

University of Groningen

Density functional theory applied to copper proteins

Swart, Marcel

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2002

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Swart, M. (2002). *Density functional theory applied to copper proteins*. [Thesis fully internal (DIV), Groningen]. s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

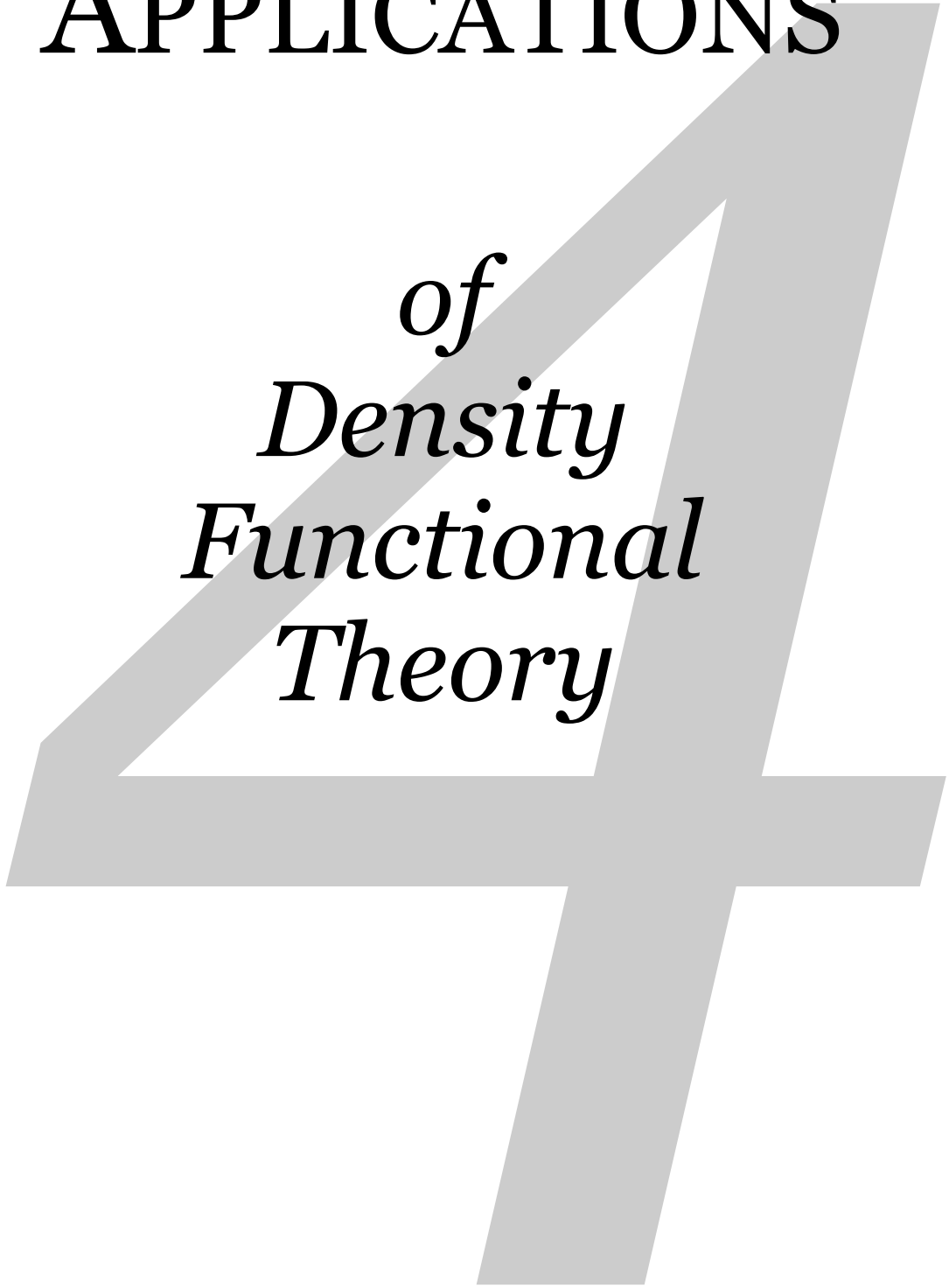
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

chapter four

APPLICATIONS

*of
Density
Functional
Theory*



CONTENTS

4.1	<i>Polarizabilities</i>	61-67
4.2	<i>Accuracy of geometries</i>	68-80
4.3	<i>Validation of charge analyses</i>	81-86

SUMMARY

The application of Density Functional Theory to three different chemical topics is discussed in Chapter 4. First the computation of molecular polarizabilities and the influence of the choice for the basis set and exchange-correlation potential on the accuracy is discussed; then the accuracy of optimized geometries of several exchange-correlation potentials in a number of basis sets is presented, where the test set consists either of a set of small molecules that was used previously by others to check the accuracy of several wavefunction based methods. Finally, the new charge analysis presented in Chapter 3 is validated by checking its use for the concepts of molecular recognition, electron withdrawing-donating groups and electrophilic substitution reactions.

Polarizabilities

A comparison of Restricted Hartree Fock, Density Functional Theory and Direct Reaction Field mean values for organic molecules

There is a growing interest in atomic and molecular polarizabilities within the scope of the development of accurate force fields to be used in QM/MM methods^{133,134,138,192-194}. Dispersion and induction forces are more and more being recognized as essential parts in the description of intermolecular interactions in chemical environments^{138,195-201}.

It has also been noticed that using a molecular polarizability located in the center of the molecule, often leads to improper behavior^{138,202}. This can be overcome by using effective atomic polarizabilities that represent the molecular value. A set of these atomic parameters have been constructed within the framework of the Direct Reaction Field (DRF) method, which predict molecular polarizabilities with a deviation comparable to the estimated experimental uncertainty^{203,204} (2-4 %). In this study, we present DRF results for a representative series of organic molecules. For comparison, we have also performed Restricted Hartree Fock (RHF) and Time Dependent Density Functional Theory (TD-DFT) calculations on the same set of molecules, the latter also at non-zero frequency.

The computation of a polarizability α is based on a Taylor expansion of the total energy about the electric field strength E :

$$U = U^{(0)} - \mu_i^{(0)} E_i - \frac{1}{2!} \alpha_{ij} E_i E_j - \frac{1}{3!} \beta_{ijk} E_i E_j E_k - \frac{1}{4!} \gamma_{ijkl} E_i E_j E_k E_l - \dots \quad (1)$$

with $U^{(0)}$ the unperturbed total energy, $\mu^{(0)}$ the permanent dipole moment, α the polarizability and β and γ the first and second hyperpolarizabilities. The appropriate derivatives of the energy with respect to the electric field then result in expressions for the (hyper)polarizabilities. This derivative can be obtained either analytically or numerically by a finite field procedure. The former is faster but not standard available for all kinds of quantumchemical methods, while the latter requires only the calculation of the energy or the dipole moment, but is in general much slower, since the energy/dipole needs to be calculated many times.

Methods

Restricted Hartree Fock

The RHF results are obtained with Coupled Perturbative Hartree Fock (CPHF) equations²⁰⁴ (as implemented in the HONDRF program¹³³), which take the response of a molecule to an applied field analytically into account. Dunning's triple zeta valence plus polarization functions basis set (TZP)²⁰³ was used for all molecules, which is the largest standard (Gaussian) basis set in the HONDRF program and has proven to give an underestimation of the polarizabilities by $\sim 22\%$ ¹³⁹. A lot of effort has been put into constructing basis sets especially designed for accurate calculations of the polarizability²⁰⁵⁻²⁰⁹. However, they are based on Gaussian type orbitals, which makes a comparison with the ADF results (using Slater type orbitals) impossible. Moreover, it is not the purpose of this study to compare *basis*

sets or to construct them for the ADF program, but rather to make an honest comparison between the three *methods* using the programs and basis sets that are readily available.

Density Functional Theory

The DFT results were all obtained by using the RESPONSE code²¹⁰⁻²¹² in the Amsterdam Density Functional (ADF) program^{117,177,213-215}. This Time Dependent DFT (TD-DFT) method is at zero frequency similar to a CPHF procedure, but then applied to DFT SCF equations which gives the same polarizabilities as when using a finite field procedure. The ADF program uses basis sets of Slater functions, of which two were used: a triple zeta valence plus polarization (TZP, basis set IV in ADF; of equal size as Dunning's TZP basis) and a special one with triple zeta valence plus double polarization functions and diffuse s, p, d functions (TZ2P++). This basis set has been constructed for improving the results and has been reported to give accurate polarizabilities²¹⁰. Several potentials were used: LDA (VWN¹¹⁹) and GGA's (Becke88 exchange¹²⁰-Perdew86 correlation¹²¹ potential, Van Leeuwen-Baerends LB94²¹⁶).

Direct Reaction Field

Within the DRF approach¹³³, a molecular polarizability is being constructed from interacting atomic polarizabilities α_p :

$$\mu_p = \alpha_p E_0 + \sum_{q \neq p}^N T_{pq} \mu_q \quad (2)$$

with the *modified* dipole field tensor T_{pq} :

$$(T_{pq})_{ij} = (3\mathbf{r}_i \mathbf{r}_j f_T - \delta_{ij} r^2 f_E) / r^5 \quad (3)$$

This can be written also as a matrix equation:

$$\mathbf{M} = \alpha(\mathbf{E} + \mathbf{TM}) \quad (4)$$

which can be solved directly with the matrix $\mathbf{A} = \alpha^{-1} - \mathbf{T}$, by using the inverse of matrix \mathbf{A} :

$$\mathbf{AM} = \mathbf{E} \quad \mathbf{M} = \mathbf{A}^{-1}\mathbf{E} \quad (5)$$

This so-called relay matrix $\mathbf{R} = \mathbf{A}^{-1}$ gives the linear response of the molecule to a given external field, i.e. its polarizability in a $3N \times 3N$ representation. It can be reduced to a "normal" 3×3 -tensor, resulting in the molecular polarizability tensor:

$$\alpha_{mn} = \sum_{i,j=1}^N (\mathbf{R}_{ij})_{mn} ; m, n \in \{x, y, z\} \quad (6)$$

The modification is present in the screening factors f_T and f_E , which represent the damping due to overlapping charge densities. Several functions have been tested for this damping²¹⁷, of which one has survived:

$$\begin{aligned}
 \rho(u) &= \frac{a^3}{8\pi} e^{-au} & u &= r_{ij} / (\alpha_i \alpha_j)^{1/6} & f_V &= 1 - \left(\frac{1}{2}v + 1\right)e^{-v} \\
 & & v &= au & f_E &= f_V - \left(\frac{1}{2}v^2 + \frac{1}{2}v\right)e^{-v} \\
 & & & & f_T &= f_E - \frac{1}{6}v^3e^{-v}
 \end{aligned}
 \tag{7}$$

In these equations the a -factor and the atomic polarizabilities are adjustable parameters obtained in the fit procedure¹³⁹ (see Table 4.1.1).

TABLE 4.1.1. DRF ATOMIC POLARIZABILITIES (BOHR³) WITH A-FACTOR 2.1304

<i>atom</i>	α	<i>atom</i>	α	<i>atom</i>	α
H	2.793	O	5.749	Cl	16.198
C	8.696	S	16.698	Br	23.571
N	6.557	F	3.001	I	36.988

Computational details

The DRF and RHF calculations were carried out on the Cray J932 supercomputer in Groningen. All ADF results were calculated at a SGI PowerChallenge workstation. The geometries of the molecules were taken as much as possible from experimental data²¹⁸. The remaining (internal) coordinates were optimized by the PM3 method in MOPAC93²¹⁹.

Results and discussion

The calculated mean polarizabilities of the 15 molecules are given in Table 4.1.2, together with experimental data. The experimental mean values are obtained mostly from the refractive index n (at 5893 Å ; sodium D-line) and the Lorentz-Lorentz equation (with M molecular weight, ρ macroscopic density, N_{av} Avogadro's number):

$$\frac{n^2 - 1}{n^2 + 2} \frac{M}{\rho} = \frac{4\pi}{3} N_{av} \bar{\alpha}
 \tag{8}$$

These values should be extrapolated to infinite wavelength (or zero frequency) to obtain the static polarizability. This scales the values down by 1-4 %²²⁰, and gives an estimate of the experimental uncertainty when using the uncorrected values.

The experimental data can also be obtained from the Kerr constants²²⁰, but this introduces, apart from the wavelength dependency, an additional uncertainty. One has to make an assumption of the geometry, thereby limiting this method to (small) symmetric molecules. The additional uncertainty has been estimated to be 5-10 %²²⁰, giving a total uncertainty of at least 6-14 %.

In Table 4.1.3, some experimental polarizability components are given, which were taken from references 220 and 221. Some other components were left out, because of spurious

assumptions that increase the uncertainty even more. If necessary, the components were uniformly scaled to reproduce the average mean polarizability. It should be noted also, that some *experimental* components reported by Applequist et al.²²¹ are extrapolated empirical values; i.e. experimental Kerr constants were obtained for some molecules which were then extrapolated to similar (larger) molecules.

TABLE 4.1.2. MOLECULAR POLARIZABILITIES (BOHR³) USING TZP BASIS SET

	exp	DRF	RHF	LDA	BP	LB94
Acetamide ^p	40.5 ^{Rl}	38.6	31.1	40.5	39.5	39.2
Acetylene ^t	22.5 ^{Kg}	21.9	18.0	19.6	19.3	19.2
Benzene ^p	70.1 ^{Rl}	61.9	61.9	66.7	65.7	65.9
Chlorine ^p	31.1 ^{Kg}	31.2	18.7	25.3	24.9	25.4
Cyclohexanol ^t	79.9 ^{Rl}	78.0	69.2	79.2	77.0	77.0
Dimethylether ^p	35.0 ^{Rg}	35.4	27.6	34.8	33.8	33.4
Formaldehyde ^p	16.5 ^{Rg}	18.3	13.5	16.6	16.3	16.1
Hydrogen ^p	5.3 ^{Kg}	4.9	2.6	4.8	4.6	4.7
Methylcyanide ^p	29.7 ^{Rl}	29.8	24.6	28.4	27.8	28.2
Neopentane ^t	69.0 ^{Rl}	65.5	59.0	67.8	65.7	65.7
Propane ^p	42.4 ^{Rl}	42.2	35.9	42.1	40.8	41.5
TCFM ^t	57.5 ^{Rl}	56.5	40.0	53.2	52.8	52.7
TCMC ^t	70.5 ^{Rl}	68.3	54.0	66.1	65.0	66.1
TFM ^p	19.0 ^{Kg}	19.0	13.1	17.9	17.5	16.8
Water ^p	9.94 ^{Rl}	10.1	5.6	8.4	8.3	8.0
Time		~ 1 s	57 hr	9 hr	14 hr	23 hr
Average deviation (%)		-1.8 ± 4.8	-24.9 ± 11.5	-6.0 ± 5.8	-8.1 ± 5.5	-8.2 ± 5.5
Average absolute dev. (%)		3.6 ± 3.6	24.9 ± 11.5	6.1 ± 5.7	8.1 ± 5.5	8.2 ± 5.5
Relative to LB94 (%)				+2.4 ± 1.9	+0.1 ± 2.0	
TCFM	trichlorofluoromethane					
TCMC	trichloromethylcyanide					
TFM	trifluoromethane					
p	Molecules used to obtain DRF-parameters ¹³⁹					
t	Molecules used to test DRF-parameters					
K	Obtained from Kerr constants (uncertainty ca. 5 %)					
R	Obtained from refractive index (uncertainty ca. 0.2 %)					
g	Gas phase					
l	Liquid phase					
DRF	Direct Reaction Field approach result ¹³⁹					
RHF	Restricted Hartree Fock					
LDA	Local Density Approximation result (Vosko-Wilk-Nusair potential ¹¹⁹)					
BP	Becke88 exchange ¹²⁰ with Perdew correlation ¹²¹ potential					
LB94	Van Leeuwen-Baerends (LB94) potential with correct asymptotic behavior ²¹⁶					
	Polarizabilities at 5893 Å					

Restricted Hartree Fock

The RHF results give an average deviation of 24.9 %, which has been reported before¹³⁹, and can be attributed to two factors. The first is the (small) basis set ; it is known that specially constructed²⁰⁵⁻²⁰⁹ or large (including very diffuse functions)²¹⁰ basis sets are needed to obtain accurate results. This is especially apparent from the results of planar and linear molecules, where the out-of-plane polarizability is very small in comparison to the experimental and DFT results.

The second factor is the absence of electron correlation in the RHF method. Inclusion of this correlation (by for instance Configuration Interaction, Multi Configuration SCF, Coupled Cluster or Møller-Plesset methods) increases the work needed enormously, and can be applied only to relatively small molecules.

TABLE 4.1.3. POLARIZABILITY COMPONENTS[#] (BOHR³)

	exp			DRF			RHF			LB94		
	xx	yy	zz	xx	yy	zz	xx	yy	zz	xx	yy	zz
Acetylene	18.5	18.5	30.5	15.1	15.1	35.6	11.5	11.5	31.0	17.7	17.7	31.8
Benzene	82.7	82.7	45.0	74.9	74.9	35.8	74.8	74.8	36.0	83.7	83.7	41.7
Chlorine	24.4	24.4	44.5	27.3	27.3	39.0	11.0	11.0	34.2	24.8	24.8	40.0
Dimethylether	33.0	29.3	42.6	34.1	32.6	39.4	25.9	25.7	31.2	32.9	32.8	41.0
Hydrogen	4.9	4.9	6.3	4.1	4.1	6.5	0.7	0.7	6.5	4.9	4.9	7.0
Methylcyanide	25.5	25.5	38.1	24.9	24.9	39.7	19.2	19.2	35.4	25.0	25.0	42.4
Neopentane	69.0	69.0	69.0	65.5	65.5	65.5	59.0	59.0	59.0	68.6	68.6	68.6
TCFM	59.7	59.7	53.0	60.0	60.0	49.4	44.3	44.3	31.4	63.3	63.3	51.3
TCMC	69.6	69.6	72.3	67.1	67.1	70.6	53.0	53.0	56.0	68.2	68.2	84.6
TFM	19.4	19.4	18.1	18.8	18.8	19.2	13.4	13.4	12.5	19.1	19.1	17.6
Av. dev. (%)						-3.4 ± 9.4			-27.7 ± 20.5			0.9 ± 7.2
Av.abs. dev. (%)						8.0 ± 5.9			28.1 ± 20.1			5.0 ± 5.2

the molecules are oriented with the main symmetrical axis along the z-axis
TCFM trichlorofluoromethane
TCMC trichloromethylcyanide
TFM trifluoromethane
exp experimental polarizability components^{220,221} (accuracy 6-14 %)
DRF Direct Reaction Field approach result¹³⁹
TZP CPHF value with TZP basis set
VII Van Leeuwen-Baerends (LB94) potential²¹⁶ in basis set VII (TZ2P++)

Density Functional Theory

Much more promising and accurate are the DFT results. The results from all three potentials in the TZP basis set give a much better accuracy (dev. 6-8 %; Table 4.1.2) for the mean values. When using the TZ2P++ basis set (Table 4.1.4), which has been reported as the basis set limit for (other) organic molecules²¹⁰, the deviations become even smaller (LDA 5.8 %, Becke-Perdew 3.6 %, LB94 potential 3.0 %). In both basis sets, the LB94 results are smaller than either the LDA or the Becke-Perdew results. The DFT polarizability components show, like the mean value, a much smaller deviation from the experimental data than the RHF results, and is far less than the experimental uncertainty of these values.

However, we are comparing with the *uncorrected* mean values. A better comparison can then be made by calculating the (frequency dependent) polarizabilities with the frequency corresponding to the wavelength (5893 Å) of the sodium D-line. The frequency dependent polarizabilities from the LB94 potential in the TZ2P++ basis set are on average 2.9 % larger than the static values, which is in good agreement with the estimated 1-4 % from experimental data²²⁰. They are also on average 4.7 % larger than the experimental values, which is a deviation comparable to the experimental uncertainty.

TABLE 4.1.4. MOLECULAR POLARIZABILITIES (BOHR³) USING TZ2P++ BASIS SET

	LDA	BP	LB94	LB94
Acetamide	43.7	42.6	41.2	43.1
Acetylene	23.8	23.5	22.4	23.1
Benzene	70.8	69.6	69.7	72.6
Chlorine	32.0	31.3	29.9	31.7
Cyclohexanol	82.1	79.9	80.7	82.7
Dimethylether	37.2	36.0	35.6	36.3
Formaldehyde	19.1	18.6	17.8	18.3
Hydrogen	5.9	5.5	5.6	5.8
Methylcyanide	31.2	30.5	30.8	31.1
Neopentane	69.5	67.4	68.6	70.3
Propane	43.7	42.3	42.1	44.1
TCFM	60.9	60.0	59.3	60.9
TCMC	74.6	73.3	73.7	75.2
TFM	20.7	20.3	18.6	18.9
Water	10.6	10.3	9.2	9.4
Time	32 hr	53 hr	80 hr	80 hr
Av. dev. (%)	5.84 ± 3.70	3.17 ± 3.55	0.89 ± 3.73	3.79 ± 3.69
Av. abs. dev. (%)	5.84 ± 3.70	3.62 ± 3.09	2.98 ± 2.41	4.65 ± 2.52
Rel. to LB94 (%)	+4.98 ± 3.75	+2.35 ± 3.91		+2.90 ± 1.35
TCFM	trichlorofluoromethane			
TCMC	trichloromethylcyanide			
TFM	trifluoromethane			
LDA	Local Density Approximation result (Vosko-Wilk-Nusair potential ¹¹⁹)			
BP	Becke88 exchange ¹²⁰ with Perdew correlation ¹²¹ potential			
LB94	Van Leeuwen-Baerends (LB94) potential with correct asymptotic behavior ²¹⁶ polarizabilities at 5893 Å			

In principle, both CPHF and TD-DFT should scale as N^3 , where N is the number of basis functions. While the CPHF method (taking about 25 % of the total CPU time) indeed practically scales as N^3 , the TD-DFT method more or less scales linearly and takes approximately 40 % of the CPU time. Apparently the current systems are too small for the N^3 behavior of TD-DFT to become dominant.

Direct Reaction Field

The DRF approach gives polarizabilities with an accuracy of 3.6 %, at a very low computational cost (< 1s) and with high transferability to other molecules¹³⁹. It should be noticed that, like the RHF method, some problems arise with linear and planar molecules, where the anisotropy of the molecular polarizability (i.e. π -bonds) cannot be obtained with interaction tensors based on the atomic positions. Adding extra fit points increases the accuracy of the polarizabilities²⁰², but leads also to bad interaction energies in classical DRF energy calculations¹³⁸.

Conclusions

The DRF approach provides mean polarizability values at low computational cost with an accuracy (3-4 %) equal to experimental uncertainty¹³⁹. Its only setback is the underestimation of the anisotropy of linear and planar molecules, but that is more than compensated with a high transferability to other molecules without the need to reparameterize.

The RHF method gives rather poor mean polarizabilities (deviation 25 % in TZP basis), with high cost for improvement upon these results. Even with the specially constructed (polarized) basis sets, one needs correlated wavefunctions, thereby limiting the applicability to small molecules.

The DFT methods give good mean values with the TZP basis set (6-8 % deviation), and accurate values (dev. 3-6 %) with the TZ2P++ basis set. The latter needs only 3-4 times more CPU time, so it is rather easy to improve the TZP results. It is also evident that the LDA gives larger values than the gradient corrected potentials (+2-5 %), while the Becke-Perdew results are substantially (2.4 %) improved by the Van Leeuwen-Baerends potential.

The impact of the frequency dependency on the polarizabilities is reflected properly by the TD-DFT method which was shown with the Van Leeuwen-Baerends potential. The frequency dependent polarizabilities are on average 2.9 % larger than the static values which is in perfect agreement with the extrapolation estimate of 1-4 %.

The polarizability anisotropy is very well reproduced by all DFT methods, and the DRF method, giving deviations (5-12 %) from experimental values within the experimental uncertainty (6-14 %). The RHF method again shows deviations (28 %) which are large in comparison to experimental deviations.

Accuracy of geometries

The influence of basis sets, xc potentials, core electrons and relativistic corrections on geometries of small molecules and metallocenes

Before one can perform a quantum chemical calculation, one needs to have a structure of the molecule; e.g. one needs to know where the atoms are positioned in space (its geometry). The accuracy with which these geometries can be predicted by different quantum chemical methods is a very useful thing to know, allowing an estimate of the reliability of a computed geometry.

In a recent paper, Helgaker et al.¹⁸⁸ presented a systematic investigation of the accuracy obtainable with wavefunction based methods, applying a hierarchy of basis sets and methods on a set of 19 small closed-shell molecules^a. They used Hartree-Fock (HF), Configuration Interaction (CISD), Møller-Plesset (MP2, MP3, MP4) and coupled cluster methods (CCSD, CCSD(T)) in Dunning's correlation consistent basis sets (cc-pVDZ, cc-pVTZ, cc-pVQZ)²²², and looked at the mean error, standard deviation, mean absolute error and maximum error. Hartree Fock is shown to result always in a too short bond distance, while the inclusion of the electron correlation tends to increase it. Increasing the basis set size tends to decrease the bond distance again. The best results are obtained when one uses the CCSD(T) method, either in the cc-pVTZ or the cc-pVQZ basis, with an accuracy (0.0022 Å) that is comparable to the experimental uncertainty. In another paper, Helgaker et al.²²³ investigated the molecular structure of ferrocene. That study showed clearly the limitations of using the CCSD or CCSD(T) methods. They could not calculate the gradients with these methods, and therefore had to *estimate* the equilibrium Fe-ring distance in ferrocene based on three single-point energy calculations.

In this study, a similar systematic investigation is presented where the accuracy of geometries as predicted by Density Functional Theory is studied. The influence of the basis set, treatment of core electrons and relativistic corrections has been investigated for several currently available exchange-correlation potentials. The same set of 19 small molecules is used in the first part, to enable not only a comparison with experimental data but also with the wavefunction based results obtained by Helgaker et al. For these molecules, the influence of including the core electrons in the optimizations has been investigated. Besides this set of molecules, also a set of metallocene molecules has been used^b, where not only the inclusion of core electrons was investigated, but also the effect of scalar relativistic corrections (with the ZORA^{117,224} hamiltonian). All of these calculations were performed in the standard available basis sets, ranging from a single zeta valence basis set (SZV, I) to a triple zeta valence basis set plus double polarization functions (TZ2P, VII). As B3LYP¹²² is not a *pure* DFT potential, and needs a portion of Hartree-Fock exchange, it can not be used in the ADF program. To check the accuracy of it, as well as to compare the ADF basis sets to the basis sets used in the paper by Helgaker et al.¹⁸⁸, the geometry optimizations were performed also with the B3LYP and BLYP potentials in the cc-pVDZ and cc-pVTZ basis sets using the HONDO98^{204,225} program.

^a The set of small molecules consists of: HF, H₂O, NH₃, CH₄, N₂, CH₂, CO, HCN, CO₂, HNC, C₂H₂, CH₂O, HNO, N₂H₂, O₃, C₂H₄, F₂, HOF, H₂O₂.

^b The metallocene set: manganocene, ferrocene, cobaltocene, nickelocene and ruthenocene.

Computational details

In the ADF program^{117,187}, standard basis sets are available ranging from small (SZV, I) to large (TZ2P, VII), with for the basis sets up to TZ2P (V) the option of either to include (all electron) or exclude (frozen core) the core electrons explicitly in the calculations. In the frozen core basis sets, there are still basis functions assigned to the core electrons; the basis functions of the valence electrons are then explicitly orthogonalized to them. As the calculations are significantly faster when the core electrons are not included, this is normally the preferred option. The following standard available exchange-correlation potentials were examined: Local Density Approximation (LDA), Becke88 exchange¹²⁰ combined with Perdew86 correlation¹²¹ (Becke-Perdew), BLAP3²²⁶, Becke88 exchange¹²⁰ with Lee-Young-Parr correlation²²⁷ (BLYP), Perdew-Burke-Ernzerhof²²⁸ (PBE), Perdew-Wang (PW91)^{189,229}, Revised Perdew-Burke-Ernzerhof²³⁰ (REVPE) and RPBE²³¹.

To compare the basis sets, the cc-pVDZ and cc-pVTZ basis sets were used also for the BLYP potential using the HONDO98 program. Moreover, also the B3LYP potential has been used with this program in the same basis sets to enable a rough comparison with the *pure* DFT xc-potentials.

Scalar relativistic corrections can be included in the calculations quite easily in ADF, using the Zeroth Order Regular Approximation (ZORA)^{117,224}, which is generally found to give an accurate description of the relativistic effects. Although spin-orbit coupling is possible also with the ZORA approach, these effects were not included since the gradients are not implemented yet; therefore it is not possible to optimize the geometries directly.

TABLE 4.2.1. EXPERIMENTAL BOND LENGTHS (pm)

Molecule	Bond	Bond length	Molecule	Bond	Bond length
C ₂ H ₂	CH	106.2	HCN	CH	106.5
C ₂ H ₂	CC	120.3	HCN	CN	115.3
C ₂ H ₄	CH	108.1	HF	HF	91.7
C ₂ H ₄	CC	133.4	HNC	NH	99.4
CH ₂	CH	110.7	HNC	CN	116.9
CH ₂ O	CH	109.9	HNO	NH	106.3
CH ₂ O	CO	120.3	HNO	NO	121.2
CH ₄	CH	108.6	HOF	OH	96.6
CO	CO	112.8	HOF	OF	143.5
CO ₂	CO	116.0	N ₂	NN	109.8
F ₂	FF	141.2	N ₂ H ₂	NH	102.8
H ₂ O	OH	95.7	N ₂ H ₂	NN	125.2
H ₂ O ₂	OH	96.7	NH ₃	NH	101.2
H ₂ O ₂	OO	145.6	O ₃	OO	127.2

Several statistical measures have been used to quantify the accuracy of the methods. The difference between the calculated (R_i^{calc}) and experimental (R_i^{exp}) bond length gives the error Δ_i :

$$\Delta_i = R_i^{calc} - R_i^{exp} \quad (1)$$

For each basis set (if possible, both with and without including the core electrons explicitly) and exchange-correlation potential, the mean error $\bar{\epsilon}$, the standard deviation in the errors ϵ_{std} , the mean absolute error $\bar{\epsilon}_{abs}$ and the maximum error ϵ_{max} were calculated:

$$\bar{\epsilon} = \frac{1}{n} \sum_{i=1}^n \epsilon_i \quad \epsilon_{std} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\epsilon_i - \bar{\epsilon})^2} \quad \bar{\epsilon}_{abs} = \frac{1}{n} \sum_{i=1}^n |\epsilon_i| \quad \epsilon_{max} = \max |\epsilon_i| \quad (2)$$

Each measure characterizes a specific aspect of the performance of the xc-potentials and basis sets. The two first measures characterize the distribution of errors about a mean value $\bar{\epsilon}$ for a given xc-potential in a certain basis set, thus quantifying both systematic and non-systematic errors. The mean absolute error represents the typical magnitude of the errors in the calculations, while the maximum error indicates how large the errors can be.

Small molecules

The experimental values for the bond lengths of the set of small molecules are given in Table 4.2.1, while the mean errors for the eight exchange-correlation potentials in basis sets I to VII are given in Table 4.2.2, both in an all-electron and frozen core basis if available.

TABLE 4.2.2. MEAN ERRORS (pm)

XC-potential	I	II	III	IV	V	VI	VII
<i>frozen core</i>							
Becke-Perdew	5.58	4.00	1.49	1.24	0.90	-	-
BLAP3	6.27	4.16	1.63	1.38	1.02	-	-
BLYP	6.09	4.32	1.82	1.55	1.21	-	-
LDA	4.65	3.26	0.83	0.61	0.28	-	-
PBE	5.44	3.97	1.43	1.19	0.87	-	-
PW91	5.47	3.81	1.31	1.08	0.74	-	-
REVPBE	5.75	4.28	1.68	1.45	1.13	-	-
RPBE	5.76	4.36	1.76	1.54	1.21	-	-
<i>average</i>	<i>5.63</i>	<i>4.02</i>	<i>1.49</i>	<i>1.26</i>	<i>0.92</i>	-	-
<i>all electron</i>							
Becke-Perdew	5.54	3.96	1.44	1.30	0.98	0.81	0.82
BLAP3	6.20	4.09	1.58	1.44	1.09	0.92	0.91
BLYP	6.00	4.24	1.76	1.61	1.26	1.09	1.08
LDA	4.55	3.16	0.74	0.62	0.31	0.17	0.17
PBE	5.36	3.93	1.37	1.29	0.93	0.78	0.78
PW91	5.38	3.76	1.25	1.14	0.82	0.66	0.65
REVPBE	5.65	4.23	1.66	1.52	1.20	1.04	1.04
RPBE	5.76	4.33	1.74	1.61	1.28	1.12	1.12
<i>average</i>	<i>5.56</i>	<i>3.96</i>	<i>1.44</i>	<i>1.32</i>	<i>0.98</i>	<i>0.82</i>	<i>0.82</i>

In all cases, the bond lengths are on average overestimated, with an improvement of the results as the basis set size is increased. For the frozen core basis sets, the average of the

mean errors for the eight xc-potentials improves gradually from 5.63 pm in basis I to 0.92 pm in basis set V. The same trend is observed for the all electron basis sets, which show still an improvement on going from basis set V to VI of some 0.2 pm. However, going from basis set VI to VII, no further improvement is observed, which indicates that the basis set limit has been reached (at least concerning the accuracy of geometries). The difference between the all electron and frozen core basis sets are quite small; the all electron basis sets seem to perform slightly better in the smaller basis sets, while the frozen core seems to give slightly better results in basis sets IV and V.

When looking at the mean errors, LDA seems to give the best performance with a mean error for instance in basis set VI/VII of only 0.17 pm. This value is of the same order of magnitude as was found for CCSD(T). The other xc-potentials are less accurate (~0.6-1.1 pm), but still give results comparable with CCSD in the largest basis set. The worst results are obtained with the BLYP, REVPBE and RPBE potentials, which are the only ones that still give a mean error that is larger than 1 pm in the largest basis set.

The standard deviation in the mean errors is given in Table 4.2.3, which show the same improving trend on increasing the basis set size.

TABLE 4.2.3 STANDARD DEVIATIONS (pm)

XC-potential	I	II	III	IV	V	VI	VII
<i>frozen core</i>							
Becke-Perdew	3.47	3.80	0.94	0.61	0.59	-	-
BLAP ₃	3.51	4.64	1.64	1.35	1.13	-	-
BLYP	3.55	4.32	1.38	1.09	0.89	-	-
LDA	3.59	2.85	1.28	1.21	1.46	-	-
PBE	3.55	3.75	0.89	0.56	0.60	-	-
PW ₉₁	3.45	3.76	0.92	0.57	0.56	-	-
REVPBE	3.49	4.03	0.97	0.61	0.53	-	-
RPBE	3.35	4.07	1.00	0.65	0.52	-	-
<i>average</i>	<i>3.50</i>	<i>3.90</i>	<i>1.13</i>	<i>0.83</i>	<i>0.79</i>	-	-
<i>all electron</i>							
Becke-Perdew	3.43	3.92	1.00	0.63	0.58	0.59	0.59
BLAP ₃	3.48	4.74	1.86	1.45	1.21	1.06	1.09
BLYP	3.53	4.40	1.55	1.16	0.95	0.83	0.85
LDA	3.56	2.94	1.04	1.17	1.39	1.53	1.50
PBE	3.55	3.88	0.92	0.63	0.57	0.61	0.61
PW ₉₁	3.44	3.87	0.96	0.62	0.57	0.57	0.58
REVPBE	3.33	4.15	1.11	0.67	0.55	0.53	0.53
RPBE	3.39	4.22	1.16	0.71	0.55	0.51	0.53
<i>average</i>	<i>3.46</i>	<i>4.02</i>	<i>1.20</i>	<i>0.88</i>	<i>0.80</i>	<i>0.78</i>	<i>0.79</i>

Also here do the frozen core and all electron basis sets perform equally well. There are some small differences, but not as large as the ones observed between basis sets of different size; i.e., whereas the former differences are of the order of 0.1 pm (comparing for instance Becke-Perdew in basis set IV), the latter are a few times larger (comparing for instance the frozen core Becke-Perdew results in basis set III and IV respectively).

LDA no longer performs best in this respect; although it had the smallest mean error of all potentials, the standard deviation belonging to it is the largest (~1.5 pm in basis sets V-VII) of all xc-potentials in the larger basis sets; in the medium basis sets, also a large standard deviation is found for the BLAP₃ potential. In fact, for the larger basis sets, all potentials give more or less equal standard deviations except three: BLAP₃, BLYP and LDA. As BLYP was also one of the worst potentials in the case of mean errors, it seems that it can not be used with great confidence for obtaining accurate geometries.

The mean absolute errors are given in Table 4.2.4. Just like the mean errors and the standard deviations, they are shown to improve gradually with increasing basis set size.

TABLE 4.2.4 MEAN ABSOLUTE ERRORS (pm)

XC-potential	I	II	III	IV	V	VI	VII
<i>frozen core</i>							
Becke-Perdew	5.68	4.00	1.51	1.24	0.93	-	-
BLAP ₃	6.27	4.17	1.64	1.38	1.03	-	-
BLYP	6.13	4.32	1.82	1.55	1.21	-	-
LDA	5.05	3.28	1.21	1.14	1.28	-	-
PBE	5.61	3.97	1.45	1.20	0.92	-	-
PW91	5.57	3.82	1.34	1.09	0.80	-	-
REVPBE	5.87	4.28	1.68	1.45	1.13	-	-
RPBE	5.78	4.36	1.76	1.54	1.21	-	-
<i>average</i>	<i>5.75</i>	<i>4.03</i>	<i>1.55</i>	<i>1.32</i>	<i>1.06</i>	-	-
<i>all electron</i>							
Becke-Perdew	5.64	3.98	1.45	1.30	1.00	0.88	0.89
BLAP ₃	6.20	4.13	1.59	1.44	1.09	0.93	0.93
BLYP	6.06	4.26	1.76	1.61	1.26	1.09	1.09
LDA	4.97	3.20	1.00	1.11	1.24	1.31	1.29
PBE	5.55	3.95	1.38	1.29	0.97	0.86	0.86
PW91	5.49	3.79	1.27	1.15	0.87	0.76	0.75
REVPBE	5.68	4.25	1.66	1.52	1.20	1.05	1.05
RPBE	5.78	4.35	1.74	1.61	1.28	1.12	1.12
<i>average</i>	<i>5.67</i>	<i>3.99</i>	<i>1.48</i>	<i>1.38</i>	<i>1.11</i>	<i>1.00</i>	<i>1.00</i>

Again there is hardly any difference between the results from either a frozen core or an all electron basis set; the difference between the averages of the frozen core and the all electron in basis set IV is for instance 0.06 pm, while the difference between the averages between basis set III and IV for the frozen core is 0.23 pm. The results of basis set V are still improved on going to VI, but increasing the basis set even more (VII) doesn't improve them any more. So, where the accuracy of geometries is concerned, the basis set limit is reached already in basis set VI.

The Becke-Perdew, PBE and PW91 are found to be the best xc-potentials, as they are the only ones that are consistently giving better results than the average for all basis sets, both in the frozen core and the all electron basis sets. The mean error in the basis set limit of 0.89 (Becke-Perdew), 0.86 (PBE) and 0.75 (PW91) pm, are comparable or slightly better than CCSD results (0.89 pm) in the largest basis set used by Helgaker et al. (cc-pVQZ). Although LDA

was performing best for the mean error, in this case it is performing the worst with an average error of ~ 1.3 pm in the larger basis sets. In fact, for the all electron basis sets, LDA does not improve upon increasing the basis set size after basis set III.

The maximum errors for the xc-potentials in a certain (frozen core/all electron) basis set are given in Table 4.2.5. These results don't show the gradual improvement as the basis set is increased; for instance, going from basis I to basis II the average maximum error increases from 10.7 to 13.4 pm. However, after increasing the basis set even more, it decreases again gradually, except for LDA that exhibits a oscillatory pattern. As the difference between the VI and VII basis sets is small, also these results show that the basis set limit has been reached.

TABLE 4.2.5 MAXIMUM ERRORS (pm)

XC-potential	I	II	III	IV	V	VI	VII
<i>frozen core</i>							
Becke-Perdew	10.52	13.16	3.35	2.51	1.81	-	-
BLAP3	11.96	15.61	6.15	5.44	4.68	-	-
BLYP	11.32	14.91	5.43	4.73	3.95	-	-
LDA	9.88	9.69	3.30	2.32	2.85	-	-
PBE	10.42	12.97	3.31	2.11	1.85	-	-
PW91	10.43	12.83	3.23	2.13	1.64	-	-
REVPBE	10.60	13.90	3.89	2.96	2.11	-	-
RPBE	10.58	14.19	4.21	3.25	2.40	-	-
<i>average</i>	<i>10.71</i>	<i>13.41</i>	<i>4.11</i>	<i>3.18</i>	<i>2.66</i>	-	-
<i>all electron</i>							
Becke-Perdew	10.42	13.42	3.98	2.68	1.82	1.75	1.75
BLAP3	11.82	15.88	6.72	5.36	4.66	4.10	4.18
BLYP	11.19	15.14	5.91	4.66	3.93	3.38	3.44
LDA	9.73	9.89	2.73	2.29	2.26	2.75	2.72
PBE	10.32	13.25	3.56	2.57	1.86	1.80	1.79
PW91	10.31	13.12	3.54	2.41	1.75	1.59	1.57
REVPBE	10.51	14.23	4.69	3.31	2.37	1.89	1.87
RPBE	10.73	14.53	5.03	3.63	2.49	2.00	2.22
<i>average</i>	<i>10.63</i>	<i>13.68</i>	<i>4.52</i>	<i>3.36</i>	<i>2.64</i>	<i>2.41</i>	<i>2.44</i>

The difference between the frozen core and all electron basis sets is not negligible for the maximum error, at least not for the small and medium sized basis sets. For basis set V, the difference is negligible again for some potentials (like Becke-Perdew, BLAP3 or PBE), while for others (LDA, REVPBE) there exists a large difference.

Just like found for the mean absolute errors, there are three potentials (Becke-Perdew, PBE and PW91) that are consistently giving better results than the average for a certain basis set. For these three, the maximum error in the basis set limit (1.75, 1.79 and 1.57 pm respectively) is comparable to MP2 (1.67 pm) and MP4 (1.48 pm) in the largest basis set studied by Helgaker et al., and is substantially better than the CCSD method in the same basis (3.07 pm).

Comparison with Dunning's basis set and the B3LYP potential

The results obtained with Dunning's basis sets for either the BLYP or the B3LYP potential are given in Table 4.2.6. Comparing the BLYP results in the cc-pVDZ basis set (e.g. [3S2P1D] for carbon) with its ADF counterpart (basis III, [4S2P1D]), the mean absolute error of cc-pVDZ is slightly larger than basis set III. However, the maximum error is much smaller for cc-pVDZ than for basis III. The same inconsistencies show up for the mean error and its standard deviation; although the mean error is more or less comparable (1.76 vs. 1.81 pm), the standard deviations are quite different (1.55 vs. 1.06 pm). Therefore, as BLYP does not perform equally well in the cc-pVDZ and the ADF-III basis set, also the B3LYP results (in the cc-pVDZ basis) can not directly compared with the *pure* xc-potentials in basis III.

The size of the cc-pVTZ basis set ([4S3P2D1F] for carbon) is more or less inbetween that of basis set V ([5S3P1D1F]) and VI ([6S4P2D1F]) of ADF. This is reflected in the mean absolute error of the BLYP potential in the cc-pVTZ basis set (1.14 pm), which is somewhere inbetween the results in the V (1.26 pm) and VI (1.09 pm) basis set. Also the maximum error shows this trend: 3.55 (cc-pVTZ), 3.93 (V), 3.38 (VI) pm. There is however a difference for the mean error, which is 0.98 pm in the cc-pVTZ basis, but slightly larger in the ADF basis sets (1.26 and 1.09 pm for respectively basis V and VI); the standard deviation on the other hand is slightly lower in the ADF basis sets: 0.95 (V) and 0.83 (VI) pm vs. 0.99 pm (cc-pVTZ). However, roughly speaking, the cc-pVTZ results for BLYP are similar to those in the V/VI basis sets, which enables a comparison of B3LYP with the *pure* DFT potentials in ADF.

The mean error of B3LYP in the cc-pVTZ basis is almost zero (0.05 pm); however, the standard deviation is almost twice as large as those of Becke-Perdew or PW91. The mean absolute error is either comparable (PW91) or slightly better (Becke-Perdew, PBE) for the B3LYP potential. Finally, the maximum error is slightly larger for B3LYP compared with either Becke-Perdew, PBE or PW91. As a whole, B3LYP performs equally well as the most accurate *pure* DFT potentials (Becke-Perdew, PBE, PW91), although with a slightly increased maximum error and standard deviation of the mean error.

TABLE 4.2.6 BLYP AND B3LYP RESULTS (pm) IN DUNNING'S BASIS SETS

	BLYP cc-pVDZ	BLYP cc-pVDZ	BLYP cc-pVDZ	BLYP cc-pVDZ
<i>mean error</i>	1.81	0.98	0.66	0.05
<i>standard deviation</i>	1.06	0.99	0.87	0.94
<i>mean absolute error</i>	1.90	1.14	0.96	0.73
<i>maximum error</i>	3.80	3.55	2.06	1.93

Metalloenes

Metalloenes are molecules where a metal atom is sandwiched between two cyclopentadiene rings, which can exist in two conformations: *staggered* or *eclipsed* (see Figure 4.2.1).

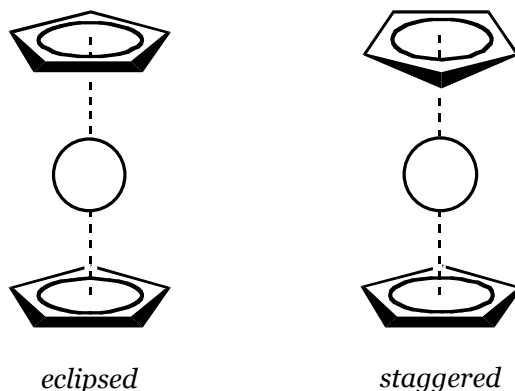


FIGURE 4.2.1. STRUCTURE OF METALLOENES

The best known example of these metalloenes is ferrocene, which has been studied in great detail in the past by theoretical methods, and showed to be a difficult molecule for which an accurate prediction of the metal-ring distance could be obtained. Early Hartree-Fock calculations reported a Fe-ring distance of 1.88 Å, which is in poor agreement with the experimental distance of 1.66 Å. This could not be improved by employing larger basis sets, as it was established²³² that the Hartree-Fock limit is only slightly better than the early HF calculation (1.872 Å). Normally, such a poor performance could be improved upon by using MP2 calculations; for instance, the systematic study by Helgaker et al. (described earlier in this section) showed a mean absolute error of 13.0 pm for Hartree-Fock, and only 2.4 pm for MP2. However, MP2 results on ferrocene²³² show a similar dramatic performance. However, unlike HF that overestimates the Fe-ring distance, MP2 underestimates it at 1.47-1.49 Å²³² (depending on the number of electrons correlated). Calculations employing the CASSCF and CASPT2 method perform much better in this respect with Fe-ring distances of 1.716 and 1.617 Å respectively²³³. Subsequently, correcting the results for Basis Set Superposition Errors, an estimated equilibrium value of 1.643 Å was obtained, which is in good agreement with the experimental value of 1.66 Å. Also CCSD and CCSD(T) calculations predict an equilibrium distance which is in good agreement with the experimental data, with respectively 1.672 Å and 1.660 Å²²³. Note however, that the coupled cluster distances were obtained by performing single point energies at three Fe-ring distances, and then fitting the potential energy curve with a second order polynomial to obtain the equilibrium distance.

The metalloenes studied in this section are manganocene (either doublet or sextet), ferrocene (singlet), cobaltocene (doublet or quartet), nickelocene (singlet or triplet) and ruthenocene (singlet). The geometry data from experimental investigations (taken from ref. 234) are given in Table 4.2.7.

TABLE 4.2.7 EXPERIMENTAL GEOMETRIES METALLOCENES (Å)

<i>metal M</i>	<i>r(C-H)</i>	<i>r(C-C)</i>	<i>r(M-ring)</i>
Mn (<i>doublet</i>)	-	1.418	1.720
Mn (<i>sextet</i>)	1.125	1.429	2.041
Fe	1.104	1.440	1.661
Co	1.095	1.430	1.722
Ni	1.083	1.430	1.823
Ru	1.130	1.439	1.823

The distance of the metal to the center of the cyclopentadienyl rings differs quite some amount for the five metal atoms as well as the multiplet state of the complex. For instance for the manganocene, a difference of more than 0.32 Å in the distance is observed between the doublet and sextet state.

The geometry of the metallocenes was optimized using all exchange-correlation potentials that had been used already in the first part of this section, employing a few different basis set sizes, ranging from a minimal basis (I) to a triple zeta valence basis plus double polarization functions (VI). The core electrons were either taken explicitly into account (*all electrons*) or frozen in the calculations; in the latter calculations, the frozen core electrons comprise the 1s electrons of carbon and for the first row transition metals up to either 3p (*frozen core 3p*; for Ru up to 4p) or up to the 2p level (*frozen core 2p*; for Ru up to 3d). For all three types, the optimizations were done in either a non-relativistic or a scalar relativistic (ZORA) calculation. The mean absolute errors of the computed distances for the eight xc-potentials in the four basis sets are given in Tables 4.2.8 (non-relativistic) and 4.2.9 (scalar relativistic).

A general improvement of the accuracy is shown by increasing the basis set size, which is most obvious for the *all electron* calculations. The mean absolute error, averaged over the eight xc-potentials, decreases from 7.70 pm to 1.77 pm. However, there are some potentials that do not show this pattern; the error of the LDA potential for instance decreases in going from basis set I (minimal basis) to II (double zeta valence), and increases if one uses larger basis sets. The same pattern had already been observed in the first part of this section, but there it emerged only after basis set III (double zeta valence plus polarization) and the error in basis set IV (triple zeta valence plus polarization) was still smaller than in basis II. Here, the LDA error is larger in basis IV than in basis II, both if one uses the *frozen core* or the *all electron* calculations.

The difference between the *frozen core 3p*, *frozen core 2p* and *all electron* results is not negligible. For the minimal basis (I), the *frozen core 3p* gives a better performance than the *all electron* calculations; for the larger basis sets, the situation is reversed. In basis II, the best performance is observed with the *frozen core 2p* option, which gives a mean absolute error that is 0.2-0.3 pm smaller than that for the *all electron* case. In basis set IV, these two options perform equally well, and considerably better (0.8-0.9 pm) than the *frozen core 3p* option.

TABLE 4.2.8. MEAN ABSOLUTE NR^a ERRORS (pm) FOR 1ST ROW METALLOCENES

XC-potential	I	II	IV	VI
<i>frozen core 3p</i>				
Becke-Perdew	7.43	2.77	2.25	-
BLAP ₃	6.88	5.66	4.99	-
BLYP	7.36	4.52	3.92	-
LDA	8.55	1.23	1.55	-
PBE	7.50	2.52	2.02	-
PW ₉₁	7.57	2.53	2.12	-
REVPBE	6.58	3.29	2.61	-
RPBE	7.00	3.53	2.81	-
<i>average</i>	<i>7.36</i>	<i>3.26</i>	<i>2.78</i>	-
<i>frozen core 2p</i>				
Becke-Perdew	-	1.77	1.39	-
BLAP ₃	-	4.58	4.05	-
BLYP	-	3.49	2.99	-
LDA	-	2.01	2.46	-
PBE	-	1.60	1.29	-
PW ₉₁	-	1.60	1.33	-
REVPBE	-	2.25	1.72	-
RPBE	-	2.49	1.92	-
<i>average</i>	-	<i>2.47</i>	<i>2.14</i>	-
<i>all electron</i>				
Becke-Perdew	7.68	2.05	1.39	1.00
BLAP ₃	7.21	4.92	4.08	3.18
BLYP	7.69	3.78	3.05	2.17
LDA	8.81	1.87	2.47	3.05
PBE	7.83	1.82	1.28	1.11
PW ₉₁	7.82	1.79	1.34	1.13
REVPBE	7.30	2.53	1.72	1.18
RPBE	7.23	2.78	1.93	1.30
<i>average</i>	<i>7.70</i>	<i>2.69</i>	<i>2.16</i>	<i>1.77</i>

a) NR=non-relativistic results

The “best” exchange-correlation potentials are for the larger basis sets and the *all electron/frozen core 2p* options, Becke-Perdew, PBE and PW₉₁, just like found in the first part of this section. For the *frozen core 3p* option in the larger basis sets, LDA performs best; in basis II, the mean absolute error is even only 1.23 pm, a value not reached by any other potential in any basis set with the *frozen core 3p* option. However, using the *frozen core 2p* or the *all electron* option, this value is reached in the larger basis sets (IV, VI) by other xc-potentials. In the minimal basis, the BLAP₃ and REVPBE give generally the “best” performance, but as the mean absolute error is about four times as large as the value in the larger basis sets, it is of limited value.

TABLE 4.2.9. MEAN ABSOLUTE SR^a ERRORS (pm) FOR 1ST ROW METALLOCENES

XC-potential	I	II	IV	VI
<i>frozen core 3p</i>				
Becke-Perdew	-	2.47	2.01	-
BLAP3	-	5.31	4.71	-
BLYP	-	4.20	3.66	-
LDA	-	1.38	1.77	-
PBE	-	2.22	1.78	-
PW91	-	2.26	1.88	-
REVPBE	-	2.97	2.35	-
RPBE	-	3.22	2.56	-
<i>average</i>	-	3.00	2.59	-
<i>frozen core 2p</i>				
Becke-Perdew	-	1.54	1.18	-
BLAP3	-	4.21	3.66	-
BLYP	-	3.14	2.63	-
LDA	-	2.31	2.76	-
PBE	-	1.40	1.07	-
PW91	-	1.41	1.11	-
REVPBE	-	1.93	1.44	-
RPBE	-	2.12	1.59	-
<i>average</i>	-	2.26	1.93	-
<i>all electron</i>				
Becke-Perdew	7.83	1.72	1.19	1.22
BLAP3	7.37	4.53	3.69	2.74
BLYP	7.85	3.43	2.67	1.76
LDA	8.97	2.18	2.80	3.42
PBE	7.92	1.58	1.06	1.38
PW91	7.97	1.59	1.15	1.39
REVPBE	7.40	2.17	1.45	1.08
RPBE	7.33	2.39	1.59	1.12
<i>average</i>	7.83	2.45	1.95	1.76

a) SR=scalar-relativistic results using ZORA approach

The influence of the scalar relativistic corrections is, apart from the *all electron* calculations in basis VI, small (0.2 pm) but improving. This effect is of the same order of magnitude as the effect observed due to a slight mismatch between the energy expression and the potential in the ZORA approach, which leads to an optimized geometry with zero gradient that may differ from the point of lowest energy by some 0.1 pm. Therefore, the effect of the relativistic corrections can safely be ignored.

The *all electron* results in basis VI show different patterns for different xc-potentials. For some, like Becke-Perdew, PBE and PW91, the mean absolute error increases relative to the non-relativistic results, while for others, like LDA, the error decreases. Still, these changes are of the same small magnitude (0.2 pm) and therefore not significant.

TABLE 4.2.10. MEAN ABSOLUTE SR/NR^a ERRORS (pm) FOR RU-RING DISTANCE

XC-potential	II/NR	IV/NR	II/SR	IV/SR	V/SR
<i>frozen core 3p</i>					
Becke-Perdew	9.01	6.69	6.84	4.54	-
BLAP3	15.64	13.43	13.02	10.89	-
BLYP	13.44	11.36	10.90	8.91	-
LDA	3.52	1.60	1.68	0.41	-
PBE	8.18	5.75	6.13	3.63	-
PW91	8.61	6.12	6.18	3.92	-
REVPBE	9.30	6.79	7.20	4.68	-
RPBE	9.72	7.17	7.63	5.04	-
<i>average</i>	<i>9.68</i>	<i>7.36</i>	<i>7.45</i>	<i>5.25</i>	-
<i>frozen core 2p</i>					
Becke-Perdew	5.59	3.57	3.99	1.82	-
BLAP3	11.90	9.91	9.89	7.77	-
BLYP	9.84	8.07	7.91	5.99	-
LDA	0.39	1.28	1.02	2.96	-
PBE	4.82	2.73	3.36	0.99	-
PW91	5.12	3.21	3.54	1.32	-
REVPBE	5.87	3.62	4.29	1.91	-
RPBE	6.25	3.97	4.65	2.25	-
<i>average</i>	<i>6.22</i>	<i>4.55</i>	<i>4.83</i>	<i>3.13</i>	-
<i>all electron</i>					
Becke-Perdew	-	-	4.82	2.47	1.18
BLAP3	-	-	10.83	8.64	7.42
BLYP	-	-	8.81	6.77	5.52
LDA	-	-	0.27	2.49	3.83
PBE	-	-	4.16	1.64	0.33
PW91	-	-	4.22	2.10	0.60
REVPBE	-	-	5.17	2.62	1.29
RPBE	-	-	5.56	2.99	1.65
<i>average</i>	-	-	<i>5.48</i>	<i>3.72</i>	<i>2.73</i>

a) NR=non-relativistic; SR=scalar-relativistic results using ZORA approach

For the first row transition metals, the effect of the inclusion of relativistic corrections is small, but for ruthenocene, it probably can no longer be safely ignored. As the relativistic corrections will have an effect mainly on the Ru-ring distance and not as much on the C-C or C-H distances, only the Ru-ring distance is taken into account for the mean absolute error. The errors for the xc-potentials in several basis sets and with different options for the core electrons are given in Table 4.2.10. As expected, the relativistic corrections now do have a significant effect on the values of the mean absolute error, which are improved by an amount of 1.4-2.2 pm. The same trends are observed as before, e.g. generally speaking the error decreases with increasing basis set size (apart from the LDA potential), the *frozen core 2p* option performs better than the *frozen core 3p* option (except for the LDA potential), the BLAP3 and BLYP perform significantly less than the other potentials.

The multiplet states of cobaltocene, nickelocene and manganocene lead in all cases to the same ground state, respectively a doublet (cobaltocene), triplet (nickelocene) and doublet (manganocene). The other states are less favored by an amount of 21 kcal/mol (quartet cobaltocene), 14 kcal/mol (singlet nickelocene) and 14 kcal/mol (sextet manganocene) with the Becke-Perdew xc-potential in basis set IV using the *all electron* option. These values differ by a few kcal/mol with different basis sets and/or xc-potentials, but the relative ordering of the multiplet states does not.

The relative ordering of the conformations of the metallocenes is also rather constant over the range of xc-potentials and basis sets; in nearly all cases the eclipsed conformation is favored. The energy difference between the two conformations however differs between the different metallocenes, respectively 0.4 kcal/mol (doublet cobaltocene), 0.1 kcal/mol (quartet cobaltocene), 1.0 kcal/mol (ferrocene), 1.2 kcal/mol (doublet manganocene), 0.01 kcal/mol (sextet manganocene), 0.1 kcal/mol (singlet or triplet nickelocene) and 0.5 kcal/mol (ruthenocene), using the Becke-Perdew xc-potential in basis set IV. For quartet cobaltocene, sextet manganocene and nickelocene, the energy difference is too small to be considered significant, and both conformations are equally favorable. Again, these values change somewhat by using another xc-potential and/or basis set, but the relative ordering of the conformations remains roughly intact.

Conclusions

The geometries of a set of small molecules were optimized using eight different exchange-correlation potential in a few different basis sets of Slater type orbitals, ranging from a minimal basis (I) to a triple zeta valence basis plus double polarization functions (VI). This enables a comparison of the accuracy of the xc-potentials in a certain basis set, which can be related to the accuracies of wavefunction based methods like Hartree-Fock and Coupled Cluster. Four different checks are done on the accuracy by looking at the mean error, standard deviation, mean absolute error and maximum error. It is shown that the mean absolute error decreases with increasing basis set size, and reaches a basis set limit at basis VI. With this basis set, the mean absolute errors of the xc-potentials are of the order of 0.7-1.3 pm, which is comparable to the accuracy obtained with CCSD and MP2/MP3 methods.

In the second part of this section, the geometry of five metallocenes is optimized with the same potentials and basis sets, either in a non-relativistic or a scalar relativistic calculation using the ZORA approach. For the first row transition metal complexes, the relativistic corrections have a negligible effect on the optimized structures, but for ruthenocene they improve the optimized Ru-ring distance by some 1.4-2.2 pm. In the largest basis set used, the absolute mean error is again of the order of 1.0 pm. As the wavefunction based methods either give a poor performance for metallocenes (Hartree-Fock, MP2), or the size of the system makes a treatment with accurate methods like CCSD(T) in a reasonable basis set cumbersome, the good performance of Density Functional Theory calculations for these molecules is very promising. Even more so as DFT is an efficient method that can be used without problems on system sizes of this kind, or larger.

Validation of charge analyses

The concepts of molecular recognition, electron withdrawing/donating groups and electrophilic substitution reactions

Assigning atomic charges in quantum chemical calculations has mainly two purposes: either they should be used in a subsequent classical mechanics calculation, or they should be used as a simple interpretation of the distribution of charge density within the molecule. An important example in both cases is the field of molecular recognition, i.e. how do molecules “feel” the presence of other molecules. The largest contribution to the way molecules “feel” each other is given by electrostatic interactions between the permanent multipole moments of the molecules. Therefore, it is important that the atomic charges give a good representation of the permanent multipole moments of the molecule. Furthermore, the atomic charges should be able to reproduce “chemical intuition”, like for instance with the concept of electron withdrawing/donating groups.

In this section, a few standard charge analyses that are available in the ADF program^{117,187} are tested on their ability to give a good description for both of the aforementioned purposes. The tested analyses are the Mulliken²³⁵ charge analysis, the Hirshfeld analysis²³⁶, the Voronoi Deformation Density analysis¹¹⁷ and the Multipole Derived Charge analysis¹⁸³ (described in more detail in Section 3.1). The charge analyses are tested on benzene and some benzene derivatives^a, which enables to check the concept of electron withdrawing/donating groups, as well as to check electrophilic substitution reactions.

All molecular multipole moments reported in this section are obtained relative to the center of mass of the molecule.

Molecular recognition

The electrostatic potential in a point r_s due to the charge density $\rho(r_i)$ of a molecule is obtained as:

$$V_{el.st.}(r_s) = \int \frac{\rho(r_i)}{|r_s - r_i|} dr_i \quad (1)$$

This can be expanded in a molecular multipole expansion within the Buckingham convention^{179,180} as:

$$V_{el.st.}(r_s) = \frac{Q}{r} + \frac{\mu \cdot \mathbf{r}}{r^3} + \frac{1}{2} \frac{\mathbf{r} \cdot \Theta \cdot \mathbf{r}}{r^5} + \dots \quad (2)$$

with the molecular charge Q , dipole moment μ , quadrupole moment Θ , and distance vector \mathbf{r} with length r . For normal purposes, the expansion can be cut off after the quadrupole

^a The substituent groups used for the benzene derivatives are: trifluoromethyl, cyanide, aldehyde, chloride, fluoride, methyl, NH₂, NO₂, OCOH, OH, OMethyl, O-minus and SH₂.

moment term, as the contribution to the electrostatic potential is decreasing with increasing multipole level due to the division by r to the power x , where x increases with increasing multipole level.

For all analyses studied here, except the Multipole Derived Charge analysis¹⁸³, there is no constraint on the charge analysis to reproduce any of the molecular multipole moments other than the total charge. This leads to rather large absolute deviations between the *expectation value* of the molecular multipole moments of the benzene derivatives, and the multipole moments obtained from either the Mulliken, Hirshfeld or Voronoi charges (see Table 4.3.1). The dipole moments deviate on average some 0.2-0.3 a.u., while it is even worse for the quadrupole moments. The Mulliken analysis is in both cases giving the largest deviation, with even an average absolute deviation for the xx , yy and zz components of the quadrupole moment of 10.3 atomic units.

The same trend is obtained for benzene, where the first non-zero multipole moment is the quadrupole moment. The expectation value for its zz -component is -5.568 , which is well reproduced by the Multipole Derived Charge analysis¹⁸³ with a value of -5.595 a.u.; the Hirshfeld and Voronoi analyses perform less with values of -2.026 and -2.272 respectively. The Mulliken analysis again performs rather poorly, and even predicts a quadrupole moment with the wrong sign: $+8.883$ a.u.

TABLE 4.3.1. AVERAGE ABSOLUTE DEVIATIONS (A.U.) FROM THE EXPECTATION VALUE OF MOLECULAR MULTIPOLES OF BENZENE AND BENZENE DERIVATIVES

analysis	charge	dipole	quadrupole	quad. $xx/yy/zz$
Mulliken	0.0	0.30	5.27	10.29
Hirshfeld	0.0	0.26	1.39	2.56
Voronoi	0.0	0.23	1.24	2.28
MDC	0.0	0.03	0.30	0.49

Although also the charges from the Multipole Derived Charge analysis¹⁸³ give molecular multipoles that deviate to a small extent from the expectation value of the multipole moments, there is a difference in the reason why these deviations occur. Unlike the other analyses, the molecular multipole moments are *by construction* constrained; however, the analysis uses the moments from the *fitted* density, not from the “*exact*” density (see Section 3.1). There are some small differences between the two sets of multipole moments, which results in the deviations reported in Table 4.3.1.

Electron withdrawing or donating effects

Atoms and functional groups can be organized according to their ability to donate or withdraw electrons either by *inductive* or *resonance* effects²³⁷ (see Figure 4.3.1). For many substituent groups, both effects work in the same direction, but there are some groups (like

the halogen atoms) that exhibit opposing effects for the two cases. These abilities to donate or withdraw electrons have a marked effect on reaction rates for several organic reactions.

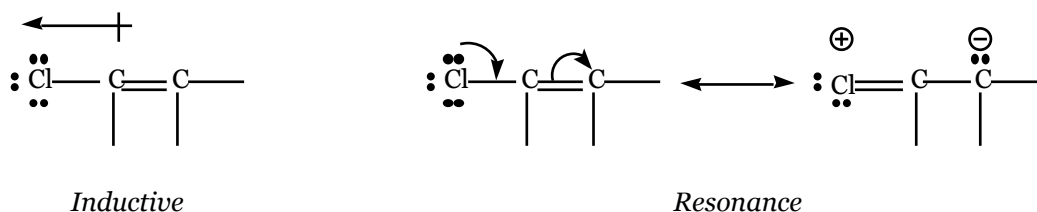


FIGURE 4.3.1. INDUCTIVE AND RESONANCE SUBSTITUENT EFFECTS

In the last part of this section, a typical organic reaction that shows a large dependence on substituent effects will be considered. The substituent effects found can either be due to (de)stabilization of the ground state or (de)stabilization of the transition state of this reaction. This part of the section only deals with the inductive effects, as these are the ones involved in the ground state properties of the molecules. In the last section, also the resonance effects will be taken into account.

In Table 4.3.2, the substituent effects on the charge distribution is presented as predicted by the four charge analyses. These effects are obtained by taking the difference between the charge of a hydrogen atom in benzene and the total charge of the substituent groups, which are corrected for the change of total charge of the molecule (as in the case of O^- and SH_2^+). A negative value indicates therefore an electron-withdrawing effect, a positive value an electron-donating effect.

TABLE 4.3.2. INDUCTIVE SUBSTITUENT EFFECTS

<i>substituent</i>	<i>Mulliken</i>	<i>Hirshfeld</i>	<i>Voronoi</i>	<i>MDC</i>
CF ₃	+0.110	-0.091	-0.088	-0.170
CN	+0.014	-0.191	-0.204	-0.273
COH	+0.046	-0.129	-0.122	-0.170
Cl	-0.043	-0.079	-0.124	-0.194
F	-0.453	-0.138	-0.129	-0.285
Me	+0.073	-0.031	-0.010	+0.058
NH ₂	+0.095	+0.019	+0.028	-0.245
NO ₂	-0.036	-0.198	-0.203	-0.378
OCOH	-0.161	-0.143	-0.156	-0.354
OH	-0.074	-0.063	-0.044	-0.204
OMe	-0.057	-0.047	-0.042	-0.240
O ⁻	+0.421	+0.485	+0.432	+0.140
SH ₂ ⁺	-0.313	-0.389	-0.429	-0.520

Experimentally²³⁷, electron-donating inductive effects are found for O- and the methyl group, while the other substituents are found to have an electron-withdrawing effect.

The Mulliken analysis does not perform completely reliable in this respect. It predicts for instance electron-donating effects for CF_3 , CN , COH , NH_2 , while this should be a withdrawing effect. However, for the other groups, the correct effect is predicted. The Hirshfeld and Voronoi analyses perform equally here, in that they both predict the correct effects except for two cases: methyl and NH_2 . For the first a donating effect is predicted and for the second a withdrawing, while in both cases it should be opposite. Generally speaking though, they perform better than the Mulliken analysis, but not as good as the Multipole Derived Charge analysis¹⁸³. This analysis gives a good prediction for the substituent effects for all cases, including the cases where the other analyses fail.

Electrophilic substitution reactions

In electrophilic aromatic substitution reactions, substituent groups have two effects. They have an effect on the reaction rate as well as on the position of the substitution. This is because the reaction may proceed at either the *ortho*, *meta* or *para* position (see Figure 4.3.2).

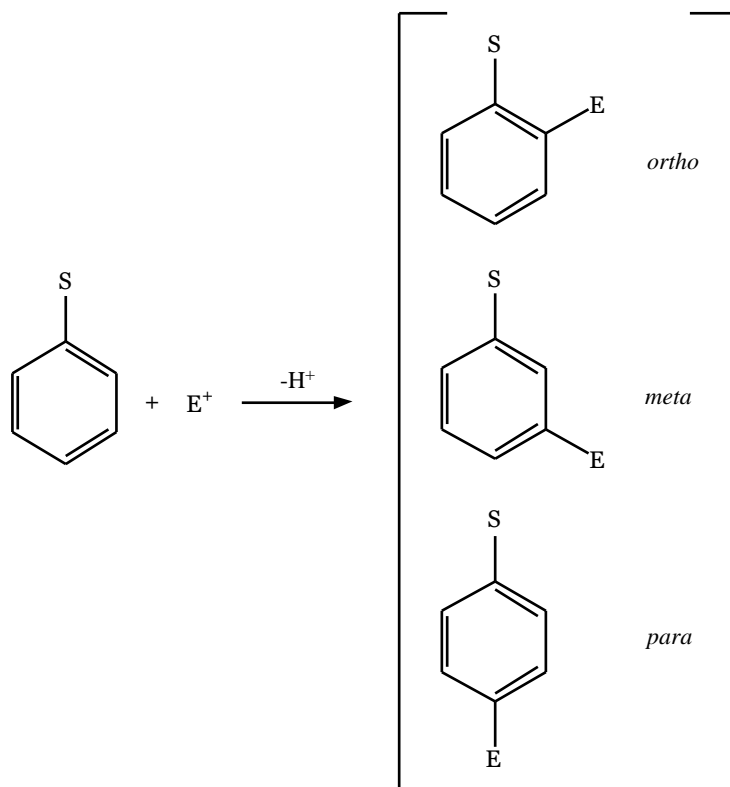


FIGURE 4.3.2. ELECTROPHILIC AROMATIC SUBSTITUTION REACTION

Upon bonding of the electrophile E^+ to the benzene ring, a cyclohexadienyl cation is formed that is stabilized by resonance (as indicated for the *ortho* position in Figure 4.3.3). For both the *ortho* and *para* positions, a resonance structure is available where the positive charge is adjacent to the substituent (Figure 4.3.3, structure on the right), while the *meta* intermediate does not have such a structure. Therefore, if the substituent is electron-donating through resonance, it will stabilize the cation intermediate and enhance the

reaction. Because of the delocalization of positive charge between the ring and substituent in the case of *ortho* or *para* addition, this substituent leads primarily to *ortho* and *para* substitution products. On the other hand, if the substituent is electron-withdrawing through resonance, the reaction will proceed slower and the product will be primarily *meta* oriented.

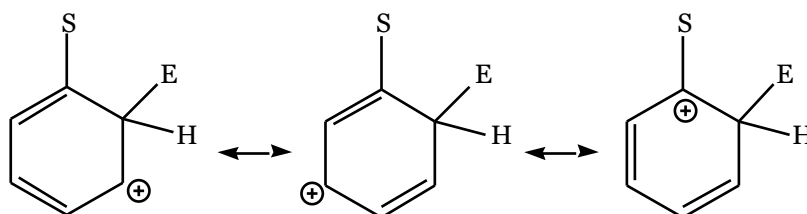


FIGURE 4.3.3. INTERMEDIATE CATION FORMED

An exception to these simple pictures is given by the halogen atoms. Chlorobenzene undergoes nitration about $\frac{1}{30}$ as fast as benzene, but the orientation is *ortho*-*para*. Here, the difference between *inductive* and *resonance* electron withdrawing/donating effects play a role. The *inductive* electron-withdrawing effect slows down the reaction, yet the *resonance* electron-donating effect favors the *ortho/para* oriented products.

TABLE 4.3.3. CHANGE IN CHARGE OF O, M AND P-POSITIONED ATOMS (0.01 A.U.)

	a	b	c	d	e	f	g	h	i	j	k	l	m
<i>Mulliken</i>													
<i>o</i>	5.0	2.6	3.2	-1.6	-6.1	2.3	-3.5	2.8	-1.1	-6.3	-6.4	-11.1	1.5
<i>m</i>	-0.6	-0.4	-0.9	-0.4	0.1	-0.8	-0.8	-0.7	-0.4	-0.6	-0.9	-3.3	2.9
<i>p</i>	1.6	2.3	2.1	0.5	-0.8	-0.4	-2.9	2.5	-0.1	-1.6	-1.8	-7.2	3.3
<i>Hirshfeld</i>													
<i>o</i>	0.7	1.8	1.7	-1.0	-1.5	-0.5	-2.7	1.3	-0.6	-2.2	-2.3	-7.1	1.7
<i>m</i>	0.9	1.1	0.7	0.5	0.6	-0.2	-0.5	1.3	0.6	0.1	-0.2	-5.1	4.4
<i>p</i>	1.3	1.5	1.6	-0.1	-0.6	-0.6	-2.7	2.0	0.1	-1.6	-1.7	-9.0	4.5
<i>Voronoi</i>													
<i>o</i>	1.3	1.7	1.8	-0.5	-2.3	-0.6	-3.0	1.6	-0.7	-2.9	-2.8	-5.7	1.4
<i>m</i>	0.8	0.8	0.4	0.6	0.9	0.1	0.3	0.8	0.7	0.4	0.4	-2.7	3.2
<i>p</i>	1.2	1.4	1.7	0.0	-0.3	-0.3	-2.3	1.7	0.1	-1.1	-1.5	-7.1	3.3
<i>MDC</i>													
<i>o</i>	0.0	2.2	0.9	0.8	-3.0	3.2	-6.1	-0.2	-3.5	-3.7	1.8	-10.0	2.6
<i>m</i>	0.4	-0.7	-0.5	-0.3	-0.8	-1.1	-0.9	-1.2	-1.0	-0.6	-0.8	-6.3	3.4
<i>p</i>	1.1	1.9	3.0	0.2	-1.6	-2.6	-4.8	2.4	-0.3	-3.1	-3.3	-9.3	3.6

a: CF₃, b: CN, c: COH, d: CL, e: F, f: Me, g: NH₂, h: NO₂, i: OCOH, j: OH, k: OMe, l: O⁻, m: SH₂⁺

Here we check the ability of the charge analyses to predict the orientational substituent effect by checking the charges on the carbon atoms at the *ortho*, *meta* or *para* positions. In Table 4.3.3, the change of charge for the carbon atoms in the *ortho*, *meta* and *para* positions is given relative to the charge of carbon in benzene. As the reaction involves a positively charged electrophile, either of the *ortho*, *meta* or *para* positions will be favored if the charge on the carbon atom will be more negative, and unfavored if it is less negative. These values can not be used in a strict sense, as both the reaction rate and orientation preference depend on the energies of the transition states leading to the cation intermediate, and not on the charges on the atoms. However, some trends can be observed for the resonance effects. The *resonance* electron-withdrawing effects²³⁷ of NO₂, CN and COH is well predicted by all charge analyses, leading to a favored *meta* oriented product. Also the *resonance* electron-donating effects²³⁷ of for instance OH or NH₂ leading to *ortho/para* oriented products is predicted well by all analyses.

Conclusions

Four charge analyses (Mulliken, Hirshfeld, Voronoi Deformation Density and Multipole Derived Charges) have been tested on three properties. The first is their ability to reproduce the molecular multipole moments, which is of great importance for molecular recognition. The Multipole Derived Charges give a good representation of the molecular multipoles *by construction*, while the other three analyses perform poorly (Mulliken) or reasonable (Hirshfeld, Voronoi). For the last three methods, the average absolute deviations of the multipole moments is of the order of 0.2-0.3 a.u. for the dipole moment, and 5.3 (Mulliken) or 1.3 (Hirshfeld, Voronoi) a.u. for the quadrupoles.

The second investigated property is the *inductive* electron-withdrawing/donating effect of substituents. The same trend is shown as for the multipole moments: the Mulliken scheme predicts a wrong substituent effect in four cases, Voronoi and Hirshfeld in two cases, while the Multipole Derived Charges predict the correct effect in all cases.

The last property studied is the substituent effect on the *ortho/meta/para*-orientation of an electrophilic aromatic substitution reaction. However, as the orientation preference is determined completely by the energy of transition states, and not charges of atoms, this can be checked only in a qualitative way. In fact, all four analyses are found to give a good prediction for the orientation of electron-withdrawing (like CN or NO₂) or electron-donating (like OH or NH₂) groups *by resonance*.