NTNU
Norwegian University of
Science and Technology

A Study of Thymine to Support Experimentalists

Interaction Energies and Core Ionisation Energies

Merete Falck

Chemical Engineering and Biotechnology
Submission date: June 2016
Supervisor: Ida-Marie Høyvik, IKJ
Co-supervisor: Henrik Koch, IKJ

Norwegian University of Science and Technology
Department of Chemistry

Abstract

This thesis aims at investigating the interaction energies between two thymine molecules at different geometries, discussing select wave function based approaches for finding interaction energies, and looking closer at the photochemical dimerisation of two thymine molecules in DNA. The motivation for this study is supporting experimental physicists wanting to investigate the cyclodimerisation of thymine closer. Their results indicate that no such product was attained during their experiments. The theoretical study of the interaction energies calculated here helped explain why. To help explain what the product was core ionisation energies were calculated. These indicate that the product present in the experiments is the most stable tautomeric form of thymine.

Sammendrag

Denne oppgaven ønsker å undersøke interaksjonen mellom to tyminmolekyler i ulike geometrier, diskutere ulike teoretiske bølefunksjons måter å finne interaksjons energier på, samt se nærmere på den fotokjemiske syklodimeriseringen som skjer mellom to tyminmolekyler i DNA. Motivasjonen bak oppgaven er å hjelpe eksperimentelle fysikere som ønsker å studere den sykliske dimerisasjonsreaksjonen mellom to tymin molekyler. Resultatene deres tyder på at ikke noe av det ønskede produktet var tilstedet etter forsøket. Det teoretiske studiet av interaksjonsenergier beregnet i denne studien kan forklare hvorfor. Teoretiske beregninger kunne deretter sannsynliggjøre at de eksperimentelle resultatene tyder på at produktet er den mest stabile tautomeren til tymin.

I cannot even imagine where I would be today were it not for that handful of friends who have given me a heart full of joy. Let's face it, friends make life a lot more fun.

- Charles R. Swindoll

Acknowledgements

The author wishes to thank her supervisor, Assoc. Prof. Ida-Marie Høyvik, for all the support and help through the work of this project. Not many fellow students describe their supervisor as so dedicated to helping and supporting her master students as she is. The regular meeting help to produce progress in the work, while also allowing us master students to get insight into each other's work. Many thanks goes to the other master students for making the working sessions and lunches fun and to other good friends for all the support. Gratitude to Achim Gerstenberg and Jørgen Erichsen for proofreading the thesis. A big thank you to Jakob Vinje for providing me with experimental data, and for explaining his master's thesis. Thank you to my parents for both financial and moral support. At last, gratitude you goes out to my wonderful rock who has kept me going and believed in me throughout this work.

Contents

Appendix

Abbreviations

aug-cc-pVXZ	augmented correlation consistent
BSIE	Basis Set Incompleteness Error
BSSE	Basis Set Superposition Error
CBS	Complete Basis Set
CCSD	Coupled-Cluster Singles and Doubles
CCSDT	Coupled-Cluster Singled, Doubles and Triples
CI	Configuration Interaction
CP	Counterpoise Corrected
CT	Charge Transfer
DFT	Density Functional Theory
FCI	Full Configuration Interaction
HF	Hartree-Fock
MBF	MidBond Functions
MP2	Second Order Møller-Plesset Perturbation
MPPT	Møller Plesset Perturbation Theory
NCI	Noncovalent Interactions
SNOOP	Same Number Of Optimized Parameters
TD-DFT	Time-Dependent Density Functional Theory
w/MBF	with MidBond Functions
wo/MBF	without MidBond Functions

1 Introduction

This study is motivated by a collaboration with experimental physicists at the Department of Physics, NTNU. They want to perform an experiment to investigate the thymine cyclobutan dimer. To do this they first absorb thymine on a MoS_2 surface, and thereafter, expose the thymines to UV radiation in an attempt to create the thymine cyclobutan dimer. Afterwards, they perform an X-ray photoemission spectroscopy (XPS) to investigate whether they have successfully generated the dimer. Such experiments are both costly, time consuming and the results can be hard to interpret. Understanding what reactions and energies to expect might help them in their work and when interpreting results. Comparing theory and experiments can help give complimentary views on the experiment. In this study, first the noncovalent interactions of two thymine molecules were investigated, and then the theoretical and experimental results were compared. From XPS spectres core ionisation energies can be obtained, and is one of the main means of interpreting the results, as the same atoms bound differently will have slightly different peaks in the spectra. These peaks can be found theoretically, with more control over the origin of the peak. The exact values often do not match perfectly, but the change in the peaks often match. To help validate the results of the experimentalists core ionisation energy calculations were performed.

Thymine is one of four different bases in our DNA. When a DNA strand contains two consecutive thymine molecules and it gets exposed to ultra violet (UV) radiation, dimerisation between the two thymines may occur, creating a cyclobutan ring between the two. This will alter the structure of DNA and it can alter the sequencing of the DNA strand, as the two molecules will interact differently with the rest of the cell. In a worst case, this dimerisation can lead to skin cancer.[1] The DNA double helix is held together by noncovalent interactions (NCI) through hydrogen bonding. NCIs are also important in many other areas in chemistry and biology, and understanding these better can lead to advances in many different important areas (e.g. drug design). The reason it is interesting here is that these forces influence how the thymine molecules arrange relative to each other on the surface used in the experimental setup. The thymine molecules are initially free to move around and knowing where these interactions are strongest, can help predict how the thymine molecules will position themselves in the experiment.

Interaction energies of noncovalent systems have been extensively studied theoretically[2–9], and several different approaches for determining them exist[5,6,8–16]. The importance of the field can be seen in its long history. The journal *Chemical Reviews* have this year produced its fourth thematic issue on the topic, with the others being publish in 1988, 1994 and

2000. A good understanding of methodology is important when deciding which method to use for this investigation.

The aim of this investigation is to offer support to the experimental physicists by helping them interpret their results. This aim is further divided into four different sub-goals. Firstly, to understand and discuss some of the different approaches for finding interaction energies theoretically, where the emphasis will be on wave function based methods. Secondly, to apply one of these methods to look closer at the noncovalent interactions of two thymine molecules in their ground state. Thirdly, there will be some discussion on the chemical aspects of the thymine dimerisation, and lastly, to compare the theoretical and experimental results. This thesis attempt to present information that can be of value for the experimentalists in reaching their aim of better understanding the dimerisation of thymine.

Computational chemistry can be a good tool for finding molecular properties. The introduction and improvement of computers have resulted in development of faster and more accurate procedures. Advances have been made both in wave function based methods[17] and non-wave function based methods[6,18]. Benchmark calculations done only about 20 years ago have been reiterated, and even some semi-empirical methods can reach almost benchmark level accuracy[19,20]. Being able to evaluate the different methods against each other in a structured way is important.

Interaction energies are a convenient tool to use in comparing the performance of the different methods. It is a relatively cheap quantity to find and its easily reproducible.[5] For consistency, different test sets[21–28] with smaller systems have been developed to evaluate the different approaches to determine the interaction energies. Often, these test sets separate their members into what forces are involved in the interaction. This separation enables evaluation of the performance across the interactions. Dispersion dominated systems often pose the greatest challenge, as many methods struggle to properly describe electron correlation correctly. The second order corrected Møller-Plesset perturbation, MP2, method, is one of the methods that overbinds the interaction energy of dispersion dominated systems. The coupled-cluster (CC) hierarchy of methods are better apt at including the long range dispersion forces, but first at the perturbative triples (CCSD(T)) level. The scaling of the CCSD(T) method makes it intractable for larger systems.

Two challenges when performing accurate interaction energy calculations, are the basis set superposition error (BSSE) and the basis set incompleteness error (BSIE). The former was sought fixed by the introduction of the counterpoise (CP) correction of Boys and Bernardi[29]. The method was met with scepticism[3,30], but has since been widely accepted and is now almost a requirement for a benchmark quality calculation[5]. The counterpoise correction has been the standard method to calculate interaction energies, but last year Kristensen et al.[31] published an article introducing the Same Number Of Optimized Parameters (SNOOP) scheme as an alternative. The SNOOP-scheme is both less computationally expensive and gives more accurate interaction energies.[31] To solve the problem of BSIE, a larger basis set can be used, which will also increase the computational cost. However, the basis set convergence is slow, compared to the cost of increasing the basis set size. Possible ways of approaching the complete basis set (CBS) energy is to use an extrapolation scheme[32–36], to use midbond functions[30,37] or to use composite schemes[38,39].

This thesis will first introduce and discuss different wave function based methods for calculating interaction energies. The concept of interaction energies will then be introduced, along with two approaches for calculating it. Thereafter, select theory on photochemical reactions will be introduced, as the thymine dimerisation is a photochemical reaction. Followed by an attempt to examine what others have discovered about the mechanism behind thymine dimerisation in DNA. Computational results describing interaction energies between two thymines in different configurations will be presented in Chapter 6, together with core ionisation energies computed for comparison to experiments. Lastly, conclusion and discussion is given.

For this investigation, the use and performance of the methods will be in focus. The more technical details about how these methods are constructed will not be discussed here. The goal is for this thesis to be accessible also for non-chemist, while at the same time be useful for other computational chemists. For readers not so familiar with density functional theory, a brief mention is added to Appendix A. For readers interested in a short review of the different ways of approaching the complete basis set limit for the CP-scheme, without increasing the basis set too much, a section on midbond functions, extrapolation schemes and composite schemes are presented in Appendix B.

2 Wave Function Methods

With the development of computers, vast possibilities in quantum chemistry have opened. There have been great developments in the field since the introduction of the Schrödinger equation in 1926. Computers are getting better, there is more memory available, and processors calculate faster. Still, many of the most accurate quantum chemical methods are out of reach for most larger systems. This chapter will discuss some of the most widely used wave function based methods, and their performance when used in interaction energy calculations. As mentioned in the introduction, the focus will be on the use and performance of the methods rather than the construction of them.

2.1 The Schrödinger Equation

The Schrödinger equation was postulated by the Austrian physicist Erwin Schrödinger in 1926.[40] This equation describes the time evolution of the wave function of a particle. It has had a substantial impact on the development of quantum mechanics, and it is said to have similar impact on this field as Newton's laws of motion has had in classical physics.[41] The Schrödinger equation forms the basis for the methods discussed here, along with many other methods. There is a limitation to the use of the Schrödinger equation, as it can only be solved exactly for one-electron systems. Hence, the need for approximations is apparent. The time-independent form of the Schrödinger equation can be expressed as:

$$\hat{H}\psi = E\psi \tag{2.1}$$

Where ψ is a time-independent wave function and $\hat{H}$ the Hamiltonian operator.

The Schrödinger equation is often solved in combination with the Born-Oppenheimer (BO) approximation. In this approximation, the nuclei are thought of as so much heavier than the electrons, that they are considered stationary while the electrons move around. This approximation allows for the Schrödinger equation to be separated into an electronic and a nuclear part, greatly simplifying calculations. The nuclei move on a potential energy surface, which is the solution to the electronic Schrödinger equation. The BO has problems when two solutions to the electronic Schrödinger equation come close in energy.[42]

2.2 The Hartree-Fock Method

It is challenging to solve the Schrödinger equation for manybody problems exactly, even in the Born-Oppenheimer approximation. In the early beginning of quantum mechanics, Hartree proposed to make an approximation where a many-electron wave function is written as a product of one-electron orbitals.[43] Hartree justified his approach in a qualitatively manner, before Slater proved it using a rigorous mathematical formulation. Slater noticed that any such product function could be obtained from the variational principle. The Hartree approximation does not comply with the Pauli exclusion principle and thus an antisymmetry requirement was needed.[43] A Slater determinant solved this issue, as interchanging rows or columns in a determinant result in a change of sign. A generalised form of a Slater determinant composed of N spinorbitals, ϕ_a, ϕ_b,..., ϕ_N is:[40]

$$\psi(1,2,\ldots,N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_a(1) & \phi_b(1) & \cdots & \phi_N(1) \\ \phi_a(2) & \phi_b(2) & \cdots & \phi_N(2) \\ \vdots & \vdots & \cdots & \vdots \\ \phi_a(N) & \phi_b(N) & \cdots & \phi_N(N) \end{vmatrix} \tag{2.2}$$

In quantum chemistry, the spatial parts of the one-electron wave functions are set up as linear combinations of atomic orbitals[44]

$$\phi_p(\mathbf{r}) = \sum_{\mu} \chi_\mu(\mathbf{r}) C_{\mu p} \tag{2.3}$$

where μ is an atomic orbital label, p the label for a molecular orbital and $C_{\mu p}$ are the molecular orbital coefficients. Considerations for choosing a basis set of atomic orbitals are described in section 3.3.

The Hartree-Fock formalism uses the assumption that each electron in their respective orbitals can move without explicit dependence of the instantaneous motion of the other electrons present, only the mean field created by them.[45]

$$\hat{F} = \sum_{i} \hat{f}(i) \tag{2.4}$$

Where $\hat{F}$ is the Fock operator for the mean field, and f(i) for each of the electrons present. The Hartree-Fock energies are eigenvalues of the Fock operator. The f(i) operator is

$$\hat{f}(i) = h(i) + \sum_{p} \{2J_p(i) - K_p(i)\} \tag{2.5}$$

where h(i) is the one-electron Hamiltonian, containing kinetic energy of the electrons and attractions between electrons and nuclei. J is the Coulomb operator and K the exchange operator, and the sum runs over the occupied molecular orbitals, p.[40] When solving the Hartree-Fock equations the orbital coefficients of Equation (2.3) are varied according to the variational principle to find the minimum value of the energy, E

$$E = \langle\psi|\hat{H}|\psi\rangle \tag{2.6}$$

The determinant corresponding to the minimum energy is called the Hartree-Fock approach wave function. In the variational method, any trial wave function, that is a linear variation of an eigenfunction to the Schrödinger equation, will give a solution higher in energy than the true value. Finding a lower energy thus means a more accurate energy.

The Hartree-Fock or self-consistent field method has long been a method of choice as it gives reasonable good results at a relatively low computational cost. It gives a reasonably good approximation of atomic systems and for interatomic bonding, but it is not able to achieve the desired accuracy for non-bonded systems.[43]

The correlation energy, is defined as the difference between the exact energy and the Hartree-Fock energy as

$$E_{\text{corr}} = E_{\text{exact}} - E_{\text{HF}} \tag{2.7}$$

within a given AO basis we can represent E_{exact} by the full configuration interaction (FCI) energy. Correlation can roughly be classified into two different concepts, dynamical or static correlation. The first is related to the instant correlation between two electrons, such as that between two electrons occupying the same orbital. The latter is related to near the degeneracy of electronic configuration. This will be dominant for almost degenerate states. There does not exist a clear way of separating these types of correlations, they can still be a useful concept.[42] The lack of correlation energy is often said to be the reason why the Hartree-Fock energy will not be able to fully describe the electronic properties of an atom or molecule, especially dispersion forces, where the correlations between the electrons are important. The Hartree-Fock theory gives good molecular geometries, and often come within picometers of the true geometry. The theory is also size-extensive, something that proves important for calculating interaction energies.[44] Hartree-Fock calculations will come very close to many molecular properties, but for an accurate description correlations must be included. This is what many of the methods building on the Hartree-Fock theory try to describe.[44]

In the Hartree-Fock theory the electrons only feel the average field created by the other electrons. Thus in quantum chemistry we define everything missing in the Hartree-Fock as electrical correlation.[46,47] The Hartree-Fock theory is the basis of many modern methods and the cornerstone of ab initio electronic-structure theory.[44] Both coupled-cluster theory and Møller-Plesset perturbation theory are based on the assumption that Hartree-Fock provides a good zero order approximation.[44] It is also routinely used for larger systems, and has been used on systems containing several hundred atoms. The Hartree-Fock wave function yields many different electronic properties to within few percentages from the true value.

Kohanoff[43] discusses possible ways of solving the issue of electron correlation; one can be to use many Slater determinants to describe the system, another to include electron correlation perturbatively above the HF solution. The Hartree-Fock wave function is the best single determinant wave function, and thus improving it is challenging - at least without imposing further approximations. Being a single reference method, the Hartree-Fock approach will not be robust enough to produce accurate results for systems where more than one electronic configuration dominates the system (static correlation).

2.3 Møller-Plesset Perturbation Theory

As noted in Section 2.2 the Hartree-Fock approximation does not include electron correlation. Perturbation theory presents a systematic way to try to improve the accuracy of the calculations as it provides a way of including different levels of perturbations.

$$\hat{H} = \hat{H}^{(0)} + \hat{U} \tag{2.8}$$

Where $\hat{H}$ is some zero order Hamiltonian and $\hat{U}$ is the perturbation. One disadvantage is that perturbation theory in general does not converge; meaning that including an ever-increasing amount of terms will not necessarily give a more accurate result.[44,48,49]

The Møller-Plesset perturbation theory[50] (MPPT) is a widely used perturbation theory in quantum chemistry, with the second order corrected perturbation, the MP2 method, being the most used of the MPPT methods. This is due to its low computational cost and satisfactory accuracy. Møller-Plesset perturbation theory uses the Fock operator as the zero order Hamiltonian. The Hartree-Fock determinant is used as the zero order determinant and the perturbations are carried out with respect to this state. The most successful application of this theory is where the Hartree-Fock wave function provides an adequate zero order approximation to the exact wave function.[44] Hence, the Møller-Plesset perturbation theory is a single-reference method that will only produce accurate answers if the Hartree-Fock state is a good first order approximation.

In the MPPT, the correlation correction arises first with the second order correction, since the energy through first order is the Hartree-Fock energy. Calculations using the third or forth order corrections, MP3 and MP4, can also be observed in literature. With MP4 having a scaling similar to CCSD, it is usually deemed better to use CCSD, than MP4.

The energy provided by each of the corrections will be size extensive, but the individual terms will not be.[44] However, the MPPT method is not a variational method, meaning that attaining a lower value will not guarantee a more accurate value.

2.3.1 Møller-Plesset Second Order Correction, MP2

MP2 is one of the simplest and least computational expensive wave function methods that include electron correlation, it is said to recover about 80-90% of the electron correlation.[43] It usually works well for hydrogen bonded and polar complexes but, tend to overestimate the interaction energy for dispersion bound and π-π stacked complexes. It is also not the best method for calculating dissociation limits, due to its incomplete ability to include electron correlation.[43] For interaction energy calculations, Řezáč and Hobza[5] suggest that using a smaller basis set will result in the error arising from the incomplete basis compensates for the MP2's incomplete ability to calculate dispersion bound complexes, and further suggest that the cc-pVTZ basis will be sufficient for this purpose. (The basis will be explained in Section 3.3.) It is most often used in combination with an extrapolation scheme, a CCSD(T) correction or one of the explicitly correlated MP2 methods. This to try to improve the accuracy without increasing the basis set too much, or to increase the computational cost too much. As they are so widely used with the CP-scheme more information about these are included in Appendix B

For more practical calculations (e.g. this investigation where many calculations have to be carried out), the low cost and relatively good accuracy makes the MP2 the method of choice. It is also a well-tested method.

Many attempts to improve the accuracy of the MP2 method itself have been suggested.[20,51–59] Some of these methods show promising improvements, while others fail even for the systems they were created for.[?] These will not be discussed further in this thesis.

2.3.1.1 Explicitly Correlated MP2 Methods

An approach that directly attempts to converge the MP2 calculations faster with respect to basis set size are the explicitly correlated methods. It is not until more recently that they have been introduced as methods for relatively large systems.[46] Different approaches for explicit correlations exist, not all suited for larger systems. The most widely used approach, at least for practical systems, is the R12 or F12 approach. The R12 approach employs two electron functions that linearly depend on the distance between two electrons, r_{ij}, to help speed up the convergence.[60,61] In the F12 approach, the linear r_{ij} term is replaced by a nonlinear function, f_{ij}. Often, a Slater type function is used as the f_{ij}

term.[47] Different Ansätze can be used for these methods, as different implementations have evolved. Ansatz 1 or 2 define which molecular space the R12/F12 functions are strongly orthogonal to.[60] Auxiliary basis functions were introduced because the RI approximation need a very large AO basis set to avoid unwanted errors. By choosing an optimised set of auxiliary basis functions, the error due to the RI approximation can be made almost negligible, without increasing the computational time too much.[60,62]

For computer implementation, different approaches, with a different amount of approximations, exist. The MP2-R12A (A'), MP2-R12B[60] and MP2-R12-3C[63,64] have all been implemented. The details of these approximations will not be described here. It is stated that the convergence of error due to truncation of the basis set in conventional calculations using correlation-consistent basis sets goes as X^{-3}, while it should be possible to achieve a convergence of about X^{-7} for the R12 method using approximation B.[60,65] Showing the great potential of the R12 methods. Řezáč and Hobza[5] claim that the explicitly correlated MP2 methods inherit some of the same flaws as the regular MP2 method. They still recommend them for larger systems as they outperform MP2, especially when using smaller basis sets.

Both Kong et al.[46] and Hättig et al.[47] have written extensive reviews of the R12/F12 methods, without going into too much details of the different Ansätze discussed above, but rather discussing the methods for MP2, CC and multireference methods. They also include an interesting evaluation of what electron correlation *really* is. Interested readers are thus referred to these reviews along with the review by Klopper et al.[66] and the summary by Ten-no[67].

2.4 The Coupled-Cluster Method

The Coupled-Cluster (CC) methods have proven to give quite accurate results for calculations in the electronic-structure theory, and CCSD(T)[68] is often quoted to be the gold standard. Due to the unfavourable N^7 scaling, it cannot always be used in calculations.[69] Coupled-cluster uses the Hartree-Fock state as a reference, and relies on the Hartree-Fock theory to provide a good zero order state. Truncated coupled-cluster method will thus only provide an accurate answer given that the Hartree-Fock state is well suited for the problem at hand.

For computer implementation, the coupled-cluster methods are given through an exponential Ansatz,[44]

$$|\mathrm{CC}\rangle = \exp(\hat{T})\,|\mathrm{HF}\rangle \tag{2.9}$$

where $\hat{T}$ is the cluster operator which excites out of the Hartree-Fock reference, and contains single, $\hat{T}_1$, double, $\hat{T}_2$, etc. excitations.

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \dots \tag{2.10}$$

Due to the exponential form, excitations can occur through both connected and disconnected excitations. (E.g quadrupole excitations can be obtained through $\hat{T}_4$, $\hat{T}_2^2$, $\hat{T}_1^4$, $\hat{T}_2\hat{T}_1^2$ and $\hat{T}_3\hat{T}_1$.) This makes the coupled-cluster method different from the configuration iteration (CI) methods, which are linearly parameterised.[44]

The coupled-cluster wave function is not variational, as the parameters are determined by projection rather than minimisation. Results obtained from coupled-cluster theory in practice usually differ with only a small amount from the FCI energy. The normally high accuracy and the size extensivity of the method outweigh the downside of loosing the variational principle.[44]

For calculations, is it impossible to include all possible excitations, and a cut off level must be specified. Because of the cut off level, the expression will no longer exactly satisfy the Schrödinger equation expressed in the exponential Ansatz. Now, it rather satisfies the Schrödinger equation in the exponential Ansatz projected onto the Hartree-Fock state and onto the projection manifold.[44] The cut off level is specified by the user, and the most common is to include excitation operators up to double or triple level (CCSD or CCSDT). The CCSDT calculations are very computationally expensive, thus hybrid models that estimates the triples contribution, have been suggested. The most commonly used is the CCSD(T) method, where the triples excitations are included perturbatively. This gives results close to those obtained for the full CCSDT, but at a lower computational cost.[44] Other methods including the triples excitations have also been developed[70,71], but will not be further discussed here.

The coupled-cluster methods are very computationally expensive. The CCSD(T) level has a scaling of N^7, compared to N^5 for MP2, putting CCSD(T) out of reach for larger systems.[31] For the investigation of the thymine dimer, a relatively large system, even CCSD will be very demanding, since many calculations will be carried out.

Complexes bound by diffusion forces are difficult to do calculations on, and it is reported that even the CCSD has problems with this.[69] As opposed to the MP2 method, CCSD underbinds the interaction, sometimes on the same order of magnitude as the overbinding of MP2.[22,69,72] It is thus sometimes stated that the triples excitations are needed for accurate calculations.[5] Řezáč and Hobza[23] investigates the error included in the gold standard CCSD(T)/CBS calculations by examining the effect of going to the CCSDT(Q) method, not using the frozen core approximation, and trying to include relativistic effects. They find that these effects changes the interactions by about 1.47% overall for the A24 test set. Confirming that the CCSD(T)/CBS values are quite accurate.

Inspired by the good behaviour of the explicitly correlated MP2 methods, methods for using explicitly correlated methods along with the coupled-cluster hierarchy have also been developed[73,74]. For a more detailed review on the coupled cluster methods, look to Bartlett and Musiał[75]. The review of Helgaker et al.[76] looks at the current status of quantum chemistry as a predictive tool. The article discusses coupled-cluster methods, explicitly correlated methods, and also methods beyond coupled-cluster theory.

In their experiments the physicists have obtained XPS spectra, these contain values for core ionisation energies. These core ionisation energies are one of their primary ways of interpreting the experimental results. It is only possible to determine what the spectral peaks signify if similar spectra for know groups are available. Core ionisation energies can also be calculated theoretically, with more control of the origin of the value. The CC hierarchy provides a possibility to perform such calculations, as long as the ground state can be well described by a single reference Hartree-Fock wave function.[77] Coriani and Koch[78] present a simple scheme to compute core ionisation energies using the CC hierarchy. This shows promising results, and will be used to find ionisation energies in this investigation. These calculations were performed to help the experimentalists interpret their XPS spectra by comparing the theoretical and experimental shifts in the peaks. The exact values often differ, but the shifts in values are still useful.

Wave function methods are viewed as the most accurate methods available in computational chemistry. These are very computationally expensive and need large basis sets to give benchmark quality results. Smaller basis sets can be used, knowing that the results will be less accurate, but more manageable. Recent developments in alternative methods have resulted in high accuracy at a lower computational cost for some of these alternative methods. DFT-based methods will not be used in this investigation. They are used in the literature on thymine, especially the time dependent form. Interested readers needing a short introduction to these methods, and a short evaluation of some functionals (performed by others) are referred to Appendix A.

Different mean field methods have developed that included dispersion correction. Grimme et al.[8] has written an extensive review about dispersion-corrected mean field methods that interested readers are referred to. Quantum Monte Carlo has been shown to give quite good interaction energies[5]. This method will not be described in this investigation, but interested readers can read the recent review by Dubecký et al.[6]. Semiempirical methods are also developing to becoming an important tool, especially for larger systems where other methods cannot reach. The article by Christensen et al.[9] reviews these methods for noncovalent interactions. Volume 116 of *Chemical Reviews* is a thematical issue about noncovalent interactions, from where many of the before mentioned reviews are published. Also other interesting topics are covered in that issue.

3 Interaction Energy

As stated in the introduction, noncovalent interactions (NCI) are a vital part of chemistry and biology. The existence of attractive forces between non-bonded molecules was first discovered by van der Waals in the 1870s, but it was not until the development of quantum mechanics that London managed to make sense of these forces, as they are based on quantum mechanical principles.[79] The NCIs are much subtler than covalent bonding, and also affect the electron structure in a different way. Highly sensitive equipment is thus needed to be able to separate them experimentally. Calculating NCIs requires the use of post-Hartree-Fock methods capable of describing electron correlations adequately. In addition to an appropriate method, a good description of the NCI often requires the use of a large basis set, as the forces are more long ranging, and diffuse than covalent forces. The noncovalent interactions become more important for larger molecules, and it thus of importance to find methods capable of doing accurate calculations on larger systems.

It is customary to categorise the noncovalent interactions into separate classes when finding interaction energies. The different classes being London dispersion, hydrogen bond, $\pi \cdots \pi$ systems, interaction between charged species, and systems with a constant dipole. The different test sets are often devised to cover all the groups, some of the interactions or they are specialised for a certain type of interaction. As discussed in Chapter 2, the different methods perform differently for the different interactions. A strict definition of how to categorise each molecule does not exist, but certain criteria are devised for classification reasons. The strength of the interactions are also of different orders, the charged or highly polarised systems are more tightly bound than the dispersion bound complexes An interesting note on hydrogen bonding is that for some hydrogen bonds charge transfer states are sometimes observed, which is a state where an almost covalent bond is formed.[5] For a deeper introduction to noncovalent bonds and to get both an experimental end theoretical view, the book by Hobza et al.[79] can be recommended.

In this investigation, the scope will be on the interactions between two, non-bonded, molecules. This can be defined by the formula:

$$\Delta E_{\mathrm{AB}} = E_{\mathrm{AB}} - E_{\mathrm{A}} - E_{\mathrm{B}} \tag{3.1}$$

Where E_{AB} is the energy of the complex and E_i the energy for the respective subsystem. This interaction is a theoretical construct, which is impossible to measure experimentally. In experiments only interaction enthalpy or dissociation energy are possible to measure. The interaction energy is incorporated into the interaction enthalpy, as this is one of three components making up this quantity, along with the deformation energy and the zero-point vibrational energies.[5] It is a quantity well suited for performing benchmark calculations, as it is an economical procedure that is easily reproducible. Finding the full interaction enthalpy would be considerably more demanding.[5] The methods to be described in the coming sections, all find only the interaction energy, and do not consider deformation energies or the zero-point vibrational energy. It is possible to include these measures, and developments are being made to more accurately calculate the full interaction enthalpy, but these methods are outside of the scope of this thesis.

Due to computational limits, it is almost impossible to do calculations using a complete basis set, something that will lead to errors occurring if Equation (3.1) is used as it is. The counterpoise correction (CP) scheme was created to alleviate this problem. To test the validity of these methods, several benchmark sets have been presented consisting of relatively small molecules bound together by a variety of forces. A well used test set is the S22 set of Jurečka et al.[21], consisting of 22 molecules containing only C, H, N and O, and with the most typical non-covalent interactions like hydrogen bonds, dispersion forces and mixed electrostatic-dispersion interactions. Several improvements to these sets have been suggested[80–83], as well as methods being tested against this set[56–58,84–86]. Such test sets are crucial for the development of accurate methods that can also be used on larger, more biologically relevant, systems. They also provide accurate results for values unattainable by experiments.[6]

When no direct experimental values are available, how can one decide which method gives the most accurate answer, and what are acceptable errors for a benchmark calculations? For some, very small systems, the FCI energy is available, and the interaction energy can be found using almost no approximation. These values can be used to approximate the error of other methods. For methods where higher levels of theory are available, as in the CC theory, the cheaper methods often get compared to the more expensive ones.[87] Also as many, very different, methods give similar answers, this serve as a validation of the methods.[6] Still, many has not been thoroughly tested for systems above a certain size, as these are too computationally expensive at the moment.[6] It is generally assumed that methods working for small systems will give good accuracy also for larger systems. Dubecký et al.[6] claim that an accuracy of 1 kcal/mol is now often required for a benchmark calculation. For systems where the entire interaction energy is less than this, it is scaled down to 0.1 kcal/mol accuracy.

In this investigation, the interest is on the interaction between two thymine molecules at their ground state. Thymine is a relatively large molecule, bound by dispersion-dominated interactions. Performing the calculations using the most accurate and computational expensive method, as described in Chapter 2, will thus pose a challenge. The coming section will go deeper into the counterpoise corrected scheme and the same number of optimized parameters scheme. Followed a description and choice of basis sets.

3.1 Counterpoise Corrected Scheme

From Equation (3.1) the definition of the interaction energy is the difference between the full system and its separated constituents. The interaction energy can be found by performing an ab initio calculation of the energy of the full system and the energy of the parts and then subtract the latter from the former. This method is called the uncorrected method and it will normally give an interaction energy that is too attractive.[31,44] The overestimation is said to be due to basis set superposition error (BSSE), as the basis set used for the dimer will be larger than for that used for the subsystems. There will most often also be some basis set incompleteness error (BSIE), as it is not possible to use the complete basis set for other than the smallest systems.

The counterpoise correction (CP) scheme, introduced by Boys and Bernardi[29], is designed to make the BSSE smaller, and thereby the interaction energy more accurate. The CP-scheme tries to minimise BSSE by calculating the energy of the subsystems A and B, using the same basis set for both monomers as used for the dimer, (i.e. the energy of monomer i is calculated with the presence of ghost functions from monomer j). Mathematically it is expressed as:[31]

$$\Delta E_{\mathrm{AB}}^{\mathrm{CP}} = E_{\mathrm{AB}}^{\{\mathrm{AB}\}} - E_{\mathrm{A}}^{\{\mathrm{AB}\}} - E_{\mathrm{B}}^{\{\mathrm{AB}\}} \tag{3.2}$$

where {AB} signifies that the combined basis is used. For variational methods, $\mathrm{E}_{\mathrm{i}}^{\mathrm{AB}}$ is expected to be less than the uncorrected energy of the monomer. Most non-variational methods are also expected to give lower values than the uncorrected energy. Since nothing is altered with the dimer energy compared to the uncorrected calculation, the interaction energy obtained from the CP-scheme is expected to be less negative than the uncorrected interaction energy. For most situations the true interaction energy will be sandwiched by the uncorrected and the CP corrected interaction energy. It is not possible to know where in the interval the true interaction energy lays. The scale of typical interaction energies is relatively small compared to other forces, the effects of incomplete basis sets can play a major role in the accuracy of the interaction energy. The calculations also rely on a large amount of cancellations of errors.[31,44].

The CP-scheme has been widely debated, mostly about whether or not it overcorrects the BSSE. This has resulted in a range of investigations into the validity of the method.[2,3,30,88–93] Due to favourable cancellation of errors, the uncorrected scheme sometimes seemingly tends to produce more accurate answers, as the error due to the incompleteness in the basis set often is of the same magnitude and different sign than the BSSE.[91] Many critics of the CP-scheme thus propose to increase the size of the basis set, instead of including the CP correction, as they claim this will reduce both the BSSE and BSIE.[89,91] The article by Alvarez-Idaboy and Galano[91] and references therein come with examples where the CP-scheme is maybe not the best approach. The CP-scheme do sometimes overcorrect the interaction, especially for atomic dimers as the basis set is optimised for a free atom calculation.[5] Also the CP-scheme does not correct for BSIE.[86] Mentel and Baerends[92] performed an investigation on the Be dimer, confirming the overcorrection of the CP scheme on atomic dimers.

It has also been suggested to use the average between the uncorrected and the counterpoise corrected values, as the true energy is said to be sandwiched between these.[85] Burns et al.[85] find that overall, both the CP and the average performs better then the uncorrected values. The use of an extrapolation scheme or the composite scheme with the CP-scheme has also been tested. The CP-scheme is very intuitive and easy to use, and it has become widely accepted as a sufficient method. In the article by Van Duijneveldt et al.[3], they give a justification of the use of the CP-scheme, giving it further validity. Combined with a sufficiently correlated method, an appropriate basis set and an appropriate extrapolation[5,32,33,94,95]/correction(composite scheme)[5,80,82,85]/addition of midbond functions[30,94,96–98] the CP-scheme is assumed to give highly accurate interaction energies. The performance of these approaches is further investigated in Appendix B, for interested readers. This as they are widely used for the CP-scheme, but not tested enough/not appropriate to used with the SNOOP-scheme.

3.2 Same Number of Optimized Parameters Scheme

The uncorrected and counterpoise corrected interaction energies are expected to be too high and too low respectively, Kristensen et al.[31] thus proposed that a more balanced approach would be to use the same number of optimized parameters (SNOOP) in the calculations of the subsystems and in the AB calculation. This is not the case when using the CP-scheme or the uncorrected scheme. The proposed SNOOP-scheme uses only the virtual orbitals of the ghost subsystem for calculating the energy of the other subsystem. In the CP-scheme both the virtual and the occupied ghost functions are included, while neither are included when using the uncorrected scheme. The number of parameters used in the SNOOP-scheme will thus be less than that used by the CP-scheme and more than what is used by the uncorrected method[31]. The SNOOP-interaction energy can be expressed as:[31]

$$\Delta E_{\mathrm{AB}}^{\mathrm{SNOOP}} = E_{\mathrm{AB}} - E_{\mathrm{A}}^{\mathrm{SNOOP}} - E_{\mathrm{B}}^{\mathrm{SNOOP}} \tag{3.3}$$

where the optimised individual energies are calculated using the optimised occupied orbitals for the system of interest along with the virtual orbitals of both the system of interest and the phantom system.

The orbital space of the uncorrected calculations is a subset of the orbitals in the SNOOP calculations, which is a subset of the CP calculations. This implies that the variational HF energy of the monomers is lowest for the uncorrected scheme, and highest for the CP-scheme. Again, this is assumed to also hold for the non-variational coupled-cluster methods. Hence, the SNOOP interaction energy should come between the uncorrected and the counterpoise corrected interaction energies, and thus hopefully closer to the true interaction energy.[31] The SNOOP orbital space being a subset of the CP orbital space the SNOOP calculations are assumed to be less computationally expensive. The scaling of the SNOOP-scheme is still at the level of the method, which is N^5 for the MP2 method and N^6 for the CCSD method.[31,99]

The SNOOP-scheme is a relatively new scheme, and has not been as rigorously tested as the CP-scheme. The author of this investigation tested the SNOOP scheme and the CP-scheme against each other in a previous project and looked at the effect of using midbond functions for finding the interaction energy between seven small dimers. The interaction energy between the seven dimers were found using aug-cc-pVDZ to aug-cc-pV5Z basis set with and without midbond functions for both the SNOOP and the CP scheme. The conclusion of that investigation was that the CP-scheme with midbond functions produced the most consistent values, with consistently low relative errors for all basis sets. The SNOOP-scheme without midbond functions produced values with similarly low mean errors. Including midbond functions to the SNOOP-scheme did not prove advantageous, and rather resulted in the interaction energies becoming too negative compared to the other methods. No significant speed-up was observed under that investigation, but the systems under consideration were also very small. This scheme should be further tested, especially against the test sets, to test the performance against the CP-scheme for other larger systems.

Rasmussen et al.[86] have tested the SNOOP-scheme further, and also implemented it with the explicitly correlated RIMP2-F12 method. They tested it against the S22 test set of Jurečka et al.[21], and found that the SNOOP-scheme at the MP2-F12/aug-cc-pVDZ level gave results at the CP/aug-cc-pVQZ level, at the cost of an aug-cc-pVTZ level method. Looking at Table V in the article by Rasmussen et al.[86], the standard deviation of the CP values is lower than for the SNOOP values. The mean and maximum absolute errors are lower for the SNOOP-scheme, than the CP or uncorrected. This is true for both the explicit methods and the normal methods. The SNOOP-F12 method provides very accurate results, compared to the extrapolated reference values. It is also interesting to see the better performance of the SNOOP-scheme even with the small aug-cc-pVDZ basis. This is promising for this investigation, where this basis set is used. The SNOOP-F12 code was not production ready, and was thus not used in this investigation.

There are also other methods that can be used to find interaction energies. Some build on the CP-scheme[100–102], while others are based on different approaches[10,11]. The simplest and most widely used is the CP-scheme, which will be the focus of this investigation along with the SNOOP-scheme. Among the other methods, finding the interaction energy using a symmetry adapted perturbation theory (SAPT) is somewhat popular. The interaction can here be broken into its constituents, and it is therefore also a popular method to use in classifying which type of interaction is present. Both wave function and DFT[12–16] based methods exist. The DFT based methods are much less computationally expensive, and is thus usually favoured, but will not be further discussed in this thesis.

3.3 Basis Sets

As mentioned above, there will be some error in the calculations due to the incompleteness of the basis set used. For benchmark calculations, Řezáč and Hobza[5] claim that this is the main source of error for practical purposes. For the wave function used in the methods mentioned in Chapter 2 to be of good quality, it is important to choose the right basis set, and to include a sufficient amount of basis functions to properly describe the phenomena of interest. The wave functions are constructed from molecular orbitals, which again are constructed as a linear combination of a finite set of simple analytical atomic basis functions as in Equation (2.3). Many different basis sets with different strengths and weaknesses have been designed and are used for different types of problems. Finding a basis set suitable for all problems has proved to be challenging.[44]

Helgaker et al.[44] gives three requirements that an ideal set of basis functions should fulfil. These requirements are that the basis functions must allow for a systematic extension towards completeness, they should allow rapid convergence to any state and they should have a form that is easy to manipulate, making the integrals easy to evaluate. For real calculations it has been difficult to find functions fulfilling these three requirements. Jensen[42] also states that a suitable basis function must go along with the physics of the problem, (i.e. as the distance between the electrons and nuclei increases the basis should be able to describe this). In theory, any type of basis function can be used.[42] The criteria point first towards the use of an exponential function as this will describe the electronic cusp behaviour close to the nucleus very well. These are complicated to do calculations with and thus Gaussian functions, which have a square in the exponential, are used for most practical calculations. These are worse at describing the electronic structure on a one-to-one basis. Exponential functions without the square become very time consuming for calculations involving 3-4 centre two-electron integrals, while the Gaussians can more easily be added together.

For calculations where electron correlations play an important role, the requirements of the basis set become slightly different and more demanding. In these calculations the basis functions must also be able to describe the virtual orbital space in a sufficient manner, recovering as much as possible of the correlation energy. For these calculations to be made simpler, it is customary to distinguish between core and valence electrons, and calculate the correlation energy only for the valence electrons. This can be physically justified by noting that the core electrons are mostly unaffected by molecular rearrangements and many physical perturbations.[44] Most calculations found in the literature are performed using the frozen core approximation.

As mentioned above, one important criteria is that the basis set has the ability to systematically recover more and more of the correlation energy and move closer to the exact energy, as more basis functions are used, given that the method used for the calculations is good. Computationally, including more functions will make the calculations more costly. The ideal basis set must thus converge fast to the exact energy without increasing the computational cost too much. In practice, one must often do a trade off between computational cost and accuracy of the calculations when choosing a basis set. This can be illustrated by using an example provided by Klopper and Samson[60], who claim that "[...]the computation time of a correlated electronic-structure calculation in a correlation-consistent cc-pVXZ basis grows as X^{12} with its cardinal number X, while basis-set truncation errors only disappear as X^{-3}"[60]. Illustrating the great increase in computational time, and low accuracy observed when increasing the basis set.

An extrapolation scheme[32–36] is sometimes used to extrapolate the values of a smaller basis set to a higher accuracy. Here the two-point extrapolation of Helgaker et. al.[32,33] is both easy to use and provide good results[5,81]. It is also possible to include small functions at the midpoint between the molecules to try to improve the description of the interaction between two molecules, hopefully giving better accuracy at a lower computational cost. This approach is not widely used, even if the ones that do report good results.[30,94,96–98] The use of midbond functions was investigated by this author in the previously mentioned study on small dimers, showing very promising results.

When selecting a basis set size two different contributions must converge. Both the HF contribution and the contribution from the correlation energy must be converged. The HF part must be large enough to correctly describe the molecular orbitals and the changes these undergo from noncovalent interactions. This part converges relatively fast, and a triple-ζ basis set size is often sufficient. The correlation energy converges slower, and a basis set size of at least quadruple-ζ must be used. For the correlation energy, it is also important to use a method capable of including correlation energy.[5]

In this investigation, the Dunning's[103] correlation consistent basis functions, or sets based on these, have been used. This family of functions is a widely used set for interaction energy calculations. The focus of the next section will thus be on these, and then some comments on other basis functions will follow.

3.3.1 The Correlation Consistent Basis Functions

In an article published in 1988 Dunning proposed the use of correlation consistent basis set functions[103]. These functions, as many others, are based on the Gaussian type orbitals. There are many different approaches to creating atomic orbitals of primitive Gaussian functions. The Dunning orbitals are constructed by a procedure that adjusted the exponents of the correlated orbitals so as to give maximum contribution to the correlation energy. With this procedure each set of correlated orbitals will contain few primitive functions and be quite compact. In his article, Dunning[103] found that these functions effectively gave values very close to results obtained from other less compact functions used at the time. Dunning's basis functions also provide a way of systematically including more basis functions, giving more accurate calculations as one moves through the set. The X in the abbreviation of the basis set name is known as the cardinal number, and it is this that when altered increases the amount of basis functions. These functions are designed to converge towards the basis set limit. Which basis functions that are included in the correlation-consistent polarised basis sets for the non-augmented cc-pVXZ basis functions is given in Table 3.1 (Table 8.11, p 310[44])

Table 3.1: Structure of the cc-pVXZ basis sets

Basis set	H-He		B-Ne		Al-Ar	
cc-pVDZ	[2s1p]	5	[3s2p1d]	14	[4s3p1d]	18
cc-pVTZ	[3s2p1d]	14	[4s3p2d1f]	30	[5s4p2d1f]	34
cc-pVQZ	[4s3p2d1f]	30	[5s4p3d2f1g]	55	[6s5p3d2f1g]	59
cc-pV5Z	[5s4p3d2f1g]	55	[6s5p4d3f2g1h]	91	[7s6p4d3f2g1h]	95

These functions were created to accurately calculate the valence-correlated wave functions of ground state neutral systems[44]. For calculations involving excited states, ions or other systems with diffuse electron distribution, the cc-pVXZ functions are too inflexible. The aug-cc-pVXZ basis functions[104] are constructed very similar to the cc-pVXZ basis functions, the aug being for augmented. The augmented version consists of more diffuse functions, describing better the more diffuse electrons. The difference can be illustrated by looking at the B-Ne function for cc-pVDZ. This basis set contains 14 functions in Table 3.1, and 23 for the aug-cc-pVDZ set. The aug-cc-pVDZ set contains 1s, 1p and 1d function extra. Going up to aug-cc-pVTZ, this set contains 1s, 1p, 1d and 1f function extra, compared to cc-pVTZ[44]. The number of functions included in the set for the first row atoms can be calculated as:[44]

$$N^1_{\text{aug}}(X) = N^1_V(X) + (X+1)^2 \tag{3.4}$$

where X is the number represented by the X in aug-cc-pVXZ, and N^1_v is the number of functions included for the cc-pVXZ set, for the first row atoms. ($\mathrm{N}^1_v = \frac{1}{3}(\mathrm{X}+1)(\mathrm{X}+\frac{3}{2})(\mathrm{X}+2)$) More functions most often provide a more accurate calculation. It also means a more expensive calculation. Finding the balance between these can be difficult.

Over the years these functions have been further studied and expanded. Doubly and triply augmented basis sets (d-aug-cc-pVXZ and t-aug-cc-pVXZ) now exists. Another version are the heavily augmented basis function, where all atoms except hydrogen are augmented. The omission of the augmentation on the hydrogen atoms will only have a small impact on the total calculation, and is therefore dropped in the heavy-aug-cc-pVXZ[105] basis set.[5] Řezáč and Hobza[5] reports that these functions give results close to, and even sometimes better than the fully augmented functions, at a less expensive calculation. Which is why these are used for the interaction energy calculations in this investigation. For more accurate results calculations correlating all electrons can be performed, for such calculations the polarised-core (cc-pCVXZ)[106] or weighted polarised-core (cc-pwCVXZ)[107] basis functions have been developed.

When the original cc-pVXZ functions were used, together with an extrapolation scheme, to find the dissociation limit of a molecule, it was discovered that they were not tight enough for the second row elements, and exhibited unacceptable errors.[108] Dunning et al.[108] discovered that the reason for this error was due to two inter-related problems. Which were a near duplication of the exponents in two of the d sets and a lack of high-exponent functions in the early members of the sets.[108] They report of a similar error for the f set, but this gives only a minor contribution when finding the dissociation limit. Simply increasing the exponent of the d functions was suggested, but this would according to Dunning et al.[108] not completely fix the problem as the valence space of the 3d and 4d functions were almost the same. The solution was to create the cc-pV(X+d)Z basis set, where a high exponent is added to the 1d and 2d sets of the cc-pVDZ and cc-pVTZ sets, eliminate the 3d function, as this is so similar to the 4d function, and then use 4d for the cc-pVQZ set, 5d for the cc-pV5Z basis sets. An augmented version of the cc-pV(X+d)Z set is also available. Both the cc-pV(X+d)Z and the aug-cc-pV(X+d)Z sets showed improved dissociation energies for the SO, O_2 and S_2, tested in the article by Dunning et al.[108]

3.3.2 Other Basis Functions

The 6-31G set of Pople[109] was introduced as a small and effective basis set. It can be extended by including polarised functions, (d (and p) functions), to the 6-31G*(*) basis set, and diffuse s and p functions to all non hydrogens with the 6-31G+ basis set. This basis set is not complete, but for ab initio calculations performed on large systems, it is sometime the only viable option. Řezáč and Hobza[5] reports that using the 6-31G$*$(0.25) and 6- 31G**(0.25,0.15) for a CCSD(T) calculation that is meant to be a correction to the MP2 calculation using a larger basis set, gives quite good results for systems that are too large for the CCSD(T) method with the aug-cc-pVDZ basis set. Where the values in parenthesis are the exponents of C, N, O and hydrogen respectively.[5]

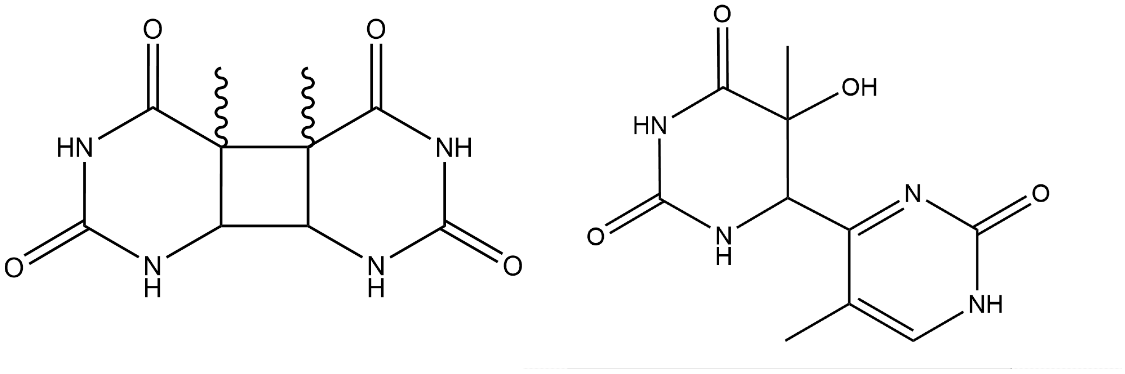
Helgaker et al.[76] state that the Dunning basis sets might not be optimal to use with the explicitly correlated methods, and basis sets especially optimised for these methods should be used. These methods need an appropriate set of auxiliary basis functions to avoid unwanted error.[60] Different sets of basis functions have thus been proposed, often these build on existing basis functions, but with the auxiliary basis functions optimised for the different approximations.[60,62,110–114] Peterson et al.[113] claim that their new cc-pVDZ-F12 basis set is able to recover 99% of the valence MP2 correlation energy for first row elements. With the inclusion of complementary auxiliary basis set, along with the RI approximation and other advances, the F12 methods can be used successfully with a small basis set, on relatively large systems.[47]

There are many other available basis sets that could be more or less suited for the task at hand. The Dunning basis sets are quite popular in the literature, and they provide both a systematic improvement and they are capable of describing electron correlation. They are also widely implemented into the different computational programs available and well tested. The thymine system of interest is a relatively large system. The optimum would be to be able to use the aug-cc-pV5Z basis, but that is prohibitively large for practical calculations on such a system. The heavy-aug-cc-pVDZ basis has thus been chosen to make the interaction energy calculations quite manageable.

4 Thymine

4.1 Short on Thymine

In the 1940s it was discovered that deoxyribonucleic acid (DNA) is the substance that contains our genes and genetic material. DNA consists of two strands of nucleotides hydrogen bonded together in a double helix shape. Each nucleotide consists of base, sugar and phosphate. There are four different nucleotides, where the difference is the base component. The four different bases are adenine (A), cytosine (C), guanine (G) and thymine (T), in RNA thymine is replaced by uracil (U). Uracil is very similar to thymine, the only difference being the methyl group on thymine. Due to complementarity of shapes adenine only binds to thymine and cytosine only binds to guanine. These bases then stack on top of each other to make up DNA. The stacking order tells the cell what it should do, and in the double helix all our genetic information is contained, using only four different bases. It is very important that the order stay intact and unaltered. When DNA is exposed to UV radiation the areas of the DNA consisting of two consecutive thymine bases are especially at risk of deformation by forming one or two covalent bonds between the two, causing great harm to the DNA.[115] Where electromagnetic radiation in the UV-B (290-320 nm) and UV-C (100-290 nm) is said to have mutagenic effect. The maximum absorption of DNA is around 260 nm, which is in the UV-C spectrum.[116] In DNA thymine can either form a cyclobutan dimer, as seen in Figure 4.1a, or a (6-4) photoproduct, as seen in Figure 4.1b. In this investigation Figure 4.1a will be of most interest, as this is the form the experimentalists are after.



(a) The thymine cyclobutan dimer

(b) The (6-4) photoproduct of thymine

Figure 4.1: The two different photoproducts of thymine in DNA

The thymine photodimer was discovered almost 60 years ago, and was the first environmentally induced lesion to a DNA component discovered.[117] Later it was shown that this also happen in cells.[118] It is possible to form six diastereomers from the [2+2] cycloaddition of thymine. In DNA, due to steric constraints, only the syn isomers can be generated.[119] Dimerisation is also possible between the other pyrimidine bases in DNA and RNA (C-C, U-U, C-T), but in experiments the thymine dimer appear to be favourable.[120] Other reactions are also possible both with thymine and the other bases. The other reactions are limited in DNA, due to the fixed placement of the bases. DNA is said to be relatively flexible, as seen by being flexible enough for dimerisation to happen, something that introduces a significant bending.[121] In this investigation the cyclobutane photoproduct of thymine will be in focus, but it is also of interest to see if other reactions are more favourable in free thymine. Since the experimentalist are looking at free thymine on a surface, it will be of interest to investigate the interactions present between two thymine molecules at different geometries relative to each other. Hopefully this information can indicate something about the likelihood of the thymine molecules obtaining the conformation needed for the product of interest.

Beukers et al.[117] report that the level of thymine dimers in a solution that has been irradiated is low, partly due to the fact that the process is reversible. The position of the equilibrium depends on the wavelength of the irradiated light, where long wavelengths favour the dimer.[117] There seems to be some debates about how the mechanism for cyclodimerisation happens, whether it happens through a singlet mechanism or a triplet mechanism. First an introduction to general photochemical reactions will come, then there will be a discussion about what others have found about the mechanism leading to both photoproducts.

A fun fact is that some species that live in high altitudes, and are exposed to high levels of strong UV radiation, have evolved DNA that contains a lower proportion of thymine residues than lower living species. Thus making them less prone to take damage from the high level of UV rays that they are exposed to.[122] Also other species have developed special enzymes to try to open up the potential deadly cyclobutane ring between two thymine molecules.[123]

4.2 Introduction to Photochemical Reactions

In DNA a photochemical reaction can produce both a cyclobutan ring between two thymines, and a (6-4) photoproduct via an oxetane ring. Luckily for us, these reactions do not occur easily in our DNA, even when exposed to sunlight. It does happen often enough for this reaction to be of interest in this investigation. No calculations will be performed on excited states or transition states related to the reaction, but having a sound understanding of the underlying reaction of interest was deemed important enough to get a proper introduction. Also others have performed these calculations, and to be able to discuss these in a proper manner, the same understanding is necessary. The coming section introduces some of the chemistry behind a photochemical reaction, and aims at containing information that can be valuable to the physicists.

Beukers et al.[117] define a photochemical reaction as "A reaction between two molecules, that normally do not react, may take place under the influence of light (photochemical reaction)." The position of the two reactants are also important, as the correct overlap of the orbitals is important for the reaction to happen.[122]

Creating an orbital correlation diagram as done by Anslyn and Dougherty[122] on page 880, Figure 15.1, and following the principles of conservation of orbital symmetry, by Woodward and Hoffmann, it is clear that the thermal cycloaddition of two ethene molecules, in their ground state, will result in the two electron excited cyclobutane with the excited electrons in the antibonding σ_3^* orbital. Using the frontier orbital theory to construct the HOMO and LUMO of ethene, as in Figure 4.2, also results in this being an unfavourable reaction, as the HOMO and LUMO do not have correct overlapping symmetry. The thermal cycloaddition reaction of a $[_\pi 2_s + _\pi 2_s]$ system, as seen in Figure 4.2, is thus deemed very unlikely. Altering the orientation of one ethene molecule and considering a $[_\pi 2_s + _\pi 2_a]$ reaction, (where the double bonds are at right angles to each-other almost creating a cross,) the cycloaddition of ethene will be an allowed reaction. It is still not a very likely reaction.

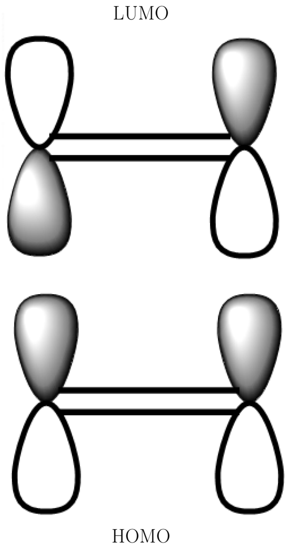


Figure 4.2: The HOMO and LUMO of ethene

Shifting to look at photochemical reactions, Anslyn and Dougherty[122] reports that very few reactions happen through a concerted mechanism, as many thermal cyclisation reactions do. Further they claim that most photochemical reactions involve triplet states and biracial intermediates. Cycloaddition reaction happening via a singlet state is also possible, but then it is claimed that formation of an exciplex (electronically excited complex) is usually involved. It is stated that the reason many reactions are termed thermally forbidden is due to a high thermal reaction barrier. This barrier often comes close in energy to the first excited state, a favourable condition for funnel formation and a photochemical reaction. The geometry of the reactants is important in whether or not the funnel will produce a product or fall back to the initial state. A small gap between the S_0 and S_1 surfaces is desired for funnel processes to happen.[122]

Anslyn and Dougherty[122] categorise [2+2] cycloadditions into the following three classes: addition of two olefins to form cyclobutane, addition of an olefin to a carbonyl to form oxetane and reaction between an α, β-unsaturated carbonyl and an olefin to form a cyclobutane.

The dimerisation of two olefins usually proceeds via triplet states, which often have a relatively high energy. Intersystem crossing can be very slow for hydrocarbon π, π^* states, and the use of a high-energy sensitiser is often recommended. Controlling the stereochemistry for such a reaction can be difficult, supporting the claim of a stepwise mechanism.[122]

Reaction of an olefin and a carbonyl to produce an oxetane, a four membered ring containing oxygen, can occur either form a singlet or triplet n, π^* state of the carbonyl. This is also a stepwise process, going through a biradical intermediate. Also here is control of stereochemistry difficult, but the stereochemistry is more retained in the singlet manifold.[122] The carbonyl group is in general a dominant chromophore in organic photochemistry. The n, π^* is readily accessible, and the fluorescence rates are relatively slow allowing time for a reaction to happen.

Reaction between an α, β-unsaturated carbonyl and an olefin to form a cyclobutane is somewhat more complicated. They can proceed through both the n, π^* or the π, π^* states of the carbonyl. These two states are often close in energy for enones, and a detailed analysis of these can thus be complicated. The overall reaction is quite efficient, with electron rich olefins speeding up the reaction further.

If formation of an exiplex/excimer (excitet dimer) is involved in the reaction mechanism, this can be detected using spectroscopy. The absorption band associated with the excited complex will then be shifted, often to longer wavelengths, compared to the original complex. The shift will also depend on the concentration of the species involved, something that will be most relevant for heterodimers.

For systems containing π-systems, π donor-acceptor states can occur. Here one system excites an electron to an empty orbital of a nearby molecule, creating a charge transfer complex. Still no formal bonds are created, and both systems can easily go back to their initial state. A solution of charge transfer systems will often be coloured, and also here spectroscopy can be used to determine the existence of such states. Both charge transfer states and excimers are interesting phenomena that are easy to verify experimentally, and that can be helpful in explaining a mechanism. These sates are also somewhat complicated to describe theoretically.

4.3 Mechanistic Insight from Literature

Many (both theoretical and experimental) studies, have been interested in the thymine dimerisation process, relevance to DNA have been very important in many of these. Some thus enforce constraints when performing geometry optimisations, as the DNA structure is not the most stable form for two thymine molecules, also seen in this investigation. Finding the correct reaction mechanism for the dimerisation process has also been of interest. Durbeej and Eriksson[120] performed an investigation of why the thymine dimer is favoured over the dimers of cytosine, uracil and a mix of these. This study is by now somewhat old, and the methods used are not the best suited for performing calculations on excited states. At the time of their investigation multireference wave function methods were not able to perform calculations on such a large system. The investigation by Durbeej and Eriksson[120] attempted to investigate whether the ground state and first excited state potential energy surfaces could be used to explain the difference in occurrence between the dimers in DNA. They found that the incentive for creating the dimer was much larger in thymine than in cytosine and the mix between the two. Uracil had a similarly favourable path to dimerisation.

Performing a similar study, Zhang and Eriksson[124] investigated the possibility of a triplet mechanism in cyclobutan thymine dimer formation. With the B3LYP functional, a triplet mechanism could not be ruled out. A reaction going through the triplet mechanism is assumed to take longer than a singlet mechanism - a claim supported by Boggio-Pasqua et al.[125].

Serrano-Pérez et al.[126] performed a CASPT2(12,12) with the ANO-S basis set, using the contraction scheme C,N,O[3s2p1d]/H[2s1p] for the thymine$\cdots$thymine and cytosine$\cdots$cytosine dimers, and found similar evidence as Durbeej and Eriksson[120]; that the thymine dimers have more incentives for formation, while cytosine more easily end up at the ground state.

A study by Boggio-Pasqua et al.[125] found using CASSCF with the 6-31G* basis set, that the thymine dimerisation proceeds through a barrierless concerted mechanism. They calculated the electronic states of the S_0 and S_1 states of thymine, and confirmed that the thermal reaction has a substantial reaction barrier. Using geometry constraints to mimic DNA, they predict that a reaction will happen if the thymines are at a configuration near the low lying conical intersection, S_0/S_1CI geometry. A study by Blancafort and Migani[127] supports the notion of the presence of a canonical intersection. Not very surprising, since this study is performed using the CASSCF(12,12)/6-31G* level of theory.

Schreier et al.[128] experimentally investigated the dimerisation of an all thymine $(dT)_{18}$ stand. They found that the cyclobutan dimer is fully formed after ~1 picosecond, giving evidence to the stated barrierless reaction pathway. The speed also indicates that the conformation must be suitable for dimerisation before photon absorption, as there is no time for conformational change. The low quantum yield in experiments suggests that this is not a very common conformation. Pilles et al.[129] also did an experimental study on the all-thymine single stand $(dT)_{18}$, where they used time resolved IR spectroscopy to study the 100 picosecond dynamics of the strand under the influence of UV light. No lesions or dimer formation were observed, but charge transfer states were found at quantum yields of ~0.07. According to Pilles et al.[129] this was the first experimental observation of CT states between two thymine bases. From the study by Serrano-Pérez et al.[126], no stable structure was found that had a S_1 below the conical intersection, something that was said to favour the products, and not the formation of excimers.

Blancafort and Migani[127] claim that the quantum yield of the thymine cyclobutan dimer in DNA is about 2-3%, and that the yield for the (6-4) photoproduct is about one order of magnitude less. The study by Schreier et al.[128] state that the yield of the (6-4) photoproduct in the $(dT)_{18}$ strand is so small that it is almost negligible. It is still of interest in this investigation, as for the free molecules it might be more reactive.

Giussani et al.[130] performed an unconstricted CASPT2//CASSCF with the atomic natural orbital (ANO) of S-type contracted to C,N,O[3s,2p,1d]/H[2s1p] to investigate the formation of the (6-4) thymine photoproduct. They report that the conformation needed to create the (6-4) photoproduct will most often lead back to the ground state, without dimerisation. Further they comment that this is different for the thymine dimer, where once it is on the right conformation and the right state, dimerisation is very likely to happen. They also find that the excitation has to be localised on a single thymine molecule, and that their usual response is to decay back to the ground state. They found that the triplet state is involved in the formation of the (6-4) photoproduct. The initial process might go through the singlet state, but Giussani et al.[130] found that the only barrierless path goes through the triplet state. A study by Yang et al.[131] is less certain in their conclusion, but they also claim to have evidence for a triplet mechanism.

Marguet and Markovitsi[132] performed an experimental study on the formation of both the cyclobutane dimer and the (6-4) photodimer. The mechanistic evidence was not completely clear from this investigation, but they found that the cyclobutane dimer is formed in less than 200 ns, and that the photoproduct is formed in 4 ms, through a reaction intermediate.

A joint experimental and theoretical study on dimerisation of thymine by Banyasz et al.[133] found that the triplet path only contributes about 10% in the formation of the cyclobutane dimer. They also claim that the two products results from two different excited states. They believe that the formation of the cyclobutane dimer proceeds through a barrierless path, involving the $^1\pi\pi^*$ state. For the (6-4) photoproduct, they claim that the formation of the intermediate oxetane ring proceeds through an excited charge transfers state, and that there is a reaction barrier present on this path. Banyasz et al.[133] claim that the $n\pi^*$ state is not involved in formation of the (6-4) photoproduct.

Something that is not so relevant for this investigation, but might be relevant for the experimental investigation, is that the (6-4) thymine photoproduct can by absorbing another photon be transformed to the Dewar structure. Structure of the Dewar product is given in Figure 4.3. The figure is included in case this will be the result of an experiment.

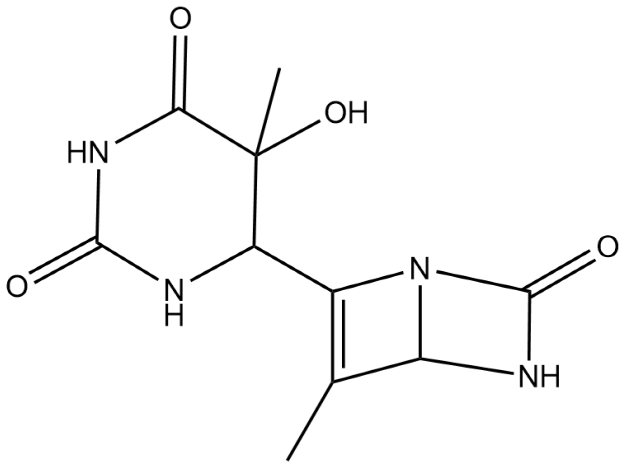
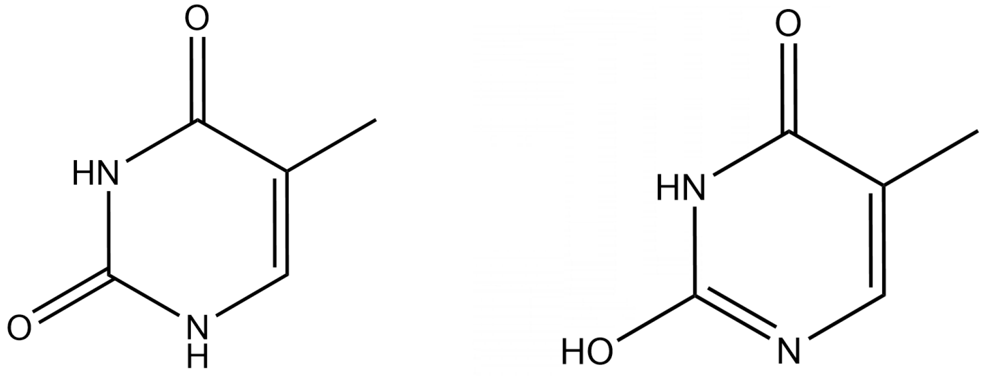
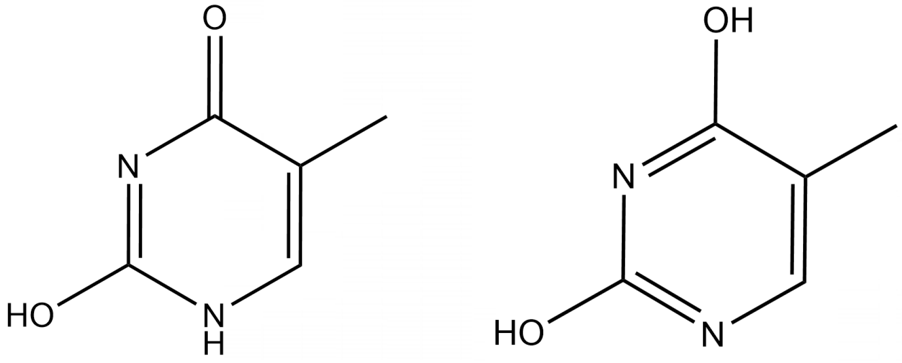


Figure 4.3: Dewar product of the (6-4) photoproduct of thymine

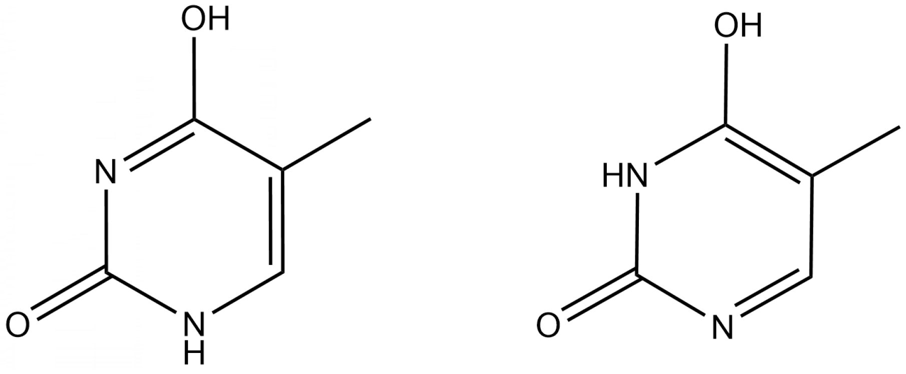
Ketones are often present in a keto-enol form, at least in solutions. The equilibrium point of this reaction depends on the pH of the solution. Where the equilibrium point for the keto-enol reaction will be under the experimental set up, is hard to tell. Knowing the structures and their existence can hopefully help during the experiments, and are therefore included here. The structures are drawn in ChemBioDraw Ultra, after Scheme 1 in the article by Morsy et al.[134]. They state that the most stable form in Figure 4.4, is the regular thymine, followed by Figure 4.4b. The study done by Morsy et al.[134] is somewhat old, and the theoretical methods used have seen further development since then. They also state that the calculations are sensitive to basis set size. The order of stabilisation is supported by Jiao et al.[135] and Piacenza and Grimme[136] for the neutral tautomers. Jiao et al.[135] also investigated the charged tautomers and found a change in stability for the anionic and cationic forms of the tautomers. Further investigation into the different tautomers should be considered, if these forms prove important in the experiments, especially as these studies use DFT based methods and/or are somewhat old. All figures in Figure 4.4 are tautomers of each other, but for simplicity, the regular thymine shape will be referred to as thymine, and not tautomer 1, and then the others are numbered from 1 to 5.



(a) Regular thymine (b) Tautomer 1



(c) Tautomer 2 (d) Tautomer 3



(e) Tautomer 4 (f) Tautomer 5

Figure 4.4: The keto-enol forms of thymine

5 Computational Details

In this investigation, the same number of optimized parameters (SNOOP) scheme was used with the heavy-aug-cc-pVDZ basis set to calculate the interaction energies. The thymine molecule was optimised using the CCSD(T) method with aug-cc-pVDZ basis. The geometry optimisation was performed with the Dalton package, and the interaction energies were performed with LSDalton.[137] The optimised geometry of thymine was used to create the input files for the stacked dimers using the Avogadro software[138]. For the other systems optimised in this investigation, the systems were optimised using Q-Chem 4.4[139] program with the MP2 method and the cc-pVDZ basis. The thymine tautomer was optimised using MP2 and the aug-cc-pVDZ basis, also in Q-Chem.

From a previous study done by the author, and the study by Rasmussen et al.[86] it is observed that the SNOOP-scheme produces more accurate values with the smaller basis set than the CP-scheme. (The mean error calculated for the small systems in the previous investigation is included in Figure 5.1 to illustrate this.) Including midbond functions to the CP-scheme will result in similar or better values for the aug-cc-pVDZ basis, but finding the optimal position for such functions for the stacked dimer was more complicated than using the SNOOP-scheme without midbond functions. As discussed in Section 3.2 the SNOOP-scheme should also be less computationally demanding. The aim of this is not to find benchmark grade interaction energies, but to help experimental physicist with their experiments. The aim is to get an overview over the interactions between two thymine molecules, to see whether the geometry favourable for making the cyclobutan dimer is likely to occur. Ideally some calculations using a larger basis set, or an explicitly correlated method could have been performed to get more accurate results, but for the overview wanted here, MP2 and the heavy-aug-cc-pVDZ basis is sufficient.

Core ionisation energies were performed in Dalton using coupled-cluster linear response theory at the CCSD level within a core-valence separated framwork.[78] The aug-cc-pVDZ basis was used for carbon and hydrogen and aug-cc-pCVDZ basis was used for oxygen and nitrogen.

All ball and stick figures are made using the Chimera Software[140], while the other molecules are drawn in ChemBioDraw. All graphs are made using MATLAB 14B[141]. All experimental graphs and results are produced by Jakob Vinje[142], stud.techn. at the Department of Physics, and used here with his permission.

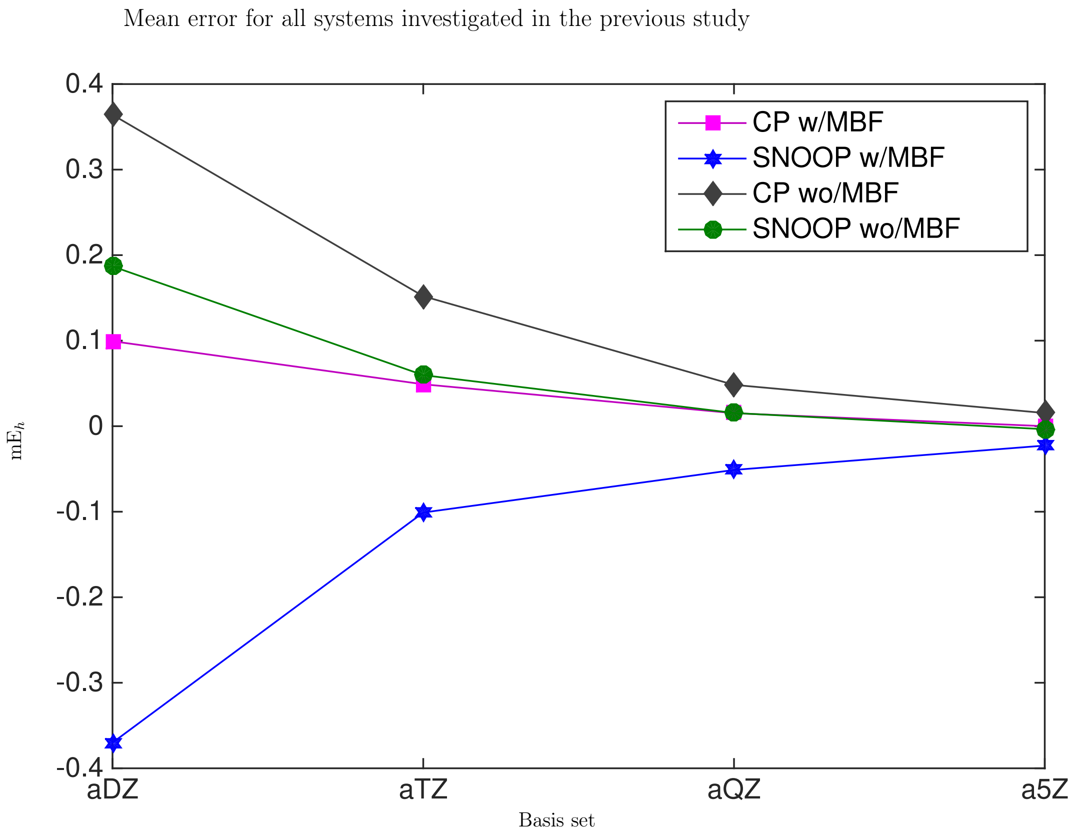


Figure 5.1: Mean error for all systems investigated in the previous study, all calculations were performed using MP2, (aXZ = aug-cc-pVXZ)

6 Results

In the coming sections the results form this investigation will be presented. First a preliminary investigation of the interactions of stacked thymine was performed, and presented in Section 6.1.1. Here, the thymine molecules were parallelly stacked on top of each other, and the top molecule rotated. Then in Section 6.1.2 three of the stacked geometries were further investigated by looking at the change in interaction energy with separation. These geometries along with a fourth geometry was then optimised and are presented in Section 6.1.3. The interaction energies obtained from the optimised geometries were then compared to the interaction energy of the parallel geometries. When placed on the MoS_2 surface the thymine molecules will be spread out on a surface. Four geometries in the plane were thus investigated and the results are presented in Section 6.1.4.

In Section 6.2 select experimental results are presented, and evaluated against the findings in this thesis. Interpreting the results were not straight forward, so to further support the experimentalists core ionisation energies were calculated and compared to the experimental core ionisation energies.

6.1 Interaction Energy Calculations

6.1.1 Parallely Stacked Thymines

In this section interaction energies for two thymine molecules stacked on top of each other will be computed, with eight different rotations relative to each other. The top thymine is rotated by 45° for each calculation, where the reference position is similar to how they will be stacked in DNA. The reason for the preliminary investigation of the parallely stacked thymines was to investigate which formation gave the largest interaction energy, and how it varied by rotation. The separation between the thymines was chosen based on values found in literature on DNA and other thymine studies[124,126,143–146]. The stacked conformations studied are depicted in Figure 6.1.

The interaction energies, as a function of rotation, are displayed in Figure 6.2. From this figure it is clear that the reference conformation has a small interaction at this separation, relative to the other conformations. This might be due to steric reasons, as the methyl groups are placed on top of each other in this arrangement. The conformation with the largest interaction energy is the one twisted 180°, which is also when the methyl groups are furthest away from each other. The thymine ring is not symmetric, but the interaction energies depicted in Figure 6.2 are nearly symmetric around 180° of rotation. Almost identical interactions are found for both the 90° and the 270° geometries, and also for

(a) Reference position (0 degrees rotation)

(b) 45 degrees rotation

(c) 90 degrees rotation

(d) 135 degrees rotation

(e) 180 degrees rotation

(f) 225 degrees rotation

(g) 270 degrees rotation

(h) 315 degrees rotation

Figure 6.1: Molecular structures used for the calculations of stacked thymine. The top thymine is rotated by X into the plane, relative to the reference position, Figure 6.1a

the 135° and the 225° geometries. The interaction energy points are connected by a line mostly for illustrative purposes. There could be outlier at other angles of rotation. It is still an interesting trend that is somewhat surprising.

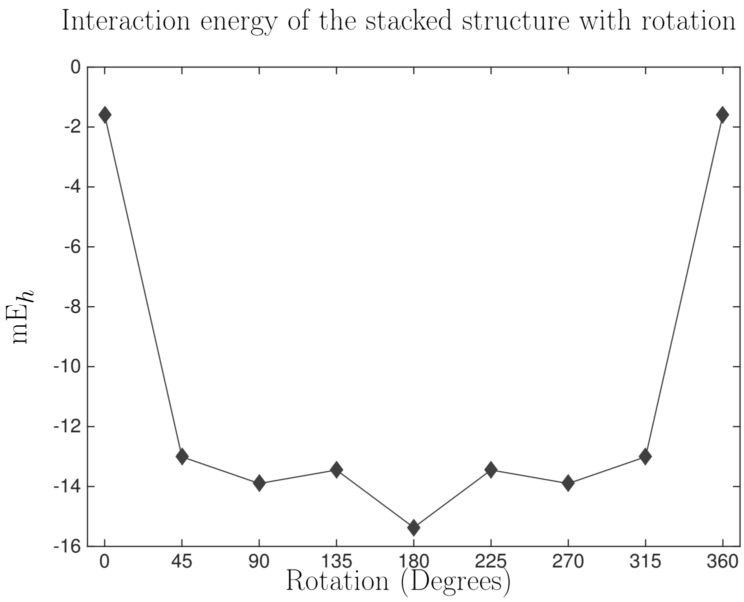


Figure 6.2: Interaction energies of the stacked thymine structures, as pictured in Figure 6.1

6.1.2 Interaction Energy of Selected Stacked Thymines

From Figure 6.1, three conformations (the reference geometry, the 90° rotated goemtry (Figure 6.1c) and the 180° rotated geometry (Figure 6.1e)) were chosen, and the separation between them varied, to investigate how the separation influences the interaction energy. The reference geometry was chosen to see whether a more favourable interaction could be found by varying the distance between the thymines. The other two were chosen as they had the largest and second largest interaction energy of the eight formations investigated from Figure 6.1. They were included to see if even larger interactions could be obtained at other distances. From Figure 6.3, it is clear that the separation influence the interaction energies in different ways. For both the 180° and 90° geometries the interaction is largest at 3.25 Å separation, while the reference geometry has its largest interaction energy at a separation of 3.75 Å. However, the interaction at this conformation is weaker than at the other two conformations. This indicated that for a parallel conformation as used here, the methyl group creates steric strains for the reference position. The 180° geometry experience the greatest effect of the separation. Since at a separation of 2.25Å, it experiences a greater repulsion than the 90° geometry (this point was later removed from the figure, so as to more clearly see the trend at more relevant interaction energies). For larger separations, the 180° geometry archives a larger interaction energy than the others, the 90° geometry still have very similar energy. The reason for this may be related to either or both the placement of the methyl group or the placements of the other atoms in the non symmetric thymine rings.

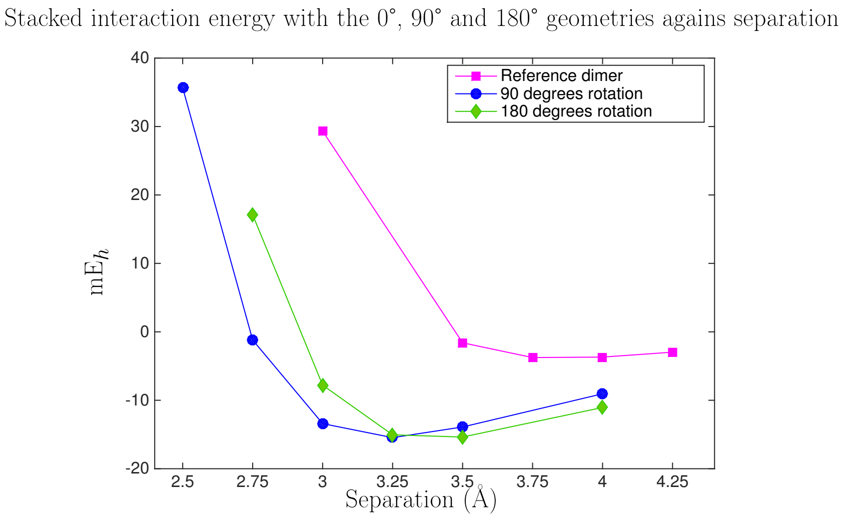


Figure 6.3: Interaction energy of the stacked dimer with the 0° (Figure 6.1a), 90° (Figure 6.1c) and 180° (Figure 6.1e) geometries

6.1.3 Optimised Geometries for Select Stacked Geometries

After the investigation in Section 6.1.2, the geometries of the three conformations (reference, 90° and 180° rotated dimers) investigated were optimised. The 270° geometry was also optimised. This to investigate whether the 90° and the 270° geometries would obtain more different interaction energies when optimised. To do the optimisation as computationally efficient as possible, the bottom molecule was kept fixed. It was of interest to see if the free molecule changed noticeably. The degree of change in the configuration was different for each conformation, as can be seen from Figure 6.4, where the optimised geometries are visualised.

The 180° geometry (Figure 6.4c), looks almost unchanged from Figure 6.1e, as the top thymine has not been tilted very noticeably. For the other three conformations the top thymine is visibly tilted. The area of the top thymine overlooking the methyl group of the bottom molecule is tilted upwards the most. Creating the largest separation between that area and the methyl group underneath.

Interaction energies were computed for the four molecules illustrated in Figure 6.4, and are plotted in Figure 6.5 together with the interaction energies of the unoptimised structures (from Section 6.1.1) for comparison. It is clear that the reference structure still is not very attractive. For the other conformations both the 90° and the 270° geometries now have a stronger interaction than the 180° geometry. The interaction energy of the 90° geometry and the 270° geometry now differ with 1 mE$_h$. This might indicate that the optimised forms might not create a curve as symmetric, as for the the parallel geometries. The theory that the methyl group creates steric constraints can be supported by the

(a) The reference position optimised (b) Optimised structure of 90 degrees rotation

(c) Optimised structure of 180 degrees rotation (d) Optimised structure of 270 degrees rotation

Figure 6.4: Optimised structures at the different rotations

optimised geometries. The 180° geometry, did not change noticeably from its constrained form in Figure 6.1e, whereas the others are visibly tilted. The top thymine seems to be tiled in a way that increase the separation of the methyl groups. Figure 6.4a is similar to the conformations observed in the literature[124,126,143–146].

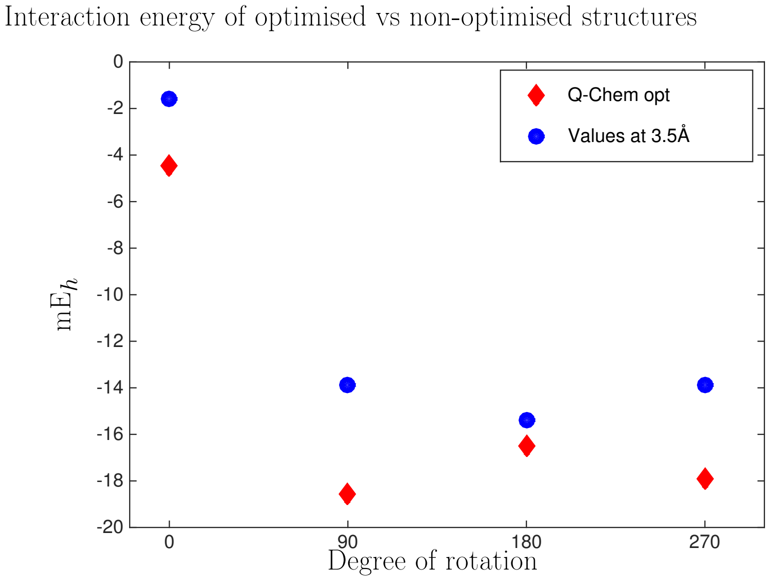
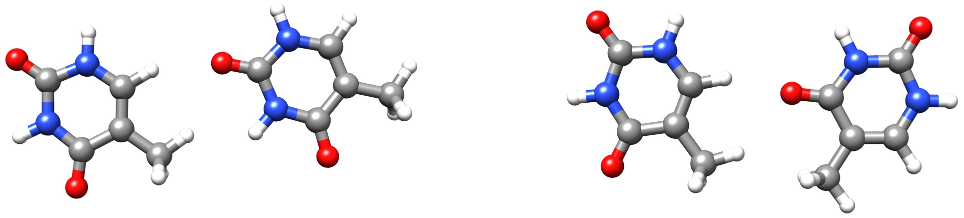


Figure 6.5: Interaction energy of the Q-Chem optimised structures v.s. the non optimised structure of Figure 6.1

6.1.4 Interaction Energies of Thymines in the Plane with Hydrogen Bonding

As stated earlier the experimentalists are performing their experiments with the thymines free to position themselves at optimum positions. In the experiments there where about three layers of thymine molecules spread out over a surface, meaning that there will both be interactions between the layers and between the thymines in each layer. In the coming subsection four different geometries where the thymine are flat and hydrogen bonded will be investigated closer.

Thymine has two carbonyl groups, and from Section 4.2 it is clear that a photochemical reaction between a carbonyl and a double bond is possible. Also, from the literature it is clear that in DNA a similar reaction happens, and create the (6-4) photoproduct. In DNA, the thymines are stacked. However, it would be interesting to see how strong the interactions are for the flat conformation, and also to see if there is a significant difference in the interactions between the two carbonyl groups with the double bond. This thesis has so far not numbered the atoms in the molecules, but from now on, the two carbonyl groups in thymine will be classified as containing O1 and O2, as explained in the caption of Figure 6.6a and 6.6b, this to increase readability. In DNA, the steric constraints will stop the thymines from being in a flat conformation, but the experimentalist will have the molecules flat on a surface, allowing the possibility of a reaction happening while flat. Considering the presence of the two methyl groups in the conformation in Figure 6.6b, there could be some steric constraints hindering a reaction. In this conformation, the thymine with the oxygen atom of interest, can be classified as an α,β-unsaturated carbonyl, which in Section 4.2 was said to be quite reactive. Furthermore, the presence of the nitrogens so close to oxygen 1 might also have an effect on its reactivity.



(a) Structure with the oxygen 1 pointing towards the double bond

(b) Structure with the oxygen 2 pointing towards the double bond

Figure 6.6: The two different optimised structures where O interacts with the double bond

The influence of moving both Figure 6.6a and 6.6b 0.25 and 0.5Å closer and further from the equilibrium position, was investigated and can be found in Figure 6.7. This was done to see how the interaction would change when changing the separation of the optimised structures. Note that the conformation in Figure 6.6b is more sensitive to being moved closer together than what Figure 6.6a is. It is likely that the methyl group is responsible for that. At the equilibrium point, the 6.6b conformation has a larger interaction than the other conformation. This can be due to the fact that one of the thymines now would be classified as an α, β-unsaturated carbonyl, which is quite reactive. Here, the presence of the methyl group does not seem to affect the interaction energy to the same degree as it has for the other conformations. Looking at Figure 6.7, the curve for Figure 6.6a (O1) has an unexpected behaviour at 0.25Å. The values at 0.25Å and 0.5Å are very similar, and this behaviour can possibly be due to the choice of method and/or basis set, or it can also be due to the choice of geometry optimisation.

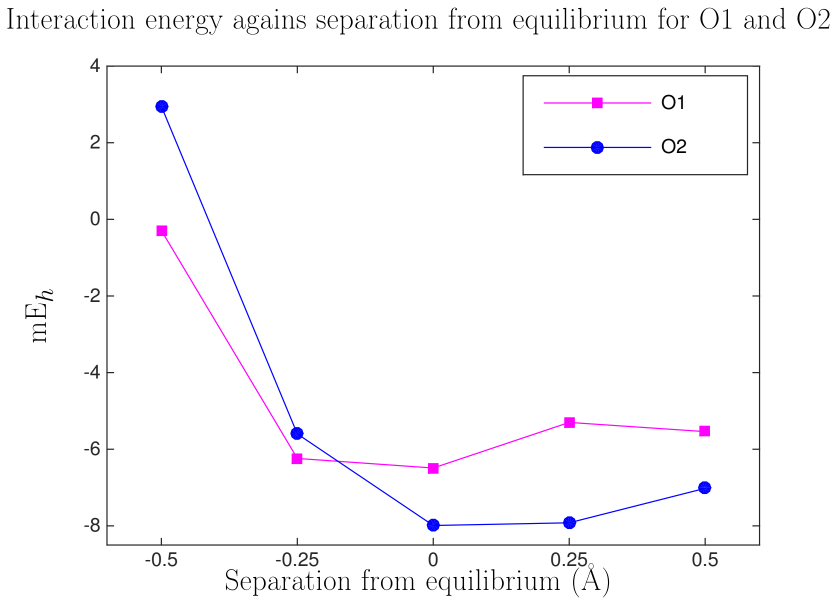


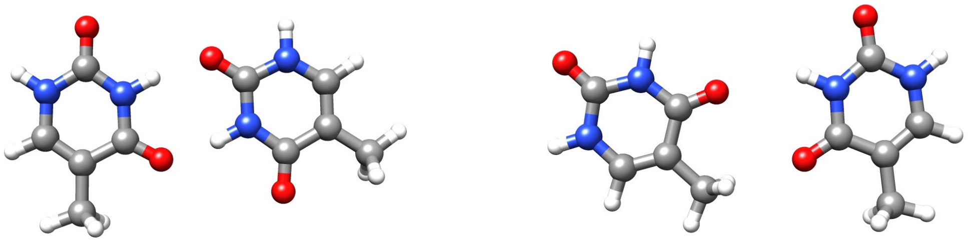
Figure 6.7: Interaction energy at different separation from equilibrium for Figure 6.6a (O1), and Figure 6.6b (O2)

Inspired by an article by Morgado et al.[147] on uracil, two other flat structures were examined. In the article, these structures are classified as hydrogen bonded, and they achieved the largest interaction energy of all the investigated dimers. The article set out to investigate the interaction of the uracil dimer at different non-equilibrium conformations, and to test different methods of achieving the energy. The other structures could also be interesting to test for the thymine dimer, but for now this investigation is mostly interested in finding the largest interactions.

In the article by Morgado et al.[147] the separations of the hydrogen bonded complexes are given, but not their optimised structures. The geometries of thymine were optimised using Q-Chem, with one molecule fixed. From the optimised structures in Figure 6.8 it is clear that the two geometries have differing separations. In Figure 6.8a the two thymines are much closer than in Figure 6.8b. This is due to the presence of the methyl group. From Table 6.1 the interaction energy of Figure 6.8b is very small at -5.26 mE_h, compared to the other geometries investigated here. For the similar structure of uracil the largest interaction is around -15 mE_h. The difference is very likely due to the presence of the methyl group creating steric strain. For Figure 6.8a the interaction is very strong, in fact it is the largest interaction energy found in this investigation. This should not be very surprising, as this structure is held together by two hydrogen bonds. Bonds that are quite strong compared to other noncovalent interactions. The findings do not give any indication about whether or not there will be any reactions here, but knowing that ketones are often found in a keto-enol equilibrium, interchange of hydrogens might be a possibility. It was also stated in Chapter 3 that hydrogen bonds sometimes are observed as charge transfer states that are almost covalently bound. Comparing the geometry in Figure 6.8a to Figure 4.4b, it is clear that the nitrogens involved in the hydrogen bonding are not the ones loosing a hydrogen in the most stable tautomer. Also, only one of the carbonyl groups involved in this bonding creates the alcohol group in Figure 4.4b.

Table 6.1: Interaction energy of the structures in Figure 6.8

Figure	Figure 6.8a	Figure 6.8b
Interaction energy mE_h	-19.12	-5.26



(a) Doubly hydrogen bonded structure (b) Singly hydrogen bonded structure

Figure 6.8: The optimised hydrogen bonding structures, inspired by Morgado et al.[147]

6.1.5 Other Geometries

The conformation with the double bonds at right angles to each other, as in Figure 6.9, was also investigated. As this is the initial conformation needed to make the reaction thermally allowed. This conformation was not optimised, due to the computational cost of this. However, the separation between the two molecules was varied using the Avogadro Software. All separations gave a repulsive interaction. Supporting that this is a very unlikely conformation, and that dimerisation from this conformation is very unlikely.

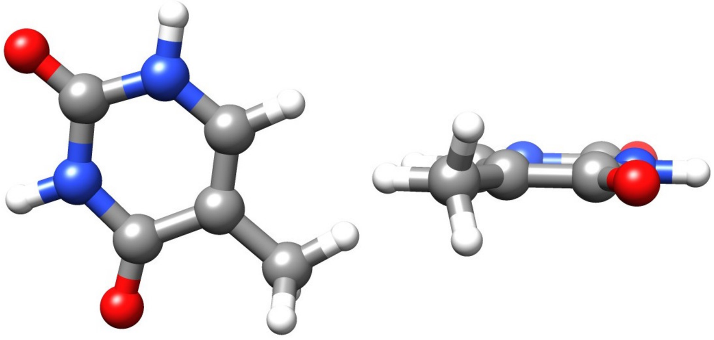


Figure 6.9: Approximate 90 degrees angle between the double bonds

6.1.6 Summary of the Interaction Energy Calculations

From the above results, it is clear that Figure 6.8a has the strongest interaction of the conformations studied here, closely followed by 6.4b. Except for the strongly hydrogen bound conformation in Figure 6.8a, the stacked conformation has the largest interaction energy. Figure 6.4a has the second lowest interaction energy in this study, only surpassed by the unoptimised Figure 6.1a. As Figure 6.9 gave only repulsive interactions it is not counted here. The results indicate that getting the thymine dimers to be in the right conformation for cyclobutan dimerisation might be somewhat challenging. It is not clear from the investigation which conformation is needed for dimerisation. However, from other literature, the conformation needed should at least be close to the optimised structure in Figure 6.4a.

It is also important to remember that in this study, the molecules are all at their ground states. Looking at the interactions with one of the molecules in its excited state might alter some of the results of this investigation. Still, it gives an indication of the most stable conformation, and where the interactions are largest. As seen in Section 4.3, the thymine molecules must be in the right conformation before excitation for dimerisation to happen. This investigation has not focused too much on the role of the nitrogens, both because the cyclobutane photoproduct has been of the greatest interest here, and also as they are not compatible with any of the most photo-reactive conformations of nitrogen. On a surface the thymine will be surrounded by other thymines, and will need to place themselves in the overall most favourable conformation. Including this in the calculations is hard. Other flat geometries could also have beed studied, but for for this investigation only the geometries with potentially the strongest interactions were studied.

6.2 Comparison to Experimental Results

6.2.1 Experimental Results

All the experimental results included in this section are made by Jakob Vinje[142], stud.techn. at the Department of Physics, NTNU, and are used here with his permission. The details concerning the experimental setup and the interpretations of the experimental data is also obtained from Jakob.

The aim of this investigation was to be able to offer support to the experimentalists performing experiments on the photochemical cyclodimerisation of thymine. The experimental work will be published in a master's thesis by Jakob Vinje[142], and interested readers are referred to his work for more information about the experimental side of this investigation. The experimentalists absorbed thymine onto a MoS_2 surface, radiated it with an unfiltered light source emitting so-called "white light" covering most of the UV region and more, trying to artificially create the cyclobutane dimer. They then perform X-ray Photoelectron Spectroscopy (XPS) combined with X-ray Absorption Spectroscopy (XAS) on the sample at deposition, after 1 hour of radiation and after 6.5 hours of radiation. From these spectra they find, amongst other things, core ionisation energies. They can also measure the relative presence π and σ bonds, which can tell if a change in bonds occur.

During the experiment the core-levels are shifted as seen in Figure 6.10. Explaining the reason for this shift, only based on these measurements, is not a trivial matter. To do this the observed peaks must be compared to literature peaks, for known substances. Performing such a literature search is easier if the product is as expected, or if one knows approximately what it is. When an experiment do not produce the expected products an investigation like this one can be of value to help explain what has happened, or just to give another viewpoint on the reaction.

In the spectra observed in Figure 6.10 only the contours of peaks can be seen. These can be further separated into which atom-type they belong to, so that the peaks can be investigated closer. It is known that the same atoms, in different environments will create slightly different signals. The apparatus used cannot make this separation. To try to separate out the different peaks, the experimentalists performed a χ^2 test, and determined the amount of and the shape of the species contributing to the larger contour peak.[142] The result of such an analysis, performed by Jakob, is included in Figure 6.11, and the binding energy of the separated peaks are found in Table 6.2. The corresponding figures for the radiated samples are found in Appendix D.

A shift as large as the one seen in Figure 6.10 can only be observed when a reaction has happened. The shifts are not large enough for the cyclobutane reaction to have happened, according to the experimentalists. Also, according to their analysis of the XAS spectra of the bonds present, no great change in the amount of carbon-carbon double bonds are seen. This support their claim that no photoproduct is formed.

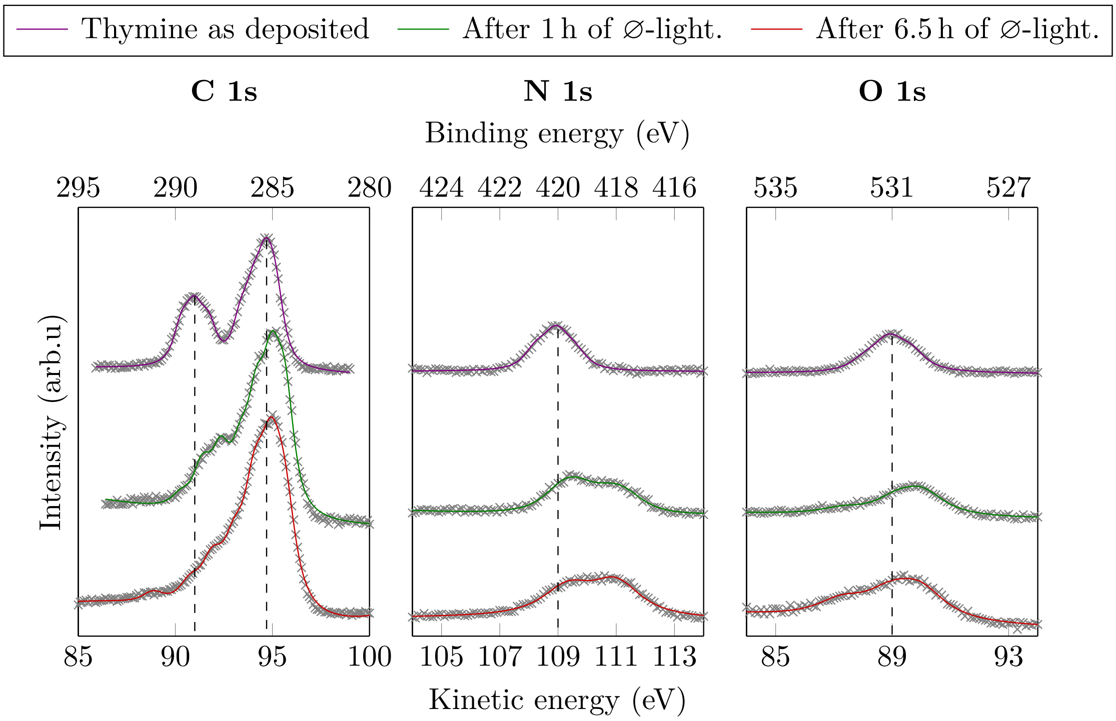


Figure 6.10: Comparison of the core-levels obtained from the experiments. The figure is made by Jakob Vinje[142]

Table 6.2: Binding energy of found components in the thymine C 1s, N 1s and O 1s core-level measurements, received from Jakob Vinje[142]

		Peak binding energy (eV)		
	Peak number	At deposition	After 1 h exposure	After 6.5 h exposure
C 1s	1	289.7	289.7	291.1
	2	289.0	288.6	289.3
	3	288.3	287.7	288.1
	4	286.6	286.6	287.0
	5	286.0	285.8	285.9
	6	285.4	285.1	285.1
	7	284.8	284.1	284.4
N 1s	1	420.7	419.6	419.6
	2	420.1	-	419.0
	3	419.4	417.9	418.0
O 1s	1	531.9	532.7	532.7
	2	531.1	530.8	530.1
	3	530.4	530.0	530.0

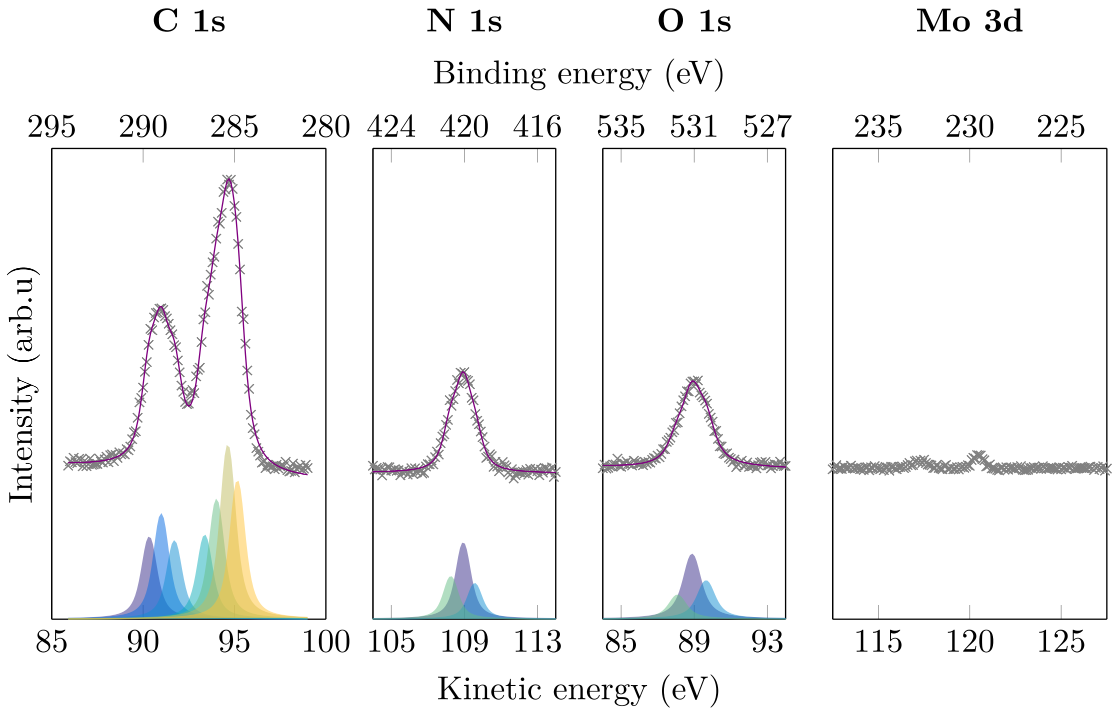


Figure 6.11: Core-levels right after deposition, here the separation of the peaks is visible. The figure is made by Jakob Vinje[142]

6.2.2 Theoretical Core Ionisation Energies

The experimental results were different from what was expected, as no cyclobutan dimer was observed. The theoretical results, along with results from the literature search, indicated that achieving the right conformation for cyclobutan dimerisation could be a challenge, but not that it would be so difficult. One possible explanation is that this is a reversible reaction, and when they shine light including the entire UV-spectrum, and more, on the sample the product can transform back into the reactants. Performing XPS experiments on more than just thymine will be very challenging. More species can lead to more disturbance in the signals, and interpreting them will be even harder. The physicists were surprised by how stabile the molecule was. Apparently, they manage to destroy many other simple organic molecules after only a short exposure to the strong radiation. In this experiment, they exposed thymine to 6.5 hours of radiation and still the molecule did not break down. This is of course good news for our genes, and not too surprising as many people choose to sunbathe for longer than that without their DNA being destroyed.

Evaluating both chemical and physical aspects as well as the experimental and theoretical aspects of the reaction, the author suggested that the shifts observed in Figure 6.10 were due to tautomerisation of the thymine molecule. From the article by Morsy et al.[134], Figure 4.4b is said to be the most stable form after regular thymine, because of time constraints for this investigation, Figure 4.4b will be the only one examined here, and simply named Tautomer in Table 6.3. Theoretical CCSD calculations on the ionisation energy of the thymine 1s orbitals of O and N where performed to see if the experimental shifts could be explained in more details by theoretical methods. The theoretical shifts for the C 1s orbital would also be of interest, but were not performed due to time constraints.

The results are show in Table 6.3. Comparing the exact values of the peaks observed here and the peak values in Table 6.2, the values do not match completely. Looking only at the shifts in energies the theoretical and experimental values seem to fit very well, as observed in Figure 6.12. In Figure 6.10, a broadening of the O 1s peak is observed. The values for O in Table 6.3 can be said to fit with that broadening, as the tautomer have both a larger and lower value than regular thymine. And after more than six hours, both larger and smaller peak values are observed in Table 6.2. The spread in the experimental values after 6.5 hours is 2.7 eV, while for the theoretical results it is 3 eV. The same trend can be seen in Table 6.3. For the N 1s orbitals, the experimental spectra gets shifted towards a lower energy, while also broadening. The shift in peaks, as seen in Table 6.2, is about 1-2 eV, the same shift is observed in Table 6.3.

Table 6.3: Ionisation energies found from CCSD calculations

		Ionisation energy (eV)	
	Orbital number	Thymine	Tautomer
O 1s	1	542.03	544.80
	2	542.11	541.80
N 1s	3	410.25	410.48
	4	409.91	408.37

To visualise the fit of the theoretical and experimental data, the theoretical values were shifted to match the experimental data. This can be justified by noting that the exact values will not be the same, and the interesting part is the shifts. In Figure 6.12, the graph at deposition contains only the thymine values, and the after radiation graphs contain only the tautomer values. The theoretical and experimental values fit well together. This good fit, indicates that it is very likely that what the experimentalists are seeing is the tautomerisation of the thymine molecule. To be more certain about which of these forms of thymine are present, the relative stability of all the tautomeric forms should be investigated, and then the shifts for all C, O and N atoms should be calculated. This was not done in this investigation, and is left as future work. In Section 6.1.4 Figure 6.8a, the hydrogen bound dimer, was compared to Figure 4.4b, the tautomer, and not found to match completely. That is not to say that Figure 4.4b is not what is seen here, as in the sample the thymines be surrounded by other thymines and hydrogen shifts can occur in several different ways[122]. As this investigation was about the mechanism of photochemical cyclisation, and not tautomerism no suggested mechanism will be presented here. Already with only the N and O 1s core-levels it is very likely that the result present after 6.5 hours of radiation is the most stable tautomer.

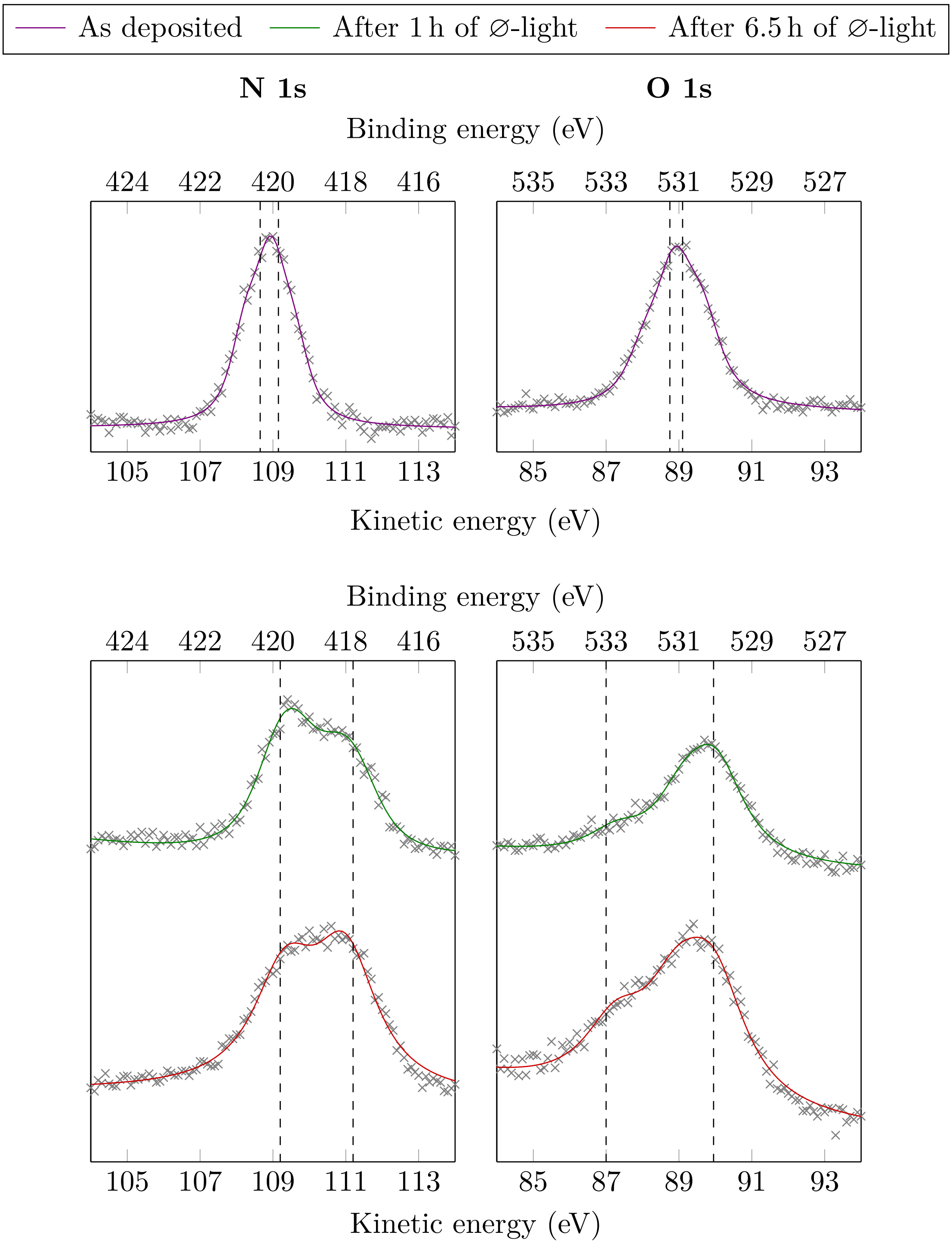


Figure 6.12: Comparison of the theoretical and experimental results. The figure is made by Jakob Vinje[142], using theoretical values from this investigation.

7 Summary and Concluding Remarks

In this investigation the interactions of two thymine molecules have been investigated using the SNOOP-scheme. It was found that the dimer has the strongest interactions either flat and hydrogen bonded, or stacked and twisted 90 degrees. The absolute values of the interactions are not of benchmark quality as only the heavy-aug-cc-pVDZ basis was used in combination with MP2. From an earlier investigation done by the author, and the investigation by Rasmussen et al.[86], the SNOOP scheme with this basis shows promising results, achieving relatively low relative and mean errors. However, the results proved to be valuable for the experimentalists, which was the goal of this investigation, and not to come with benchmark level calculations. The relative values of the interaction energies, could help explain why no photoproducts were observed, as the needed conformation for formation of the cyclobutane dimer had a very low interaction. The literature study on thymine also proved valuable, as it was through this work the tautomers were discovered.

Based on the experimental results the presence of the most stable thymine tautomer was investigated, by calculating the excitation energy of the 1s orbitals of O and N in both regular thymine and the conformation simply named tautomer in this investigation. Based on these calculations it is very probable that the tautomerisation of thymine is the reaction observed in the experiments. Finding the relative stability and the shifts of all C, N and O atoms in all the tautomers would be interesting, and if done, it would either further confirm or debunk the presence of the tautomers.

This thesis aimed at assisting experimental work on the thymine dimerisation. The experimentalists did not achieve to observe the desired photoproduct, the theoretical results helped explain the reason for this, and suggested the most likely outcome of the experiments. This thesis has gone very deep into the photochemical reaction trying to find information about its mechanism, something that will hopefully be valuable to the experimentalists in their coming experiments. Having different perspectives on the reaction proved to give a fruitful discussion about the experimental results. For the author seeing such a good fit between the experience and the theory has been very motivating. The good fit is also good validation that the theoretical methods provide valuable information and good results.

8 Future Work

This investigation only calculated ionisation energies for the N and O 1s orbitals of thymine and what is supposedly its most stable tautomer. This gave sufficient evidence to support the claim that this is the reaction observed by the experimentalists. Performing new calculations on the relative stability of the tautomers and find the ionisation energies of all C, N and O 1s orbitals would further support or debunk this theory.

The experimentalists did not manage to create the desired cyclobutane dimer during their experiments. The theoretical investigation helped explain why, and also helped in explaining the X-ray spectra. It is very likely that the experimentalists will attempt to create the desired photoproduct again, buy it and perform experiments on it, or perform experiments on a similar structure. Theoretical support will very likely also be helpful in the later experiments. Especially as the XPS spectra are not trivial to interpret.

Bibliography

[1] John Stephen Taylor. Unraveling the molecular pathway from sunlight to skin cancer. *Accounts of Chemical Research*, 27(3):76–82, 1994. ISSN 0001-4842. doi: 10.1021/ar00039a003. URL `http://dx.doi.org/10.1021/ar00039a003`.

[2] Grzegorz Chalasinski and Maciej Gutowski. Weak interactions between small systems. models for studying the nature of intermolecular forces and challenging problems for ab initio calculations. *Chemical Reviews*, 88(6):943–962, 1988. ISSN 0009-2665. doi: 10.1021/cr00088a007. URL `http://dx.doi.org/10.1021/cr00088a007http://pubs.acs.org/doi/pdf/10.1021/cr00088a007`.

[3] Frans B Van Duijneveldt, Jeanne GCM van Duijneveldt-van de Rijdt, and Joop H van Lenthe. State of the art in counterpoise theory. *Chemical Reviews*, 94(7): 1873–1885, 1994. ISSN 0009-2665.

[4] Klaus Müller-Dethlefs and Pavel Hobza. Noncovalent interactions: a challenge for experiment and theory. *Chemical Reviews*, 100(1):143–168, 2000. ISSN 0009-2665. doi: 10.1021/cr9900331. URL `http://dx.doi.org/10.1021/cr9900331http://pubs.acs.org/doi/pdfplus/10.1021/cr9900331`.

[5] Jan Řezáč and Pavel Hobza. Benchmark calculations of interaction energies in noncovalent complexes and their applications. *Chemical Reviews*, 2016. ISSN 0009-2665. doi: 10.1021/acs.chemrev.5b00526. URL `http://dx.doi.org/10.1021/acs.chemrev.5b00526http://pubs.acs.org/doi/pdfplus/10.1021/acs.chemrev.5b00526`.

[6] Matúš Dubecký, Lubos Mitas, and Petr Jurečka. Noncovalent interactions by quantum monte carlo. *Chemical Reviews*, 2016. ISSN 0009-2665. doi: 10.1021/acs.chemrev.5b00577. URL `http://dx.doi.org/10.1021/acs.chemrev.5b00577http://pubs.acs.org/doi/pdfplus/10.1021/acs.chemrev.5b00577`.

[7] Pavel Hobza. Calculations on noncovalent interactions and databases of benchmark interaction energies. *Accounts of Chemical Research*, 45(4):663–672, 2012. ISSN 0001-4842. doi: 10.1021/ar200255p. URL `http://dx.doi.org/10.1021/ar200255p`.

[8] Stefan Grimme, Andreas Hansen, Jan Gerit Brandenburg, and Christoph Bannwarth. Dispersion-corrected mean-field electronic structure methods. *Chemical Reviews*, 2016. ISSN 0009-2665. doi: 10.1021/acs.chemrev.5b00533. URL `http://dx.doi.org/10.1021/acs.chemrev.5b00533http://pubs.acs.org/doi/pdfplus/10.1021/acs.chemrev.5b00533`.

[9] Anders S. Christensen, Tomáš Kubař, Qiang Cui, and Marcus Elstner. Semiempirical quantum mechanical methods for noncovalent interactions for chemical and biochemical applications. *Chemical Reviews*, 2016. ISSN 0009-2665. doi: 10.1021/acs.chemrev.5b00584. URL `http://dx.doi.org/10.1021/acs.chemrev.5b00584http://pubs.acs.org/doi/pdfplus/10.1021/acs.chemrev.5b00584`.

[10] Bogumil Jeziorski, Robert Moszynski, and Krzysztof Szalewicz. Perturbation theory approach to intermolecular potential energy surfaces of van der waals complexes. *Chemical Reviews*, 94(7):1887–1930, 1994. ISSN 0009-2665. doi: 10.1021/cr00031a008. URL `http://dx.doi.org/10.1021/cr00031a008`.

[11] I. Mayer. Towards a "chemical" hamiltonian. *International Journal of Quantum Chemistry*, 23(2):341–363, 1983. ISSN 0020-7608. doi: 10.1002/qua.560230203.

[12] Andreas Heßelmann and Georg Jansen. First-order intermolecular interaction energies from kohn–sham orbitals. *Chemical Physics Letters*, 357(5–6):464–470, 2002. ISSN 0009-2614. doi: http://dx.doi.org/10.1016/S0009-2614(02)00538-9. URL `http://www.sciencedirect.com/science/article/pii/S0009261402005389http://ac.els-cdn.com/S0009261402005389/1-s2.0-S0009261402005389-main.pdf?_tid=18f823a4-2268-11e6-bd62-00000aacb35f&acdnat=1464174297_c629553da2c3d33b0d0d832273415b9a`.

[13] Andreas Heßelmann and Georg Jansen. Intermolecular induction and exchange-induction energies from coupled-perturbed kohn–sham density functional theory. *Chemical Physics Letters*, 362(3–4):319–325, 2002. ISSN 0009-2614. doi: http://dx.doi.org/10.1016/S0009-2614(02)01097-7. URL `http://www.sciencedirect.com/science/article/pii/S0009261402010977http://ac.els-cdn.com/S0009261402010977/1-s2.0-S0009261402010977-main.pdf?_tid=3093225c-2268-11e6-ba17-00000aacb35d&acdnat=1464174336_6c5dae5353a4a5be2da607553d4ee69b`.

[14] Andreas Heßelmann and Georg Jansen. Intermolecular dispersion energies from time-dependent density functional theory. *Chemical Physics Letters*, 367(5–6):778–784, 2003. ISSN 0009-2614. doi: http://dx.doi.org/10.1016/S0009-2614(02)01796-7. URL `http://www.sciencedirect.com/science/article/pii/S0009261402017967http://ac.els-cdn.com/S0009261402017967/1-s2.0-S0009261402017967-main.pdf?_tid=620975e8-2268-11e6-b4e7-00000aab0f02&acdnat=1464174419_cfb3a9937328175755904ccbd026856c`.

[15] Alston J. Misquitta and Krzysztof Szalewicz. Intermolecular forces from asymptotically corrected density functional description of monomers. *Chemical Physics Letters*, 357(3–4):301–306, 2002. ISSN 0009-2614. doi: http://dx.doi.org/10.1016/S0009-2614(02)00533-X. URL `http://www.sciencedirect.com/science/article/pii/S000926140200533X`.

[16] Alston J. Misquitta and Krzysztof Szalewicz. Symmetry-adapted perturbation-theory calculations of intermolecular forces employing density-functional description of monomers. *The Journal of Chemical Physics*, 122(21):214109, 2005. doi: doi:http://dx.doi.org/10.1063/1.1924593. URL `http://scitation.aip.org/content/aip/journal/jcp/122/21/10.1063/1.1924593`.

[17] Trygve Helgaker, Sonia Coriani, Poul Jørgensen, Kasper Kristensen, Jeppe Olsen, and Kenneth Ruud. Recent advances in wave function-based methods of molecular-property calculations. *Chemical Reviews*, 112(1):543–631, 2012. ISSN 0009-2665.

doi: 10.1021/cr2002239. URL `http://dx.doi.org/10.1021/cr2002239http://pubs.acs.org/doi/pdfplus/10.1021/cr2002239`.

[18] Ágnes Nagy. Advances in dft. *International Journal of Quantum Chemistry*, 116(11):801–801, 2016. ISSN 1097-461X. doi: 10.1002/qua.25138. URL `http://dx.doi.org/10.1002/qua.25138http://onlinelibrary.wiley.com/store/10.1002/qua.25138/asset/qua25138.pdf?v=1&t=iopifl9w&s=ab39b6df7f94231c1b2fd7e9c4b82263ffb9a3db`. This is a citation to what is a special issue on recent advances in DFT.

[19] Amanda Li, Hari S. Muddana, and Michael K. Gilson. Quantum mechanical calculation of noncovalent interactions: A large-scale evaluation of pmx, dft, and sapt approaches. *Journal of Chemical Theory and Computation*, 10(4):1563–1575, 2014. ISSN 1549-9618. doi: 10.1021/ct401111c. URL `http://dx.doi.org/10.1021/ct401111chttp://www.ncbi.nlm.nih.gov/pmc/articles/PMC3985464/pdf/ct401111c.pdf`.

[20] Lucie Gráfová, Michal Pitonak, Jan Řezáč, and Pavel Hobza. Comparative study of selected wave function and density functional methods for noncovalent interaction energy calculations using the extended s22 data set. *Journal of chemical theory and computation*, 6(8):2365–2376, 2010. ISSN 1549-9618.

[21] Petr Jurečka, Jiří Šponer, Jiří Černý, and Pavel Hobza. Benchmark database of accurate (mp2 and ccsd (t) complete basis set limit) interaction energies of small model complexes, dna base pairs, and amino acid pairs. *Physical Chemistry Chemical Physics*, 8(17):1985–1993, 2006.

[22] Jan Rezác, Kevin E Riley, and Pavel Hobza. S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures. *Journal of chemical theory and computation*, 7(8):2427–2438, 2011. ISSN 1549-9618. URL `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3152974/pdf/ct2002946.pdf`.

[23] Jan Řezáč and Pavel Hobza. Describing noncovalent interactions beyond the common approximations: How accurate is the "gold standard," ccsd (t) at the complete basis set limit? *Journal of Chemical Theory and Computation*, 9(5):2151–2155, 2013. ISSN 1549-9618.

[24] Tait Takatani and C. David Sherrill. Performance of spin-component-scaled moller-plesset theory (scs-mp2) for potential energy curves of noncovalent interactions. *Physical Chemistry Chemical Physics*, 9(46):6106–6114, 2007. ISSN 1463-9076. doi: 10.1039/B709669K. URL `http://dx.doi.org/10.1039/B709669K`.

[25] C. David Sherrill, Tait Takatani, and Edward G. Hohenstein. An assessment of theoretical methods for nonbonded interactions: Comparison to complete basis set limit coupled-cluster potential energy curves for the benzene dimer, the methane dimer, benzenemethane, and benzeneh2s. *The Journal of Physical Chemistry A*, 113(38):10146–10159, 2009. ISSN 1089-5639. doi: 10.1021/jp9034375. URL `http://dx.doi.org/10.1021/jp9034375http://pubs.acs.org/doi/pdfplus/10.1021/jp9034375`.

[26] Edward G. Hohenstein and C. David Sherrill. Effects of heteroatoms on aromatic interactions: Benzenepyridine and pyridine dimer. *The Journal of Physical Chemistry A*, 113(5):878–886, 2009. ISSN 1089-5639. doi: 10.1021/jp809062x. URL `http://dx.doi.org/10.1021/jp809062xhttp://pubs.acs.org/doi/pdfplus/10.1021/jp809062x`.

[27] Kanchana S. Thanthiriwatte, Edward G. Hohenstein, Lori A. Burns, and C. David Sherrill. Assessment of the performance of dft and dft-d methods for describing distance dependence of hydrogen-bonded interactions. *Journal of Chemical Theory and Computation*, 7(1):88–96, 2011. ISSN 1549-9618. doi: 10.1021/ct100469b. URL `http://dx.doi.org/10.1021/ct100469bhttp://pubs.acs.org/doi/pdfplus/10.1021/ct100469b`.

[28] John C. Faver, Mark L. Benson, Xiao He, Benjamin P. Roberts, Bing Wang, Michael S. Marshall, Matthew R. Kennedy, C. David Sherrill, and Kenneth M. Merz. Formal estimation of errors in computed absolute interaction energies of proteinligand complexes. *Journal of Chemical Theory and Computation*, 7(3):790–797, 2011. ISSN 1549-9618. doi: 10.1021/ct100563b. URL `http://dx.doi.org/10.1021/ct100563b`.

[29] S. F. Boys and F. Bernardi. The calculation of small molecular interactions by the differences of separate total energies. some procedures with reduced errors. *An International Journal at the Interface Between Chemistry and Physics*, 19(4):553–566, 1970. ISSN 0026-8976. doi: 10.1080/00268977000101561.

[30] Fu-Ming Tao. Bond functions, basis set superposition errors and other practical issues with ab initio calculations of intermolecular potentials. *International Reviews in Physical Chemistry*, 20(4):617–643, 2001. ISSN 0144-235X.

[31] Kasper Kristensen, Patrick Ettenhuber, Janus Juul Eriksen, Frank Jensen, and Poul Jørgensen. The same number of optimized parameters scheme for determining intermolecular interaction energies. *The Journal of chemical physics*, 142(11):114116, 2015. ISSN 0021-9606.

[32] Trygve Helgaker, Wim Klopper, Henrik Koch, and Jozef Noga. Basis-set convergence of correlated calculations on water. *The Journal of Chemical Physics*, 106(23):9639–9646, 1997. doi: doi:http://dx.doi.org/10.1063/1.473863. URL `http://scitation.aip.org/content/aip/journal/jcp/106/23/10.1063/1.473863http://scitation.aip.org/docserver/fulltext/aip/journal/jcp/106/23/1.473863.pdf?expires=1460366377&id=id&accname=2120809&checksum=0C5C4FCB48831501EA929EAFFCCF8ADF`.

[33] Asger Halkier, Trygve Helgaker, Poul Jørgensen, Wim Klopper, Henrik Koch, Jeppe Olsen, and Angela K. Wilson. Basis-set convergence in correlated calculations on ne, n2, and h2o. *Chemical Physics Letters*, 286(3–4):243–252, 1998. ISSN 0009-2614. doi: http://dx.doi.org/10.1016/S0009-2614(98)00111-0. URL `http://www.sciencedirect.com/science/article/pii/S0009261498001110http://ac.els-cdn.com/S0009261498001110/1-s2.0-S0009261498001110-main.pdf?_tid=4f8c1590-fa42-11e5-8c0b-00000aab0f27&acdnat=1459760021_8e4aaa6b2c217c3c7c556e728634462b`.

[34] Donald G. Truhlar. Basis-set extrapolation. *Chemical Physics Letters*, 294(1–3):45–48, 1998. ISSN 0009-2614. doi: http://dx.doi.org/10.1016/S0009-2614(98)00866-5. URL `http://www.sciencedirect.com/science/article/pii/S0009261498008665`.

[35] Dirk Bakowies. Accurate extrapolation of electron correlation energies from small basis sets. *The Journal of Chemical Physics*, 127(16):164109, 2007. doi: doi:http://dx.doi.org/10.1063/1.2768359. URL `http://scitation.aip.org/content/aip/journal/jcp/127/16/10.1063/1.2768359`.

[36] Masaki Okoshi, Teruo Atsumi, and Hiromi Nakai. Revisiting the extrapolation of correlation energies to complete basis set limit. *Journal of Computational Chemistry*, 36(14):1075–1082, 2015. ISSN 1096-987X. doi: 10.1002/jcc.23896. URL `http://dx.doi.org/10.1002/jcc.23896`.

[37] Fu-Ming Tao and Yuh-Kang Pan. Møller-plesset perturbation investigation of the he 2 potential and the role of midbond basis functions. *The Journal of Chemical Physics*, 97(7):4989–4995, 1992. ISSN 00219606.

[38] Pavel Hobza and Jiří Šponer. Toward true dna base-stacking energies: Mp2, ccsd (t), and complete basis set calculations. *Journal of the American Chemical Society*, 124(39):11802–11808, 2002. ISSN 0002-7863. URL `http://pubs.acs.org/doi/pdfplus/10.1021/ja026759n`.

[39] Seiji Tsuzuki, Kazumasa Honda, Tadafumi Uchimaru, Masuhiro Mikami, and Kazutoshi Tanabe. Origin of attraction and directionality of the / interaction: model chemistry calculations of benzene dimer interaction. *Journal of the American Chemical Society*, 124(1):104–112, 2002. ISSN 0002-7863. URL `http://pubs.acs.org/doi/pdfplus/10.1021/ja0105212`.

[40] P. W. Atkins and R. S. Friedman. *Molecular quantum mechanics.* Oxford University Press, Oxford, 5th ed. edition, 2011. ISBN 9780199541423.

[41] Encyclopaedia Britannica. Schrødinger equation, 2016. URL `http://academic.eb.com/EBchecked/topic/528298/Schrodinger-equation`. Date accessed: 23/02-16.

[42] Frank Jensen. *Introduction to computational chemistry.* Wiley, Chichester, 2nd ed. edition, 2007. ISBN 9780470011874.

[43] Jorge Kohanoff. *Electronic Structure Calculations for Solids and Molecules : Theory and Computational Methods.* Cambridge University Press, Cambridge, 2006. ISBN 0-521-81591-6.

[44] Trygve Helgaker, Poul Jørgensen, and Jeppe Olsen. *Molecular electronic-structure theory.* John Wiley & Sons, 2012. ISBN 9781118531471.

[45] C David Sherrill. An introduction to hartree-fock molecular orbital theory. *School of Chemistry and Biochemistry, Georgia Institute of Technology*, 2000.

[46] Liguo Kong, Florian A. Bischoff, and Edward F. Valeev. Explicitly correlated r12/f12 methods for electronic structure. *Chemical Reviews*, 112(1):75–107, 2012. ISSN 0009-2665. doi: 10.1021/cr200204r. URL `http://dx.doi.org/10.1021/cr200204rhttp://pubs.acs.org/doi/pdfplus/10.1021/cr200204r`.

[47] Christof Hättig, Wim Klopper, Andreas Köhn, and David P. Tew. Explicitly correlated electrons in molecules. *Chemical Reviews*, 112(1):4–74, 2012. ISSN 0009-2665. doi: 10.1021/cr200168z. URL `http://dx.doi.org/10.1021/cr200168zhttp://pubs.acs.org/doi/pdfplus/10.1021/cr200168z`.

[48] Jeppe Olsen, Ove Christiansen, Henrik Koch, and Poul Jørgensen. Surprising cases of divergent behavior in møller–plesset perturbation theory. *The Journal of Chemical Physics*, 105:5082, 1996. ISSN 00219606. doi: 10.1063/1.472352. URL `http://scitation.aip.org/docserver/fulltext/aip/journal/jcp/105/12/1.472352.pdf?expires=1461939562&id=id&accname=2120809&checksum=FAF7AE0C626B6047EF029CD99509758C`.

[49] Ove Christiansen, Jeppe Olsen, Poul Jørgensen, Henrik Koch, and Per-Åke Malmqvist. On the inherent divergence in the møller-plesset series. the neon atom — a test case. *Chemical Physics Letters*, 261(3):369–378, 1996. ISSN 0009-2614. doi: http://dx.doi.org/10.1016/0009-2614(96)00974-8. URL `http://www.sciencedirect.com/science/article/pii/0009261496009748`.

[50] Christian Møller and Milton S Plesset. Note on an approximation treatment for many-electron systems. *Physical Review*, 46(7):618, 1934.

[51] Stefan Grimme. Improved second-order møller–plesset perturbation theory by separate scaling of parallel- and antiparallel-spin pair correlation energies. *The Journal of Chemical Physics*, 118(20):9095–9102, 2003. doi: doi:http://dx.doi.org/10.1063/1.1569242. URL `http://scitation.aip.org/content/aip/journal/jcp/118/20/10.1063/1.1569242http://scitation.aip.org/docserver/fulltext/aip/journal/jcp/118/20/1.1569242.pdf?expires=1461694315&id=id&accname=2120809&checksum=F087EE74C1ED34E4609C8648D20BBB0D`.

[52] J. Grant Hill and James A. Platts. Spin-component scaling methods for weak and stacking interactions. *Journal of Chemical Theory and Computation*, 3(1):80–85, 2007. ISSN 1549-9618. doi: 10.1021/ct6002737. URL `http://dx.doi.org/10.1021/ct6002737`.

[53] Robert A Distasio Jr and Martin Head-Gordon. Optimized spin-component scaled second-order møller-plesset perturbation theory for intermolecular interaction energies. *Molecular Physics*, 105(8):1073–1083, 2007. ISSN 0026-8976.

[54] Kevin E. Riley, Michal Pitoňák, Petr Jurečka, and Pavel Hobza. Stabilization and structure calculations for noncovalent interactions in extended molecular systems based on wave function and density functional theories. *Chemical Reviews*, 110(9):5023–5063, 2010. ISSN 0009-2665. doi: 10.1021/cr1000173. URL `http://dx.doi.org/10.1021/cr1000173http://pubs.acs.org/doi/pdfplus/10.1021/cr1000173`.

[55] Jens Antony and Stefan Grimme. Is spin-component scaled second-order møller-plesset perturbation theory an appropriate method for the study of noncovalent interactions in molecules? *The Journal of Physical Chemistry A*, 111(22):4862–4868, 2007. ISSN 1089-5639. doi: 10.1021/jp070589p. URL http://dx.doi.org/10.1021/jp070589phttp://pubs.acs.org/doi/pdfplus/10.1021/jp070589p.

[56] Matthew Goldey and Martin Head-Gordon. Attenuating away the errors in inter- and intramolecular interactions from second-order møller–plesset calculations in the small aug-cc-pvdz basis set. *The Journal of Physical Chemistry Letters*, 3(23):3592–3598, 2012. ISSN 1948-7185. doi: 10.1021/jz301694b. URL http://dx.doi.org/10.1021/jz301694bhttp://pubs.acs.org/doi/pdfplus/10.1021/jz301694b.

[57] Matthew Goldey, Anthony Dutoi, and Martin Head-Gordon. Attenuated second-order moller-plesset perturbation theory: performance within the aug-cc-pvtz basis. *Physical Chemistry Chemical Physics*, 15(38):15869–15875, 2013. ISSN 1463-9076. doi: 10.1039/C3CP51826D. URL http://dx.doi.org/10.1039/C3CP51826Dhttp://pubs.rsc.org/en/content/articlepdf/2013/cp/c3cp51826d.

[58] Matthew B. Goldey and Martin Head-Gordon. Convergence of attenuated second order møller–plesset perturbation theory towards the complete basis set limit. *Chemical Physics Letters*, 608:249–254, 2014. ISSN 0009-2614. doi: http://dx.doi.org/10.1016/j.cplett.2014.05.092. URL http://www.sciencedirect.com/science/article/pii/S0009261414004746http://ac.els-cdn.com/S0009261414004746/1-s2.0-S0009261414004746-main.pdf?_tid=aa9bd882-fb0d-11e5-92de-00000aacb35f&acdnat=1459847362_3b3f39f54e36b36fde229b6c696fb292.

[59] Michal Pitoňák, Pavel Neogrády, Jiří Černý, Stefan Grimme, and Pavel Hobza. Scaled mp3 non-covalent interaction energies agree closely with accurate ccsd(t) benchmark data. *ChemPhysChem*, 10(1):282–289, 2009. ISSN 1439-7641. doi: 10.1002/cphc.200800718. URL http://dx.doi.org/10.1002/cphc.200800718.

[60] Wim Klopper and Claire C. M. Samson. Explicitly correlated second-order møller–plesset methods with auxiliary basis sets. *The Journal of Chemical Physics*, 116(15):6397–6410, 2002. doi: doi:http://dx.doi.org/10.1063/1.1461814. URL http://scitation.aip.org/content/aip/journal/jcp/116/15/10.1063/1.1461814http://scitation.aip.org/docserver/fulltext/aip/journal/jcp/116/15/1.1461814.pdf?expires=1459240206&id=id&accname=2120809&checksum=E0A8D7E8159ED390D01D0655C302AB6C.

[61] Werner Kutzelnigg. r12-dependent terms in the wave function as closed sums of partial wave amplitudes for large l. *Theoretica chimica acta*, 68(6):445–469, 1985. ISSN 1432-2234. doi: 10.1007/bf00527669. URL http://dx.doi.org/10.1007/BF00527669http://download.springer.com/static/pdf/423/art%253A10.1007%252FBF00527669.pdf?originUrl=http%3A%2F%2Flink.springer.com%2Farticle%2F10.1007%2FBF00527669&token2=exp=1460888644~acl=%2Fstatic%2Fpdf%2F423%2Fart%25253A10.1007%25252FBF00527669.pdf%3ForiginUrl%3Dhttp%253A%252F%252Flink.

`springer.com%252Farticle%252F10.1007%252FBF00527669*~hmac=d4c91938e79f9c0b273ecf4a7a07396c756ace46ba6231230caa4bbf8d410312`.

[62] Florian Weigend, Marco Häser, Holger Patzelt, and Reinhart Ahlrichs. Ri-mp2: optimized auxiliary basis sets and demonstration of efficiency. *Chemical Physics Letters*, 294(1–3):143–152, 1998. ISSN 0009-2614. doi: http://dx.doi.org/10.1016/S0009-2614(98)00862-8. URL `http://www.sciencedirect.com/science/article/pii/S0009261498008628`.

[63] Stanislav Kedžuch, Matúš Milko, and Jozef Noga. Alternative formulation of the matrix elements in mp2-r12 theory. *International Journal of Quantum Chemistry*, 105(6):929–936, 2005. ISSN 1097-461X. doi: 10.1002/qua.20744. URL `http://dx.doi.org/10.1002/qua.20744`.

[64] Hans-Joachim Werner, Thomas B. Adler, and Frederick R. Manby. General orbital invariant mp2-f12 theory. *The Journal of Chemical Physics*, 126(16):164102, 2007. doi: doi:http://dx.doi.org/10.1063/1.2712434. URL `http://scitation.aip.org/content/aip/journal/jcp/126/16/10.1063/1.2712434`.

[65] Werner Kutzelnigg and Wim Klopper. Wave functions with terms linear in the interelectronic coordinates to take care of the correlation cusp. i. general theory. *The Journal of Chemical Physics*, 94(3):1985–2001, 1991. doi: doi:http://dx.doi.org/10.1063/1.459921. URL `http://scitation.aip.org/content/aip/journal/jcp/94/3/10.1063/1.459921`.

[66] Wim Klopper, Frederick R Manby, Seiichiro Ten-No, and Edward F Valeev. R12 methods in explicitly correlated molecular electronic structure theory. *International Reviews in Physical Chemistry*, 25(3):427–468, 2006. ISSN 0144-235X.

[67] Seiichiro Ten-no. Explicitly correlated wave functions: summary and perspective. *Theoretical Chemistry Accounts*, 131(1):1–11, 2012. ISSN 1432-881X.

[68] Krishnan Raghavachari, Gary W. Trucks, John A. Pople, and Martin Head-Gordon. A fifth-order perturbation comparison of electron correlation theories. *Chemical Physics Letters*, 157(6):479–483, 1989. ISSN 0009-2614. doi: http://dx.doi.org/10.1016/S0009-2614(89)87395-6. URL `http://www.sciencedirect.com/science/article/pii/S0009261489873956`.

[69] Lori A. Burns, Michael S. Marshall, and C. David Sherrill. Appointing silver and bronze standards for noncovalent interactions: A comparison of spin-component-scaled (scs), explicitly correlated (f12), and specialized wavefunction approaches. *The Journal of Chemical Physics*, 141(23):234111, 2014. doi: doi:http://dx.doi.org/10.1063/1.4903765. URL `http://scitation.aip.org/content/aip/journal/jcp/141/23/10.1063/1.4903765http://scitation.aip.org/docserver/fulltext/aip/journal/jcp/141/23/1.4903765.pdf?expires=1459244721&id=id&accname=2120809&checksum=036933D61762941839539CA59E74AABB`.

[70] Yoon S. Lee, Stanislaw A. Kucharski, and Rodney J. Bartlett. A coupled cluster approach with triple excitations. *The Journal of Chemical Physics*, 81(12):5906–5912, 1984. doi: doi:http://dx.doi.org/10.1063/1.447591. URL `http://scitation.aip.org/content/aip/journal/jcp/81/12/10.1063/1.447591`.

[71] Jozef Noga, Rodney J. Bartlett, and Miroslav Urban. Towards a full ccsdt model for electron correlation. ccsdt-n models. *Chemical Physics Letters*, 134(2):126–132, 1987. ISSN 0009-2614. doi: http://dx.doi.org/10.1016/0009-2614(87)87107-5. URL `http://www.sciencedirect.com/science/article/pii/0009261487871075`.

[72] Brian W. Hopkins and Gregory S. Tschumper. Ab initio studies of ··· interactions: the effects of quadruple excitations. *The Journal of Physical Chemistry A*, 108(15):2941–2948, 2004. ISSN 1089-5639. doi: 10.1021/jp0369084. URL `http://dx.doi.org/10.1021/jp0369084http://pubs.acs.org/doi/pdfplus/10.1021/jp0369084`.

[73] Josef Noga, Werner Kutzelnigg, and Wim Klopper. Cc-r12, a correlation cusp corrected coupled-cluster method with a pilot application to the be2 potential curve. *Chemical Physics Letters*, 199(5):497–504, 1992. ISSN 0009-2614. doi: http://dx.doi.org/10.1016/0009-2614(92)87034-M. URL `http://www.sciencedirect.com/science/article/pii/000926149287034M`.

[74] Michael S. Marshall and C. David Sherrill. Dispersion-weighted explicitly correlated coupled-cluster theory [dw-ccsd(t**)-f12]. *Journal of Chemical Theory and Computation*, 7(12):3978–3982, 2011. ISSN 1549-9618. doi: 10.1021/ct200600p. URL `http://dx.doi.org/10.1021/ct200600p`.

[75] Rodney J. Bartlett and Monika Musiał. Coupled-cluster theory in quantum chemistry. *Reviews of Modern Physics*, 79(1):291–352, 2007. URL `http://link.aps.org/doi/10.1103/RevModPhys.79.291`.

[76] Trygve Helgaker, Wim Klopper, and David P. Tew. Quantitative quantum chemistry. *An International Journal at the Interface Between Chemistry and Physics*, 106(16-18):2107–2143, 2008. ISSN 0026-8976. doi: 10.1080/00268970802258591. URL `http://www.tandfonline.com/doi/pdf/10.1080/00268970802258591`.

[77] Ove Christiansen, Henrik Koch, Asger Halkier, Poul Jørgensen, Trygve Helgaker, and Alfredo Sánchez de Merás. Largescale calculations of excitation energies in coupled cluster theory: The singlet excited states of benzene. *The Journal of Chemical Physics*, 105(16):6921–6939, 1996. doi: doi:http://dx.doi.org/10.1063/1.471985. URL `http://scitation.aip.org/content/aip/journal/jcp/105/16/10.1063/1.471985http://scitation.aip.org/docserver/fulltext/aip/journal/jcp/105/16/1.471985.pdf?expires=1465043197&id=id&accname=2120809&checksum=D28200C8894EA66B151221E2A7ED5A5D`.

[78] Sonia Coriani and Henrik Koch. Communication: X-ray absorption spectra and core-ionization potentials within a core-valence separated coupled cluster framework. *The Journal of Chemical Physics*, 143(18):181103, 2015. doi: doi:http://dx.doi.org/10.1063/1.4935712. URL `http://scitation.aip.org/content/aip/journal/jcp/143/18/10.1063/1.4935712http:`

//scitation.aip.org/docserver/fulltext/aip/journal/jcp/143/18/1.4935712.pdf?expires=1465310562&id=id&accname=guest&checksum=468B042324B841BC7C70ADFECD6FA8AChttp://scitation.aip.org/docserver/fulltext/aip/journal/jcp/143/18/1.4935712.pdf?expires=1465310569&id=id&accname=guest&checksum=61F6EA05143248E107A40FD83CD0A159.

[79] Pavel Hobza, Klaus Muller-Dethlefs, and Jonathan Hirst. *Non-Covalent Interactions : Theory and Experiment*, volume v.2 of *RSC Theoretical and Computational Chemistry Series, 2*. Royal Society of Chemistry, Cambridge, 2009. ISBN 1-84755-853-4.

[80] Michael S. Marshall, Lori A. Burns, and C. David Sherrill. Basis set convergence of the coupled-cluster correction, δ(mp2)(ccsd(t)): best practices for benchmarking non-covalent interactions and the attendant revision of the s22, nbc10, hbc6, and hsg databases. *The Journal of chemical physics*, 135(19):194102, 2011. doi: 10.1063/1.3659142. URL `http://scitation.aip.org/docserver/fulltext/aip/journal/jcp/135/19/1.3659142.pdf?expires=1459257765&id=id&accname=2120809&checksum=C099DA2AEC5EF00D6FF5CB24AF844E13`.

[81] J. Řezáč, M. Dubecky, P. Jurečka, and P. Hobza. Extensions and applications of the a24 data set of accurate interaction energies. *Phys. Chem. Chem. Phys.*, 17(29):19268–19277, 2015. ISSN 1463-9076. doi: 10.1039/c5cp03151f. URL `http://pubs.rsc.org/en/content/articlepdf/2015/cp/c5cp03151f`.

[82] Rafal Podeszwa, Konrad Patkowski, and Krzysztof Szalewicz. Improved interaction energy benchmarks for dimers of biological relevance. *Physical chemistry chemical physics : PCCP*, 12(23):5974, 2010. doi: 10.1039/b926808a. URL `http://pubs.rsc.org/en/content/articlepdf/2010/cp/b926808a`.

[83] Tait Takatani, Edward G. Hohenstein, Massimo Malagoli, Michael S. Marshall, and C. David Sherrill. Basis set consistent revision of the s22 test set of noncovalent interaction energies. *The Journal of Chemical Physics*, 132(14):144104, 2010. doi: doi:http://dx.doi.org/10.1063/1.3378024. URL `http://scitation.aip.org/content/aip/journal/jcp/132/14/10.1063/1.3378024http://scitation.aip.org/docserver/fulltext/aip/journal/jcp/132/14/1.3378024.pdf?expires=1459607882&id=id&accname=2120809&checksum=6D5EEEA6FC355069EDEE8DC97F7AC6CF`.

[84] Joanna C. Flick, Dmytro Kosenkov, Edward G. Hohenstein, C. David Sherrill, and Lyudmila V. Slipchenko. Accurate prediction of noncovalent interaction energies with the effective fragment potential method: Comparison of energy components to symmetry-adapted perturbation theory for the s22 test set. *Journal of Chemical Theory and Computation*, 8(8):2835–2843, 2012. ISSN 1549-9618. doi: 10.1021/ct200673a. URL `http://dx.doi.org/10.1021/ct200673ahttp://pubs.acs.org/doi/pdfplus/10.1021/ct200673a`.

[85] Lori A. Burns, Michael S. Marshall, and C. David Sherrill. Comparing counterpoise-corrected, uncorrected, and averaged binding energies for benchmarking noncovalent interactions. *Journal of Chemical Theory and Computation*, 10(1):49–57,

2014. ISSN 1549-9618. doi: 10.1021/ct400149j. URL http://dx.doi.org/10.1021/ct400149jhttp://pubs.acs.org/doi/pdfplus/10.1021/ct400149j.

[86] Troels Hels Rasmussen, Yang Min Wang, Thomas Kjærgaard, and Kasper Kristensen. The explicitly correlated same number of optimized parameters (snoop-f12) scheme for calculating intermolecular interaction energies. *The Journal of Chemical Physics*, 144(20):204102, 2016. doi: doi:http://dx.doi.org/10.1063/1.4950846. URL http://scitation.aip.org/content/aip/journal/jcp/144/20/10.1063/1.4950846.

[87] Lucia Šimová, Jan Řezáč, and Pavel Hobza. Convergence of the interaction energies in noncovalent complexes in the coupled-cluster methods up to full configuration interaction. *Journal of Chemical Theory and Computation*, 9(8):3420–3428, 2013. ISSN 1549-9618. doi: 10.1021/ct4002762. URL http://dx.doi.org/10.1021/ct4002762.

[88] M. Gutowski, J. H. Van Lenthe, J. Verbeek, F. B. Van Duijneveldt, and G. Chałasinski. The basis set superposition error in correlated electronic structure calculations. *Chemical Physics Letters*, 124(4):370–375, 1986. ISSN 0009-2614. doi: http://dx.doi.org/10.1016/0009-2614(86)85036-9. URL http://www.sciencedirect.com/science/article/pii/0009261486850369.

[89] Maciej Gutowski, Frans B Van Duijneveldt, Grzegorz Chałasiński, and Lucjan Piela. Proper correction for the basis set superposition error in scf calculations of intermolecular interactions. *Molecular Physics*, 61(1):233–247, 1987. ISSN 0026-8976.

[90] I. Mayer and L. Turi. An analytical investigation into the bsse problem. *Journal of Molecular Structure: THEOCHEM*, 227:43–65, 1991. ISSN 0166-1280. doi: http://dx.doi.org/10.1016/0166-1280(91)85271-8. URL http://www.sciencedirect.com/science/article/pii/0166128091852718.

[91] Juan Raul Alvarez-Idaboy and Annia Galano. Counterpoise corrected interaction energies are not systematically better than uncorrected ones: comparison with ccsd (t) cbs extrapolated values. *Theoretical Chemistry Accounts*, 126(1-2):75–85, 2010. ISSN 1432-881X.

[92] Ł M. Mentel and E. J. Baerends. Can the counterpoise correction for basis set superposition effect be justified? *Journal of Chemical Theory and Computation*, 10(1):252–267, 2014. ISSN 1549-9618. doi: 10.1021/ct400990u. URL http://dx.doi.org/10.1021/ct400990uhttp://pubs.acs.org/doi/pdfplus/10.1021/ct400990u.

[93] E. Miliordos and S. S. Xantheas. On the validity of the basis set superposition error and complete basis set limit extrapolations for the binding energy of the formic acid dimer. *J. Chem. Phys.*, 142(9), 2015. ISSN 0021-9606. doi: 10.1063/1.4913766.

[94] Małgorzata Jeziorska, Robert Bukowski, Wojciech Cencek, Michał Jaszuński, Bogumił Jeziorski, and Krzysztof Szalewicz. On the performance of bond functions and basis set extrapolation techniques in high-accuracy calculations of interatomic potentials. a helium dimer study. *Collection of Czechoslovak chemical communications*, 68(3):463–488, 2003. ISSN 1212-6950.

[95] Małgorzata Jeziorska, Wojciech Cencek, Konrad Patkowski, Bogumił Jeziorski, and Krzysztof Szalewicz. Complete basis set extrapolations of dispersion, exchange, and coupledclusters contributions to the interaction energy: a helium dimer study. *International Journal of Quantum Chemistry*, 108(12):2053–2075, 2008. ISSN 1097-461X.

[96] Rudolf Burcl, Grzegorz Chałasiński, Robert Bukowski, and Małgorzata M. Szczęśniak. On the role of bond functions in interaction energy calculations: Ar···hcl, ar···h2o, (hf)2. *The Journal of Chemical Physics*, 103(4):1498–1507, 1995. doi: doi:http://dx.doi.org/10.1063/1.469771. URL `http://scitation.aip.org/content/aip/journal/jcp/103/4/10.1063/1.469771http://scitation.aip.org/docserver/fulltext/aip/journal/jcp/103/4/1.469771.pdf?expires=1463052356&id=id&accname=2120809&checksum=562685EC94C4688C4C8BB56FE49D3629http://scitation.aip.org/docserver/fulltext/aip/journal/jcp/103/4/1.469771.pdf?expires=1463052363&id=id&accname=2120809&checksum=6A1CF6F933092A42A24A0CA86459FD21`.

[97] Henrik Koch, Berta Fernández, and Ove Christiansen. The benzene–argon complex: A ground and excited state ab initio study. *The Journal of Chemical Physics*, 108(7):2784–2790, 1998. doi: doi:http://dx.doi.org/10.1063/1.475669. URL `http://scitation.aip.org/content/aip/journal/jcp/108/7/10.1063/1.475669http://scitation.aip.org/docserver/fulltext/aip/journal/jcp/108/7/1.475669.pdf?expires=1461672334&id=id&accname=2120809&checksum=469B783569D21EE45F9E877722B34A05http://scitation.aip.org/docserver/fulltext/aip/journal/jcp/108/7/1.475669.pdf?expires=1461672340&id=id&accname=2120809&checksum=16977DB91720868B18064F27FF59B345`.

[98] Jiandong Zhang, Shujin Li, and Fuming Tao. Ab initio calculations of the ar–ethane intermolecular potential energy surface using bond function basis sets. *Journal of Computational Chemistry*, 34(8):673–680, 2013. ISSN 0192-8651. doi: 10.1002/jcc.23179. URL `http://onlinelibrary.wiley.com/store/10.1002/jcc.23179/asset/23179_ftp.pdf?v=1&t=imhmjocx&s=52a2bcad86af1ea84a2d5e767cc8d0169762a873`.

[99] Ove Christiansen, Henrik Koch, and Poul Jørgensen. The second-order approximate coupled cluster singles and doubles model cc2. *Chemical Physics Letters*, 243(5):409–418, 1995. ISSN 0009-2614. doi: 10.1016/0009-2614(95)00841-Q.

[100] Joop H van Lenthe, Teus van Dam, Frans B van Duijneveldt, and Loes MJ Kroon-Batenburg. Ab initio calculations on weakly bonded systems. In *Faraday Symposia of the Chemical Society*, volume 19, pages 125–135. Royal Society of Chemistry, 1984.

[101] Martin Schütz, Steve Brdarski, Per-Olof Widmark, Roland Lindh, and Gunnar Karlström. The water dimer interaction energy: Convergence to the basis set limit at the correlated level. *The Journal of Chemical Physics*, 107(12):4597–4605, 1997. doi: doi:http://dx.doi.org/10.1063/1.474820. URL `http://scitation.aip.org/content/aip/journal/jcp/107/12/10.1063/1.474820`.

[102] Annia Galano and J. Raúl Alvarez–Idaboy. A new approach to counterpoise correction to bsse. *Journal of Computational Chemistry*, 27(11):1203–1210, 2006. ISSN 1096-987X. doi: 10.1002/jcc.20438. URL `http://dx.doi.org/10.1002/jcc.20438`.

[103] Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *The Journal of Chemical Physics*, 90(2):1007–1023, 1989. doi: doi:http://dx.doi.org/10.1063/1.456153. URL `http://scitation.aip.org/content/aip/journal/jcp/90/2/10.1063/1.456153`.

[104] Rick A. Kendall, Thom H. Dunning, and Robert J. Harrison. Electron affinities of the firstrow atoms revisited. systematic basis sets and wave functions. *The Journal of Chemical Physics*, 96(9):6796–6806, 1992. doi: doi:http://dx.doi.org/10.1063/1.462569. URL `http://scitation.aip.org/content/aip/journal/jcp/96/9/10.1063/1.462569http://scitation.aip.org/docserver/fulltext/aip/journal/jcp/96/9/1.462569.pdf?expires=1461694899&id=id&accname=2120809&checksum=ED842541FC467114E7A522C1D777BCD7`.

[105] Ewa Papajak, Jingjing Zheng, Xuefei Xu, Hannah R. Leverentz, and Donald G. Truhlar. Perspectives on basis sets beautiful: Seasonal plantings of diffuse basis functions. *Journal of Chemical Theory and Computation*, 7(10):3027–3034, 2011. ISSN 1549-9618. doi: 10.1021/ct200106a. URL `http://dx.doi.org/10.1021/ct200106a`.

[106] David E. Woon and Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. v. corevalence basis sets for boron through neon. *The Journal of Chemical Physics*, 103(11):4572–4585, 1995. doi: doi:http://dx.doi.org/10.1063/1.470645. URL `http://scitation.aip.org/content/aip/journal/jcp/103/11/10.1063/1.470645http://scitation.aip.org/docserver/fulltext/aip/journal/jcp/103/11/1.470645.pdf?expires=1462525598&id=id&accname=2120809&checksum=84243F5BB2D781D2628A95483E1E21B2`.

[107] Kirk A. Peterson and Thom H. Dunning. Accurate correlation consistent basis sets for molecular core–valence correlation effects: The second row atoms al–ar, and the first row atoms b–ne revisited. *The Journal of Chemical Physics*, 117(23):10548–10560, 2002. doi: doi:http://dx.doi.org/10.1063/1.1520138. URL `http://scitation.aip.org/content/aip/journal/jcp/117/23/10.1063/1.1520138`.

[108] Thom H. Dunning, Kirk A. Peterson, and Angela K. Wilson. Gaussian basis sets for use in correlated molecular calculations. x. the atoms aluminum through argon revisited. *The Journal of Chemical Physics*, 114(21):9244–9253, 2001. doi: doi:http://dx.doi.org/10.1063/1.1367373. URL `http://scitation.aip.org/content/aip/journal/jcp/114/21/10.1063/1.1367373`.

[109] P. C. Hariharan and J. A. Pople. The influence of polarization functions on molecular orbital hydrogenation energies. *Theoretica chimica acta*, 28(3):213–222, 1973. ISSN 1432-2234. doi: 10.1007/bf00533485. URL `http://dx.doi.org/10.1007/BF00533485`.

[110] David E. Bernholdt and Robert J. Harrison. Fitting basis sets for the ri-mp2 approximate second-order many-body perturbation theory method. *The Journal of Chemical Physics*, 109(5):1593–1600, 1998. doi: doi:http://dx.doi.org/10.1063/1.476732. URL `http://scitation.aip.org/content/aip/journal/jcp/109/5/10.1063/1.476732`.

[111] Florian Weigend, Andreas Köhn, and Christof Hättig. Efficient use of the correlation consistent basis sets in resolution of the identity mp2 calculations. *The Journal of Chemical Physics*, 116(8):3175–3183, 2002. doi: doi:http://dx.doi.org/10.1063/1.1445115. URL `http://scitation.aip.org/content/aip/journal/jcp/116/8/10.1063/1.1445115`.

[112] Christof Hattig. Optimization of auxiliary basis sets for ri-mp2 and ri-cc2 calculations: Core-valence and quintuple-[small zeta] basis sets for h to ar and qzvpp basis sets for li to kr. *Physical Chemistry Chemical Physics*, 7(1):59–66, 2005. ISSN 1463-9076. doi: 10.1039/B415208E. URL `http://dx.doi.org/10.1039/B415208E`.

[113] Kirk A. Peterson, Thomas B. Adler, and Hans-Joachim Werner. Systematically convergent basis sets for explicitly correlated wavefunctions: The atoms h, he, b–ne, and al–ar. *The Journal of Chemical Physics*, 128(8):084102, 2008. doi: doi:http://dx.doi.org/10.1063/1.2831537. URL `http://scitation.aip.org/content/aip/journal/jcp/128/8/10.1063/1.2831537`.

[114] Masato Tanaka, Michio Katouda, and Shigeru Nagase. Optimization of rimp2 auxiliary basis functions for 631g** and 6311g** basis sets for first, second, and thirdrow elements. *Journal of Computational Chemistry*, 34(29):2568–2575, 2013. ISSN 0192-8651. doi: 10.1002/jcc.23430.

[115] William H. Elliott, Daphne C. Elliott, and Despo Papachristodoulou. *Biochemistry and molecular biology*. Oxford University Press, Oxford, 5th ed. edition, 2014. ISBN 9780199609499.

[116] Stefan Weber. Light-driven enzymatic catalysis of dna repair: a review of recent biophysical studies on photolyase. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1707(1):1–23, 2005. ISSN 0005-2728. doi: http://dx.doi.org/10.1016/j.bbabio.2004.02.010. URL `http://www.sciencedirect.com/science/article/pii/S0005272804000830http://ac.els-cdn.com/S0005272804000830/1-s2.0-S0005272804000830-main.pdf?_tid=4cda06a2-baa2-11e5-8fbf-00000aab0f6c&acdnat=1452764374_5c4bdc696c378ca2daac626704373799`.

[117] R. Beukers, Apm Eker, and Phm Lohman. 50 years thymine dimer. *DNA Repair*, 7(3):530–543, 2008. ISSN 1568-7864. doi: 10.1016/j.dnarep.2007.11.010. URL `http://ac.els-cdn.com/S1568786407004119/1-s2.0-S1568786407004119-main.pdf?_tid=8967f5ea-c10d-11e5-8c1f-00000aacb361&acdnat=1453470138_c4097a70ed8e249d0b23fb4cf3fd21db`.

[118] Adolf Wacker, Hanswerner Dellweg, and Diether Jacherts. Thymin-dimerisierung und Überlebensrate bei bakterien. *Journal of Molecular Biology*, 4(5):410–412, 1962.

ISSN 0022-2836. doi: http://dx.doi.org/10.1016/S0022-2836(62)80022-9. URL `http://www.sciencedirect.com/science/article/pii/S0022283662800229`.

[119] Jean-Luc Ravanat, Thierry Douki, and Jean Cadet. Direct and indirect effects of uv radiation on dna and its components. *Journal of Photochemistry and Photobiology B: Biology*, 63(1–3):88–102, 2001. ISSN 1011-1344. doi: http://dx.doi.org/10.1016/S1011-1344(01)00206-8. URL `http://www.sciencedirect.com/science/article/pii/S1011134401002068`.

[120] Bo Durbeej and Leif A Eriksson. On the formation of cyclobutane pyrimidine dimers in uvirradiated dna: Why are thymines more reactive? *Photochemistry and photobiology*, 78(2):159–167, 2003. ISSN 1751-1097. URL `http://onlinelibrary.wiley.com/store/10.1562/0031-8655(2003)0780159OTFOCP2.0.CO2/asset/0031-8655(2003)0780159OTFOCP2.0.CO2.pdf?v=1&t=in8u2r1m&s=52e2af7d29362a67c462c7907d749fe835eaea14`.

[121] Intisar Husain, Jack Griffith, and Aziz Sancar. Thymine dimers bend dna. *Proceedings of the National Academy of Sciences of the United States of America*, 85(8): 2558–2562, 1988. ISSN 00278424.

[122] Eric V. Anslyn and Dennis A. Dougherty. *Modern physical organic chemistry.* University Science Books, Sausalito, Cal, 2006. ISBN 1891389319.

[123] Aziz Sancar. Structure and function of dna photolyase and cryptochrome blue-light photoreceptors. *Chemical Reviews*, 103(6):2203–2238, 2003. ISSN 0009-2665. doi: 10.1021/cr0204348. URL `http://dx.doi.org/10.1021/cr0204348`.

[124] Ru Bo Zhang and Leif A. Eriksson. A triplet mechanism for the formation of cyclobutane pyrimidine dimers in uv-irradiated dna. *The journal of physical chemistry. B*, 110(14):7556, 2006. ISSN 1520-6106. URL `http://pubs.acs.org/doi/pdfplus/10.1021/jp060196a`.

[125] Martial Boggio-Pasqua, Gerrit Groenhof, Lars V. Schäfer, Helmut Grubmüller, and Michael A. Robb. Ultrafast deactivation channel for thymine dimerization. *Journal of the American Chemical Society*, 129(36):10996–10997, 2007. ISSN 0002-7863. doi: 10.1021/ja073628j. URL `http://dx.doi.org/10.1021/ja073628jhttp://pubs.acs.org/doi/pdfplus/10.1021/ja073628j`.

[126] Juan José Serrano-Pérez, Israel Gonzalez-Ramirez, Pedro B. Coto, Manuela Merchán, and Luis Serrano-Andrés. Theoretical insight into the intrinsic ultrafast formation of cyclobutane pyrimidine dimers in uv-irradiated dna: Thymine versus cytosine. *The Journal of Physical Chemistry B*, 112(45):14096–14098, 2008. ISSN 1520-6106. doi: 10.1021/jp806794x. URL `http://dx.doi.org/10.1021/jp806794xhttp://pubs.acs.org/doi/pdfplus/10.1021/jp806794x`.

[127] Lluís Blancafort and Annapaola Migani. Modeling thymine photodimerizations in dna: mechanism and correlation diagrams. *Journal of the American Chemical Society*, 129(47):14540–14541, 2007. ISSN 0002-7863. doi: 10.1021/ja074734o. URL `http://dx.doi.org/10.1021/ja074734ohttp://pubs.acs.org/doi/pdfplus/10.1021/ja074734o`.

[128] Wolfgang J. Schreier, Tobias E. Schrader, Florian O. Koller, Peter Gilch, Carlos E. Crespo-Hernández, Vijay N. Swaminathan, Thomas Carell, Wolfgang Zinth, and Bern Kohler. Thymine dimerization in dna is an ultrafast photoreaction. *Science*, 315(5812):625–629, 2007. doi: 10.1126/science.1135428. URL `http://science.sciencemag.org/content/sci/315/5812/625.full.pdf`.

[129] Bert M. Pilles, Dominik B. Bucher, Lizhe Liu, Peter Gilch, Wolfgang Zinth, and Wolfgang J. Schreier. Identification of charge separated states in thymine single strands. *Chem. Commun.*, 50(98):15623–15626, 2014. ISSN 1359-7345. doi: 10.1039/c4cc07663j. URL `http://pubs.rsc.org/en/content/articlepdf/2014/cc/c4cc07663j`.

[130] Angelo Giussani, Luis Serrano-Andrés, Manuela Merchán, Daniel Roca-Sanjuán, and Marco Garavelli. Photoinduced formation mechanism of the thymine-thymine (6-4) adduct. *The journal of physical chemistry. B*, 117(7):1999, 2013. doi: 10.1021/jp307200g. URL `http://pubs.acs.org/doi/pdfplus/10.1021/jp307200g`.

[131] Zhao Bo Yang, Ru Bo Zhang, and Leif A. Eriksson. A triplet mechanism for the formation of thymine–thymine (6-4) dimers in uv-irradiated dna. *Phys. Chem. Chem. Phys.*, 13(19):8961–8966, 2011. ISSN 1463-9076. doi: 10.1039/c0cp02830d. URL `http://pubs.rsc.org/en/content/articlepdf/2011/cp/c0cp02830d`.

[132] Sylvie Marguet and Dimitra Markovitsi. Time-resolved study of thymine dimer formation. *Journal of the American Chemical Society*, 127(16):5780–5781, 2005. ISSN 0002-7863. doi: 10.1021/ja050648h. URL `http://dx.doi.org/10.1021/ja050648hhttp://pubs.acs.org/doi/pdfplus/10.1021/ja050648h`.

[133] Akos Banyasz, Thierry Douki, Roberto Improta, Thomas Gustavsson, Delphine Onidas, Ignacio Vayá, Marion Perron, and Dimitra Markovitsi. Electronic excited states responsible for dimer formation upon uv absorption directly by thymine strands: Joint experimental and theoretical study. *Journal of the American Chemical Society*, 134(36):14834–14845, 2012. ISSN 0002-7863. doi: 10.1021/ja304069f. URL `http://dx.doi.org/10.1021/ja304069fhttp://pubs.acs.org/doi/pdfplus/10.1021/ja304069f`.

[134] M. A. Morsy, A. M. Al-Somali, and A. Suwaiyan. Fluorescence of thymine tautomers at room temperature in aqueous solutions. *The Journal of Physical Chemistry B*, 103(50):11205–11210, 1999. ISSN 1520-6106. doi: 10.1021/jp990858e. URL `http://dx.doi.org/10.1021/jp990858ehttp://pubs.acs.org/doi/pdfplus/10.1021/jp990858e`.

[135] Dongsheng Jiao, Hongyan Wang, Yanlan Zhang, and Yu Tang. A dft study of thymine and its tautomers. *Canadian Journal of Chemistry*, 87(2):406–415, 2009. ISSN 0008-4042. doi: 10.1139/v08-168.

[136] M. Piacenza and S. Grimme. Systematic quantum chemical study of dna-base tautomers. *Journal of Computational Chemistry*, 25(1):83–99, 2004. ISSN 1096-987X. doi: 10.1002/jcc.10365. URL `http://dx.doi.org/10.1002/jcc.10365`.

[137] Kęstutis Aidas, Celestino Angeli, Keld L. Bak, Vebjørn Bakken, Radovan Bast, Linus Boman, Ove Christiansen, Renzo Cimiraglia, Sonia Coriani, Pål Dahle, Erik K. Dalskov, Ulf Ekström, Thomas Enevoldsen, Janus J. Eriksen, Patrick Ettenhuber, Berta Fernández, Lara Ferrighi, Heike Fliegl, Luca Frediani, Kasper Hald, Asger Halkier, Christof Hättig, Hanne Heiberg, Trygve Helgaker, Alf Christian Hennum, Hinne Hettema, Eirik Hjertenæs, Stinne Høst, Ida-Marie Høyvik, Maria Francesca Iozzi, Branislav Jansík, Hans Jørgen Aa. Jensen, Dan Jonsson, Poul Jørgensen, Joanna Kauczor, Sheela Kirpekar, Thomas Kjærgaard, Wim Klopper, Stefan Knecht, Rika Kobayashi, Henrik Koch, Jacob Kongsted, Andreas Krapp, Kasper Kristensen, Andrea Ligabue, Ola B Lutnæs, Juan I. Melo, Kurt V. Mikkelsen, Rolf H. Myhre, Christian Neiss, Christian B. Nielsen, Patrick Norman, Jeppe Olsen, Jógvan Magnus H. Olsen, Anders Osted, Martin J. Packer, Filip Pawlowski, Thomas B. Pedersen, Patricio F. Provasi, Simen Reine, Zilvinas Rinkevicius, Torgeir A. Ruden, Kenneth Ruud, Vladimir V. Rybkin, Pawel Sałek, Claire C. M. Samson, Alfredo Sánchez de Merás, Trond Saue, Stephan P. A. Sauer, Bernd Schimmelpfennig, Kristian Sneskov, Arnfinn H. Steindal, Kristian O. Sylvester-Hvid, Peter R. Taylor, Andrew M. Teale, Erik I. Tellgren, David P. Tew, Andreas J. Thorvaldsen, Lea Thøgersen, Olav Vahtras, Mark A. Watson, David J. D. Wilson, Marcin Ziolkowski, and Hans Ågren. The Dalton quantum chemistry program system. *WIREs Comput. Mol. Sci.*, 4(3):269–284, 2015. doi: 10.1002/wcms.1172. LSDalton, a linear scaling molecular electronic structure program, Release Dalton2015.X (2015), see `http://daltonprogram.org`.

[138] Marcus D Hanwell, Donald Ephraim Curtis, David C Lonie, Tim Vandermeersch, Eva Zurek, and Geoffrey R Hutchison. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminformatics*, 4(1):17, 2012.

[139] Y. Shao, Z. Gan, E. Epifanovsky, A. T. B. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kús, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock III, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. DiStasio Jr., H. Dop, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. D. Hanson-Heine, P. H. P. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, P. A. Pieniazek, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, N. Sergueev, S. M. Sharada, S. Sharmaa, D. W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A. J. W. Thom, T. Tsuchimochi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, V. Vanovschi, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhou, B. R. Brooks, G. K. L. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard III,

M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer III, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xua, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. Van Voorhis, J. M. Herbert, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon. Advances in molecular quantum chemistry contained in the q-chem 4 program package. *Mol. Phys.*, 113:184–215, 2015.

[140] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. Ucsf chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004. ISSN 1096-987X. doi: 10.1002/jcc.20084. URL `http://dx.doi.org/10.1002/jcc.20084`.

[141] MATLAB. *version 8.4.0.150421 (R2014b)*. The MathWorks Inc., Natick, Massachusetts, 2014.

[142] Jakob Vinje. Photochemistry of a nucleobase relevant to the formation of malignant melanoma - using spectroscopic techniques to study the uv-induced alternation of thymine on MoS_2. Master's thesis, NTNU, 2016. Written at the Department of Physics.

[143] Bo Durbeej and Leif A. Eriksson. Reaction mechanism of thymine dimer formation in dna induced by uv light. *Journal of Photochemistry Photobiology, A: Chemistry*, 152(1):95–101, 2002. ISSN 1010-6030. doi: 10.1016/S1010-6030(02)00180-6. URL `http://ac.els-cdn.com/S1010603002001806/1-s2.0-S1010603002001806-main.pdf?_tid=fe93f514-c345-11e5-bb27-00000aacb35f&acdnat=1453714289_01cc6cb643f5477b9e7681a15385e300`.

[144] Hiroto Tachikawa and Hiroshi Kawabata. A direct ab initio molecular dynamics (md) study on the repair reactions of stacked thymine dimer. *Chemical Physics Letters*, 462(4–6):321–326, 2008. ISSN 0009-2614. doi: http://dx.doi.org/10.1016/j.cplett.2008.07.107. URL `http://www.sciencedirect.com/science/article/pii/S0009261408010762http://ac.els-cdn.com/S0009261408010762/1-s2.0-S0009261408010762-main.pdf?_tid=21cadbb8-c34d-11e5-b061-00000aacb361&acdnat=1453717355_63f5a408186987eed8bb537676a4a3a7`.

[145] Jiande Gu, Jing Wang, Jerzy Leszczynski, Yaoming Xie, and Henry F. Schaefer Iii. To stack or not to stack: Performance of a new density functional for the uracil and thymine dimers. *Chemical Physics Letters*, 459(1–6):164–166, 2008. ISSN 0009-2614. doi: http://dx.doi.org/10.1016/j.cplett.2008.05.049. URL `http://www.sciencedirect.com/science/article/pii/S0009261408006465http://ac.els-cdn.com/S0009261408006465/1-s2.0-S0009261408006465-main.pdf?_tid=98d55480-c34e-11e5-a3b1-00000aab0f02&acdnat=1453717984_f829334ae95e27e52a4ce6e2ed202b8b`.

[146] Jesús González-vázquez, Leticia González, Elena Samoylova, and Thomas Schultz. Thymine relaxation after uv irradiation: the role of tautomerization and * states.

Physical Chemistry Chemical Physics, 11(20):3927–3934, 2009. ISSN 1463-9076. doi: 10.1039/b815602f. URL `http://pubs.rsc.org/en/content/articlepdf/2009/cp/b815602f`.

[147] Claudio A. Morgado, Petr Jurečka, Daniel Svozil, Pavel Hobza, and Jiří Šponer. Balance of attraction and repulsion in nucleic-acid base stacking: Ccsd(t)/complete-basis-set-limit calculations on uracil dimer and a comparison with the force-field description. *Journal of Chemical Theory and Computation*, 5(6):1524–1544, 2009. ISSN 1549-9618. doi: 10.1021/ct9000125. URL `http://dx.doi.org/10.1021/ct9000125http://pubs.acs.org/doi/pdfplus/10.1021/ct9000125`.

[148] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical Review*, 136(3B): B864–B871, 1964. URL `http://link.aps.org/doi/10.1103/PhysRev.136.B864`.

[149] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140(4A):A1133–A1138, 1965. URL `http://link.aps.org/doi/10.1103/PhysRev.140.A1133`.

[150] Laszlo Fusti Molnar, Xiao He, Bing Wang, and Kenneth M. Merz. Further analysis and comparative study of intermolecular interactions using dimers from the s22 database. *The Journal of Chemical Physics*, 131(6):065102, 2009. doi: doi:http://dx.doi.org/10.1063/1.3173809. URL `http://scitation.aip.org/content/aip/journal/jcp/131/6/10.1063/1.3173809http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2738737/pdf/JCPSA6-000131-065102_1.pdf`.

[151] Qin Wu and Weitao Yang. Empirical correction to density functional theory for van der waals interactions. *The Journal of Chemical Physics*, 116(2):515–524, 2002. doi: doi:http://dx.doi.org/10.1063/1.1424928. URL `http://scitation.aip.org/content/aip/journal/jcp/116/2/10.1063/1.1424928`.

[152] Stefan Grimme. Accurate description of van der waals complexes by density functional theory including empirical corrections. *Journal of computational chemistry*, 25(12):1463–1473, 2004. ISSN 1096-987X.

[153] Petr Jurečka, Jiří Černý, Pavel Hobza, and Dennis R. Salahub. Density functional theory augmented with an empirical dispersion term. interaction energies and geometries of 80 noncovalent complexes compared with ab initio quantum mechanics calculations. *Journal of Computational Chemistry*, 28(2):555–569, 2007. ISSN 1096-987X. doi: 10.1002/jcc.20570. URL `http://dx.doi.org/10.1002/jcc.20570`.

[154] Stefan Grimme. Semiempirical ggatype density functional constructed with a longrange dispersion correction. *Journal of computational chemistry*, 27(15): 1787–1799, 2006. ISSN 1096-987X. URL `http://onlinelibrary.wiley.com/store/10.1002/jcc.20495/asset/20495_ftp.pdf?v=1&t=inppi9l6&s=865f8c958075282832b656511eb98d3f1968e036`.

[155] Stefan Grimme, Jens Antony, Tobias Schwabe, and Christian Mück-lichtenfeld. Density functional theory with dispersion corrections for supramolecular structures, aggregates, and complexes of (bio)organic molecules. *Organic Biomolecular Chemistry*, 5(5):741–758, 2007. ISSN 1477-0520. doi: 10.1039/b615319b. URL `http://pubs.rsc.org/en/content/articlepdf/2007/ob/b615319b`.

[156] Erich Runge and E. K. U. Gross. Density-functional theory for time-dependent systems. *Physical Review Letters*, 52(12):997–1000, 1984. URL http://link.aps.org/doi/10.1103/PhysRevLett.52.997.

[157] Andreas Dreuw and Martin Head-Gordon. Single-reference ab initio methods for the calculation of excited states of large molecules. *Chemical Reviews*, 105(11):4009–4037, 2005. ISSN 0009-2665. doi: 10.1021/cr0505627. URL http://dx.doi.org/10.1021/cr0505627http://pubs.acs.org/doi/pdfplus/10.1021/cr0505627.

[158] Felix Plasser, Mario Barbatti, Adélia J. A. Aquino, and Hans Lischka. Electronically excited states and photodynamics: a continuing challenge. *Theoretical Chemistry Accounts*, 131(1):1–14, 2012. ISSN 1432-2234. doi: 10.1007/s00214-011-1073-y. URL http://dx.doi.org/10.1007/s00214-011-1073-yhttp://download.springer.com/static/pdf/621/art%253A10.1007%252Fs00214-011-1073-y.pdf?originUrl=http%3A%2F%2Flink.springer.com%2Farticle%2F10.1007%2Fs00214-011-1073-y&token2=exp=1461055445~acl=%2Fstatic%2Fpdf%2F621%2Fart%25253A10.1007%25252Fs00214-011-1073-y.pdf%3ForiginUrl%3Dhttp%253A%252F%252Flink.springer.com%252Farticle%252F10.1007%252Fs00214-011-1073-y*~hmac=89c49db61c6ad5342f90a3567403afc2830ef97b7053028775ed7093344d6e59.

[159] Michael J. G. Peach, Peter Benfield, Trygve Helgaker, and David J. Tozer. Excitation energies in density functional theory: An evaluation and a diagnostic test. *The Journal of Chemical Physics*, 128(4):044118, 2008. doi: doi:http://dx.doi.org/10.1063/1.2831900. URL http://scitation.aip.org/content/aip/journal/jcp/128/4/10.1063/1.2831900.

[160] Michael JG Peach, C Ruth Le Sueur, Kenneth Ruud, Maxime Guillaume, and David J Tozer. Tddft diagnostic testing and functional assessment for triazene chromophores. *Physical Chemistry Chemical Physics*, 11(22):4465–4470, 2009. URL http://pubs.rsc.org/en/Content/ArticleLanding/2009/CP/b822941dhttp://pubs.rsc.org/en/content/articlepdf/2009/cp/b822941d.

[161] Gerhard Hirsch, Pablo J. Bruna, Sigrid D. Peyerimhoff, and Robert J. Buenker. Ab initio ci study of the stability and electronic spectrum of the hocl molecule. *Chemical Physics Letters*, 52(3):442–448, 1977. ISSN 0009-2614. doi: http://dx.doi.org/10.1016/0009-2614(77)80483-1. URL http://www.sciencedirect.com/science/article/pii/0009261477804831.

[162] Fu-Ming Tao and Yuh-Kang Pan. An accurate ab initio calculation of the ne 2 potential. *Chemical Physics Letters*, 194(3):162–166, 1992. ISSN 0009-2614. doi: 10.1016/0009-2614(92)85527-H.

[163] Fu-Ming Tao. The use of midbond functions for ab initio calculations of the asymmetric potentials of he-ne and he-ar. *Journal of Chemical Physics*, 98(4):3049–3059, 1993. ISSN 00219606.

[164] Krzystof Szalewicz, Bogumil Jeziorski, Hendrik J. Monkhorst, and John G. Zabolitzky. A new functional for variational calculation of atomic and molecular second-order correlation energies. *Chemical Physics Letters*, 91(3):169–172,

1982. ISSN 0009-2614. doi: http://dx.doi.org/10.1016/0009-2614(82)83634-8. URL `http://www.sciencedirect.com/science/article/pii/0009261482836348`.

[165] J. Grant Hill, Kirk A. Peterson, Gerald Knizia, and Hans-Joachim Werner. Extrapolating mp2 and ccsd explicitly correlated correlation energies to the complete basis set limit with first and second row correlation consistent basis sets. *The Journal of Chemical Physics*, 131(19):194105, 2009. doi: doi:http://dx.doi.org/10.1063/1.3265857. URL `http://scitation.aip.org/content/aip/journal/jcp/131/19/10.1063/1.3265857http://scitation.aip.org/docserver/fulltext/aip/journal/jcp/131/19/1.3265857.pdf?expires=1459240271&id=id&accname=2120809&checksum=1E6EAF634F1C138E607D249F3E76C9FChttp://scitation.aip.org/docserver/fulltext/aip/journal/jcp/131/19/1.3265857.pdf?expires=1459240297&id=id&accname=2120809&checksum=C33926CA25338D84E1D93DEDDFC8E56B`.

[166] Oregon State University Libraries Special Collections Archives Research Center (sarc). The pauling point, 2008. URL `https://paulingblog.wordpress.com/2008/07/10/the-pauling-point/`. Date accessed: 13/04-16.

[167] Konrad Patkowski and Krzysztof Szalewicz. Argon pair potential at basis set and excitation limits. *The Journal of chemical physics*, 133(9):094304, 2010. ISSN 0021-9606. URL `http://scitation.aip.org/docserver/fulltext/aip/journal/jcp/133/9/1.3478513.pdf?expires=1458294850&id=id&accname=2120809&checksum=979B0AB30BA2CA0A94EC72526D3370FE`.

A Additional Information on Density Theory Methods

A.1 The Density Functional Theory Methods

The Density Functional Theory (DFT)[148,149] is the most widely used alternative to the wave function methods. These methods have long been viewed as too inaccurate for quantum chemical computations, but are now able to reach quite high accuracies. The largest change from wave function methods is the use of electron density, consisted of only three spatial coordinates, to determine the molecular properties. Many different functionals and corrections to the DFT energy are available, and to appropriately use DFT, some preliminary work on the methods available is advised. The different test sets discussed for the wave function methods are also valuable for the DFT methods, and the sets are used to test the performance of different DFT functionals.[27,150] Hence, it is possible to get some hints to the performance of the specific functionals. The abundance of available functionals make the DFT calculations a labyrinth of possibilities. The computational time of these methods are considerably lower, and is often the only possibility for larger systems. With the availability of the test sets, it has also been shown that these methods now achieve very high accuracies and are thus a force to be reckoned with in the field of computational chemistry.

One of the biggest disadvantages with the original DFT methods, with respect to calculation of noncovalent interactions, is their inability to describe London dispersions.[5] As the dispersion energy is rather well separated from the DFT energy, a solution to this problem is to add it independently. This is what is done in the DFT-D[151–155] methods. Even a simple pairwise empirical model can be added and result in a high degree of accuracy.[5]

Thanthiriwatte et al.[27] tested several counterpoise corrected classical DFT functional as well as DFT-D functionals on whether they correctly describe the potential energy surface of the members of the HBC6 hydrogen bonded test set, using CCSD(T) reference values. They found that even though DFT is said to compute accurate values for H-bonded systems, some of the classical methods failed to reach an acceptable accuracy. The functionals: B3LYP, BP86, and B970 were shown to not be able to describe the H-bonding interaction quantitatively,[27] while the PBE and PBE0 functional were able to come close to the CCSD(T) values. These performed so well alone that the empirically corrected dispersion term values resulted in the method performing less well. Showing that one needs to evaluate the results before adding any corrections and claiming better results, and that often the methods rely on a favourable cancellation of errors. The DFT-D methods M05-2X, M06-2X, B970-D2, and ωB97X-D gave low mean average error values, and showed that appropriately corrected DFT-D methods can perform very well.

In a study performed by Gráfová et al.[20] on the S22 test set, different wave function based methods and DFT functionals was tested by calculating the interaction energy at the equilibrium separation, as well as one compressed and three extended geometries. Most of the compressed geometries had to be removed, in order not to contaminate the statistics. This happened for both the wave function based and DFT based methods. Five criteria were then proposed, that would test the overall performance across binding motif in the S22 test set. Each criterion was given a high and low threshold, and then the methods were scored according to how many of the criteria they fulfilled. None of the DFT methods fulfilled the statistical criteria set by the authors, and they do not recommend using them as a black box method. That is not to say that Gráfová et al.[20] warns against the use of DFT methods, they just emphasise the point that they do not work well as black box methods at this moment. For the functionals tested Gráfová et al.[20] states that the counterpoise scheme is well suited also for the DFT methods, when the DFT methods provides a good description of the problem. With increasing basis set, the CP-scheme was shown to be less important. It also did not perform equally well for the more empirically DFT-D methods.[20]

A.2 Time Dependent Density Functional Theory

Time Dependent Density Functional Theory (TD-DFT)[156] is now one of the most used methods in calculations of excited states properties on systems of medium to large size.[157] The foundation of this methods builds on the Kohn-Sham formalism, in that the time-dependent electron density determines the time-dependent external field, up to a spatially constant, time-dependent function, and thus also the time-dependent wave function.[157] As with time-independent DFT different functionals are available, some even work both for the time-dependent and time-independent cases. And as with the time-independent case, the results are often very dependent on the choice of functional. Thus, it is important to have good wave function based calculations or accurate experimental values to compare to, and also know the sensitivity of the system to the choice of functional.[157,158] Given the right system, an appropriate exchange correlation functional the TD-DFT method can to a good degree of accuracy, compared to experiments, perform relatively computationally cheap calculations on systems containing up to 300 second row atoms, compared to the wave function methods that become very expensive with more than 20 atoms.[157]

Dreuw and Head-Gordon[157] states that the TD-DFT methods are usually incapable of correctly describing charge transfer (CT) states when a "traditional" DFT exchange correlation functional is used. The excitation energy of such states is underestimated and the 1/R dependence along a charge-separation coordinate, R, is not correctly described. The time-dependent HF methods being built on a different assumption manage these calculations better, and inclusion of a non-local HF exchange in the exchange correlation potential in the TD-DFT method might improve the calculation.[157] New exchange correlation functional, including a non-local HF, has been developed, but a need for further investigation into the photo-physics of the problem is needed.[158] Plasser et al.[158] performed an investigation of the adenine$\cdots$thymine dimer, at its ground-state optimised geometry of the charge-transfer state using the RI-ADC(2)/TZVP method, and compared to TD-DFT results using the M06-2X functional. This functional is show to give good agreements with the RI-ADC(2) results, and avoids the overstabilisation of the CT states that the B3LYP functional does.[158] For long range CT states, the troublesome determinants vanish, and both TD-DFT and TD-HF provides the same answer, and better description. Something that is interesting to note is that Banyasz et al.[133] suggests that a CT state is important in the thymine dimerisation process.

When testing the TD-DFT methods, it is often customary to compare the values to a wave function based method. Managing to describe excited states and charge transfer state accurately is challenging for many approaches. Peach et al.[159] devised a simple diagnostics test based on spatial orbital overlap to assess the reliability of the calculated excited state from different TD-DFT functionals. The test still needs some further refinements, but shows promising behaviour so far.[160]

B Additional Information on Approaches for Coming Closer to the CBS Interaction Energy

B.1 Midbond Functions

Many attempts have been made to reduce the BSSE. After investigations of the Rydberg atoms H_3, NH_4 and H_3O, where diffuse s and p functions were used as an addition to the basis set, the notion of non-atomic basis functions were created for all molecules.[30] The idea of bond functions had been around longer, but not widely implemented.[161] To obtain an accurate interaction energy, large basis sets with highly diffuse functions are usually required. This is to better describe the interactions holding the two molecules together. This bond is said to be made up of two excited electrons, one from each subsystem, at least for the noble gases.[162] These excited electrons form a bonding orbital in the mid region between the two atoms.[162] The use of bond functions, centred off the atoms, to reduce the need of large highly augmented atomic functions and thus using a smaller atomic basis set, while obtaining the accuracy was suggested.[37,162] This as bond functions placed at the mid point between two atoms/molecules are better at describing the excited configuration. A criterion posed by Tao and Pan[162] is that the ΔE^{HF} should remain relatively unaffected by the addition of the midbond functions.

Another way to justify the use of midbond functions is to consider the highly excited electrons on each system. For them, there is an effective nucleus in the midpoint between the two systems. This system can be regarded as a Rydberg atom. The most efficient way of describing a Rydberg atom is by having the basis functions centred in the middle of the system. This Rydberg system will thus be well described by the midbond functions.[37,162]

Tao and Pan[37] used primitive Gaussians as midbond functions, with some variation in the exponent. They discovered that the calculated interaction energy was not so sensitive to changes in the exponent or the placement of the midbond functions. Tao[163] also found that for asymmetrical potential the placement of the midbond functions did not influence the interaction energy. As discussed in the same article, the midbond function will not work if the BSSE is not removed properly as in the CP-scheme. Tao and Pan[37] states that midbond functions must be used with some care, and that the core electrons must be properly described before including bond functions. This is to avoid unphysical shifts of the electrons due to the extra bond functions. They also state that an appropriate amount of the polarised atom centred functions need to be included for a proper description of intersystem correlations.[37]

Midbond functions have proved valuable for many calculations[30,94,96–98]. One of the greatest advantages is that including midbond functions does not increase the computational cost as much as increasing the basis set size, while at the same time giving interaction energies at a similar level as an increased basis set. The previous unpublished study of interaction energies between small systems, found that including midbond functions gave consistently low relative errors for the CP-scheme. The CP-values with the largest basis set were the references in this study, but all values seemed to converge towards the same value. The low relative error indicates that even the smaller basis set were able to come quite close to the value given by much larger basis sets. The inclusion of midbond functions did not have the same effect on the SNOOP-scheme. In fact the inclusion of midbond functions resulted in less accurate values from the SNOOP-scheme. The SNOOP-scheme without midbond functions gave values close to the CP-values where midbond functions were included. For all the molecules in that study there was a marked difference between including midbond functions and not, especially for the smaller basis sets. Still, midbond functions are not very widely used, and an extrapolation scheme is often the chosen method towards the CBS energy. It is possible to include midbond functions when also using an extrapolation scheme. Zhang et al.[98] reports very good results for such an approach. Řezáč and Hobza[5] have provided a very detailed review of different benchmark methods regarding interaction energies, but does not discuss the use of bond functions in regards to benchmark interaction energies.

B.2 Extrapolation Schemes

As mentioned in section 3.3, one way to try to come closer to the CBS energy is to use an extrapolation scheme. For this to give a good result, it is important to use a series of functions with increasing highest angular momentum number l. This is due to the electron correlation being better described with increasing the number of l, and the increase in basis set size should increase the accuracy in a systematic way, making it possible to use an extrapolation scheme.[5] The correlation consistent basis functions of Dunning possess these qualities and can be used in combination with an extrapolation scheme.

It is possible to derive a result showing that the convergence of the correlation energy follows approximately with the inverse of the third power of the basis set size expressed by the cardinal number.[5] This lead Helgaker et. al.[32,33] to introduce a function giving the basis set dependence of the correlation energy expressed using the cardinal number X.

$$\Delta E_{CBS}^{corr} = E_{CBS}^{corr} + aX^{-3} \tag{B.1}$$

Where a is a system dependent parameter, a parameter that it is desirable to eliminate. To do this an extrapolation between at least two points must be performed, this is usually done using cardinal numbers X and X+1. The expression for the correlation energy will then look like this:

$$\Delta E_{CBS}^{corr} = \frac{(X+1)^3 E_{X+1}^{corr} - X^3 E_{X}^{corr}}{(X+1)^3 - X^3} \tag{B.2}$$

An expression widely used in the literature to extrapolate results to the CBS limit.[5] A more general form of Equation (B.2) can be used by replacing the 3 by p in the exponent. Other exponentials have been investigated[81] without much luck. Three point extrapolation schemes have also been suggested, also with minor luck.

The Helgaker extrapolation scheme assumes that the only error stems from the limited description of the correlation error. An assumption that work well for larger basis sets, but which can give inaccurate results when used for smaller basis sets. Extrapolations performed on values obtained by using midbond functions perform especially well.[95] In general, this scheme should not be used on results extrapolated between the double and triple-ζ basis sets, as this might results in poorer answers than the triple-ζ itself.[5,33] The extrapolation should be done using the two larges basis sets. Others have tried to fix this deficiency of the Helgaker extrapolation scheme, but so far no one has succeeded.[5,94]

From two studies on the He dimer, Jeziorska et al.[94],[95] studied the convergence of different extrapolation schemes, and also the effect of including midbond functions for these calculations. Only the first study corrected for the BSSE using the CP-scheme, which is a drawback of the other paper. Both studies found that for general use the X^{-3} extrapolation scheme is the overall best performing one. When breaking the different contributions down, some schemes will outperform it, but for all practical purposes its easier to stick to one scheme. Their reference values comes from calculations using the Gaussian-type geminals (GTG)[164], which make the conclusions harder to compare to other studies. One of the interesting things from these studies is the effect of including midbond functions to an extrapolation scheme.

The latest of the studies by Jeziorska et al.[95] also investigated the effect of separation on both the extrapolation scheme and the inclusion on midbond functions. Both depend on the separation, and not all approaches are able to handle either or both the repulsion when the systems are moved very close and the weakening forces when the systems are moved far apart. Moreover, the benefit of including midbond functions changes with the separation. For small R, the atomic basis functions themselves have sufficiently large amplitudes in the midbond region. For large R the midbond functions become less important due to the distance and the already small interaction.[95]

For small systems it is also possible to use the results from a whole series of increasing basis functions results in the extrapolation, here the double-ζ value if often emitted. For these large systems it is possible to fit the values to Equation (B.1) directly. Another approach is to vary the exponent in the power law formula and then fit it in a series of the three largest basis sets. The latter approach being the better justified one, as the scaling does not follow Equation (B.1) directly, just closely. Again, for practical purposes, the easier Helgaker approach is often well enough.

Burns et al.[85] tested whether or not an average between the uncorrected and the counterpoise corrected value would give better results than the counterpoise corrected values themselves. For the smallest basis sets, they found that the averaging avoided the worst errors in both methods, but that at the larger basis sets, and especially when extrapolation was used, the error became so small that the averaging had less effect. Still, the CP corrected values performed somewhat better. Both the averaging and the CP corrected values were better than the uncorrected values for the 27 test cases considered in the article.[85]

In their article Kristensen et al.[31] claim that the SNOOP-scheme is not suitable for extrapolations like the X^{-3} scheme of Halkier et al.[33]. But they do claim that the values obtained using the SNOOP-scheme are of similar quality to the extrapolated CP-values of similar computational cost.[31] Especially SNOOP implemented with the explicitly correlated F12 method seems to come very close to extrapolated CP results. The article by Rasmussen et al.[86], the SNOOP-F12 values are said to be so good that one cannot say whether they or the extrapolated CP values are the most accurate. The effect of composite schemes on the SNOOP-scheme is not known, and is one of the areas where the SNOOP scheme needs further investigation.

Hill et al.[165] introduced an extrapolation scheme for explicitly correlated methods, which was optimised using a Schwenke-style approach for first and second row elements (and both combinded). They report good agreement to their reference values for their new method, but this method will not be discussed much further and interested readers are referred to the work by them, Ref[165]. Other investigations report that the extrapolated MP2 method and unextrapolated MP2-F12 perform similarly[81], and thus see no need to use an extrapolation scheme when using an explicitly correlated method.

B.3 Composite CCSD(T)/CBS Schemes

The CCSD(T) method is often called the gold standard of interaction energy calculations, due to its ability to incorporate the electron correlation. Řezáč and Hobza[5] call it the first method to include correlations at a satisfactory level. The CCSD(T) is a very computationally expensive method, making the largest basis sets unavailable for all but the smallest systems. As discussed in Section B.2 the first reliable extrapolation results come from basis sets higher than the double-ζ, and most often an accuracy higher than the triple-ζ is sought, making extrapolation schemes in the CCSD(T) method unavailable for all but the smallest systems. For larger systems, composite schemes is the most favourable solution.[5]

For noncovalent systems, the favourite solution to the extrapolation scheme problems of CCSD(T) is to decompose the CCSD(T) energy[38,39] into the three terms, Hartree-Fock energy, MP2 correlation energy and a higher order correction.[81]

$$E^{\mathrm{CCSD(T)}} = E^{\mathrm{HF}} + E^{\mathrm{MP2}} + \delta E^{\mathrm{CCSD(T)}} \tag{B.3}$$

Where the $\delta E^{\mathrm{CCSD(T)}}$ is defined as

$$\delta E^{\mathrm{CCSD(T)}} = E^{\mathrm{CCSD(T)}} - E^{\mathrm{MP2}} \tag{B.4}$$

Both $\mathrm{E}^{\mathrm{CCSD(T)}}$ and $\mathrm{E}^{\mathrm{MP2}}$ are calculated in the same basis here. In Equation (B.3), the different contributions to the CCSD(T) energy can be found using different basis sets. The HF contribution converges faster than the other contributions, so fast that it does not need to be extrapolated further. The MP2 method is capable of including a lot of the correlation energy, but is also known for overestimating the dispersion energy. Still, MP2 is computationally cheaper than CCSD(T), and a relatively large basis can be used, explicitly correlated methods and/or an extrapolation scheme can be used to further improve the MP2 energy. This low computational cost makes MP2 a preferred method for used in such composite schemes. The main weakness of the MP2 method is the overestimation of the interaction energy, something the CCSD(T) correction tries to balance out. Řezáč and Hobza[5] report of attempts of using the more expensive MP3 or CCSD methods, but to little improvements in the accuracy in the final composite scheme.

Podeszwa et al.[82] claim that the accuracy of the composite schemes depend on how well the MP2 method covers the correlation energy, and thus the ratio between the $\delta E_{\text{int}}^{\text{MP2/FC}}$ and $\delta E_{\text{int}}^{\text{CCSD(T)/FC}}$ correction and the size of the basis set used for the CCSD(T) correction. For small hydrogen bound systems, the MP2 method covers a lot of the correlation energy, and it is possible to use a relatively large basis set for the CCSD(T) calculation, giving good accuracy. For dispersion bound complexes the two corrections are about the same order of magnitude, making the correction more sensitive to the basis set used. For larger systems, only smaller CCSD(T) basis sets can be used. Podeszwa et al.[82] claim that the original values for the dispersion bound complexes of the S22 test set of Jurečka et al.[21], where only small basis sets were used for the CCSD(T) correction was off by about 10%, a value too high for a benchmark calculation. Podeszwa et al.[82] attempt to improve on the accuracy by increasing the basis set used, and also including midbond function for the CCSD(T) calculations. Marshall et al.[80] report that the double-ζ values can be better if the explicitly correlated methods are used both for the MP2 and CCSD(T) corrections, but that otherwise the small basis sets usually are not capable of reaching benchmark quality corrections.

In the composite schemes it has been customary to use MP2 extrapolation from the aug-cc-pVTZ to the aug-cc-pVQZ level, making it accurate enough not to make this the limiting factor in accuracy. It is also possible to calculate the MP2 energy using the explicitly correlated MP2 methods. Řezáč and Hobza[5] state that for the double-ζ basis, the explicitly correlated methods outperforms the extrapolation schemes. For values obtained using the quadruple-ζ basis set, using an extrapolation scheme from aug-cc-pVTZ to aug-cc-pVQZ might be less computationally demanding than the explicitly correlated method. They are reported to give similar answers. The need for a composite scheme when an explicitly correlated method is used is sometimes debated, as the CCSD(T)-F12 method is capable of producing quite accurate results on its own simply using the aug-cc-pVDZ basis.

Marshall et al.[80] investigate the convergence of the CCSD(T) correction. In an earlier paper, Takatani et al.[83] reported that using a $\delta_{\text{MP2}}^{\text{CCSD(T)}}$ correction obtained from an CCSD(T) aug-cc-pVTZ→aug-cc-pVQZ extrapolated calculation was very similar, but slightly preferable to a aug-cc-pVTZ$\delta_{\text{MP2}}^{\text{CCSD(T)}}$ correction. This is questioned in the later paper by Marshall et al.[80], as the CBS(aug-cc-pVTZ, aug-cc-pVQZ) itself is not always fully converged.[80] According to Marshall et al.[80], one would need a aug-cc-pVQZ→aug-cc-pV5Z extrapolated estimate to achieve a nearly converged CBS $\delta_{\text{MP2}}^{\text{CCSD(T)}}$. Something that is only possible to achieve for small systems. Based on their calculations on the dimers of water, formic acid and formamide Marshall et al.[80] states that a nice Pauling point[166] would be to use the aug-cc-pVTZ $\delta_{\text{MP2}}^{\text{CCSD(T)}}$ correction. Generally, Marshall et al.[80] recommend a single value correction over an extrapolated correction, as the extrapolations would need to be performed in very large basis sets. Burns et al.[85] claim that the CCSD(T) correction is less sensitive to the basis set used, and even whether or not the CP scheme is used, as long as the basis set aug-cc-pVTZ or higher is used. They also claim that the convergence beyond the aug-cc-pVTZ basis for the CCSD(T) correction is so slow that it is not worth it. Hence, showing that calculations on different systems might yield different conclusions.

The composite scheme has been tested against the Ar_2 systems for which the basis set limit has been reached[167] using the CCSD(T) method alone. This system is dispersion bound, and the $\delta E_{int}^{\mathrm{CCSD(T)}}$ correction is large and positive (17%)[82]. For this system it is found that the composite scheme performs worse than the CCSD(T) method alone with a medium basis set. Generally, if the CCSD(T) correction makes up a significant part of the interaction energy, it is advisable to perform the CCSD(T) calculation in a large basis set, otherwise the accuracy can be diminished.[80] Podeszwa et al.[82] thus also report the pure CCSD(T) values for some of the dispersion bound members of the S22 set with midbond functions included in the basis set. There are no great differences between the values from the composite schemes and the pure CCSD(T) method for the larger systems of the S22 set, but they report that this is something that should be further investigated.[82] They still claim that their values on the S22 test set are of benchmark quality, with an estimated error of only about 3%.[82]

Řezáč et al.[81] also attempt to investigate the effect of basis set and method used on the composite schemes by testing on the A24 test set[23]. They find that using extrapolated MP2 and the MP2-F12 methods gives very similar answers. The MP2-F12 outperforms the extrapolated MP2 for the smaller aug-cc-pVDZ basis, making the MP2-F12 aug-cc-pVDZ interaction corrected by the δCCSD(T) correction using the 6-31G**(0.25,0.15) basis set, especially economical and well behaved. They comment on the possibility of there being advantageous error cancellation when the small 6-31G**(0.25,0.15) basis set is used, and that they might not be universal. Other extrapolation schemes were also tested for use with a composite scheme, but no improvements to the X^{-3} scheme was found.

C Interaction Energies

Table C.1: Interaction energy of the optimised stacked structures from Figure 6.1

Figure	Figure 6.1a	Figure 6.1b	Figure 6.1c	Figure 6.1d
Interaction energy mE_h	-1.60	-13.01	-13.90	-13.45
Figure	Figure 6.4c	Figure 6.1f	Figure 6.4d	Figure 6.1h
Interaction energy mE_h	-15.38	-13.45	- 13.90	- 13.01

Table C.2: Interaction energy of the optimised stacked structures from Figure 6.4

Figure	Figure 6.4a	Figure 6.4b	Figure 6.4c	Figure 6.4d
Interaction energy mE_h	-4.45	-18.55	-16.48	-17.90

Table C.3: Interaction energy of the flat structures Figure 6.8 and 6.6

Figure	Figure 6.8a	Figure 6.8a	Figure 6.6a	Figure 6.6b
Interaction energy mE_h	-19.12	-5.26	-6.49	-7.99

D Core-levels After Radiation

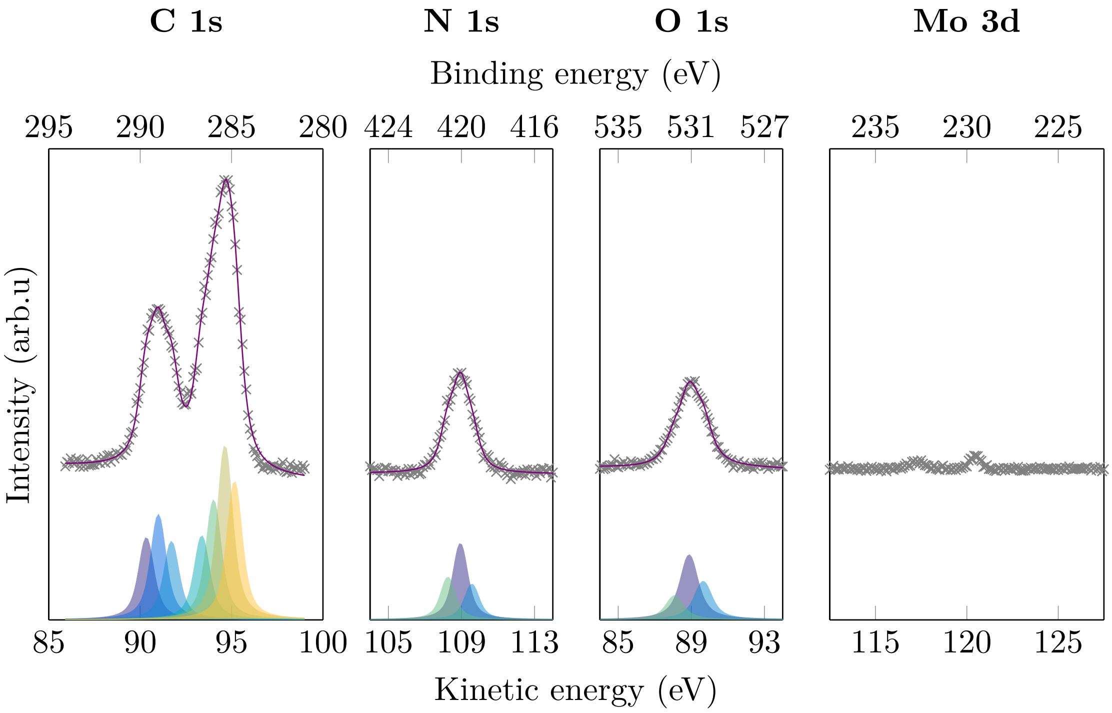


Figure D.1: Core-levels right after deposition. The figure is made by Jakob Vinje

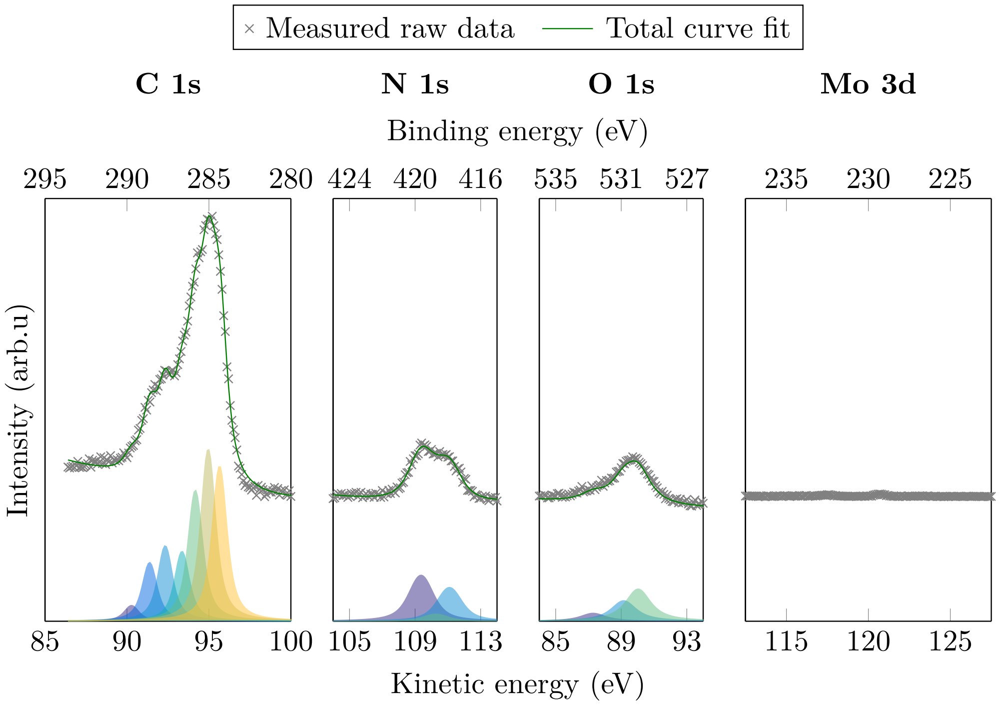


Figure D.2: Core-levels after 1 hour of light. The figure is made by Jakob Vinje

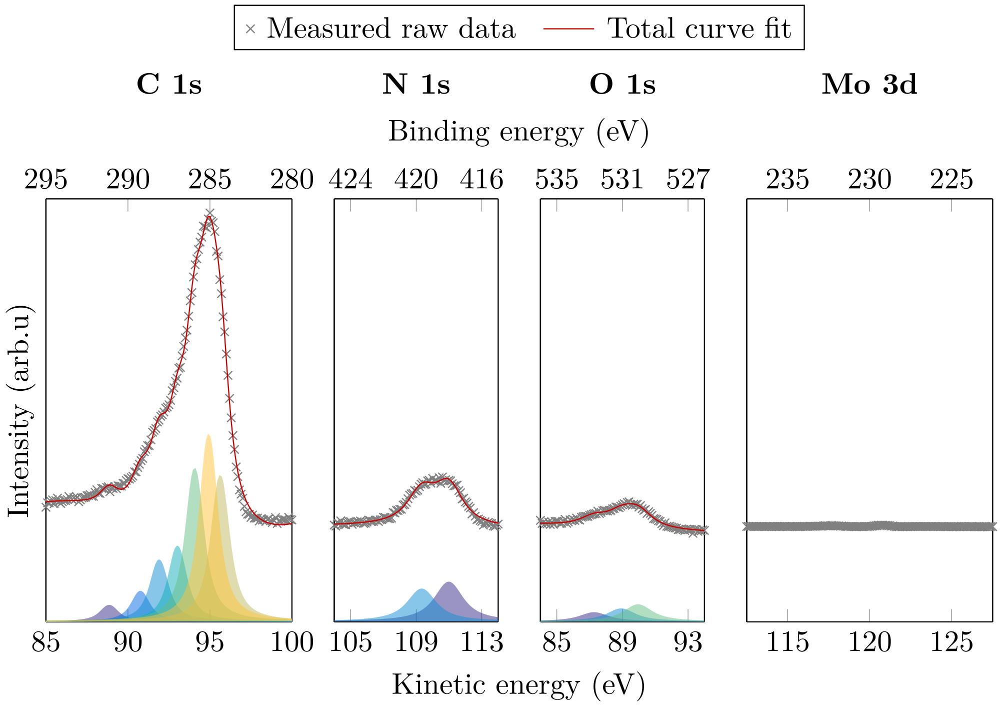


Figure D.3: Core-levels after 6.5 hour of light. The figure is made by Jakob Vinje