	1.939	1.939	2.099	2.099	2.099	2.099
hom2”	2.013	1.947			2.196	2.142		
hom3a	1.869	1.907	2.065		2.011	2.203	2.301	
hom3$_b$’	2.174	1.957	1.880		2.415	2.087	2.024	
hom3	1.866	1.952	1.942		1.987	2.153	2.159	
hom3’	1.923	1.959	1.948		2.055	2.172	2.173	

SCS-LMP2/aug-cc-pVTZ, H=cc-pVTZ

Molecular System	with dispersion				without dispersion			
het2”	2.071	2.071			2.215	2.215		
het3$_b$’	2.232	2.060	1.928		2.386	2.166	2.035	
het3	1.976	2.050	1.965		2.120	2.167	2.109	
het3’	2.031	2.019	1.993		2.171	2.150	2.138	
het4	2.005	2.005	2.005	2.005	2.132	2.132	2.132	2.132
hom2”	2.091	2.015			2.248	2.174		
hom3a	1.935	2.046	2.142		2.045	2.229	2.344	
hom3$_b$’	2.257	2.007	1.934		2.443	2.113	2.052	
hom3	1.916	2.020	2.016		2.014	2.182	2.192	
hom3’	1.985	2.028	2.019		2.085	2.197	2.204	

B3LYP-D3(BJ,abc)/ma-def2-TZVP

Molecular System	with dispersion				without dispersion			
het2”	1.959	1.959			2.007	2.007		
het3$_b$’	2.137	1.961	1.853		2.175	2.036	1.908	
het3	1.877	1.950	1.849		1.921	2.005	1.926	
het3’	1.912	1.922	1.868		1.947	1.986	1.939	
het4	1.928	1.928	1.928	1.928	1.988	1.988	1.988	1.988
hom2”	1.967	1.915			2.015	2.002		
hom3a	1.830	1.930	2.005		1.851	2.010	2.079	
hom3$_b$’	2.102	1.925	1.850		2.124	2.007	1.905	
hom3	1.834	1.913	1.896		1.860	2.009	1.993	
hom3’	1.860	1.917	1.907		1.898	2.017	2.003	

Table A.4: Intermolecular H-bonding (HO···H) of CHexDO systems at different levels of theory. Units in Å.

LMP2/aug-cc-pVTZ, H=cc-pVTZ

Molecular System	with dispersion				without dispersion			
het2”	1.982	1.982			2.394	2.394		
het3$_b$’	2.192	2.069	1.883		2.361	2.298	2.081	
het3’	1.975	1.980	1.995		2.223	2.207	2.256	
het4	1.963	1.963	1.963	1.963	2.185	2.185	2.185	
hom3$_b$’	1.906	2.099	2.090		2.193	2.306	2.313	
hom4	1.887	1.887	2.343	2.343	2.091	2.091	2.733	2.733

SCS-LMP2/aug-cc-pVTZ, H=cc-pVTZ

Molecular System	with dispersion				without dispersion			
het2”	2.054	2.054			2.430	2.430		
het3$_b$’	2.232	2.144	1.948		2.364	2.331	2.109	
het3’	2.056	2.048	2.075		2.243	2.235	2.277	
het4	2.035	2.035	2.035	2.035	2.212	2.212	2.212	2.212
hom3$_b$’	2.029	2.155	2.159		2.220	2.322	2.335	
hom4	1.953	1.953	2.473	2.475	2.113	2.113	2.758	2.757

B3LYP-D3(BJ,abc)/ma-def2-TZVP

Molecular System	with dispersion				without dispersion			
het2”	1.965	1.964			2.066	2.066		
het3$_b$’	2.173	2.014	1.853		2.226	2.138	1.937	
het3’	1.909	1.948	1.923		1.999	2.053	2.055	
het4	1.950	1.950	1.950	1.950	2.036	2.036	2.036	2.036
hom3$_b$’	1.916	1.986	2.094		2.014	2.055	2.234	
hom4	1.864	1.878	2.462	2.181	1.966	1.975	2.788	2.359

Table A.5: Distance of the center of masses R(CM-CM) of different systems optimized at different levels of theory with and without dispersion (disp and nodisp, respectively). Units in Å. Optimizations were done at (SCS)LMP2/aug-cc-pVTZ,H=cc-pVTZ and B3LYP-D3(BJ,abc)/ma-def2-TZVP.

	SCSLMP2		LMP2		B3LYP	
EDO	disp	nodisp	disp	nodisp	disp	nodisp
het2”	3.46	3.60	3.40	3.58	3.38	3.48
het3$_\text{b}$’	3.41	3.57	3.34	3.54	3.30	3.41
het3	3.27	3.46	3.18	3.42	3.14	3.26
het3’	3.35	3.53	3.26	3.50	3.21	3.32
het4	3.06	3.23	2.98	3.20	2.97	3.04
hom2”	3.39	3.53	3.33	3.51	3.32	3.40
hom3a	3.50	3.71	3.41	3.68	3.38	3.48
hom3a’	3.44	4.06	3.33	4.01	3.30	3.45
hom3$_\text{b}$’	3.41	3.59	3.33	3.56	3.25	3.41
hom3	3.31	3.51	3.22	3.47	3.14	3.30
hom3’	3.33	3.55	3.23	3.52	3.17	3.33
Average	3.36	3.58	3.27	3.54	3.23	3.35
CHexDO						
het2”	5.24	5.43	5.17	5.41	5.18	5.37
het3$_\text{b}$’	5.65	5.82	5.56	5.78	5.53	5.68
het3	5.66	5.84	5.57	5.80	5.53	5.66
het3’	5.66	5.85	5.56	5.81	5.50	5.66
het4	5.49	5.67	5.40	5.62	5.39	5.48
hom2”	5.14	5.34	5.07	5.32	5.09	5.25
hom3a	5.47	5.79	5.36	5.76	5.38	5.61
hom3$_\text{b}$’	5.24	5.52	5.12	5.48	5.07	5.39
hom3	5.50	5.76	5.36	5.72	5.27	5.53
hom3’	5.35	5.69	5.20	5.65	5.19	5.49
Average	5.44	5.67	5.34	5.64	5.31	5.51
Pinacol						
het2”	4.88	5.29	4.78	5.26	4.75	4.92
het3$_\text{b}$’	4.82	5.03	4.74	5.00	4.69	4.83
het3’	4.81	5.03	4.72	5.01	4.65	4.81
het4	4.56	4.79	4.47	4.76	4.45	4.57
hom3$_\text{b}$’	4.96	5.19	4.87	5.17	4.80	4.98
hom4	4.68	4.90	4.58	4.87	4.56	4.73
Average	4.78	5.04	4.69	5.01	4.65	4.81

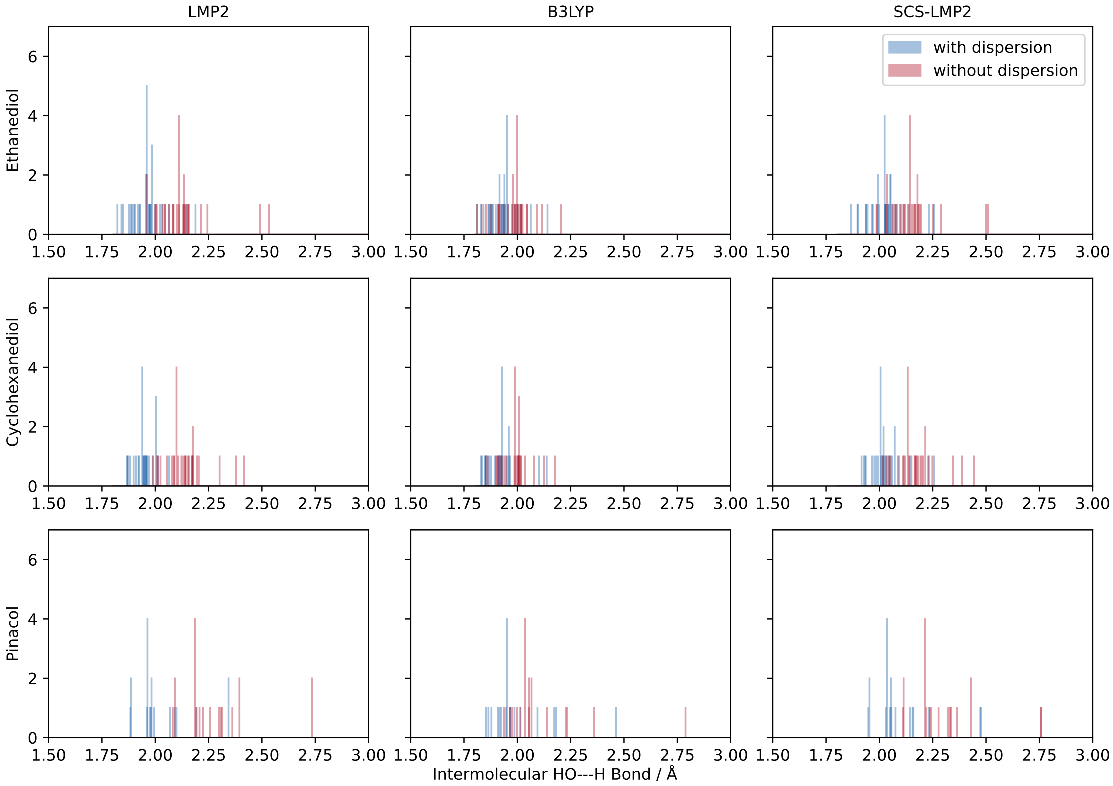


Figure A.1: Histogram counterpart of Figure 3.3 from Section 3.3

Generation of KDE Plots. To demonstrate the impact of omitting dispersion contributions on the intermolecular HO···H bond distance, a kernel density estimate (KDE) plot was utilized. This plot provides a visual representation of the distribution of O–H values, akin to a histogram. The KDE presents the data through a continuous probability density curve in one or more dimensions.

Given that the HO···H bond distances in the preceding tables are reported with precision up to three decimal places, a bin width of 0.001 was selected to depict the data. This choice ensures that each value finds representation within its respective bin, with higher frequencies in the plots indicating multiple occurrences of the same HO···H bond distances. In this section, a parallel histogram is presented to illustrate the data's appearance before applying the continuous probability density curve through KDE.

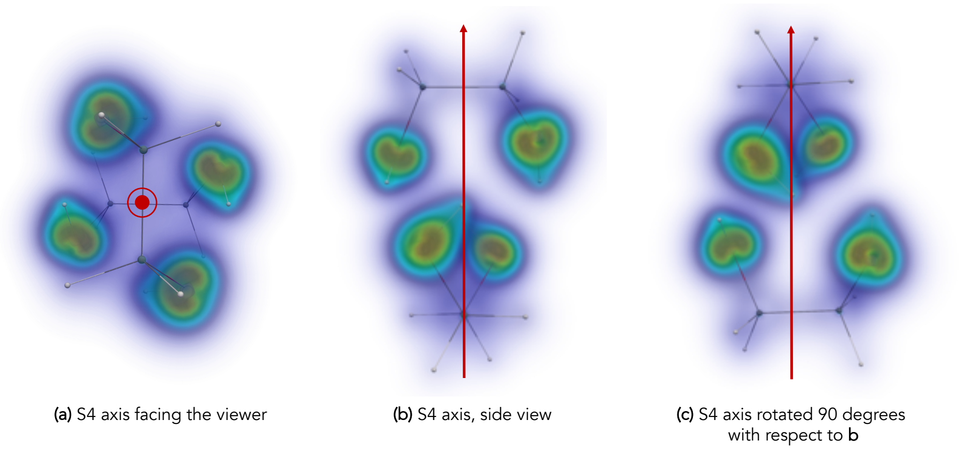

(a) S4 axis facing the viewer
(b) S4 axis, side view
(c) S4 axis rotated 90 degrees with respect to **b**

Figure A.2: Dispersion interaction density of het4 EDO viewed in different perspectives of its S_4 symmetry. Visualization was done using ParaView 5.4.0.

Optimized Structures. All the optimized structures of EDO, CHexDO and Pinacol are deposited in the university repository. Collection of data can be found here: DOI:10.25625/UQWQKZ

A.2 Energy Values and Experimental Data for Fenchone-Fenchol Systems

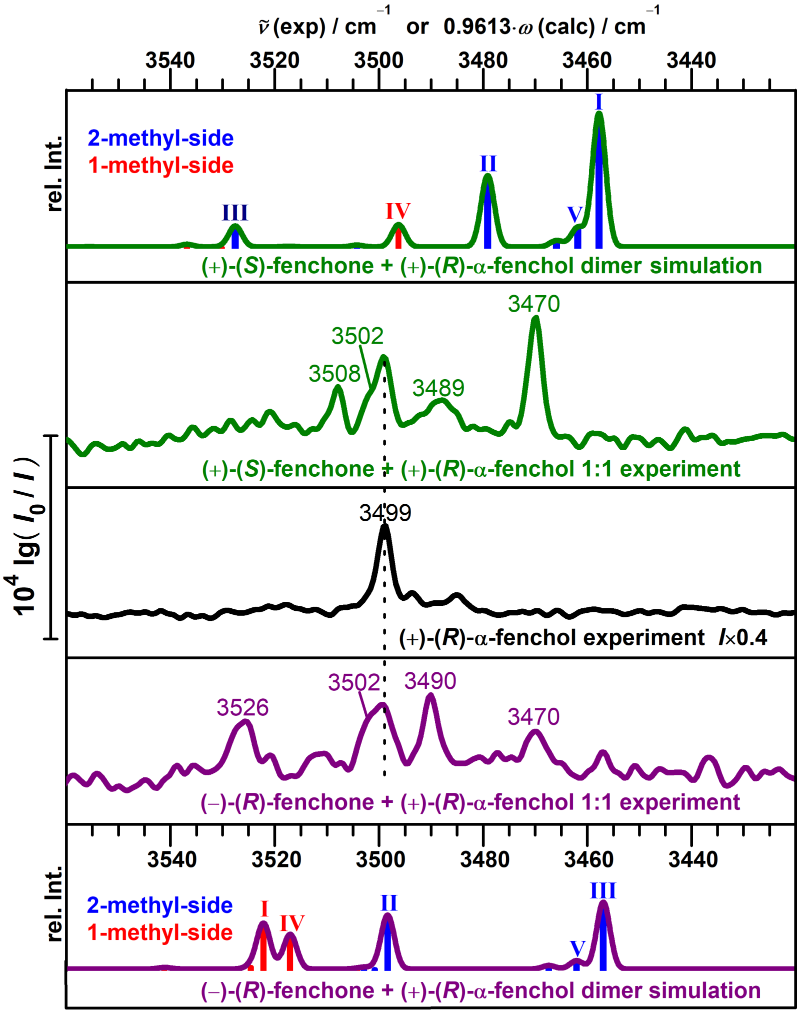


Figure A.3: FTIR jet spectroscopy spectra, both experimental and theoretical results are reflected (supplement information for Section 3.4. Theoretical results were obtained from calculations using B3LYP/may-cc-pVTZ with D3BJ corrections. Results provided by Dr. Robert Medel from the group of Prof. Dr. Martin Suhm.

Optimized Structures. All the optimized structures of EDO, CHexDO and Pinacol are deposited in the university repository. Collection of data can be found here: DOI:10.25625/UQWQKZ

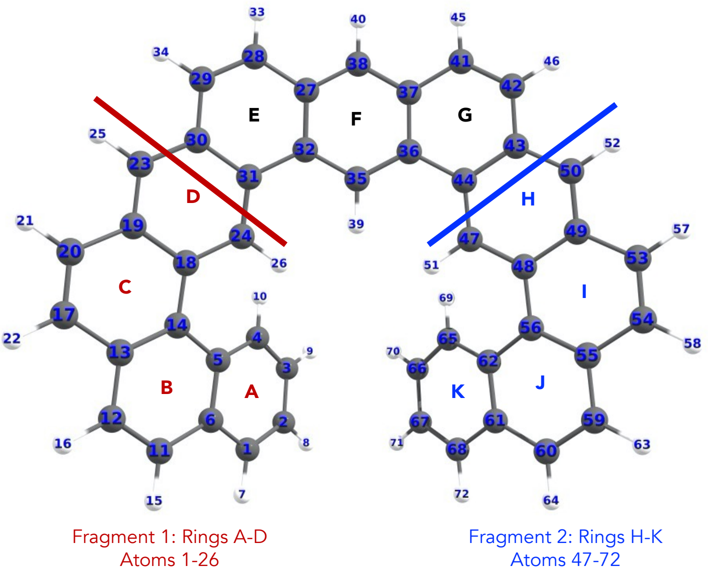
Table A.6: Final fits from the *SR* and *RR* systems, alongside the two similarly predicted structures from the B3LYP-D3(BJ)/may-cc-PVTZ level of theory using Gaussian 16[202] (supplement information for Section 3.4. Experimental data provided by Johann Benedikt Meyer, Moritz Niessner, Pablo Pinacho, Jun. Prof. Daniel Obencahin and Prof. Melanie Schnell, collected in Hamburg. Experiments were ran using COMPACT spectrometer.

	Rotational Constants (exp)		Rotational Constants (theo)	
Rotational parameters	*SR* system	*RR* system	*SR*-I isomer	*RR*-I isomer
A/MHz	589.54132(28)	521.84039(21)	590.47	522.66
B/MHz	139.076286(45)	171.675050(47)	141.56	175.02
C/MHz	130.289328(42)	160.028737(52)	132.48	162.07
D_J/kHz	0.003870(27)	0.009598(50)	0.003707	0.009575
D_{JK}/kHz	0.01322(85)	0.02470(57)	0.008742	0.032014
D_K/kHz	−0.0141(67)	−0.0284(36)	−0.002202	−0.030713
d_1/kHz	−0.000320(20)	0.000228(36)	−0.000371	0.000435
d_2/kHz	−0.000056(15)	[0]	−0.000046	−0.000135
N	370 (248/0/122)	307 (226/0/81)		
rms/kHz	6.9	6.9		
μ(a/b/c) /D			−3.9/0.2/0.8	−3.5/−0.2/−1.9

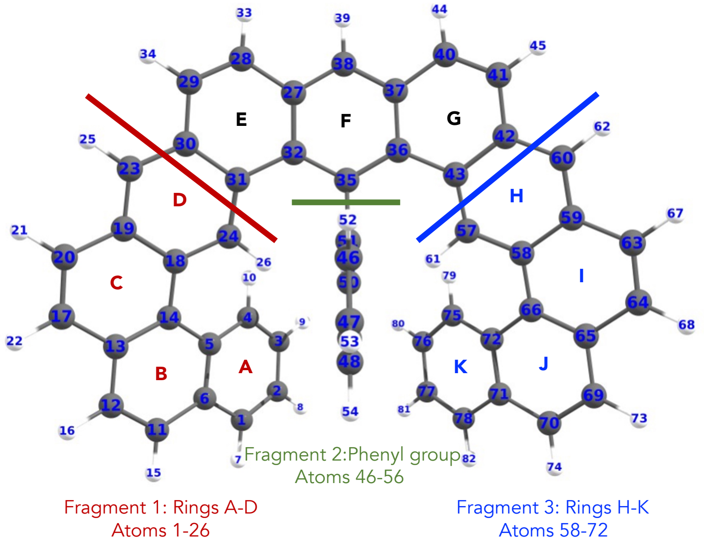
Table A.7: The specifics of energy decomposition analysis (EDA) are elucidated in Table 3.1, wherein the distinct energy components ΔE_{ionic} and ΔE_{intra} have been individually outlined. Note that these components, previously collected under the $\Delta E_{\text{non-disp}}$ term in the primary text, have now been segregated for detailed examination.

	ΔE (HF)			ΔE (PNO-SCS-LMP2)				E_{int}
Isomer	ΔE_{elec}	ΔE_{exch}	ΔE_{rep}	ΔE_{pol}	ΔE_{disp}	ΔE_{ionic}	ΔE_{intra}	
SR-I	−54.2	−75.6	135.2	−23.4	−17.7	−66.0	65.9	−35.8
SR-II	−50.5	−70.2	125.1	−22.0	−17.9	−59.8	59.7	−35.5
SR-III	−44.8	−70.5	123.5	−19.3	−21.4	−57.0	56.4	−33.1
SR-IV	−48.9	−72.9	129.1	−20.9	−19.3	−60.9	60.6	−33.2
RR-I	−44.3	−72.2	126.2	−19.6	−22.5	−57.1	55.8	−33.7
RR-II	−49.0	−71.2	125.9	−21.2	−19.5	−59.4	59.2	−35.1
RR-III	−54.1	−75.1	134.4	−23.2	−17.7	−65.5	65.6	−35.6
RR-IV	−45.4	−74.0	129.1	−19.3	−22.3	−57.4	56.2	−33.2

A.3 Fragmentation of Helicenes



(a) Compound **[4]**$_{\mathrm{H}}$



(b) Compound **[4]**$_{\mathrm{Ph}}$

Figure A.4: The partitioning scheme applied to Compounds **[4]**$_{\mathrm{H}}$ and **[4]**$_{\mathrm{Ph}}$ for the DID visualization in Figure 3.19. Each image illustrates the fragment partitioning considered in the DID visualization, excluding rings E-G which serves as backbone in both compounds. DID visualization was done at the SCS-LMP2/cc-pVTZ level of theory using Molpro 2021.2.

Appendix B

Supplemental Materials for Chapter 4

B.1 Geometries, Sample Calculations and Details of Pair Orbital Analysis

Optimized Coordinates of Argon Dimer, Figure 4.4

```
  2
Energy:  -1054.041738286098
Ar   0.0000000000   -1.9425700937   -0.0218263000
Ar   0.0000000000    1.9425700937    0.0218263000
```

Optimized Coordinates of Water Dimer, Figure 4.4

```
 6
Energy:  -152.657596241358
O   -0.0104947690    1.5497297660   -0.0000001037
H   -0.0877621489    0.5854519968    0.0000004501
H   -0.9184599534    1.8630206428   -0.0000001205
O    0.0213336613   -1.4000084528   -0.0000001150
H    0.5094776987   -1.7291486597    0.7609377676
H    0.5094776812   -1.7291489631   -0.7609378784
```

Optimized Coordinates of Methane Dimer, Figure 4.4

```
  10
Energy:  -80.843830700383
C     0.0000002216     0.0000000000     1.8918007919
H     0.0000000451     0.0000000000     2.9793383957
H     1.0250198064     0.0000000001     1.5285616208
H    -0.5125094820     0.8876929479     1.5285613702
H    -0.5125094819    -0.8876929480     1.5285613703
C    -0.0000002216    -0.0000000000    -1.8918007919
H    -0.0000000455     0.0000000001    -2.9793383958
H    -1.0250198063     0.0000000000    -1.5285616204
H     0.5125094820    -0.8876929481    -1.5285613705
H     0.5125094822     0.8876929479    -1.5285613703
```

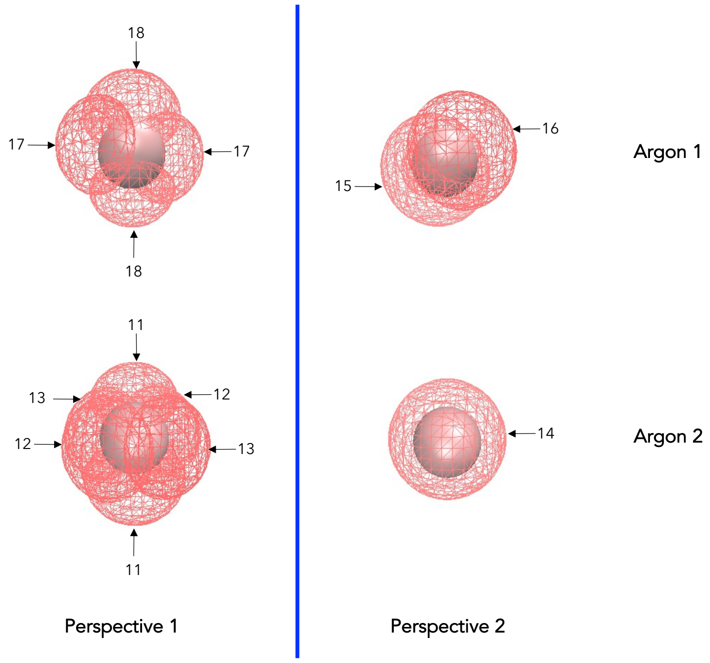
Optimized Coordinates of Benzene-Acetylene System, Figure 4.4

```
 16
Energy:  -308.894597826092
C    -2.3293162906     0.7342620543    -1.5946266252
C    -3.0779601496    -0.4173327486    -1.3442317014
C    -0.9385648289     0.6558776950    -1.6895064544
H    -4.1560848293    -0.3558127943    -1.2662468804
H    -0.3576820314     1.5496253822    -1.8792671039
C    -2.4358527041    -1.6473123469    -1.1887181082
C    -0.2964572553    -0.5741016941    -1.5339925477
H    -3.0160510414    -2.5395881650    -0.9901408024
H     0.7823518298    -0.6341491509    -1.6031601089
C    -1.0451010377    -1.7256967257    -1.2835977429
H    -0.5468321624    -2.6787556440    -1.1585943349
H    -2.8269002945     1.6887932099    -1.7108130588
C    -1.3268089448     0.2798573647     3.2055552436
C    -1.4192496567     0.0807922207     2.0135942491
H    -1.5004657325    -0.0941007482     0.9663731850
H    -1.2457481336     0.4544170129     4.2507740484
```

Unoptimized Water Dimer for Single Orbital Density Matrix Analysis, Figure 4.2

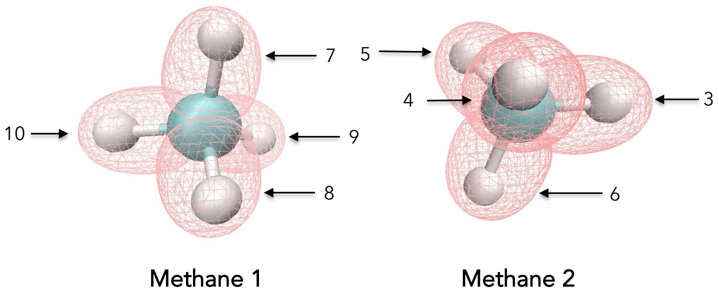
```
 6
Unoptimized Water Dimer
O    0.0    0.0    0.0
H    1.0    0.0    0.0
H    0.0    1.0    0.0
O    0.0    0.0    2.2
H    1.0    0.0    2.2
H    0.0    1.0    2.2
```

Table B.1: Details of the pair orbital analysis of argon dimer in Figure 4.6. The picture above shows the valence orbital densities for each monomer in a dimer. These numbers are the valence orbitals of every system automatically assigned in the procedure. Two perspectives are shown in order to clearly see the some valence orbitals which are hidden.



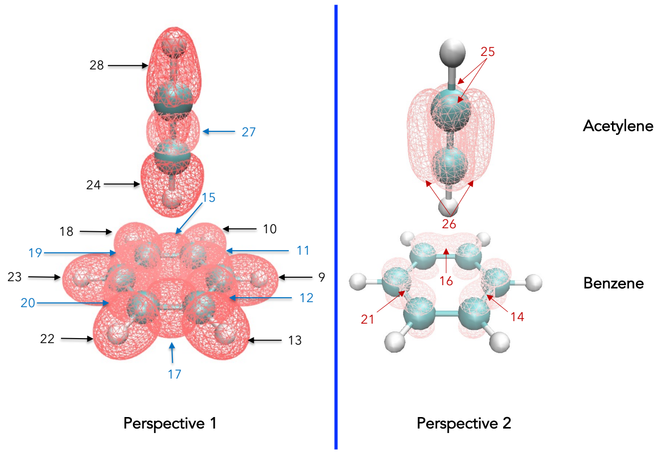
Orbital pair	ΔE_{disp}, kJ mol^{-1}	% ΔE_{disp} per orbital pair
15–11	−0.05	4.1
15–12	−0.02	2.1
15–13	−0.03	2.2
15–14	−0.01	0.6
16–11	−0.11	9.2
16–12	−0.06	5.2
16–13	−0.07	5.7
16–14	−0.01	1.2
17–11	−0.12	10.3
17–12	−0.08	6.5
17–13	−0.07	5.9
17–14	−0.02	1.4
18–11	−0.26	22.0
18–12	−0.12	10.3
18–13	−0.12	10.3
18–14	−0.04	3.0
TOTAL	−1.19	100.0

Table B.2: Details of the pair orbital analysis for methane dimer in Figure 4.7. The picture above shows the valence orbital densities for each monomer in a dimer. These numbers are the valence orbitals of every system automatically assigned in the procedure.



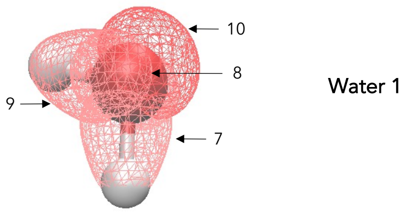
Orbital pair	ΔE_{disp}, kJ mol^{-1}	% ΔE_{disp} per orbital pair
7–4	-0.17	8.7
7–5	-0.17	8.7
7–6	-0.12	6.1
8–4	-0.17	8.7
8–5	-0.12	6.1
8–6	-0.17	8.7
9–4	-0.12	6.1
9–5	-0.17	8.7
9–6	-0.17	8.7
7–3	-0.09	4.4
8–3	-0.09	4.4
9–3	-0.09	4.4
10–4	-0.09	4.4
10–5	-0.09	4.4
10–6	-0.09	4.4
10–3	-0.06	3.1
TOTAL	-1.98	100.0

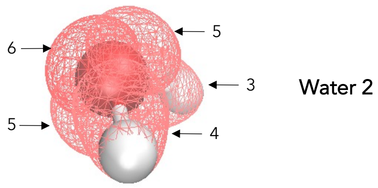
Table B.3: Details of the pair orbital analysis for benzene-acetylene system in Figure 4.8. The picture above shows the valence orbital densities for each monomer in a dimer. These numbers are the valence orbitals of every system automatically assigned in the procedure.



Orbital pair	ΔE_{disp}, kJ mol^{-1}	% ΔE_{disp} per orbital pair	Orbital pair	ΔE_{disp}, kJ mol^{-1}	% ΔE_{disp} per orbital pair
24–16	−0.81	9.3	25–18	−0.06	0.7
24–21	−0.81	9.3	25–13	−0.06	0.7
24–14	−0.81	9.3	25–10	−0.06	0.7
25–16	−0.41	4.7	25–22	−0.06	0.7
26–21	−0.40	4.6	26–22	−0.06	0.6
26–14	−0.39	4.5	26–10	−0.06	0.6
25–14	−0.38	4.4	26–13	−0.05	0.6
25–21	−0.37	4.3	26–18	−0.05	0.6
26–16	−0.37	4.2	25–9	−0.05	0.6
24–19	−0.18	2.0	25–23	−0.05	0.6
24–11	−0.18	2.0	27–19	−0.04	0.4
24–17	−0.18	2.0	27–11	−0.04	0.4
24–15	−0.16	1.9	27–17	−0.04	0.4
24–20	−0.16	1.9	27–15	−0.03	0.4
24–12	−0.16	1.9	27–20	−0.03	0.4
27–16	−0.15	1.7	27–12	−0.03	0.4
27–21	−0.15	1.7	27–18	−0.02	0.3
27–14	−0.15	1.7	27–10	−0.02	0.3
24–23	−0.09	1.0	27–23	−0.02	0.3
24–18	−0.09	1.0	27–9	−0.02	0.3
24–10	−0.09	1.0	27–22	−0.02	0.3
24–22	−0.09	1.0	27–13	−0.02	0.3
24–13	−0.09	1.0	28–21	−0.02	0.2
24–9	−0.09	1.0	28–14	−0.02	0.2
25–17	−0.07	0.8	28–16	−0.02	0.2
26–11	−0.07	0.8	28–19	0.00	0.0
26–19	−0.07	0.8	28–11	0.00	0.0
25–19	−0.07	0.8	28–17	0.00	0.0
25–11	−0.07	0.8	28–15	0.00	0.0
26–17	−0.07	0.8	28–20	0.00	0.0
26–15	−0.07	0.8	28–12	0.00	0.0
25–20	−0.07	0.8	28–23	0.00	0.0
25–12	−0.07	0.8	28–22	0.00	0.0
26–12	−0.07	0.8	28–18	0.00	0.0
26–20	−0.07	0.8	28–13	0.00	0.0
25–15	−0.07	0.8	28–10	0.00	0.0
26–23	−0.06	0.7	28–9	0.00	0.0
26–9	−0.06	0.7	TOTAL	−8.75	100.0

Table B.4: Details of the pair orbital analysis for water dimer in Figure 4.9. The picture above shows the valence orbital densities for each monomer in a dimer. These numbers are the valence orbitals of every system automatically assigned in the procedure.





Orbital pair	ΔE_{disp}, kJ mol^{-1}	% ΔE_{disp} per orbital pair
7–3	−0.34	9.9
7–4	−0.34	9.9
7–5	−0.83	23.7
7–6	−0.63	18.0
8–3	−0.11	3.2
8–4	−0.11	3.2
8–5	−0.24	7.0
8–6	−0.17	4.7
9–3	−0.07	2.0
9–4	−0.07	2.0
9–5	−0.18	5.3
9–6	−0.10	3.0
10–3	−0.05	1.4
10–4	−0.05	1.4
10–5	−0.12	3.3
10–6	−0.07	2.1
TOTAL	−3.49	100.0

Appendix C

Supplemental Materials for Chapter 5

C.1 Structures for CH_3OH and NCOH

Optimized Structure for CH_3OH, calculated at the MP2/cc-pVTZ using Gaussian 16.[202]

```
 6
Optimized Structure
C    0.663101000   -0.019517000    0.000010000
H    1.078461000    0.983414000   -0.000924000
H    1.027551000   -0.539098000    0.887769000
H    1.027858000   -0.540728000   -0.886725000
O   -0.748406000    0.121767000   -0.000016000
H   -1.125229000   -0.760617000   -0.000056000
```

Optimized Structure for NCOH, calculated at the MP2/cc-pVTZ using Gaussian 16.[202]

```
 4
Optimized Structure
N   -0.060221000    1.347167000    0.000000000
C    0.000000000    0.176201000    0.000000000
O    0.143887000   -1.119336000    0.000000000
H   -0.729554000   -1.532686000    0.000000000
```

C.2 Values of MR Diagnostics

Table C.1: Values of different diagnostic tests for several systems found in Figures 5.2-5.4. Optimized geometries and diagnostic values reflected in this table are found in References 50, 248.

Systems	T_1	T_2	D_1(MP2)	D_1(CCSD)	D_2(MP2)	D_2(CCSD)
HCl	0.00566	0.00429	0.00753	0.01072	0.12399	0.14338
HF	0.00668	0.00609	0.01026	0.01168	0.11003	0.11805
CO	0.01841	0.01671	0.03579	0.03883	0.14638	0.16155
CO_2	0.01794	0.01677	0.04313	0.04643	0.15078	0.15015
HCN	0.01421	0.01295	0.02651	0.02862	0.16775	0.17800
C_2	0.03861	0.02930	0.05139	0.08611	0.26411	0.37730
O_3	0.02522	0.02087	0.05795	0.07177	0.26972	0.26086
F_2	0.01025	0.00883	0.02330	0.02738	0.18164	0.21254

Table C.2: Individual D_2(CCSD) values of a dissociation reaction, where ΔMR_{De} and max D_2(CCSD) were plotted in Figure 5.6.

	Diatomic	Metal ion	Main Group Ion	$\lvert\Delta MR_{De}\rvert$	max D_2(CCSD)
CASE 1: All SR					
$AgCl \longrightarrow Ag^+ + Cl^-$	0.153	0.094	0.145	0.2	0.153
$AgF \longrightarrow Ag^+ + F^-$	0.133	0.094	0.117	0.1	0.133
$AuCl \longrightarrow Au^+ + Cl^-$	0.161	0.101	0.145	0.1	0.161
$AuF \longrightarrow Au^+ + F^-$	0.150	0.101	0.117	0.1	0.150
$CuF \longrightarrow Cu^+ + F^-$	0.169	0.086	0.117	0.5	0.169
CASE 2: All MR					
$PtC \longrightarrow Pt^{2+} + C^{2-}$	0.201	0.290	0.344	1.6	0.344
$ScH \longrightarrow Sc^+ + H^-$	0.183	0.294	0.236	1.9	0.294
CASE 3: Mix of SR and MR					
$AgH \longrightarrow Ag^+ + H^-$	0.178	0.094	0.236	0.3	0.236
$AgO \longrightarrow Ag^{2+} + O^{2-}$	0.143	0.093	0.183	0.5	0.183
$AuH \longrightarrow Au^+ + H^-$	0.173	0.101	0.236	0.5	0.236
$AuO \longrightarrow Au^{2+} + O^{2-}$	0.220	0.104	0.183	1.1	0.220
$AuS \longrightarrow Au^{2+} + S^{2-}$	0.179	0.104	0.184	0.6	0.184
$CoCl \longrightarrow Co^+ + Cl^-$	0.186	0.103	0.145	3.7	0.186
$CoH \longrightarrow Co^{2+} + O^{2-}$	0.210	0.103	0.236	2.8	0.236
$CoO \longrightarrow Co^{2+} + O^{2-}$	0.125	0.089	0.183	7.7	0.183
$CuCl \longrightarrow Cu^+ + Cl^-$	0.236	0.086	0.145	1.1	0.236
$CuH \longrightarrow Cu^+ + H^-$	0.154	0.086	0.236	1.4	0.236
$CuO \longrightarrow Cu^+ + O_2{}^-$	0.218	0.078	0.183	0.8	0.218
$CuS \longrightarrow Cu^{2+} + S^{2-}$	0.205	0.078	0.184	0.3	0.205
$FeCl \longrightarrow Fe^+ + Cl^-$	0.204	0.118	0.145	1.8	0.204
$FeH \longrightarrow Fe^+ + H^-$	0.199	0.115	0.236	0.7	0.236
$FeO \longrightarrow Fe^{2+} + O^{2-}$	0.225	0.098	0.183	2.1	0.225
$IrC \longrightarrow Ir^{2+} + C^{2-}$	0.153	0.206	0.344	2	0.344
$IrO \longrightarrow Ir^{2+} + O^{2-}$	0.122	0.206	0.183	15.3	0.206
$NiCl \longrightarrow Ni^+ + Cl^-$	0.187	0.095	0.145	3.7	0.187
$NiF \longrightarrow Ni^+ + F^-$	0.254	0.095	0.117	3.4	0.254
$NiH \longrightarrow Ni^+ + H^-$	0.213	0.095	0.236	3	0.236
$NiO \longrightarrow Ni^{2+} + O^{2-}$	0.172	0.083	0.183	0.9	0.183
$PtH \longrightarrow Pt^+ + H^-$	0.191	0.124	0.236	0.6	0.236
$PtO \longrightarrow Pt^{2+} + O^{2-}$	0.189	0.111	0.183	5	0.189
$RuC \longrightarrow Ru^{2+} + C^{2-}$	0.302	0.145	0.344	5	0.344
$RuO \longrightarrow Ru^{2+} + O^{2-}$	0.325	0.120	0.183	0.6	0.325
$ScF \longrightarrow Sc^+ + F^-$	0.154	0.294	0.117	1.6	0.294
$ZnCl \longrightarrow Zn^+ + Cl^-$	0.185	0.081	0.145	0.1	0.185
$ZnH \longrightarrow Zn^+ + H^-$	0.248	0.081	0.236	0.4	0.248
$ZnO \longrightarrow Zn^{2+} + O^{2-}$	0.210	0.072	0.183	1.1	0.210