

AIX-MARSEILLE UNIVERSITÉ

ECOLE DOCTORALE

Institut des sciences moléculaires de Marseille

Thèse présentée pour obtenir le grade universitaire de docteur

Discipline: Sciences chimiques

Spécialité: Chimie théorique

Alexander PUNTER

Développement de pseudopotentiels moléculaires adaptés à
l'étude de propriétés moléculaires

Pseudopotentials for the study of Molecular Properties

Soutenue le 16/10/2019 devant le jury composé de:

Fabienne ALARY	Professeur	Rapportrice
Laurent JOUBERT	Professeur	Rapporteur
Christophe MORELL	Professeur	Examineur
Carine CLAVAGUÉRA	Chargé de recherche	Examineur
Yannick CARISSAN	Maître de conférences	Directeur de thèse
Paola NAVA	Maître de conférences	Directrice de thèse

Abstract

A pseudopotential method for replacing small hydrocarbon fragments in molecular quantum chemistry calculations is derived and tested. These fragments contain only one or two explicitly-treated electrons.

The first is an sp^2 carbon atom fragment optimised on ethylene. This potential is found to transfer well to all-trans-polyene and PAH systems, reproducing the HOMO and 1st ionisation and excitation energies to within around 20% of all-electron DFT calculations. Additional pseudopotentials are created for other sp^2 and sp^3 hybridised carbon-based fragments. They are able to form bonds with all-electron atoms, and are also found to transfer well to larger systems, provided the systems with which they interact are similar enough to those on which they were optimised. The pseudopotentials are found to be able accurately to reproduce molecular UV and ECD spectra, and are tested on a variety of larger and more complex systems including helicenes, nanotubes and a metallocomplex.

A program for the optimisation of these pseudopotentials is developed in Python, which is able to extract pseudopotentials from a variety of all-electron reference criteria.

Unrelated to the above, the effects of a lactone group on a particular cobalt-mediated cycloaddition are also examined.

Keywords: pseudopotentials, model potentials, sp^2 carbon, sp^3 carbon, PAH, helicene, twistacene, hemicryptophane, s - p hybridisation

I should like to acknowledge both the extensive wisdom and the generous patience of my supervisors, Dr Yannick Carissan and Dr Paola Nava.

To Claire, and to my parents

Contents

Abstract	2
Contents	5
List of Figures	8
List of Tables	11
1 Introduction	13
1.1 Preamble	14
1.2 The Problem	14
1.2.1 Making the Schrödinger Equation Soluble	14
1.2.2 The Hartree-Fock Method	18
1.2.3 The Self-Consistent Field Procedure	19
1.2.4 Post-Hartree-Fock and Other Animals	20
1.2.5 Density Functional Theory	21
1.3 Historical Notes	24
1.3.1 The Pseudopotential	24
1.3.2 The Model Core Potential	27
1.3.3 Other Flavours of Pseudopotential	31
2 Methodology	33
2.1 Pseudopotential Form	33
2.2 Extraction of Parameters	36

3	Results	38
3.1	Prototype work	40
3.1.1	Initial Experiments and Thoughts	40
3.2	Results for all-trans-polyenes and Polycyclic Aromatic Hydrocarbons	48
3.2.1	All-trans-polyenes	48
3.2.2	Polycyclic Aromatic Hydrocarbons	56
3.2.3	CASPT2	59
3.3	Expanding the Method: The MOO program	60
3.3.1	The Optimisation Criteria	61
3.3.2	The Optimisation Procedure	63
3.3.3	Interpreting the Results	65
3.4	Simple Molecular Systems	66
3.4.1	β potentials: sp^2 two-electron pseudofragment	67
3.4.2	γ potentials: sp^3 one-electron pseudofragment	72
3.4.3	δ potentials: sp^3 two-electron pseudofragment	76
3.5	Absorption spectra of Pseudomolecules	82
3.5.1	Optimisation for Spectra	87
3.6	Pseudopotential Studies of Complex Systems	93
3.6.1	Complex Spectra: Helicene	93
3.6.2	Complex Spectra: Twistacene	104
3.6.3	Complex Spectra: Dodecaphenyltetracene	114
3.6.4	Complex Spectra: Nanotube-embedded Coronene	117
3.6.5	Complex Spectra: Hemi-Cryptophane	122
3.6.6	Results and Discussion	123
3.7	Geometry Optimisation	131
3.7.1	Development of the Algorithm	131
3.7.2	Results and Discussion	134
3.8	Timings	139

4	A Study of Cobalt-Mediated Cycloaddition: Lactone and Lactame	142
4.1	Background	142
4.2	Experiment	143
4.3	DFT Study	145
4.4	Lactame <i>vs</i> Lactone	147
4.5	Conclusions	148
4.5.1	Thoughts on Further Development	153
	Conclusion	155
	Bibliography	156
	Appendices	166
A	Physical Meaning of the p_z Pseudopotential	166
B	Topological Analysis of Electron Density	168
	Résumé	181

List of Figures

1.1	Google Scholar results for ‘pseudopotential’ by year	32
3.1	Diagram of initial s potential setup.	41
3.2	Energy level diagram of ethylene.	46
3.3	Diagram of pseudo-all-trans-polyene unit.	48
3.4	Short all-trans-polyene energy level results.	49
3.5	Long all-trans-polyene energy level results.	50
3.6	Comparison of S^2 expectation values for all-electron and pseudosystems.	53
3.7	Comparison of short all-trans-polyenes with a previous study.	54
3.8	Hartree-Fock ionisation energies for frozen-orbital all-trans-polyenes.	56
3.9	Diagrams of PAH molecules.	57
3.10	Energy level results for PAH molecules.	58
3.11	TDDFT and CASPT2 results comparison.	60
3.12	β potential diagram.	68
3.13	Energy level results for β pseudopotentials.	70
3.14	γ potential diagram.	72
3.15	Energy level results for γ pseudopotentials.	75
3.16	δ potential diagram.	78
3.17	Energy level results for δ pseudopotentials.	80
3.18	PAH and pseudo-PAH UV spectra: <i>set4</i>	84
3.19	PAH and pseudo-PAH UV spectra: <i>geom₁</i>	86

3.20	Images from a least-squares pseudopotential fit.	89
3.21	The structure of (P)-helicene, from 6 to 9 rings.	94
3.22	Computed and experimental spectra for helicenes.	95
3.23	Computed all-electron and pseudopotential spectra for helicenes: <i>set4</i>	97
3.24	Computed all-electron and pseudopotential spectra for helicenes:	
	<i>geom</i> ₁	99
3.25	[10]helicene transition densities.	101
3.26	<i>geom</i> ₁ pseudo[10]helicene transition densities.	102
3.27	Definition of dihedral torsion for benzene	104
3.28	Twistacene structure diagrams.	105
3.29	Twistacene all-electron and pseudopotential spectra.	107
3.30	Zoomed twistacene all-electron and <i>geom</i> ₁ ECD spectra.	109
3.31	Transition densities for twistacene.	112
3.33	Dodecaphenyltetracene structure diagram.	115
3.34	Dodecaphenyltetracene spectra: All-electron and pseudopotentials. .	116
3.35	Nanotube-embedded coronene structure diagram.	119
3.36	UV spectrum of nanotube-embedded coronene.	120
3.37	Transition densities of nanotube-embedded coronene.	121
3.38	Diagram of Cu ²⁺ in hemicryptophane cage.	123
3.39	Labelling scheme for Cu(II)hemicryptophane complex.	124
3.40	The ‘no heterobonding’ rule for upper π ring pseudopotentials. . . .	125
3.41	All-electron and pseudoCu(II)hemicryptophane spectra.	127
3.42	Transition densities for Cu(II)hemicryptophane.	130
3.43	All-electron and pseudoethane dissociation curves.	133
3.44	A collapsed pseudoaspirin molecule.	139
3.32	Transition densities for pseudotwistacene.	141
4.1	Theoretical cobalt-mediated [2+2+2] cycloaddition graph, with triplet states marked in green.	143

4.2	The synthesis and X-ray structure of Cobalt(III) complex P1a . . .	144
4.3	Expected standard outcome of Cobalt(I)-mediated process vs experimental result.	145
4.4	Reaction profile for cobalt complex with lactone.	146
4.5	Reaction profile for cobalt complex with lactame.	147
.1	AIMs diagrams of all-electron and pseudobenzene.	168
.2	Transition densities for all-electron and <i>set4</i> pseudopyrene.	170
.3	Comparaison des spectres pour les HAP.	178
.4	Résultat standard attendu du processus utilisant le cobalt (I) par rapport au résultat expérimental.	181

List of Tables

3.1	<i>s</i> -only CH ₃ [•] pseudopotentials, $r = 2.0$ au.	42
3.2	<i>s</i> -only CH ₃ [•] pseudopotentials, $r = 0.5$ au.	42
3.3	Early ethylene pseudopotentials.	44
3.4	Pseudoethylene results using the <i>set4</i> pseudopotentials.	47
3.5	Errors across calculation types for all-trans-polyenes C _{<i>n</i>} H _{<i>n</i>+2}	51
3.6	Hartree-Fock ionisation energies for frozen-orbital all-trans-polyenes.	55
3.7	Errors across calculation types for PAH molecules.	59
3.8	Boundary conditions for pseudopotential optimisation.	64
3.9	An example pseudopotential optimisation table.	65
3.10	Summary of α , β , γ and δ potentials.	67
3.11	β test molecule set diagrams.	68
3.12	Error results for β pseudopotentials.	71
3.13	Optimisation criteria and parameters for the best β potential set.	71
3.14	γ test molecule set diagrams.	73
3.15	Error results for γ pseudopotentials.	77
3.16	Optimisation criteria and parameters for the best γ potential set.	77
3.17	δ test molecule set diagrams.	79
3.18	Error results for δ pseudopotentials.	82
3.19	Optimisation criteria and parameters for the best δ potential set.	82
3.20	Optimisation table for the <i>leastsq₁</i> pseudopotential set.	88
3.21	Optimisation of <i>geom₁</i> pseudopotentials.	91
3.22	Planar distortion of benzene compared to helicene.	111

3.23	Breakdown of pseudoCu(II)hemicryptophane structures.	126
3.24	Additional molecules used to test the optimisation of geometries with γ pseudopotentials.	135
3.25	Geometry optimisation results for β potentials.	136
3.26	Geometry optimisation results for γ potentials.	137
3.27	Time per SCF iteration for all-electron and pseudopotential calcula- tions.	140
.1	Orbital weights for all-trans-polyene excitations.	171
.2	Erreurs dans les calculs pseudopotentiels pour les trans-polyènes. . .	174
.3	Un résumé des types de potentiels créé.	176

1. Introduction

Sommaire

1.1	Preamble	14
1.2	The Problem	14
1.2.1	Making the Schrödinger Equation Soluble	14
1.2.1.1	The Variational Principle	17
1.2.2	The Hartree-Fock Method	18
1.2.3	The Self-Consistent Field Procedure	19
1.2.4	Post-Hartree-Fock and Other Animals	20
1.2.5	Density Functional Theory	21
1.3	Historical Notes	24
1.3.1	The Pseudopotential	24
1.3.1.1	Extraction of Parameters	26
1.3.2	The Model Core Potential	27
1.3.3	Other Flavours of Pseudopotential	31

1.1. Preamble

This introduction is divided into two parts. The first part consists of a brief theoretical exploration of the problems to be solved. The purpose of this part is not to be a primer to Quantum Chemistry (those interested in such a primer should look elsewhere, for example to Reference [1]), though some of the foundations will necessarily be touched upon. The aim is to establish some general ideas, terms and equations that will be of direct relevance in later sections, and more generally to emphasise the way in which the subject as a whole relies upon the use of chemical, physical and mathematical intuition in order to make approximations, as well as the computational necessity of doing so.

The second part consists of some historical notes on pseudopotentials in their various incarnations. This is not a literature review, either of pseudopotentials in general or non-relativistic molecular pseudopotentials in particular. The aim is to trace the ideas behind this particular work and to follow their historical thread. As such, large regions of theory and development in the field are mentioned only in passing or not mentioned at all. This is particularly true of pseudopotentials for solid states, and relativistic pseudopotentials. Those in search of literature reviews on the subject are encouraged to refer to those of Dolg [2], Dolg and Cao [3] and Klobukowski *et al.* [4], as well as the references therein.

1.2. The Problem

1.2.1. Making the Schrödinger Equation Soluble

In order to make a complete *ab initio* calculation of the energetics of a molecule, one would analytically solve the time-independent Schrödinger eigenvalue equation [5]

$$\hat{H}|\Psi\rangle = E|\Psi\rangle \quad (1.1)$$

where \hat{H} is a Hamiltonian operator, Ψ is the wavefunction of the system, and E is the total energy. One would solve this equation making sure to include all particles in the system in Ψ . This would include the spatial positions of all particles in the molecule, as well as every possible combination of electron occupation. Moreover, one would additionally need to take into account the environment of the molecule (such as temperature, any external fields present, the influence of any other molecules *e.g.* solvents and so on).

This is a task so daunting that in practice it is only possible for the very simplest systems. Mercifully however, such an approach is not necessary for practical considerations, with a great many approximations and simplifications available to the chemist which bring detailed knowledge of chemical systems, at the quantum level, into the realm of the possible.

The first among these is the Born-Oppenheimer approximation. Born and Oppenheimer proposed [6] that while electrons and protons experience similar forces (due to having equal but opposite charges), the mass difference between nucleons and electrons was large enough that they were effectively dynamically decoupled. The electrons move so much more quickly than the heavy nucleons, that for the purposes of calculation it could be assumed that there is no coupling of their motion with that of the electrons. This means that the total wavefunction of the molecule can be separated into nuclear and electronic terms, with the nuclear terms becoming simply a potential field acting on the electrons. Solutions for these variables can then be found for individual spatial arrangements of the nuclei. This new electron-only wavefunction therefore depends on the positions of the nuclei *parametrically*, rather than *explicitly*, reducing the number variables in the overall wavefunction.

This new electron-only wavefunction however, is still not much less intractable,

and must be broken down further. This is done by finding a way in which Ψ_N , the wavefunction describing N electrons, can be approximated by a product of wave functions describing each electron individually:

$$\Psi_N = \psi_1\psi_2\dots\psi_N \quad (1.2)$$

This is known as a Hartree product. In order to describe the electrons each with their own individual wavefunction, it must be temporarily assumed that they do not interact with one another. This is not an intuitive assumption, as the electrons obviously do interact, both classically as charged particles, and quantum-mechanically through the Pauli exclusion principle [7]. However, it will be seen that the electron-electron interactions can be worked back into the theory afterwards, allowing for the recovery of much of the accuracy lost in this initial assumption.

In order for it accurately to conform to the Pauli principle, that electrons are indistinguishable and anti-symmetric, a correction must be added. Alternatively, the anti-symmetry criterion can be met in a more natural way by re-writing the above as a matrix determinant. In order to do this, we introduce the idea of an 'orbital'. This is a function which describes the wavefunction of a single particle, in our case an electron. A so-called 'spatial orbital' is a function of the position vector \vec{r} of an electron. However, in order to describe an electron fully it is necessary to specify its spin. Therefore we can introduce the so-called 'spin-orbital', χ , which is a function of both spatial and spin coordinates, and which we shall call \vec{x} .

These orbitals can then form our matrix determinant, with the general form for N electrons of

$$\Psi(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N) = (N!)^{-\frac{1}{2}} \begin{vmatrix} \chi_i(\vec{x}_1) & \chi_j(\vec{x}_1) & \dots & \chi_k(\vec{x}_1) \\ \chi_i(\vec{x}_2) & \chi_j(\vec{x}_2) & \dots & \chi_k(\vec{x}_2) \\ \vdots & \vdots & & \vdots \\ \chi_i(\vec{x}_N) & \chi_j(\vec{x}_N) & \dots & \chi_k(\vec{x}_N) \end{vmatrix} \quad (1.3)$$

where $(N!)^{-\frac{1}{2}}$ is a normalisation factor. The mathematical advantage of the determinant format is that the exchange of any two rows or columns of the determinant (that is, the exchange of any two electrons in the system) results in the change of the sign of the determinant, thus fulfilling the anti-symmetry condition. This is known as a Slater determinant, after John Slater [8].

1.2.1.1. The Variational Principle

While finding the orbitals χ_i is not trivial, it is possible to be sure of having the correct orbitals by invoking the variational principle. Taken from a standard textbook [1], the variational principle states that given a normalised wavefunction ψ that satisfies appropriate boundary conditions, the expectation value of the Hamiltonians is an upper bound to the exact ground state energy E_0 . That is, if

$$\langle \psi | \psi \rangle = 1 \quad (1.4)$$

then

$$\langle \psi | \hat{H} | \psi \rangle \geq E_0 \quad (1.5)$$

where the equality holds only when $\psi = \psi_0$, where ψ_0 is the eigenfunction of the ground state. What this means is that the wavefunction ψ that gives the lowest overall expectation value for the total energy must be the true ground state wavefunction. ψ is often referred to as a trial function.

Putting the above together, we have an approximation to Equation 1.1. A

Slater determinant with a normalised set of orbitals χ_i which gives the lowest possible expected total energy must describe the ground state of the system. One very common way to make use of this approximation is with the Hartree-Fock method [9].

1.2.2. The Hartree-Fock Method

Established above is the idea that Equation 1.1 can be re-written in terms of a Slater determinant of orbital functions, and that the set of orbital functions that minimise the total expected energy represent the ground state (the variational principle), and that a proposed set of such orbital functions could be referred to as a *trial function*. We thus have all the ingredients necessary to understand the essence of the Hartree-Fock method. In the Hartree-Fock method, Equation 1.1 is replaced by the Hartree-Fock equation:

$$\hat{f}(i)\chi_i(\vec{x}_i) = \epsilon_i\chi_i(\vec{x}_i) \quad (1.6)$$

where $\chi_i(\vec{x}_i)$ are the spin orbitals approximating the original wavefunction Ψ , and where $\hat{f}(i)$ is known as the Fock operator

$$\hat{f}(i) = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^M \frac{Z_A}{r_{iA}} + v^{HF}(i) \quad (1.7)$$

where the first term gives us the kinetic energy of the electron and the second the interaction with the M nuclei of the molecule. The third term comes about as a consequence of our earlier assumption that the electrons do not interact. This lack of interaction is now corrected for by assuming instead that an individual electron experiences only the *mean* potential field of all the other electrons. This is the term $v^{HF}(i)$. Since this term must depend on the orbitals of all electrons other than electron i , the Hartree-Fock equation is non-linear and must be solved iteratively, in what is known as the Self-Consistent Field procedure.

1.2.3. The Self-Consistent Field Procedure

Above it is shown how the Schrödinger equation for a molecule could, by a process of assumption and approximation, be rendered in the form given in Equation 1.6. We know that there exists a set of orbital functions χ_i which give us the eigenvalues ϵ_i , their respective energies, and thus the total energy of the system. We also know that the variational principle means that the set of orbitals χ_i that give us the lowest overall expectation energy must be the ground state. This means that finding the correct orbitals becomes a minimisation problem. Through starting with a set of trial orbital functions χ_i , calculating the expectation of the energy, altering the orbitals and recalculating, we eventually arrive at the lowest possible energy, the ground state of the system.

The combination of all of the above approximations is a near-complete description of the Hartree-Fock Method. This has allowed us to reformulate the intractable problem of Equation 1.1 in a way which

1. reduces the overall wavefunction Ψ so that it describes only the electrons of the system explicitly (the Born-Oppenheimer approximation).
2. allows us to write the overall electronic wavefunction as an anti-symmetrised product of one-electron wavefunctions in the form of a single matrix determinant, or Slater determinant.
3. reduces the problem of electron-electron interaction to the treatment of each electron as acting in the mean field of all the others (the mean field theory).

In order to describe the Hartree-Fock method fully, two further concepts need to be introduced. The first is the assumption that the effects of relativity in the system are small enough that they can be neglected, which for molecules containing only light elements is generally true. The second governs the nature of the orbital functions introduced above. By introducing constraints on the form of

the trial orbital functions χ_i so as to restrict the SCF procedure to solely linear variations, the finding of the optimal set of orbitals can be made vastly simpler. The means by which this is done are known as Molecular Orbital theory, and is outside the scope of this introduction. Interested readers should turn to the references given in Section [1.1](#).

1.2.4. Post-Hartree-Fock and Other Animals

The above summaries of the Hartree-Fock approximation and SCF procedure are brief, and lack a full discussion of the mathematical background. They also gloss over certain important details, such as the starting guesses commonly used to begin SCF calculations, as well as the full molecular integrals used in them. They also do not give any real account of Molecular Orbital theory. What they do however, is begin to give an impression of the nature of computational chemistry. Specifically they aim to give an unfamiliar reader first an idea of the problems faced in the analysis of large and complex quantum systems, and also some feeling for the nature of the solutions used. These solutions rely extensively on intuitions and approximations that are based on both mathematical and chemical understanding of the problems.

Overall however, the Hartree-Fock method is only one of a number of common approaches to solving the Schrödinger equation, though these other methods frequently make use of the same basic concepts. The Hartree-Fock method is itself often used as the starting basis for further and more accurate refinements of the orbitals it generates (so-called 'post-Hartree-Fock' methods). Still other methods overlap with the Hartree-Fock method, but replace some of its assumptions with different treatments of the physics. Since in this work extensive use is made of one particular such method, namely Density Functional Theory (DFT), a brief examination of this theory is made below.

1.2.5. Density Functional Theory

Recalling Equation 1.2, the Hartree product, it is noted that the reason one is able to factorise the wavefunction into a product of individual one-electron wavefunctions is because of an initial assumption that the electrons do not interact with one another. It is then later noted that this assumption is dropped, and the electron-electron interaction re-incorporated in the form of the mean field theory.

The assumption that electrons do not interact with one another is plainly false. Electrons interact with one another quantum-mechanically to produce the restrictions of the Pauli principle, that no two electrons can occupy the same quantum state. This is also sometimes known as the exchange effect or Fermi correlation, or even simply as Pauli repulsion. The electrons also interact with one another as negative charges that repel one another. In the Hartree-Fock method, the exchange effect is included via the properties of the Slater determinant *i.e.* it recovers the anti-symmetry of the electrons. The electron-electron interactions however, are more complex. In the definition of the Fock operator (Equation 1.7), it is explained that the Hartree-Fock theory treats the interaction between electrons as if each electron sees all other electrons only as a mean field, incorporated in the term $v^{HF}(i)$. This is a useful concept, and allows the Hartree-Fock method to achieve quantitatively useful results. Despite this, the so-called Mean Field Theory is not accurate enough for the purposes of most theoretical chemists. It assumes the electrons move independently of one another's motion when of course, the motion of an individual electron will also affect the motion of all others. Their motion is *correlated*. Since this correlation is not included in the Mean Field Theory, the Hartree-Fock method is said to ignore the electron correlation. Theoretical chemists therefore describe attempts to include this additional level of electron-electron interaction, over and above the mean field level of the Hartree-Fock method, as *including the electron correlation*.

Density Functional Theory begins with the two Hohenberg-Kohn theorems [10].

The first of these states that the external potential acting on an electron, and thus the total energy, can be uniquely defined in terms of the overall electron density $n(\vec{r})$ alone. The benefit of this reformulation is to reduce the dimensionality of the problem. To calculate the external potential felt by an electron in a field of N electrons, one might be tempted to write the solution in terms of the positions of all the electrons individually, making for $3N$ separate variables. With DFT, there is only the electron density $n(\vec{r})$, which varies only according to the three spatial coordinates. We have thus gone from $3N$ variables to just 3.

The above statement that the density $n(\vec{r})$ determines the external potential $v_{ext}(\vec{r})$ means that the density also determines the form of the overall Hamiltonian and thus what the energy will be. This means that the Schrödinger eigenvalue problem can also be re-written in terms of the density $n(\vec{r})$, and that therefore we can use the variational principle (see Section 1.2.1.1) just as for the Hartree-Fock approach. For an appropriately normalised trial electron density $n(\vec{r})$ such that $n(\vec{r}) \geq 0$:

$$E_0 \leq E[n(\vec{r})] \quad (1.8)$$

This then leads to the second Hohenberg-Kohn theorem, which is that because the density uniquely determines the total energy, and because the variational principle can be applied, the electron density $n_0(\vec{r})$ that minimises the expected total energy $\langle E \rangle$ is the ground-state density.

Finding the overall electron density $n(\vec{r})$ in an analytical fashion is, once again, impossible. Instead, the various contributions to the total energy E are separated into various contributions which are themselves written as functionals of the electron density $n(\vec{r})$. The $n(\vec{r})$ density is then found through an iterative procedure. The result is a series of one-electron-like orbitals not unlike those of the Hartree-Fock equation (1.6). These are known as the Kohn-Sham orbitals [11]. The split contributions to the total energy can be expressed as follows:

$$E[n(\vec{r})] = \underbrace{T_s[n(\vec{r})]}_{\text{kinetic}} + \overbrace{\int d\vec{r} n(\vec{r}) v_{ext}(\vec{r})}^{\text{external potential}} + \underbrace{E_H[n(\vec{r})]}_{\text{uncorrelated Coulomb}} + \overbrace{E_{xc}[n(\vec{r})]}^{\text{exchange-correlation}} \quad (1.9)$$

where the first term is the kinetic energy of a system of non-interacting electrons (once again, we see the correlated part of the electrons' behaviour is separated out). The second, 'external potential' term is the electron-nuclei interaction and, being the only term that depends on the nuclei, is commonly referred to as the *system-dependent* part, whereas the others are referred to as *universal*. All of these terms can be computed exactly, save for the final one, that of exchange-correlation. There are various approximations which are used to compute this term, and they are mostly graded by number of derivatives of $n(\vec{r})$ which they take into consideration.

Collectively these functionals are the density functionals of Density Functional Theory. New functionals with new methods of treating the final, exchange-correlation term of Equation 1.9 are being continuously developed. In this work several different density functionals are used. Those interested in a full development of Density Functional Theory are encouraged to refer to Reference [12].

In summary then, the development of computational chemistry methods is ongoing, and this is true from the most basic level upwards. New ideas and new formulations of older ideas, as well as new computational implementations of both, are appearing continuously. In this work, we investigate and attempt to develop further just one method, that of the pseudopotential.

1.3. Historical Notes

1.3.1. The Pseudopotential

The key to the pseudopotential approximation is the notion of the separability of quantum systems. Experimentally, this means recognising that certain behaviours of such systems are attributable to specific pieces of the system. Theoretically, this is the recognition that a wavefunction describing a system can be factorised into parts approximately governing identifiable pieces of the system. This is however, only an approximation, as in reality quantum subsystems are always coupled, even if only weakly.

The particular separation of interest to us is that of the active electrons of a system (*i.e.* those that take part in chemistry in which we are interested) from the inactive. This generally refers to the separation of the valence electrons of an atom (those of the highest principal quantum number) from the core electrons. This idea is a very familiar one. As it has been expressed previously by Huzinaga [13]:

The very existence of the periodic table attests to the validity of the notion that the valence electrons determine most of the chemical properties of atoms and molecules and the core electrons are inactive and dormant, just shielding the nuclei and providing an effective potential for the valence electrons.

The pseudopotential method aims to supply this ‘effective potential’, removing the core electrons from a system and thereby reducing the number of electronic orbitals that require explicit calculation. They are instead replaced by simple potential functions, or ‘pseudopotentials’.

The pseudopotential in its original formulation was published in 1935 [14, 15], in the early years of quantum mechanics. This pseudopotential treated the core

electrons according to Thomas-Fermi and the single valence electron according to Schrödinger.

From here the theoretical basis was refined and became more mathematically rigorous. First Phillips and Kleinman gave us a one-electron formulation in 1959 [16], in which they considered a one-electron Hamiltonian \hat{H} with ‘core’ and ‘valence’ eigenfunctions ψ_c and ψ_v . They go on to demonstrate that one can transform the basic eigenvalue problem

$$\hat{H}\psi_v = E_v\psi_v$$

to the form

$$\left(\hat{H} + \hat{V}_R^{PK}\right)\chi_v = E_v\chi_v$$

where χ_v is a ‘pseudowave function’ for the valence system only, and \hat{V}_R^{PK} is a repulsive non-local potential. In a one-electron framework such as the Hartree-Fock method, orbitals containing valence electrons have a nodal structure near the core, which is how they maintain their orthogonality to the core orbitals. The pseudopotential method of Phillips and Kleinman chooses not to reproduce this structure. What this means is that the orthogonality constraints of the eigenvalue problem can be replaced by the addition of a pseudopotential, which allows the recovery of the valence solutions.

To create these potentials, a cutoff radius r_c is chosen, inside which the nodal structure is disregarded, and replaced with a nodeless polynomial expansion $f_{lj}(r)$, and outside which the pseudovalence orbitals $\psi_{p,lj}$ are to be kept the same as the original valence orbitals $\psi_{v,lj}$:

$$\psi_{p,lj} = \begin{cases} \psi_{v,lj} & \text{for } r > r_c \\ f_{lj}(r) & \text{for } r < r_c \end{cases}$$

The boundary at r_c should be smooth. This one-electron pseudopotential formalism was then in 1969 generalised by Weeks and Rice [17], resulting in the Generalised Phillips-Kleinman equation

$$\left(\hat{H}_0 + \hat{V}^{GPK}\right)|\Psi_p\rangle = E_v|\Psi_p\rangle$$

in which \hat{H}_0 is the valence electron Hamiltonian, E_v is the total valence energy, and $|\Psi_p\rangle$ denotes a ‘pseudovalence’ eigenfunction for many electrons. \hat{V}^{GPK} is the Generalised Phillips-Kleinman pseudopotential. This formulation was originally developed for use in solid state calculations, though was later adapted to molecular calculations [18].

These potentials can then be fitted, pointwise by a least-squares method, for example. Effects such as relativity are thus included implicitly.

1.3.1.1. Extraction of Parameters

Multiple methods have been used in the extraction of the pseudopotential \hat{V}^{GPK} . They have historically been defined in terms of their consistency either with the ‘shape’ of the reference valence orbitals, or with their energy levels.

The ‘energy consistency’ school of thought is the idea that pseudopotential reference data should be based directly on quantum mechanical observables. In practice this means total energies and differences between total energies over a number of relevant electron configurations and charge states. Restricting the reference data to quantum mechanical observables has an obvious appeal in that unlike the ‘shape consistency’ below, it doesn’t rely on a ‘one-electron picture’ of the wave function. In practice, the criteria are usually total energy differences including the ground state, low-charged cations, anions (if present), and low-energy excited states. Examples of such potentials include the Stuttgart potentials [19].

The ‘shape consistency’ school of thought originates in the notion that the shape of the pseudovalence orbitals should match that of their all-electron equivalents. This makes an obvious physical sense if one hopes to reproduce the

bonding properties of the system to be replaced. This means that in contrast to energy-consistent potentials, which use only quantum-mechanical observables, shape-consistent potentials must make use of all-electron orbital energies as reference criteria *i.e.* quantities "defined within an effective one-electron picture" [3]. These reference orbitals will include both occupied and virtual orbitals. The loss of the nodal structure above for the core region means that the shape-consistency will be true only for the pseudovalence orbitals. Examples of such potentials include the popular Hay and Wadt potentials [20, 21, 22].

In the development of more recent pseudopotentials, there have been efforts to ensure that they can be described as both shape and energy consistent, and a theoretical framework laid out as to how this can be achieved [23].

1.3.2. The Model Core Potential

The term ‘Model Core Potential’ (MCP), whilst it has often been used interchangeably with ‘Effective Core Potential’ and ‘pseudopotential’, has come in general to refer to an alternative formulation of the pseudopotential method defined by Bonifacic and Huzinaga in the 1970’s [24], which has then been progressively refined since [25, 26, 27, 28].

The MCP method relies, in the same way as the pseudopotential method, on the concept of active electrons, which are separable from inactive, or dormant, electrons. In laying out this theory, we shall try to keep to the notation and structure as used in its final form [4].

The heart of the method lies in the modified one-electron Hamiltonian

$$\hat{h}_p(i) = -\frac{1}{2}\nabla_i^2 + \sum_{J=1}^K [\hat{V}_J^{core}(i) + \hat{\Omega}_J(i)] \quad (1.10)$$

where the first term represents the kinetic energy as usual, and where J runs over the K dormant regions of the molecule, in which the electrons are frozen. This section will go on to discuss the origins and roles of the terms $\hat{\Omega}_J$, the ‘energy

shift' operator, and \hat{V}_J^{core} , the 'core-valence' operator. We start with the former, and move on to the latter.

We begin by considering a standard one-electron system with the Schrödinger equation as follows:

$$\hat{H}|\psi_n\rangle = E_n|\psi_n\rangle \quad (1.11)$$

We then define the energy shift operator

$$\hat{\Omega} = \Delta_E|\psi_1\rangle\langle\psi_1| \quad (1.12)$$

where ψ_1 is the one-electron wave function of the ground state. This is then added to the one-electron Hamiltonian \hat{H} to give us the modified Hamiltonian

$$\hat{H}_p = \hat{H} + \Delta_E|\psi_1\rangle\langle\psi_1| \quad (1.13)$$

This Hamiltonian applied to the one-electron wavefunction $|\psi_1\rangle$ gives us

$$\begin{aligned} \hat{H}_p|\psi_1\rangle &= \left(\hat{H} + \Delta_E|\psi_1\rangle\langle\psi_1|\right)|\psi_1\rangle \\ &= \hat{H}|\psi_1\rangle + \Delta_E|\psi_1\rangle\langle\psi_1|\psi_1\rangle \\ &= (E_1 + \Delta_E)|\psi_1\rangle \end{aligned} \quad (1.14)$$

and for all other one-electron wavefunctions $|\psi_m\rangle$

$$\begin{aligned} \hat{H}_p|\psi_m\rangle &= \left(\hat{H} + \Delta_E|\psi_1\rangle\langle\psi_1|\right)|\psi_m\rangle \\ &= \hat{H}|\psi_m\rangle + \Delta_E|\psi_1\rangle\langle\psi_1|\psi_m\rangle \\ &= E_m|\psi_m\rangle \end{aligned} \quad (1.15)$$

i.e. the operator $\hat{\Omega}$ shifts E_1 by Δ_E , and does not affect the other E_m . Provided

this shift is large enough, this means that E_2 is now the ground state energy, with E_1 pushed out of the way.

The implementation of this method for larger systems within the Hartree-Fock regime is relatively straightforward. Since the one-electron picture is already defined (see Section 1.2.1), we can choose dormant orbitals to be shifted in the same way. The Hartree-Fock equations become

$$\begin{aligned}\hat{F}_p|\phi_c\rangle &= (\epsilon_c + \Delta_E)|\psi_c\rangle \\ \hat{F}_p|\phi_v\rangle &= \epsilon_v|\psi_v\rangle\end{aligned}\tag{1.16}$$

where ϕ_c and ϕ_v are the core and valence orbitals respectively, and \hat{F}_p is the modified Hartree-Fock operator

$$\hat{F}_p \equiv \hat{F} + \Delta_E|\phi_c\rangle\langle\phi_c|\tag{1.17}$$

As one can see, this still requires the dormant orbitals ϕ_c . The MCP method makes use of the frozen orbital approximation. This means that the orbitals ϕ_c are taken from a previously-solved set of Hartree-Fock equations. Generally these will be atomic orbitals for the atoms in question. The frozen orbital approximation means that while this method will not give us much in the way of computational gain for atomic systems, calculations on molecular systems will be more efficient, as there are fewer molecular orbitals to converge via the SCF procedure (Section 1.2.3).

The final part of the MCP method comes from noting that even with the dormant orbitals frozen as above, there would still be various two-electron integrals to evaluate involving Coulomb and exchange interactions between the core and valence electrons. In order to avoid this computational cost, the Coulomb and exchange operators are replaced an approximate local function, the so-called ‘core-valence’ operator \hat{V}^{core} , which has taken a number of forms.

Putting all of the above steps together, this results in a molecular Hamiltonian \hat{H}_p that can be applied to a molecule with N_v valence electrons and K non-overlapping dormant regions, of which each pair I and J has $N_{I,c}$ and $N_{J,c}$ electrons, respectively:

$$\hat{H}_p(1, \dots, N_v) = \sum_{i=1}^{N_v} \hat{h}_p(i) + \sum_{i>j}^{N_v} \frac{1}{r_{ij}} + \sum_{I>J}^K \frac{(Z_I - N_{I,c})(Z_J - N_{J,c})}{R_{IJ}} \quad (1.18)$$

Here the final term represents the nuclear repulsion between dormant regions, the second term represents the two-electron interactions between the active electrons, and

$$\hat{h}_p(i) = -\frac{1}{2}\nabla_i^2 + \sum_{J=1}^K [\hat{V}_J^{core}(i) + \hat{\Omega}_J(i)] \quad (1.19)$$

is the one-electron Hamiltonian, complete with the shift operator $\hat{\Omega}_J(i)$ and the core-valence operator $\hat{V}_J^{core}(i)$ for each core J . That which in the Hartree-Fock formalism of the Hamiltonian would be written

$$-\frac{Z}{r_i} + \sum_j 2\hat{J}_{j,c}(i) - \sum_j \hat{K}_{j,c}(i) \quad (1.20)$$

where the last two terms contain the electrostatic and exchange interactions, is instead approximated by $\hat{V}_J^{core}(i)$, a local potential function which has the form

$$V_{mp}(r_i) = -\frac{(Z - N_D)}{r_i} \left[1 + \sum_I A_I e^{-\alpha_I r_i^2} + \sum_J A_J r_i e^{-\alpha_J r_i^2} \right] \quad (1.21)$$

where N_D is the number of dormant electrons in the dormant region. One important point to note here is that these dormant regions referred to above do not necessarily have to be single atoms, and can refer also to groups of atoms. The term ‘core-valence’ operator can therefore be misleading.

The fact that the dormant orbitals are shifted however does not remove them

entirely from our concern. These ‘intruder orbitals’ are now embedded in the virtual space, which is of some concern in calculations that might involve the virtual space, such as multi-configurational calculations or excitation calculations. However, it has been shown that these orbitals (a) can be identified and removed [29], and that (b) if far enough removed from the valence orbitals, have very low occupation numbers, and thus have little effect on results [30].

Overall the distinctive feature of MCPs is the way in which they preserve the nodal structure of the valence electron wavefunction in the core region. This does mean that the basis sets of MCPs are larger and thus do not result in the same computational savings as PPs. They are consequently less used. However, it also means they are not as dependent on the basis set as PPs, and can use full all-electron basis sets with no loss of accuracy.

Extraction of parameters: The extraction of parameters for MCPs, including the frozen core orbitals as well as the potential coefficients and exponents, is done on atomic systems. Such extractions have been done for many elements [31, 32]. Particular variations of MCP theory have been created, for example for relativistic calculations [33, 34, 35, 36] or for DFT [37, 38, 39, 40]. However, in the same way as for the pseudopotential methods above, these effects can also be captured implicitly in model potentials with an appropriate choice of reference system on which to extract parameters. For example, so-called ‘quasi-relativistic’ potentials have been extracted by using Cowan-Griffin relativistic Hartree-Fock equations [41] as a reference [42, 43].

1.3.3. Other Flavours of Pseudopotential

With the advent of modern computing, it has become more common for pseudopotential methods to be used to provide enhancements or corrections to other methods, with examples including their use as link atoms in QM/MM calculations,

[44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65] or the efforts to use them alongside DFT calculations in correcting for DFT errors in the effect of van der Waals forces [66]. Their primary use, however, has for the most part been that of increasing computational efficiency as compared with all-electron solutions for the same systems. Nor have pseudopotentials have been employed only in the treatment of heavy atoms. There have been efforts to create potentials for use on light atoms from the early work of Topiol *et al.* or Gresh and Pullman [67, 68] to the libraries spanning the whole periodic table (see the review by Dolg [69] and references therein) which are still under active development [70].

As Figure 1.1 shows, there is no sign that the importance of pseudopotential methods will diminish in the foreseeable future.

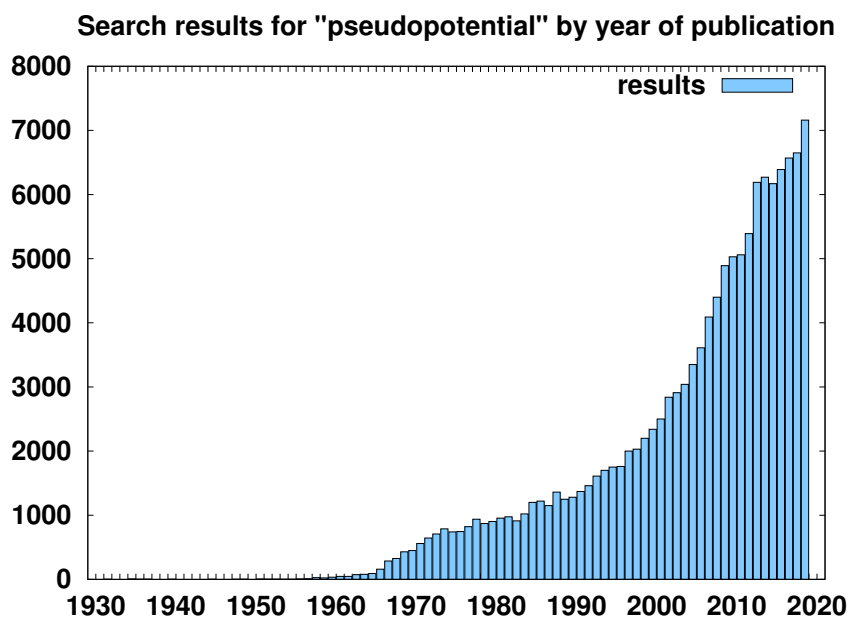


Figure 1.1. Number of Google Scholar results for 'pseudopotential' by year of publication, from 1930 to 2018 [71].

2. Methodology

Sommaire

2.1 Pseudopotential Form	33
2.2 Extraction of Parameters	36

2.1. Pseudopotential Form

In all of the PP and MCP methods described in Section 1.3, the techniques have revolved around separating the electrons of an atom into core and valence electrons. We propose a different formulation. Our goal is to investigate the feasibility of pseudopotentials reproducing electronic behaviour accurately, having been designed not only to replace the core electrons of small molecular fragments, but also to replace specific valence electrons that do not take part in the electronic behaviour in which we are interested. Such potentials could find ready use in performing QM calculations on large hydrocarbons and polycyclic aromatic hydrocarbons, as well as large organic molecules in general.

Before beginning we had decided to restrict ourselves to creating an ‘out of the box’ method, that is to say that no modification of quantum chemistry package source code should be done. This was done so as to keep the potentials independent of the quantum chemistry package used.

In practice, most of the implementation in molecular calculations of the various methods in Section 1.3 amount to the application of sums of modified Gaussian

functions. Sometimes they begin as some other function before undergoing a Gaussian expansion in the manner of Kahn *et al.* [18], other times they are created directly as Gaussians. Since the main quantum chemistry package we used was Turbomole [72], this means our potentials are also of this form. Specifically, they can be written in the form

$$\sum_k A_k (r^{n_k-2}) e^{-\alpha_k r^2} \quad (2.1)$$

where n_k is an integer and where the total number of expressions k can run across multiple such expressions for each angular momentum l . This means that any such potentials can be described in pseudopotential terms as ‘semi-local’, meaning that they have an angular momentum component, but that they are not expanded in the atomic Gaussian basis sets. Equation 2.1 means we have a choice of three variables per potential function, A_k , n_k and α_k . Throughout this work, the value of n_k is fixed at 1. We limit ourselves to a maximum of one potential function per angular momentum l for each potential centre. This is done only for the sake of simplicity and the possibility of increasing this limit is discussed in Section 4.5.1. There is however a precedent of the main group elements being sufficiently well-described by a single potential function per angular momentum [19].

The next departure from most pseudopotential methods is that we remove whole atoms from the molecular fragments for which we create potentials, as well as some further protons. This will become clearer at the beginning of Section 3.1. This means that the frozen orbital approximation cannot be used as in the MCP method. In order then to achieve the correct valence orbital shapes, we use non-atom-centred potentials. This particular choice means it is difficult to make physically intuitive guesses as to what the potential parameters ought to be, though an example is detailed in Appendix A. This means that the pseudowave function is empirical rather than derived. In full then, the multi-centred

pseudopotential operator can be written as

$$\hat{W} = \underbrace{\frac{A}{r} \exp(-\alpha r^2) \sum_m |Y_{1,m}\rangle \langle Y_{1,m}|}_{\text{atom-centred potentials}} + \underbrace{\sum_J \frac{C_J}{r - r_J^0} \exp(-\gamma_J (r - r_J^0)^2) |Y_{0,0}\rangle \langle Y_{0,0}|}_{\text{non-atom-centred potentials}} \quad (2.2)$$

with $Y_{0,0}$ the s spherical harmonic, $Y_{1,m}$ the p spherical harmonics and r_J^0 the relative fixed position of the J^{th} potential with respect to the origin of the pseudoatom to which the potentials are assigned. This \hat{W} operator can then be added to the mono-electronic operator

$$\hat{h}(i) = -\frac{1}{2}\Delta_i - \frac{1}{r_i} + \sum_K \hat{W}_k \quad (2.3)$$

where K is the number of pseudofragments.

We can, albeit rather loosely, say that the atom-centred potentials recover the Coulombic interaction between the active electrons of the fragment and the dormant-electron-screened nucleus (see Appendix A), while the effect of the non-atom-centred potentials is the recovery of part of the bi-electronic interaction between the active electrons and the dormant electrons themselves. This explains their positioning around the pseudoatom. The fact that the extraction of all successful pseudopotentials in this work gave us attractive atom-centred potentials and repulsive non-atom-centred potentials further confirms this intuition.

We draw on the MCP formalism for the theoretical foundation of this work. In 1991 Huzinaga emphasised that the ‘dormant’ electrons of the inactive region had three main effects upon the active electrons, which any pseudopotentials should maintain; these were the Coulomb and exchange interactions, as well as a ‘no-collapse’ term that prevents active electrons collapsing into the dormant region (see Section 1.3.2).

In our method the Coulomb and exchange forces, and usually the no-collapse term, become implicit in the optimisation, which is described more fully below.

2.2. Extraction of Parameters

In previous PP and MCP optimisations, certain physical effects such as relativistic effects in heavy atoms have been incorporated implicitly, and we take a similar approach here. Indeed, we aim to capture almost all of our physics empirically by the choice of reference, all-electron calculations used for the extraction. Specifically this includes the Coulomb interaction and the Fermi correlation implied by the Pauli principle between dormant and active electrons, though as described above the aim is that this is aided to an extent by the spatial arrangement of the pseudopotentials themselves.

One exception to this theme of implicit extraction is in the treatment of the electron correlation. Our potentials are deliberately extracted on Hartree-Fock calculations in order to keep them correlation-independent. This means that the potentials are not ‘biased’ toward any particular post-Hartree-Fock method, allowing them to be used with any. In this work, we mostly use DFT. A consequence of this is that while correlation effects between active electrons are treated at the level of applied theory (DFT), correlation effects resulting from the dormant electrons are implicit in the potential only at the Hartree-Fock level. We believe they should be small enough that the results will not be too-heavily impacted.

Since all our molecular fragments are made up of light atoms we ignore relativistic effects, and extract our parameters on non-relativistic calculations. Finally, while we remove whole atoms from the systems under consideration, and our non-atom-centred potentials do not carry basis sets, we do not optimise the basis sets for the atom-centred potentials (though for certain potentials basis functions are removed (see Section [3.5.1](#))). Instead, we use all-electron basis sets. This is discussed as a possible area for improvement of the method in Section [4.5.1](#).

All of the above leaves us then with only a few empirical parameters to extract.

We use a variety of all-electron system properties as reference criteria, including those which are quantum mechanical observables (such as total energy differences including ionisation and singlet-triplet gap energies) and those which are not (such as orbital energies, excitation energies and least-squares spectra fits).

Overall then, this method draws inspiration from both PP theories (including characteristics of both energy and shape-consistent schools of thought), as well as from MCP theory. While it is true that above we draw on MCP theory for the theoretical foundation of this work, it is far enough removed from it that the general term of 'pseudopotential' seems more appropriate to describe that which we create.

3. Results

Sommaire

3.1	Prototype work	40
3.1.1	Initial Experiments and Thoughts	40
3.1.1.1	Ethylene	43
3.1.1.2	Ionisation and Singlet-Triplet gap Energies	44
3.2	Results for all-trans-polyenes and Polycyclic Aromatic Hydrocarbons	48
3.2.1	All-trans-polyenes	48
3.2.1.1	Triplet Instability	52
3.2.1.2	Comparison with a previous study	53
3.2.1.3	Method Comparison: Frozen Orbitals	55
3.2.2	Polycyclic Aromatic Hydrocarbons	56
3.2.3	CASPT2	59
3.3	Expanding the Method: The MOO program	60
3.3.1	The Optimisation Criteria	61
3.3.2	The Optimisation Procedure	63
3.3.3	Interpreting the Results	65
3.4	Simple Molecular Systems	66
3.4.1	β potentials: sp^2 two-electron pseudofragment	67
3.4.1.1	Optimisation	72
3.4.2	γ potentials: sp^3 one-electron pseudofragment	72
3.4.2.1	Optimisation	76

3.4.3	δ potentials: sp^3 two-electron pseudofragment	76
3.4.3.1	Optimisation	81
3.5	Absorption spectra of Pseudomolecules	82
3.5.1	Optimisation for Spectra	87
3.6	Pseudopotential Studies of Complex Systems	93
3.6.1	Complex Spectra: Helicene	93
3.6.2	Complex Spectra: Twistacene	104
3.6.2.1	Results and Discussion	105
3.6.3	Complex Spectra: Dodecaphenyltetracene	114
3.6.4	Complex Spectra: Nanotube-embedded Coronene	117
3.6.5	Complex Spectra: Hemi-Cryptophane	122
3.6.6	Results and Discussion	123
3.7	Geometry Optimisation	131
3.7.1	Development of the Algorithm	131
3.7.2	Results and Discussion	134
3.7.2.1	Geometric Collapse	138
3.8	Timings	139

3.1. Prototype work

3.1.1. Initial Experiments and Thoughts

As discussed in Section 2, the engineering of any approximation, including a pseudopotential, requires an *a priori* knowledge of the system to be replaced. To begin an investigation of drastically-reduced systems, it made sense to choose as a starting point the sp^2 -hybridised carbon, because the $\sigma - \pi$ separation is very clear. This is well-documented [73]. We began with the CH_3^\bullet radical, it being the smallest sp^2 -hybridised carbon system. We wished ultimately to keep only one electron in the whole molecule, the carbon p_z electron. We removed all other electrons, the hydrogen nuclei, and five of the six protons belonging to the carbon atom.

The σ electron density, whilst it does not contribute directly to the π molecular orbital, will affect the shape and energy of the π orbital by its presence. Therefore we choose to have non-atom-centred potentials in order to increase the flexibility of the potentials in shaping the electronic density, hopefully leading to an accurate description.

The next step was to place our potentials. Placing a single s potential on the sp^2 bond axis itself would result in no overlap between the s potentials and the p_z orbital. Therefore we reasoned that the simplest setup would have two separate s potentials along the σ bond axis as in Figure 3.1.

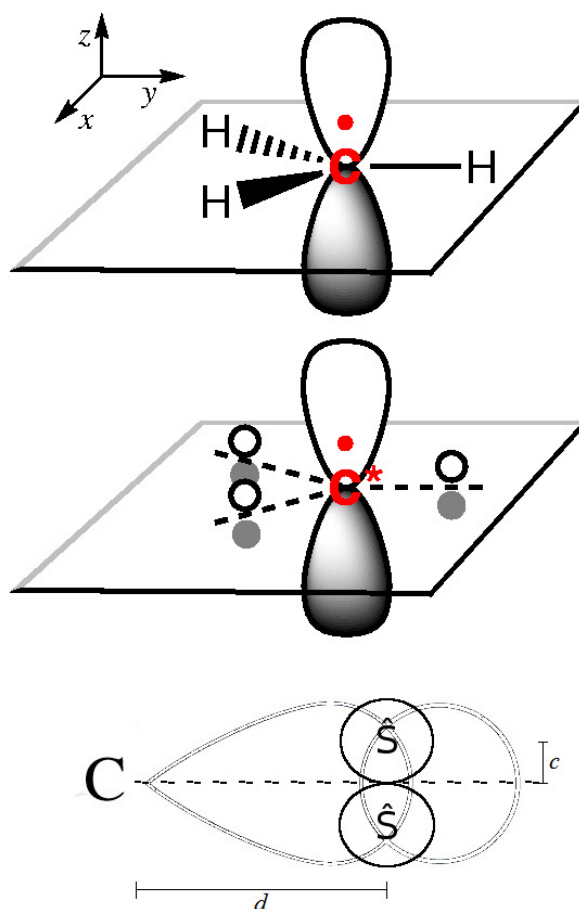


Figure 3.1. Diagram of initial s potential setup.

In the first few attempts, the potentials were placed at the point where the electronic density would have been greatest had the σ electrons been present. This was done by plotting the electronic density of an all-electron calculation and estimating the distance by eye as $d=2.0$ *a.u.*. This worked for the lone CH_3^\bullet . Potentials could be extracted that gave the p_z orbital the same energy as in an equivalent all-electron calculation. Example values are given in Table 3.1.

This setup, however, would present problems in other systems, as the potentials would be far enough away from the pseudocarbon that they would have significant effects on adjacent atoms, often greater than the effect on the pseudo-

Table 3.1. Optimised s potentials for CH_3^\bullet , with $d=2.0$ $a.u.$ and $c=0.25$ $a.u.$. These potentials were optimised to give the all electron energy for an all-electron CH_3^\bullet molecule at the HF level and def-SV(P) basis set.

Calculation Type	Coefficient	Exponent
HF	-7.7041	1.0
DFT-PBE0	-7.1534	1.0

Table 3.2. Optimised s potentials for CH_3^\bullet , with $d=0.5$ $a.u.$ and $c=0.25$ $a.u.$. These potentials were optimised to give the all electron energy for an all-electron CH_3^\bullet molecule at the HF level and def-SV(P) basis set.

Calculation Type	Coefficient	Exponent
HF	-2.5940	1.0
	-4.7888	5.0
	-7.5242	10.0
DFT-PBE0	-2.6050	1.0
	-4.8730	5.0
	-7.6786	10.0

carbon itself. For example, the C(1)-C(2) bond length in the all-electron ethylene in Section 3.1.1.1 is 2.5 $a.u.$, meaning that a $d=2.0$ $a.u.$ would make the potentials belonging to C(1) closer to (and thus more influential on) C(2) than on C(1). It was therefore decided to approach the problem differently. At the expense of the possible physical significance of placing the potentials at the maximum σ density, generality would be aimed for in placing the potentials as close to the pseudo-carbon as possible, thus hoping that they might be more easily-transferable to other systems. Noting that the shortest inter-atomic distance possible in the Turbomole [72] package before geometry problems are encountered is a little under 0.5 $a.u.$, this value was chosen as our potential set distance d , and also as our total s potential split distance, *i.e.* $c=0.25$ $a.u.$ (see Figure 3.1).

With this setup, a range of potentials were optimised with varying parameters, as shown in Table 3.2. In each case, an exponent value was chosen for the potentials and then the coefficient optimised to give the correct p_z orbital energy

for the CH_3^\bullet .

At the same time there was another problem to be solved, to implement the 'no-collapse' term described by Huzinaga (see Section 1.3.2). In the hope of affecting the p_z electron as directly as possible, in this work a p -shaped potential was added to the central carbon.

As described above, the s potentials in the system were of only limited physical significance, and so it would be hard to find values for these potential parameters intuitively. By contrast, the p potential on the central carbon is intended to overlap with the p_z orbital, in order to represent the effect of the electron-screened carbon nucleus. It is thus possible to make some inferences about this potential and what kind of parameters it might have in order to be physically grounded. The details are described in Appendix A, and led to a p potential with a coefficient $A=-3.2680$ a.u. and an exponent $\alpha=0.295$. Ultimately, it was found that either p or s potentials could by themselves ensure the carbon p orbitals were the orbitals lowest in energy, with p potentials lowering in energy all orbitals with p components and s potentials raising in energy the σ orbitals.

With this setup, of a central p potential and non-centred s potentials, we were able to take the CH_3^\bullet system and optimise some new potentials that gave the us the correct HOMO (p_z orbital) energy, these are shown and discussed further below.

3.1.1.1. Ethylene

Having successfully created several CH_3^\bullet potentials that give the correct HOMO energy, a pseudoethylene system was created to see how well they reproduced the π systems we eventually hoped to simulate. Both the s -only potential sets and potential sets that combined both non-atom-centred s and atom-centred p potentials were transferred to the ethylene.

Using only s potentials optimised for CH_3^\bullet , the results were of roughly the cor-

Table 3.3. Pseudoethylene results using both an atom-centred p potential with coefficient $A=-3.2680$ $a.u.$ and exponent $\alpha=0.2950$, and sets of non-atom-centred s potentials. The potentials are optimised to give the all electron energy for an all-electron CH_3^\bullet molecule, using a def-SV(P) basis set, $d=0.5$ $a.u.$ and $c=0.25$ $a.u.$.

Calculation Type	s coefficient	s exponent	π orbital energy (eV)
all-electron system			
HF	-	-	-10.363
DFT-PBE0	-	-	-6.632
pseudopotential system			
HF	2.7721	1.0	-13.654
	6.1732	5.0	-14.011
	10.3810	10.0	-14.061
DFT-PBE0	3.4835	1.0	-10.325
	9.8012	5.0	-10.409
	18.3511	10.0	-12.543

rect order of magnitude but even the better of them were not accurate enough to be of any interest. Clearly the s potential system for CH_3^\bullet does not capture enough of the underlying physics of the system to transfer well to ethylene.

The results for potential sets using both s and p potentials are shown in Table 3.3, and are more interesting. One first thing to note is that when using a p potential centred on the pseudocarbon itself, the s potentials change sign, from negative to positive, *i.e.* they are now repulsive rather than attractive. The next thing to note is that each of the s potential combinations, alongside the p potential, has generated an ethylene π orbital within a few electron-volts of the reference all-electron values. This is promising, but still not accurate enough for much useful chemistry. Therefore, rather than using the CH_3^\bullet system as the base for our optimisations, ethylene itself was used instead.

3.1.1.2. Ionisation and Singlet-Triplet gap Energies

Something one notes from Table 3.3 is that there is a wide range of pseudopotential parameters that give us a suitable HOMO energy for ethylene. This was

a promising sign, as it suggested other constraints than the HOMO energy could be added as criteria for the optimisation of potentials, without forcing us to compromise on the accuracy of the HOMO energy itself. With this in mind we decided to choose further criteria on which to optimise potentials, and this choice is explored below.

Previous pseudopotential methods have made use of different optimisation criteria, based on the intended function of the pseudopotentials. The pseudobond method, for example, is intended to facilitate geometry optimisation in QM/MM calculations, and uses a series of reference data, including bond lengths and angles, across a training set of different molecules.[49]

In our case, the aim is to simulate the electronic behaviour of π systems, and since our reference system is the planar CH_3^\bullet radical (with the four atoms in the xy plane), the HOMO energy of CH_3^\bullet , $\epsilon_{\text{HOMO}(\text{CH}_3)}$, *i.e.* the p_z orbital energy, is an obvious choice, and should ensure the orbital of the pseudosystem is well-positioned to interact with other chemical moieties.

We then take our CH_3^\bullet pseudopotentials and create from them a pseudoethylene, the smallest π system. In addition to the HOMO energy of the pseudoethylene, we decided we would also compare the $\pi - \pi^*$ singlet-triplet energy gap (Δ_{ST}), and the 1st ionisation energy. These energies and energy gaps are shown in Figure 3.2.

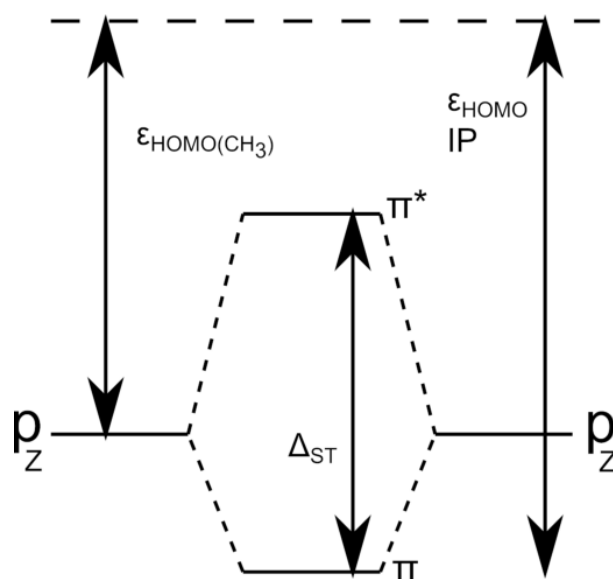


Figure 3.2. Energy level diagram of ethylene.

The reproduction of Δ_{ST} ensures that the first virtual orbitals are well-positioned, which is an indicator of the quality of the system for the reproduction of spectroscopic and chemical interactions, whilst the reproduction of the ionisation potential shows the extent to which the pseudosystem can still behave like the reference system upon losing an electron. As discussed below, the reproduction of this latter property is a particular challenge, as all σ electrons are replaced by the pseudopotentials, and therefore not treated explicitly. Additionally, previous pseudopotentials in the literature, such as those of Igelmann *et al.* [19], have used progressively more ionised molecules to extract potential parameters. The success of these potentials makes plain that this is a natural choice. Finally, there is also the argument from the point of view of shape-consistency. Since it is very important that any active orbitals in the pseudomolecule have the same shape as those of the all-electron system, and since the shape of these orbitals is restricted by other orbitals that are spatially nearby, it makes sense to incorporate into the optimisation process in some way at least some of these other orbitals. Here this is done by the inclusion of the π^* orbital as part of the Δ_{ST} .

Table 3.4. Pseudoethylene results using the *set4* pseudopotentials.

Calculation		$E_{\Delta_{ST}}$ (eV)	E_{HOMO} (eV)	$1^{st} E_{ionisation}$ (eV)
HF				
All-electron		-3.5334	-10.3633	-9.0911
Pseudosystem		-3.5329	-10.0620	-9.8063
DFT-PBE0				
All-electron		4.3374	-7.7960	10.5098
Pseudosystem		4.0428	-7.9782	11.1118
Pseudopotential Parameters				
Potential	Coefficient	Exponent	d (a.u.)	c (a.u.)
p	-3.9096	0.6245	-	-
s	1.5	0.5	0.5	0.25

Later on in this work, it is done by the inclusion of virtual orbital energies.

It was found during the attempts to optimise the pseudoethylene that the parameter that affected the singlet-triplet gap the most was the diffuseness of the p potential. It is suggested that this is because varying the potential affects its relative overlap with the π and π^* orbitals. The assumption in Appendix A (that the p potential exponent should be fixed to maximise the overlap with the p_z orbital of CH_3^\bullet) was therefore ultimately abandoned.

The s potentials were varied by hand and the p potential simultaneously optimised to produce the final set of potentials, whose results are shown in Table 3.4, with HOMO, singlet-triplet gap and 1^{st} ionisation energies for both pseudosystem and all-electron reference system. The relative differences between pseudosystem and all-electron system for the singlet-triplet gap, HOMO and 1^{st} ionisation energies for Hartree-Fock are 0.0%, 2.9% and 7.7% respectively, and 8.3%, 2.3% and 6.1% for DFT-PBE0. These are hereafter referred to as the *set4* potentials.

3.2. Results for all-trans-polyenes and Polycyclic Aromatic Hydrocarbons

3.2.1. All-trans-polyenes

One of the predecessors to this work was a 2013 study by Carissan and Drujon [74], in which another sp^2 pseudopotential fragment was built with non-atomic potentials. It was first decided to test the potentials created in Section 3.1 in a similar manner to those of the 2013 work, and so a range of pseudo-all-trans-polyenes of the form C_nH_{n+2} were created (see Figure 3.3), up to a length of $C_{100}H_{102}$ with a selection of results displayed in Figures 3.4 and 3.5.

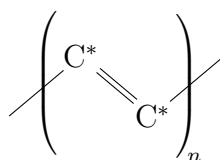


Figure 3.3. Diagram of pseudo-all-trans-polyene unit, with C^* representing a pseudocarbon.

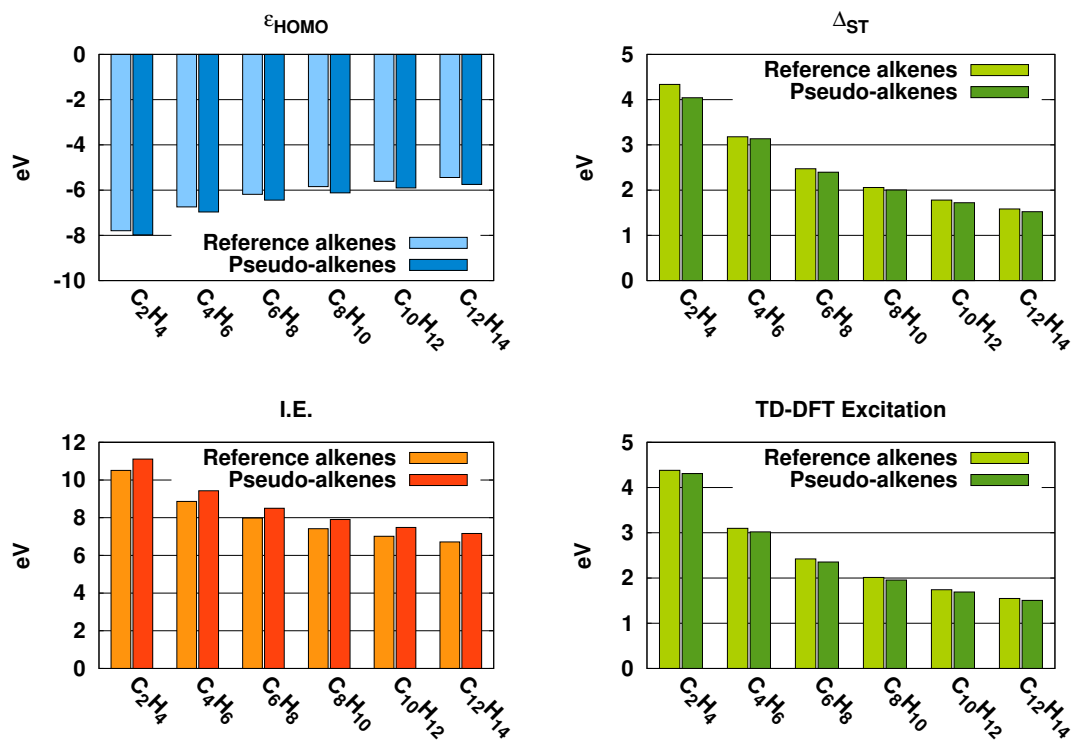


Figure 3.4. All-trans-polyene energy level results. Calculations carried out at the DFT-PBE0 and TDDFT-PBE0 level with def-SV(P) basis sets. The TDDFT calculation is for the first triplet ($\pi - \pi^*$) excitation.

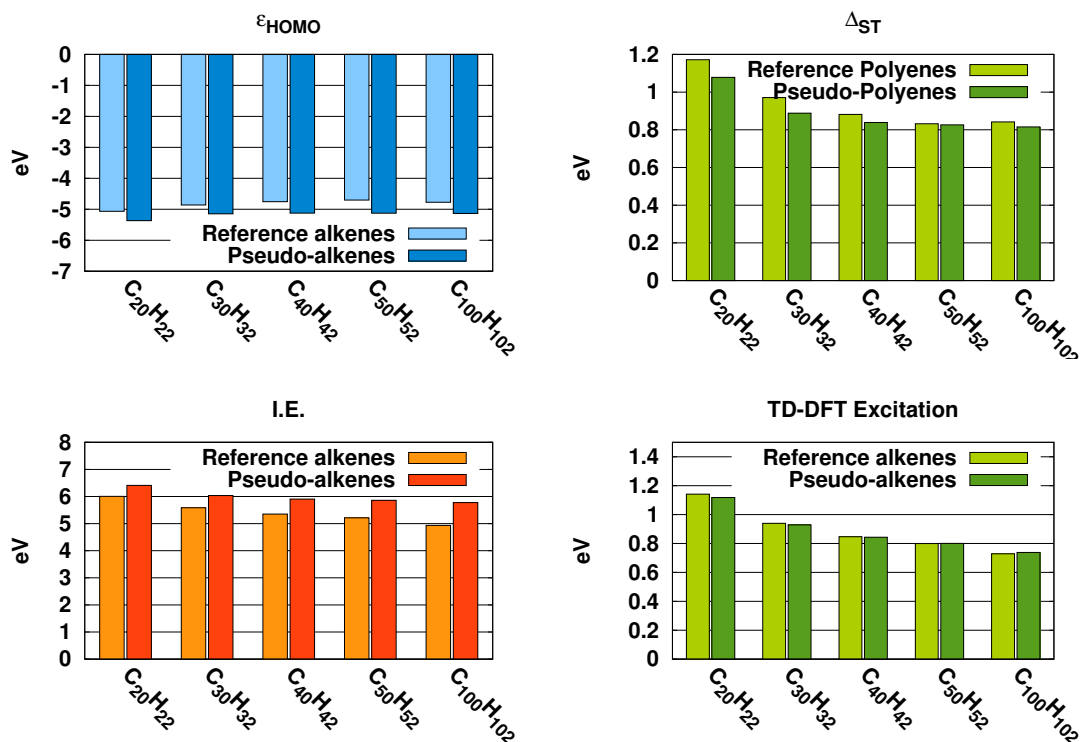


Figure 3.5. Long all-trans-polyene energy level results. Calculations carried out at the DFT-PBE0 and TDDFT-PBE0 level with def-SV(P) basis sets. The TDDFT calculation is for the first triplet ($\pi - \pi^*$) excitation.

Having stated in Section 2 that the potentials should be independent of electron correlation, a trial of different DFT functionals makes sense. Table 3.5 displays the percentage differences between all-electron and pseudosystem calculations, averaged across the trans-polyene systems, for HF, and for DFT and Time-Dependent Density Functional Theory (TDDFT) for a number of different DFT functionals. These include one GGA (PBE) and one meta-GGA (TPSS), along with their hybrid counterparts PBE0 and TPSSh, as well as the hybrid functional B3-LYP.

The non-hybrid PBE functional leads to acceptable results and the addition of exact exchange improves the overall agreement. Indeed, the best agreement is found with the PBE0 functional. The quality of the results does not degrade

Table 3.5. Errors across calculation types for all-trans-polyenes C_nH_{n+2} . For each functional, each molecular geometry was optimised in the singlet configuration, and then the singlet geometry taken for the ion and excitation calculations *i.e.* all excitations are vertical.

Mean Difference (%)	PBE0	PBE	TPSS	TPSSh
Short polyenes (n=2,4,6,8,10)				
1 st Ionisation	7.0	8.8	11.3	9.9
Singlet HOMO	4.2	9.3	13.7	10.7
1 st TDDFT Excitation	2.6	2.7	5.9	6.4
Long polyenes (n=20, 30, 40, 50, 100)				
1 st Ionisation	6.6	10.3	11.3	11.2
Singlet HOMO	7.8	12.8	20.1	15.7
1 st TDDFT Excitation	1.0	6.8	6.1	2.9

with the size of the chains. TPSS and TPSSh however, display larger errors, up to 20% in the energy of the HOMO, which impacts the ionisation potential. Additionally, the TDDFT results are very good.

Figure 3.4 shows in detail the results obtained with the PBE0 functional for the small polyenes. In the pseudosystems, the HOMO systematically has a slightly lower energy than in the reference, but absolute errors are small (between 0.18 and 0.30 eV) and the pattern of increasing HOMO energies with the size of the chain is well-replicated. As a consequence of this underestimation, ionisation energies are slightly higher than in reference calculations, even if tendencies are well-reproduced. We recall here that the ionisation energy will be particularly challenging for any method such as ours, in which only the π electrons are treated explicitly, as we cannot account for the reorganisation of the σ electrons in the cation.

For comparison, the mean ionisation errors obtained with our pseudopotentials are between 7% (PBE0) and 11% (TPSS) and an average error of 9.93% is found in all-electron ROHF calculations for the small all-trans-polyenes where all the σ doubly-occupied orbitals are frozen in the cations, as they are obtained in the calculations for the neutral systems (see Section 3.2.1.3).

A very good agreement between reference and pseudosystems is obtained for TDDFT calculations (for the $\pi - \pi^*$ excitation), which is interesting as it suggests that the virtual space is of good quality. We also compare TDDFT results for this system and results of the 2013 paper, which match to within 3% (see Section 3.2.1.2).

Figure 3.5 refers to longer alkene chains (n up to 100). As for the smaller chains, the properties of large all-trans-polyenes are well-reproduced, which indicates that these results do not degrade with the size of the system. This is confirmed by examining Table 3.5, which displays the average percentage differences between all-electron and pseudosystem calculations across a range of properties, and which shows that these differences are consistent with the small all-trans-polyenes.

3.2.1.1. Triplet Instability

There is one complication in the calculation of pseudo-all-trans-polyenes. As the molecules start to become very long, spin contamination becomes non-negligible, both in the reference and the pseudopotential calculations, and the Δ_{ST} approach is not viable as it is for small polyenes. This is demonstrated by the increasing S^2 expectation value of the pseudosystems compared to the all-electron systems shown in Figure 3.6.

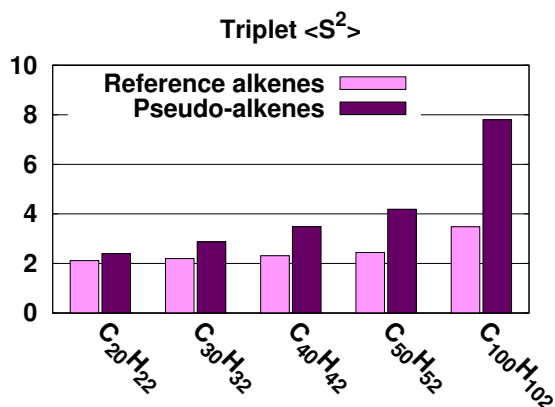


Figure 3.6. Comparison of S^2 expectation values obtained for the calculation of the first π^* triplet configuration in a Hartree Fock formalism, for reference and pseudosystems.

The correct treatment of electronic correlation is essential to recover the physics of the ground and excited states of all-trans-polyenes [75, 76, 77, 78, 79, 80, 81]. Thus, the computation of the excited state has to be done through TDDFT, which leads to an excellent agreement between all-electron and pseudopotential calculations.

3.2.1.2. Comparison with a previous study

One of the predecessors to this work was a 2013 study by Carissan and Drujon [74], in which another sp^2 pseudopotential fragment was built with non-atomic potentials. Here some results of that study are compared with the updated methods of this work.

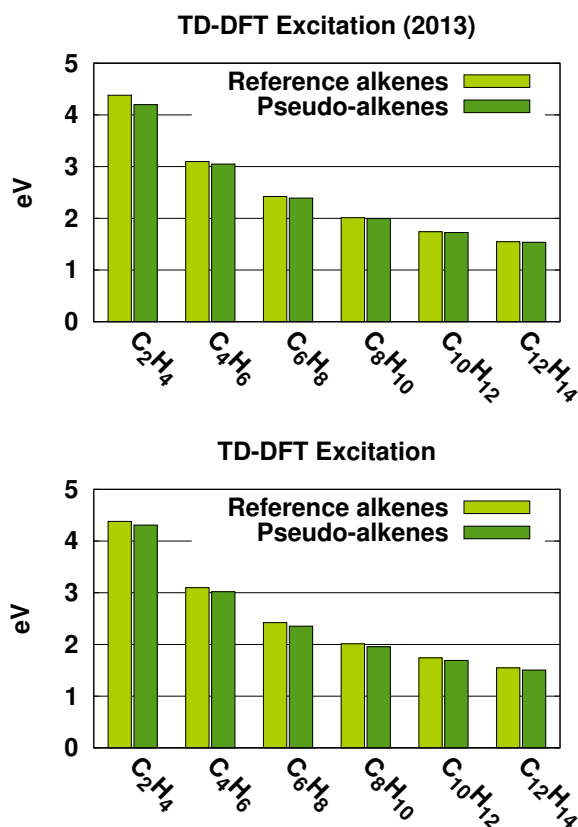


Figure 3.7. Comparison of short all-trans-polyenes with reference [74] and current potentials using TDDFT excitation energies for the first triplet ($\pi - \pi^*$) excitation.

Figure 3.7 shows the excitation energy of the first triplet π excited state for six short all-trans-polyenes. These excitation energies are obtained at the same level of theory both in the 2013 work and the current work. As can be seen there is no degradation of the agreement between reference and pseudosystem calculations. The main difference between both these pseudosystems lies in the fact that in 2013, the pseudopotentials had to be placed exactly at the centre of the C-C bond whereas in the current work, pseudopotentials are placed at a position relative to the centre of the pseudosystem and do not involve the C-C distance (the bond direction is still used).

3.2.1.3. Method Comparison: Frozen Orbitals

We computed the ionisation potentials of the small alkenes at the Hartree-Fock level in the frozen orbital approximation in order to quantify the effect of neglecting the reorganisation of the σ molecular orbitals in the cation. To do that, we froze the σ doubly-occupied orbitals in the cations, as they were obtained in the calculations for the neutral system, Table 3.6 and Figure 3.8. These calculations were performed with the OpenMOLCAS program package [82].

Table 3.6. All-electron Hartree-Fock ionisation energies computed for short all-trans-polyenes. The frozen values are obtained by freezing the doubly-occupied σ molecular orbitals in the cation as they obtained in the neutral system.

	Ionisation Energies (IE) in <i>eV</i>			
	IE Reference	IE Frozen	Error	% Error
C ₂ H ₄	10.06	9.11	0.95	10.38
C ₄ H ₆	8.82	8.09	0.73	9.04
C ₆ H ₈	8.10	7.38	0.72	9.70
C ₈ H ₁₀	7.65	6.96	0.69	9.86
C ₁₀ H ₁₂	7.35	6.67	0.68	10.21
C ₁₂ H ₁₄	7.15	6.47	0.67	10.40
Average			0.74	9.93

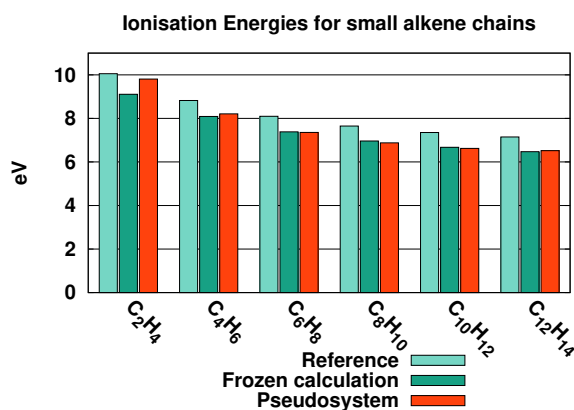


Figure 3.8. Computed ionisation energies for small alkene-chains. Reference calculations are all-electron Hartree-Fock calculations, frozen calculations are all-electron calculations where the σ orbitals were frozen in the cation as they were obtained in the neutral system. Pseudosystem calculations are Hartree-Fock calculations using the *set4* pseudopotentials.

As can be seen from Table 3.6 and Figure 3.8, the difference between regular all-electron Hartree-Fock calculations and Hartree-Fock calculations with frozen σ orbitals for short all-trans-polyenes is 9.93%, and the percentage difference is fairly consistent across the range of polyenes tested.

Figure 3.8 also shows the equivalent Hartree-Fock calculations for pseudopotential systems. The pseudosystems compare well with both frozen and regular all-electron systems, with small differences of consistent size across all molecules. The average difference in ionisation energy between the regular all-electron calculations and pseudosystem calculations in Figure 3.8 is 7.66%.

3.2.2. Polycyclic Aromatic Hydrocarbons

The potentials derived above are also tested on Polycyclic Aromatic Hydrocarbons (PAH). This introduces a new challenge, as it allows us to see if pseudopotentials optimised to reproduce the π bonding of ethylene are also able to reproduce the effect of aromaticity, not present in the ethylene system on which

the *set4* pseudopotentials are optimised. Figure 3.9 shows the set of molecules studied.

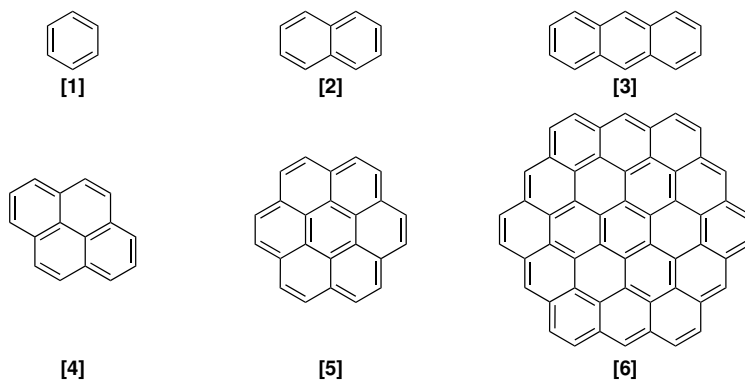


Figure 3.9. Polycyclic Aromatic Hydrocarbons studied in this work: [1] Benzene, [2] Naphthalene, [3] Anthracene, [4] Pyrene, [5] Coronene, [6] $C_{54}H_{18}$.

Figure 3.10 shows the excitation energies, the IE and HOMO energy values for several all-electron and pseudo-PAH, calculated at the DFT level. As with the polyenes, the general trend of the results is well-replicated by the pseudosystems. The differences between the all-electron and pseudosystems are small, and we see that these difference remain of a consistent size across all the systems tested. The results do not degrade as the number of rings increases.

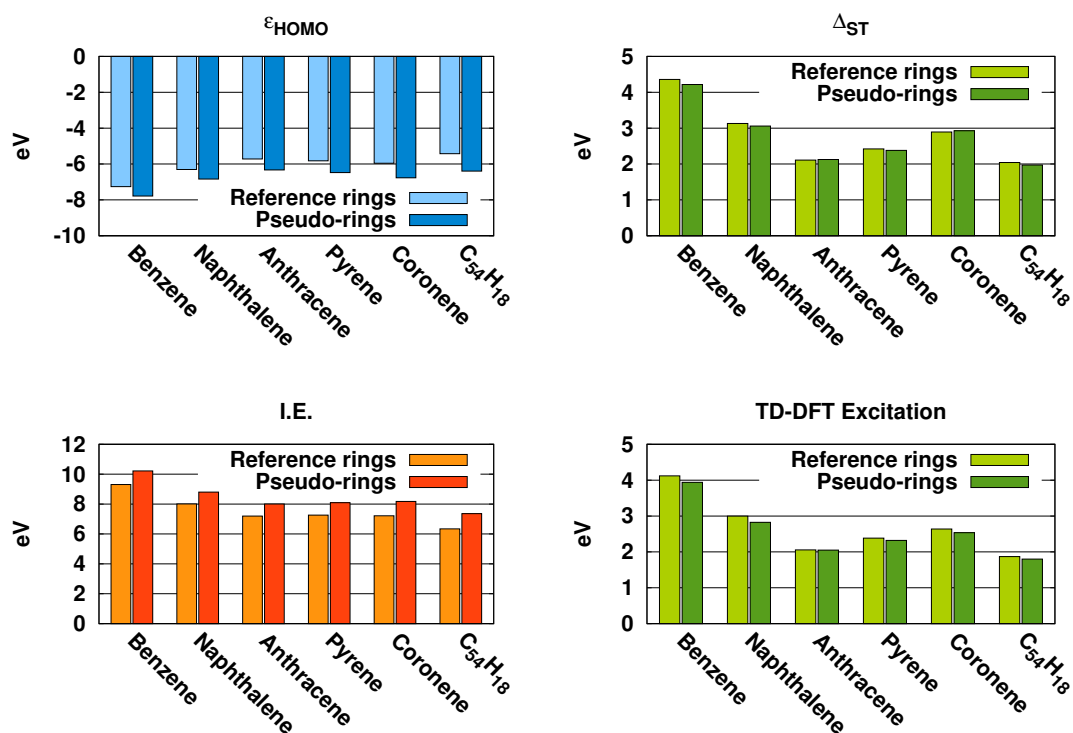


Figure 3.10. DFT and TDDFT (PBE0) comparison of all-electron and pseudosystem energies across a range of Polycyclic Aromatic Hydrocarbons.

Table 3.7 displays the average percentage difference between all-electron and pseudosystems across the same range of PAH systems. We see that the results are similar to those of the all-trans-polyenes. The size of the average differences between all-electron and pseudosystems is slightly larger than for the all-trans-polyenes, ranging up to 17.1% for the PBE HOMO and 22.3% for the TPSS HOMO. Once more we find that the TPSS and TPSSh results are slightly worse than those of the PBE and PBE0 functionals, across all properties.

These results suggest that the pseudopotentials do indeed reproduce the π rings well. This makes physical sense, as aromaticity arises from the fact that the p orbitals overlap and form a ring. Since we reproduce the p orbitals well, the aromatic character of the π ring appears. A more detailed study of the electron density in pseudo-PAH compounds can be found in Appendix B.

Table 3.7. Errors across calculation types for PAH molecules.

Mean Difference (%)	PBE0	PBE	TPSS	TPSSh
1 st Ionisation	12.0	15.4	18.0	16.1
Singlet HOMO	11.5	17.1	22.3	18.7
1 st Excitation (TDDFT)	4.4	2.6	5.9	6.5

3.2.3. CASPT2

The use of our potentials is not restricted to DFT. To demonstrate this, we compute at the CASPT2 level the first singlet-singlet and singlet-triplet excitation energies for benzene, naphthalene and anthracene. For both reference and pseudopotential calculations, the active spaces are (6,6), (10,10), and (10,10), respectively. Results are reported and compared to the TDDFT data in Figure 3.11: trends are well-reproduced and errors with respect to the all-electron calculations are of a similar magnitude to those of the TDDFT calculations.

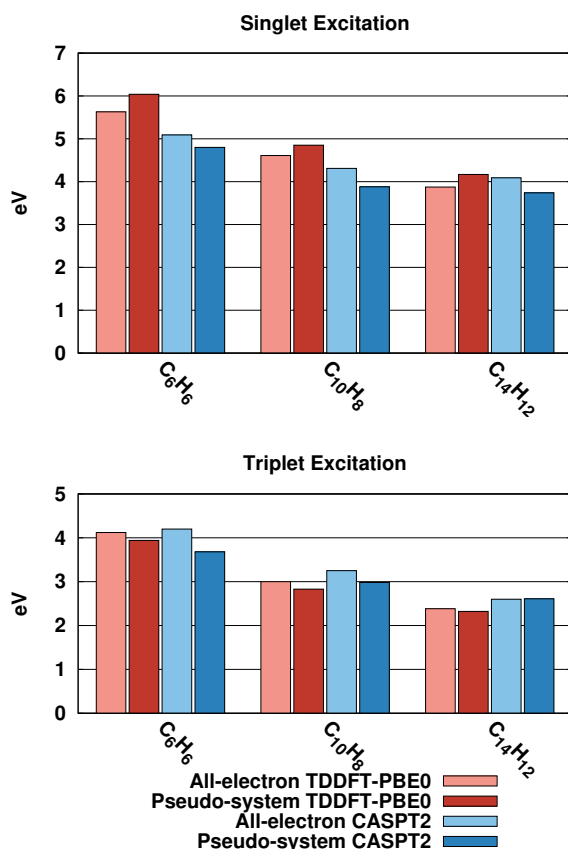


Figure 3.11. TDDFT-PBE0 and CASPT2 comparison of all-electron and pseudo-system first singlet-singlet and singlet-triplet transition energies for benzene, naphthalene and anthracene. The active spaces are (6,6), (10,10), and (10,10), respectively for benzene, naphthalene and anthracene.

3.3. Expanding the Method: The MOO program

We wished to generalise the method developed above, so that it might be used for creating other pseudosystems. Building on the scripts we had developed above we created a general minimisation program, the Multiple Orbital Optimiser (MOO), which could minimise the error between all-electron and pseudosystems. This program is written in Python [83], and works on the following procedure:

1. Accept a series of pseudosystem calculations as an input.
2. Accept a range of observable molecular properties as input, as well as reference, all-electron values for these properties.¹
3. Attempt to minimise these differences iteratively by
 - a) altering the pseudopotential parameters
 - b) running the calculation(s) in Turbomole [72]
 - c) measuring the properties specified in step 2
 - d) calculating the differences between reference and pseudosystems

The minimisation itself is carried out via a Sequential Quadratic Programming (SQP) algorithm implemented in SciPy [84]. This method was chosen for the flexibility of the boundary constraints implemented in SciPy.

3.3.1. The Optimisation Criteria

A number of optimisation criteria were proposed and used in extracting potentials. They are listed below, along with the reasoning behind them.

1. **Molecular orbital energies:** Matching the pseudomolecular orbital energies is a straightforward test of whether the remaining active electrons are subject to the same physical effects as they would be in an all-electron calculation.
 - a) **Occupied orbitals:** Most of the interesting chemistry of any molecule is governed by the highest-energy electrons, and of these, the Highest Occupied Molecular Orbital will be the most crucial orbital to get right if we want a pseudomolecule to behave as an all-electron molecule would.

¹In addition to the above, the user can also weight the various criteria as desired, thus altering the priority of the optimisation criteria.

- b) **Unoccupied Orbitals:** Since the virtual orbital space has so great an effect on the reactivity of the molecule, and since we aim to reproduce electronic structure, it may make sense to fit potentials not only on occupied orbitals, but also on unoccupied orbitals, particularly low-lying ones, such as the Lowest Unoccupied Molecular Orbital.
 - c) **Orbitals of neighbouring atoms:** Since we use only small molecules for optimisation, and since our pseudofragments contain only one or two explicit electrons, many occupied molecular orbitals are likely to be made up mainly of contributions from atomic orbitals belonging to all-electron atoms. If we expect pseudofragments to interact successfully with all-electron atoms, using these orbitals for optimisation may help to ensure the long-range effects of our potentials are correct. It can thus make sense to use the orbitals of relevant all-electron atoms for the optimisation of pseudofragments.
2. **TDDFT excitation energies:** In order to achieve a good reproduction of the virtual space in the pseudopotentials, trying to fit them to TDDFT energies of reference molecules is a logical choice.
 3. **UV spectra fitting:** The production of UV spectra being so crucial to molecular chemistry in general, it may make sense to fit pseudomolecular spectra to all-electron molecular spectra directly. This is done via a least-squares method.
 4. **Total energy differences:** The program is able to run multiple pseudosystem calculations at once, and to measure their total energy differences. This is used to characterise a number of effects:
 - a) **Singlet-Triplet gap:** As already used in Section 3.1, getting the singlet-triplet gap energy correct, even if only for a vertical excitation, is a solid step toward being able to handle molecules with a more complex

electronic structure.

- b) **Ionisation energy:** As used in Section 3.1, ionisation energies are a proven method of optimising pseudopotentials [19].
- c) **Bond stretching:** The potentials in this work are mostly optimised assuming pseudocarbon bonds are to other carbon atoms. In an effort to make the potentials more reliably transferable to systems in which this might not be the case, it was decided to try optimising potentials against carbon-carbon bonds that had been stretched or squeezed from their relaxed geometry.

3.3.2. The Optimisation Procedure

Deciding on the molecular properties to be used as optimisation criteria (along with any weighting judged to be necessary) is only the first step. Next the choice must be made as to which properties of the pseudosystem may be altered in the minimisation. As seen in Section 2, the pseudopotential functions have both coefficients and exponents that can be modified. There is also the spatial arrangement of the pseudopotentials themselves. For the potentials described in Section 3.1, there are the distances d and c , which can be altered by the MOO program in the course of the minimisation. The other potential setups in the following sections have their own geometric properties which may be altered.

A minimisation algorithm hopes to find the global minimum of a function within any boundary conditions specified. However, for complex functions where the boundary conditions are hard to define it is very easy for such algorithms to become ‘trapped’ in local minima. They are thus far more reliable when starting the minimisation from a point on the surface that is already close to the global minimum. As seen from Appendix A, intelligent guesses about where this minimum may lie are sometimes possible. With some intuition, we were able to find a good starting guess for the coefficient and exponent of the p potential of the

Table 3.8. Boundary conditions for pseudopotential optimisation. d is the distance between a pseudocarbon and its s potentials. c is the distance from the xy plane of s potentials in α and β potentials (see Figure 3.1).

Parameter	Lower Bound	Upper Bound
coefficients	-50.0	50.0
exponents	0.001	50.0
d (a.u.)	0.5	1.0
c (a.u.) (α and β potentials)	0.25	1.0

pseudocarbon. However, such intuition is not always feasible. With the MOO program, we also wished to test to what extent brute force was an option for extracting potentials.

An entirely random guess for pseudopotential parameters would likely give us a pseudosystem in very poor agreement with an all-electron equivalent. Indeed, a quantum chemistry calculation on such a system would likely fail entirely. But if some simple and physically reasonable boundaries can be set on the range of allowable guesses, and if a great many guesses were then made within these boundaries, then one might be reasonably sure that at least one of these ‘seeds’ may be close enough to the global minimum as to allow the algorithm to find it.

The boundary conditions used for the selection of pseudopotential parameters are listed in Table 3.8. It was observed in optimisations that weaker and more diffuse potentials that gave the correct results for the chosen optimisation system always transferred better to other systems than stronger and more concentrated potentials. This led to the decision that, rather than taking guesses entirely at random for the values of the potential coefficients and exponents, the values should be chosen from a normal distribution with a mean of $\mu=0$ and with a standard deviation half of that of the upper bound (*i.e.* a standard deviation of $\sigma=25.0$).

Table 3.9. An example optimisation run on a hypothetical pseudosystem. Column ‘Pseudocarbon l ’ shows the angular momenta for which potential functions have been applied. The ‘Criteria’ column lists what reference values have been supplied for the program to attempt to match, as well as any weighting applied in brackets (). Finally, the ‘Best H_{total}/N ’ gives a normalised total error for the best result found by the minimisation program.

Carbon l	Basis	Criteria	Best H_{total}/N	Notes
s, p	def-SV(P)	HOMO; HOMO-1($\times 2$); IE	0.001	None

3.3.3. Interpreting the Results

With the range of criteria and parameters possible for optimising pseudopotentials, it becomes easier to follow results if the details are condensed into a regular format, described in this section.

Table 3.9 displays the results of an example optimisation in a format designed to be as readable as possible. ‘Pseudocarbon l ’ is the list of all the angular momenta for which potential functions are applied to the central pseudocarbon atom. In this example, there is an s and a p potential, each with its corresponding coefficient and exponent. Under the ‘Criteria’ heading, we see the reference criteria against which the minimiser works. In this example, our criteria are ‘HOMO’ (the HOMO orbital energy), ‘HOMO-1’ (the energy of the orbital *below* the HOMO, the ‘HOMO-1th’ orbital) and ‘IE’, the ionisation energy. We also see that the ‘HOMO-1’ criteria has a $\times 2$ weighting. Most of the various criteria on which the potentials are optimised can be written in terms of energy, which allows us to normalise the final error H_{total} by dividing it by the number of criteria (multiplied by any individual weighting applied to them).

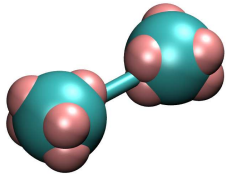
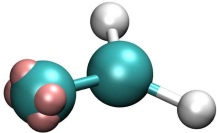
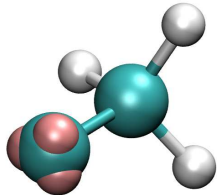
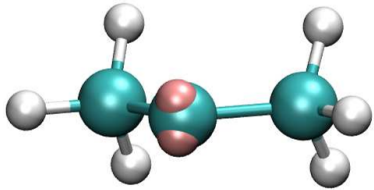
In the sections that follow, results generated using optimised pseudopotentials will be presented first. They will then be followed by the details of the optimisation.

3.4. Simple Molecular Systems

While there is no reason to think that the procedure outlined above would not work for fragments involving elements other than carbon and hydrogen, we did not foresee ourselves trying to create pseudopotentials for many different elements. Therefore, in order that we might simulate as wide a variety of systems as possible, it was necessary to create potentials that could interact with all-electron atoms.

In keeping with our philosophy described in Section 2 of retaining only the chemically-relevant parts of the system, we designed three further potential set-ups that kept only the bonding atoms. These consisted of one more for sp^2 carbon atoms, and two for sp^3 carbon atoms. These we termed β , γ and δ potentials respectively, in addition to the potentials described in Section 3.2, which became the α potentials. All four set-ups are summarised in Table 3.10.

Table 3.10. A summary of different pseudopotential set-ups presented in this work. Non-atom-centred potentials are highlighted in red.

Diagram	Designation	Description
	α	sp^2 1-electron
	β	sp^2 2-electron
	γ	sp^3 1-electron
	δ	sp^3 2-electron

3.4.1. β potentials: sp^2 two-electron pseudofragment

All of the test systems used above for the α potentials contained only sp^2 carbon atoms, meaning that the entire system could be replaced with pseudocarbons. If however we wished to include carbons in any other bonding configuration, or if we wished to include other elements in the system, it would be necessary to treat them explicitly. In order to have an explicit, all-electron system interact successfully with a pseudosystem, an sp^2 carbon potential that is capable of interacting with all-electron atoms is required. The easiest way to create one is to add one more electron to our original potential setup, allowing it to bond to a nearest neighbour. This meant that one of the s potential sets could be removed, as these were intended to recapture the effect of σ electrons which, in the original potential setup, were no longer there. The sp^2 , two-electron pseudocarbon is

depicted in Figure 3.12, as one half of an ethylene molecule.

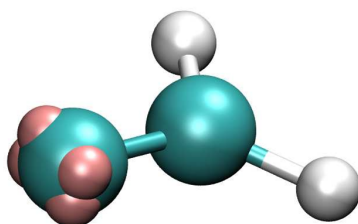


Figure 3.12. An ethylene molecule created using a β pseudocarbon. The near end is the pseudocarbon, the far end is the all-electron CH_2 moiety. Non-atom-centred potentials are highlighted in red.

Once optimised, these pseudopotentials were then transferred to a range of different molecules for testing, chosen to expose the pseudopotentials to a range of different chemical environments. The molecules, along with the sites of the pseudopotentials, are shown in Table 3.11.

Table 3.11. Molecules used to test β pseudopotentials. The pseudofragments are denoted by C^* .

Name	Formula	Pseudomolecular Structure
Ethene	C_2H_4	$\text{C}^* = \text{C}$
Formaldehyde	CH_2O	$\text{C}^* = \text{O}$
Ethenol	CH_2CHOH	$\text{C}^* = \text{CH-OH}$
Methanimine	CH_2NH	$\text{C}^* = \text{NH}$
Ethylamine	CH_2CHNH_2	$\text{C}^* = \text{CH-NH}_2$

Figure 3.13 displays results for the HOMO energy, first ionisation energy, singlet-

triplet gap energy and first excitation energy across the range of different test molecules using β potentials. The majority of the results can be seen immediately to be of a similar quality to the first pseudopotentials of Sections 3.1 and 3.2, and well within half an electron-volt of the all-electron values. The pattern which emerges is that for systems where the potentials are bonded to carbon, the error is relatively small, but where the pseudocarbon bond is to another element such as oxygen (CH_2O), or nitrogen (CH_2NH), then the errors are larger, up to around 30%. Given the potentials are optimised on a potential system bonded to a real carbon, this is probably to be expected. Bonds between carbon atoms have a very different character to bonds between carbon and oxygen or nitrogen atoms.

Figure 3.13. DFT and TDDFT-PBE0 comparison of all-electron and pseudosystem energies across a range of β potential systems.

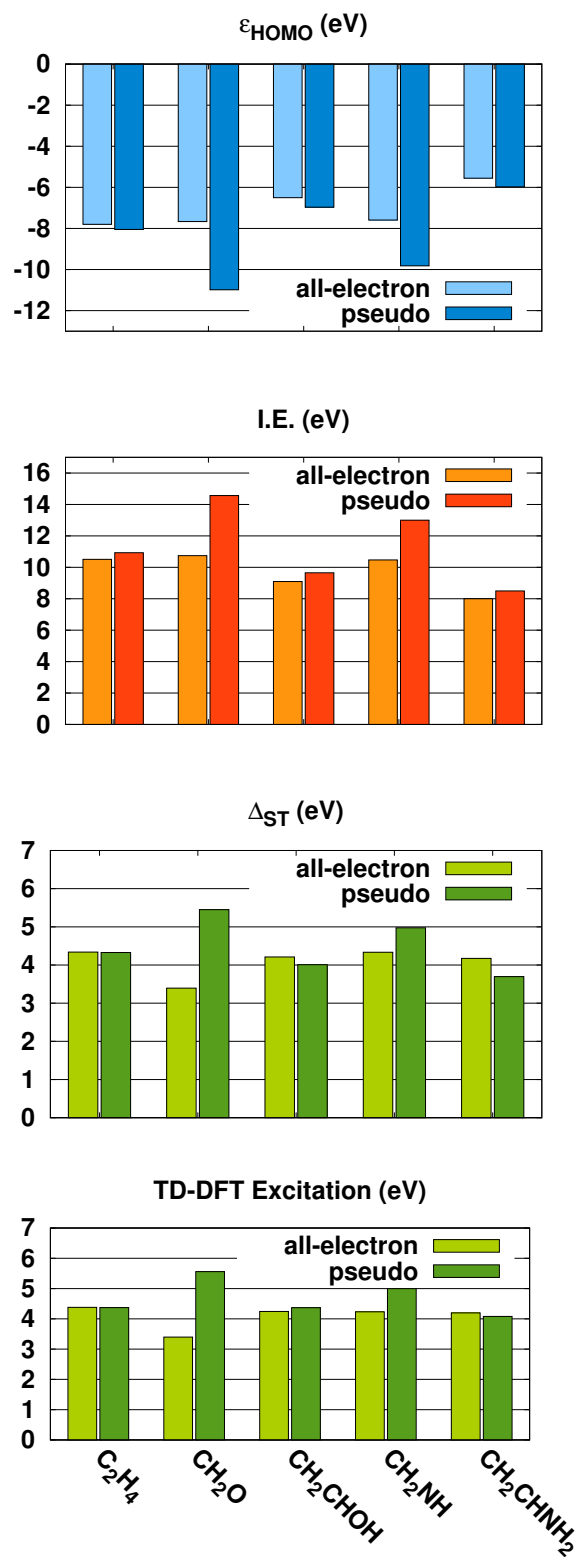


Table 3.12. Average errors for molecules using β potentials, arranged by pseudocarbon-X bond type, for HF, DFT-PBE0, TD-HF and TDDFT calculations.

Bond Type (C*-X)	Mean Error (%)			
	$\Delta_{\text{Singlet-Triplet}}$	HOMO	1 st Ionisation	1 st Excitation
Hartree-Fock				
All	92.4	8.6	29.7	9.3
C*-C	19.1	1.9	12.4	4.1
C*-O	397.6	28.6	90.0	24.0
C*-N	7.4	8.8	21.1	10.0
DFT-PBE0				
All	18.4	18.1	15.2	17.5
C*-C	5.5	5.9	5.5	2.0
C*-O	60.6	43.3	35.6	63.7
C*-N	14.7	29.4	24.1	18.1

Table 3.13. Optimisation criteria and parameters for the best β potential set.

Carbon l	Basis	Criteria		Best H_{total}/N (eV)
p	def-SV(P)	HOMO(*3); HOMO-1; HOMO-2		0.90670
potential	Coefficient	Exponent	d (a.u.)	c (a.u.)
p	-2.0031	0.4694	-	-
s	0.4376	0.4946	0.5	0.25

This trend is quantified in Table 3.12, where the mean difference between all-electron and pseudosystems across a range of properties is shown, broken down by the type of C*-X bond at the pseudo/all-electron interface. When X is carbon, pseudosystem values for DFT-PBE0 calculations are consistently within 6% of the all-electron system values, for all recorded properties. Where X is oxygen or nitrogen however, the DFT-PBE0 mean differences range from 14.7% to 63.7%.

3.4.1.1. Optimisation

Contained in Table 3.13 are the details of the β optimisation. The optimisation used as its reference the top three occupied orbitals in a closed-shell Hartree-Fock calculation on ethylene, with a $3\times$ weighting on the HOMO orbital, and the normalised total error is 0.9067 eV . Overall, this pseudopotential setup does not differ greatly from the α pseudopotentials, other than that these potentials are slightly weaker and more diffuse than the α potentials.

3.4.2. γ potentials: sp^3 one-electron pseudofragment

At this stage we felt the methods developed so far were robust enough to try something different. It was decided to create a one-electron pseudopotential for an sp^3 pseudocarbon, and to optimise it to mimic a methyl group bonded to a carbon atom. This setup has a p -shaped potential on the central carbon, along with three s -shaped potentials replacing the hydrogen atoms at a distance from the central carbon of 0.5 a.u. . The setup is shown in Figure 3.14.

Once optimised, the pseudopotentials were again transferred to a set of test molecules, shown in Table 3.14 along with the sites of the pseudopotentials.

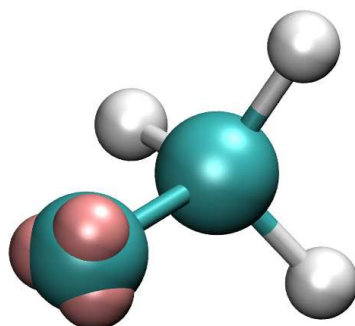
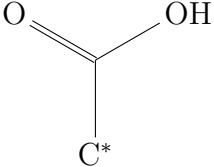
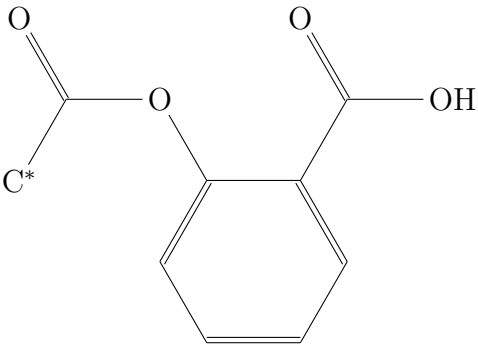
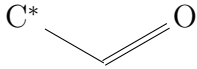
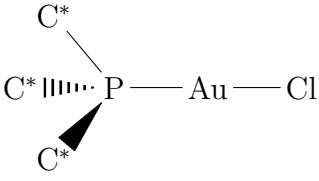


Figure 3.14. An (eclipsed) ethane molecule created using a γ pseudocarbon. The near is the pseudocarbon, the far end is the all-electron CH_3 moiety. Non-atom-centred potentials are highlighted in red.

In Figure 3.15 we see results for the HOMO, first ionisation, singlet-triplet gap

Table 3.14. Molecules used to test γ pseudopotentials. The pseudofragments are denoted by C*.

Name	Formula	Pseudomolecular Structure
Ethane (eclipsed)	CH ₃ CH ₃	C* — CH ₃
Methanol	CH ₃ OH	C* — OH
Methylamine	CH ₃ NH ₂	C* — NH ₂
Ethanoic Acid	CH ₃ COOH	
Aspirin	C ₉ H ₈ O ₄	
Methane	CH ₄	C* — H
Ethanal	C ₂ H ₄ O	
Adapted from Reference [85]	ClAuP(CH ₃) ₃	

and first TDDFT excitation energies for the different molecular systems incorporating a γ potential. These molecules included pseudofragments bonded to groups including -OH, -NH₂, -COOH and an aromatic ring. Figure 3.15 shows us that the results are mostly close to the reference system values. There are however, notable differences in the ionisation and singlet-triplet gap energies of methanol in the pseudosystem with respect to the all-electron system. In this case the γ pseudocarbon is bonded to an oxygen atom. In the β potential system described above, a pseudocarbon being bonded to a non-carbon atom resulted in significant differences between all-electron and pseudosystems. Whilst in Figure 3.15 this remains true for the singlet-triplet gap and ionisation energy of the one pseudocarbon-oxygen bond in our test set, the HOMO and first-excitation energies appear to be sound. One notes also that the C*-N bond of CH₃NH₂ does not appear to have caused any problems for this pseudopotential, with all the pseudo/reference mean differences being comparable in all cases to those of the C*-C systems.

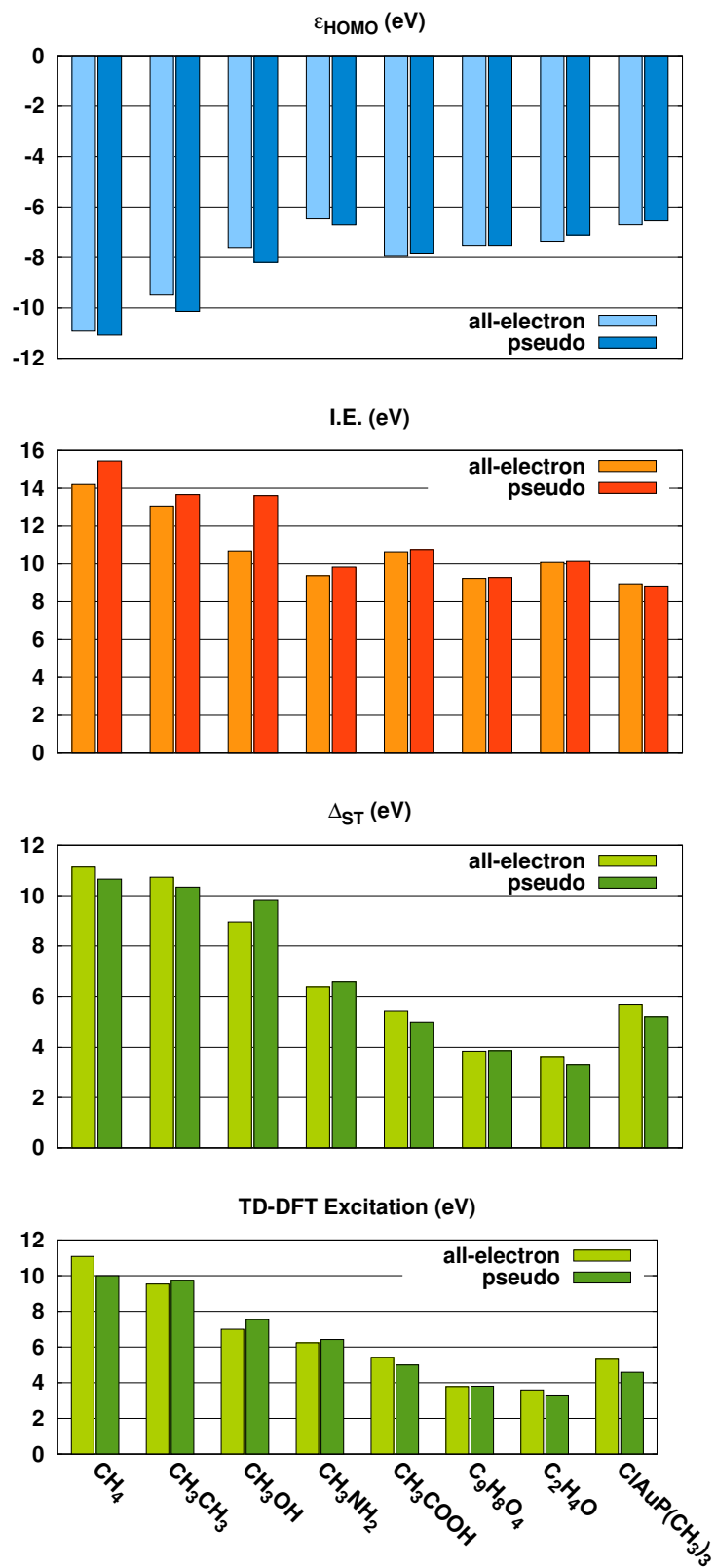


Figure 3.15. DFT and TDDFT-PBE0 comparison of all-electron and pseudosystem energies across a range of γ potential systems.

Table 3.15 breaks down the results above more clearly, showing the reader the mean difference between all-electron and pseudosystems across a range of properties (HOMO, 1st ionisation, singlet-triplet gap and 1st TDDFT excitation energies), broken down by the type of C*-X bond at the pseudo/all-electron interface. The worst results are to be seen where X is either oxygen or phosphorus, with mean differences from the all-electron systems ranging up to 13.7% (the 1st excitation energy of the CIAuPC(CH₃)₃ molecule). In the particular case of the C*-P results, it is worth noting that the pseudosystem is built with three separate sets of γ potentials, which would likely have compounded any error in the potentials themselves. The C*-H bond of methane has also caused difficulties, with ionisation and excitation energies significantly above those of C*-C systems. The fact however that the C*-N bond gives mean differences similar to those of C*-C systems, combined with the fact that the overall differences between all-electron and pseudosystems for non-carbon-bonded pseudocarbons are much lower than those of the β potentials, suggests that either the γ potentials we have optimised are more transferable than the β potentials, or that methyl groups are simply easier to reproduce than sp^2 -CH₂ groups.

3.4.2.1. Optimisation

Contained in Table 3.16 are the details of the γ optimisation. This optimisation was carried out on an eclipsed ethane molecule, using s and p central potentials, with the top two occupied orbitals as a reference (with the HOMO doubly-weighted). These s potentials are far more concentrated than those of either of the sp^2 potentials.

3.4.3. δ potentials: sp^3 two-electron pseudofragment

The final pseudopotentials designed were another sp^3 pseudocarbon configuration. In this pseudopotential setup, there are two electrons, so it is able to act

Table 3.15. Average errors for molecules using γ potentials, arranged by pseudocarbon-X bond type, for both HF and DFT-PBE0 calculations.

Bond Type (C*-X)	Mean Difference (%)			
	$\Delta_{\text{Singlet-Triplet}}$	HOMO	1 st Ionisation	1 st Excitation
Hartree-Fock				
All	12.2	2.9	6.6	4.1
C*-C	6.6	2.4	1.6	5.8
C*-O	27.5	6.2	26.0	1.9
C*-N	19.0	1.6	2.7	6.1
C*-H	5.4	1.2	6.9	5.7
C*-P	4.0	2.8	11.9	0.2
DFT-PBE0				
All	5.1	3.9	7.7	4.3
C*-C	5.3	2.8	1.7	4.6
C*-O	9.5	7.8	12.4	7.7
C*-N	3.1	3.8	4.9	2.9
C*-H	4.3	1.5	8.8	9.8
C*-P	9.0	2.3	1.3	13.7

Table 3.16. Optimisation criteria and parameters for the best γ potential set.

Carbon l	Basis	Criteria		Best H_{total}/N (eV)
s, p	def-SV(P)	HOMO($\times 2$); HOMO-1		0.4179
Potential	Coefficient	Exponent	d (a.u.)	c (a.u.)
p	-2.0949	0.3951	-	-
s	38.5052	14.8328	0.5	-

as a bridge between two other all-electron, bonding atoms. This time the central carbon has not only a p -shaped potential, but also its own s -shaped potential, and we have s potentials replacing two of the all-electron bonds at a distance from the pseudocarbon of $d=0.5$ a.u.. The setup is shown in Figure 3.16 as part of a propane molecule, on which it is optimised.

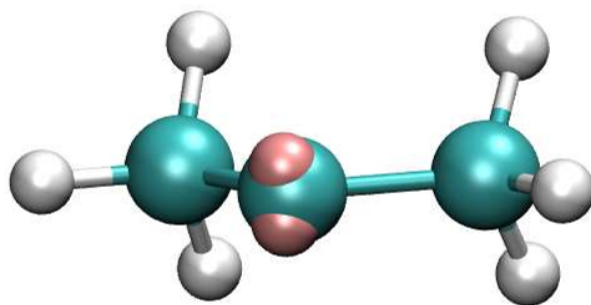
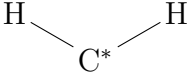
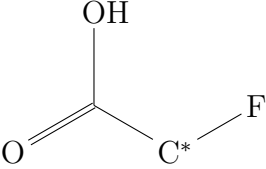
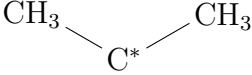
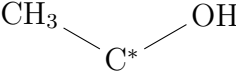
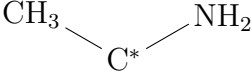
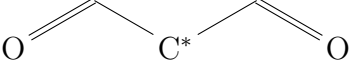
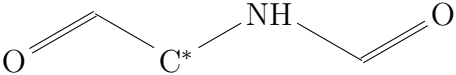


Figure 3.16. A propane molecule with a central δ pseudocarbon between two all-electron CH_2 moieties. Non-atom-centred potentials are highlighted in red.

As above, these pseudopotentials were tested in a variety of small molecules, shown in Table 3.17 along with the sites of the pseudopotentials.

Displayed in Figure 3.17 are HOMO, 1^{st} ionisation, singlet-triplet gap and 1^{st} TDDFT excitation energies for the test molecules and pseudomolecules using δ potentials. Among them are pseudopotentials bonded to groups including $-\text{OH}$, $-\text{NH}_2$ and $-\text{COOH}$. For the great majority of these results the all-electron values are matched closely by the pseudosystems. Notable differences between all-electron and pseudosystem results are found however in methane, and in the 1^{st} excitation energy of ethanol. This remains in keeping with the pattern seen in the other potential sets. In methane, the pseudoatom is bonded to two explicit hydrogen atoms. In ethanol, the pseudoatom is bonded to one explicit carbon atom and one explicit oxygen atom. This pseudopotential setup was optimised on propane, and so both methane and ethanol require the pseudofragment to form bonds for which it was not optimised.

Table 3.17. Molecules used to test δ pseudopotentials. The pseudofragments are denoted by C*.

Name	Formula	Pseudomolecular Structure
Methane	CH ₄	
Fluoroacetic Acid	CH ₂ FCOOH	
Propane	CH ₃ CH ₂ CH ₃	
Ethanol	C ₂ H ₅ OH	
Ethylamine	C ₂ H ₅ NH ₂	
Malonaldehyde	CHOCH ₂ CHO	
-	CHONH ₂ CH ₂ CHO	

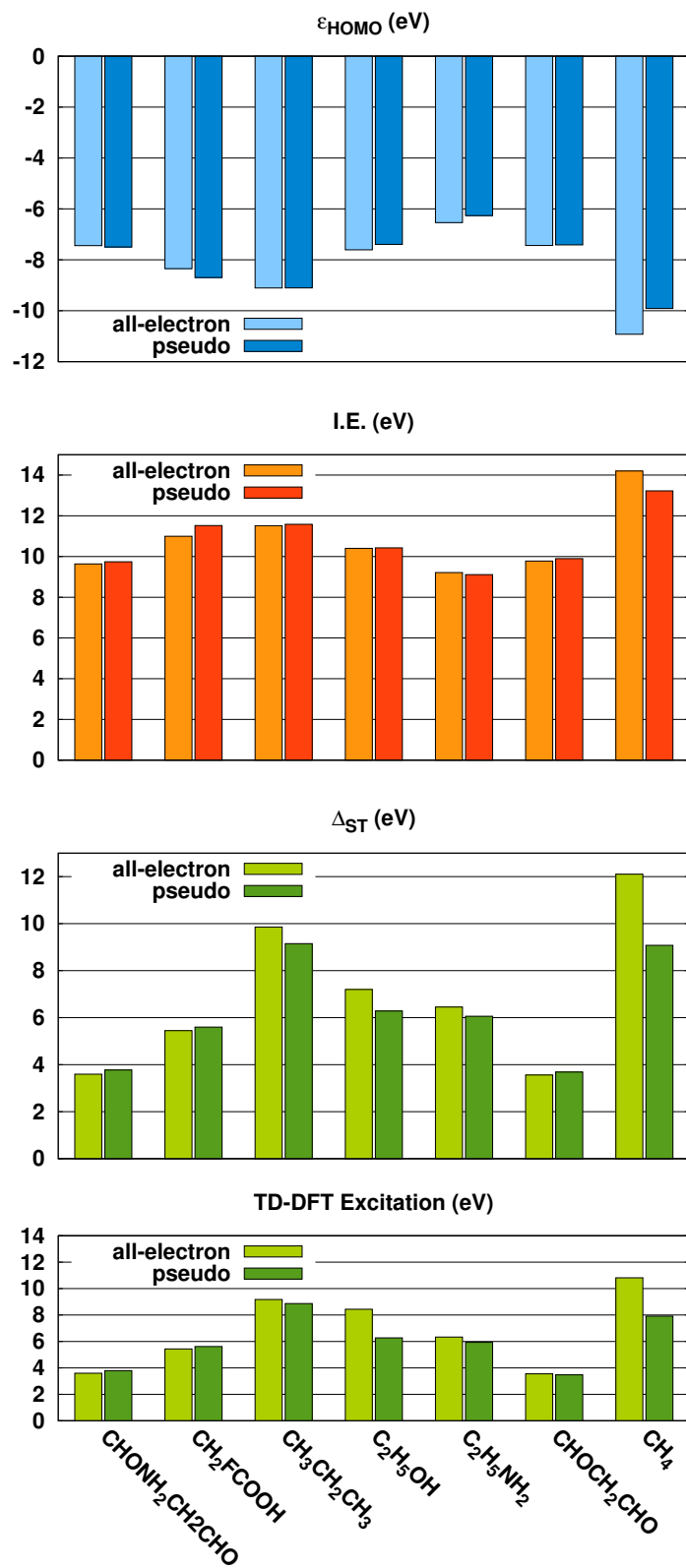


Figure 3.17. DFT and TDDFT (PBE0) comparison of all-electron and pseudosystem energies across a range of δ potential systems.

Further details are shown in Table 3.18, where the results are broken down according to the R-C*-X bond type present between the pseudocarbon and its neighbours. The worst results are those of methane (H-C*-H) with a 23.6% difference in singlet-triplet gap energy and a 34.5% difference in 1st excitation energy between pseudosystem and all-electron system for Hartree-Fock calculations, which does not reduce by much with DFT added. One reason for the particularly poor performance of methane is likely the fact that it is the only system for which both of the bonds of the pseudocarbon are formed with hetero-atoms. The C-C*-N bonds, (those of ethylamine and CHONH₂CH₂CHO) give a poor singlet-triplet result under HF, but a good one under DFT-PBE0. This high average HF singlet-triplet gap error for molecules containing C-C*-N bonds comes overwhelmingly from the CHONH₂CH₂CHO molecule (a 48% difference in singlet-triplet gap energy as compared to the 7.1% of ethylamine), and the reason for this may well be that in CHONH₂CH₂CHO both pseudocarbon bonds, while formed with explicit carbon atoms, are formed with *sp*² carbons, as opposed to the *sp*³ carbons we see in propane (see Figure 3.16). Finally we note that under HF the propane pseudosystem itself has a 17.7% singlet-triplet gap energy difference with its all-electron counterpart, but returns to a 5.3% difference under DFT-PBE0. The reason for this is explored below.

3.4.3.1. Optimisation

Table 3.19 details the optimisation criteria and parameters for the δ potential setup. These are optimised on propane and use both a *p* potential and an *s* potential on the central pseudocarbon, in addition to the non-atom-centred *s* potentials. The reference criteria used are the six highest-occupied molecular orbitals of an unrestricted Hartree-Fock calculation.

Optimisation on this molecule is complicated by the fact that the highest occupied molecular orbitals are all within ~ 1 eV of one another making it very easy for

Table 3.18. Average errors for molecules using δ potentials, arranged by pseudocarbon-X bond type, for both HF and DFT-PBE0 calculations.

Bond Type (R-C*-X)	Mean Difference (%)			
	$\Delta_{\text{Singlet-Triplet}}$	HOMO	1 st Ionisation	1 st Excitation
Hartree-Fock				
All	18.1	1.6	5.0	11.5
H-C*-H	23.6	5.3	9.6	34.5
C-C*-C	17.7	0.2	4.1	8.9
C-C*-O	9.4	0.8	4.2	13.3
C-C*-N	27.6	0.9	4.5	6.2
C-C*-F	3.0	2.9	3.8	2.7
DFT-PBE0				
All	8.9	3.1	2.3	10.5
H-C*-H	25.0	9.3	6.9	26.8
C-C*-C	5.3	0.2	0.9	2.8
C-C*-O	12.7	2.9	0.3	25.7
C-C*-N	5.5	2.5	1.1	5.9
C-C*-F	2.8	4.2	4.8	3.4

Table 3.19. Optimisation criteria and parameters for the best δ potential set.

Carbon l	Basis	Criteria		Best H_{total}/N (eV)
s, p	def-SV(P)	HOMO;HOMO-1;-2;-3;-4;-5		0.40067
Potential	Coefficient	Exponent	d (a.u.)	c (a.u.)
p	-6.9569	3.4066	-	-
s	5.3934	6.4912	0.5	-
s (carbon)	0.6391	0.9382	-	-

the orbitals to end up in the wrong order if care is not taken. This is the reason for the significant change in accuracy between the propane singlet-triplet gap under HF and under DFT-PBE0 in Table 3.18.

3.5. Absorption spectra of Pseudomolecules

Given that the potentials described thus far were mostly capable of reproducing the first excited states of their respective systems, across a variety of different

chemical environments, we decided to make a study of the ability of the potentials to reproduce higher-energy excitations, and to what extent they were capable of replicating molecular electronic spectra. In this study, we decided to continue to keep our focus on the potentials we had studied the most, that is, the α potentials. Reproduction of spectra goes on to become a major theme of this work.

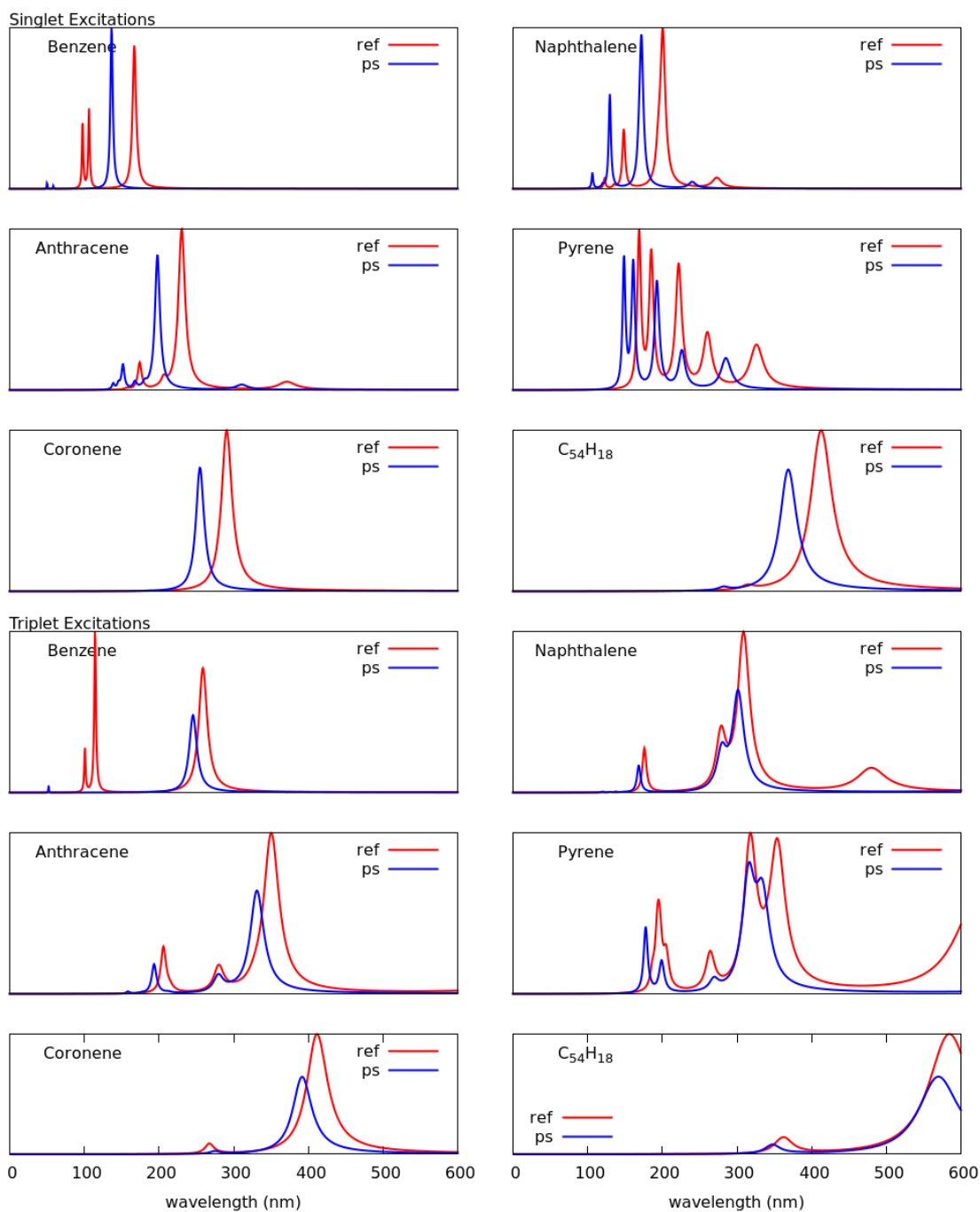


Figure 3.18. Comparison of the oscillator strengths for the 20 first singlet (above) and triplet (below) excitations obtained with *set4* pseudopotentials (ps) and all-electron (ref) calculations (def2-SV(P)/TD-PBE0) within the RPA framework.

Figure 3.18 shows the absorption spectra calculated with the original pseu-

dopotentials of Section 3.1 and all-electron calculations for a selection of PAH, covering in each case the first 20 first singlet π excited states, as well as the corresponding all-electron spectra. It then displays the equivalent oscillator strengths for the triplet excitations. The PBE0 functional was chosen for these calculations as it displayed the most early promise in Section 3.2.

As the pseudopotential calculations contain only π orbitals, excitations to or from σ orbitals cannot be reproduced, two such excitations are visible in the spectra of benzene in the 100-150 *nm* range, for both singlet and triplet excitations. Since we asked for the same number of excitations in both the all-electron and the pseudosystem, and since the pseudobenzene cannot reproduce these peaks, we instead see that the pseudobenzene spectra has two additional peaks at 70-80 *nm*. These peaks therefore correspond to real excitations in the all-electron spectrum which have simply not been calculated. They are not the product of the 'intruder orbitals' mentioned in Section 2.

Other than this however, it can be seen that the spectra of the pseudosystems are excellent reproductions of the all-electron spectra. All the peaks of the reference spectra are clearly identifiable in the pseudosystem spectra, and these peaks in the pseudosystem spectra have very similar intensities and relative frequencies to their all-electron counterparts. This is particularly so for the singlet excitations. One sees that the pseudosystem spectra are consistently shifted by a 30-40 *nm* as compared to the all-electron spectra.

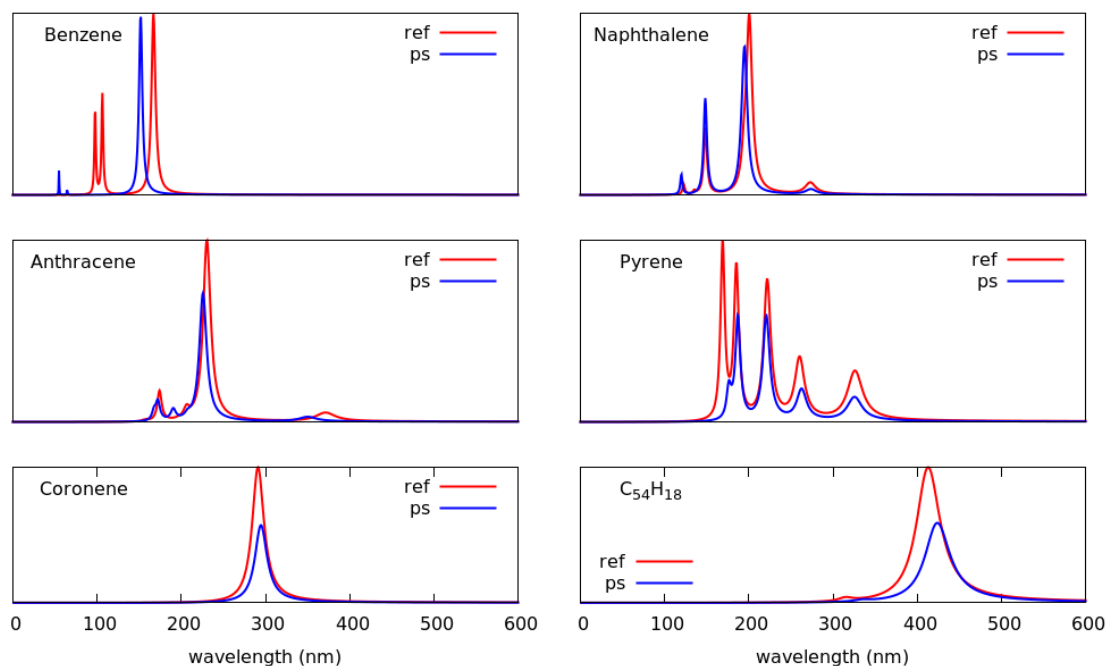


Figure 3.19. Comparison of the absorption spectra for the 20 first singlet excitations obtained with $geom_1$ pseudopotentials (ps) and all-electron (ref) calculations (def2-SV(P)(s-less)/TD-PBE0) within the RPA framework.

Figure 3.19 displays the same excitations for the same series of ring systems as Figure 3.18, however this time the pseudopotentials are a new set, $geom_1$. This set of pseudopotentials is optimised specifically with the reproduction of UV spectra in mind, and the details of this optimisation are described below in Section 3.5.1.

The improvement we desired for these spectra was to shift them closer to the all-electron spectra, and from Figure 3.19 we can see this has been achieved. Most of the peaks of the pseudosystem spectra overlap exactly with their all-electron counterparts, though it does seem there has been a very slight visible decrease in the overall intensity of the pseudosystem spectra, for example in the highest-energy peak in the pyrene spectrum.

Unfortunately, the $geom_1$ pseudopotentials were unable to produce oscillator

strengths for triplet excitations in the manner of the *set4* potentials, as the π^* triplet state was lower in energy than the singlet state. This is discussed further below.

Overall this is a very pleasing result, given not only the degree of overall simplification of the pseudosystems compared to the all-electron systems, but also given the fact that the only ‘training set’ used to optimise these potentials was a lone ethylene molecule treated only at the Hartree-Fock level. This pseudopotential method can be said to accurately reproduce UV absorption spectra.

Intruder Orbitals: In Section 1.3.2 it was noted that the MCP method had the problem of the so-called ‘intruder orbitals’. These are the dormant orbitals which have been projected into the virtual space by the potentials. In Section 3.2 only the first excitation was calculated, and the intruder orbitals were not to be seen. At higher-energy excitations however, there is a risk that the pseudosystems will produce extra peaks not present in the all-electron system. Ultimately we did not see any such peaks for the PAHs. Nevertheless, this is discussed as a possible area for improvement of the method in Section 4.5.1.

3.5.1. Optimisation for Spectra

So far in this work, the measure of success for potentials has been their ability to reproduce the HOMO, 1st ionisation and 1st excitation energies of the all-electron systems they replace. Following the success of the *set4* pseudopotentials in reproducing the spectra of PAHs, an effort was made to optimise pseudopotentials specifically for the reproduction of UV spectra. It was decided that three approaches should be tested for the creation of such potentials:

1. Using the virtual orbitals as reference criteria
2. Using TDDFT excitation energies as reference criteria

3. Fitting the spectra directly via a least-squares method

Using these criteria we generated a great many potentials, most of which were poor fits, either to the reference system on which they were optimised or when transferred to other systems. Here the findings are summarised:

Using TDDFT excitations proved quickly to be a poor optimisation criteria, at least on its own. Fitting pseudopotentials to the first few excitations of ethylene or benzene was possible, but they generally had very poor oscillator strengths and no peaks were ultimately visible in the spectra created. It may be possible to use TDDFT excitation energies as optimisation criteria in conjunction with other criteria, but ultimately no results were generated that warranted inclusion in this work.

Fitting the spectra directly was achieved by adapting the MOO code to use the Peak ANalysing MACHine (PANAMA) program [86] to generate UV spectra on the fly, during the optimisation. Table 3.20 shows an example least-squares optimisation. The H_{total}/N value cannot be calculated for these optimisations as for those using other criteria, as the least-squares difference obviously is not in eV and so cannot be easily compared with them.

Table 3.20. Optimisation table for the *leastsq₁* pseudopotential set.

carbon l	Basis	Criteria		Best H_{total}/N (eV)
p	def-SV(P)	HOMO; UV-spectrum (1^{st} ex)		N/A
potential	coefficient	exponent	d ($a.u.$)	c ($a.u.$)
p	-12.6984	8.6022	-	-
s	-8.8917	37.2870	0.5	0.25

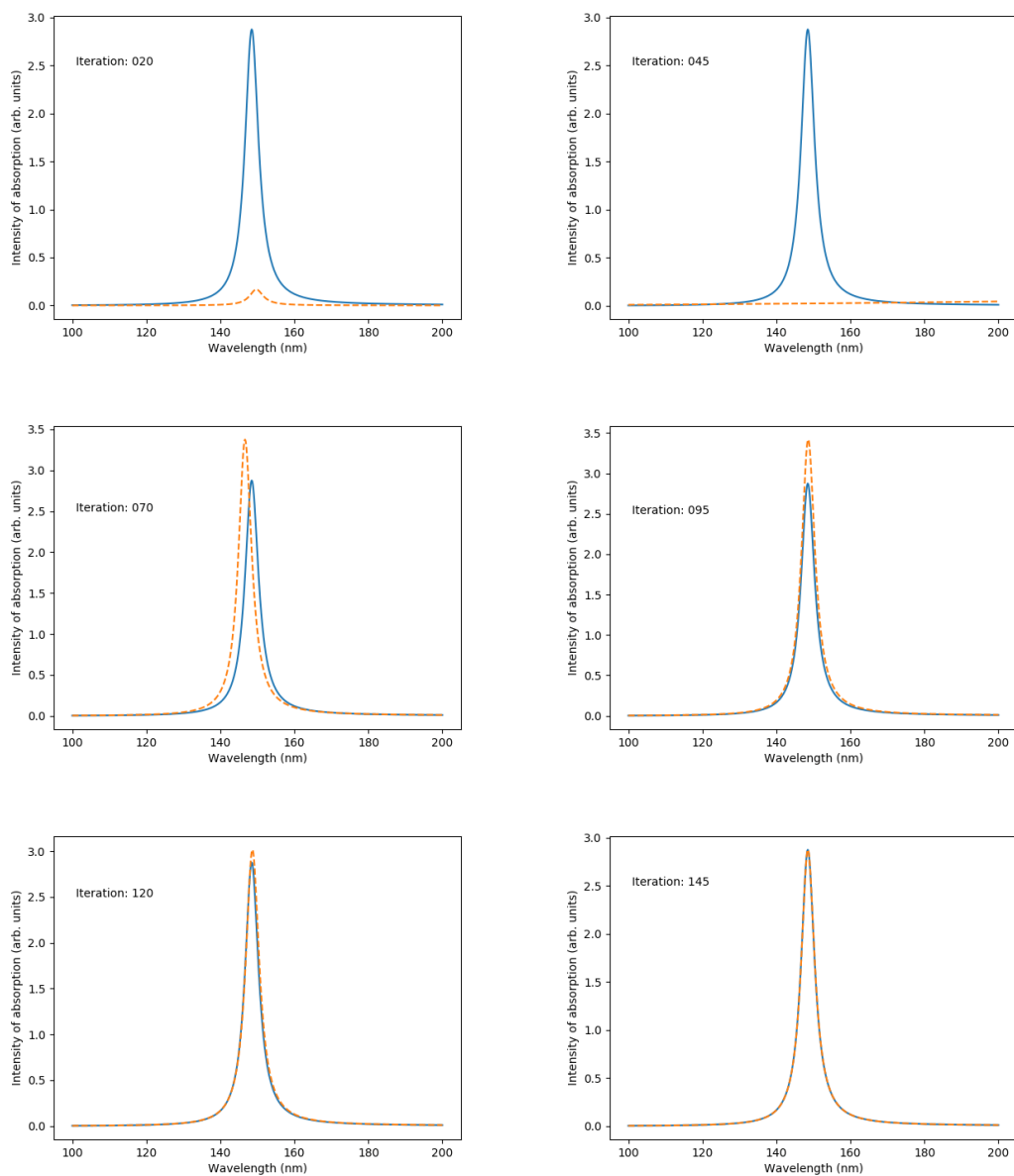


Figure 3.20. A series of images from a least-squares fit of the first $\pi - \pi^*$ excitation of ethylene, with the reference spectrum in blue and the pseudosystem spectrum in dotted orange.

Choosing reference spectra for fitting requires that the spectra will contain only peaks that are present in both all-electron and pseudopotential systems. For ethylene, this means $\pi - \pi^*$ excitations. Another problem encountered was

that the least-squares difference between all-electron and pseudosystem spectra only begins to improve once peaks began to overlap. This means the procedure takes a great many iterations if there is no spectral overlap to begin with, and also that it is very easy for the optimiser to become ‘stuck’, trying to fit partially-overlapping spectra. In practice, this means that least-squares fitting only works well when the spectra are already almost correct. Figure 3.20 displays a series of images taken from a least-squares of the first $\pi - \pi^*$ excitation of ethylene. Once an overlap of the pseudosystem and all-electron spectra is found, the algorithm is quickly able to fit the peak perfectly.

A further difficulty encountered with least-squares fitting was the insufficiency of least-squares fitting as a lone criterion. A successful optimisation using only the spectra (such as that of Figure 3.20) will generate a potential for which the excitations are the correct energy, but for which the orbital energies themselves can still be very wrong. It is therefore necessary to combine optimisation criteria in the minimisation. The energy gaps between states can be found by the least-squares fitting, but it is necessary to pin these states overall to a particular reference energy. The HOMO orbital energy was found to be the most successful one for this task.

The largest problem with this technique however, is the nature of the potentials it generated. We found that even where they were able to reproduce the all-electron spectra accurately as well as the orbital energies, they would not fulfil the ‘no-collapse’ criteria of Section 2. The occupied orbitals would no longer be the ground state. Looking at Table 3.20 we see that these potentials are much more concentrated than those of previous sections, and we suspect this is part of the reason for the failure of these potentials to project the dormant orbitals into the virtual space. This problem is manageable for molecules for which the orbital occupation can easily be forced by the use of symmetry, but would render the potentials unusable for molecules where this is not the case.

Overall this particular technique has potential, and seems an interesting av-

enue for further development, but to be effective it will require a better way of getting rid of the dormant orbitals, either by projecting them into the virtual space or by removing them entirely. It may also require the writing of more specialised code for the generation and measuring of spectra. This is discussed further in Section 4.5.1.

Using the virtual orbitals as reference criteria was found to be the most successful of the techniques tested. Table 3.21 summarises the optimisation of the $geom_1$ potential seen in the UV spectra comparison above. This potential set has precisely the same setup as the original potentials, with one p potential centred on the carbon and six non-atom-centred s -shaped potentials around the central pseudocarbon. One notes looking at the results that the pseudopotential parameters found by the optimiser are not dissimilar to those of the original $set4$ potentials. The coefficients and exponents are all within 0.5 of those of $set4$, and incorporating the distances d and c (see Section 3.1) into the optimisation procedure has resulted in the non-atom-centred s potentials being moved by less than 0.1 $a.u.$ in both directions.

Table 3.21. List of optimisation criteria and results for the $geom_1$ pseudopotential set. This is an α potential using ethylene as a reference.

carbon l	Basis	Criteria		Best H_{total}/N (eV)
p	def-SV(P)	HOMO; LUMO; LUMO+1		0.04037
potential	coefficient	exponent	d (a.u.)	c (a.u.)
p	-3.9020	0.6914	-	-
s	1.2266	0.5448	0.5821	0.2689

The significant difference between the original $set4$ potentials and the $geom_1$ potentials of Table 3.21 is in the optimisation criteria chosen. Rather than the HOMO, 1st ionisation and singlet-triplet gap energies chosen for $set4$, the $geom_1$ potentials are optimised on the HOMO energy, as well as the first two virtual

orbitals. This includes the π^* orbital. The success of this potential as shown above means that ensuring these virtual orbitals are correct is crucial for the accurate reproduction of the spectrum, as one would expect.

All all-trans-polyene and PAH molecules shown in Section 3.2 were then retested using the *geom₁* pseudopotentials. As noted above in Section 3.5, the *geom₁* pseudopotentials were not able to reproduce triplet excitations, and so the results were obtained only for the 1st ionisation energy and the HOMO energy of each molecule. The average percentage difference between all-electron and pseudosystems for the 1st ionisation and HOMO energies respectively were 0.8% and 15.2%. It is apparent from these results that while the use of virtual orbitals improved the ability of the *geom₁* potentials to reproduce the spectra of the PAH molecules, their ability to reproduce the HOMO energy for the same test molecules was decreased, with errors significantly larger compared to the *set4* percentage errors of Section 3.2. This will be a consequence of changing the reference criteria. The 1st ionisation energy of ethylene used in the *set4* optimisation will still depend heavily on the π orbital of ethylene even if it is only half-occupied. The same is true of the singlet-triplet gap energy. This may not be so true of the excitation energies. The fact that the *geom₁* potentials do not use any part of the ethylene triplet energy surface in their optimisation criteria, unlike the *set4* potentials, also seems a likely explanation for the fact that the *geom₁* potentials are unable to reproduce the triplet excitation oscillator strengths above.

Overall then, it can be said that the use of virtual orbitals as reference criteria for optimisation improves the UV spectra produced. Simultaneously however, losing the singlet-triplet gap and 1st ionisation energies as reference criteria has an adverse effect on the ability of pseudopotential systems to reproduce these characteristics.

The above suggests that tuning pseudopotentials to reproduce certain physical characteristics of molecules is possible, but that it is easy to bias potentials with the choice of reference criteria. Finding all-electron reference criteria that

will allow the optimisation of general, ‘all-purpose’ pseudopotentials is therefore difficult.

In the rest of this work we are primarily interested in molecular spectra and so use the *geom*₁ pseudopotentials as a preference, though we also use and make comparisons with the original *set4* potentials.

Removal of basis functions: We found that it was possible to remove the *s* basis functions of the def-SV(P) *geom*₁ pseudocarbon without altering any of the *geom*₁ results. This is promising, as fewer basis functions means a greater gain in computational efficiency. Henceforth in this work, the *geom*₁ potentials are used without *s* functions.

3.6. Pseudopotential Studies of Complex Systems

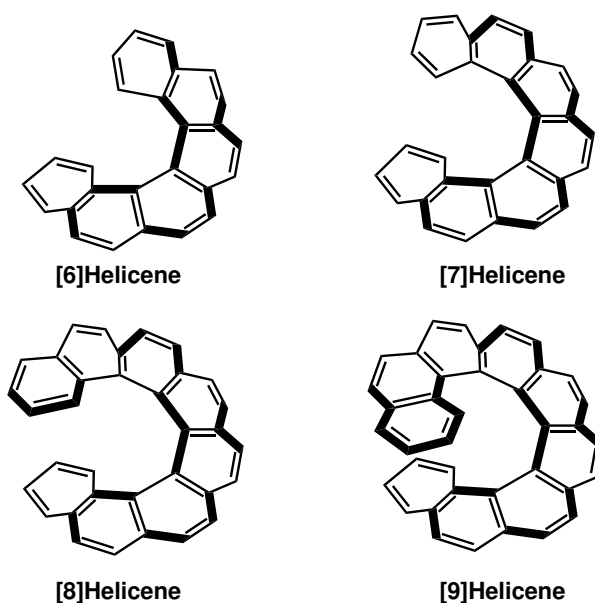
Thus far, all potentials have been tested on systems related closely to those for which they have been optimised. In this section, the various potentials are tested in a series of more complex systems involving heavy atoms, distorted π systems, and neighbouring π rings. These systems were selected from recent literature with the aim of testing the limits of the potential methods thus-far described with their complex molecular and electronic structures. They are focused most heavily on the α -type potentials. All four pseudosystems detailed in Section 3.4 make an appearance, however.

3.6.1. Complex Spectra: Helicene

We were interested in looking at some yet larger and more complex molecules. In particular, we were curious to see how the potentials would perform in a system that wasn’t planar, as distortions in the plane would, in an all-electron system, alter the influence of the *sp*²-hybridised electrons. In a comparison between

CC2 theoretical calculations and experimental results, Nakai *et al.* have shown that Electronic Circular Dichroism (ECD) spectra of helicenes can be accurately computed (see Figure 3.22) [87]. Helicene is an excellent test candidate for this work, as we have already tested pseudopotentials on extended π systems, and so the only untested chemical environment we would be introducing with helicene would be the non-planarity of the system.

Figure 3.21. The structure of (P)-helicene, from 6 to 9 rings.



First of all let us show here that for small helicenes (6 to 9 rings), it is possible accurately to reproduce the experimental spectra with all-electron calculations (DFT-PBE0/def2-SV(P)). In this work, we are interested in carbon based helicenes made of ortho-fused six-member-rings only, like those displayed in Figure 3.21.

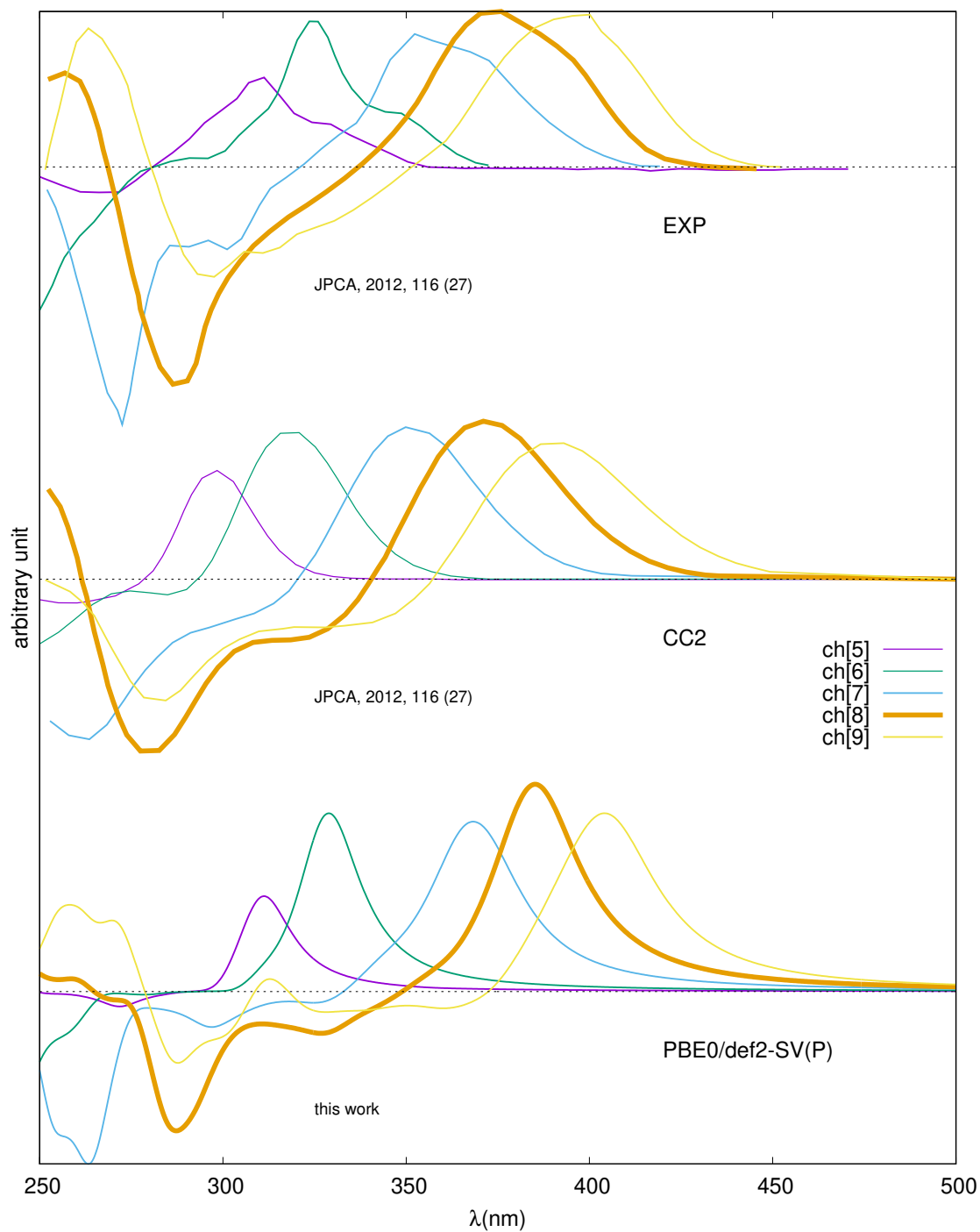


Figure 3.22. Comparison of experimental and all-electron CC2 and all-electron PBE0/def2-SV(P) computed ECD spectra for helicenes of length 5 to 9 rings.

Figure 3.22 shows a comparison of ECD spectra of helicenes from a length of 5 rings to 9 rings, between experimental results, and CC2 and DFT-PBE0/def2-SV(P) calculations. The agreement between experiment, and all-electron CC2 and DFT calculations is very good. Relative intensities, sign and positions of the absorption peaks are reproduced very well by the DFT calculations, with no shift in the wavelengths. Recall that when the helix followed by a helicene rotates in the counter-clockwise direction, the helicene is labelled P (plus) and M (minus) otherwise. The ECD spectra of P helicenes feature a positive absorption at low energy followed by a negative absorption at a wavelength about 100 *nm* further (the spectrum of [8]helicene is plotted with a thicker line to emphasise this feature). This characteristic behaviour (in the above defined framework) is indicative of the chirality M or P of the helicene.

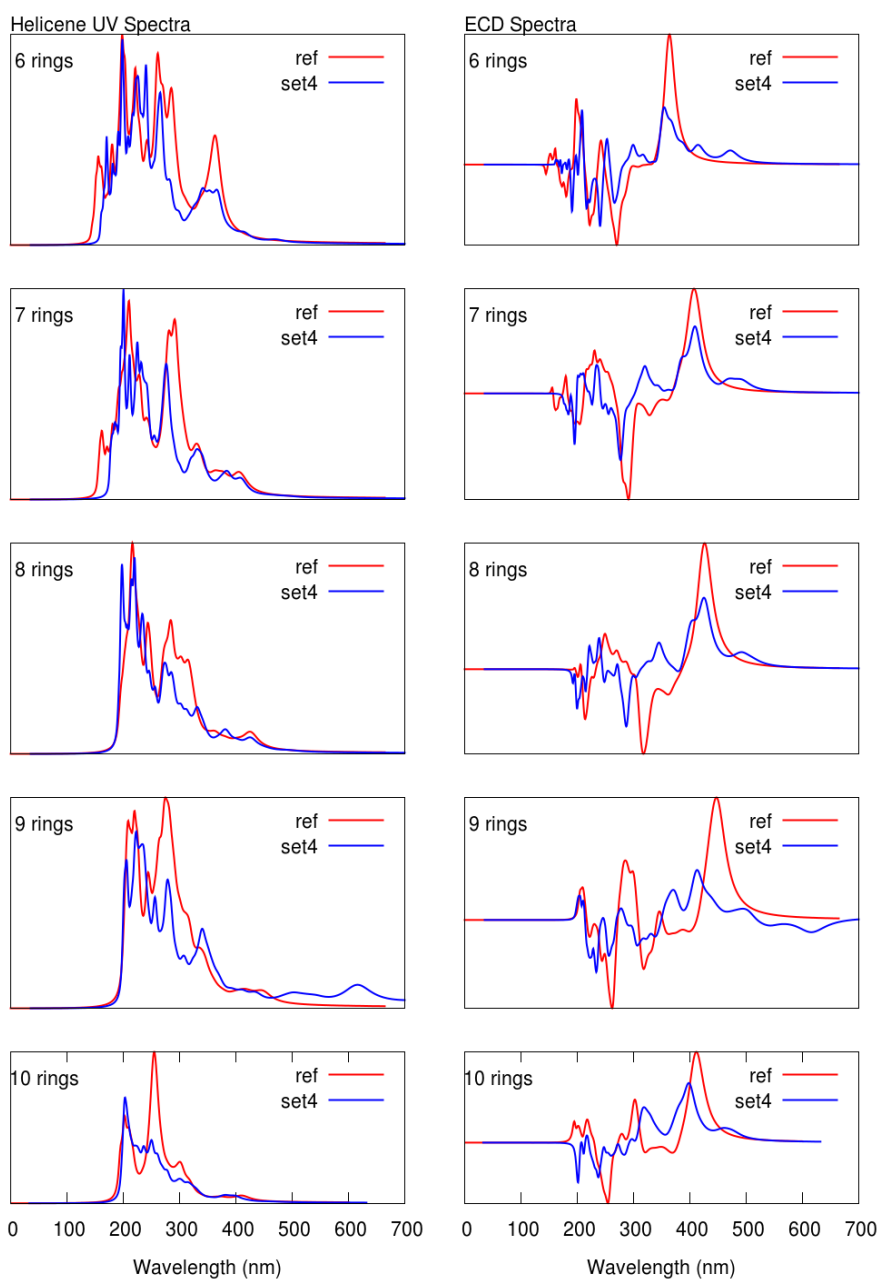


Figure 3.23. Helicene UV and ECD spectra, for all-electron (red) and *set4* pseudopotential (blue) systems. Calculations are performed at the TDDFT-PBE0 level.

Figure 3.23 displays the UV and ECD spectra for a series of P-[*n*]helicene molecules of $n = 6$ to $n = 10$, along with their pseudomolecular counterparts. After red-shifting the pseudosystem spectra by 32 nm, we can see the general

shapes of the pseudosystem UV spectra are qualitatively very similar to the all-electron spectra, both in their spread and their intensities. We can also see that many of the peaks of the all-electron spectra appear to be clearly identifiable in the pseudosystem spectra.

The ECD spectra are more complex. The general shapes of the pseudosystem spectra are distorted compared to the all-electron results, with the distortion increasing from slight to more severe as the helicene becomes longer. However, one can still clearly identify the largest peaks for the smaller systems, and the chiral signature of the dichroism remains clear throughout.

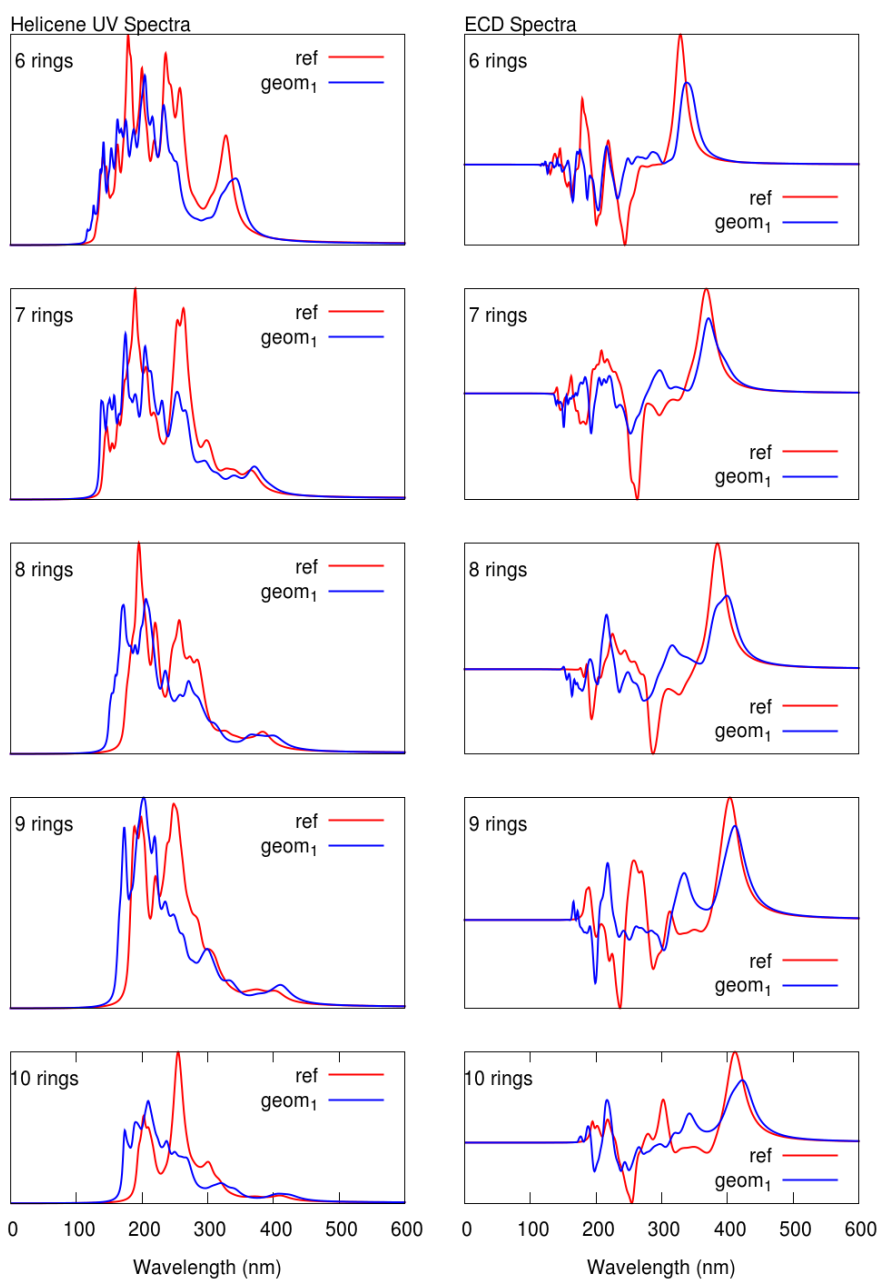


Figure 3.24. Helicene UV and ECD spectra, for all-electron (red) and $geom_1$ pseudopotential (blue) systems. Calculations are performed at the TDDFT-PBE0 level.

Figure 3.24 displays UV and ECD for the same range of molecules as Figure 3.23, with spectra for both all-electron and pseudomolecules. This time however, the pseudopotentials used are the $geom_1$ potentials from Section 3.5, as

opposed to the original *set4* potentials from Section 3.1. Comparing the spectra for *geom₁* and *set4* potentials, the first thing to note in is that the *geom₁* potentials do not require a shift in wavelength in order to line up the pseudosystem spectra with the all-electron spectra, which is an improvement. Other differences between the two pseudopotentials are present, but are more subtle. *set4* appears to generate more low-energy peaks that are not present in the reference spectra (particularly for the ECD spectra), whereas *geom₁* has additional peaks at the high-energy end of the spectra. As noted in Section 3.5, the high-energy peaks are not necessarily unphysical. Aside from this the two pseudopotential spectra are very similar, both in the distribution of peaks and in their intensity.

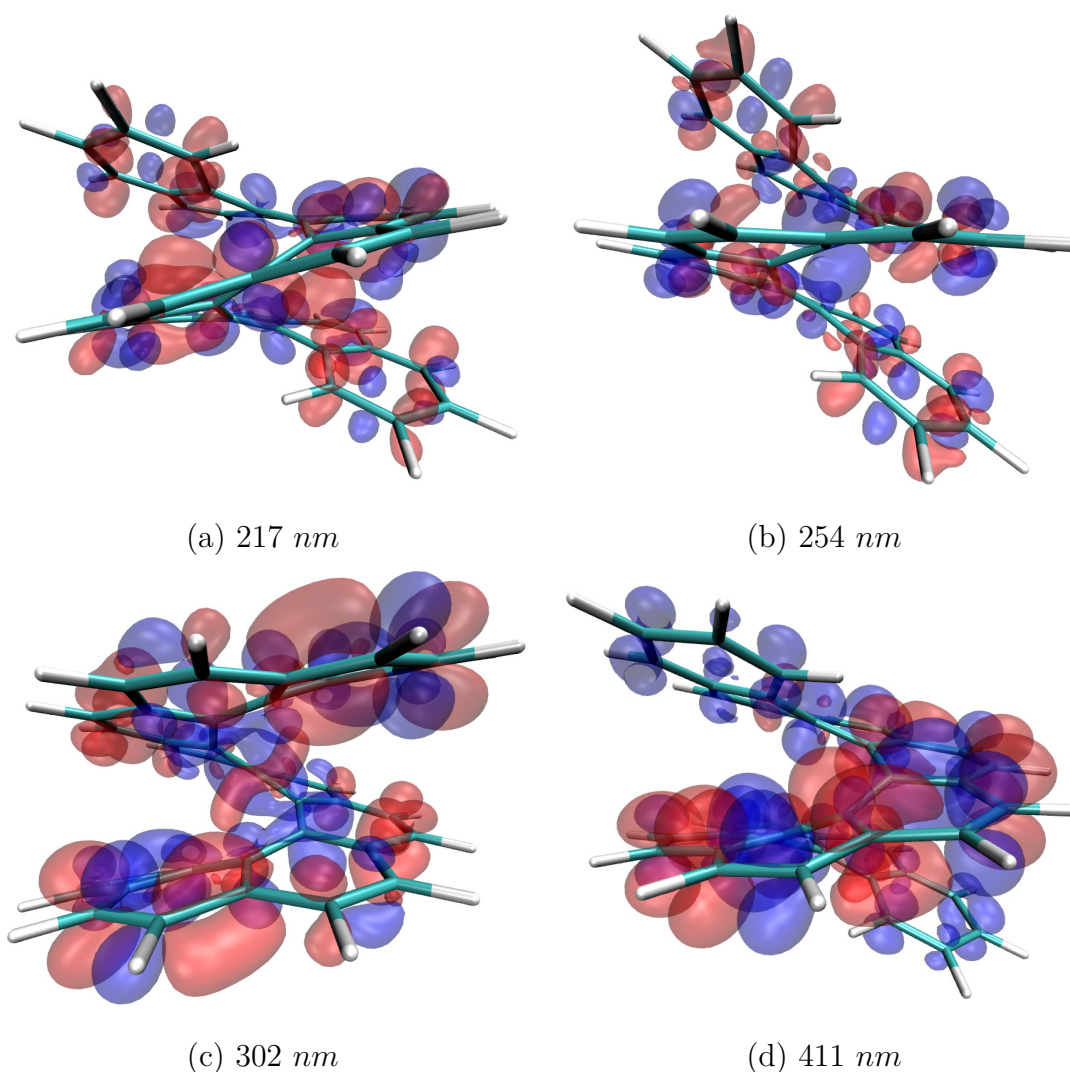


Figure 3.25. Transition densities for [10]helicene. Excitations from left to right, top to bottom: (a) 217 *nm*, (b) 254 *nm*, (c) 302 *nm*, (d) 411 *nm*. In each case, the charge transfer is from the blue regions to the red. These are all-electron calculations performed at the TDDFT-PBE0 level.

Figure 3.25 displays transition densities for the four most distinct peaks in the all-electron [10]helicene ECD spectrum. One sees that for all four, most of the electron density exhibits a strong π character. The peak for which this is least true is the excitation at 302 *nm*, Figure 3.25c, which has a noticeable degree of electron delocalisation around the centre of the helix. This may go some way to explaining why it is this particular peak that is least well-reproduced by the

pseudopotentials, particularly $geom_1$.

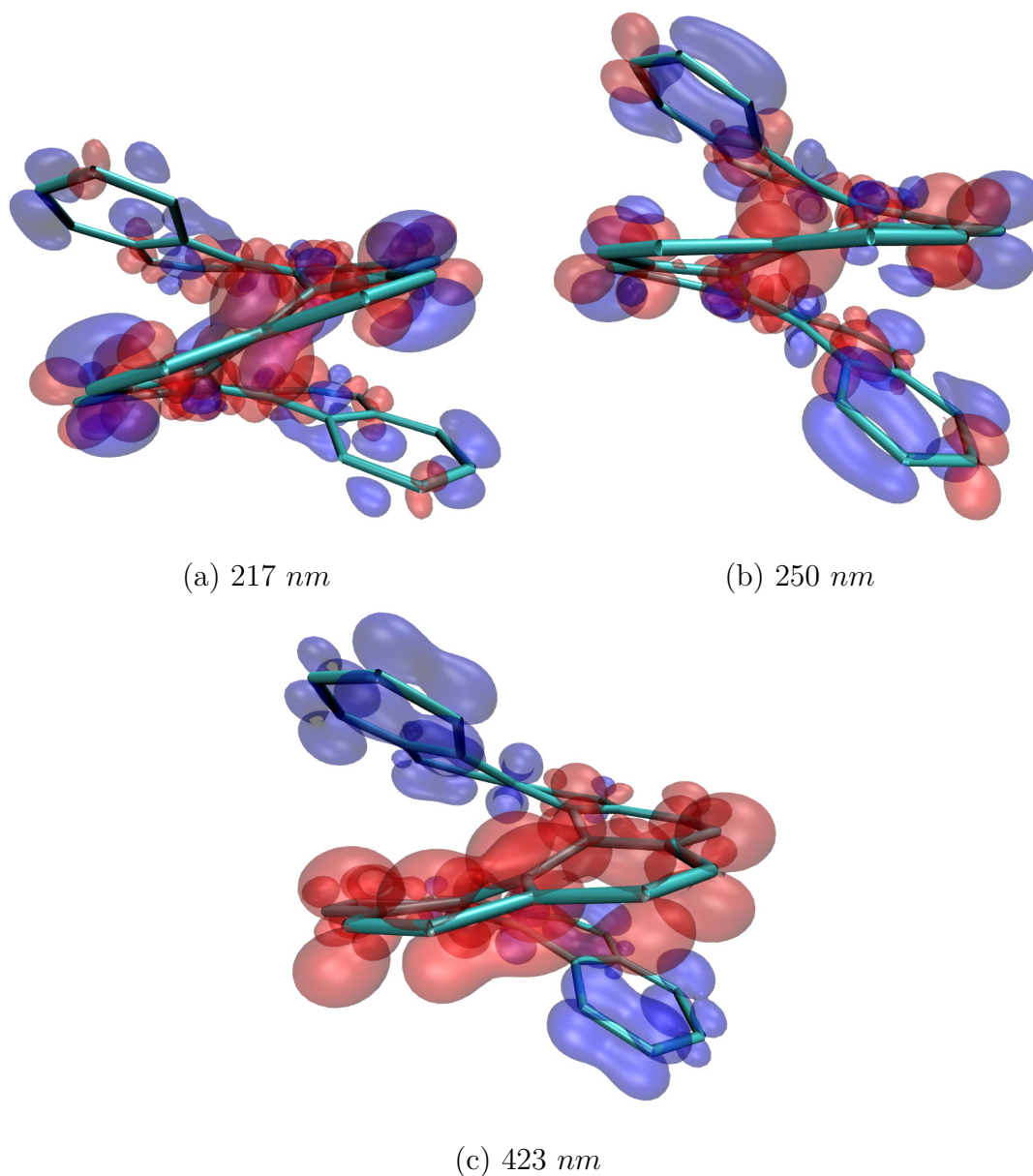


Figure 3.26. Transition densities for pseudo[10]helicene using $geom_1$ pseudopotentials. Excitations from left to right, top to bottom: (a) 217 nm, (b) 250 nm, (c) 423 nm. In each case, the charge transfer is from the blue regions to the red. These calculations are performed at the TDDFT-PBE0 level.

Figure 3.26 displays transition densities for $geom_1$ pseudo[10]helicene, for the three peaks which appear to line up best with those of the all-electron spectrum

in Figure 3.24. Figures 3.26a, b and c therefore correspond to Figures 3.25a, b, and d respectively. While the shapes of the lobes do not align as neatly with the all-electron transition densities of Figure 3.25 as transition densities for planar molecules (see Appendix B), we see that the densities all correspond to transitions of similarly π - π -like character. A difference however, between the pseudosystem and all-electron transition densities is that while it was noted above that Figure 3.25c contained some delocalisation of the electron density in the centre of the helix, this is true for all of the pseudosystem transitions in Figure 3.26. It is possible therefore that the π electrons are not tightly-enough bound by the pseudopotentials, allowing for a greater electron delocalisation than that permitted by the all-electron systems.

We suggest two further possible explanations for the degradation of the ECD spectra: (1) with $n > 6$, the ends of the helicene overlap more and more, resulting in a much more complex electronic structure, and (2) that as the helicene length increases the steric effects of the overlapping π rings causes an increasing distortion of individual benzene rings, *i.e.* the helicene ‘stretches’ [88].

These spectra show that we can be confident that our potentials can retain much of the physics of complex π systems, even allowing for some distortion of the molecular plane. Using the definition of dihedral torsion shown in Figure 3.27, [6]helicene has a maximum dihedral torsion-per-ring of 14° , while for [10]helicene this figure is 16° . These calculations show that the ECD spectra of helicenes are mainly due to the π -like electrons and that our pseudopotentials allow for the reproduction of properties which are much more difficult to reproduce than UV spectra as they are much more sensitive to the environment. We can also say that the *geom*₁ pseudopotentials prove to be more accurate than the *set4* potentials for the reproduction of helicene spectra.

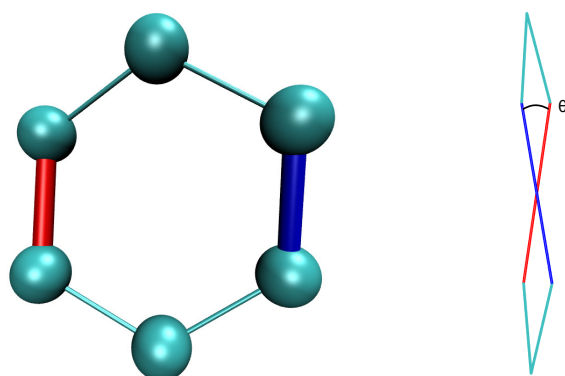


Figure 3.27. Dihedral torsion of a benzene ring, viewed from the front (left) and the side (right). In this work, the dihedral torsion of an individual benzene ring is defined as the maximum angle θ between any two opposing bonds in the ring.

3.6.2. Complex Spectra: Twistacene

The phrase ‘twisted acene’, later ‘twistacene’, was introduced in 2004 [89], then more thoroughly described in 2006 [90]. The electronic and optical properties of their parent acenes are altered by this twisting; notably, they are more soluble [91]. The twisting also introduces chirality to the molecule, and can produce bathochromic shifts [92].

Study of these molecules is complicated by the fact that the twisting must be induced by other chemical fragments acting on the acene, making them hard to study directly. However, efforts have been made to study the electronic structure of twistacenes in a systematic manner, such as that of Bedi *et al.* [92]. In this paper, they describe a system allowing them to ‘helically-lock’ an acene at a specific torsion angle, and it is to this publication that we turn for geometric data. Figure 3.28 shows the Ant- c_n molecule (where $n=3-6$ is the length of the carbon bridge at the top of the molecule), which allows the helical locking of anthracene. The increasing length n of the carbon bridge at the top of the molecule allows the anthracene to relax back toward a planar alignment.

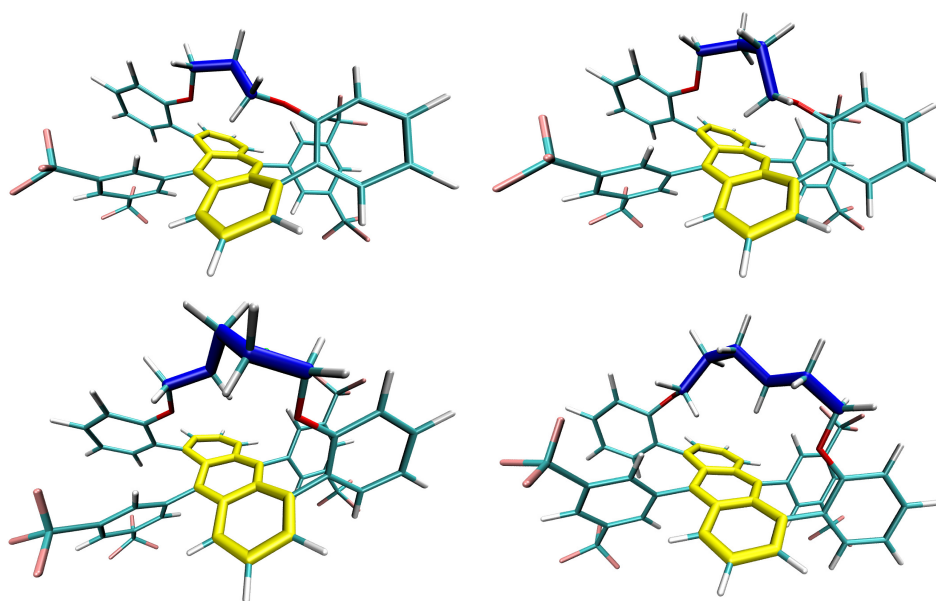


Figure 3.28. Structures of Ant- cn , where $n = 3 - 6$. From left to right and top to bottom: Ant-c3 (torsion angle 38°), Ant-c4 (torsion angle 32°), Ant-c5 (torsion angle 30°), Ant-c6 (torsion angle 23°). In each case, the anthracene itself is highlighted in yellow, while the carbon bridge that determines the torsion is highlighted in blue.

This molecule is of specific interest to us for similar reasons to the helicene described in Section 3.6.1, *i.e.* that the distorting of π rings starts to break the separation of σ and π orbitals, which will in turn make our α potentials less and less physically descriptive of the system, likely leading to a poorer result. With a systematic test of increasingly distorted π rings therefore, we can gain some idea of the limits of the pseudopotentials used.

3.6.2.1. Results and Discussion

As our interest was solely in the part of the molecule we wished to model with pseudopotentials, *i.e.* the twisted anthracene itself, we first optimised the molecular geometries (at the DFT-PBE0 level, with def-SV(P) basis sets) before removing the atoms surrounding the central anthracene. For the all-electron twistacene

calculations, hydrogen atoms were added to dangling bonds and optimised. This way, we could calculate the UV spectra without the risk of the anthracene peaks being obscured by excitations in parts of the molecule in which we were not interested.

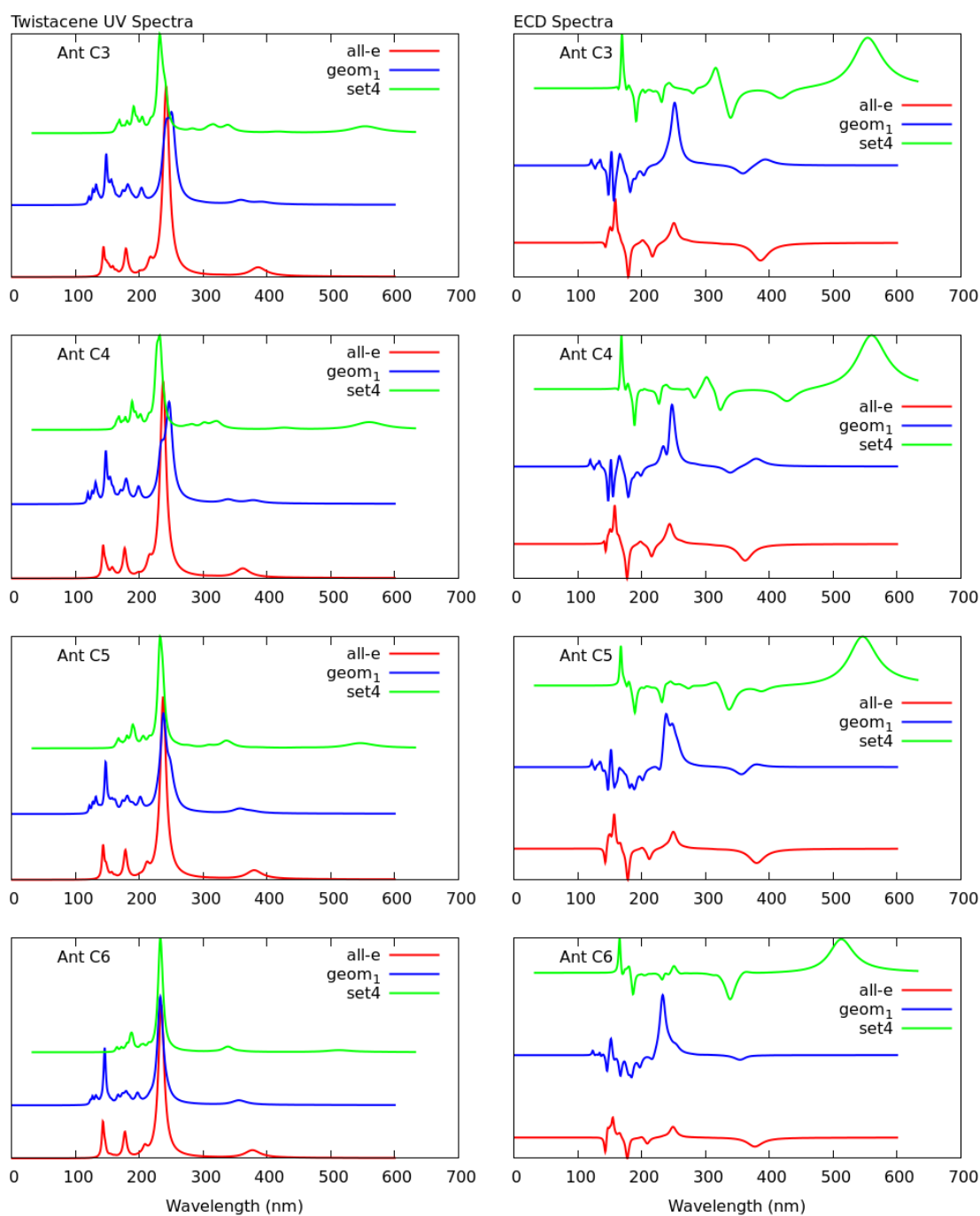


Figure 3.29. UV and ECD spectra for twisted anthracenes Ant- cn (for $n = 3 - 6$), calculated at the TDDFT-PBE0 level for the first 20 singlet excitations. These include the all-electron spectra (red), the *set4* potential spectra (green), and the *geom*₁ potential spectra (blue).

Figure 3.29 displays the spectra for Ant- cn , for $n = 3 - 6$, using all-electron

results as well as the *geom*₁ pseudopotential set. The all-electron UV spectra are very similar to one another and have peaks in approximately the 120 *nm* to 400 *nm* range, with most of the peaks clustered between roughly 120 *nm* and 280 *nm*, and the main peak at around 240 *nm* in each molecule. The pseudopotentials reproduce the broad shape of the all-electron spectra, though have an extra cluster of peaks at the high end of the spectra in the 100-150 *nm* range. This is likely due to the phenomenon noted in Section 3.5, whereby calculating *n* excitations in a pseudopotential and an all-electron system will result in more higher-energy peaks in the pseudosystem, due to excitations that are possible in the all-electron system being prevented in the pseudosystem by the missing electrons. For the pseudopotentials, the majority of the peaks in the all-electron systems appear to be present in the pseudosystems, though it is hard to see exactly which pseudopotential peaks correspond to the lowest-energy peak in the all-electron spectra at around 380 *nm*. The relative intensities of the peaks also seem broadly correct. The pseudopotential sets differ in that the *set4* spectra are red-shifted by 32 *nm* in order to align with the all-electron spectra, whereas the *geom*₁ spectra require no such shift. This is keeping with earlier results for *geom*₁ and *set4* spectra. The other difference between the two is that the *set4* potentials appear to generate more peaks at longer wavelengths, for example between 500 and 600 *nm*. These peaks will be unphysical as they do not exist in the all-electron system, as we see that there do not appear to be enough excitations calculated to reach the shorter-wavelength excitations reached by the all-electron and *geom*₁ spectra. The fact that the *geom*₁ potentials produce peaks at shorter wavelengths than the all-electron spectra with the same number of excitations is therefore a good sign, as it suggests that fewer spurious excitations are calculated overall. This is therefore a second reason for saying that the *geom*₁ potentials are superior to the *set4* potentials for the recreation of molecular spectra.

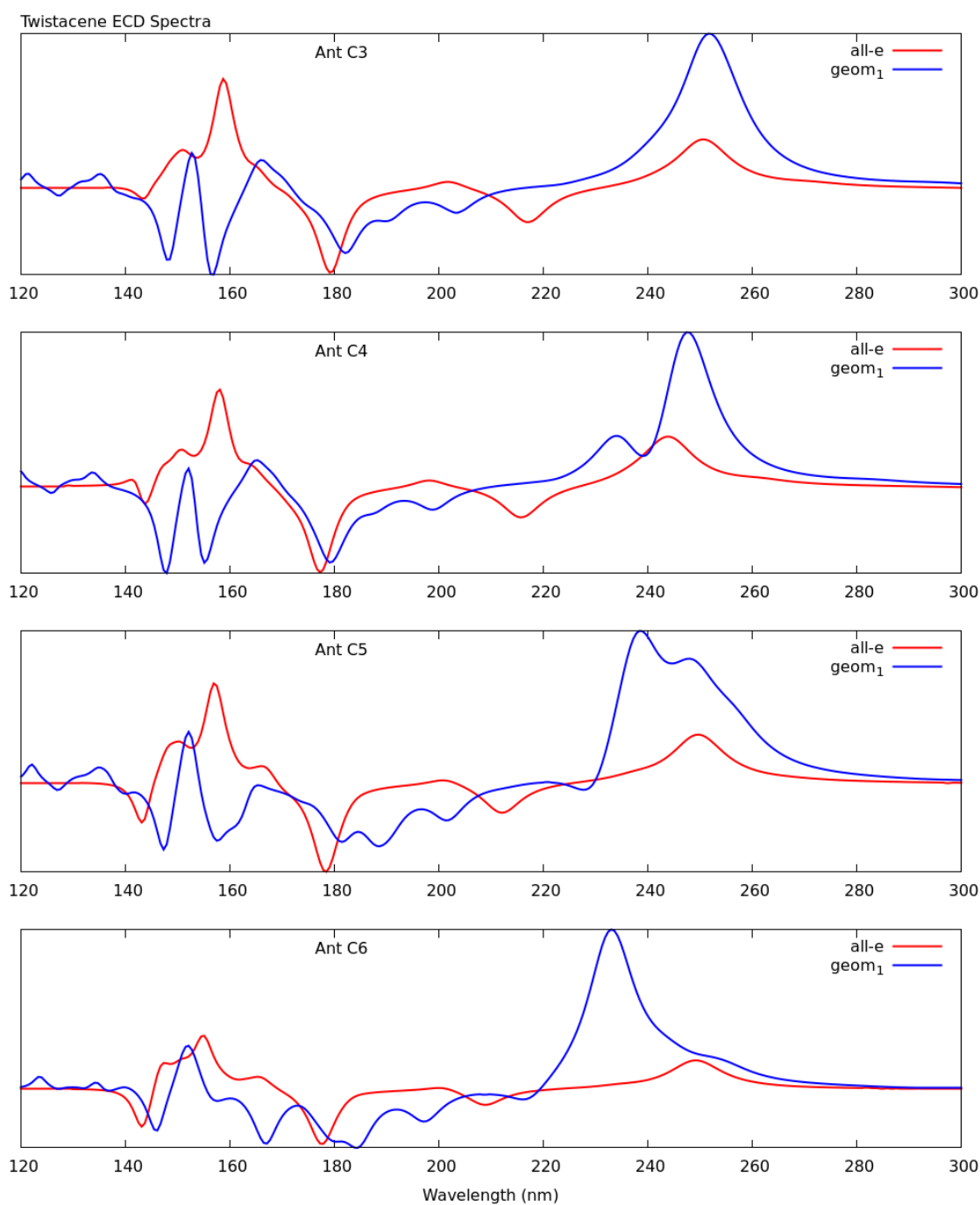


Figure 3.30. A close-up of Figure 3.29, showing only ECD spectra for all-electron (red) and $geom_1$ (blue) calculations for Ant- cn (for $n = 3 - 6$).

Turning to the ECD results we see a slightly different picture. Figure 3.30 displays a closer comparison of all-electron and $geom_1$ ECD spectra in the 120

to 260 nm range. The *geom*₁ potential spectra peaks are spread across a similar range to those of the all-electron spectra, but there are some marked differences between the spectra themselves. Some peaks, particularly the large peaks in the range of around 170 nm to 270 nm, are consistently identifiable in both all-electron and pseudosystem spectra, albeit that the intensity of the main peak at around 260 nm is several times higher for the pseudosystem than for the all-electron system. We do however see that the negative-to-positive shift indicative of chirality is found in this range, and remains clear even for the most distorted system, Ant-c3. Outside this range identification of peaks becomes harder. At the shorter wavelengths of 140 nm to 180 nm we see a large negative peak in the pseudosystem spectra that is not present in the all-electron spectra, and in the 360 nm to 400 nm we see a positive peak present in the *geom*₁ spectra that is absent in the all-electron spectra. One interesting feature of these spectra is that while these spurious peaks increase slightly in intensity as the system becomes more distorted, we do not see any additional spurious peaks appear, nor do we see much further distortion in the peaks present. Overall then, we can say that even in the wavelength ranges for which the *geom*₁ potential produce a poor spectrum, they at least remain fairly consistent as the system becomes more distorted. On the other hand, the *set4* spectra of Figure 3.29 are very different to both all-electron and *geom*₁ spectra. The only consistently-identifiable peaks appear to be the positive peak around 180 nm and the negative peaks around 190 and 330 nm, all of which are present in the all-electron spectra (even if shifted). However the peak at 250 nm which allows us to identify the molecule as chiral is absent in the *set4* spectra. One can though say that the *set4* potentials appear to capture the positive 180 nm excitation where the *geom*₁ potentials do not.

This result is consistent with the helicene results of Section 3.6.1. Table 3.22 shows the maximum and mean distortion angles per benzene unit (see Figure 3.27) across both helicene and twistacene molecules used in this work. The

Table 3.22. Maximum and mean distortions per benzene unit for helicene and twistacene molecules used in this work. The maximum distortion per benzene unit is chosen to be the largest dihedral torsion angle in the molecule (see Figure 3.27).

System	Distortion (°)	
	Mean	Maximum
[6]helicene	9.3	14.5
[7]helicene	10.7	15.6
[8]helicene	11.1	15.5
[9]helicene	11.5	15.7
[10]helicene	12.0	15.9
Ant-c3	14.1	17.9
Ant-c4	13.6	16.9
Ant-c5	10.4	13.3
Ant-c6	6.2	8.1

maximum and average distortion rates per benzene unit across both ranges of molecules are similar.

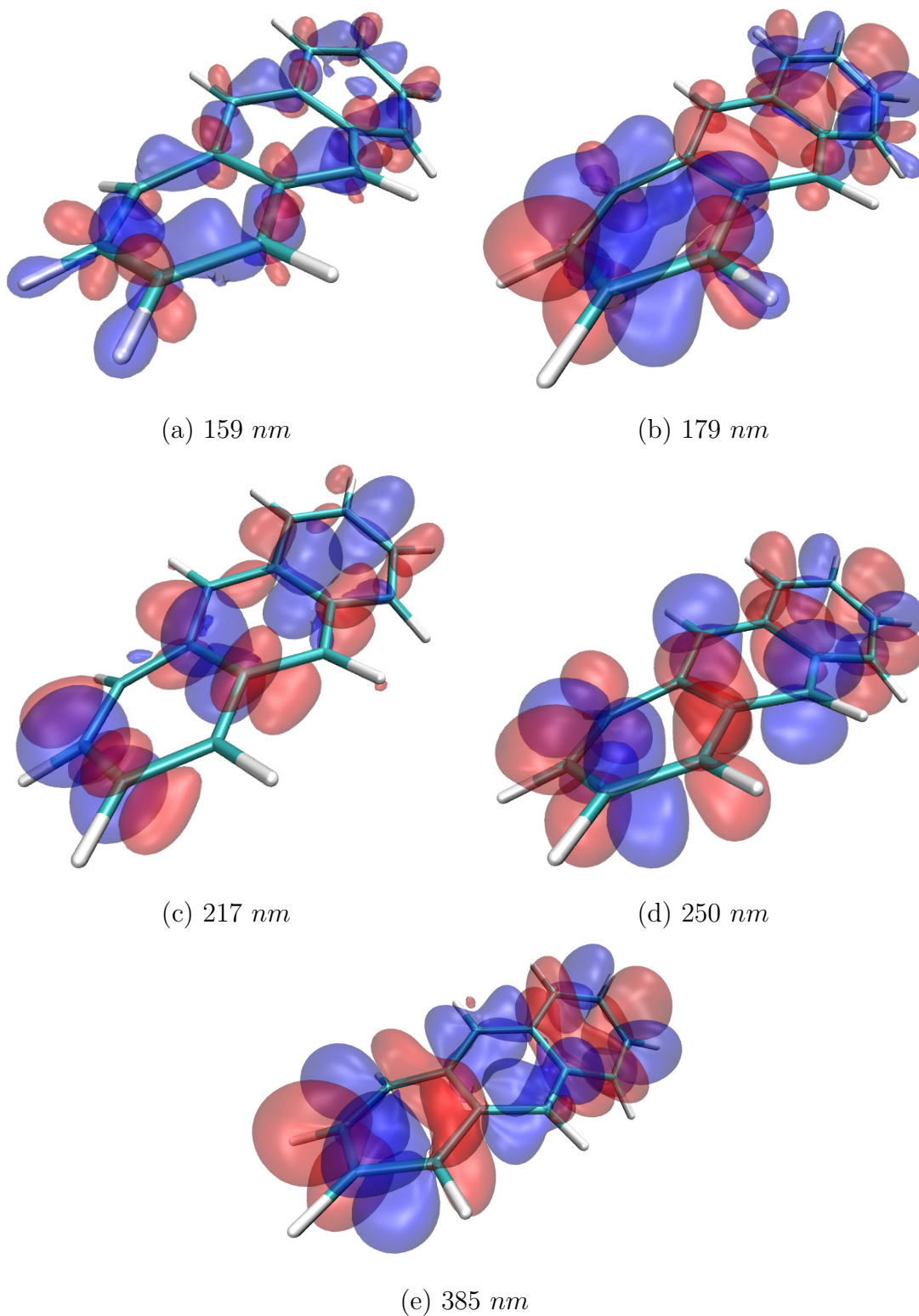


Figure 3.31. Transition densities for Ant-c3. Excitations from left to right, top to bottom: (a) 159 *nm*, (b) 179 *nm*, (c) 217 *nm*, (d) 250 *nm*, (e) 385 *nm*. In each case, the charge transfer is from the blue regions to the red. These are all-electron calculations performed at the TDDFT-PBE0 level.

Figure 3.31 displays transition densities for the five most distinct peaks in the all-electron Ant-c3 ECD spectrum. The first peak, Figure 3.31a, at 159 nm is one which is not reproduced by the pseudopotentials. The transition density shows us why this might be. It shows a strong σ character, which the pseudomolecule will of course not be able to capture. The other peaks, Figures 3.31b, c, d, and e at 179, 217, 250 and 385 nm, are of a much more π -like character, which explains why most of them, with the possible exception of the peak at 217 nm, can be identified in the Ant-c3 pseudomolecular spectra.

Figure 3.32 displays transition densities for the five most distinct peaks in the pseudo-Ant-c3 ECD spectrum. We see immediately that in Figures 3.32a and 3.32b the electron density is distorted in a mostly un- π -like manner, and so will not be physically representative of any transition present in the all-electron system. It is thought that Figures 3.32c, d, and e correspond to Figures 3.31b, d, and e respectively. They are broadly π -like, although the shapes of the electron densities do not correspond neatly to the respective all-electron densities in the way that the densities of transitions in the planar molecules do (see Appendix B), other than perhaps those of Figures 3.31e and 3.32e.

In conclusion then, we can say that the pseudosystems were successful at recreating the UV spectra of the distorted anthracene π systems, up to and including a distortion of around 14.0° per benzene unit (Ant-c3), with the *geom*₁ potentials proving the most successful. However, even at the comparatively low rate of twisting of around 6.2° per benzene unit (Ant-c6), some spurious excitations are introduced into the ECD spectra of pseudotwistacene, and that this is particularly true of the *set4* potentials. This is in contrast to the helicene examined previously, where the pseudopotential ECD spectra were of good quality and did not contain any obviously spurious peaks, despite in many cases a greater rate of planar distortion. The reason for this seems to be in part a result of the nature of the particular excitations that dominate the spectra, and how much of a π orbital character they have, as well as local planar distortion in regions of the

molecule to and from which the charge transfer takes place.

3.6.3. Complex Spectra: Dodecaphenyltetracene

In the process of searching through various acenes for suitable test molecules for the pseudopotentials (see Section [3.6.2](#)), we came across a recent synthesis of dodecaphenyltetracene [[93](#)]. The complexity of this molecule, along with the fact the team had reported a UV spectrum to which theoretical spectra could be compared, made this an attractive challenge. The molecular structure is shown in [Figure 3.33](#).

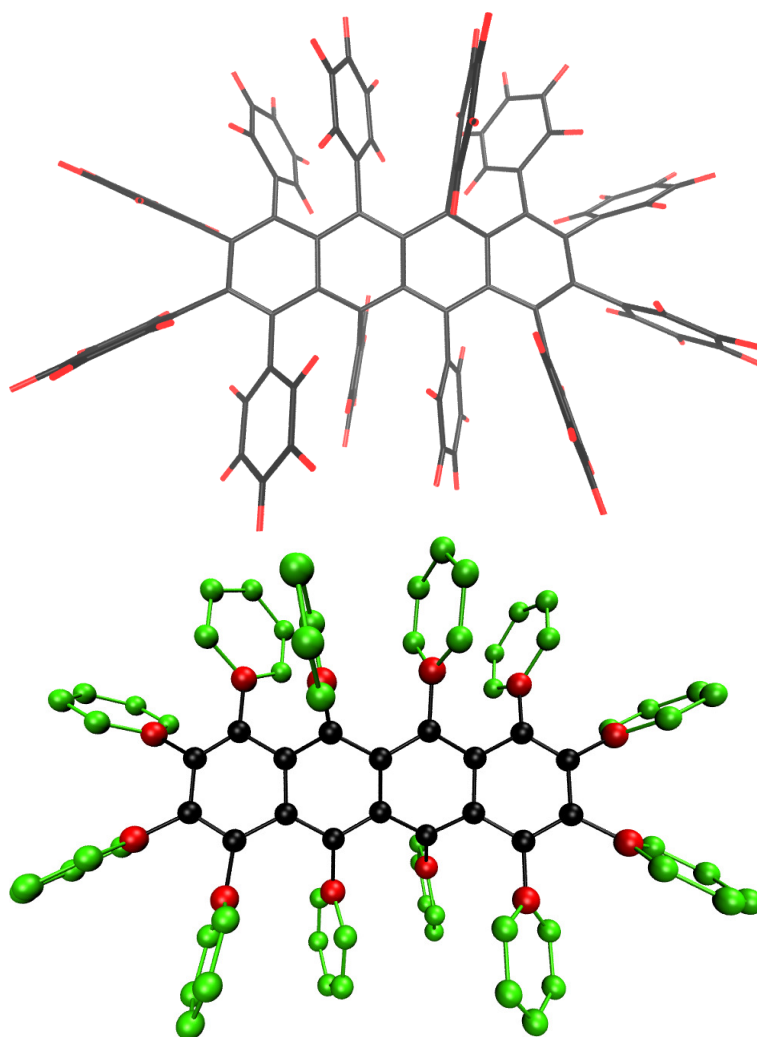


Figure 3.33. Dodecaphenyltetracene, with carbon in black and hydrogen in red (above). Pseudododecaphenyltetracene, with α pseudoatoms in green and β pseudoatoms in red (below). All-electron atoms remain in black.

The reader will see that this molecule could encapsulate two different challenges. First are the phenyl groups, *i.e.* the presence of many π systems facing and overlapping one another at different angles. This is sure to involve the overlapping of many higher-energy orbitals, making for a complicated electronic structure. Second is the fact that the central tetracene is distorted by the phenyl groups. Since we already have both the twistacene molecule in Section [3.6.2](#)

and the helicene molecules in Section 3.6.1 to investigate distorted π systems, we decided to ‘pseudopotentialise’ only the phenyl groups so as not to confuse the results. Figure 3.33 shows the pseudomolecular system. This setup uses both α potentials, along with β potentials connecting the phenyl rings to the central, all-electron tetracene.

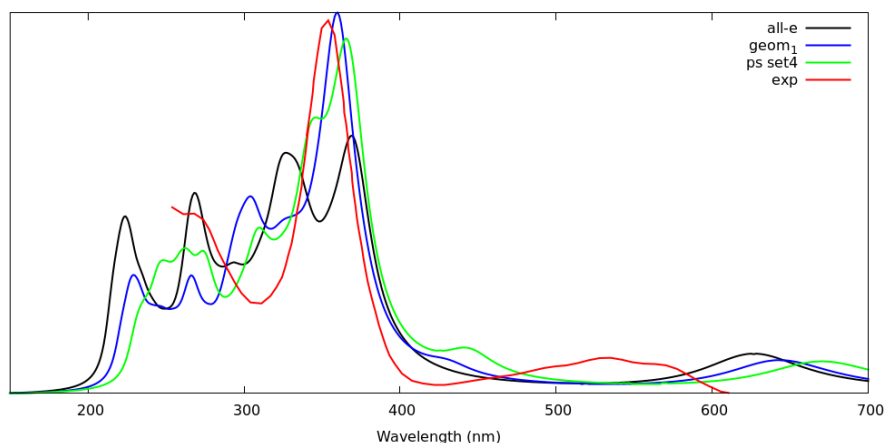


Figure 3.34. UV spectrum of dodecaphenyltetracene using all-electron DFT-PBE0 (black), as well as spectra using both *set4* (green) and *geom₁* (blue) sets of pseudopotentials, also calculated at the DFT-PBE0 level. The experimental spectrum is also shown (red).

At first glance the comparison of both pseudopotential spectra with the all-electron spectrum seems favourable. The *geom₁* potentials show five clear peaks in near-agreement with the all-electron peaks, with the possible exception of the peak at around 300 nm (*geom₁*) as compared to around 330 nm (all-electron). The relative intensities of these peaks are also similar to those of the all-electron spectrum, although they are somewhat skewed toward the lower-energy excitations by comparison. The *set4* spectrum is more difficult to interpret. Despite containing excitations across roughly the same wavelengths and at roughly the same intensities as the all-electron spectrum, the peaks are not so clearly defined, with the exceptions of the peaks at 680 nm and 360 nm. There is also a peak at 450 nm that is not present in the all-electron spectrum.

The next comparison to be made is that of these theoretical results with the experimental ones in the original synthesis. The experimental spectrum extends from around 650 *nm* to 250 *nm* and has a spread of shallow excitations across the 600-450 *nm* range, a clear and strong peak at 360 *nm*, and at least one further peak in the 250 *nm* region. The stronger peaks (below 450 *nm*) are reproduced in the theoretical spectra, both in all-electron and pseudosystem calculations. The range of peaks between 600 and 400 *nm* is not reproduced in any of the theoretical spectra. Comparison of theoretical and experimental spectra is complicated by the fact in the experimental results the molecules are in a CH₂Cl₂ solution.

Finally, we note that the HOMO energies of the all-electron system, the *geom*₁ and *set4* systems are -4.941 *eV*, -5.476 *eV* and -5.522 *eV* respectively, making for percentage errors that are not dissimilar to those of the higher polyenes modelled previously (see Table 3.5).

In conclusion, these results for dodecaphenyltetracene show that this pseudopotential technique is effective at recreating the UV spectra of overlapping and interacting π systems.

3.6.4. Complex Spectra: Nanotube-embedded Coronene

In a recent work, Nakamura *et al.* have studied the photoexcitation spectrum of a coronene polymer encapsulated in a carbon nanotube [94]. In this work, they show that two absorption bands (at around 1.7 and 3.4 *eV*) lead to a charge transfer from the coronene polymer to the nanotube. This charge transfer process is backed up by DFT calculations in the plane wave formalism using boundary conditions. These calculations do not allow for the computation of excited states. Instead, the nature of the transition is assumed from the calculation of band structure of a model system, which is a (19,0) nanotube. We tried to investigate this problem with our pseudopotentials in the cluster approximation to

see if we could find discrete electronic excitations, which could confirm a charge transfer between the coronene and the nanotube.

The challenge here for the pseudopotentials is that the encapsulating tube is not planar. As with the helicene and twistacene molecules, this means the assumption of σ - π separation does not hold, and so the pseudopotentials are a much less physical description of the system than they would be for a planar system. However, if the curvature is not too great, the pseudopotentials could still lead to physically relevant results.

The comparison between experiment and such a cluster approach is risky as even without pseudopotentials the system under study is a very crude approximation of the experimental system. Experimentally, Nakamura *et al.* make measurements over a variety of nanotubes (their diameter ranges from 1.0 to 1.5 *nm*), each encapsulating one coronene. We can study a model system of the experimental system using a coronene molecule encapsulated in a nanotube slice as can be seen in Figure 3.35.

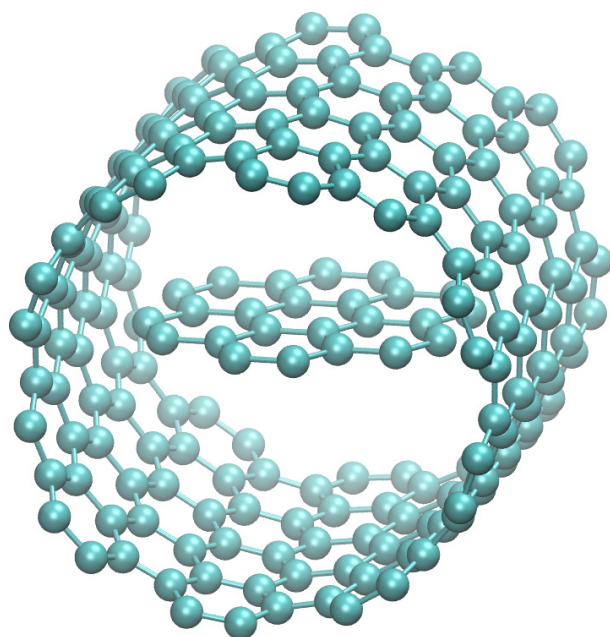


Figure 3.35. Model system of a coronene-polymer-encapsulating single wall carbon nanotube. It is made of a slice (1,14) armchair nanotube of length 1.10 *nm* and of diameter 1.25 *nm* encapsulating a coronene molecule. Hydrogen atoms were removed from the representation.

We propose that if we are able to obtain any sensible results with our method, then the π -like electrons must be mainly responsible for the photoelectronic properties of the system at low energy. We decided to refer to an orbital as π -like on a carbon atom if it exhibited one lobe above and one lobe below the plane made by the three nearest neighbours of the atom considered.

We chose a non-metallic chiral (1,14) nanotube of an diameter similar to those used experimentally to be as close as possible to the experimental system. For an all-electron version of this model system, the calculation of the excited states would not be routine at the TDDFT level as it would require the calculation of a huge number of excited states, probably some thousands, as one must include both the σ and π excitations. Because of the reduction of the system when using our pseudopotentials (only π -like electrons treated explicitly), we could do such a calculation in the π -like-only space. We performed the calculation of the

photoabsorption spectrum with TDDFT from the ground state to the first 200 excited states. The results are shown in Figure 3.36.

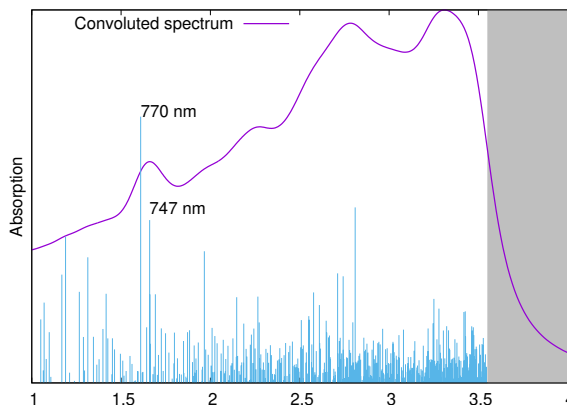


Figure 3.36. Photoabsorption spectrum (absorption in arbitrary unit in terms of energy in eV) of the model system described in Figure 3.35 obtained with TDDFT (with the PBE0 functional) asking for 200 excited states (impulses in blue). Excitations relevant for the charge transfer process between the nanotube and the coronene moieties are labelled with their wavelength values. The grey zone indicates where no excitation is computed, which explains the decay of the spectrum.

The computation of these 200 excitation energies leads to a maximum energy of roughly $3.6 eV$. A peak appears in the region of interest of the spectrum around $1.7 eV$ and two excitations (at $747 nm$ and $770 nm$) have large oscillator strengths.

This allows for a study of the peak at $1.7 eV$ as two excitations appear to have an intensity much above the average. However, there are no such clear features around $3.4 eV$. It is likely that the level of modelling is not sufficient in this case. Either, the basis set is not large enough or the number of states to be computed is larger or TDDFT cannot handle such high-energy excited states.

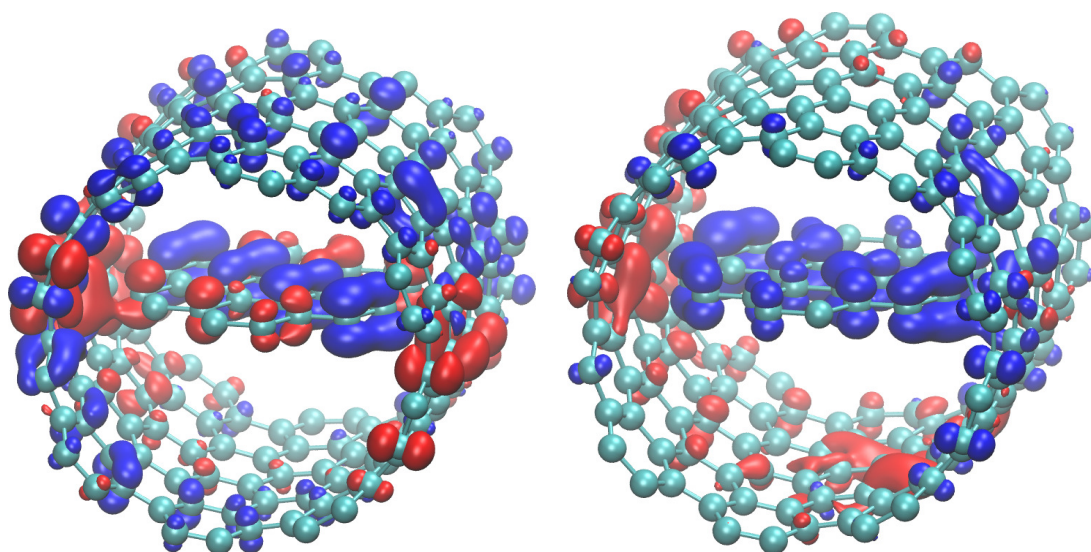


Figure 3.37. Transition densities for the excitation at 770 *nm* (left) and 747 *nm* (right) at the TDDFT-PBE0 level with the pseudopotentials developed in this work. Blue indicates a diminution of electronic density and red indicates an increase of electronic density.

The transition densities of these excitations (at 747 *nm* and 770 *nm*) are shown in Figure 3.37. The first excitation at 747 *nm* shows a charge transfer from the coronene to the nanotube (there is only blue π density on the coronene and only red π density on the tube). Such a result confirms the assignment made by Nakamura *et al.* The second excitation with the higher oscillator strength at 770 *nm* shows electronic depletion and increase on both the coronene and the nanotube. This rearrangement is the reason for the large oscillator strength but cannot be assigned to a charge transfer process.

Overall, one can say that this method allows for the use of quantum chemistry methods on systems for which such calculations, even if strictly not impossible, would be normally be computationally intensive in the extreme. With a system whose character puts it so far outside the design scope of the pseudopotentials, the authors do not believe it would be advisable to rely solely on pseudopotential calculations to draw firm conclusions about it. However, in conjunction with the experimental results, we feel this method can nonetheless yield useful in-

formation about the electronic structure of this system, and so could do so for others, provided that any properties of interest involve electrons with a strong π character.

3.6.5. Complex Spectra: Hemi-Cryptophane

The interest in cryptophanes and hemicyptophanes lies in their propensity to form Van der Waals complexes as hosts. Some of these complexes have interesting catalytic properties [95]. The fact that their cavities and catalytic sites are lipophilic has led to an increasing interest in their use as biomimetics of enzymatic systems [96, 97].

Our own interest in such molecules is that they combine a range of different carbon environments with a complex electronic structure. This means we can test a variety of different pseudopotentials on the same molecule, as well as see how they perform in the presence of metal atoms. A further consideration is that such 'cage molecules' are often characterised by their UV spectra, and so the ability to produce an accurate UV spectrum with a reduced pseudocage molecule would be useful practically.

The molecule we have adopted as a test subject is a Cu(II)hemicryptophane complex, shown in Figure 3.38.

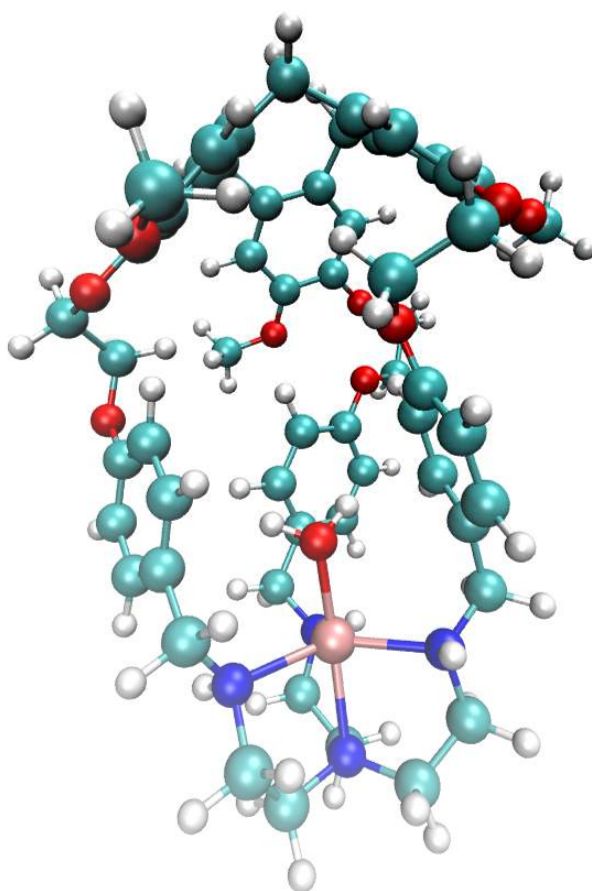


Figure 3.38. Diagram of Cu^{2+} in hemicryptophane cage, where oxygen is in red, nitrogen in blue, and copper is pink.

One can see from Figure 3.38 that there are enough different chemical environments to test each kind of potential we have developed simultaneously. By creating a series of different 'pseudopotentialisations' of this molecule, we can build up an idea of what the pseudofragments can reproduce well, and where they fail, an idea of where the most chemically-sensitive parts of the molecule lie.

3.6.6. Results and Discussion

In order to try to make the investigation as systematic as possible, we decided to group parts of the molecule together as shown in Figure 3.39, before building

up the number of pseudopotentials used.

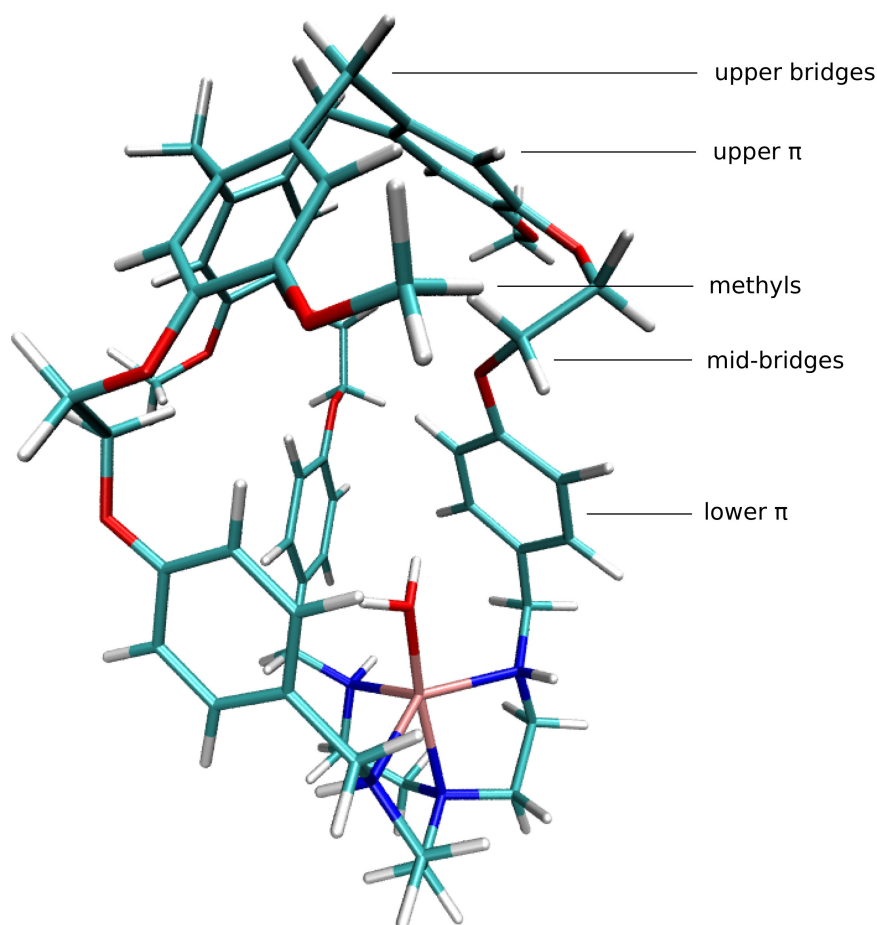


Figure 3.39. Labelling scheme for Cu(II)hemicryptophane complex.

In light of the results from Section 3.4, where we saw that the results were markedly less accurate when pseudofragments were permitted to bond with heteroatoms, it was decided to institute a 'no-heterobonding' rule for certain parts of certain pseudocomplexes, in order to see what the effects might be on the resulting spectra. An example of the 'no-heterobonding' rule for the 'upper π ' section of the cage is shown in Figure 3.40. Figure 3.40a shows the regular-all-electron upper π section of the cage. Figure 3.40b shows a version in which all carbon atoms have been replaced by pseudoatoms. Finally, Figure 3.40c shows the upper π section following the no-heterobonding rule, in which there must be at least

one all-electron carbon atom in between pseudofragments and heteroatoms.

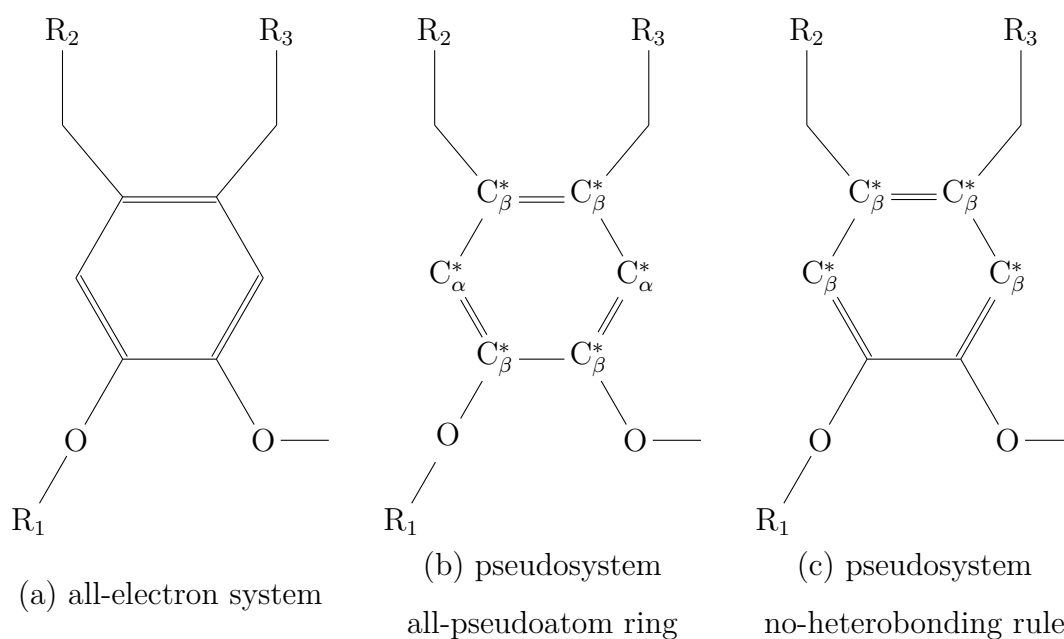


Figure 3.40. The ‘no heterobonding’ rule for upper π ring pseudopotentials.

Combining Figure 3.39 with Table 3.23 allows us to visualise each complex created with its different pseudopotentialisation scheme. If these are viewed alongside Figure 3.41 one can begin to see which parts of the molecule are responsible for producing the different parts of the spectra, and how well the pseudopotentials at each point have replicated the effects present in the all-electron system.

Table 3.23. Structures of pseudohemicryptophane complexes. Each column represents a part of the molecule (see Figure 3.39) that has been replaced with potentials. The ‘heterobonding’ column shows whether or not heterobonding has been permitted, *i.e.* whether pseudofragments have been placed next to non-carbon atoms in the complex. For those systems with the ‘N (π)’ designation, heterobonding is forbidden for carbons which are members of π rings, but has been permitted elsewhere. The ‘~’ symbol in the final column represents a molecule whose spectrum may contain neither, either or both peaks but which is distorted enough that we do not feel confident in assigning them.

ID	upper π	upper bridges	methyls	mid-bridges	lower π	heterobonding	Peaks
2	Y	Y	Y	Y	N	Y	2
3	Y	Y	N	N	N	N	1, 2
4	Y	N	Y	N	N	Y	2
5	Y	N	Y	N	N	N (π)	1, 2
6	N	N	N	N	Y	Y	~
3+5	Y	Y	Y	N	N	N (π)	1, 2
7	Y	N	Y	Y	N	N (π)	1, 2
8	Y	Y	Y	Y	N	N (π)	1, 2
9	Y	Y	Y	Y	Y	N (π)	~

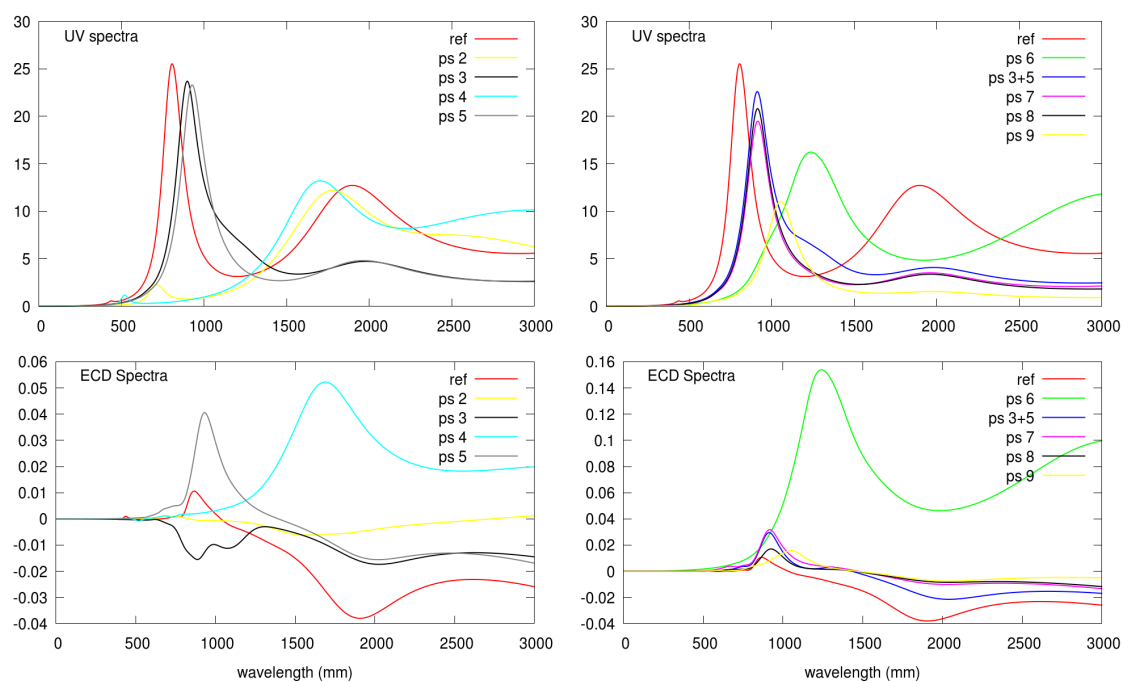


Figure 3.41. All-electron and pseudomolecular UV (above) and ECD (below) spectra for Cu(II)hemicryptophane. Calculations were carried out at the TDDFT-B3LYP level, with the first 20 singlet excitations.

The reference spectra has two major features, a sharp, intense peak at around 800 *nm* and a shallower and less intense peak centred at around 1900 *nm*. In the final column of Table 3.23, these are labelled peaks 1 and 2, respectively. One thing that is immediately striking from Figure 3.41 is the range of different results. Since none of the pseudocomplexes uses any pseudopotentials below the ‘lower π ’ section of the molecule, we know therefore that a good part of the spectral activity of this molecule must arise from the hemicryptophane itself, and not just the metal ion and its immediate neighbours. This is hinted at in the original article, where electro-chemical studies revealed that the oxidation and reduction potentials of the complexes were dependent on the precise structure of the cage. What is revealed by pseudoreults however, is the magnitude of the effect of the ‘upper π ’ rings on the electronic structure. While most of the pseudocomplexes have replaced the ‘upper π ’ rings with pseudoatoms, only those

which have respected the 'no heterobonding' rule with respect to the π atoms have been able to reproduce both peaks (pseudocomplexes 3, 5, 3+5 7, 8 and 9), whereas those which do not follow the no-heterobonding rule and have β pseudopotentials bonded to the oxygens adjacent to the upper π ring (that is, pseudocomplexes 2 and 4) are unable to reproduce peak 1. Furthermore, we see that structures which reproduce both peaks 1 and 2 have a much shallower peak 2 than the all-electron reference spectrum. These results tells us three things:

1. A heteroatom bond (*i.e.* a C-O bond) in the upper π region of the molecule is largely responsible for peak 1. We can then compare pseudocomplex 3 (whose spectrum contains peak 1 and which has no pseudoatoms replacing the methyl groups) to pseudocomplex 5 (whose spectrum contains a near-identical peak 1 but does replace the methyl groups with pseudoatoms) to see that peak 1 must arise from the upper π carbon-oxygen bond that connects to the mid-bridges (the oxygen bonded to R_1 in Figure 3.40). In the original article, this peak was attributed only to copper transitions, independent of the cage. Here we see the cage is in fact necessary for peak 1 to be produced.
2. Peak 2 has a strong upper π component.
3. The difference between these results lies in the difference between the pseudopotentials used in the upper π rings. The different structures of the upper π ring pseudopotentials is displayed in Figure 3.40 for clarification.

Further to point 3, since we see that peak 2 is present in the all-electron spectrum (with upper π rings as shown in Figure 3.40a), and also in the spectra 2 and 4 (with all-pseudopotential upper π rings as in Figure 3.40b) and is also present but at a diminished intensity in spectra 3, 5, 7 and 8 and (with upper π rings respecting the 'no-heterbonding' rule as shown in Figure 3.40c), this leads

us to conclude that the reduced intensity of peak 2 is a result of the setup in Figure 3.40c reproducing the upper π system less well than that in Figure 3.40b.

Pseudocomplexes 6 and 9 are the only ones which have placed pseudopotentials on the lower π rings of the molecule. The fact that spectrum 6 is shifted significantly from the reference spectrum, and that pseudocomplex spectrum 9 is shifted from pseudocomplex spectrum 8 (their respective counterparts with all-electron lower π rings), suggests that delocalised electron density from the lower π rings also contributes to the dominant excitations present in the spectra, though the fact that pseudocomplex 6 does not respect the heterobonding rule, and that pseudocomplex 9 has so many other pseudopotentials present in the system, makes it hard to draw conclusions. Further investigation would be needed to determine the exact nature of this effect.

The chirality of this molecule arises from the top of the hemicryptophane itself, and so one can expect to see the ECD spectra altered by the presence of the pseudopotentials. Looking at Figure 3.41, this is indeed the case. Pleasingly, with the exception of pseudocomplexes 4 and 6, all the ECD spectra retain a clear positive-to-negative shift, *i.e.* they are still obviously chiral. The most distinctive features of the all-electron system are a positive peak at around 800 nm, leading into a large, shallow negative peak at around 1900 nm. Pseudocomplex spectra 5, 3+5, 7, 8 and 9 broadly share these traits, albeit with intensities differing by up to a factor of around three. Pseudocomplex spectra 3 is somewhat distorted, and has no positive peak. The difference between pseudocomplex 3 and pseudocomplex 3+5 is in the pseudopotentialisation of the methyl groups. However, other pseudocomplexes which replace the methyl groups, e.g. 5, seem to reproduce this positive peak very well, and so it is hard to see the reason for the performance of pseudocomplex 3. Once more, the pseudocomplex spectra 6 and 9 appear shifted, and the pseudocomplex whose ECD spectrum matches best that of the all-electron calculation, in both wavelengths and intensities, is 3+5.

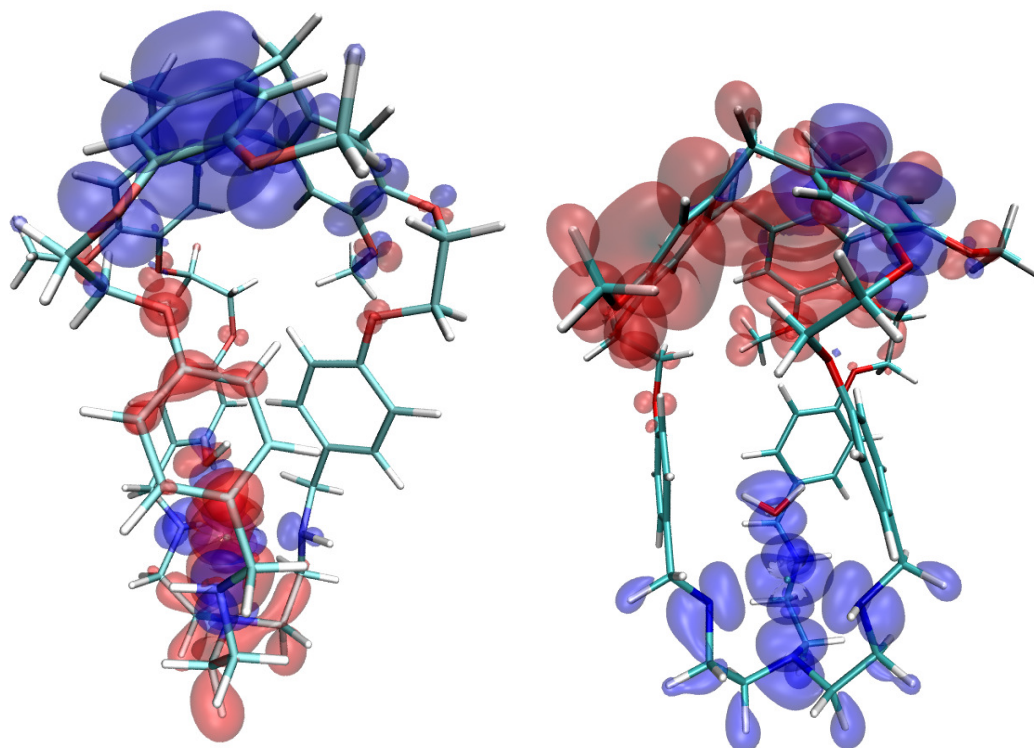


Figure 3.42. Transition densities based on all-electron TDDFT-B3LYP calculations for excitation peaks 1 (left) and 2 (right). In both cases, electron density is reduced in the blue zones and increased in the red.

Figure 3.42 displays the all-electron transition densities for peaks 1 and 2. These provide further evidence that deductions above regarding the nature of the two peaks are correct. Peak 1 broadly shows an electron transfer between upper and lower parts of the molecules, and the density on the upper part is indeed focused on the upper π rings and the oxygen atoms directly below them. Peak 2 similarly contains a transfer of electron density between the top and bottom of the molecule, with a strong upper π component.

We set out to replace as much of the hemicryptophane cage as possible with pseudopotentials, with the stipulation that the features of the UV and ECD spectra should still be identifiable. Many of the pseudocomplexes tested meet this criterion. The one which meets this criterion while removing the largest number of explicit electrons however, is pseudocomplex 8. From Table 3.23, we see that

this pseudocomplex applies the pseudopotentials from the upper bridges all the way down to the mid-bridges, while making sure to respect the no-heterobonding rule for the upper π systems. This makes for an overall reduction in the number of explicit electrons in the complex of 132, from 545 to 413.

In conclusion then, this Cu(II)hemicryptophane complex is reproducible with simple carbon pseudopotentials, as we were able to recreate the key features of the complex's spectrum. This is particularly impressive given the heavy delocalisation of electron density over the whole molecule. However, it should be noted that it was necessary to derive a new rule in order to be sure of retaining the necessary electronic complexity, which is that bonds between pseudocarbons and all-electron atoms should be restricted to carbon-carbon bonds only, and that bonding pseudocarbons to explicit heteroatoms should be avoided. Given the results seen in Section 3.4, and given that all pseudocarbons in this work are optimised on carbon-carbon bonds, this seems reasonable.

3.7. Geometry Optimisation

3.7.1. Development of the Algorithm

Using pseudopotential calculations for geometry optimisation presents some difficulties. Designing pseudopotentials such that the explicitly-treated parts of the molecule experience the correct attraction and repulsion at a particular geometry is one thing, designing them such that the same is true at any (reasonable) geometry is quite another. With a little knowledge of the all-electron system however, we can ensure that the pseudosystem will fall into the correct geometry. We do this by adding a correction to the energy gradients along the direction of the bond between pseudocarbons and their neighbours.

We begin by finding curves of dissociation for the explicit and pseudopotential parts of the molecule, as well as another for the same parts of the all-electron

molecule (see Figure 3.43). We then use a nonlinear least-squares Marquardt-Levenburg algorithm to fit a simple, exponentially-decreasing function of the form

$$g(x) = ae^{-b(x-x_0)} \quad (3.1)$$

to the difference between these two curves. We can now use this to make an energy correction to the pseudosystem (see again Figure 3.43).

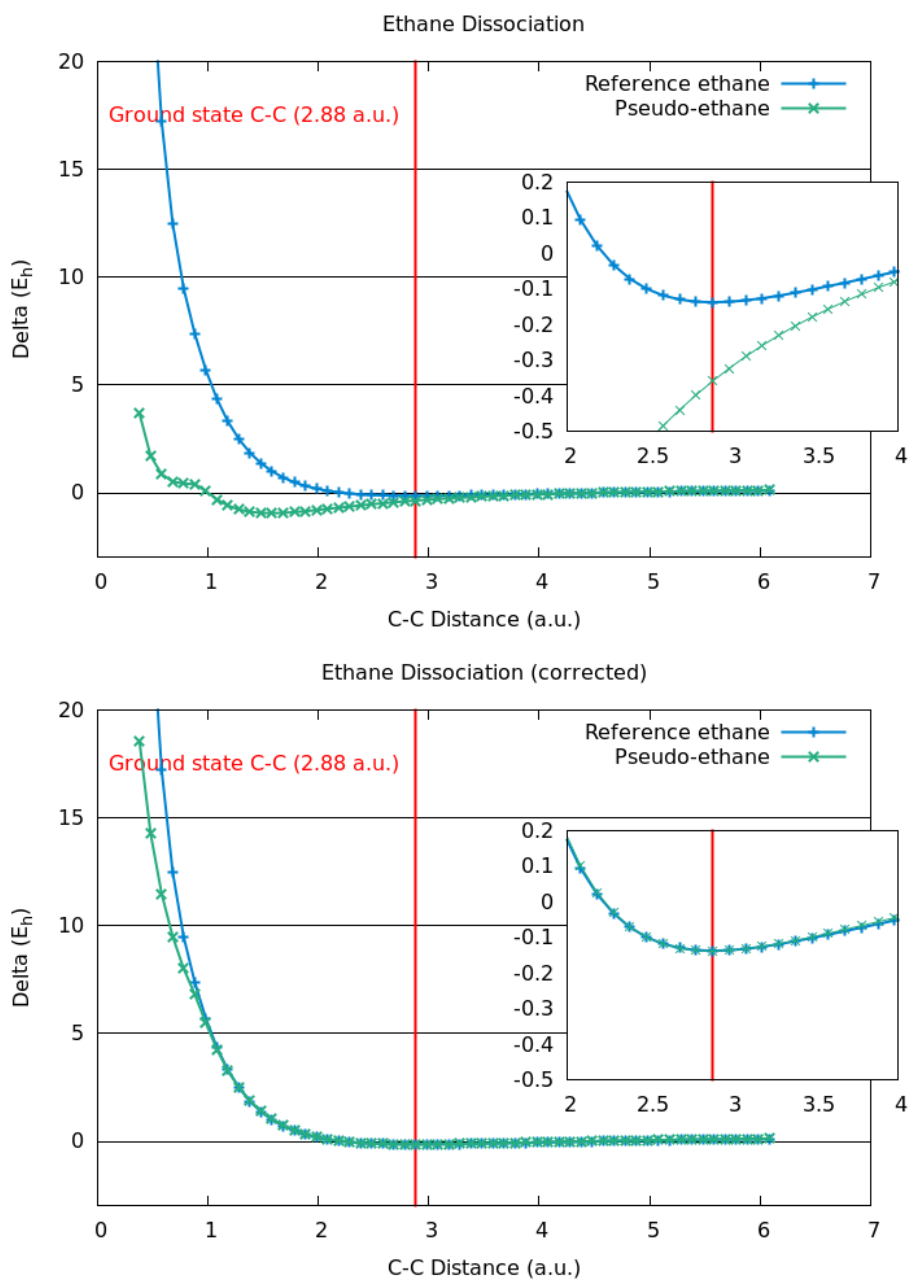


Figure 3.43. Diagrams of dissociation curves for all-electron and pseudomolecular calculations of ethane, first with just the pseudopotentials themselves (above), then with the energy correction (below). Both graphs have insets showing the behaviour of the reference and pseudosystem total energies near the true C-C distance.

We want the total energy of the system to be a minimum and the energy gradi-

ents on the explicitly-treated atoms to be zero at the true geometry (this needn't be true of the pseudoatoms, see below). We have the correction for the total energy, and by taking the derivative of the fitted Function 3.1, we have a measure of whether the explicit and pseudopotential parts of the molecule experience an overall attraction or repulsion, as well as an estimate of its magnitude. We assume the effect of the potentials on the explicit hydrogen atoms is small, and add our gradient correction directly to the potential felt along the carbon-pseudocarbon axis by one carbon, whilst subtracting it from the potential felt by the other. By doing this at every step of the optimisation, the carbon-pseudocarbon distance should naturally reach the correct value. As one can see from Figure 3.43 and especially the insets, the corrected total energy minimum is indeed at the correct distance of 2.88 *a.u.*.

In addition to the above corrections, we also require that the pseudoatoms and dummy atoms present in our pseudopotential setups remain in position relative to each other. This must be done manually.

3.7.2. Results and Discussion

Two sets of molecules were chosen to test the β and γ potentials, respectively. For the β potentials, these consisted of the previous set of β test molecules shown in Table 3.11. For γ potentials, they consisted of the previous γ molecules shown in Table 3.14, as well as the extra two molecules shown in Table 3.24. These molecules were chosen, as before, because they represented a range of different chemical environments.

The results for the β test set are shown in Table 3.25. These results include the C*-X bond lengths and C*-X-R bond angles. For most of these molecules it was necessary to fix a C*-X-R bond angle, in addition to the geometry-fixing scheme described above, in order for the geometry to be minimised successfully. This is discussed further below in Section 3.7.2.1. For the most part, the differences

Table 3.24. Additional molecules used to test the optimisation of geometries with γ pseudopotentials.

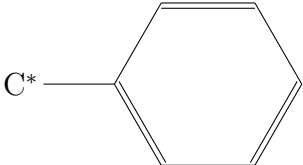
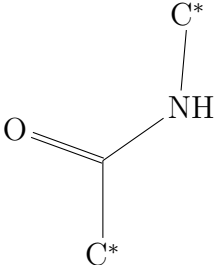
Name	Formula	Pseudomolecular Structure
Toluene	C_7H_8	
Peptide	$CH_3OCNHCH_3$	

Table 3.25. Results of geometry optimisation for molecules containing β potentials, at HF and PBE0 levels, using the ethene fit.

Molecule	Fixed Angle	Measured Angle	Difference from all- e^- (%)	
	C*-X-R ($^\circ$)	C*-X-R ($^\circ$)	C*-X Length	Bond Angle
HF				
C ₂ H ₄	C*-C-H	C*-C-H	4.4	6.1
CH ₂ O	-	-	10.2	-
CH ₂ COH	C*-C-O	C*-C-H	1.0	5.6
CH ₂ NH	C*-N-H	C*-N-H	4.9	2.1
CH ₂ CHNH ₂	C*-C-N	C*-C-H	0.8	4.6
DFT-PBE0				
C ₂ H ₄	C*-C-H	C*-C-H	22.0	7.7
CH ₂ O	-	-	9.5	-
CH ₂ COH	C*-C-O	C*-C-H	1.5	7.5
CH ₂ NH	C*-N-H	C*-N-H	4.8	1.1
CH ₂ CHNH ₂	C*-C-N	C*-C-H	5.5	7.1

between bond lengths and angles in the pseudosystems as compared to their all-electron systems was within a few percent, though were as high as 10.2% for HF optimisations and 22.0% for DFT optimisations. This seems to reflect a general trend for the geometry optimisation of pseudosystems of DFT results being poorer than HF results, in contrast to earlier applications of the pseudopotentials. The reason for this is that the fit was carried out on HF calculations (on the ethane system), and will thus bias the calculations toward the bond lengths obtained with HF. These geometry optimisations are therefore not independent of the calculation method used.

Other than the differences between HF and DFT results, we see a similar pattern to that of the electronic structure results, where molecules in which the pseudoatom is bonded to a non-carbon element, CH₂O and CH₂NH, give bond lengths and angles that are poorer compared to those of the C*-C systems, though the errors are still mostly small. This phenomenon is explained in the same way, that the energy correction fit is performed on a C*-C bond. On

the whole, the optimisations are successful, with the caveats discussed in Section 3.7.2.1.

Table 3.26. Results of geometry optimisation for molecules containing γ potentials, at HF and PBE0 levels, using the ethane fit. Molecules which suffered a geometric failure or collapse during optimisation are marked by the † symbol.

Molecule	Fixed Angle C*-X-R (°)	Measured Angle C*-X-R (°)	Difference from all-e ⁻ (%)	
			C*-X Length	Bond Angle
HF				
CH ₃ CH ₃ ^{eclipsed}	-	C*-C-H	0.8	0.0
CH ₃ OH	-	C*-O-H	4.4	1.5
CH ₃ NH ₂	-	C*-N-H	5.6	13.8
CH ₃ COOH	C*-C-O	C*-C-O	1.7	2.3
C ₉ H ₈ O ₄ †	-	-	-	-
CH ₄	-	-	68.8	-
C ₂ H ₄ O	C*-C-O	C*-C-O	3.0	2.8
Toluene	-	C*-C-C	0.4	1.2
Peptide (C*-C)	C*-C-N, C*-N-C	C*-C-O	1.6	3.2
Peptide (C*-N)	C*-C-N, C*-N-C	C*-N-H	2.8	2.5
ClAuP(CH ₃) ₃	C*-P-Au	C*-P-Au	12.7	0.0
DFT-PBE0				
CH ₃ CH ₃ ^{eclipsed}	-	C*-C-H	0.3	3.9
CH ₃ OH	-	C*-O-H	3.9	22.8
CH ₃ NH ₂ †	-	-	-	-
CH ₃ COOH	C*-C-O	C*-C-O	2.4	0.0
C ₉ H ₈ O ₄ †	-	-	-	-
CH ₄	-	-	73.0	-
C ₂ H ₄ O	C*-C-O	C*-C-H	3.1	2.9
Toluene	C*-C-C	C*-C-C	38.2	1.8
Peptide †	-	-	-	-
ClAuP(CH ₃) ₃ †	-	-	-	-

Table 3.26 displays results for γ potential optimisations, including C*-X bond lengths and C*-X-R bond angles. For the successfully-optimised systems, we can make similar observations as for the β potentials. The use of DFT in calculations increases the differences between all-electron and pseudosystems, likely

for the same reason, that the energy correction is fitted on a HF system. The errors are nevertheless mostly small (under 5%) with the exception of C*-X heteroatoms, where they are increased, sometimes very significantly, as in the case of methane.

What is different for the γ potential results is the number of failures, mostly due to geometry collapses. This is so despite the use of further fixed C*-X-R bond angles, and is discussed in further depth below.

3.7.2.1. Geometric Collapse

Fixing the pseudocarbon and dummy atoms relative to each other is sometimes enough to prevent a collapse of the molecular geometry. However there are situations where this is insufficient. The problem originates in the fact that for all the systems tested above the overall effect of the pseudopotentials, taken at a distance, is too attractive. This means that all other parts of the molecule are attractive to the pseudofragments, and should another part of the molecule approach close enough to be caught in the potentials' sphere of influence, they will become 'stuck' to it, warping the molecule into an unphysical geometry.

The systems we have tested are generally small, and the moieties that are attracted to the pseudopotentials are generally separated from them by only two or three bonds. For these systems, the problem of the potentials' 'stickiness' can be avoided by fixing the bonding angle C*-X-R, in addition to fixing the pseudopotential atoms and dummy atoms relative to one another. The systems where this was necessary are marked in Tables [3.25](#) and [3.26](#).

Finally however, there are systems where even this measure cannot eliminate the risk of collapse. The problem is illustrated by the case of aspirin, displayed in Figure [3.44](#). In this system, even after fixing the bond angle C*-C-O, there are enough degrees of freedom that all-electron and pseudoatoms are able to approach one another closely enough to cause a collapse.

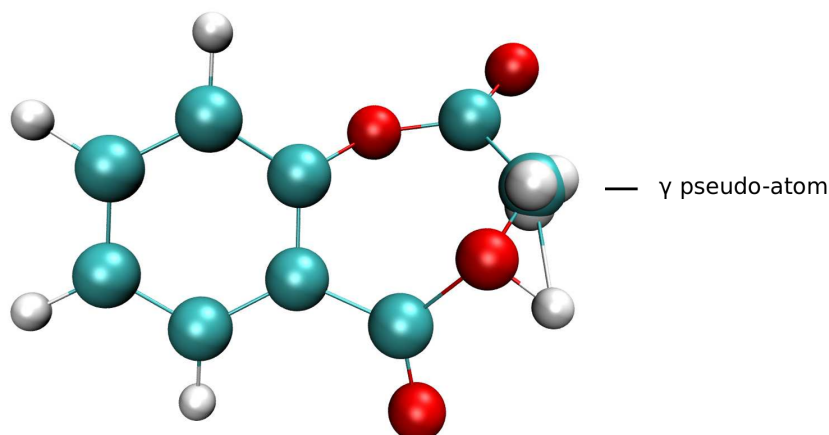


Figure 3.44. An aspirin molecule with a γ potential. The γ potential has become ‘stuck’ to the O-H group.

We did not develop this part of the pseudopotential method further. We suggest that the problem of geometric collapse might be partially mitigated by the more careful selection of potential parameters. A relatively diffuse repulsive potential might help avoid such collapses, and could be taken into account when extracting potentials so as to ensure that it does not affect the electronic structure. Overall however, without considerable further development, this will remain an intrinsic limitation of the method, rendering it too unreliable for general use in geometry optimisation.

3.8. Timings

The model that we develop in this work allows significant computational gains. To illustrate this, we provide in Table 3.27 a small study performed on the $C_{50}H_{52}$ alkene chain. Calculations were performed with two different basis sets: def2-SV(P) and QZVPPD. When using the pseudopotential, the basis set was truncated by removing all basis functions not necessary to reproduce the π system. In other words, only p functions or those of higher angular momentum were kept.

As can be seen from Table 3.27, the gain increases with the size of the basis set as expected. The gain ranges from 2.4 (def2-SV(P)) to 8 (QZVPPD).

Table 3.27. Time (in seconds) per SCF iteration and relative gain (time ratio all-electron/pseudopotential) for a standard and a large basis set.

Basis Set	def2-SV(P)	QZVPPD
All-electron SCF iteration (s)	67	68928
Pseudopotential SCF iteration (s)	28	8566
Gain (no unit)	2.4	8.0

The authors also note that, as demonstrated in Section 3.5, it is also possible to reduce the basis sets of certain potential sets, namely the α potentials, as they contain no explicit electrons in orbitals with a significant s character. In the case of the def-SV(P), this would amount to removing all s basis functions, halving the size of the basis set. This could be exploited for further gains in efficiency.

It is typical when performing *ab-initio* calculations to make use of 'starting guess' orbitals supplied by, for example, Hückel theory. This is harder to do when using pseudopotentials however, and at the time of writing is not implemented in Turbomole. This means the number of SCF iterations needed in pseudopotential calculations is increased. In the Turbomole package, this is normally mitigated by the fact that pseudopotentials supply their own starting guess orbitals. This is something that has not been incorporated into this work, but could be (see Section 4.5.1).

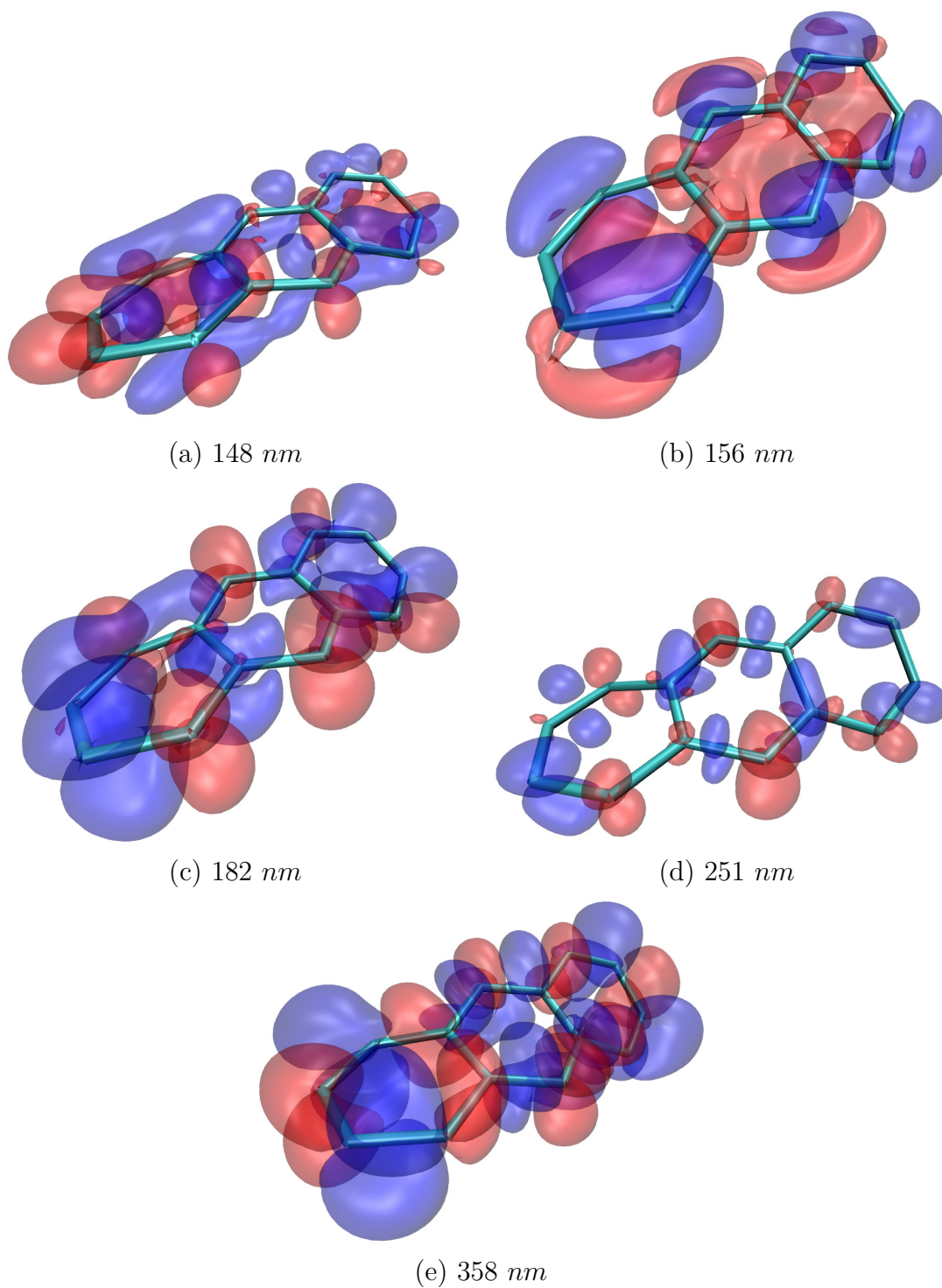


Figure 3.32. Transition densities for pseudoant-c3. Excitations from left to right, top to bottom: (a) 153 *nm*, (b) 177 *nm*, (c) 215 *nm*, (d) 225 *nm*. In each case, the charge transfer is from the blue regions to the red. These calculations are performed at the TDDFT-PBE0 level.

4. A Study of Cobalt-Mediated Cycloaddition: Lactone and Lactame

Sommaire

4.1	Background	142
4.2	Experiment	143
4.3	DFT Study	145
4.4	Lactame <i>vs</i> Lactone	147
4.5	Conclusions	148
4.5.1	Thoughts on Further Development	153

4.1. Background

Cycloaddition methods, in which more than two moieties are fused in one step, are a powerful way of creating cyclical systems. Particularly useful mediators for such reactions are Cobalt complexes, which can allow for a great flexibility in the systems created [98, 99, 100]. CpCoL_2 (where $\text{L}=\text{CO}$, PR_{R3} , alkenes) is one such complex and has been used in a variety of cycloadditions [101, 102].

In Gandon *et al.*'s DFT study of the linear 2:1 co-oligomerisation of alkynes

with alkenes [101], we see the reaction proceed as shown in Figure 4.1, whereby the two alkyne molecules are first fused in the presence of the CpCo, before a Minimum-Energy Crossing Point (MECP) allows the system to ‘hop’ from the singlet to the triplet potential energy surface. The system then passes through a second MECP in order to allow the insertion of an ethylene molecule, leaving us with molecule **I**. This metallacycle is a common intermediate from which to go either to the cyclohexadiene product **II** via a reductive elimination, or to the triene **III** via a β H-elimination and reductive elimination.

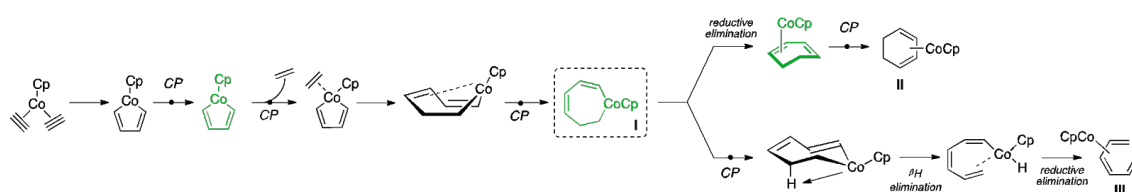


Figure 4.1. Theoretical cobalt-mediated [2+2+2] cycloaddition graph, with triplet states marked in green.

Prior to this work, a known Co(I)-mediated [2+2+2] cycloaddition reaction was taken and attempted with a new substrate, only to discover that the reaction proceeded in a radically different direction to that expected. An attempt is being made to make full sense of this reaction with the aid of theoretical calculations.

4.2. Experiment

The study began with the complex **2a** as seen in Figure 4.2, synthesised from the known triyne **1a** [103, 104, 105]. Complex **2a** is exposed to a stoichiometric amount of CpCo(CO)₂ under irradiation, in toluene, at 110°C.

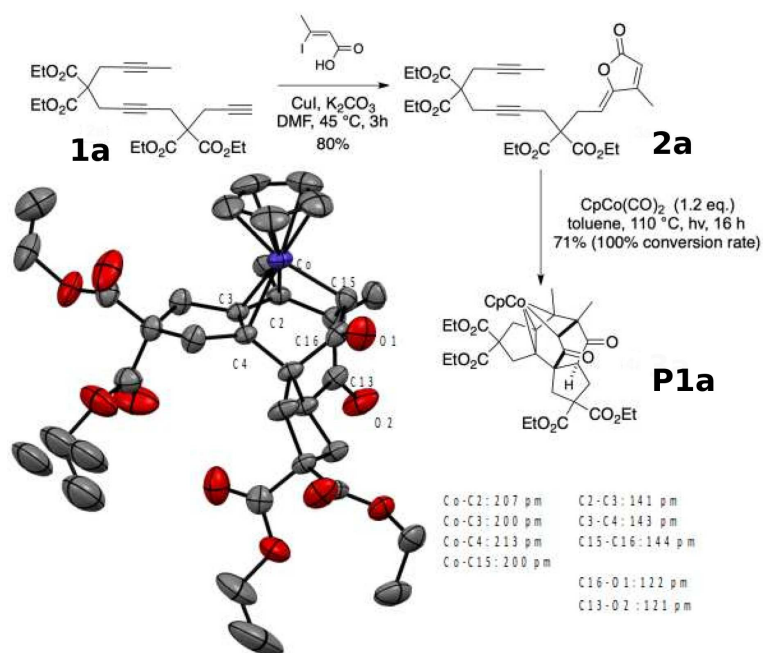


Figure 4.2. The synthesis and X-ray structure of Cobalt(III) complex **P1a**.

From the work of Gandon *et al.*, one expects the reaction to proceed as in Figure 4.1 to form a complex with either a central triene or central six-membered ring. Simplified versions of the expected complexes, **P2** and **P3**, are shown in Figure 4.3. Instead of this however, a tetracyclic cobalt complex is formed (**P1a** in Figure 4.2, **P1** in Figure 4.3) is isolated at a yield of 71%, as the sole diastereomer. The structure of this molecule has been confirmed by X-ray diffraction.

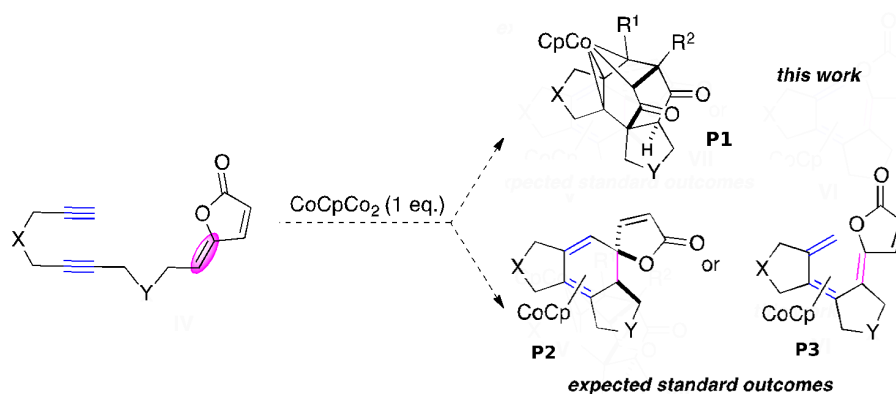


Figure 4.3. Expected standard outcome of Cobalt(I)-mediated process vs experimental result.

4.3. DFT Study

In order to understand this process better a theoretical study was carried out, attempting to find suitable reaction paths that might help explain the experimental results. Since the work by Gandon *et al.* shows that state-crossing is a feature of reactions involving such cobalt complexes, both singlet and triplet potential energy surfaces must be considered. A pair of calculated paths is shown in Figure 4.4.

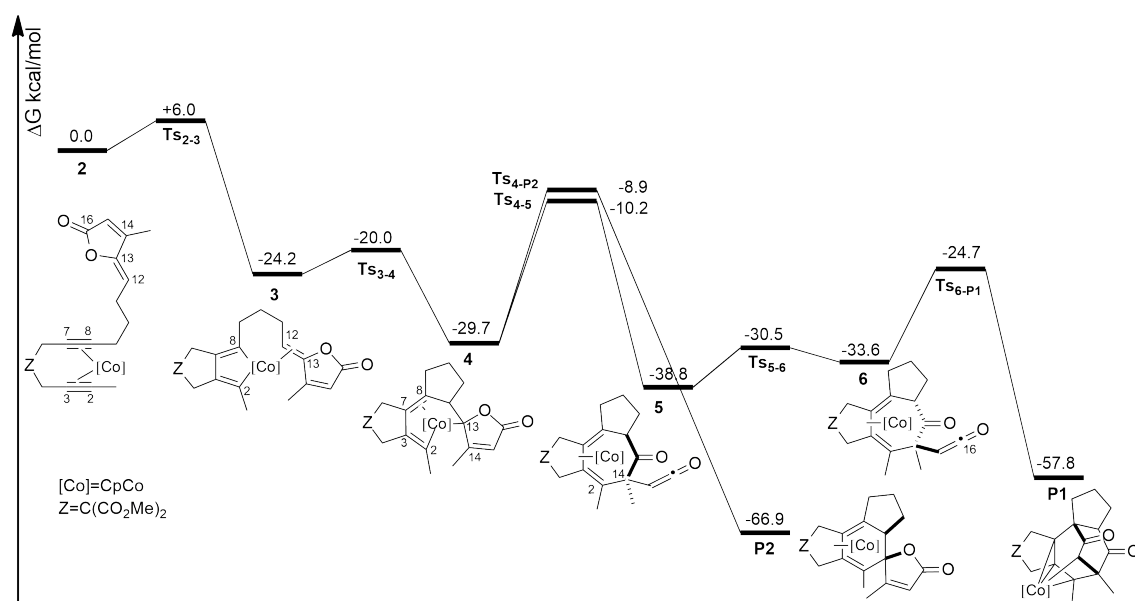


Figure 4.4. Computed Gibbs Energy ($T=110\text{C}$) reaction profile, where $[\text{Co}] = \text{CpCo}$ and $Z=\text{C}(\text{CO}_2\text{Me})_2$. Energies are given in reference to **2** and are in kcal/mol. Solvation effects are calculated using the COnductor-like Screening Model (COSMO).

It can be seen that the divergence in the reaction pathway occurs at structure **4**. One would expect C(2) to form a bond with C(13), in which case the reaction can immediately end with the expected spiro lactone **P2**. Instead, C(2) forms a bond with C(14), the lactone is broken, and the result is structure **5**, from which there is a path to **P1**. The difficulty then lies in explaining why the reaction should proceed to **5** rather than **P2**.

At the time of writing, work on establishing the precise mechanism of this reaction remains ongoing. However, this study has suggested that there are three important components of the complex that allow the reaction to proceed as it does, and to diverge from the expected outcome. These include firstly the role of the cobalt ion itself, which makes the unsaturated bonds reactive. Secondly, there is the role of the organic link that connects C(8) to C(12), imposing a constraint on the coordination of the alkene moiety. This constraint means that the bond C(8)-C(12) is easily accessible, and the formation of the bond C(2)-C(14)

is avoided. Finally, there is the role of the lactone moiety. My role in this study was to aid in the investigation of the role of the lactone by comparing it to a hypothetical alternative complex in which the lactone was replaced by a lactame.

4.4. Lactame vs Lactone

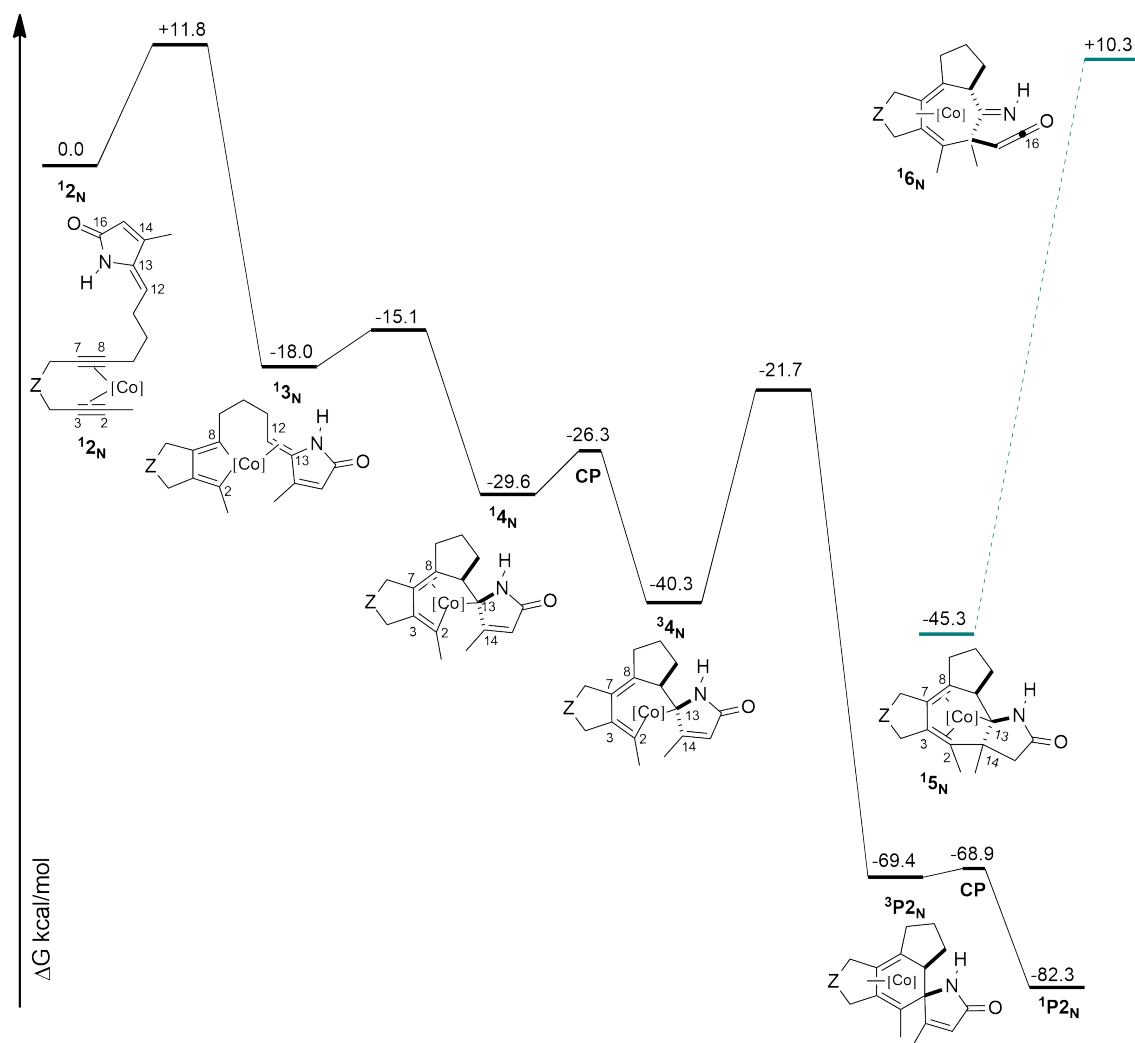


Figure 4.5. Computed Gibbs Energy (T=110C) reaction profile as in Figure 4.4, but with the lactone group replaced by a lactame. Energies are given in reference to 12_N and are in kcal/mol. Calculations carried out with PBE0-D3/def2-TZVP, with solvation effects from COSMO treatment as single point calculations.

Figure 4.5 displays the reaction pathway starting from the ${}^1\mathbf{2}_N$ geometry. For the initial two steps (the oxidative coupling and the insertion) are similar to those computed for the lactone of Figure 4.4. After this however, the paths diverge by with the creation of the bond between C(2) and C(14). The N-C(16) bond does not break, as the O-C(16) bond does in the lactone version. This implies that the corresponding ${}^1\mathbf{6}_N$ intermediate shown in Figure 4.5 is very high in energy and not reachable. Instead, we have computed the standard [2+2+2] cycloaddition mechanism leading to ${}^1\mathbf{P2}_N$ by following the proposal by Gandon *et al.*. The system continues from ${}^1\mathbf{E}_N$ to ${}^3\mathbf{E}_N$, where the C(2)-C(13) bond is generated, and from there to the final, spiro lactame product.

This allows us to conclude that the breaking of the O-C(16) bond in the lactone version of the molecule is in part a consequence of the nature of the C-O bond itself.

4.5. Conclusions

The reaction of what should be a standard [2+2+2] cycloaddition reaction of an diene/ene diverges significantly from what is expected when using γ -alkylidene-butenolide and a Cobalt mediator, instead yielding a tetracyclic Cobalt complex. A theoretical investigation was performed in the hopes of understanding this reaction. In this work, we focus on the role of the lactone moiety of the alkylidene-butenolide, and compare it with equivalent calculations in which a lactame moiety is instead used. It is found that with the use of a lactame group the reaction proceeds as expected, in contrast to the lactone. This has been confirmed experimentally. From this we conclude that the lactone is important in altering the path of the overall reaction, and that this difference between the lactone and lactame moieties may be exploitable, helping to broaden the range of possibilities for such complexes.

Conclusion

In this dissertation, a pseudopotential method for replacing small hydrocarbon fragments in molecular quantum chemistry calculations was derived and tested. The framework for these pseudopotentials borrowed elements from various current and historical pseudopotential methods, including Phillips-Kleinman and the Model Core Potential method. It begins from the same concept as all other pseudopotential methods, that the electrons can be divided into those which are active and which take part in relevant chemical interactions, and those which are dormant and do not. However, rather than choosing this divide to be between core and valence shells, it chooses as active electrons only one or two specific electrons that are involved in the chemical behaviour in which we are interested. This method removes almost all of the electrons from the fragments, in addition to some of the nuclei, and some further protons. The basis functions from the nuclei that remain are (for the most part) kept. This method makes use of non-atom-centred potentials. It incorporates in the optimisation procedure elements of both shape and size-consistency. Finally, it should be easily compatible with most standard computational chemistry packages.

First we created an sp^2 hybridised carbon fragment (nicknamed the α potentials) which contained only one nucleus and one electron (the p electron in the z axis, at a normal to the xy molecular plane). This was then optimised to fit the HOMO, 1st ionisation and singlet-triplet gap energies of ethylene, which it was successfully able to do, with differences from equivalent all-electron calculations of 7.7%, 2.9% and 0.0% respectively. This pseudo- sp^2 fragment was then taken

and tested in a series of all-trans-polyenes up to a length of 100 carbon atoms, as well as a series of Polycyclic Aromatic Hydrocarbons of up to 19 concentric six-membered rings. Both of these tests were successful, with the π orbitals successfully reproduced in all test molecules. This test was repeated with several popular DFT functionals, and found average differences from all-electron calculations in the HOMO, 1st ionisation and 1st excitation energies of up to 7.8% for the best functional (PBE0), and 20.1% for the worst (TPSS). There was no increase in these differences with all-trans-polyene length and differences were similar for PAH systems, meaning we had captured the physical effects of π delocalisation on all-trans-polyenes, as well as those of aromaticity on the PAHs. They were also shown to work with CASPT2 calculations.

Next we broadened and refined our optimisation algorithm by developing our own computer code which was able to take a variety of reference criteria from all-electron systems, supplied by the user, and to attempt to fit to these criteria new pseudopotentials by repeated calculations on pseudosystems. This could be done either by supplying a starting guess for the pseudopotential parameters, or by a process of making semi-random guesses.

We then went on to create a series of new hydrocarbon fragments and to fit them to reference data with our program. These fragments featured a two-electron sp^2 carbon (another ethylene fragment, but one which this time could form both a π bond and a single σ bond), a one-electron sp^3 carbon (a methyl group) and a two-electron sp^3 carbon (a 'link' atom between all-electron atoms, fitted to the central carbon of propane). These were termed the β , γ and δ potentials respectively. Each of these pseudopotentials was then tested in a series of small molecules in order to observe their responses to different chemical environments. It was found across all the pseudopotential types (β , γ , δ) that they performed well when the bonds they formed were homoatomic (that is, to an all-electron carbon atom), with average difference between all-electron and pseudosystem of up to around 20%, but generally on the order of a few per

cent. It was also found that errors were considerably increased (up to several hundred per cent in the extreme case of formaldehyde, where the molecular orbitals are incorrectly-ordered in the pseudosystem) when made to form bonds with heteroatoms such as Nitrogen, Oxygen or Fluorine.

We made a study of using the original α pseudopotentials to generate UV spectra, which we found was highly effective. The pseudopotentials, optimised only on ethylene, were able to recreate singlet and triplet excitation spectra for a range of PAH molecules. The spectra were clearly defined, with all relevant all-electron spectra peaks identifiable in the pseudosystem spectra, with mostly-correct intensities and with only minimal and consistent shifts in wavelength. We found that they are able to lead to the same AIM view of the bonding pattern as all-electron systems and to reproduce physically-reasonable absorption spectra together with physically-grounded electronic transition densities.

Using the improved optimisation code, we were successfully able to generate α pseudopotentials with improved spectra by using all-electron virtual orbitals as reference criteria, though this tended to worsen their fits to the previous criteria.

Next, we took all pseudopotential types created (α , β , γ and δ) and tested their behaviour across a series of much larger and more complex molecules in order to find the limits of the method. These included the UV and ECD spectra of systems with increasingly distorted π systems ([*n*]helicene, twistacene, Coronene embedded in a carbon nanotube), many overlapping π rings (dodecaphenyltetracene), and with transition metal atoms (a Copper (II) complex in a hemi-cryptophane cage). Several further conclusions were drawn from these exercises:

1. From helicene, twistacene and the nanotube systems we learned that when using pseudopotentials for aromatic rings with only π electrons treated explicitly (α potentials), distortions in the π plane do not worsen the spectra as long as the distortion is only of a few degrees. Above this value however, the ECD spectra quickly become less reliable.

2. From helicene, twistacene and dodecaphenyltetracene we learned that UV spectra of a consistently high quality can be generated even after replacing most or all of the explicit atoms in the molecule with pseudopotentials, and that this remains true even up to the most distorted system tested (see Section 3.6.2).
3. From the cage system we learned that allowing pseudopotential fragments to bond directly to heteroatoms risked creating poor UV and ECD spectra. In order to ensure that both UV and ECD spectral features were properly reproduced, it was necessary to keep at least one explicit all-electron carbon atom between a pseudopotential fragment and a heteroatom.

Finally, we created and tested a technique for performing geometry optimisation with the pseudopotentials. This involved creating dissociation curves between pseudoatoms and all-electron atoms for β and γ potentials, and then fitting a correction to both atoms to ensure they remain at the correct bond distance during the geometry optimisation. This technique was tested across a range of different small molecules and the results were not dissimilar to those of the earlier investigations, *i.e.* where the pseudocarbons were bonded to all-electron carbons the approximation worked well, with small bond length and angle errors. Where the all-electron atom was a heteroatom, the approximation worked more poorly, with larger errors on bond length and angle. However, the main problem was that the potential fragments were attractive to other atoms in the vicinity, causing not-infrequent geometric collapse of the molecule. Some possible improvements to the method are suggested, but in its current form the geometry optimisation technique is too unreliable for general use.

In summary, we have created a pseudopotential technique which manages to retain many basic properties of the molecular fragments they replace, including orbital energies, ionisation energy, and excitation energies sufficient for the production of full UV and ECD spectra. This is in spite of the fact that most of the

electrons and some of the nuclei are removed from these fragments. We have tested these potentials and derived some general principles for their use. We have also created a general algorithm for creating these potentials, and investigated the kinds of molecular properties that make suitable reference criteria, based intuitions of the kinds of molecular properties that will help us capture the underlying physics one wishes to simulate. Finally, we have established that, provided they are used carefully, these potentials can replicate or merely provide insight into some surprisingly complex systems.

Unrelated to the pseudopotential work above, a small study of a Cobalt-mediated cycloaddition reaction was carried out, in order to try to understand an unexpected experimental result. In this study, the particular importance of a lactone moiety was established in provoking the experimental results observed, by contrasting it with the behaviour of a theoretical non-lactone equivalent molecule.

4.5.1. Thoughts on Further Development

While the results achieved in this work seem self-contained enough to be satisfactory, there are further avenues for development that, time permitting, could have been explored, and may be in future.

Probably the most straightforward way to achieve better pseudopotential fits would be simply to use more potential functions per angular momentum l . As noted in Section 2, using a maximum of only one potential functions per angular momentum per pseudopotential centre was a choice made purely for the sake of simplicity, and the MOO code could easily be expanded to include more.

Another option we did not have time to explore was the optimisation of our own basis sets, as is standard for most pseudopotentials. This may have helped improve their accuracy, but the real benefit would likely be that the basis sets could be smaller than all-electron basis sets (as for the α potentials where the s basis functions were removed, see Section 3.5.1), meaning a smaller overlap

matrix and thus further gains in computational efficiency.

While attempting to fit pseudosystem and all-electron UV spectra via a least-squares method, we found that the procedure was hampered by the fact that the least-squares difference between a peak that was wrong by say, 5000 *nm* and a peak that was wrong by say, 50 *nm* was the same. One answer to this could be to artificially broaden the peaks so as to allow them to overlap at a much greater distance, making the fitting algorithm more efficient.

Something noted in Section 3.8 is that our pseudopotentials could not make use of the Hückel guess method for obtaining starting orbital guesses at the beginning of the SCF procedure, and that this increased the number of SCF iterations needed to converge pseudopotential calculations. Many supplied pseudopotentials in the Turbomole package come with their own starting orbital guesses. This is something that could also be done for the method in this work.

The largest arbitrary constraint placed on the development was that the method should be easily compatible with standard quantum chemistry software without any source code modification. If this stricture were to be lifted then drawing on previous work by Krause *et al.* on the MCP method, it should be possible to identify ‘intruder orbitals’ in a similar way, and to remove them [30]. While this proved not to be a major problem in this work, it could prove useful when studying the spectra of other molecules.

One possibility that was discussed but not explored during this work was that this method could very easily be adapted to produce link atoms for Quantum Mechanical/Molecular Mechanical (QM/MM) calculations. They would provide an interface between parts of molecules treated with QM and MM formalisms, in that they are treated in the QM framework but would have no ‘dangling electrons’. They would function not dissimilarly to the Quantum Capping Potential method [50].

Some final improvements not directly related to the Chemistry itself could be in the MOO program. Over the course of this work the time spent creating pseudo-

molecule geometries was vastly decreased by the development of code to place potentials automatically. One further step which could be added to this process would be the full automation of pseudopotential testing. This would incorporate test sets of molecules such as those in this work, against which to benchmark any newly-developed pseudopotential sets, allowing one to know quickly not only how well a new pseudopotential performed against the optimisation criteria, but how transferable it was to a great variety of other systems.

Bibliography

- (1) A. Szabo and N. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Dover Publications, 1996 (cit. on pp. [14](#), [17](#)).
- (2) M. Dolg, *NIC Series*, 2000, **1**, 479–508 (cit. on p. [14](#)).
- (3) M. Dolg and X. Cao, *WIREs Comput Mol Sci*, 2011, **1**, 200–210 (cit. on pp. [14](#), [27](#)).
- (4) M. Klobukowski, S. Huzinaga and Y. Sakai, in *Computational Chemistry: Reviews of Current Trends*, 1999, ch. 2, pp. 49–74 (cit. on pp. [14](#), [27](#)).
- (5) E. Schrödinger, *Physical review*, 1926, **28**, 1049 (cit. on p. [14](#)).
- (6) M. Born and R. Oppenheimer, *Annalen der Physik*, 1927, **389**, 457–484 (cit. on p. [15](#)).
- (7) W. Pauli, *Zeitschrift für Physik A Hadrons and Nuclei*, 1925, **31**, 765–783 (cit. on p. [16](#)).
- (8) J. C. Slater, *Physical Review*, 1929, **34**, 1293 (cit. on p. [17](#)).
- (9) D. R. Hartree and W Hartree, *Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences*, 1935, **150**, 9–33 (cit. on p. [18](#)).
- (10) P. Hohenberg and W. Kohn, *Physical review*, 1964, **136**, B864 (cit. on p. [21](#)).
- (11) W. Kohn and L. J. Sham, *Physical review*, 1965, **140**, A1133 (cit. on p. [22](#)).

- (12) R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules (International Series of Monographs on Chemistry)*, Oxford University Press, USA, 1994 (cit. on p. [23](#)).
- (13) S. Huzinaga, *Can. J. Chem.*, 1995, **73**, 619–628 (cit. on p. [24](#)).
- (14) H. Hellmann, *J. Chem. Phys.*, 1935, **3** (cit. on p. [24](#)).
- (15) P. Gombás, *Z. Phys.*, 1935, **94**, 473–488 (cit. on p. [24](#)).
- (16) J. C. Phillips and L. Kleinman, *Phys. Rev.*, 1959, **116**, 287–294 (cit. on p. [25](#)).
- (17) J. D. Weeks, A. Hazi and S. A. Rice, 1969, **XVI**, 283–342 (cit. on p. [25](#)).
- (18) L. R. Kahn, P. Baybutt and D. G. Truhlar, *The Journal of Chemical Physics*, 1976, **65**, 3826–3853 (cit. on pp. [26](#), [34](#)).
- (19) G. Igel-Mann, H. Stoll and H. Preuss, *Mol. Phys.*, 1988, **65**, 1321–1328 (cit. on pp. [26](#), [34](#), [46](#), [63](#)).
- (20) P. J. Hay and W. R. Wadt, *The Journal of chemical physics*, 1985, **82**, 270–283 (cit. on p. [27](#)).
- (21) P. J. Hay and W. R. Wadt, *The Journal of Chemical Physics*, 1985, **82**, 299–310 (cit. on p. [27](#)).
- (22) W. R. Wadt and P. J. Hay, *The Journal of Chemical Physics*, 1985, **82**, 284–298 (cit. on p. [27](#)).
- (23) J. R. Trail and R. J. Needs, *The Journal of Chemical Physics*, 2017, **146**, 204107 (cit. on p. [27](#)).
- (24) V. Bonifacic and S. Huzinaga, *J. Chem. Phys.*, 1987, **86**, 2132 (cit. on p. [27](#)).
- (25) Y. Sakai, *The Journal of Chemical Physics*, 1981, **75**, 1303–1308 (cit. on p. [27](#)).

- (26) Y. Sakai and S. Huzinaga, *The Journal of Chemical Physics*, 1982, **76**, 2537–2551 (cit. on p. 27).
- (27) S. Huzinaga, L. Seijo, Z. Barandiarain et al., *The Journal of Chemical Physics*, 1987, **86**, 2132–2145 (cit. on p. 27).
- (28) S. Katsuki and S. Huzinaga, *Chemical Physics Letters*, 1988, **152**, 203–206 (cit. on p. 27).
- (29) C. W. Bauschlicher, *The Journal of Chemical Physics*, 1980, **72**, 880–885 (cit. on p. 31).
- (30) D. Krause and M. Klobukowski, *Canadian Journal of Chemistry*, 1996, **74**, 1248–1252 (cit. on pp. 31, 154).
- (31) Y. Sakai, E. Miyoshi, M. Klobukowski et al., *Journal of Computational Chemistry*, 1987, **8**, 226–255 (cit. on p. 31).
- (32) Y. Sakai, E. Miyoshi, M. Klobukowski et al., *Journal of Computational Chemistry*, 1987, **8**, 256–264 (cit. on p. 31).
- (33) U. Wahlgren, *Chemical Physics*, 1978, **29**, 231–240 (cit. on p. 31).
- (34) U. Wahlgren, *Chemical Physics*, 1978, **32**, 215–221 (cit. on p. 31).
- (35) L. G. Pettersson and C. W. Bauschlicher Jr, *Chemical Physics*, 1989, **131**, 267–279 (cit. on p. 31).
- (36) L. G. Pettersson, U. Wahlgren and O. Gropen, *The Journal of chemical physics*, 1987, **86**, 2176–2184 (cit. on p. 31).
- (37) S. Katsuki and H. Taketa, *International Journal of Quantum Chemistry*, 1980, **18**, 25–29 (cit. on p. 31).
- (38) J. Andzelm, E. Radzio and D. R. Salahub, *The Journal of chemical physics*, 1985, **83**, 4573–4580 (cit. on p. 31).
- (39) D. R. Salahub, *Ab Initio Methods in Quantum Chemistry Part II of Advanced in Chemical Physics*, 1987, **9**, 447–505 (cit. on p. 31).

- (40) D. Salahub, R Fournier, P Mlynarski et al., *by JK Labanowski, JW Andzelm, SpringerVerlag, New York*, 1991, 77 (cit. on p. 31).
- (41) R. D. Cowan and D. C. Griffin, *J. Opt. Soc. Am.*, 1976, **66**, 1010–1014 (cit. on p. 31).
- (42) Y. Sakai, E. Miyoshi, M. Klobukowski et al., *The Journal of Chemical Physics*, 1997, **106**, 8084–8092 (cit. on p. 31).
- (43) E. Miyoshi, Y. Sakai, K. Tanaka et al., *Journal of Molecular Structure: THEOCHEM*, 1998, **451**, 73–79 (cit. on p. 31).
- (44) U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, 1986, **7**, 718–730 (cit. on p. 32).
- (45) X. Assfeld and J.-L. Rivail, *Chem. Phys. Lett.*, 1996, **263**, 100–106 (cit. on p. 32).
- (46) S. Goedecker, M. Teter and J. Hutter, *Phys. Rev. B*, 1996, **54**, 1703–1710 (cit. on p. 32).
- (47) J. Gao, P. Amara, C. Alhambra et al., *J. Phys. Chem. A*, 1998, **102**, 4714–4721 (cit. on p. 32).
- (48) C. Hartwigsen, S. Goedecker and J. Hutter, *Phys. Rev. B*, 1998, **58**, 3641–3662 (cit. on p. 32).
- (49) Y. Zhang, T.-S. Lee and W. Yang, *J. Chem. Phys.*, 1998, **110**, 46–54 (cit. on pp. 32, 45).
- (50) G. A. DiLabio, M. M. Hurley and P. A. Christiansen, *J. Chem. Phys.*, 2002, **116**, 9578–9584 (cit. on pp. 32, 154).
- (51) O. A. von Lilienfeld, I. Tavernelli, U. Rothlisberger et al., *J. Chem. Phys.*, 2004, **122**, 014113 (cit. on p. 32).
- (52) O. A. von Lilienfeld, I. Tavernelli, U. Rothlisberger et al., *Phys. Rev. Lett.*, 2004, **93**, 153004 (cit. on p. 32).

- (53) Y. Zhang, *J. Chem. Phys.*, 2004, **122**, 024114 (cit. on p. 32).
- (54) O. A. von Lilienfeld, I. Tavernelli, U. Rothlisberger et al., *Phys. Rev. B*, 2005, **71**, 195119 (cit. on p. 32).
- (55) G. A. DiLabio, R. A. Wolkow and E. R. Johnson, *J. Chem. Phys.*, 2005, **122**, 044708 (cit. on p. 32).
- (56) C. R. Jacob and L. Visscher, *J. Chem. Phys.*, 2006, **125**, 194104 (cit. on p. 32).
- (57) J. M. Parks, H. Hu, A. J. Cohen et al., *J. Chem. Phys.*, 2008, **129**, 154106 (cit. on p. 32).
- (58) J. L. Lewin and C. J. Cramer, *J. Phys. Chem. A*, 2008, **112**, 12754–12760 (cit. on p. 32).
- (59) A. C. Ihrig, C. Schiffmann and D. Sebastiani, *J. Chem. Phys.*, 2011, **135**, 214107 (cit. on p. 32).
- (60) O. A. von Lilienfeld, *Mol. Phys.*, 2013, **111**, 2147–2153 (cit. on p. 32).
- (61) M. Hitzenberger and T. S. Hofer, *J. Comput. Chem.*, 2015, **36**, 1929–1939 (cit. on p. 32).
- (62) M. A. Collins and R. P. A. Bettens, *Chem. Rev.*, 2015, **115**, 5607–5642 (cit. on p. 32).
- (63) S. Pezeshki and H. Lin, *Mol. Sim.*, 2015, **41**, 168–189 (cit. on p. 32).
- (64) L. W. Chung, W. M. C. Sameera, R. Ramozzi et al., *Chem. Rev.*, 2015, **115**, 5678–5796 (cit. on p. 32).
- (65) M. Hitzenberger, M. Ratanasak, V. Parasuk et al., *Theor. Chem. Acc.*, 2016, **135**, 1–7 (cit. on p. 32).
- (66) G. A. DiLabio, *Chem. Phys. Lett.*, 2008, **455**, 348–353 (cit. on p. 32).
- (67) S. Topiol, A. A. Frost, M. A. Ratner et al., *J. Chem. Phys.*, 1976, **65**, 4467 (cit. on p. 32).

- (68) N. Gresh and A. Pullman, *Theoret. Chim. Acta.*, 1978, **49**, 283–294 (cit. on p. [32](#)).
- (69) M. Dolg and X. Cao, *Chem. Rev.*, 2012, **112**, 403–480 (cit. on p. [32](#)).
- (70) W. Zou, Z. Cai, J. Wang et al., *J. Comput. Chem.*, 2018, **39**, 1697–1706 (cit. on p. [32](#)).
- (71) V. Strobel, *Pold87/academic-keyword-occurrence: First release*, 2018 (cit. on p. [32](#)).
- (72) *TURBOMOLE V7.1 2016*, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from <http://www.turbomole.com>., 2016 (cit. on pp. [34](#), [42](#), [61](#), [175](#)).
- (73) K. Jug, P. C. Hiberty and S. Shaik, *Chemical Reviews*, 2001, **101**, PMID: 11710230, 1477–1500 (cit. on p. [40](#)).
- (74) J. Drujon and Y. Carissan, *J. Comput. Chem.*, 2013, **34**, 49–59 (cit. on pp. [48](#), [53](#), [54](#)).
- (75) J. Paldus and E. Chin, *Int. J. Quantum Chem.*, 1983, **24**, 373–394 (cit. on p. [53](#)).
- (76) K. N. Houk, P. S. Lee and M. Nendel, *J. Org. Chem.*, 2001, **66**, 5517–5521 (cit. on p. [53](#)).
- (77) M. E. Colvin, C. L. Janssen, E. T. Seidl et al., *Chem. Phys. Lett.*, 1998, **287**, 537–541 (cit. on p. [53](#)).
- (78) O. Castano and P. Karadakov, *Int. J. Quantum Chem.*, 1982, **22**, 367–383 (cit. on p. [53](#)).
- (79) I. I. Ukrainsky, *Int. J. Quantum Chem.*, 1972, **6**, 473–489 (cit. on p. [53](#)).
- (80) A. Y. Sokolov, S. Guo, E. Ronca et al., *J. Chem. Phys.*, 2017, **146**, 244102 (cit. on p. [53](#)).

- (81) K. Nakayama, H. Nakano and K. Hirao, *International Journal of Quantum Chemistry*, 1998, **66**, 157–175 (cit. on p. 53).
- (82) F. Aquilante, J. Autschbach, R. K. Carlson et al., *Journal of computational chemistry*, 2016, **37**, 506–541 (cit. on p. 55).
- (83) G. Rossum, *Python Reference Manual*, tech. rep., Amsterdam, The Netherlands, The Netherlands, 1995 (cit. on pp. 60, 174).
- (84) E. Jones, T. Oliphant and P. Peterson, *SciPy: Open source scientific tools for Python*, en, 2001 (cit. on pp. 61, 175).
- (85) P. Nava, D. Hagebaum-Reignier and S. Humbel, *ChemPhysChem*, 2012, **13**, 2090–2096 (cit. on p. 73).
- (86) M. Kühn and F. Weigend, *The Journal of Chemical Physics*, 2014, **141**, 224302 (cit. on p. 88).
- (87) Y. Nakai, T. Mori and Y. Inoue, *J. Phys. Chem. A*, 2012, **116**, 7372–7385 (cit. on p. 94).
- (88) M. Rickhaus, M. Mayor and M. Juriček, *Chem. Soc. Rev.*, 2016, **45**, 1542–1556 (cit. on p. 103).
- (89) J. Lu, D. M. Ho, N. J. Vogelaar et al., *Journal of the American Chemical Society*, 2004, **126**, PMID: 15355095, 11168–11169 (cit. on p. 104).
- (90) R. A. Pascal, *Chemical Reviews*, 2006, **106**, PMID: 17165675, 4809–4819 (cit. on p. 104).
- (91) J. Xiao, H. M. Duong, Y. Liu et al., *Angewandte Chemie International Edition*, 2012, **51**, 6094–6098 (cit. on p. 104).
- (92) A. Bedi, L. J. W. Shimon and O. Gidron, *Journal of the American Chemical Society*, 2018, **140**, PMID: 29905480, 8086–8090 (cit. on p. 104).
- (93) Y. Xiao, J. T. Mague, R. H. Schmehl et al., *Angewandte Chemie International Edition*, 2019, **58**, 2831–2833 (cit. on p. 114).

- (94) A. Nakamura, K.-i. Yamanaka, K. Miyaoura et al., *J. Phys. Chem. C*, 2018, **122**, 16940–16949 (cit. on p. 117).
- (95) S. A. Ikbali, C. Colombari, D. Zhang et al., *Inorganic Chemistry*, 2019, **58**, 7220–7228 (cit. on p. 122).
- (96) O. Perraud, J.-B. Tommasino, V. Robert et al., *Dalton Trans.*, 2013, **42**, 1530–1535 (cit. on p. 122).
- (97) I. Gosse, K. Robeyns, C. Bougault et al., *Inorganic Chemistry*, 2016, **55**, PMID: 26771438, 1011–1013 (cit. on p. 122).
- (98) D. Leboeuf, L. Iannazzo, A. Geny et al., *Chemistry: A European Journal*, 2010, **16**, 8904–8913 (cit. on p. 142).
- (99) P. T. Snee, C. K. Payne, K. T. Kotz et al., *Journal of the American Chemical Society*, 2001, **123**, PMID: 11456872, 2255–2264 (cit. on p. 142).
- (100) B. J. Fallon, V. Corcé, M. Amatore et al., *New J. Chem.*, 2016, **40**, 9912–9916 (cit. on p. 142).
- (101) V. Gandon, N. Agenet, K. P. C. Vollhardt et al., *Journal of the American Chemical Society*, 2006, **128**, PMID: 16802817, 8509–8520 (cit. on pp. 142, 143, 180).
- (102) N. Agenet, V. Gandon, K. P. C. Vollhardt et al., *Journal of the American Chemical Society*, 2007, **129**, PMID: 17585766, 8860–8871 (cit. on p. 142).
- (103) P. Jungk, F. Fischer, I. Thiel et al., *The Journal of Organic Chemistry*, 2015, **80**, PMID: 26334453, 9781–9793 (cit. on p. 143).
- (104) S. Inack-Ngi, R. Rahmani, L. Commeiras et al., *Advanced Synthesis & Catalysis*, 2009, **351**, 779–788 (cit. on p. 143).
- (105) S. Inack-Ngi, K. Cherry, V. Héran et al., *Chemistry: A European Journal*, 2011, **17**, 13692–13696 (cit. on p. 143).

- (106) T. Helgaker, P. Jørgensen and J. Olsen, *Molecular Electronic-Structure Theory*, John Wiley and Sons Ltd, Baffins Lane, Chichester, West Sussex PO19 1UD, England, 2002 (cit. on p. [166](#)).
- (107) R. F. W. Bader, *Atoms in Molecules: A Quantum Theory*. Oxford University Press., 1994 (cit. on p. [168](#)).

Appendices

A. Physical Meaning of the p_z Pseudopotential

The α pseudosystem of Section 3.1 contains only one proton. However, it is plain that the effective charge felt by the p_z electron in a carbon atom should be larger than this, as the screening effect of the other electrons is not so complete that the p_z electron would feel a charge of only one. Slater's rules, for example, suggest an effective charge of $Z_{eff} = 2.4$. Thus we must account for this difference. We shall see that this role is filled by the p potential once optimised.

In order to evaluate the screening effect in the real CH_3^\bullet system, we computed the expectation value of the distance of the electron from the nucleus $\langle r \rangle$:

$$\langle r \rangle = \langle \psi_{p_z} | r | \psi_{p_z} \rangle = 1.80 \text{ a.u.} \quad (.1)$$

where ψ_{p_z} is the molecular orbital extracted from a reference calculation of CH_3^\bullet .

Since the CH_3^\bullet pseudo system has only one electron (as in a hydrogen-like atom), the analytical form of the p_z orbital is [106]:

$$\phi_{210} = \frac{1}{\sqrt{\pi}} \frac{Z_{eff}^{\frac{5}{2}}}{2a_0} r e^{-\frac{Z_{eff}r}{2a_0}} \cos \theta \quad (.2)$$

where Z_{eff} is the total nuclear attraction the electron would feel in the real CH_3^\bullet system, taking into account the screening effect of the core electrons that would be present in the all-electron system. This leads to the following expression for $\langle r \rangle(Z_{eff})$:

$$\langle \phi_{210} | r | \phi_{210} \rangle = \frac{5a_0}{Z_{eff}} \quad (.3)$$

From Equations .1 and .3, we obtain the theoretical value of $Z_{eff} = 2.77$.

The ψ_{p_z} molecular orbital is influenced strongly by the p pseudopotential, because of their overlap. In this paragraph we use a simplified definition \hat{p} of the p

pseudopotential, which allows for a quick evaluation of overlap effects with the basis set:

$$\chi = e^{-\frac{\alpha}{2}r^2} \sum_{m=-1}^{+1} |Y_{1,m}\rangle, \quad \hat{p} = -A|\chi\rangle\langle\chi| \quad (.4)$$

We define

$$S = \langle\psi_{p_z}|\chi\rangle \quad (.5)$$

the overlap between a molecular orbital ψ_{p_z} , and χ , leading us to evaluate the effect of the pseudopotential on the ψ_{p_z} molecular orbital as:

$$\langle\psi_{p_z}|\hat{p}|\psi_{p_z}\rangle = \langle\psi_{p_z}|-A|\chi\rangle\langle\chi|\psi_{p_z}\rangle = -AS^2 \quad (.6)$$

Finally, knowing that our hydrogen-like pseudosystem already contains a charge, $Z_{nucleus} = 1$, we expect that:

$$A = (Z_{eff} - Z_{nucleus})S^{-2} \quad (.7)$$

Ultimately, we used this expression to make informed guesses for the starting coefficients and exponents of the p Gaussian functions, from which we optimised the final pseudopotential. By using Equation .7, and the optimised exponent of $\alpha = 0.624$, an estimated value of $A = 3.614$ is obtained. The optimised coefficient $A = 3.909$ of our the *set4* potential set is very close to this estimation, supporting the idea that the p potential retrieves the incomplete screening effect of the missing electrons.

B. Topological Analysis of Electron Density

An analysis can be performed on the electronic density obtained with the pseudopotentials to create a visualisation of the π -only electronic cloud [107]. In the Atoms in Molecules (AIMs) framework, critical points are characterised by the sign of the eigenvalues of the Hessian of the electronic density and are classified as follows: as the maxima of the electronic density are located at the nuclear positions, atom critical points are characterised by three negative eigenvalues, noted (3, -3); a bond critical point is a saddle point, showing a minimum along the bond direction, thus noted (3, -1); cycle critical points are minima on the hypersurface of the cycle and maxima along the direction perpendicular to the cycle, noted (3, +1). A last kind of critical point (cage critical point) is characterised by three positive eigenvalues (3, +3). The critical paths are obtained by connecting the critical points along a direction of zero gradient.

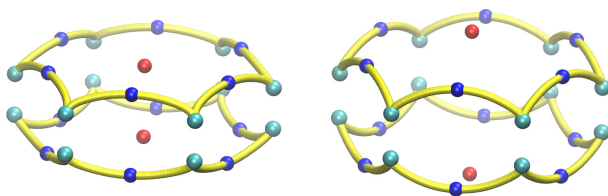


Figure .1. AIMs diagrams of benzene, all-electron (left) and with *set4* pseudopotentials (right). Electronic density maxima are in green, bond critical points are in blue, and cycle critical points are in red.

For benzene, we compare the critical points of the π cloud obtained in a reference calculation with pseudopotentials at the same level of theory. The π system of the all-electron calculation is obtained by populating the converged π molecular orbitals only. As the AIM procedure deals with the electronic density, no reoptimisation of the orbitals is necessary. In our case, as the critical point analyses were obtained with the π system only, the electronic density maxima are not located at the nuclei but rather at the maxima of the π density, *i.e.* above and below the nuclei. Accordingly, twelve bond critical points rather than six are

expected, above and below the molecular plane. Moreover, due to the topology of the density, two cycle critical points are found above and below the molecular plane on the C_6 symmetry axis. To fulfil the Poincaré-Hopf relationship, a cage critical point is required. However, this cage critical point at the centre of the molecule could not be found, either in the reference or in the pseudopotential calculations. After a careful inspection of the results, it turns out that the electronic density is below computer accuracy at this point (it is evaluated to be around 10^{-31}). Therefore we deduce there is a cage critical point at the centre of the molecule, but that it can not be detected.

Finally, Figure .1 shows that the description of the bond pattern by the pseudopotential calculation corresponds to that by the reference, in terms of types of found critical points. The position of the critical points are not exactly the same and the pseudopotential calculation leads to a slightly more diffuse π bonding pattern. However, as both images have the same scale, one can see that the topologies of the electronic densities are very similar. This confirms that the results obtained are not artefactual but that the pseudopotentials lead to a physically relevant electronic density.

Furthermore, it is possible to see that the physics of the excitations is reproduced as well by an analysis of the transition densities. Figure .2 shows the transition densities for the two lowest-energy singlet excitations of pyrene, and it is plain that the pseudopotentials have faithfully reproduced the all-electron molecular orbital shapes.

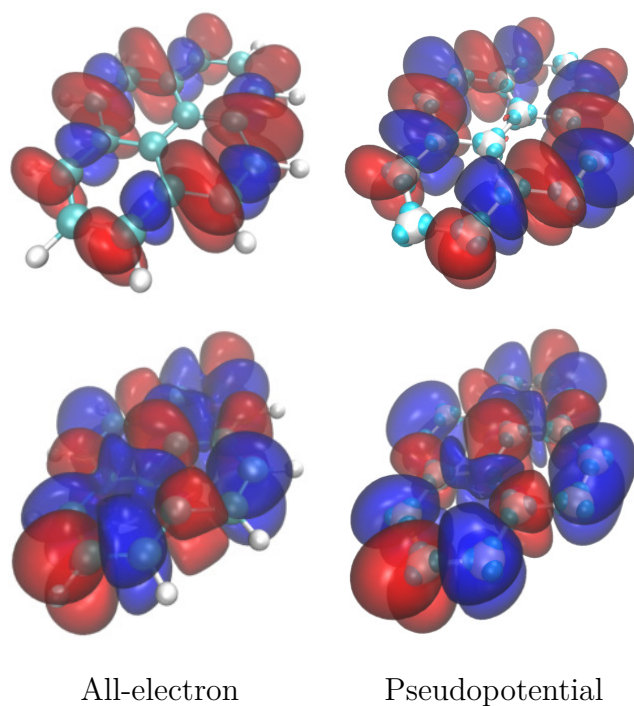


Figure .2. Transition densities of the first two (lowest-energy) pyrene absorption peaks, for singlet-singlet excitation spectra, as seen in Figure 3.18, with all-electron spectra (left) and *set4* pseudomolecule spectra (right). In each transition, the electron density is decreased in the blue-coloured lobes and increased in the red.

Table .1 contains the relative molecular orbital weights of the singlet-triplet excitation of $C_{50}H_{52}$, for both all-electron and pseudopotential TDDFT calculations. Once more it is possible to see that the pseudopotential excitation uses the correct molecular orbitals, with a very similar weighting to that of the all-electron system.

The remarkable feature is the neatness of the agreement of the relative excitation intensities in both singlet and triplet cases. This demonstrates that these pseudopotentials capture a lot of the physics of the reference system, as the intensities are related to the transition dipole moment.

Table .1. Comparison of the weights (all-electron *vs.* *set4* pseudopotentials) of the excitations obtained with TD-DFT to represent the triplet excited state from the closed shell singlet state. Example case of C₅₀H₅₂.

Excitation MO	Weight(%)	
	Ref.	Pseudo.
25 a" → 26 a"	77.0	67.1
24 a" → 27 a"	10.5	13.1
23 a" → 28 a"	3.6	5.2

Résumé

Dans cette thèse, une méthode de pseudopotentiels est développée et testée pour le remplacement de petits fragments dans des calculs moléculaires de chimie quantique.

Cette méthode emprunte des principes à une sélection de méthodes de pseudopotentiels historiques et actuelles. Elle a pour origine le même concept, à savoir que les électrons peuvent être divisés entre ceux qui sont actifs, et participent aux interactions chimiques d'intérêt, et ceux qui sont inactifs, et ne participent pas. Précisément, ces pseudopotentiels sont de la forme :

$$\sum_k A_k (r^{n_k-2}) e^{-\alpha_k r^2} \quad (.8)$$

ou A est un coefficient, α est un exposant, et k un entier.

Cependant, au lieu de placer la division entre les électrons de cœur et de valence, la méthode choisit comme électrons actifs seulement un ou deux électrons spécifiques qui participent au comportement chimique qui nous intéresse. Cette méthode élimine un grand nombre d'électrons, ainsi que plusieurs noyaux et protons. Par conséquent, pour récupérer les formes correctes de la densité électronique, quelques uns des pseudopotentiels ne sont pas centrés sur les atomes ce qui donne l'expression mathématique finale :

$$\hat{W} = \underbrace{\frac{A}{r} \exp(-\alpha r^2) \sum_m |Y_{1,m}\rangle \langle Y_{1,m}|}_{\text{centré sur l'atome}} + \underbrace{\sum_J \frac{C_J}{r - r_J^0} \exp(-\gamma_J (r - r_J^0)^2) |Y_{0,0}\rangle \langle Y_{0,0}|}_{\text{pas centré sur l'atome}} \quad (.9)$$

avec $Y_{0,0}$ l'harmonique sphérique s , $Y_{1,m}$ les harmoniques sphériques p et r_J^0 la position relative fixée du $J^{\text{ème}}$ potentiel par rapport à l'origine du pseudoatome auquel les potentiels sont assignés. Cet opérateur \hat{W} peut ensuite s'ajouter aux opérateurs monoélectroniques

$$\hat{h}(i) = -\frac{1}{2}\Delta_i - \frac{1}{r_i} + \sum_K \hat{W}_k \quad (.10)$$

ou K représente le nombre de pseudofragments.

Les fonctions de base utilisée ne sont pas modifiées par rapport aux calculs tous électrons. On inclut dans la procédure de l'optimisation des éléments des deux écoles 'shape-consistent' et 'size-consistent'. Finalement, cette méthode devrait être compatible avec la plupart des logiciels de chimie quantique.

D'abord, nous avons créé un fragment de carbone hybridé sp^2 (surnommé le pseudopotentiel α), avec un seul noyau et un électron. Ce système a été optimisé pour reproduire les énergies de la HOMO, l'énergie de 1^{ère} ionisation et l'écart singulet-triplet du système éthylène. Dans cette entreprise, nous avons réussi, avec les différences entre les systèmes calculé pseudo et tous électrons de 7,7%, 2,9% et 0,0% respectivement. Ce fragment sp^2 a été ensuite transféré dans une série de trans-polyènes, jusqu'à une longueur de 100 atomes carbones. Un travail identique a été effectué avec une série d'hydrocarbures aromatiques polycycliques (HAP), jusqu'à 19 anneaux de six carbones. Les orbitales π étaient bien reproduites dans tous les cas. L'étude a été étendue à plusieurs fonctionnelles de la densité standards, et il a été trouvé que les différences moyennes entre les résultats tous électrons et pseudopotentiels pour les valeurs de l'énergie de la HOMO, la 1^{ère} ionisation et la 1^{ère} excitation var-

ées de 7,8% pour la meilleur fonctionnelle (PBE0) à 20,1% pour la pire (TPSS). Nous n'avons pas observé de détérioration de ces résultats avec la longueur du polyène, et les différences étaient similaires pour les systèmes HAP. Ce résultat signifie que nous avons capturé les effets physiques de la délocalisation π sur les trans-polyènes, en même temps que l'aromaticité sur les HAP. Les résultats DFT sont résumés dans la Table .2, où l'on compare les systèmes et les fonctionnels.

Table .2. Erreurs dans les calculs pseudopotentiels pour les trans-polyènes (C_nH_{n+2}) et les HAP, pour une sélection des fonctionnelles de la densité. Les excitations sont verticales.

différence moyenne (%)	PBE0	PBE	TPSS	TPSSh
polyènes courts (n=2,4,6,8,10)				
1 ^{ère} Ionisation	7,0	8,8	11,3	9,9
HOMO Singulet	4,2	9,3	13,7	10,7
1 ^{ère} Excitation TDDFT	2,6	2,7	5,9	6,4
longs polyènes (n=20, 30, 40, 50, 100)				
1 ^{ère} Ionisation	6,6	10,3	11,3	11,2
HOMO Singulet	7,8	12,8	20,1	15,7
1 ^{ère} Excitation TDDFT	1,0	6,8	6,1	2,9
les HAP				
1 ^{ère} Ionisation	12,0	15,4	18,0	16,1
HOMO Singulet	11,5	17,1	22,3	18,7
1 ^{ère} Excitation TDDFT	4,4	2,6	5,9	6,5

Il a aussi été démontré que pseudopotentiels fonctionnaient avec les calculs CASPT2.

Nous avons souhaité généraliser la méthode développée ci-dessus, afin qu'elle puisse être utilisée pour créer d'autres pseudosystèmes. À partir des scripts que nous avons développés ci-dessus, nous avons créé un programme général de minimisation, l'optimiseur multiple orbital (MOO), qui permet de minimiser l'erreur sur des critères choisis entre les calculs avec des pseudosystèmes et ceux avec tous les électrons. Ce programme est écrit en Python [83] et fonctionne comme suit:

1. Soit en entrée une série de calculs de pseudosystèmes.
2. Soit en entrée une gamme de propriétés moléculaires observables, ainsi qu'en référence leurs valeurs tous électrons.
3. Minimiser les différences tous électrons/pseudosystèmes de manière itérative en
 - a) modifiant des paramètres des pseudopotentiels (coefficients, exposants, positions);
 - b) exécutant les calculs dans Turbomole [72];
 - c) calculant les propriétés spécifiées à l'étape 2;
 - d) mettant à jour les différences entre référence et pseudosystèmes.

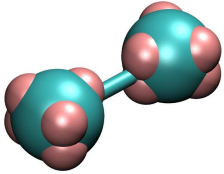
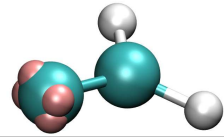
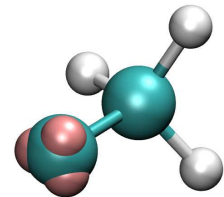
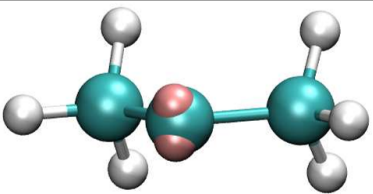
La minimisation elle-même est effectuée via un algorithme de programmation quadratique séquentielle (PQS) implémenté dans SciPy [84]. Cette méthode a été choisie pour la flexibilité des contraintes de limites implémentées dans SciPy.

Cette procédure fonctionne avec une variété de critères de références spécifiés par l'utilisateur. Il peut commencer par une estimation, ou par un processus semi-aléatoire.

Nous avons continué par créer une série de nouveaux fragments et à les optimiser avec notre logiciel. Ces fragments sont un carbone sp^2 de deux électrons, (un deuxième fragment d'éthylène, mais qui peut, cette fois, former une liaison σ et une liaison π), un fragment carbone sp^3 d'un électron (un groupe méthyl), et un carbone sp^3 de deux électrons (un 'lien atomique' entre deux atomes explicites, optimisé sur l'atome central du propane). Ils ont pris les sobriquets de β , γ et δ respectivement. Chacun de ces potentiels a été pris et examiné dans une sélection de molécules afin de découvrir leur réponse à des environnements chimiques différents. Tous ces potentiels ont bien performé quand leurs voisins étaient homoatomiques (cad avec un atome carbone explicite), avec une erreur entre systèmes tous électrons et pseudopotentiels allant jusqu'à environ 20%

(énergie de la HOMO en TPSS), mais généralement de l'ordre de quelques %. Il a été aussi trouvé que les erreurs augmentent considérablement (d'environ 100% dans le cas extrême du formaldéhyde), quand les pseudopotentiels étaient obligés de former leur liaisons avec des hétéroatomes comme l'azote, l'oxygène ou le fluor. On peut voir les diagrammes et résumé des potentiels créés dans la Table .3.

Table .3. Un résumé des types de potentiels créé. Les atomes qui ne sont pas posés sur les centres atomiques sont soulignés en rouge.

Diagramme	Nom	Description
	α	sp^2 1-électron
	β	sp^2 1-électron
	γ	sp^3 2-électron
	δ	sp^3 2-électron

Nous avons par la suite utilisé les pseudopotentiels α pour reproduire des spectres ultra-violet, ce que nous avons trouvé être très efficace. Les pseudopotentiels, optimisé seulement sur éthylène, étaient capables de reproduire les spectres d'excitation singulet et triplet pour une gamme de molécules HAP. Les spectres étaient clairement définis, avec tous les pics de tous électrons explicites identifiables dans les spectres pseudosystèmes, avec les intensités cor-

rects (pour la plupart) et avec seulement un décalage en fréquence faible et constant.

Les pseudopotentiels pouvaient aussi reproduire le même résultat que AIM (Atoms In Molecules) pour la distribution de la densité électronique, et pouvaient reproduire les densités de transitions électroniques fondées physiquement.

Avec le code d'optimisation amélioré, nous avons créé des potentiels α en utilisant différents critères de référence dans le but de déterminer ceux qui conduisent à la meilleure reproduction des spectres ultraviolets. Nous avons testé l'utilisation des énergies TDDFT, des énergies d'orbitales virtuelles et l'ajustement des spectres pseudopotentiels aux spectres de référence directement via une méthode des moindres carrés. Il a été constaté que (1) les énergies TDDFT n'étaient pas fiables, du moins en elles-mêmes (2) les ajustements des moindres carrés étaient prometteurs, mais créaient d'autres problèmes en soi, et enfin (3) que l'utilisation des orbitales virtuelles permettait une amélioration des spectres par rapport au pseudopotentiel d'origine, bien que l'absence simultanée de nouveaux critères de référence à partir de la surface d'énergie potentielle du triplet signifiait que, contrairement aux pseudopotentiels d'origine, ils ne pouvaient pas produire d'excitations triplets du tout. En même temps, ces potentiels avaient tendance à détériorer les critères originaux. Un exemple est inclus dans la Figure .3, pour quelques spectres HAP créés avec les meilleurs pseudopotentiels.

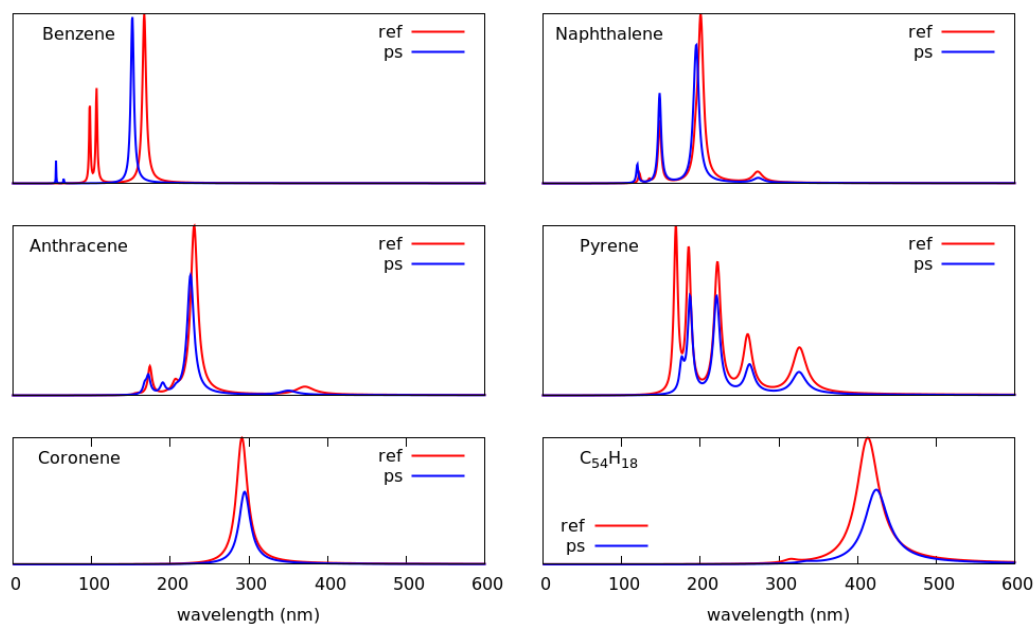


Figure .3. Comparaison des spectres pour les 20 premières excitations obtenues avec un système tous électrons (ref) et pseudopotentiel (ps). Calculs effectués au niveau TD-PBE0 dans le cadre RPA.

Ensuite, tous les types de pseudopotentiels créés (α , β , γ , δ) ont été testés dans une gamme de systèmes beaucoup plus grands et complexes. De cette manière nous espérons trouver les vrais limites de cette méthode. Ces systèmes incluaient les spectres ultra-violet et de dichroïsme circulaire électronique de systèmes avec une distorsion π de plus en plus forte ([n]héliène, twistacène, coronène intégré dans un nanotube carbone), les anneaux π qui se chevauchent (dodécaphényltétracène), et avec des métaux (un ion cuivre dans une cage de type hemi-cryptophane). Plusieurs conclusions ont été tirées de ces expériences :

1. Des systèmes hélicènes, twistacènes et nanotube, nous avons appris que lors de l'utilisation de pseudopotentiels pour des cycles aromatiques avec seulement des électrons π traités explicitement (potentiels α), les distorsions dans le plan π ne détériorent pas les spectres tant qu'elles ne sont

que de quelques degrés. Au-dessus de cette valeur cependant, les spectres ECD deviennent rapidement moins fiables.

2. À partir de l'hélicène, du twistacène et du dodécaphényltétracène, nous avons appris que des spectres ultraviolets d'une qualité constante peuvent être générés même après remplacement de la plupart ou de la totalité des atomes explicites de la molécule avec des pseudopotentiels, et que cela reste vrai même jusqu'au système le plus déformé testé (twistacène, 38 par anneau de benzène).
3. A partir du système de cage, nous avons appris que permettre à des fragments pseudopotentiels de se lier directement aux hétéroatomes risquait de créer des spectres ultra-violet et dichroïsme circulaire électronique médiocres. Afin de s'assurer que les caractéristiques spectrales ultra-violet et de dichroïsme circulaire électronique ont été correctement reproduit, il était nécessaire de conserver au moins un atome carbone explicite entre un fragment pseudopotentiel et un hétéroatome.

Ensuite, nous avons créé et testé une technique permettant d'optimiser la géométrie avec les pseudopotentiels. Cela impliquait de créer des courbes de dissociation entre pseudoatomes et atomes tous électrons pour les potentiels β et γ , puis d'ajuster une correction sur les deux atomes afin de garantir qu'ils restent à la distance de liaison correcte lors de l'optimisation géométrique. Cette technique a été testée sur une gamme de différentes petites molécules et les résultats ne diffèrent pas de ceux des études précédentes, c'est-à-dire que si les pseudocarbones étaient liés à des atomes de carbone tous électrons, l'approximation fonctionnait bien, avec des erreurs faibles pour la longueur et l'angle de liaison. Lorsque l'atome tous électrons était un hétéroatome, l'approximation fonctionnait moins bien, avec des erreurs plus grandes sur la longueur et l'angle de la liaison. Cependant, le principal problème résidait dans le fait que les fragments potentiels étaient attractifs pour d'autres atomes à proximité, provoquant

fréquemment un effondrement géométrique de la molécule. Certaines améliorations possibles de la méthode sont suggérées, mais dans sa forme actuelle, la technique d'optimisation géométrique est trop peu fiable pour une utilisation générale.

Enfin, nous avons mené une petite étude sur les gains de calcul rendus possibles par l'utilisation de cette méthode pseudopotentiels. Nous avons constaté que pour la petite base def2-SV(P), un gain de 2,4 était réalisé dans le temps par itération SCF pour un calcul sur $C_{50}H_{52}$. Pour un ensemble de base QZVPPD plus important, le gain a été augmenté à 8,0. Cette partie de l'étude n'a pas été poussée plus loin, car il a été noté qu'un travail supplémentaire serait nécessaire pour tirer pleinement parti des gains rendus possibles par les pseudopotentiels.

En résumé, nous avons créé une technique pseudopotentielle qui parvient à conserver de nombreuses propriétés de base des fragments moléculaires remplacés, notamment les énergies des orbitales, d'ionisation et d'excitation suffisantes pour la production de spectres UV et ECD complets. Ceci en dépit du fait que la plupart des électrons et une partie des noyaux sont retirés de ces fragments. Nous avons testé ces potentiels et modifié certains principes généraux pour leur utilisation. Nous avons également créé un algorithme général pour la création de ces potentiels et étudié les types de propriétés moléculaires qui constituent des critères de référence appropriés, basés sur des intuitions des types de propriétés moléculaires qui nous aideront à capturer la physique sous-jacente que l'on souhaite simuler. Enfin, nous avons établi que, s'ils sont utilisés avec précaution, ces potentiels peuvent reproduire des propriétés électroniques de systèmes étonnamment complexes.

Sans lien avec le travail ci-dessus, une étude d'une réaction de cycloaddition en présence de cobalt a été réalisée afin de tenter de comprendre un résultat expérimental inattendu. Les molécules à la fois attendues et obtenues expérimentalement, sont présentées dans la Figure .4. Sur la base de travaux expérimentaux antérieurs [101], on pourrait s'attendre à ce que cette cycloaddi-

tion en présence de cobalt produise un complexe avec un triène central ou une spirolactone avec un cycle central à six chaînons. Au lieu de cela, le complexe tétracyclique **P1** est formé en tant que seul diastéréomère.

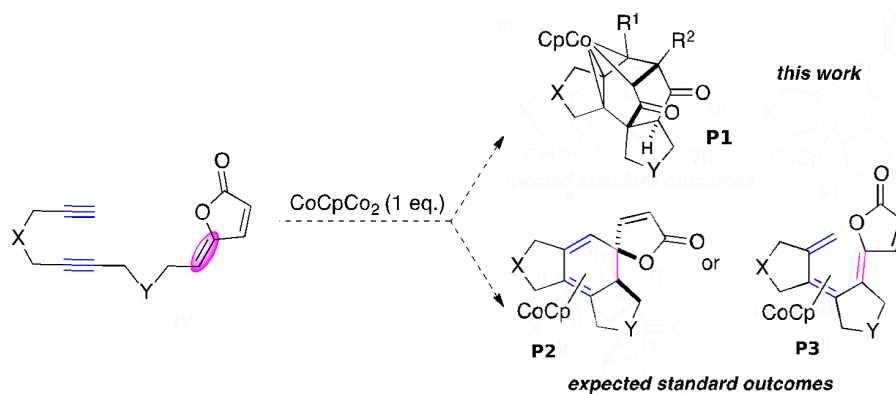


Figure .4. Résultat standard attendu du processus utilisant le cobalt (I) par rapport au résultat expérimental.

Dans cette étude, l'importance particulière de la lactone a été établie pour provoquer les résultats expérimentaux observés, en la comparant avec le comportement d'une molécule théorique non équivalente à une lactone.

Mots clés: pseudopotentiels, potentiels modèles, carbone sp^2 , carbone sp^3 , HAP, hélicène, twistacène, hemi-cryptophane, hybridation $s - p$