NTNU
Norwegian University of
Science and Technology

Using Localized Molecular Orbitals in the Multi-level Coupled Cluster Approach

Linda Stakvik

Chemistry
Submission date: May 2016
Supervisor: Ida-Marie Høyvik, IKJ
Co-supervisor: Henrik Koch, IKJ

Norwegian University of Science and Technology
Department of Chemistry

Abstract

In multi-level coupled cluster theory local orbitals have traditionally been generated using Cholesky decomposition of the Hartree-Fock density matrix. Orbitals obtained through Cholesky decomposition are not very local compared to state-of-the-art localized orbitals, and furthermore dependent on the locality of the Hartree-Fock density matrix. The objective in this master project is to make it possible to run MLCC calculations with any set of local orbitals, and analyze how the choice of orbitals impact the results. The idea is to generate localized orbitals in LSDALTON, in which a variety of localization schemes are already available, and use these orbitals for the multi-level coupled cluster calculation in DALTON. The scheme is tested by calculating excitation energies and total energies for various molecules, both conjugated systems and non-conjugated systems. The results obtained from the different orbital sets are expected to significantly improve the results using the multi-level coupled cluster approach. Results have shown that active spaces for non-conjugated systems can be chosen straight-forward, but active spaces for conjugated systems requires more insight.

Sammendrag

Denne oppgaven er en del av den pågående utviklingen av *multi level coupled cluster* (MLCC) teori ved institutt for kjemi, Norges Tekniske Naturvitenskapelige universitet, NTNU. I MLCC blir orbitalrommet delt opp i flere underrom, og om disse rommene behandles med forskjellige nivå i *Coupled cluster*(CC) hierarkiet, kan det spares mye i beregningstid uten å miste betydelig nøyaktighet. På grunn av at elektronkorrelasjon er en lokal egenskap er det mulig å kunne benytte lokaliserte orbital i et molekylært system. Når orbitalene er lokaliserte kan forskjellige deler av molekylet bli behandlet med ulike nivå av teori, uten å miste nøyaktighet for den interessante regionen.

Tradisjonelt er Cholesky dekomponering av Hartree-Fock tetthetsmatrisen blitt brukt for å generere lokale orbital til MLCC, men i denne oppgaven er det forsøkt å bruke både *powers of fourth central moment*(PFM), *powers of second central moment*(PSM) og Boys lokaliseringsfunksjon. Ved å bruke disse lokaliseringsfunksjonene i tillegg til den opprinnelige Cholesky dekomponeringen av tettheten har eksitasjonsenergiene for ulike molekyl blitt utregnet og sammenlignet. Total energi for molekyl, og dets avhengighet av størrelse og valg av orbitalrom er også blitt studert. Flere metoder for å velge orbitaler som skal inkluderes i det aktive rommet er også blitt testet ut.

Ved å regne ut eksitasjonsenergier for både konjugerte og ikke-konjugerte organiske molekyl har forskjellige lokaliseringsmetoder blitt utprøvd for ulike aktive rom. Resultatene fra disse utregningene viser at Cholesky dekomponering generelt inkluderer mange orbitaler i aktivt rom, og ikke gir betydelig forskjell i eksitasjonsenergi for ulike rom for konjugerte system. I stedet gir Cholesky dekomponering bedre resultater for ikke-konjugerte system, men inkluderer fortsatt mange orbitaler i det aktive rommet. De tre andre lokaliseringsfunksjonene gir energier som er nærmere referanseenergien og med færre orbital inkludert i det aktive rommet. Fra disse resultatene gir bruk av PFM, PSM og Boys lokaliseringsfunksjon resultater med bedre nøyaktighet.

Preface

Acknowledgements

First off I would like to thank my supervisor Associate Professor Ida-Marie Høyvik for regular meetings, support and interesting discussions. For showing such enthusiasm for the project when I didn't myself.

A special thanks to Ulla Eidissen Jensen, Jon Åge Stakvik and Karen Dundas for proof reading. Also big thanks to Jon Åge Stakvik for always trying to assist me with my issues with LaTeX, there have been a lot over the past two years.

I would also like to thank my friends, both at NTNU and other places, and I hope that we all can stay in touch even when we're spread all over the country(or the world).

And at last I would like to thank my parents, Sissel and Steinar. For always believing in me even when I haven't myself. You haven't always been able to understand my studies, but you have always been there for me. Without your love and support I would not be where I am today.

Miscellaneous

All orbital and molecule visualizations in this thesis have been made using the UCSF Chimera Package, developed by the Resource for Biocomputing, Visualization, and Informatics atthe University of California, San Fransisco [1].

Contents

Abbrevations

AO	**A**tomic **O**rbital
CC	**C**oupled **C**luster
cc	**c**orrelation **c**onsistent
CD	**C**holesky **D**ecomposition
CCSD	**C**oupled **C**luster **S**ingles and **D**oubles
CMO	**C**anonical **M**olecular **O**rbital
cs	**c**losed **s**hell
DZ	**D**ouble **Z**eta
ER	**E**dminston-**R**uedenberg
FCI	**F**ull **C**onfiguration **I**nteraction
GTO	**G**aussian **T**ype **O**rbital
HF	**H**artree-**F**ock
HOMO	**H**ighest energy **O**ccupied **M**olecular **O**rbital
LCC	**L**ocal **C**oupled **C**luster
LUMO	**L**owest energy **U**noccupied **M**olecular **O**rbital
MLCC	**M**ulti-**L**evel **C**oupled **C**luster
MLCC3	**M**ulti-**L**evel **CC3**
MO	**M**olecular **O**rbital
PAO	**P**rojected **A**tomic **O**rbital
PM	**P**ipek-**M**ezey
SCF	**S**elf-**C**onsistent **F**ield
STO	**S**later **T**ype **O**rbital
TZ	**T**riple **Z**eta

"Begin at the beginning,"
the King said, gravely,
"and go on till you come to the end;
then stop."

LEWIS CARROLL, Alice in Wonderland

Chapter 1

Introduction

The foundation of quantum chemistry is the Schrödinger equation, but it is impossible to solve the equation analytically for systems containing more than one electron. Thus the solution for molecular wave functions have to approximated. Coupled cluster(CC) is a popular method for electronic structure calculations[2–4], and generally produce quite accurate results. Because the method is size-extensive it is suitable for calculations on large molecules. The CC method is a post Hartree-Fock(HF) method. As HF only accounts for Fermi correlation, the objective of post-HF methods is to capture correlation effects. Electron correlation includes both short-range interaction of electrons that stems from their mutual repulsion, and dispersion effects, which are both local phenomena. Traditionally delocalized canonical HF-orbitals have been used in conventional CC, but use of these orbitals causes the computational cost of the CC method to scale steeply with increasing molecule size. An infinite choice of molecular orbitals are related by unitary transformations, the invariance of these properties can be exploited to generate a new set of HF orbitals that are local in space. When using local molecular orbitals (LMOs), the scaling problem can be counteracted, and the methods based on the use of LMOs are called local correlation methods.

The standard procedures that have been used to generate local orbitals is the Boys [5], Edminston-Ruedenberg (ER) [6] and Pipek-Mezey (PM) schemes [7]. However, these procedures have been proven impossible to localize virtual molecular orbital (MOs). Therefore, in the absence of local virtual HF orbitals, the set of projected atomic orbitals (PAOs) have been used instead. Jansík et al.[8] and Høyvik et al.[9] have shown that by using orbital localization functions with powers of the second (PSM) and fourth (PFM) central moment, respectively, it is possible to obtain a good locality for both occupied and virtual orbitals.

Various methods have been developed in the attempt to reduce the scaling wall of electron correlation methods, where molecular orbitals(MOs) have been localized [10–19]. The multi-

level coupled cluster (MLCC) formalism, developed by Myhre et al.[20], is an example of such a method. It is a framework that makes it possible to divide the molecular system in various subspaces and apply different levels of CC theory to these. For the multi-level CC3 (MLCC3)[21] model the active space of the molecule is treated with CC3[22, 23], while the rest of the molecule is treated with CCSD. The CC3 models scale as N^7. But when treating just a small part of the molecular system with CC3, the overall scaling becomes equal to the CCSD scaling, N^6, while the accuracy is retained as in CC3. Even when a significant part of the system is included in the active space the computational requirements can be reduced significantly.

In the MLCC formalism the localized orbitals have originally been obtained from Cholesky decomposition (CD) [24, 25]. The Cholesky orbitals can be computed non-iteratively and each orbital coincides with an atomic center. The Cholesky orbitals are however not very local, as opposed to orbitals generated by optimization of a localization function. The problem with using localized orbitals is that they are highly dependent on how the active space is chosen, therefore it is necessary to find a good way to chose the orbital space.

In this work a routine to read matrices from LSDALTON and using these for the MLCC calculation in DALTON [26] has been implemented. By using this routine the MLCC3 has been tested on orbitals localized using powers of the fourth and second central moment, and also using the Boys localization scheme on different systems. Potential energy surfaces of ferrocene have also been studied. Ferrocene is a complicated system, and the problem have been to create potential energy surfaces without ”bumps”, when using local correlation methods. To be able to describe these potential energy surfaces an analysis of how the orbital space is included had to be carried out. With the ultimate objective being to find a general method to include orbitals in active space and still maintain good results, without much dependence of user input.

Chapter 2 begins with an introduction to electronic structure theory, with a summary of HF and CC theory, followed by a discussion of different localization functions. Chapter 3 shows the outline of the different methods used for assigning orbitals into active spaces. The Results from the MLCC3 calculations are presented in Chapter 4, and are further discussed. Chapter 5 summaries and concludes the result and further work is discussed in Chapter 6

Chapter 2

Theory

2.1 The Schrödinger equation

The non-relativistic time-independent Schrödinger equation, within the Born-Oppenheimer approximation, can be expressed as

$$H|\Psi\rangle = E|\Psi\rangle \tag{2.1.1}$$

where $|\Psi\rangle$ is the exact electronic wave function describing the system, E is the energy, and H is the Hamilton operator(here represented in atomic units)

$$H = -\frac{1}{2}\sum_{i}^{N_e}\nabla_i^2 - \sum_{i}^{N_e}\sum_{I}^{N_n}\frac{Z_I}{|\boldsymbol{r}_i - \boldsymbol{R}_I|} + \frac{1}{2}\sum_{i\neq j}^{N_e}\frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} + \sum_{I\neq J}^{N_n}\frac{Z_I Z_J}{|\boldsymbol{R}_I - \boldsymbol{R}_J|} \tag{2.1.2}$$

where N_e is the number of electrons, N_n is the number of nuclei, Z_I is the atomic number of nucleus I, $\boldsymbol{R}_I$ is the coordinates of nucleus I and $\boldsymbol{r}_i$ is the coordinates of electron *i*. ∇_i^2 is the Laplace operator which is defined as $\left(\frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}\right)$.

The spin-free, non-relativistic Hamiltonian can be written in the second quantization formalism as

$$H = \sum_{pq} h_{pq}E_{pq} + \frac{1}{2}\sum_{pqrs} \mathrm{g}_{pqrs}e_{pqrs} + h_{\mathrm{nuc}} \tag{2.1.3}$$

where the indices *p, q,...* indicate unspecified orbital occupations, E_{pq} is the singlet excitation operator and h_{nuc} represents the repulsion between nuclei.

Only one-electron systems can be solved analytically, thus for larger systems it is most convenient to use approximate methods for solving these classes of problems. As approximate methods is unable to solve the problem accurately the solutions given will only have complete accuracy. However the exact solution for the given one-electron basis in N_e-electron Fock

space can be obtained by Full Configuration Interaction(FCI), which can be represented as a linear combination of all determinants

$$|\text{FCI}\rangle = \sum_{M=1}^{D} C_M |M\rangle \tag{2.1.4}$$

The expansion coefficients, C_M, are found from the variation principle. The number of determinants, D, has a factorial dependence of the number of spin orbitals, and this makes FCI impossible to use on anything but small systems in small basis sets. [2]

2.2 Hartree-Fock Theory

The original procedure of finding atomic orbitals numerically was introduced by Hartree, and was known as the self-consistent field (SCF) method. Fock and Slater further improved this method by including electron exchange effects, and the orbitals that are obtained by this method are called the Hartree-Fock orbitals [27]. In HF theory the wave function is written as a linear combination of Slater determinants, or configuration state functions(CSFs), and in restricted HF, the electronic state is described by a single CSF. The assumption behind Hartree Fock is that an electron is moving in a spherical average potential that is caused by the nucleus and the other electrons. Only closed-shell HF will be considered in this thesis, denoted by cs. The HF-method assumes that the exact wave function of the system can be approximated by a single determinant with fully occupied MOs

$$|\text{cs}\rangle = |\Phi\rangle \tag{2.2.1}$$

written in the second quantization formalism. To obtain the HF state the closed shell wave function can be parametrized in terms of orthogonal transformations among the MOs.

$$|\text{cs}(\boldsymbol{\kappa})\rangle = \exp(-\hat{\kappa})|\text{cs}\rangle \tag{2.2.2}$$

where $\hat{\kappa}$ is the anti-Hermitian one-electron operator

$$\hat{\kappa} = \sum_{p>q} \kappa_{pq}(E_{pq} - E_{qp}) = \sum_{p>q} \kappa_{pq} E^{-}_{pq} \tag{2.2.3}$$

Due to the presence of redundant parameters there are infinitely many ways to represent each stationary point of the energy, $E(\boldsymbol{\kappa})$

$$E(\boldsymbol{\kappa}) = \langle \text{cs}(\boldsymbol{\kappa})|H|\text{cs}(\boldsymbol{\kappa})\rangle \tag{2.2.4}$$

Minimizing the energy expression in Eq. 2.2.4 through the variational principle, will give the HF wave function. The value of the Hamiltonian is stationary for the HF state with respect to unitary variations in the MOs:

$$\delta E(\boldsymbol{\kappa}) = \delta\langle \mathrm{cs}(\boldsymbol{\kappa})|H|\mathrm{cs}(\boldsymbol{\kappa})\rangle = 0 \tag{2.2.5}$$

The solution to this equation, $\boldsymbol{\kappa}^{\mathrm{HF}}$, defines the HF state

$$|\mathrm{HF}\rangle = |\mathrm{cs}(\boldsymbol{\kappa}^{\mathrm{HF}})\rangle \tag{2.2.6}$$

As mentioned above, the redundant parameters give infinite possibilities for representing each stationary point $E(\boldsymbol{\kappa}) = \langle \mathrm{cs}(\boldsymbol{\kappa})|H|\mathrm{cs}(\boldsymbol{\kappa})\rangle$. The definition of a redundant parameter is that it is not needed for a general first-order transformation of the wave function, or equivalently that

$$E_{pq}|\mathrm{HF}\rangle = 0, \qquad p \neq q \tag{2.2.7}$$

For a closed-shell wave function non-redundant parameters are achieved when virtual and occupied orbital mix. The non-redundant HF conditions are

$$\langle \mathrm{cs}|[E_{ai}, H]|\mathrm{cs}\rangle = 0 \tag{2.2.8}$$

The optimization condition given in Eq. (2.2.8) is the same as requiring that the Fock matrix is block diagonal, meaning that off-diagonal elements are zero. The Fock matrix is defined as

$$F_{pq} = h_{pq} + \sum_i (2\mathrm{g}_{pqii} - \mathrm{g}_{piiq}) \tag{2.2.9}$$

In optimization schemes, the Fock matrix is usually diagonalized, and the MOs obtained from this diagonalization are called the canonical molecular orbitals (CMOs). This set of MOs is completely delocalized over the entire molecular system. The parameters that are redundant in the energy optimization, can however, be used in the localization of orbitals. Thus resulting in a block-diagonal Fock matrix, rather than diagonal.

In addition to being an approximation method in its own right, the HF approximation is an important starting point for more accurate approximations, that include the effects of electron correlation, e.g. the Coupled Cluster method which is explained in more detail in section 2.4. [2, 28]

2.3 Electron correlation

The difference between the exact energy, i.e., energy obtained by Full Configuration Interaction (FCI) in given basis, and the HF energy is called the electron correlation energy as shown in Eq. 2.3.1. This energy corresponds to the motion of correlated electrons [29].

$$E_{\text{corr}} = E_{\text{exact}} - E_{\text{HF}} \tag{2.3.1}$$

Due to the fact that electrons have the same electronic charge, they will have a mutual repulsion, or in other words, correlated motion. The HF energy will be higher than the exact energy because HF is a variational method, and due to this the correlation energy will be negative, $E_{\text{corr}} < 0$ [30].

Electron correlation can be separated into dynamic and static correlation. Dynamic correlation is associated with capturing the effect of the instantaneous electron repulsion, such as for those electrons occupying the same spatial orbital and thus has opposite spin. Static correlation is associated with small differences in energy between different states and more than one determinant is required to cover the electronic structure. Multi-Configurational Self Consistent Field (MCSCF), or other multi-reference methods, are needed to describe static correlation [31].

2.4 The Coupled Cluster method

As mentioned in Section 2.2, an important post HF method is CC. It alleviates some of the problems in the HF method due to electron correlation as demonstrated in eq. 2.4.1. However the CC method is only able to give a good description of dynamic correlation and the HF state is used as a reference state to include static correlation. It is therefore necessary that the HF state is a good description of the system. CC is a size-extensive method, i.e, energies calculated vary linearly with the number of particles when the system size increases. Unlike the HF method the CC method is not a variational method, and as a consequence the electronic energy obtained may be lower than the true energy. [28].

$$E_{\text{CC}} = E_{\text{HF}} + E_{\text{corr}} \tag{2.4.1}$$

The wave function for the CC method can be written as follows

$$|\text{CC}\rangle = \exp(\hat{T})|\text{HF}\rangle \tag{2.4.2}$$

where the exponential operator $\exp(\hat{T})$ is defined by the Taylor series expansion, this expression is known as the exponential ansatz.

$$\exp(\hat{T}) = 1 + \hat{T} + \frac{1}{2!}\hat{T}^2 + \frac{1}{3!}\hat{T}^3 + \cdots = \sum_{k=0}^{\infty} \frac{1}{k!}\hat{T}^k \tag{2.4.3}$$

The Cluster operator, $\hat{T}$ is defined as

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \cdots + \hat{T}_e = \sum_{n=1}^{N_e} \hat{T}_n \tag{2.4.4}$$

The cluster operator is a sum of excitation operators up to an N_e-fold excitation for an N_e-electron system. $\hat{T}_1$ denotes the one-electron excitation operator, $\hat{T}_2$ denotes the two-electron operator, and so on. These excitation operators are given by the following equation

$$\hat{T}_1 = \sum_{ia} t_i^a E_{ai} \qquad \hat{T}_2 = \frac{1}{2} \sum_{\substack{ij \\ ab}} t_{ij}^{ab} E_{ai} E_{bj} \tag{2.4.5}$$

where t_i^a and t_{ij}^{ab} are the singles and doubles amplitudes, respectively. E_{pq} is the singlet excitation operator. It is possible to make the CC wave function equivalent to the FCI wave function if all the excitations in the cluster operator (Eq. 2.4.4) are retained. The coupled-cluster wave function, for a given orbital basis, satisfies the Schrödinger equation

$$H \exp(\hat{T})|\mathrm{HF}\rangle = E \exp(\hat{T})|\mathrm{HF}\rangle \tag{2.4.6}$$

As CC is not a variational method, the energy and amplitudes in the CC method are determined rather by subspace projections than variationally. Which is done by multiplying Eq. 2.4.6 from the left with $\exp(-\hat{T})$. The subspace projections are against the HF state, $\langle \mathrm{HF}|$, and the excited state determinants, $\langle \mu|$, which will give the following equations for the energy and the amplitudes

$$\langle \mathrm{HF}| \exp(-\hat{T}) H \exp(\hat{T})|\mathrm{HF}\rangle = E \tag{2.4.7}$$

$$\langle \mu| \exp(-\hat{T}) H \exp(\hat{T})|\mathrm{HF}\rangle = 0 \tag{2.4.8}$$

A common approximation in CC theory is to truncate the cluster operator $\hat{T}$ to include only certain types of terms. E.g. in the approach referred to as "coupled cluster singles and doubles" (CCSD), $\hat{T}$ is truncated after $\hat{T}_2$, in coupled cluster doubles (CCD), only $\hat{T}_2$ is used, and in the approach coupled cluster singles, doubles and triples (CCSDT), the $\hat{T}$ is

truncated after $\hat{T}_3$ [2].

CCSD combined with an approximate, non-iterative treatment of the excitations, called CCSD(T) [32], is often referred to as the "gold standard" of computational chemistry because of the high accuracy obtained compared to computational cost [33]. In CCSD(T) the triples are calculated using perturbation theory. This method is quite popular due to relatively low computation cost and that it is an accurate approximation to CCSDT. CCSD(T) also gives very accurate equilibrium structures and equilibrium properties [34, 35]. Higher orders of the CC hierarchy have also been developed, like CCSDTQ [36] and CCSDTQ56 [37].

The conventional CC models are expressed in terms of the completely delocalized CMOs, and this results in a steep computational scaling of these methods, making them intractable for anything but small molecular systems. However, as electron correlation effects are an example of a local phenomena, the scaling wall may be bypassed by expressing the correlated wave function in a set of local HF orbitals.

The groundwork for local correlation methods were laid by Pulay and Sæbø[10], and an important early contribution is the local coupled cluster method presented by Werner et al.[11–14]. Other methods to reduce the scaling in correlation methods include atomic orbital based CC[15, 16], the natural linear scaling approach[17], and local CC using bump functions[18, 19]. Most of these methods use local occupied HF orbitals, but none of them have been able to apply local virtual HF orbitals, as they have previously been impossible to generate and as a result projected atomic orbitals(PAOs) have been used instead. Another alternative to local correlation methods have been to parallelize the CC method over multiple nodes[38–40]. This have been made possible due to advancements in supercomputers over the last years and is a useful tool, but this technique does not overcome the scaling obstacle.

2.4.1 CC3

The CC3 model, like the CCSD(T) model, is an approximation to the CCSDT model. CC3 shows a reduction in scaling compared to the CCSDT (N^7 as opposed to N^8). The definition of the CCSDT state is

$$|\text{CCSDT}\rangle = \exp(\hat{T}_1 + \hat{T}_2 + \hat{T}_3)|\text{HF}\rangle \tag{2.4.9}$$

The cluster amplitudes are determined by projecting the Schrödinger equation onto the space of single, double and triple excitations from the HF reference state. In order to obtain the CCSDT state the complete set of equations must be solved

$$\langle \mu_i | \exp(-\hat{T}_1 - \hat{T}_2 - \hat{T}_3) H \exp(\hat{T}_1 + \hat{T}_2 + \hat{T}_3)|\text{HF}\rangle = 0 \tag{2.4.10}$$

Introducing the modified two-electron Hamiltonian, $\hat{H}$, as

$$\hat{H} = \exp(-\hat{T}_1) H \exp(T_1) \tag{2.4.11}$$

The singles and doubles equation for both CC3 and CCSDT can be written as

$$\langle \mu_1 | [\hat{H}, \hat{T}_2] | \mathrm{HF} \rangle + \langle \mu_1 | [\hat{H}, \hat{T}_3] | \mathrm{HF} \rangle = 0 \tag{2.4.12}$$

$$\langle \mu_2 | \hat{H} + [\hat{H}, \hat{T}_2] + \frac{1}{2} [[\hat{H}, \hat{T}_2], \hat{T}_2] | \mathrm{HF} \rangle + \langle \mu_2 | [\hat{H}, \hat{T}_3] | \mathrm{HF} \rangle = 0 \tag{2.4.13}$$

where $\hat{T}_1$ transformed operators are defined through

$$\hat{O} = \exp(-\hat{T}_1) O \exp(\hat{T}_1) \tag{2.4.14}$$

Equations 2.4.12, 2.4.13 and 2.4.14 define the CC3 energy in a system where there are no external perturbations. The singles and doubles excitations are included to an infinite order and the triples excitations are treated perturbatively.

The Hamiltonian can be divided into two terms, The Fock operator ($\hat{F}$) and the fluctuation potential ($\hat{\Phi}$).

$$H = \hat{F} + \hat{\Phi} + h_{nuc} \tag{2.4.15}$$

In CC3 there are no approximations to the treatment of single excitations, this is due to the fact that they act as orbital relaxation parameters. This results in that the singles amplitude is treated as zeroth order in the fluctuation potential, while the triple excitations are treated correctly to second order. The CC3 model is a good approximation to the CCSDT model, as the computational scaling of CCSDT and CC3 is N^8 and N^7, as mentioned above. It can be seen that the CC3 model has a reduction in scaling from the CCSDT model, and scales to the same order as CCSD(T). CC3 also has the same accuracy and robustness as the CCSD(T). The difference is however that CC3 is better for calculations of time-independent properties [22, 23].

2.4.2 Multi-level Coupled Cluster

A variant of a local CC method is Multi-level Coupled Cluster (MLCC) where the orbital space is divided into different subspaces, and each of these subspaces is treated with different levels of the CC hierarchy. This makes it possible to reduce computational cost by treating the most important part of the system with a higher level of accuracy. One way of partitioning the system is to divide it into active and inactive space. The easiest way to do this is by assigning the highest energy occupied molecular orbital (HOMO) and the lowest energy

unoccupied molecular orbital (LUMO) to the active space. Other procedures to assign active and inactive spaces is by Cholesky decomposition, or multiple methods that have been tested in this thesis. MLCC treats the active part of the molecular system with a higher level of CC theory, while the rest of the molecule is treated with a more approximative method [20]. For MLCC3 the active space is treated with CC3 and the inactive part is treated with CCSD. As the MLCC3 method treats only a small part of the molecule with CC3, and the remaining with CCSD, the overall scaling will be equal to CCSD, but the accuracy will be comparable to that of CC3 [21].

MLCC is similar to active space CC by Olsen and Köhn[41, 42], but in comparison to this method MLCC can include several levels of theory which makes it possible to gradually increase accuracy. In MLCC the Cluster operator, $\hat{T}$, is divided into $\hat{X} + \hat{S}$, which gives the following CC-equation

$$|\mathrm{CC}\rangle = \exp(\hat{T})|\mathrm{HF}\rangle = \exp(\hat{X} + \hat{S})|\mathrm{HF}\rangle \tag{2.4.16}$$

Each of the operators, $\hat{X}$ and $\hat{S}$, are associated with the subspace projections $\mu^{\hat{X}}$ and $\mu^{\hat{S}}$, in the same fashion as in eq. 2.4.8. The subspace projections against $\mu^{\hat{S}}$ are solved perturbatively for the amplitudes in the $\hat{S}$ operator, and the amplitudes for $\hat{X}$ are determined without any approximations.

If the excitations only involves active orbitals, they are considered internal(I) and are included in $\hat{X}$. Excitations involving active and inactive orbitals are labeled semi-external(SE), and the excitations involving only inactive orbitals are labeled external(E). Both the semi-external and the external excitations are included in $\hat{S}$. The different excitations are illustrated in Figure 2.4.1

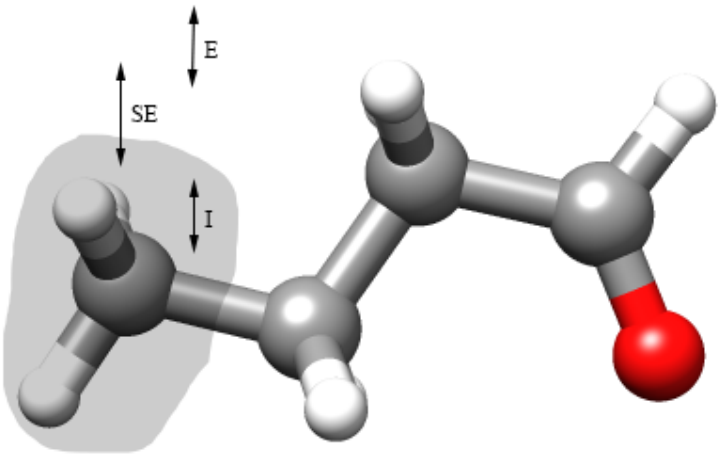


Figure 2.4.1: Example of an active space for butanal, with illustrations for internal(I), semi-external(SE) and external(E) excitations

It is possible to exploit the fact that the electron correlation is a local property by combining MLCC with localized orbitals, either by using Cholesky orbitals or orbitals obtained by optimizing localization functions, such as those mentioned in Section 2.6.1. The advantage of Cholesky orbitals are that they can be computed without optimization [20, 21], the downside of Cholesky orbitals is that they are not as local as orbitals obtained by using explicit localization through optimizing an orbital localization function.

2.5 Basis set

One of the approximations ingrained in almost all ab initio calculation methods is the use of basis sets. It will not be an approximation if an unknown function, e.g. and MO, is expanded in a set of known functions, while having a complete basis set. However, because the complete basis consists of an infinite number of functions, this will be impossible to achieve in practice. It is common to represent an MO as a function in an infinite coordinate system that is spanned by a complete basis. And if a finite basis set is used, the only components of the MO that are represented are those along the axes.

Especially two types of basis sets are most commonly used in ab initio calculations, Slater Type Orbitals(STOs) and Gaussian Type Orbitals(GTOs). The STOs do not have any radial nodes, these only occur when linear combinations of STOs are created. They are also fast convergent due to the exponential dependence. The drawback of STOs is that they cannot be used for calculations on three- and four-center two-electron integrals. Due to this they are most often used for atomic and diatomic systems where high accuracy is necessary. GTOs have a r^2 dependence in the exponent which is the cause of two drawbacks. Firstly, they have a zero slope at the nucleus, where the STOs have a cusp, and this causes a problem in the representation of proper behavior close to the nucleus. Secondly, the GTOs also fall off too rapidly far from the nucleus which leads to a poor representation of the tail. These drawbacks of the GTOs are the reason that more of them are needed to achieve the same level of accuracy as with STOs. A rule of thumb is that three GTOs are needed for each STO. However, even if more GTOs are needed for a calculation their integrals are fairly easy to compute. For this reason the GTOs are preferred as basis functions in electronic structure calculations.

2.5.1 Classification of Basis sets

A minimal basis set contains just enough functions that are required to contain all the electrons of neutral atoms, i.e., one basis function is used to describe each atomic orbital.

To improve the basis set further, the basis functions can be doubled, and this will produce a Double Zeta(DZ) type basis. For a Triple Zeta(TZ) basis set, the basis functions are tripled, and so on. Another possibility is to only double the valence functions and keep the core functions in minimal basis, which will create a split valence basis set. Higher angular momentum is vital if electron correlation methods are being used. As stated in Section 2.3, there are two types of electron correlation. To describe the radial correlation, the basis set needs functions of the same type, but with different exponents. While the basis set needs functions with exponents of the same magnitude, but with different angular momentum, to describe angular correlation.

2.5.2 Correlation consistent basis sets

The goal of correlation consistent (cc) basis sets, developed by Dunning et al. [43], is to recover the correlation energy of the valence electrons. These basis sets are known by the acronym cc-pVXZ (X = D,T,Q,5), for a polarized X-zeta Gaussian basis set. The basis sets are designed to include functions that contribute similar amounts of correlation energy at the same stage, independent of function type. For the cc-basis sets the error in energy from the sp-basis should be comparable with the error in correlation energy that is caused when the polarization space is incomplete, and as the polarization space is increased the polarization space has to be extended accordingly. The cc-basis sets can also be optimized with augmented functions, which are indicated by adding aug- to the beginning of the prefix. The basis set is then augmented with one extra function that has a smaller exponent for each angular momentum. And finally, it it possible to add functions with large exponents, known as tight functions, if the goal is to recover core-core and core-valence correlation, the acronym will then be cc-pCVXZ (X = D,T,Q,5). [29, 30, 44]

2.6 Orbital localization

CMOs are used in correlated calculations, and because these MOs are completely delocalized the scaling will become big for standard electronic correlation calculations on large molecules. To overcome this scaling problem, the standard has been to localize the canonical HF orbitals and by this exploiting the locality of a large molecular system [45]. As stated in Section 2.2, there are an infinite amount of MOs describing the same HF wave function, which are related by unitary transformations. The goal is to exploit these MOs to make HF orbitals that are local in space, and then use these local orbitals to reduce the scaling in CC methods [46]. Localized orbitals are determined by performing rotations among the occupied and virtual HF orbitals by optimizing a localization function and exploiting that the HF energy

is invariant with respect to such rotations [47]. One way to optimize a localization function is by using the trust region minimization method. This method is defined where the function to be optimized is approximated by a quadratic function [48]. The standard way of localizing orbitals have been by using either the Boys, Edminston-Ruedenberg or Pipek-Mezey scheme [8].

2.6.1 Localization functions

From early on the localization functions have been focused on localizing the occupied MOs, as local virtual MOs have been impossible to generate. The occupied orbitals describe the electronic density and therefore contain information on the localizability of electrons and bonding properties. In this section the standard localization procedures Boys, Edminston-Ruedenberg and Pipek-Mezey scheme will be presented. Furthermore, more recent methods like powers of the fourth and second central moment are also presented.

For the Boys scheme [5] the orbitals' second central moment is minimized, which is done by maximizing the distance between the centers of the orbitals. This is achieved by maximizing the function

$$\xi^{\text{Boys}} = \sum_p \sum_x \langle p|\hat{x}|p\rangle^2 \tag{2.6.1}$$

where $|p\rangle$ refers to a set of occupied MOs and $\hat{x}$ is a component of the position operator. The Edminston-Ruedenberg (ER) localization scheme [6] maximizes the sum of the self-repulsion energies of the orbitals

$$\xi^{\text{ER}} = \sum_p (pp|pp) \tag{2.6.2}$$

where

$$(pp|pp) = \int \varphi_p^*(\boldsymbol{r}_1)\varphi_p(\boldsymbol{r}_1)\frac{1}{r_{12}}\varphi_p^*(\boldsymbol{r}_2)\varphi_p(\boldsymbol{r}_2)d\boldsymbol{r}_1 d\boldsymbol{r}_2 \tag{2.6.3}$$

The ER orbitals have however not been used for large molecular systems [49]. The Pipek-Mezey (PM) scheme is based on a measure of how many atomic centers there are in a MO, and it is a sum of squared Mulliken charges, for both occupied and virtual orbitals. The function is then maximized to get local orbitals.

$$\xi^{\text{PM}} = \sum_p \sum_A |\langle p|\hat{P}_A|p\rangle|^2 \tag{2.6.4}$$

where $\hat{P}_A$ is a projection operator that projects onto the atomic orbital space centered on atom A.

All of the three procedures mentioned above are formulated as optimization problems. A localization functional is maximized with respect to rotations among the occupied orbitals,

thus making the orbital localization into an iterative procedure, traditionally done by Jacobi sweeps. This method have been effective for localizing occupied orbitals, but is not able to generate local virtual orbitals.

The powers of the second central moment (PSM) function is obtained by using powers of the Boys localization function:

$$\xi_m^{\mathrm{SM}} = \sum_p \langle p|(\hat{\boldsymbol{r}} - \bar{\boldsymbol{r}}_p)^2|p\rangle^m \tag{2.6.5}$$

where m is an integer, $\hat{\boldsymbol{r}}$ contains the $\hat{x}$, $\hat{y}$, and $\hat{z}$ Cartesian coordinates, and $\bar{\boldsymbol{r}}_p = \langle p|\hat{\boldsymbol{r}}|p\rangle$. For m = 1 in Eq. 2.6.5 the Boys localization function is recovered. The PSM of an orbital is often referred to as the variance of the orbital. The goal of the powers of the fourth central moment (PFM) localization function is to reduce the orbital tails, and thus achieving a better spatial locality. The PFM localization function to the second power is defined as

$$\xi_m^{\mathrm{FM}} = \sum_p \langle p|(\hat{\boldsymbol{r}} - \bar{\boldsymbol{r}}_p)^4|p\rangle^2 \tag{2.6.6}$$

The trust region minimization method, in comparison to Jacobi sweeps, have been able to produce local virtual orbitals, and the local virtual orbitals obtained from this method are more local than the PAOs that have traditionally been used to localize virtual orbitals [8].

The one-electron density matrix is a sparse matrix in the AO basis and it defines the interaction region between the AOs, and by this it indirectly also defines the locality of MOs. Exploiting the fact that the density matrix is positive semidefinite, where the rank of the matrix is equal to the number of occupied orbitals, the local MOs can be defined by using Cholesky decomposition on this matrix [24]. The Cholesky orbitals are discussed in more detail in section 2.6.2.

2.6.2 Cholesky Orbitals

Cholesky orbitals are obtained by decomposition of the one-electron density matrix in the AO basis for both occupied and virtual orbitals

$$D_{\mu\nu} = \sum_i^{occ} C_{\mu i} C_{\nu i} \tag{2.6.7}$$

$$D^V_{\mu\nu} = \sum_a^{occ} C_{\mu a} C_{\nu a} \tag{2.6.8}$$

where the $\mathbf{C}$ matrix contains the MO coefficients, i and a label the occupied and the virtual orbitals, respectively. And μ and ν label the AOs. The sum of the matrices $\mathbf{D}$ and $\mathbf{D}^V$ are equal to the inverse of the overlap matrix $\mathbf{S}$.

$$\mathbf{D} + \mathbf{D}^V = \mathbf{S}^{-1} \tag{2.6.9}$$

This makes it possible to create the pseudo density matrix, $\mathbf{D}^V$, without explicit knowledge of the orbitals.

If the density matrix is positive definite this would result in N Cholesky vectors, but density matrices are almost always positive semidefinite. Decomposition of all diagonal elements over a defined threshold is performed, and ended when no remaining elements are larger than this threshold. This will result in less than N Cholesky vectors for a positive semidefinite matrix.

The Cholesky decomposition gives a positive semidefinite matrix with a corresponding rank of the number of occupied orbitals and the number of virtual orbitals. The matrices are further decomposed into a lower-triangular matrix, $\mathbf{L}$, and its transpose $\mathbf{L}^T$

$$\mathbf{D} = \mathbf{L}\mathbf{L}^T \tag{2.6.10}$$

The density matrix is positive semidefinite and symmetric for real orbitals, which makes the Cholesky decomposition well defined. However, the Cholesky decomposition is not unique for $\mathbf{D}$ and $\mathbf{D}^V$. Different decomposition can also be done depending on the pivoting scheme, i.e., in which order the diagonals are chosen. The pivoting is done to avoid mathematical instability, and the largest diagonal elements are chosen for pivoting.

Active Cholesky orbitals are chosen by restricting the decomposition to the diagonal elements, and the active parts are not fully decomposed. Only the elements over a given threshold are decomposed, and the same is the case for the inactive part, with the difference that the threshold for the inactive part is about a 100th of the active threshold. Each of the diagonal elements in the density matrix corresponds to an AO centered on an atom. To generate the localized active space atoms of interest are set as active. The decomposition is then executed on the diagonal elements that corresponds to the AO centered on atoms set as active. The elements are then pivoted until there are no diagonal elements larger than the set threshold.

The Cholesky decomposition is a fast and numerically stable algorithm. It scales linearly for matrices that have a linear scaling of non-zero elements [50]. The decomposition is not iterative, this means that there is no need for complicated optimization techniques. It is also not necessary to give initial orbitals, which means that the method is suitable for determining

local MOs from the density matrix. The disadvantage of Cholesky decomposition is the possibility of generating unbalanced sets. This is caused by the fact that orbitals that describe bonds between atoms may be designated entirely to one atom. [20, 24, 25, 51].

Chapter 3

Implementation and calculations

This chapter presents the key modifications of the DALTON software package [26] to make it possible to run MLCC calculations with any set of local orbitals. The software package includes two different excecutables, DALTON and LSDALTON. DALTON includes a powerful tool to calculate a range of molecular properties at different levels of theory, while LSDALTON includes a linear-scaling HF and Density Functional Theory code for large molecular systems [26]. The objective of this routine was to be able to use localization schemes from LSDALTON in DALTON. The implemented routine was then used for calculations of excitation energies and toal energies for various molecules.

3.1 Using MLCC for any sets of local orbitals

The DALTON 2013 software package already had an existing code for MLCC, but the current localzation scheme in DALTON can only localize occupied orbitals. A localization scheme that also localizes virtual orbitals is desireable. This was the motivation for implementing a code that could take orbitals localized in LSDALTON, where there are a lot of localization schemes, and use these for the MLCC calculation in to DALTON.

The first thing that had to be done was to make a routine that could read the orbital matrix from LSDALTON in DALTON. The existing code in DALTON is in the AO basis, therefore the code had to be consistent with this basis, and it had to be ensured that everything was correct in the MO basis to be used in the new basis. To ensure that the MLCC are given all the information that is necessary, it has to be made sure that the Fock matrix for every room (this means active occupied, active virtual, inactive occupied and inactive virtual) are transformed to the local MO basis. The MO coefficient matrix for each space should then be $C_{\text{orb}} = C_{\text{local}}U$, where U diagonalizes the Fock matrix in the basis of C_{local} for each

space.

How the local orbitals are assigned to active spaces can be complicated. For Cholesky decomposition, this comes quite naturally from the method, as described in Section 2.6.2. Five other methods were evaluated in the scope of this work, which will be explained in more detail. An overview of these five methods are presented in Table 3.1.1.

To be able to use any set of orbitals in MLCC a routine to read matrices from file was implemented. The calculations performed in this work have been MLCC3 for the excitation energies and MLCCSD(T) for the total energy. MLCC3 treats the active space of the molecule with CC3, and the inactive space with CCSD. MLCCSD(T) also treats the inactive space with CCSD, but the active space is treated with CCSD(T).

Cholesky decomposition

The standard way of choosing active orbitals in the MLCC code was by Cholesky decomposition. In Cholesky decomposition the diagonal elements are decomposed as explained in Section 2.6.2.

Method 2

The first attempt of choosing active spaces was by looking at the orbital coefficients from the HF-calculation, and assign all orbitals with a coefficient over a certain threshold to the active space. Method 2 was considered because it seemed to be the most simple and naive approach. The method was soon discarded because the orbital coefficients were problematic to control, and no results for this method will be included.

Method 3

One new way of assigning orbitals to the active space was by looking at the orbitals centers. The core and virtual orbitals was assigned one atom center, while the valence orbitals were assigned two atom centers.

The spatial location of each orbital, ϕ_p, was found by calculating $\mathbf{r}_p$

$$\mathbf{r}_p = \langle p|\mathbf{r}|p\rangle \tag{3.1.1}$$

which assigns the orbital to the atom located closest in space. Valence orbitals are shared between two atoms since they represent bonding between the two atoms, and valence orbitals

are therefore assigned two atomic centers. Core and virtual orbitals are only assigned one atomic center.

Method 4

Method 4 consists of Method 3 and plus laso including additional orbitals by adding orbitals with an orbitals spread higher than a given value. The orbital spread, σ_2^p, of an orbital $|p\rangle$, in three dimensions, is defined as

$$\sigma_2^p = \sqrt{\langle p|\mathbf{r}^2|p\rangle - \langle p|\mathbf{r}|p\rangle^2} \tag{3.1.2}$$

All obitals with an orbital spread over the average are included in the active space.

Method 5

Finally, a fifth way of assigning orbitals to active space was to include a certain number of occupied and virtual orbitals. This mehtod was based on the degree of contribution of each orbital to the AO density elements for the indexes that belonged to the basis functions of the active atoms. As described in the following equation

$$D_{\mu\nu}^{\mathrm{AO}} = \sum_{i}^{\mathrm{occ}} C_{\mu i} C_{\nu i} = \sum_{i}^{\mathrm{occ}} D_{\mu\nu}^{i} \tag{3.1.3}$$

where the last term in this equation indicates the contribution of the density elements from orbital *i*. When given a number of active orbitals method 5 will include those orbitals with the highest contribution to the AO density elements.

$$\sum_{\mu \in \mathrm{active}} D_{\mu\mu}^{i} \tag{3.1.4}$$

Table 3.1.1: Overview of the five different methods for assigning orbitals to active space considered in this thesis

Method 1	Cholesky decomposition diagonals are decomposed
Method 2	Assign active space for $C_{\mu i} >$ threshold
Method 3	Assign active space by orbital centers core, virtual assigned 1 atom center valence assigned 2 atom centers
Method 4	Method 3 plus including additional orbitals by adding orbitals with σ_2 higher than a given value
Method 5	Include certain number of occupied and virtual orbitals based on orbitals contribution to AO density elements for indexes belonging to basis functions of active atoms

Chapter 4

Results and Discussion

4.1 Excitation energies

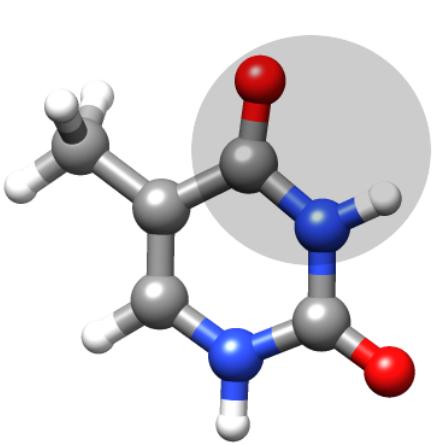
Excitation energies are obtained using the MLCC3 framework in DALTON, and the orbital localization used for the PFM, PSM and Boys scheme, as described in Section 2.6.1, are executed in LSDALTON, while the local orbitals generated using Cholesky decomposition, as described in Section 2.6.2, were from DALTON. Electronic excitations are highly localized, and this fact make them highly suitable for local electron correlation calculations. Since the excitations inspected in this thesis are from core to valence, these energies are calculated in the core-valence separated coupled cluster framework, which is described in more detail by Coriani et al. [52]. All calculations for the excitation energies are done with Method 3, from Table 3.1.1, of assigning orbitals to the active space.

4.1.1 Thymine

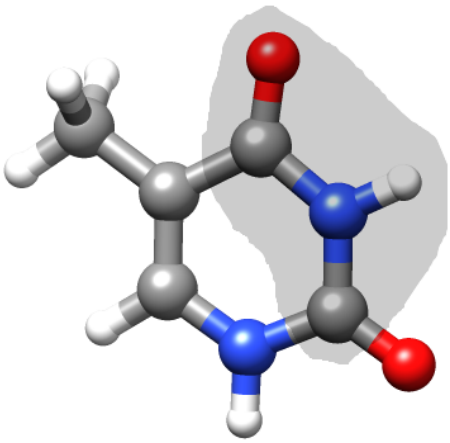
The structure for thymine was optimized using CCSD(T) with the basis set aug-cc-pVDZ for all atoms except oxygen where aug-cc-pCVDZ was used, and these basis sets are also used in the calculations.

The excitation energy for the oxygen closest to the methyl group was calculated using MLCC3. The full CC3 excitation energy was found to be 532.960 eV. The excitation energy for thymine was calculated for various active spaces, as illustrated in Figure 4.1.1. Where the number of active atoms ranging from 4 to 11 atoms. The motivation for investigating so many active spaces for the thymine was to check if how the active space was chosen would impact the result of the calculation. Furthermore, a number of large active spaces were investigated for thymine, this was done because thymine is a conjugated system, and it was

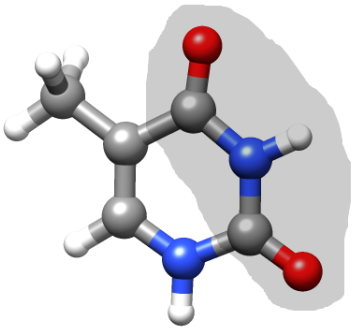
interesting to see how the conjugation of the molecule impacted the final results.



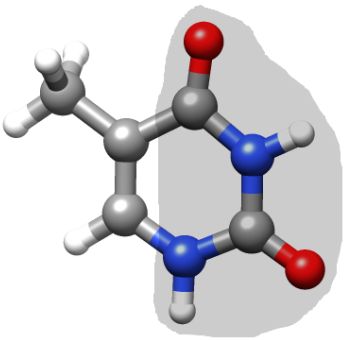
(a) 4 atoms in active space



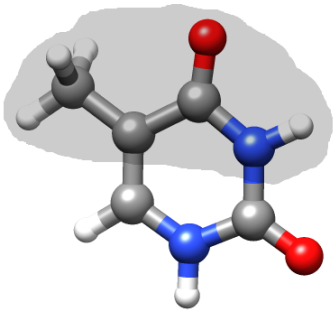
(b) 5 atoms in active space



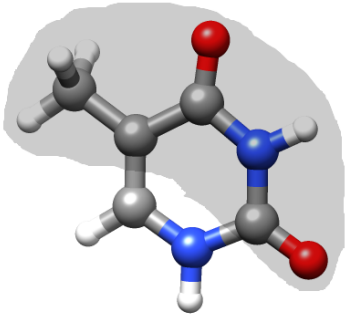
(c) 6 atoms in active space



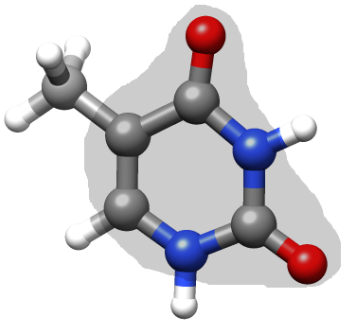
(d) 8 atoms in active space



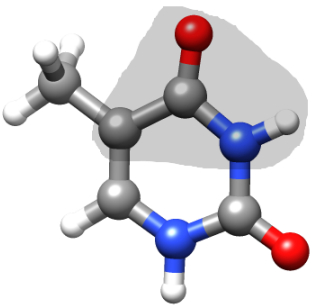
(e) 9 atoms in active space



(f) 11 atoms in active space



(g) 8 atoms in active space



(h) Space e minus the methyl group

Figure 4.1.1: Thymine molecule with active atoms highlighted

Table 4.1.1 shows the calculation of the excitation energies for thymine with local orbitals generated by Cholesky decomposition, or localized using the PFM, PSM and Boys local-

ization functions for eight different active spaces, as shown in Figure 4.1.1. The table also shows the number of orbitals that are in the active space, for both occupied and virtual, and also the difference in excitation energy from full CC3, denoted ΔCC3.

We first consider the Cholesky decomposition results of thymine, as illustrated in the top section of Table 4.1.1. From these results it can be seen that the deviation from full CC3 does not vary considerably when the active space is increased. To investigate if these deviations are acceptable, the CC3 contribution was calculated, illustrated in Table 4.1.2. These contributions are calculated by subtracting the CC3 excitation energy from the CCSD energy. Comparing the CC3 contribution with the CC3 deviation, shows that they are not significantly different. For example, for space b the deviation from full CC3 is 1.352 eV, while the CC3 contribution of the energy is 1.180 eV. Therefore, the MLCC3 results from where the active orbitals are generated with the Cholesky decomposition are not acceptable at all for thymine. To get results that are acceptable, there has to be a significant difference between the CC3 deviation and the CC3 contribution to the energy. This can be linked towards that the locality of the Cholesky orbitals are dependent of the locality of the HF density matrix. The best result is obtained for space g, where the CC3 deviation is 1.082 eV and the CC3 contribution is 1.422 eV. However, this is still not an acceptable result due to the difference between the two still being small.

The three other localization schemes are discussed together, as they follow a similar trend. For the PFM, PSM and Boys scheme the deviation from full CC3 shows a decreasing trend for increasing active space. By using the local orbitals from the PFM scheme it can be seen a significant drop in the CC3 deviation from space d to e of about 0.4 eV. From this it seems that the active space chosen in case e is able to describe more of the conjugated system than the one chosen in case d. The CC3 deviation for space d is 1.186 eV and for space e it is 0.810 eV. From the table it can also be seen that space d has more active orbitals than space e. Space d has 23 active occupied orbitals and 142 active virtual orbitals, in comparison to space e that has 19 active occupied orbitals and 137 active virtual orbitals. The CC3 contribution has also been investigated for these three localization schemes. Most of these results are also not acceptable, until space g is reached, which have a CC3 deviation around 0.50 eV for all three methods, while having a CC3 contribution of about 2 eV. The results for space g can therefore be deemed acceptable because there is a significant difference between the CC3 deviation and contribution.

If we then finally compare the results for the local orbitals generated with Cholesky decomposition with the three other localization schemes. It can be seen that Cholesky decomposition generally includes more active orbitals, especially virtual orbitals, than in the case for the other three orbital localization schemes. Even though there are more orbitals included in the active space when using Cholesky decomposition the results are not better.

Table 4.1.1: Number of active (and inactive) orbitals for both occupied and virtual orbitals for the active spaces illustrated in Figure 4.1.1 for the thymine molecule. ΔCC3 is the deviation from the full CC3 value, which was found to be 532.960 eV for thymine.

	Active atoms	#occ orbital	#virt orbital	ΔCC3 [eV]
CD	a	15(18)	100(136)	1.182
	b	19(14)	109(127)	1.325
	c	22(11)	134(102)	1.239
	d	26(7)	159(77)	1.148
	e	21(12)	149(87)	1.387
	f	28(5)	193(43)	1.198
	g	30(3)	183(53)	1.082
PFM	a	12(21)	70(166)	1.301
	b	18(15)	89(147)	1.234
	c	19(14)	114(122)	1.216
	d	23(10)	142(94)	1.186
	e	19(14)	137(99)	0.810
	f	27(6)	180(56)	0.717
	g	29(4)	147(89)	0.496
	h	16(17)	89(147)	0.906
PSM	a	12(21)	70(166)	1.245
	b	19(14)	85(151)	1.179
	c	20(13)	114(122)	1.158
	d	23(10)	142(94)	1.120
	e	20(13)	135(101)	0.865
	f	28(5)	179(57)	0.766
	g	29(4)	142(94)	0.515
Boys	a	12(21)	70(166)	1.237
	b	18(15)	85(151)	1.172
	c	19(14)	114(122)	1.152
	d	23(10)	141(95)	1.113
	e	20(13)	135(101)	0.870
	f	27(6)	179(57)	0.776
	g	29(4)	142(94)	0.515

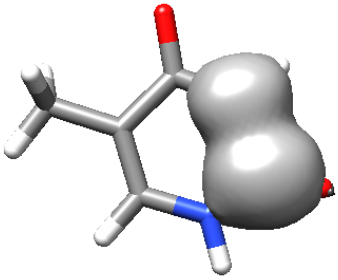
Table 4.1.2 shows the CC3 contribution for the thymine molecule. The CC3 contribution is calculated by subtracting the MLCC3 excitation energy from the the CCSD excitation energy. From this table it can be seen that the CC3 contribution for the Cholesky decomposition does not vary much for the different active spaces, except for spaces b and e which have a somewhat lower contribution. The Cholesky decomposition actually gives a higher contribution than the three other localization scheme for spaces a and d. For the PFM, PSM and Boys scheme the CC3 contribution increases when the active space is increased.

Table 4.1.2: CC3 contribution for the active spaces illustrated in Figure 4.1.4 for the thymine molecule. The CC3 contribution is calculated by subtracting the MLCC3 excitation energy from the CCSD excitation energy.

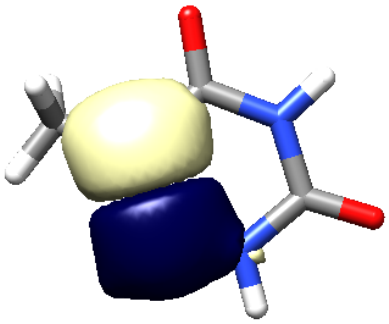
	Active atoms	Contribution CC3 [eV]
CD	a	1.322
	b	1.180
	c	1.268
	d	1.356
	e	1.117
	f	1.312
	g	1.422
PFM	a	1.204
	b	1.271
	c	1.289
	d	1.319
	e	1.695
	f	1.788
	g	2.009
	h	1.599
PSM	a	1.259
	b	1.326
	c	1.347
	d	1.385
	e	1.640
	f	1.739
	g	1.990
Boys	a	1.268
	b	1.333
	c	1.353
	d	1.391
	e	1.634
	f	1.729
	g	1.990

Localization of thymine using PFM

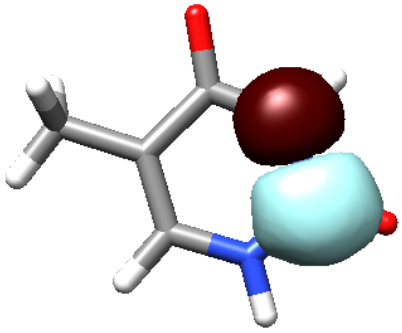
The first localization of the orbitals for the PFM scheme was found out not to be correct. Closer inspection of the orbital localization output showed that the core orbitals had a too high orbital spread, than what was expected. The four first core orbitals had orbital spread values of about 1.3, that should ideally be around 0.3. A large orbital spread indicates a delocalized orbital, because it shows how much the orbital density deviates from the orbital's mean position [8]. Plotting of these orbitals showed that there had been a mixing between the core and valence orbitals, the illustration of these orbitals can be seen in Figure 4.1.2, while subfigure e shows the correct shape of a local core orbital. For local core orbitals the contribution of the 1s function from the atom it belongs to will dominate, and thus a local core orbital will look like an 1s orbital centered on this atom. The problem was discovered to be a wrong start-guess for the calculation, and the localization was run again. The calculations for spaces a-c, e and f, shown in Figure 4.1.1, were run again with a correct orbital localization, but the remaining spaces should also be recalculated.



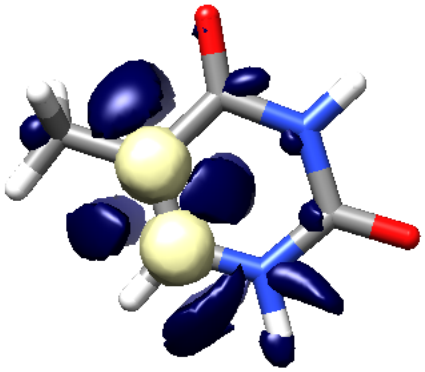
(a) First core orbital



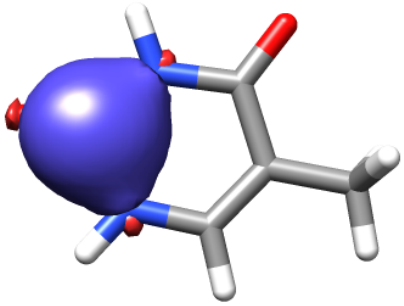
(b) Second core orbital



(c) Third core orbital



(d) Fourth core orbital



(e) Correct core orbital

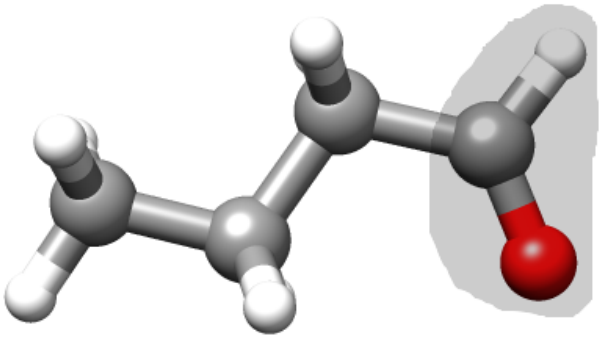
Figure 4.1.2: Subfigures a-d show the four wrongly localized core orbitals, and Subfigure e shows the shape of a correct core orbital

4.1.2 Butanal

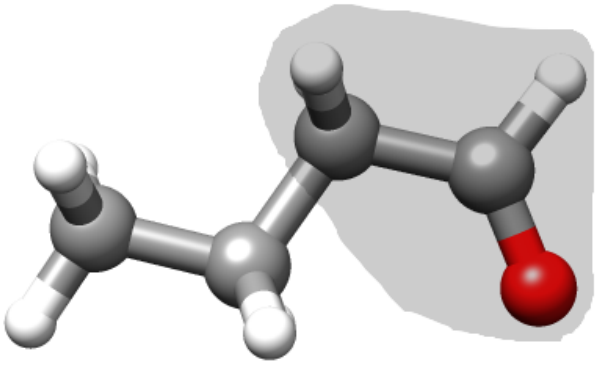
The structure for the butanal molecule is taken from PubChem [53]and the molecule input file was generated using the Dalton Input Plugin for Avogadro [54]. The excitation energy calculations were performed using the basis set aug-cc-pCVDZ. The excitation energy for the oxygen atom was found after running a MLCC3 calculation on the molecule using a start guess for the energy found by CCSD calculations. The start guess for the excitation energy was 532.969 eV. Table 4.1.3 shows the results from the MLCC3 calculations on the active spaces a-c, as illustrated in Figure 4.1.3. The table also shows the number of both occupied and virtual orbitals in the active spaces, and the deviation in the excitation energy from full CC3, denoted ΔCC3. The deviation in the excitation energy shows a decreasing trend for both the Cholesky decomposition and the PFM, PSM and Boys localization schemes. However, the deviation for the Cholesky decomposition does not decrease as much as the deviations from the PFM, PSM and Boys scheme as the active space is expanded. Taking active space c as an example, for Cholesky decomposition the deviation from full CC3 was 0.065 eV and for the PFM scheme it was 0.010 eV. The CC3 contribution was also inspected to analyze if the results were acceptable. For Cholesky decomposition this contribution was 1.649 eV, while for the PFM scheme it was 1.704 eV. Due to the significant difference between the CC3 deviation and contribution, the results for space c are acceptable for both the Cholesky decomposition and the other three localization schems.

Further inspection of the results from the Cholesky decomposition shows, however, that they are not acceptable for active space a and b. The difference in the CC3 deviation and contribution of these to spaces are about 0.23 eV. This then means that to get acceptable results when using Cholesky decomposition the two carbon atoms closest to the functional group should be included in the active space when performing calculations on non-conjugated organic compounds.

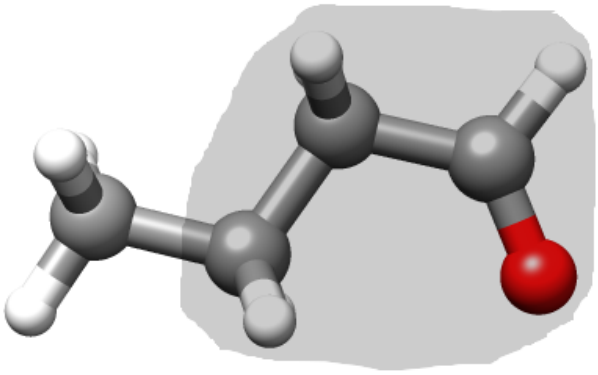
To investigate further the dependence of the occupied and virtual orbitals, two additional analyses were performed for butanal. One where all the occupied orbitals were included in the active space while increasing the number of virtual orbitals, and one where all virtual orbitals were included in the active space while increasing the number of occupied orbitals. The motivation for these analyses was to investigate if either the occupied or virtual orbitals had a bigger impact on the calculations. The results for these calculations are presented in Table 4.1.5, for all occupied included, and in Table 4.1.6, for all virtual included.



(a) 3 atoms in active space



(b) 6 atoms in active space



(c) 9 atoms in active space

Figure 4.1.3: Butanal molecule with active atoms highlighted

Table 4.1.3: Number of active (and inactive) orbitals for both occupied and virtual orbitals for the active spaces shown in Figure 4.1.3 for the butanal molecule. ΔCC3 for the excitation energy is the deviation from the full CC3 value, which was found to be 532.969 eV for butanal.

	Active atoms	#occ orbital	#virt orbital	ΔCC3 [eV]
CD	a	8(12)	74(113)	0.769
	b	12(8)	108(79)	0.743
	c	16(4)	145(42)	0.065
PFM	a	8(12)	55(132)	0.454
	b	12(8)	96(91)	0.140
	c	16(4)	137(50)	0.010
PSM	a	8(12)	55(132)	0.440
	b	12(8)	96(91)	0.110
	c	16(4)	137(50)	0.009
Boys	a	8(12)	55(132)	0.431
	b	12(8)	96(91)	0.105
	c	16(4)	136(51)	0.010

Table 4.1.4: CC3 contribution for the active spaces shown in Figure 4.1.3 for the pentanal molecule. Where the contribution is calculated by subtracting the MLCC3 excitation energy from the CCSD excitation energy.

	Active atoms	Contribution CC3 [eV]
CD	a	0.945
	b	0.972
	c	1.649
PFM	a	1.260
	b	1.574
	c	1.704
PSM	a	1.275
	b	1.604
	c	1.705
Boys	a	1.283
	b	1.609
	c	1.705

All occupied orbitals included in active space

For the calculation where all the occupied orbitals for butanal were included in the active space, and the number of virtual varied, the results are not significantly different for the original calculation where both occupied and virtual orbitals are varied. If the PSM calculation for space b is taken as an example, when all the occupied orbitals are included in the active space, the excitation energy is 0.108 eV, which it also is for the calculation were both the occupied and virtual orbitals were varied. It can be seen from these results that the occupied orbitals do not contribute very much to the excitation energy. These calculations were performed with local orbitals from the PFM, PSM and Boys scheme.

Table 4.1.5: Deviation in excitation energies from the full CC3 value, ΔCC3, for the butanal molecule, when all occupied orbitals are chosen as active for the PFM, PSM and Boys scheme, for the corresponding active spaces illustrated in Figure 4.1.3.

	Active atoms	#occ orbital	#virt orbital	ΔCC3 [eV]
PFM	a	20(0)	55(132)	0.439
	b	20(0)	96(91)	0.138
	c	20(0)	137(50)	0.009
PSM	a	20(0)	55(132)	0.424
	b	20(0)	96(91)	0.108
	c	20(0)	137(50)	0.009
Boys	a	20(0)	55(132)	0.415
	b	20(0)	96(91)	0.103
	c	20(0)	136(51)	0.009

All virtual orbitals included in active space

The calculation were all the virtual orbitals were included in the active space, however, showed a significant difference in the CC3 deviation than the case where all occupied orbitals were included. If the PSM calculation for space b is further inspected, with all virtual orbitals included in the active space the CC3 deviation was 0.013 eV. This is a reduction of 0.1 eV in the CC3 deviation from where all occupied orbitals are included. As stated above, the case where all occupied were included in the active space and where both occupied and virtual were varied showed similar results. This means that there is also a reduction in 0.1 eV from that calculation. It is therefore possible to get a much better results, i.e., less deviation from full CC3, with a smaller active space when including more virtual orbitals.

Table 4.1.6: Deviation in excitation energies from the full CC3 value, ΔCC3, for the butanal molecule, when all virtual orbitals are chosen as active for the PFM, PSM and Boys scheme, for the corresponding active spaces as shown in Figure 4.1.3.

	Active atoms	#occ orbital	#virt orbital	ΔCC3 [eV]
PFM	a	8(12)	187(0)	0.086
	b	12(8)	187(0)	0.012
	c	16(4)	187(0)	0.003
PSM	a	8(12)	187(0)	0.087
	b	12(8)	187(0)	0.013
	c	16(4)	187(0)	0.003
Boys	a	8(12)	187(0)	0.088
	b	12(8)	187(0)	0.013
	c	16(4)	187(0)	0.003

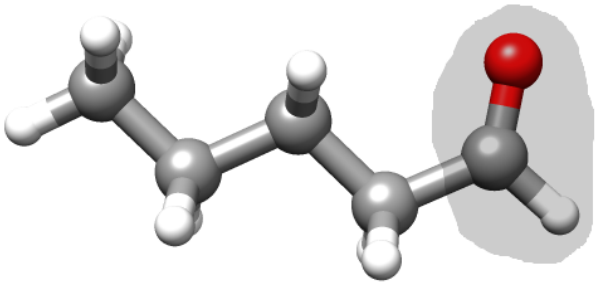
4.1.3 Pentanal

To investigate if the size of the molecule would affect the results for the excitation energies, they were also calculated for pentanal. The main objective for these calculations was to examine if the same active spaces chosen as for butanal, would give significantly different results. The structure for the pentanal molecule is taken from PubChem [55], and the input file was generated using the Dalton Input Plugin for Avogadro [54]. The calculations for pentanal were performed using the basis set aug-cc-pCVDZ. The full CC3 excitation energy was 532.970 eV. The results from the calculations for pentanal for three different active spaces, as shown in Figure 4.1.4, are presented in Table 4.1.7. The table shows the number of occupied and virtual orbitals in active space, and the deviation from full CC3, denoted ΔCC3. For both the Cholesky decomposition and the other three localization schemes the deviation from full CC3 shows a decreasing trend when the active space is increased, as illustrated in Table 4.1.7. For the Cholesky decomposition more virtual orbitals were added than when the PFM, PSM and Boys scheme were used.

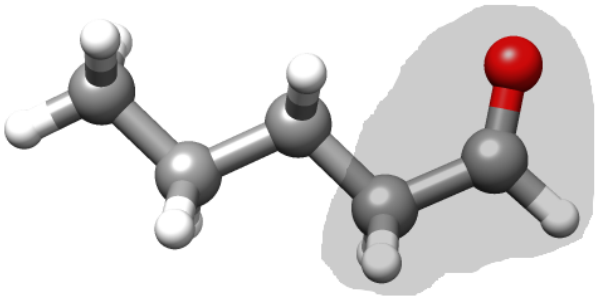
To analyze if the MLCC3 results for pentanal were acceptable, the CC3 contribution was also considered for this case. For pentanal the results the Cholesky decomposition shows good results for all spaces excluding space a. Space a has a CC3 deviation of 0.784 eV and a CC3 contribution of 0.903 eV. Both space b and c have a significant difference between the deviation from full CC3 and the CC3 contribution, therefore only space a in this case gives an excitation energy that is not acceptable.

The deviation from full CC3 for the excitation energy is, however, higher for the Cholesky decomposition than for the other three localization schemes. In addition, the calculations where the Cholesky decomposition was used, more orbitals were included in the active space

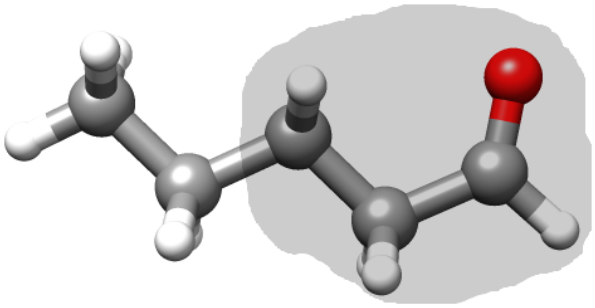
than when the other three were used. Taking space a as an example, the calculation where the localization was performed with the Boys scheme the deviation from full CC3 was 0.392 eV, with 8 occupied and 57 virtual orbitals in the active space. Comparing this to the case where the local orbitals were generates using Cholesky decomposition, the deviation was 0.784 eV, with 8 occupied and 78 virtual orbitals in the active space.



(a) 3 atoms in active space



(b) 6 atoms in active space



(c) 9 atoms in active space

Figure 4.1.4: Pentanal molecule with active atoms highlighted

The same problem with orbital localization for the PFM scheme that was apparent for thymine was also discovered for pentanal. Due to this the localization should be performed again with the correct start-guess, and the excitation energies recalculated.

Table 4.1.7: Number of active(and inactive) orbitals for both occupied and virtual orbitals for the active spaces shown in Figure 4.1.4 for the pentanal molecule. ΔCC3 is the deviation in the excitation energy from the full CC3 value, which was found to be 532.970 eV for pentanal.

	Active atoms	#occ orbital	#virt orbital	ΔCC3 [eV]
CD	a	8(16)	78(150)	0.784
	b	12(12)	109(119)	0.288
	c	16(8)	146(82)	0.136
PFM	a	9(15)	55(173)	0.461
	b	12(12)	97(131)	0.137
	c	16(8)	137(91)	0.011
PSM	a	8(16)	55(173)	0.445
	b	12(12)	96(132)	0.118
	c	16(8)	138(90)	0.011
Boys	a	8(16)	57(171)	0.392
	b	12(12)	96(132)	0.110
	c	16(8)	137(91)	0.011

Table 4.1.8: CC3 Contribution for the active spaces shown in Figure 4.1.4 for the pentanal molecule. The contribution is calculated by subtracting the MLCC3 energy from the CCSD energy.

	Active atoms	Contribution CC3 [eV]
CD	a	0.930
	b	1.426
	c	1.579
PFM	a	1.253
	b	1.577
	c	1.704
PSM	a	1.270
	b	1.598
	c	1.704
Boys	a	1.323
	b	1.605
	c	1.703

Comparison of butanal and pentanal

To conclude the analyses of butanal and pentanal, the results for the two molecules were compared. Comparing Table 4.1.3 for butanal with Table 4.1.7 for pentanal. It can be seen that the deviation from full CC3 is quite close to each other for the two molecules, only differing with about 0.1 eV. Due to the similarity in these results it seems like non-conjugated systems are highly local, and that it is sufficient to chose the active space to just include the functional group and the closest carbon atoms even if the molecule is large.

For the butanal calculation it was found out that only active space c was acceptable for use in calculations, when the local orbitals were generated using Cholesky decomposition, in comparison for pentanal active space b and c both gave acceptable results. It is possible that active space b also would have given an acceptable result for butanal if a more accurate start guess had been used for the MLCC3 calculation. The start guess used for this calculation was from the CCSD calculation, while the start guess for the pentanal calculation was found from full CC3.

4.2 Total energy

4.2.1 Ferrocene

The minimum of the potential energy curve for ferrocene does shift when it its calculated with either CCSD or CCSD(T), and Cholesky decomposition is employed, as illustrated in Figure 4.2.2. Thus the objective for these calculations was to see if this was also the case when the orbitals were localized using the PFM function. However, when using local correlation methods the potential energy surface seem to have several minima, instead of looking like a second order polynomial, with one minimum.

The structure of ferrocene was obtained by optimizing the geometry at the CCSD(T)/frozen core level, and can be seen in Figure 4.2.1. The basis set used for the calculations was cc-pVDZ. The total energy for ferrocene with different geometries were calculated using MLCCSD(T), and the iron atom was the only atom chosen as active. These calculations were done using methods 3, 4 and 5, as described in Table 3.1.1, to assign orbitals to active space, the orbitals were localized using PFM for both methods. A comparison was done between the local orbitals generated from the Cholesky decomposition and the orbitals localized with PFM. In Table 4.2.1 the results for three different geometries are presented, along with the number of occupied and virtual orbitals in the active space, and the deviation from full CCSD(T), denoted ΔCCSD(T). The different geometries considered are the equilibrium geometry (0) and structures where the cyclopentadienyl rings are 0.02 atomic units closer to the iron atom (-0.02) and 0.02 atomic units farther away from the iron atom (0.02). With the reference values for CCSD(T) being -1648.93789 a.u. for geometry 0, -1648.93796 a.u. for geometry -0.02, and -1648.93762 a.u. for geometry 0.02.

From Table 4.2.1 one can see that the energies from the PFM calculation deviates considerably more from the CCSD(T) calculation than what the CD calculation does. The PFM deviations is about 1.9 a.u. from the CCSD(T) calculation, and the CD calculation is about 0.008 a.u. from the CCSD(T) calculation. The number of active orbitals is also very different for the two methods. Where method 3 was used to assign orbitals in active space, there were 9 active occupied orbitals for geometry -0.02, 8 for geometry 0 and 8 for geometry 0.02. The active virtual orbitals were 23 for all three geometries. Method 4 was also tested for the ferrocene molecule. With this method of assigning orbitals in active space, geometry -0.02 had 12 active occupied and 23 active virtual orbitals, geometry 0 had 28 active occupied and 23 active virtual orbitals, and geometry 0.02 had 12 active occupied and 24 active virtual orbitals. For method 5 the number of active occupied orbitals was 20, and the number of active virtual orbitals was 80, with the lowest energy being for geometry -0.02 at -1658.450374 a.u..

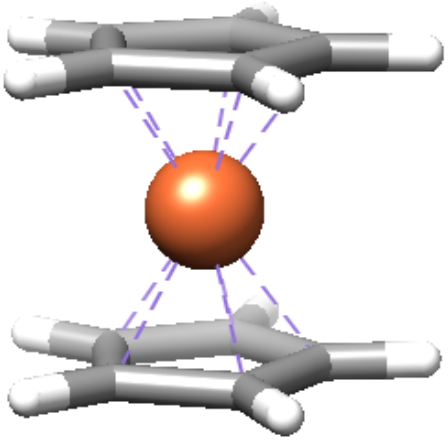
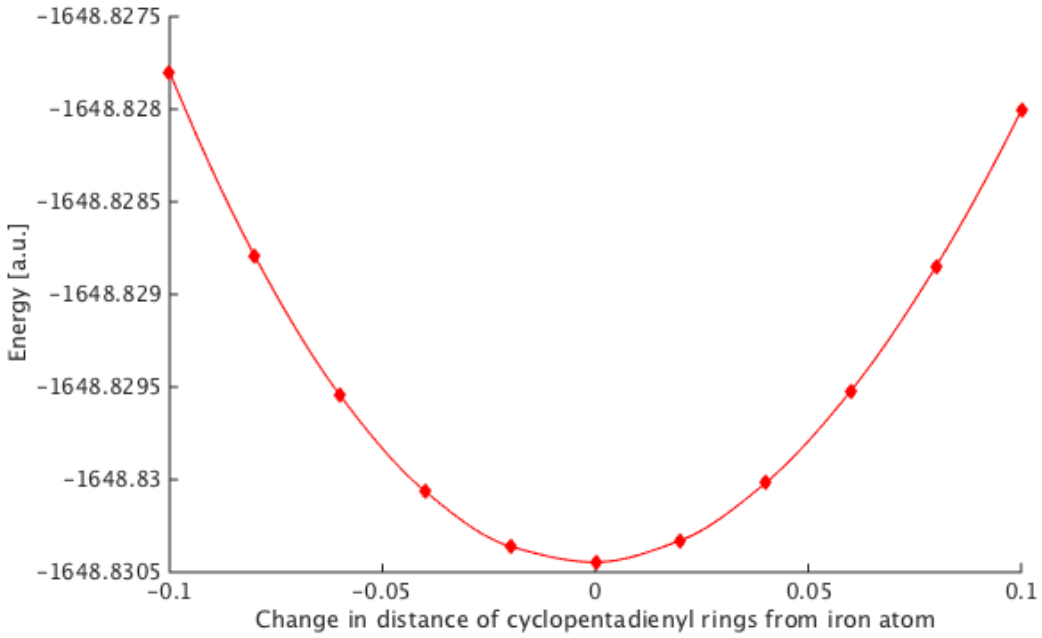


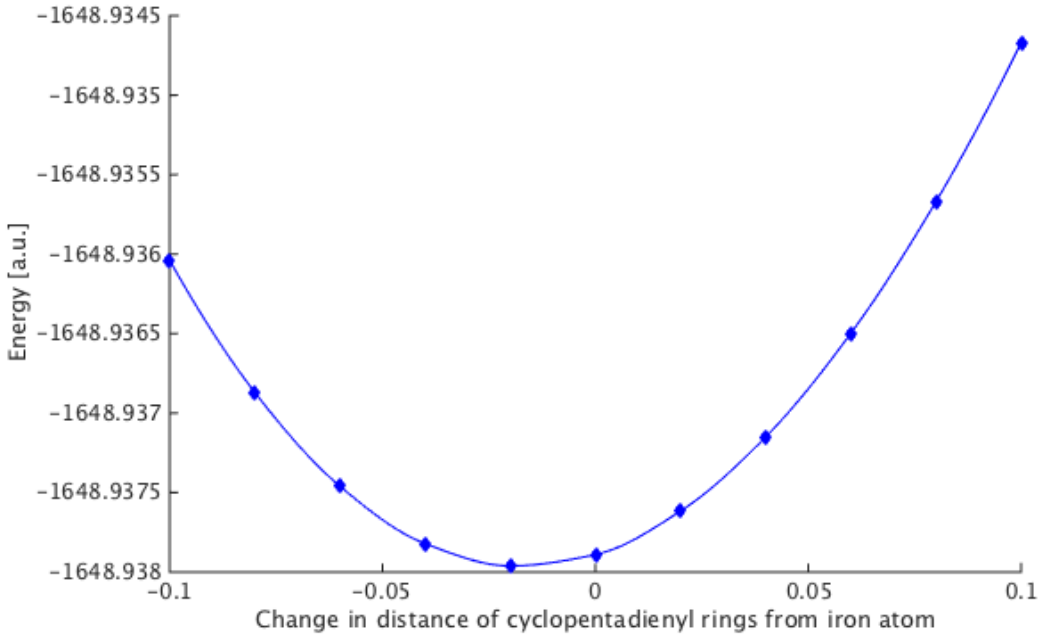
Figure 4.2.1: The structure of the ferrocene molecule

Table 4.2.1: Total energy for three different geometries for ferrocene. 0 is the equilibrium geometry, for -0.02 the two cyclopentadienyl rings are 0.02 a.u. closer together, and for 0.02 they are 0.02 a.u. farther away

	Geometry	#occ orbital	#virt orbital	MLCCSD(T) [a.u.]	ΔCCSD(T) [a.u.]
CD	-0.02	33(15)	180(5)	-1648.92968	-0.00828
	0	33(15)	180(5)	-1648.92962	-0.00827
	0.02	33(15)	180(5)	-1648.92936	-0.00825
Method 3	-0.02	9(39)	23(162)	-1646.99597	-1.94200
	0	8(40)	23(162)	-1646.98413	-1.95376
	0.02	8(40)	23(162)	-1646.99406	-1.94355
Method 4	-0.02	12(36)	23(162)	-1647.07038	-1.86758
	0	28(20)	23(162)	-1648.43933	-0.49856
	0.02	12(36)	24(161)	-1647.09045	-1.84717
Method 5	-0.02	20(28)	80(105)	-1658.45037	9.51241
	0	20(28)	80(105)	-1657.68590	8.74800
	0.02	20(28)	80(105)	-1657.59364	8.65603



(a) Total energy from CCSD



(b) Total energy from CCSD(T)

Figure 4.2.2: The potential energy curve for ferrocene calculated with CCSD, red curve, and CCSD(T), blue curve.

The potential energy curve calculated with MLCCSD(T) for local orbitals generated from Cholesky decomposition has a correct shape. It does resemble a second order polynomial. However, it is not perfect, as the curve is not completely smooth, a slight edge can be seen for the energy at -0.08. Calculations with MLCCSD(T) is calculated using CCSD(T) for the active space, and CCSD for the inactive space. The potential energies for MLCCSD(T) were calculated with 33 occupied and 180 virtual orbitals in the active space, which is almost a complete space. Lacking 15 occupied orbitals and 5 virtual orbitals to be a full CCSD(T) calculation. When this many orbitals are included in the active space, there is not much saving in running a local method, and a conventional CC could have been performed instead. The results from the MLCCSD(T) calculation are gotten from Rolf H. Myhre.

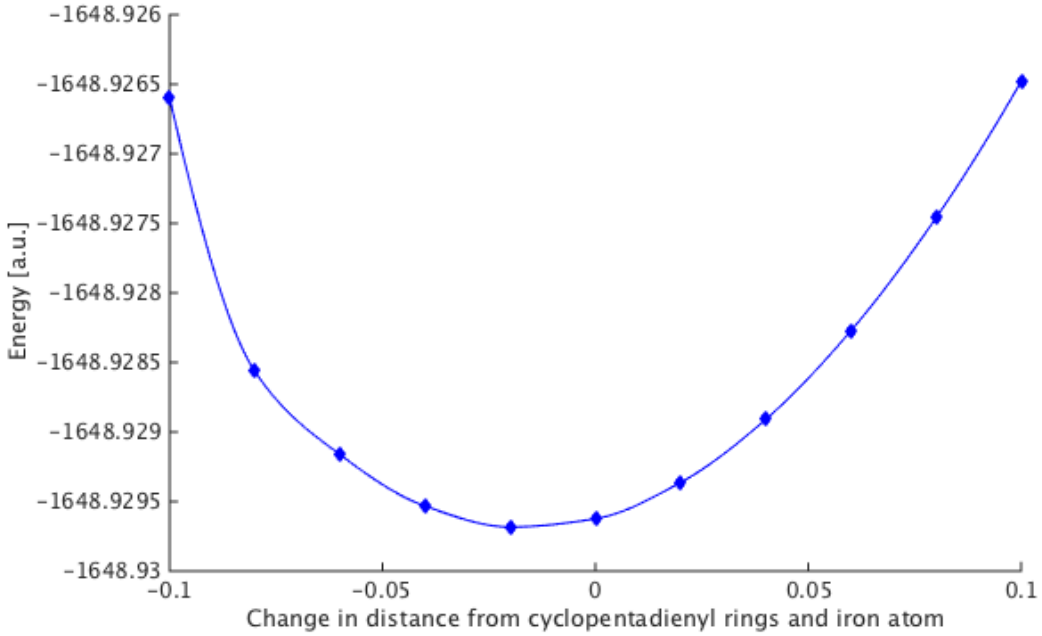


Figure 4.2.3: The potential energy curve calculated with MLCCSD(T)

The motivation for these calculations was to investigate if local orbitals would work for these kind of calculations and reduce the scaling. Further investigations for the potential energy surface was performed, with orbital localized with the PFM scheme and orbitals assigned with Method 5. Method 5 was chosen for these calculations, due to that the other methods did not include consistent number of orbitals in the active space. The potential energy was therefore calculated using MLCCSD, where the active space was treated with CCSD and the inactive space was treated with HF. For this calculation 25 occupied and 100 virtual orbitals were included in the active space. The potential energy curve, did however not resemble a second order polynomial. The curve has two minima instead of one, as can be seen in Figure 4.2.4. Because of the shape of this potential curve, the results from these calculations are not usable. This could possibly be because there is such a large difference between the CCSD

and HF calculations, and maybe there was too much electron correlation to consider. To investigate this further an MLCC3 calculation was performed on the system. The MLCC3 calculation had the same number of orbitals included in the active space as the MLCCSD calculation, 25 occupied and 100 virtual orbitals. The curve for the results from MLCC3 is illustrated in Figure 4.2.5. From the curve it can be seen that this curve also has two minima, and it is no better than the curve where MLCCSD was used.

Table 4.2.2 and Table 4.2.3 shows the total energy calculated with MLCCSD and MLCC3, respectively, for different distances between the iron atom and the cyclopentadienyl rings. These energies are plotted as a potential energy curve in Figure 4.2.4 for MLCCSD and in Figure 4.2.5 for MLCC3.

Table 4.2.2: Total energy for ferrocene calculated with MLCCSD for different distances between the iron atom and the cyclopentadineyl rings. For -0.08 they are 0.08 a.u. closer together than the equilibrium geometry, and for 0.08 they are 0.08 a.u. farther apart. In this calculation there are 25 occupied and 100 virtual orbitals in the active space.

Geometry	MLCCSD [a.u.]
-0.08	-1647.56475
-0.06	-1647.58084
-0.04	-1647.56385
-0.02	-1647.51905
0	-1647.57481
0.02	-1647.60003
0.04	-1647.63932
0.06	-1647.59441
0.08	-1647.61977

Table 4.2.3: Total energy for ferrocene calculated with MLCC3 for different distances between the iron atom and the cyclopentadineyl rings. For -0.06 they are 0.06 a.u. closer together than the equilibrium geometry, and for 0.06 they are 0.06 a.u. farther apart. In this calculation there are 25 occupied and 100 virtual orbitals in the active space.

Geometry	MLCC3 [a.u.]
-0.06	-1648.85046
-0.04	-1648.85072
-0.02	-1648.85126
0	-1648.85048
0.02	-1648.85081
0.04	-1648.84992
0.06	-1648.84970

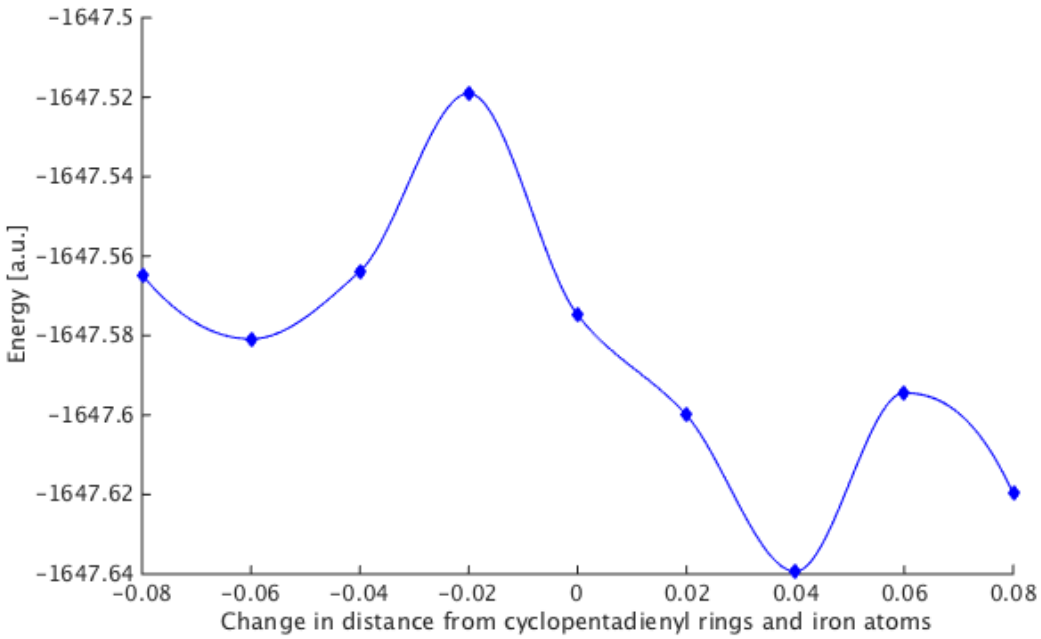


Figure 4.2.4: The potential energy curve calculated with MLCCSD

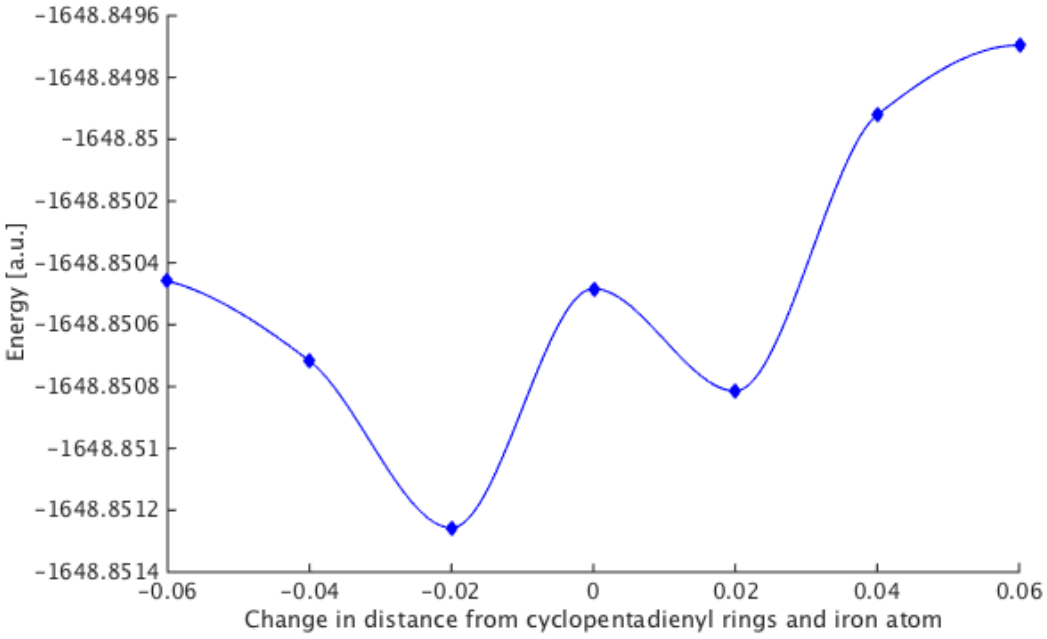


Figure 4.2.5: The potential energy curve calculated with MLCC3

Chapter 5

Summary and Concluding remarks

In the current MLCC implementation the orbitals are localized using Cholesky decomposition, which assigns the orbital to different spaces depending on which atom they are localized. Since the Cholesky orbitals are not very local in comparison to state-of-the-art localized orbitals, and they are more dependent on the locality of the HF density matrix, a routine was implemented to be able to use any kind of sets fo localized orbitals. The orbitals used in this work are localized using LSDALTON, and the excitation energies are calculated using the MLCC framework in DALTON. The orbitals studied in this work are generated using Cholesky decomposition or localized using either the PFM, PSM or Boys scheme. To be able to assign orbitals into the active space, 5 different methods were tested out for the calculations. For the excitation energies several different active spaces were studied to investigate how much the result depended on the selection of active space. This analysis was performed for both a conjugated system, thymine, and two non-conjugated systems, butanal and pentanal. Finally, the dissociation for ferrocene was studied.

5.1 Excitation energies

5.1.1 Local orbitals

When varying the number of orbitals included in the active space the orbitals localized using the PFM, PSM and Boys scheme generally gave better results than the orbitals localized with the Cholesky decomposition. The deviations for these three localization schemes from a full CC3 calculation were in almost all cases lower than for the orbitals generated with Cholesky decomposition. From the results for the excitation energies calculated for thymine there were no significant decrease in the deviation from full CC3 when the local orbitals were generated with Cholesky decomposition. Furthermore, the Cholesky decomposition did not

give any results that had a significant difference between the deviation from full CC3 and CC3 contribution, and for this reason none of these results were acceptable. However, the three other localization schemes gave acceptable results for the largest active space. When comparing the inclusion of occupied and virtual orbitals in the active space, the case where all the virtual orbitals were included gave a much larger reduction in the deviation from full CC3 than when all occupied orbitals were included. From these results it would seem that inclusion of more virtual orbitals in the active space will be more beneficial in order to improve the results.

5.1.2 Comparison of conjugated and non-conjugated systems

In the scope of this work excitation energies were calculated for both a conjugated system, thymine, and for non-conjugated systems, butanal and pentanal. The results showed that not any of the calculations were local orbitals were generated with Cholesky decomposition gave acceptable results, but by using the PFM, PSM and Boys scheme acceptable results were achieved for the largest active space. From these results it seems that the active spaces for conjugated systems have to be carefully chosen, as they tend to be delocalized. Additionally that the active space should cover most of the conjugated system to get acceptable results. However, non-conjugated systems have a more local character and selection of active spaces in these systems are straight-forward and by including the functional group and the closest atoms should be sufficient. From the results of the non-conjugated systems, it can also be seen that the size of the molecule does not affect the result significantly, and the active space can be chosen independently of the size of the molecule.

5.2 Total energy

The initial challenge for the total energy calculation was to be able to include the same number of orbitals in the active space for the different calculations. Because of these challenges several methods were investigated. This was achieved by including a certain number of occupied and virtual orbitals based on the orbitals contribution to the AO density elements.

For the calculations done to model the dissociation of ferrocene, potential energy curves were plotted for different levels of theory. Only the MLCCSD(T) calculation, with almost complete space, gave a potential energy curve that resembled a second order polynomial. The MLCCSD calculation and the MLCC3 calculation, both gave curves with two minima instead of one. From these results for the dissociation of ferrocene, it can be seen that the

local models used in this work are not able to describe the proper behavior unless a lot of orbitals are included in the active space. This means that how the active space is chosen for these kind of reactions are not trivial, and a lot more analyses should be performed for this case.

Chapter 6

Future Work

The orbital localization performed with the PFM scheme, caused some problems because a wrong start guess was used. Due to this, the PFM orbital localization for pentanal should be run again to avoid the mixing of the core and valence orbitals. And the following calculation for the excitation energy should also be done again to get the right energies. The remaining spaces for thymine that has not been recalculated should also be done again. This means the spaces d and g.

At the time of writing the local orbitals used in this thesis are not very viable for describing dissociation curves, as they are not able to illustrate the proper behavior of dissociation, and the results give several minima in the potential energy surface. This should be looked into further, and further analyses should be done for these kind of reactions. It is of interest to reduce the scaling for the calculation of dissociation curves as well as getting good results. A starting point for further development could be to investigate how many orbitals have to be included before the potential energy surface that is calculated gives a curve with only one minimum.

How the active space is specified, by this date, is of great importance to achieve acceptable results, especially for conjugated systems. At this date the framework relies heavily on that the user has knowledge of the system and chemical intuition, to achieve usable results. To avoid inaccuracies and not acceptable results, it should be highly prioritized to make the MLCC framework independent of user input.

Bibliography

[1] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004.

[2] T. Helgaker, P. Jørgensen, and J. Olsen. *Molecular Electronic Structure Theory*. John Wiley & Sons, LTD, 2000.

[3] G. D. Purvis and R. J. Bartlett. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *The Journal of Chemical Physics*, 76(4):1910–1918, 1982.

[4] H. J. Monkhorst. Calculation of properties with the coupled-cluster method. *International Journal of Quantum Chemistry*, 12(S11):421–432, 1977.

[5] S. F. Boys. Construction of some molecular orbitals to be approximately invariant for changes from one molecule to another. *Rev. Mod. Phys.*, 32:296–299, Apr 1960.

[6] C. Edmiston and K. Ruedenberg. Localized atomic and molecular orbitals. *Rev. Mod. Phys.*, 35:457–464, Jul 1963.

[7] J. Pipek and P. G. Mezey. A fast intrinsic localization procedure applicable for abinitio and semiempirical linear combination of atomic orbital wave functions. *The Journal of Chemical Physics*, 90(9):4916–4926, 1989.

[8] B. Jansík, S. Høst, K. Kristensen, and P. Jørgensen. Local orbitals by minimizing powers of the orbital variance. *The Journal of Chemical Physics*, 134(19):–, 2011.

[9] I.-M. Høyvik, B. Jansik, and P. Jørgensen. Orbital localization using fourth central moment minimization. *The Journal of Chemical Physics*, 137(22), 2012.

[10] S. Sæbø and P. Pulay. Local treatment of electron correlation. *Annual Review of Physical Chemistry*, 44(1):213–236, 1993.

[11] M. Schütz, G. Hetzer, and H.-J. Werner. Low-order scaling local electron correlation methods. i. linear scaling local mp2. *The Journal of Chemical Physics*, 111(13):5691–5705, 1999.

[12] G. Hetzer, M. Schütz, H. Stoll, and H.-J. Werner. Low-order scaling local correlation methods ii: Splitting the coulomb operator in linear scaling local second-order møller–plesset perturbation theory. *The Journal of Chemical Physics*, 113(21):9443–9455, 2000.

[13] M. Schütz. Low-order scaling local electron correlation methods. iii. linear scaling local perturbative triples correction (t). *The Journal of Chemical Physics*, 113(22):9986–10001, 2000.

[14] M. Schütz and H.-J. Werner. Low-order scaling local electron correlation methods. iv. linear scaling local coupled-cluster (lccsd). *The Journal of Chemical Physics*, 114(2):661–681, 2001.

[15] G. E. Scuseria and P. Y. Ayala. Linear scaling coupled cluster and perturbation theories in the atomic orbital basis. *The Journal of Chemical Physics*, 111(18):8330–8343, 1999.

[16] O. Christiansen, P. Manninen, P. Jørgensen, and J. Olsen. Coupled-cluster theory in a projected atomic orbital basis. *The Journal of Chemical Physics*, 124(8), 2006.

[17] N. Flocke and R. J. Bartlett. A natural linear scaling coupled-cluster method. *The Journal of Chemical Physics*, 121(22):10935–10944, 2004.

[18] J. E. Subotnik and M. Head-Gordon. A local correlation model that yields intrinsically smooth potential-energy surfaces. *The Journal of Chemical Physics*, 123(6), 2005.

[19] J. E. Subotnik, A. Sodt, and M. Head-Gordon. A near linear-scaling smooth local coupled cluster algorithm for electronic structure. *The Journal of Chemical Physics*, 125(7), 2006.

[20] R. H. Myhre, A. M. J. Sánchez de Merás, and H. Koch. Multi-level coupled cluster theory. *The Journal of Chemical Physics*, 141(22):–, 2014.

[21] R. H. Myhre and K. Henrik. The multi-level cc3 coupled cluster model. *Submitted*, 2016.

[22] O. Christiansen, H. Koch, and P. Jørgensen. Response functions in the cc3 iterative triple excitation model. *The Journal of Chemical Physics*, 103(17):7429–7441, 1995.

[23] H. Koch, O. Christiansen, P. Jørgensen, A. M. Sanchez de Merás, and T. Helgaker. The cc3 model: An iterative coupled cluster approach including connected triples. *The Journal of Chemical Physics*, 106(5):1808–1818, 1997.

[24] F. Aquilante, T. Bondo Pedersen, A. Sánchez de Merás, and H. Koch. Fast noniterative orbital localization for large molecules. *The Journal of Chemical Physics*, 125(17):–, 2006.

[25] A. M. J. Sánchez de Merás, H. Koch, I. G. Cuesta, and L. Boman. Cholesky decomposition-based definition of atomic subsystems in electronic structure calculations. *The Journal of Chemical Physics*, 132(20), 2010.

[26] K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, E. K. Dalskov, U. Ekström, T. Enevoldsen, J. J. Eriksen, P. Ettenhuber, B. Fernández, L. Ferrighi, H. Fliegl, L. Frediani, K. Hald, A. Halkier, C. Hättig, H. Heiberg, T. Helgaker, A. C. Hennum, H. Hettema, E. Hjertenæs, S. Høst, I.-M. Høyvik, M. F. Iozzi, B. Jansík, H. J. A. Jensen, D. Jonsson, P. Jørgensen, J. Kauczor, S. Kirpekar, T. Kjærgaard, W. Klopper, S. Knecht, R. Kobayashi, H. Koch, J. Kongsted, A. Krapp, K. Kristensen, A. Ligabue, O. B. Lutnæs, J. I. Melo, K. V. Mikkelsen, R. H. Myhre, C. Neiss, C. B. Nielsen, P. Norman, J. Olsen, J. M. H. Olsen, A. Osted, M. J. Packer, F. Pawlowski, T. B. Pedersen, P. F. Provasi, S. Reine, Z. Rinkevicius, T. A. Ruden, K. Ruud, V. V. Rybkin, P. Sałek, C. C. M. Samson, A. S. de Merás, T. Saue, S. P. A. Sauer, B. Schimmelpfennig, K. Sneskov, A. H. Steindal, K. O. Sylvester-Hvid, P. R. Taylor, A. M. Teale, E. I. Tellgren, D. P. Tew, A. J. Thorvaldsen, L. Thøgersen, O. Vahtras, M. A. Watson, D. J. D. Wilson, M. Ziolkowski, and H. Ågren. The dalton quantum chemistry program system. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 4(3):269–284, 2014.

[27] P. Atkins and R. Friedman. *Molecular Quantum Mechanics*. Oxford University Press, 5 edition, 2011.

[28] A. Szabo and N. S. Ostlund. *Modern Quantum Chemistry, Introduction to advanced electronic structure theory*. Dover Publications, Inc., 1 edition, 1996.

[29] F. Jensen. *Introduction to Computational Chemistry*. John Wiley & Sons, Ltd, 2 edition, 2007.

[30] J. P. Lowe and K. A. Peterson. *Quantum Chemistry*. Elsevier Academic Press, 3 edition, 2006.

[31] B. O. Roos, P. R. Taylor, and P. E. Siegbahn. A complete active space scf method (casscf) using a density matrix formulated super-ci approach. *Chemical Physics*, 48(2):157 – 173, 1980.

[32] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon. A fifth-order perturbation comparison of electron correlation theories. *Chemical Physics Letters*, 157(6):479–483, 1989.

[33] J. Řezáč and P. 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. PMID: 26583708.

[34] G. E. Scuseria. The open-shell restricted hartree—fock singles and doubles coupled-cluster method including triple excitations ccsd (t): application to c+3. *Chemical Physics Letters*, 176(1):27 – 35, 1991.

[35] G. E. Scuseria and T. J. Lee. Comparison of coupled-cluster methods which include the effects of connected triple excitations. *The Journal of Chemical Physics*, 93(8):5851–5855, 1990.

[36] S. A. Kucharski and R. J. Bartlett. The coupled-cluster single, double, triple and quadruple excitation method. *The Journal of Chemical Physics*, 97(6):4282–4288, 1992.

[37] L. Thøgersen. *Optimizations of Densities in Hartree-Fock and Density-functional Theory Atomic Orbital Based Response Theory and Benchmarking for Radicals*. PhD thesis, University of Aarhus, 2005.

[38] S. Hirata*. Tensor contraction engine: abstraction and automated parallel implementation of configuration-interaction, coupled-cluster, and many-body perturbation theories. *The Journal of Physical Chemistry A*, 107(46):9887–9897, 2003.

[39] J. L. Bentz, R. M. Olson, M. S. Gordon, M. W. Schmidt, and R. A. Kendall. Coupled cluster algorithms for networks of shared memory parallel processors. *Computer Physics Communications*, 176(9–10):589 – 600, 2007.

[40] R. M. Olson, J. L. Bentz, R. A. Kendall, M. W. Schmidt, and M. S. Gordon. A novel approach to parallel coupled cluster calculations: combining distributed and shared memory techniques for modern cluster based systems. *Journal of Chemical Theory and Computation*, 3(4):1312–1328, 2007.

[41] J. Olsen. The initial implementation and applications of a general active space coupled cluster method. *The Journal of Chemical Physics*, 113(17):7140–7148, 2000.

[42] A. Köhn and J. Olsen. Coupled-cluster with active space selected higher amplitudes: Performance of seminatural orbitals for ground and excited state calculations. *The Journal of Chemical Physics*, 125(17), 2006.

[43] T. 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.

[44] A. R. Leach. *Molecular Modelling, Principles and Applications.* Pearson Education Limited, 2 edition, 2001.

[45] M. Ziółkowski, B. Jansík, P. Jørgensen, and J. Olsen. Maximum locality in occupied and virtual orbital spaces using a least-change strategy. *The Journal of Chemical Physics*, 131(12), 2009.

[46] I.-M. Høyvik, K. Kristensen, T. Kjærgaard, and P. Jørgensen. A perspective on the localizability of hartree–fock orbitals. *Thom H. Dunning, Jr.*, 2015.

[47] I.-M. Høyvik and P. Jørgensen. Localized orbitals from basis sets augmented with diffuse functions. *The Journal of Chemical Physics*, 138(20), 2013.

[48] I.-M. Høyvik, B. Jansik, and P. Jørgensen. Trust region minimization of orbital localization functions. *Journal of Chemical Theory and Computation*, 8(9):3137–3146, 2012. PMID: 26605725.

[49] I.-M. Høyvik. *Local Hartree-Fock orbitals: Characterization, structure and optimization.* PhD thesis, Aarhus University, 2013.

[50] J. M. Millam and G. E. Scuseria. Linear scaling conjugate gradient density matrix search as an alternative to diagonalization for first principles electronic structure calculations. *The Journal of Chemical Physics*, 106(13), 1997.

[51] R. H. Myhre. Development and implementation of extended cc2 models. Master's thesis, Norwegian University of Science and Technology.

[52] S. Coriani and H. 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), 2015.

[53] Butanal - PubChem, 2016. `https://pubchem.ncbi.nlm.nih.gov/compound/261`, 2016.

[54] Avogadro: an open-source molecular builder and visualization tool. version 1.1.1. `http://avogadro.openmolecules.net/`, 2000.

[55] Pentanal - PubChem, 2016. `https://pubchem.ncbi.nlm.nih.gov/compound/8063`, 2016.