

Université de Montréal

A new SOS-DFPT approximation for NMR shielding calculations:  
the Loc.3 correction applied to the catalytic mechanism  
of Serine Proteases

par  
Elisa Fadda

Département de Chimie  
Faculté des Arts et des Sciences

Thèse présentée à la Faculté des études supérieures  
en vue de l'obtention du grade de  
Philosophiæ Doctor (Ph.D.)  
en chimie

Octobre 2003

© Elisa Fadda, 2003



QD

3

U54

2004

v.005

**Direction des bibliothèques**

**AVIS**

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

**NOTICE**

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal  
Faculté des études supérieures

Cette thèse intitulée:

A new SOS-DFPT approximation for NMR shielding calculations:  
the Loc.3 correction applied to the catalytic mechanism  
of Serine Proteases

présentée par

Elisa Fadda

a été évaluée par un jury composé des personnes suivantes:

Tucker Carrington Jr. (président-rapporteur)

Dennis R. Salahub (directeur de recherche)

Matthias Ernzerhof (membre du jury)

David A. Case (examineur externe)

Tucker Carrington Jr. (représentant du doyen de la FES)

## Summary

Nuclear Magnetic Resonance (NMR) is unquestionably one of the most powerful and extensively used experimental techniques for molecular structure elucidation. The key to the success of NMR spectroscopy is firmly related to two main characteristics: its virtually non-existent interference with the sample, and the extreme sensitivity of NMR parameters. Chief among these is the chemical shielding, also denoted as the chemical shift with respect to a standard.

In recent years, *ab initio* calculations have been playing an increasingly important role in support of experimental NMR research. Advances attained in computer data processing now allow one to implement and carry out increasingly more sophisticated and accurate NMR shielding calculations. Among the available approaches, Density-Functional Theory (DFT) based methods have been shown to represent a good compromise between accuracy and computational cost.

This thesis reports the results of the development and extensive testing of an approximation for the determination of the NMR shielding, namely the “local approximation 3” (Loc.3). This approximation has its roots in “Time-Dependent Density-Functional Response Theory” (TDDFRT) and it also represents the physical basis of “Sum-Over-States Density-Functional Perturbation Theory” (SOS-DFPT). The latter is a very successful DFT-based methodology for NMR shield-

ing calculations, which has been the subject of criticism in the past years specifically for the lack of a clear physical basis.

The first part of this work gives a broad perspective on the theoretical framework within which the Loc.3 approximation has been developed. In particular, Chs. 1 and 2 are devoted to the introduction of the basis of time-independent and of time-dependent DFT, respectively. In Ch. 3 we explore the theory of NMR shielding and the fundamentals of the most used *ab initio* implementations, namely “coupled Hartree-Fock” (CHF), “uncoupled Kohn-Sham” (UKS) and SOS-DFPT.

A more technical part then follows, in which a brief digression on the general characteristics of the implementation of the DFT working equations in the *deMonKS* program is presented. Next, we describe the computational details concerning the calculations performed.

The three following chapters are dedicated to the description and discussion of the three stages of testing of the Loc.3 approximation. In the first chapter, the first set of benchmark tests is shown, where Loc.3 is compared to the other SOS-DFPT approximations (UKS, Loc.1 and Loc.2). Moreover, the test is also extended to other powerful DFT-based methodologies. In the second chapter, we discuss the Loc.3 performance against experimental  $^{14,15}\text{N}$  chemical shifts in liquid media. This very important test, based on the results of 132 different nitrogen nuclei, allows us to determine the maximum error associated with the SOS-DFPT NMR shielding calculation of each of ten major types of nitrogen atoms. In the third chapter, we provide a critical assessment of Loc.3 as a support tool for experimental research. This step is concerned with the study of the catalytic mechanism of Serine Protease enzymes. In particular, a suitable model of the active-site of the enzyme has been constructed and the fundamental first steps of the catalytic

mechanism have been reproduced according to two main pathways proposed on the basis of, among others, NMR experimental data. In light of the information obtained through the previous tests, the results will show that the calculation supports only one of the two catalytic mechanisms proposed.

The last part of this work is devoted to the conclusions and perspectives for future development of new DFT-based methodologies for the calculation of NMR properties.

**Keywords:** Density-Functional Theory, Sum-Over-States Density-Functional Perturbation Theory, Loc.3, Time-Dependent Density-Functional Response Theory, Nuclear Magnetic Resonance, Nitrogen Compounds, Serine Proteases, “Ring Flip” Catalytic Mechanism.

# Sommaire

La résonance magnétique nucléaire (RMN) est une technique expérimentale très puissante et entre les plus utilisées actuellement pour la détermination de la structure moléculaire. Les avantages les plus importants que l'utilisation de la RMN présente, sont notamment deux. Le premier est relié à l'absence des interférences avec l'échantillon et le second concerne la haute sensibilité de ses paramètres par rapport au milieu électronique. Le déplacement chimique est certainement un des grandeurs RMN les plus étudiés du point de vue expérimental aussi que théorique. Récemment les méthodes *ab initio* ont commencé à jouer un rôle important comme support dans les études RMN expérimentales. Particulièrement, la performance offerte par les ordinateurs d'aujourd'hui, permet de développer des nouvelles méthodes et subséquemment d'exécuter des calculs RMN de plus en plus précis et sophistiqués. Parmi les méthodes théoriques disponibles, la théorie de la fonctionnelle de la densité (TFD) montre un bon compromis entre précision et coût des calculs.

Dans cette thèse, je présenterai le développement et la validation d'une nouvelle approximation pour le calcul du déplacement chimique en RMN, définie comme "approximation locale 3" ou plus simplement "Loc.3". Loc.3 a été dérivée à partir de la théorie de la fonctionnelle de la densité dépendant du temps (TFD-



DT) et représente le fondement physique de “Sum-Over-States Density-Functional Perturbation Theory” (SOS-DFPT). Cette dernière est une théorie parmi les plus utilisées pour le calcul des paramètres RMN, toutefois, récemment elle a été sujet à des critiques précisément pour l’absence d’un fondement physique clair.

La première partie de cette thèse est dévolue à fournir une ample perspective sur la méthode théorique dans laquelle l’approximation Loc.3 a été développée. En particulier, les chapitres 1 et 2 sont dédiés, respectivement, à l’introduction de la base de la TDF indépendant et dépendent du temps. La théorie de la RMN est présentée dans le chapitre 3, aussi que les aspects fondamentaux des méthodes de calcul les plus utilisées.

Une partie plus technique suit. Ici sont présentées les caractéristiques générales de l’implémentation des équations de la TFD dans le programme de Mon-KS. De plus, le choix des paramètres de calcul est décrit en détail.

Les trois chapitres suivants sont dévolus à l’exposition et l’analyse des résultats obtenus dans les tests de validation et dans la phase d’application de la méthode. Les deux premières séries des vérifications présentées concernent la détermination de la compétence de Loc.3 par rapport aux autres approximations dans la TPDFD-SSE, c’est-à-dire UKS, Loc.1 et Loc.2, et par rapport aux autres fonctionnelles développées précisément pour les calculs RMN. Dans le troisième test, les déplacements chimiques de l’atome d’azote, obtenues avec Loc.3, sont comparés aux références proviennent d’expérimentes dans le milieu liquide. L’importance de ce test est relié à la dimension de l’échantillon étudié. Sur la base des résultats obtenus pour 132 types différents d’atomes d’azotes, on était capable d’établir les erreurs maximales relatives à chacun des dix groupements majeurs des systèmes azotés.

L'étape suivante voit Loc.3 dans un rôle de support pour l'interprétation des données expérimentales RMN. Cette étude concerne l'analyse du mécanisme catalytique d'une classe d'enzymes connues comme protéases à serine. Un modèle adéquat du site actif de l'enzyme a été conçu. En-suite, les premières étapes du mécanisme catalytique ont été reproduit en accord avec les deux chemins de réaction proposées sur la base des données expérimentales RMN. Sur la base des informations relatives à la précision de la méthode obtenues à partir des tests précédents, les résultats appuient seulement un des deux mécanismes proposés.

Le dernier chapitre est dévolu aux conclusions et aux perspectives sur le développement des nouvelles méthodes basées sur la TFD pour le calcul des paramètres RMN.

**Mots Clés:** théorie de la fonctionnelle de la densité, sum-over-states density-functional perturbation theory, Loc.3, théorie de la fonctionnelle de la densité dépendent du temps, résonance magnétique nucléaire, composés azotés, protéases à serine, mécanisme catalytique "ring flip".

# Table of Contents

<b>Summary</b>	<b>1</b>
<b>Sommaire</b>	<b>4</b>
<b>Table of Contents</b>	<b>10</b>
<b>List of Figures</b>	<b>14</b>
<b>List of Tables</b>	<b>21</b>
<b>List of Acronyms</b>	<b>22</b>
<b>Acknowledgements</b>	<b>28</b>
<b>Introduction</b>	<b>30</b>
<b>I Theoretical Background</b>	<b>38</b>
<b>1 Brief Introduction to Density-Functional Theory</b>	<b>39</b>
1.1 The Hohenberg-Kohn Theorems . . . . .	41
1.1.1 The Levy Constrained Search . . . . .	46
1.2 The Kohn-Sham Approach . . . . .	47

1.3	Approximate Exchange-Correlation Functionals . . . . .	49
1.3.1	The Exchange-Correlation Hole . . . . .	50
1.3.2	The Adiabatic Connection . . . . .	52
1.3.3	The Local Density Approximation . . . . .	54
1.3.4	The Generalized Gradient Approximation . . . . .	56
1.3.5	Hybrid Functionals . . . . .	58
1.3.6	Meta-GGA Functionals . . . . .	60
1.3.7	Optimized Effective Potential . . . . .	61
1.3.8	Asymptotically Corrected Functionals . . . . .	63
<b>2</b>	<b>The Time Domain</b>	<b>67</b>
2.1	The Time-Dependent Kohn-Sham Equations . . . . .	67
2.2	Time-Dependent Linear Response Theory . . . . .	70
2.2.1	Density-Functional Response Theory . . . . .	72
2.3	Working Equations for Real and Imaginary Perturbations . . . . .	75
2.3.1	Linear Response from a Real Perturbation . . . . .	76
2.3.2	Linear Response from an Imaginary Perturbation . . . . .	78
2.3.3	The Tamm-Dancoff Approximation . . . . .	81
<b>3</b>	<b>Theory of the NMR Shielding Tensor</b>	<b>83</b>
3.1	The Electronic Hamiltonian in a Magnetic Field . . . . .	85
3.1.1	The Gauge Origin Problem . . . . .	88
3.2	The Coupled Hartree-Fock (CHF) Method . . . . .	91
3.3	From CHF to Uncoupled Density-Functional Theory (UDFT) . . . . .	95
3.3.1	The Sum-Over-States Density-Functional Perturbation Theory . . . . .	99

---

3.3.2	The Loc.3 Approximation . . . . .	103
<b>II</b>	<b>Computational Methodology</b>	<b>106</b>
<b>4</b>	<b>Solution of the KS Equations: the <i>deMon-KS</i> Program</b>	<b>107</b>
4.1	The Coulomb Integral . . . . .	110
4.2	The Exchange-Correlation Integral . . . . .	111
4.3	Response Module: Calculation of the Coupling Matrix Elements . . . . .	112
4.3.1	Implementation of the TDA in the <i>dynaRho</i> Program . . . . .	114
<b>5</b>	<b>Computational Details</b>	<b>116</b>
5.1	Exchange-Correlation Functionals . . . . .	116
5.2	Basis-Sets . . . . .	118
5.3	Integration Grid . . . . .	119
<b>III</b>	<b>Results and Discussion</b>	<b>121</b>
<b>6</b>	<b>Strengths and Limitations of the Loc.3 Correction</b>	<b>122</b>
6.1	Loc.3 vs. Classic SOS-DFPT . . . . .	123
6.2	Loc.3 vs. other DFT-based Methodologies . . . . .	133
<b>7</b>	<b>Nitrogen Shielding in Different Molecular Structures</b>	<b>140</b>
7.1	Amines . . . . .	145
7.2	Hydrazines . . . . .	149
7.3	Amides, Ureas and Guanidines . . . . .	150
7.4	Cyanides and Isocyanides . . . . .	152
7.5	Azoles . . . . .	156

**TABLE OF CONTENTS** **10**

---

7.6	Azines . . . . .	161
7.7	Azine N-Oxides . . . . .	167
7.8	Nitrates . . . . .	170
<b>8</b>	<b>The Catalytic Mechanism of Serine Proteases</b>	<b>173</b>
8.1	The Active Site . . . . .	176
8.2	The Generally Accepted Catalytic Mechanism . . . . .	178
8.3	The “Ring Flip” Mechanism . . . . .	183
8.4	Construction of a Model System for the Tetrahedral Intermediate .	186
8.5	<sup>14,15</sup> N NMR Shielding in His <sub>57</sub> . . . . .	190
8.6	H-bond Induced Shift . . . . .	194
8.7	<sup>14,15</sup> N Shieldings in TI-1 . . . . .	199
8.7.1	Loc.3 <sup>14,15</sup> N Shieldings of the TI-1 Model Systems . . . . .	202
<b>9</b>	<b>Conclusions and Perspectives</b>	<b>206</b>
	<b>Bibliography</b>	<b>216</b>
	<b>Appendix</b>	<b>236</b>
<b>A</b>	<b>Tables</b>	<b>237</b>
<b>B</b>	<b>Figures</b>	<b>263</b>

## List of Figures

- 1.1 Electron density map from X-Ray diffraction of a protein side-chain containing the phenylalanine residue. The numbers at the bottom represent three different degrees of resolution. (Copyright 1997 Christophe Verlinde) . . . . . 40
- 6.1  $^{13}\text{C}$  NMR shieldings from SOS-DFPT and *ab initio* calculations compared to experimental results. The color of the symbols depends upon the functional chosen . . . . . 137
- 6.2  $^{14,15}\text{N}$  NMR shieldings from SOS-DFPT and *ab initio* calculations compared to experimental results. The color of the symbols depends upon the functional chosen. The results obtained for  $\text{NH}_3$  are shown on the left-top corner of the graph. . . . . 138
- 6.3  $^{17}\text{O}$  NMR shieldings from SOS-DFPT and *ab initio* calculations compared to experimental results. The color of the symbols depends upon the functional chosen. The left portion of the diagram shows the down-field region of the spectrum, where are indicated the results for  $\text{OF}_2$  and  $\text{HCHO}$ . Here the small ticks mark 100 ppm. The bisecting line is shown in both regions of the diagram. . 139

- 7.1 Portion of the  $^{14,15}\text{N}$  NMR shielding spectrum (vs.  $\text{CH}_3\text{NO}_2$ ) analyzed in this study. The Loc.3 results, indicated with red  $\star$ , and the UKS results, indicated with black  $\Delta$ , are shown for comparison. The sections associated with the different atom types are also indicated. . . . . 146
- 7.2  $^{14,15}\text{N}$  NMR shielding in various amines. The different SOS-DFPT approximations are represented by the symbols listed on the legend. The linear regression line, relative to each SOS-DFPT approach, is also shown. . . . . 148
- 7.3 Three system studied for the evaluation of the  $\beta$ -effect. a) *n*-butyl-amine, b) *s*-butyl-amine and c) *t*-butyl-amine. . . . . 149
- 7.4 Distribution of the calculated results on cyanides (left part) and isocyanides (right part) against experimental references. The symbols,  $\Delta$  (black),  $\square$  (blue),  $\circ$  (green) and  $\star$  (red), represent the results for UKS, Loc.1, Loc.2 and Loc.3 respectively. The black continuous line on both parts represents the bisecting line. The linear regression lines are shown only on the right part. . . . . 154
- 7.5 Distribution of the calculated  $^{14,15}\text{N}$  shieldings in azoles against the experimental references. The linear regression lines have been derived from the analysis of both pyrrole- and pyridine-type  $^{14,15}\text{N}$  shieldings. . . . . 158
- 7.6 Distribution of the calculated  $^{14,15}\text{N}$  shieldings in azines against the experimental references. The black continuous line indicates the bisecting line. . . . . 163



- 7.7 Calculated  $^{14,15}\text{N}$  shieldings of a pyridine molecule in different water clusters. The  $^{14,15}\text{N}$  shieldings are expressed as a function of the number of water molecules in the cluster. . . . . 166
- 7.8 Distribution of the calculated  $^{14,15}\text{N}$  shieldings in various azine N-oxides against the experimental references. The points corresponding to the 2-OMe-pyridine N-oxide are indicated within a box in the top right corner of the graph. The continuous black line corresponds to the bisecting line . . . . . 169
- 7.9 Distribution of the calculated  $^{14,15}\text{N}$  shieldings in nitrates against experimental references. The black continuous line indicates the bisecting line. . . . . 171
- 8.1 Porcine pancreatic elastase acyl enzyme (PDB code 1GVK). The catalytic triad is represented schematically on the top-right corner of the picture. Its position within the enzyme is also indicated (His in cyan, Asp in red and Ser in yellow). . . . . 175
- 8.2 Active site of porcine pancreatic elastase acyl enzyme (PDB code 1GVK). Besides the residues composing the catalytic triad, Ser<sub>214</sub> and the water molecule H<sub>2</sub>O<sub>305</sub> are also shown. . . . . 177
- 8.3 Qualitative reaction profile of the generally accepted catalytic mechanism of serine proteases. The information on the relative stability of each reaction stage has been based on the quantitative diagram in the work of West and coworkers.<sup>241</sup> . . . . . 180
- 8.4 Formation of the TI-1 during the aminolysis according to the “ring flip” mechanism. . . . . 184

8.5	Formation of the TI-2 from the flipped acyl-enzyme during the hydrolysis. . . . .	186
8.6	Model of the tetrahedral intermediate 1 (TI-1 <sub>A</sub> ) built according to the generally accepted mechanism. . . . .	188
8.7	Model of the tetrahedral intermediate 1 (TI-1 <sub>B</sub> ) built according to the “ring flip” mechanism. . . . .	188
8.8	The two models indicated as a) and b) have been constructed in order to evaluate approximately the energy required for the breaking of the two strong H-bonds in the active site of serine protease. . . .	189
8.9	Types of nitrogen nuclei of the catalytic His involved in the different steps of the serine protease reaction path. . . . .	191
8.10	Three model systems built to analyze the H-bond induced shifts on the imidazole <sup>14,15</sup> N shieldings. . . . .	195
8.11	N <sub>δ1</sub> NMR shielding dependence upon the varying N <sub>δ1</sub> -H distance. The NMR calculations have been performed on the two systems, indicated as 1) and 2), shown schematically below the diagram. . . .	198
8.12	Active site of γ-chymotrypsin complexed with DIFP (PDB identifier 1GMH). His <sub>57</sub> in cyan, Asp <sub>102</sub> in red, Ser <sub>195</sub> in yellow and DISP in orange. . . . .	201
8.13	Reduced version of the TI-1 <sub>A</sub> and TI-1 <sub>B</sub> models used to perform the NMR calculations. . . . .	203
B.1	Structure of some of the cyclic amides examined in Tab. A.XVIII.	264
B.2	Structure of the azoles examined in Tab. A.XXI. . . . .	265
B.3	Structure of some of the azines examined in Tab. A.XXII. . . . .	266

## List of Tables

3.I	Corrections to the excitation energy in SOS-DFPT. . . . .	105
5.I	Absolute average deviation ( $\text{\AA}$ ) from experimental references of bond lengths calculated with different $V_{xc}$ and different basis-sets. . . . .	117
6.I	Statistical analysis of the SOS-DFPT (SVWN) and high quality <i>ab initio</i> $^{13}\text{C}$ absolute shieldings, relative to experimental references. See Tab. A.I for details. . . . .	124
6.II	Statistical analysis of the SOS-DFPT (PW91-PW91) and high quality <i>ab initio</i> $^{13}\text{C}$ absolute shieldings, relative to experimental references. See Tab. A.II for details. . . . .	125
6.III	Statistical analysis of the SOS-DFPT (ACLDA) and high quality <i>ab initio</i> $^{13}\text{C}$ absolute shieldings, relative to experimental references. See Tab. A.III for details. . . . .	126
6.IV	Statistical analysis of the SOS-DFPT (SVWN) and high quality <i>ab initio</i> $^{14,15}\text{N}$ absolute shieldings, relative to experimental references. See Tab. A.IV for details. . . . .	128

6.V	Statistical analysis of the SOS-DFPT (PW91-PW91) and high quality <i>ab initio</i> $^{14,15}\text{N}$ absolute shieldings, relative to experimental references. See Tab. A.V for details. . . . .	129
6.VI	Statistical analysis of the SOS-DFPT (ACLDA) and high quality <i>ab initio</i> $^{14,15}\text{N}$ absolute shieldings, relative to experimental references. See Tab. A.VI for details. . . . .	130
6.VII	Statistical analysis of the SOS-DFPT (SVWN) and high quality <i>ab initio</i> $^{17}\text{O}$ absolute shieldings, relative to experimental references. See Tab. A.VII for details. . . . .	131
6.VIII	Statistical analysis of the SOS-DFPT (PW91-PW91) and high quality <i>ab initio</i> $^{17}\text{O}$ absolute shieldings, relative to experimental references. See Tab. A.VIII for details. . . . .	131
6.IX	Statistical analysis of the SOS-DFPT (ACLDA) and high quality <i>ab initio</i> $^{17}\text{O}$ absolute shieldings, relative to experimental references. See Tab. A.IX for details. . . . .	132
6.X	Statistical analysis of the $^{13}\text{C}$ absolute shieldings calculated with the SOS-DFPT Loc.3 (LDA), other DFT-based and high quality <i>ab initio</i> methods, relative to experimental references. See Tab. A.X and A.XI for details. . . . .	134
6.XI	Statistical analysis of the $^{14,15}\text{N}$ absolute shieldings calculated with the SOS-DFPT Loc.3 (LDA), other DFT-based and high quality <i>ab initio</i> methods, relative to experimental references. See Tab. A.XII and A.XIII for details. . . . .	135

6.XII	Statistical analysis of the $^{17}\text{O}$ absolute shieldings calculated with the SOS-DFPT Loc.3 (LDA), other DFT-based and high quality <i>ab initio</i> methods, relative to experimental references. See Tab. A.XIV and A.XV for details. . . . .	136
7.I	Statistical analysis of the calculated $^{15}\text{N}$ chemical shifts (ppm) in amines relative to neat liquid nitromethane . . . . .	147
7.II	Statistical analysis of the calculated $^{15}\text{N}$ chemical shifts (ppm) in hydrazines relative to neat liquid nitromethane . . . . .	150
7.III	Statistical analysis of the calculated $^{15}\text{N}$ chemical shifts (ppm) in amides, ureas and guanidines relative to neat liquid nitromethane	151
7.IV	Statistical analysis of the calculated $^{15}\text{N}$ chemical shifts (ppm) in cyanides relative to neat liquid nitromethane . . . . .	153
7.V	Statistical analysis of the calculated $^{15}\text{N}$ chemical shifts (ppm) in isocyanides relative to neat liquid nitromethane . . . . .	155
7.VI	Statistical analysis of the calculated pyrrole-type $^{15}\text{N}$ chemical shifts (ppm) in azoles relative to neat liquid nitromethane . . .	159
7.VII	Statistical analysis of the calculated pyridine-type $^{15}\text{N}$ chemical shifts (ppm) in azoles relative to neat liquid nitromethane .	161
7.VIII	Statistical analysis of the calculated $^{15}\text{N}$ chemical shifts (ppm) in azines relative to neat liquid nitromethane . . . . .	162
7.IX	Statistical analysis of the calculated $^{15}\text{N}$ chemical shifts (ppm) in azine N-oxides relative to neat liquid nitromethane . . . . .	168
7.X	Statistical analysis of the calculated $^{15}\text{N}$ chemical shifts (ppm) in nitrates relative to neat liquid nitromethane . . . . .	170

8.I	$^{14,15}\text{N}$ shieldings (ppm from neat liquid $\text{MeNO}_2$ ) of $\alpha$ , $\alpha^+$ and $\beta$ nitrogen types in imidazole, N(Me)-imidazole and in His. . . . .	192
8.II	$^{14,15}\text{N}$ shieldings (ppm from neat liquid $\text{MeNO}_2$ ) of $\alpha$ , $\alpha^+$ and $\beta$ nitrogen types in 5-Me-imidazole. . . . .	193
8.III	H-bond effect on the $^{14,15}\text{N}$ shielding of the $\alpha$ and $\beta$ nitrogen types in imidazole and 5-Me-imidazole and of the $\alpha^+$ nitrogen type in imidazolium. The systems studied are represented in Fig. 8.10. . . . .	196
8.IV	Experimental $^{14,15}\text{N}$ shieldings (ppm from neat liquid $\text{MeNO}_2$ ) of the $\alpha$ -lytic protease inhibited complexes in a full protonation state compared to the Loc.3 calculations on models of the TI-1. . . . .	204
A.I	$^{13}\text{C}$ shielding constant (ppm) calculated with the four SOS-DFPT approximations and the SVWN $v_{xc}$ . . . . .	238
A.II	$^{13}\text{C}$ shielding constant (ppm) calculated with the four SOS-DFPT approximations. The PW91 $v_x$ is used in combination with the PW91 $v_c$ . . . . .	239
A.III	$^{13}\text{C}$ shielding constant (ppm) calculated with the four SOS-DFPT approximations and the ACLDA $v_{xc}$ . . . . .	240
A.IV	$^{14,15}\text{N}$ shielding constant (ppm) calculated with the four SOS-DFPT approximations and the SVWN $v_{xc}$ . . . . .	241
A.V	$^{14,15}\text{N}$ shielding constant (ppm) calculated with the four SOS-DFPT approximations. The PW91 $v_x$ is used in combination with the PW91 $v_c$ . . . . .	242

A.VI	$^{15}\text{N}$ shielding constant (ppm) calculated with the four SOS-DFPT approximation and the AC-LDA $v_{xc}$ . . . . .	243
A.VII	$^{17}\text{O}$ shielding constant calculated with the four SOS-DFPT approximations and the SVWN $v_{xc}$ . . . . .	244
A.VIII	$^{17}\text{O}$ shielding constant calculated with the four SOS-DFPT approximations and the PW91-PW91 $v_{xc}$ . . . . .	245
A.IX	$^{17}\text{O}$ shielding constant calculated with the four SOS-DFPT approximations and the AC-LDA $v_{xc}$ . . . . .	246
A.X	$^{13}\text{C}$ shielding constants (ppm) calculated with the SOS-DFPT Loc.3 approximation (LDA) and with other DFT-based methodologies. High quality <i>ab initio</i> and experimental data are shown as reference . . . . .	247
A.XI	$^{13}\text{C}$ shielding constants (ppm) calculated with the SOS-DFPT Loc.3 approximation (LDA) and with the KT1 and KT2 functionals. High quality <i>ab initio</i> and experimental data are shown as reference . . . . .	248
A.XII	$^{15}\text{N}$ shielding constant (ppm) calculated with the Loc.3 (LDA) and with other DFT-based methodologies. High quality <i>ab initio</i> and experimental data are shown as reference . . . . .	249
A.XIII	$^{14,15}\text{N}$ shielding constants (ppm) calculated with the SOS-DFPT Loc.3 approximation (LDA) and with the KT1 and KT2 functionals. High quality <i>ab initio</i> and experimental data are shown as reference . . . . .	250

A.XIV	$^{17}\text{O}$ shielding constant calculated with the SOS-DFPT Loc.3 approximation (LDA) and with other DFT-based methodologies. High quality <i>ab initio</i> and experimental data are shown as reference . . . . .	251
A.XV	$^{17}\text{O}$ shielding constants (ppm) calculated with the SOS-DFPT Loc.3 approximation (LDA) and with the KT1 and KT2 functionals. High quality <i>ab initio</i> and experimental data are shown as reference . . . . .	252
A.XVI	Calculated and experimental $^{15}\text{N}$ chemical shifts (ppm) in amines relative to neat liquid nitromethane . . . . .	253
A.XVII	Calculated and experimental $^{15}\text{N}$ chemical shifts (ppm) in hydrazines relative to neat liquid nitromethane . . . . .	254
A.XVIII	Calculated and experimental $^{15}\text{N}$ chemical shifts (ppm) in amides, ureas, guanidines and related structures relative to neat liquid nitromethane . . . . .	255
A.XIX	Calculated and experimental $^{15}\text{N}$ chemical shifts (ppm) in cyanides relative to neat liquid nitromethane . . . . .	256
A.XX	Calculated and experimental $^{15}\text{N}$ chemical shifts (ppm) in isocyanides relative to neat liquid nitromethane . . . . .	257
A.XXI	Calculated and experimental $^{15}\text{N}$ chemical shifts (ppm) in azoles relative to neat liquid nitromethane . . . . .	258
A.XXII	Calculated and experimental $^{15}\text{N}$ chemical shifts (ppm) in azines relative to neat liquid nitromethane . . . . .	260
A.XXIII	Calculated and experimental $^{15}\text{N}$ chemical shifts (ppm) in azine N-oxides relative to neat liquid nitromethane . . . . .	261



---

A.XXIV Calculated and experimental  $^{15}\text{N}$  chemical shifts (ppm) in nitrates relative to neat liquid nitromethane . . . . . 262

# List of Acronyms

2LM Two Level Model

ACLDA Asymptotically-Corrected Local Density Approximation

ALDA Adiabatic Local Density Approximation

AO Atomic Orbital

B Becke 88 exchange functional (also known as B88)

B3LYP three parameter Becke (B) exchange plus LYP correlation hybrid functional

B86 Becke 86 exchange functional

B97 Becke 97 exchange-correlation functional

B97-1 reparametrized B97 exchange-correlation functional

BFGS Broyden-Fletcher-Goldfarb-Shanno geometry optimization algorithm

BSSE Basis Set Superposition Error

CCSD Coupled-Cluster Singles and Doubles

---

CCSD(T) Coupled-Cluster Singles and Doubles augmented with Triples excitations

COSY CORrelation SpectroscopY

CHF Coupled Hartree-Fock

CI Configuration-Interaction

DFT Density-Functional Theory

FF Finite Field theory

FP Finite Perturbation theory

GGA Generalized Gradient Approximation

GIAO Gauge-Independent Atomic Orbitals

GTO Gaussian-Type Orbital

HH Half-and-Half hybrid functional

HF Hartree-Fock

HOMO Highest Occupied Molecular Orbital

HCTH Hamprecht-Cohen-Tozer-Handy exchange functional

IGLO Individual Gauge for Localized Orbitals

KLI Krieger-Li-Iafrate approximation

KS Kohn-Sham

LAP meta-GGA correlation functional

LBHB Low Barrier Hydrogen Bond

LCAO Linear Combination of Atomic Orbitals

LDA Local Density Approximation

LMO Localized Molecular Orbital

Loc.n Local approximation “n”

LORG localized orbital Local ORiGin

LSD Local Spin-Density approximation

LUMO Lowest Unoccupied Molecular Orbital

LYP Lee-Yang-Parr correlation functional

MO Molecular Orbital

mGGA meta Generalized Gradient Approximation

MKS Multiplicative Kohn-Sham method

MPn nth order Möller-Plesset theory

NMR Nuclear Magnetic Resonance

NOESY Nuclear Overhauser Enhancement effect Spectroscopy

OEP Optimized Effective Potential

PDB Protein Data-Bank

P86 Perdew 86 correlation functional

PBE Perdew-Burke-Ernzerhof exchange and/or correlation functional

PBE0 Perdew-Burke-Ernzerhof zero parameters hybrid functional

PW86 Perdew-Wang 86 exchange functional

PW91 Perdew-Wang 91 exchange and/or correlation functional

RHF Restricted Hartree-Fock

RPA Random Phase Approximation

S Slater exchange functional

SCF Self-Consistent Field

SIC Self-Interaction Correction

SOLO Second-Order LORG

SOS Sum-Over-States

SOS-DFPT Sum-Over-States Density-Functional Perturbation Theory

STO Slater-Type Orbital

TDA Tamm-Dancoff Approximation

TDDFT Time-Dependent Density-Functional Theory


TDDFRT Time-Dependent Density-Functional Response Theory

TDHF Time-Dependent Hartree-Fock

UDFT Uncoupled Density-Functional Theory

UKS Uncoupled Kohn-Sham

VWN Vosko-Wilk-Nusair LSD correlation functional



*To my father and my mother,  
for being my life's greatest inspiration.*

*To Andrea,  
for all the love and support.*



## Acknowledgements

My experience as a Ph.D. student started almost 5 years ago. During this time I have had the fortune and the privilege of meeting many people that, knowingly or not, helped me throughout my studies.

Firstly, I would like to thank my research director, Prof. Dennis R. Salahub. He accepted me in his research group when its “most celebrated” days at the University of Montréal were about to end. Shortly after, in fact, he took the position as director of the NRC Stacie Institute and moved to Ottawa. Nonetheless, he trusted me to stay in Montréal to continue my research project. During the whole time I have been his student, he has allowed me the freedom to take the path I chose to follow to conduct my research, and has always been available whenever I have needed help and guidance. Furthermore, he offered me the opportunity to be part of the “*deMon*\* developers family”.

During the annual meetings of the *deMon* developers group, I had the privilege of meeting some of the most bright and talented researchers in DFT. Among these I would like to acknowledge in particular Mark E. Casida, who I not only admire as a scientist, but also who I consider a very dear friend. I am especially thankful to Mark for all the time he invested in teaching me all that I know about time-

---

\**deMon* is a suite of DFT-based programs developed by the Salahub research group.



dependent DFT and for all the support he provided through the years.

I would like to thank Alain Rochefort, Steeve Chrétien, Seongho Moon and Serguei Patchkovskii for all the help and the good advice, and Jean-François Truchon, who introduced me to NMR calculations.

I would also like to acknowledge all the people at the *centre de recherche en calcul appliqué* (CERCA), where I studied for all these years, and particularly Michel Béland, Richard Lefebvre and Pierre Hamelin for their invaluable help.

Furthermore, I'm especially grateful to Lyne Laurin at the University of Montréal for helping me with all the paperwork I had to prepare every year for my scholarship.

My Ph.D. studies have been funded by a grant from the Government of Sardinia, Italy. I would like to acknowledge Rag. Massimo Lallai for his help and understanding throughout these years.

The *réseau québécois de calcul de haute performance* (RQCHP) is also gratefully acknowledged for the computational time.

# Introduction

One of the primary challenges of today's chemistry pertains to the determination of the link between structure and function of large biomolecular systems. Among these, a particularly difficult case is represented by proteins. Their three-dimensional (3D) structure is, in fact, an apparently chaotic arrangement of peptide chains. A closer analysis leads the observer to realize that this 3D structure is the result of a perfect equilibrium of weak interactions between different atom groups in the peptide residues. The mechanistic significance of this arrangement is shown by the simple fact that active proteins *in vivo* assume only one configuration, known as native.

The determination of protein structure is nowadays an interdisciplinary effort. The experimental techniques most widely used for this purpose are X-ray diffraction and nuclear magnetic resonance (NMR). Through the analysis of high resolution density maps, X-ray crystallography allows one to obtain very accurate protein structures, to the extent that presently the position of many hydrogen atoms can be resolved. The only shortcoming of this method is that it reveals the structure of the crystallized form of the protein, which can differ from that of the *in vivo* one. On the other hand, NMR based techniques are able to produce a less refined 3D structure, which, nevertheless, corresponds more closely to the native

configuration of the protein.

The resolution of a protein structure through NMR is an elaborate process involving several methods. The first step consists usually in the acquisition of a series of one-dimensional (1D) spectra to construct a two-dimensional (2D) correlation spectroscopy (COSY) spectrum. The homonuclear COSY spectrum, together with heteronuclear assignment experiments, are necessary for the attribution of the signals. The distance constraints can be determined through the nuclear Overhauser enhancement (NOE) effect. The intensities of the cross peaks in a 2D NOESY spectrum are correlated to the distance between the protons which they represent. The technique is sensitive for distances up to 5 Å. Further, information on the backbone and side-chain torsional angles can be derived from the three-bond coupling constants. The combination of these methods provides accurate results for smaller and medium-sized proteins (i.e. proteins with molecular weight <25kDa or with less than 200 residues). However, for larger systems, problems due to NOE peaks overlap and line broadening can compromise the refinement of the final structure.

The limitations of experimental NMR have given rise to the study of alternative methods, one of which is the analysis of the correlation between structure and NMR properties on a smaller scale. The comparison of the NMR shieldings and coupling constants to known structural features in proteins provides us with critical information. Furthermore, recent developments in theoretical methods for NMR properties offer a very powerful alternative. The careful assessment of a theoretical methodology through the comparison with experimental data can result in the use of theory as a supporting tool for the interpretation of the NMR spectra.

The size of biomolecular systems is of major concern for the application of theoretical methods. However, the level of compromise between size and accuracy is lowered progressively by constant advances in computer data processing. In addition, the development of hybrid theoretical methods enables one to consider systems of increased size compared to the full quantum mechanical or semiempirical approaches. One of the methodologies that has the highest potentials for treating large biomolecules is QM/MM. In QM/MM, the largest portion of the system is taken into account by means of an empirical force field (MM), while the part of most interest is treated with higher level quantum mechanics (QM). For NMR calculations, the ONIOM-NMR approach of Karadakov and Morokuma<sup>1</sup> offers a very interesting prospect. Highly accurate correlated methods (i.e. higher order perturbation theory or coupled-cluster approaches) can be used to describe the central part of the system, while the remainder can be treated with coupled Hartree-Fock (CHF). However, the computational cost of higher correlated methods for routine application in biomolecules, even if on a small portion of the system, is prohibitive as yet. In fact, recent applications of the ONIOM-NMR method on proteins involve uncoupled density-functional theory (UDFT) or CHF with different quality of basis sets for the central and peripheral regions.<sup>2,3</sup>

Generally, UDFT is preferred to CHF since it allows the inclusion of correlation effects at the same computational cost. For some important nuclei in biomolecular NMR, such as nitrogen, correlation effects are crucial in order to obtain accurate results. The shortcoming of UDFT is related to the underestimation of the eigenvalue difference, which becomes an important issue in the determination of the NMR properties of unsaturated nuclei. The least accurate results are obtained with exchange-correlation functionals defined within the local density

approximation (LDA). Nonetheless, the impact of this problem can be reduced in various ways, for instance, through the use of more sophisticated exchange-correlation functionals, the construction of new functionals for NMR calculations (e.g. through an *ad hoc* parametrization), or by the explicit correction of the excitation energy through the addition of appropriate terms. The latter approach has been introduced by Malkin and coworkers<sup>4</sup> and it is known as sum-over-states density-functional perturbation theory (SOS-DFPT). Two exchange-only correction terms have been introduced in the original paper, namely the local approximation 1 (Loc.1) and the local approximation 2 (Loc.2). Although SOS-DFPT improves significantly the accuracy of the results compared to the uncoupled approach, maintaining at the same time its computational efficiency, many criticized it for the lack of a clear physical basis.

The subject of this thesis is the introduction of a newly developed SOS-DFPT approximation, namely the local approximation 3 (Loc.3), for NMR shielding calculations. Loc.3 has its roots in time-dependent density-functional response theory, a formally exact theory, and is defined within the Tamm-Dancoff approximation (TDA) and the two level model (2LM). According to the exact time-dependent perturbation theory, both magnetic and electric perturbations may be described using the same SOS expression<sup>†</sup>. However, in adiabatic time-dependent DFT (TDDFT), the absence of explicit current-density dependence of the exchange-correlation functional causes the magnetic and electric SOS expressions to differ. The TDA allows us to keep electric and magnetic perturbations on the same footing. The 2LM has been considered because, by approximating the excited states

---

<sup>†</sup>SOS expressions for the corresponding time-independent response properties are obtained by considering the static limit of their time-dependent derivation.

by singlet states obtained through the excitation of one electron from an occupied orbital to a virtual orbital, it is computationally more efficient. Moreover, the 2LM brings Loc.3 to the same level as Loc.1 and Loc.2. One of the most important results of this thesis work is indeed the elucidation of the physical basis of SOS-DFPT through the comparison of Loc.3 to Loc.1 and Loc.2.

The quality of Loc.3 as an independently determined approximation is another crucial aspect of the assessment of SOS-DFPT. The first set of tests entails the calculation of the absolute shielding of  $^{13}\text{C}$ ,  $^{14,15}\text{N}$  and  $^{17}\text{O}$  of small organic and inorganic molecules. The comparison against the uncoupled approach and the SOS-DFPT “classic” local corrections demonstrates that Loc.3 generally enhances the quality of the NMR shieldings especially in the case of unsaturated nuclei. A second step consists in the evaluation of Loc.3 against the most recent and promising DFT-based approaches for NMR calculations. The results indicate that Loc.3 is rather weak in the approximation of the  $^{13}\text{C}$  NMR shielding, however it is competitive in the case of  $^{17}\text{O}$  and especially in the case of  $^{14,15}\text{N}$ .

The good performance obtained in the calculation of the  $^{14,15}\text{N}$  shielding is a significant result. In fact, the accurate calculation of its NMR properties requires the introduction of correlation effects and Loc.3 is the only SOS-DFPT correction term that explicitly accounts for correlation. Furthermore,  $^{14,15}\text{N}$  NMR plays a central role in the study of the structure and function of biomolecules. Hence, the careful assessment of the Loc.3 strengths and limitations in the NMR shielding calculation of various types of nitrogen atoms can allow us to use SOS-DFPT as a supporting tool for the experimental NMR of molecules of biological interest. In order to accomplish this, I tested all the SOS-DFPT approximations, including Loc.3, against 132 experimental  $^{14,15}\text{N}$  NMR data on the chemical shift scale.

The major obstacle of this analysis is the lack of explicit inclusion of solvent effects in the calculation. In fact, while the experiment is conducted at room temperature and in the liquid-phase, the calculation reproduces more closely a zero-temperature, zero-pressure state. Furthermore,  $^{14,15}\text{N}$  NMR is very sensitive to the presence of the solvent. Since the explicit inclusion of solvent molecules in the calculation is prohibitively expensive, in the absence of a QM/MM module in our program, the computational approach chosen to limit the solvent effect is a pragmatic one. According to the Kamlet-Taft theory (see Ch.7 page 143) solvents with similar characteristics induce comparable shifts to the NMR shielding of similar types of  $^{14,15}\text{N}$  nuclei. Therefore, the 132 calculated  $^{14,15}\text{N}$  shieldings have been analyzed within eight different nitrogen classes (i.e. amines, hydrazines, amides, cyanides and isocyanides, azoles, azines, azine N-oxides and nitrates). The results, which cover a vast portion of the  $^{14,15}\text{N}$  NMR spectrum (i.e.  $\sim 450$  ppm), confirm that the accuracy of the calculations is strongly dependent on the type of nitrogen nucleus. Loc.3 represents a significant improvement for highly unsaturated nitrogens, such as cyanides and isocyanides, and for aromatic systems with multiple nitrogen nuclei within the ring. The most difficult cases are represented by the highly delocalized systems, where ring currents play a crucial role. Particularly large errors are observed for aryl cyanides and for pyridine-type nitrogens. In the latter case the solvent effect has been simulated explicitly by the progressive addition of water molecules. The results show that the addition of a minimum of 25 water molecules is necessary to reproduce the  $\sim +40$  ppm of solvent-induced shift.

The knowledge of the margins of error of Loc.3 for each nitrogen type offers the opportunity to employ it as a predictive tool, as in the study of the catalytic

mechanism of a class of proteolytic enzymes known as serine proteases. The general features of the catalytic reaction path have been proven beyond any doubt by X-ray diffraction and by  $^1\text{H}$  and  $^{14,15}\text{N}$  NMR studies. However some important details involving transition states and unstable intermediates are still a matter of debate. Recently, a new version of the generally accepted catalytic mechanism has been introduced, namely the “ring flip” mechanism.<sup>5</sup> According to this theory, the imidazole ring of the catalytic histidine residue (His) in the active site performs a rotation of  $180^\circ$ , presumably in a stage near the formation of the first reaction’s intermediate. The role of the rotated His is to promote the hydrolysis of the amine group from the intermediate, a step that has been difficult to explain in terms of the most widely accepted version of the catalytic mechanism. Unfortunately, there are considerable difficulties in finding experimental evidence for the “ring flip” theory. In fact, the flipped rotamer is thought to appear only in intermediate structures, which have a very short lifetime. The only data supporting the presence of a rotated His in the active site are  $^{14,15}\text{N}$  NMR shieldings coming from a low pH structure of particular inhibited complexes. The molecules acting as inhibitors in these complexes (i.e. diisopropylfluorophosphates or phenylmethanesulfonylfluorides) are known to block the catalysis of serine proteases in a stage that mimics the conformation of the reaction intermediate of the free enzyme.

The advantage of the theoretical approach to this problem is that any type of structure, regardless of the extent of its lifetime, can be studied. Therefore, a suitable prototype of the active site in the first reaction intermediate has been built according to the two catalytic mechanisms, i.e. the generally accepted one and the “ring flip”. Subsequently, the Loc.3  $^{14,15}\text{N}$  shieldings have been calculated for all the His conformers obtained and the results compared to the experimental data



relative to the inhibited complexes. Based on the tests performed to gauge the dependence of the  $^{14,15}\text{N}$  shieldings upon the H-bond characteristics, and on the margins of errors determined for the model chosen for the His residue, the calculations prove that the formation of the reaction intermediate involves a rotation of the imidazolium ring in the catalytic His. Although the methodology's margin of error does not allow one to state with certainty if this rotation is partial or complete, the calculations do provide compelling evidence for the flipped conformer which shows the best agreement with the experimental results.

This thesis is organized as follows. The first part is dedicated to the theoretical background. The basis of time-independent and time-dependent DFT are demonstrated, respectively, in Ch. 1 and 2. The content of these chapters has been organized according to two of the most important and recent textbooks on DFT.<sup>6,7</sup> Ch. 3 is devoted to the basics of the NMR theory and of its most widely used implementations, such as CHF and UDFT. The fundamentals of SOS-DFPT are shown in Sec. 3.3.1, while the Loc.3 approximation is introduced subsequently in Sec. 3.3.2.

In the second part all the computational details are explained, from the description of the DFT program used throughout this study, to the choice of the parameters for each type of calculation (i.e. exchange-correlation functionals, basis sets and type of grid).

The third and last part is dedicated to the presentation and discussion of the results. Ch. 6 and Ch. 7 are dedicated to the benchmark tests of Loc.3, while Ch. 8 is dedicated to the study of the catalytic mechanism of serine protease. Finally, conclusive remarks are presented in Ch. 9, together with some ideas for the future development of this work.



## **Part I**

# **Theoretical Background**



# Chapter 1

## Brief Introduction to Density-Functional Theory

The probability density  $\rho(\mathbf{r})$ , more commonly known as the “electron density”, associated with a wave function  $\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$  of an  $N$ -electron system, is defined as the integral over  $N$  spin coordinates and  $(N - 1)$  spatial coordinates of the square of the wave function  $\Psi$ ,

$$\rho(\mathbf{r}_1) = N \int \cdots \int |\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)|^2 d\sigma_1 d\mathbf{x}_2 \cdots d\mathbf{x}_N \quad (1.1)$$

$\rho(\mathbf{r})$  reflects the probability per unit volume  $d\mathbf{r}$  of finding an electron of arbitrary spin, while all the other  $(N - 1)$  electrons have arbitrary positions and spins. The electron density’s fundamental characteristics follow directly from this interpretation.  $\rho(\mathbf{r})$  is always positive and decays to zero at infinity with an asymptotic behavior given by,

$$\rho(\mathbf{r}) \propto \exp \left[ -2(2I_{\min})^{\frac{1}{2}} \mathbf{r} \right] \quad (1.2)$$

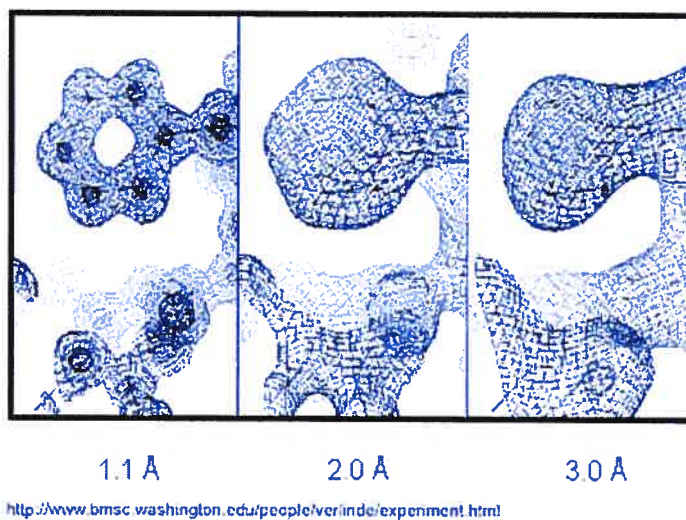


Figure 1.1: Electron density map from X-Ray diffraction of a protein side-chain containing the phenylalanine residue. The numbers at the bottom represent three different degrees of resolution. (Copyright 1997 Christophe Verlinde)

where  $I_{\min}$  is the exact first ionization potential.<sup>7</sup> Furthermore, by integration  $\rho(\mathbf{r})$  provides us with the number  $N$  of electrons of the system,

$$\int \rho(\mathbf{r}) d\mathbf{r} = N \quad (1.3)$$

Unlike the wave function, the electron density is an observable. Electron density maps, such as the one shown in Fig. 1.1, are determined from X-Ray diffraction data and play a crucial role in molecular structure determination. The position of the nuclei corresponds to a peak in the map. In fact, here, due to the positive charge  $Z_A$  of the nucleus, the electron density value reaches a maximum. At the

same point the gradient of the electron density presents a discontinuity due to the singularity of the nuclear-electron attraction term shown where  $\mathbf{r}_A \rightarrow 0$ . The cusp condition is given by:

$$\frac{\partial}{\partial \mathbf{r}_A} \bar{\rho}(\mathbf{r}_A)|_{\mathbf{r}_A=0} = -2Z_A \bar{\rho}(0) \quad (1.4)$$

where  $\bar{\rho}$  indicates the electron density spherical average.

As indicated by the equations presented above, the electron density  $\rho(\mathbf{r})$  contains essential information on the system, such as the number of electrons  $N$  and the charge and the position of the nuclei. Its central role in defining all the properties of an electronic system has been delineated by Hohenberg and Kohn in their ground-breaking paper.<sup>8</sup>

## 1.1 The Hohenberg-Kohn Theorems

The first Hohenberg-Kohn theorem recognizes that the electron density  $\rho(\mathbf{r})$  is the only function required to derive all the information concerning an arbitrary electron system. The same role was previously believed to be held exclusively by the wave function  $\Psi$ . Quoting directly from the original paper:<sup>8</sup>

*“the external potential  $V_{\text{ext}}(\mathbf{r})$  is (to within a constant) a unique functional of  $\rho(\mathbf{r})$ ; since in turn  $V_{\text{ext}}(\mathbf{r})$  fixes the Hamiltonian  $H$  we see that the full many particle ground state is a unique functional of  $\rho(\mathbf{r})$ .”*

The proof of this theorem is extremely simple and is based on *reductio ad absurdum*. From the electron density  $\rho_0(\mathbf{r})$ , corresponding to the non-degenerate ground state wave function  $\Psi$  of a generic  $N$ -electron system, we can derive the

number of electrons  $N$  from Eq.(1.3). From Eq.(1.4) we can also determine the nuclear charges  $Z_A$  and their positions. Therefore,

$$V_{\text{ext}}(\mathbf{r}) \Rightarrow V_{\text{Ne}}(\mathbf{r}) = - \sum_i^N \sum_A^M \frac{Z_A}{r_{iA}} \quad (1.5)$$

is uniquely defined. Note that the external potential  $V_{\text{ext}}$  may include other kinds of potentials in addition to the nuclear-electron Coulomb attraction potential  $V_{\text{Ne}}(\mathbf{r})$ .

The first Hohenberg-Kohn theorem's proof starts with the antithetic assumption that two external potentials,  $V_{\text{ext}}$  and  $V'_{\text{ext}}$ , correspond to the same ground-state electron density  $\rho_0(\mathbf{r})$ , although differing by more than a constant.  $V_{\text{ext}}$  and  $V'_{\text{ext}}$  fix two Hamiltonians,  $H$  and  $H'$  that differ only in the external potentials,

$$H = T + V_{ee} + V_{\text{ext}} \neq T + V_{ee} + V'_{\text{ext}} = H' \quad (1.6)$$

where  $T$  is the kinetic energy term and  $V_{ee}$  the electron-electron repulsion term.  $H$  and  $H'$  are also associated with two different ground-state eigenfunctions,  $\Psi_0$  and  $\Psi'_0$ , and therefore with two different ground-state energies,  $E_0$  and  $E'_0$ . However, these two states, defined by  $\Psi_0$  and  $\Psi'_0$  with energies  $E_0$  and  $E'_0$  respectively, lead to the same electron density  $\rho_0(\mathbf{r})$ . We can apply the variational principle using  $\Psi'_0$  as a trial function for  $H$ ,

$$E_0 \equiv \langle \Psi_0 | H | \Psi_0 \rangle < \langle \Psi'_0 | H | \Psi'_0 \rangle \quad (1.7)$$

Since  $H$  and  $H'$  differ only in the external potentials we can write,

$$E_0 < \langle \Psi'_0 | H' | \Psi'_0 \rangle + \langle \Psi'_0 | H - H' | \Psi'_0 \rangle \quad (1.8)$$

which corresponds to,

$$E_0 < E'_0 + \int \rho_0(\mathbf{r}) (V_{\text{ext}} - V'_{\text{ext}}) d\mathbf{r} \quad (1.9)$$

If we used  $\Psi_0$  as a trial function for  $H'$  we would have obtained,

$$E'_0 < E_0 + \int \rho_0(\mathbf{r}) (V'_{\text{ext}} - V_{\text{ext}}) d\mathbf{r} \quad (1.10)$$

Adding Eq.(1.9) and (1.10) we obtain the contradiction,

$$E_0 + E'_0 < E'_0 + E_0 \quad (1.11)$$

which shows the validity of the first Hohenberg-Kohn theorem. Therefore, the ground-state electron density  $\rho_0(\mathbf{r})$  fixes unambiguously the Hamiltonian of the system, thus determining the ground-state energy,  $E_0$ . In other words, the ground-state energy  $E_0$  is a functional of the ground-state electron density  $\rho_0(\mathbf{r})$ . The expression in terms of the different energy components is given by\*,

$$E_0[\rho_0] = T[\rho_0] + E_{ee}[\rho_0] + E_{\text{ext}}[\rho_0] \quad (1.12)$$

$T[\rho_0]$  is the kinetic energy functional,  $E_{ee}[\rho_0]$  the electron-electron interaction energy functional and  $E_{\text{ext}}[\rho_0]$  the external potential energy functional, which, in the absence of external perturbation fields, is reduced to the Coulomb nucleus-electron attraction energy,

$$E_{\text{ext}}[\rho_0] \xrightarrow{\text{No Ext. Field}} E_{\text{Ne}}[\rho_0] = \int V_{\text{Ne}}(\mathbf{r}) \rho_0(\mathbf{r}) d\mathbf{r} \quad (1.13)$$

$V_{\text{Ne}}(\mathbf{r})$  is defined by Eq.(1.5).  $T[\rho_0]$  and  $E_{ee}[\rho_0]$  in Eq.(1.12) together define a functional known as “Hohenberg-Kohn functional”  $F_{\text{HK}}[\rho_0]$ ,

$$F_{\text{HK}}[\rho_0] = T[\rho_0] + E_{ee}[\rho_0] \quad (1.14)$$

\*In terms of expectation value of the exact ground state wave function,

$$E_0[\rho_0] = \langle \Psi_0 | \hat{T} + \hat{V}_{ee} + \hat{V}_{\text{ext}} | \Psi_0 \rangle$$

where the operators  $\hat{T}$ ,  $\hat{V}_{ee}$  and  $\hat{V}_{\text{ext}}$  represent, respectively, the kinetic energy of the electrons, the electron-electron repulsion potential and the external potential

valid for any number of electrons and any external potential.<sup>8</sup> The explicit form of  $T[\rho_0]$  and  $E_{ee}[\rho_0]$  is not known in a practical form. However, isolating the classic Coulomb electron-electron repulsion energy component from  $E_{ee}[\rho_0]$ , we can write Eq.(1.14) as,

$$F_{\text{HK}}[\rho_0] = T[\rho_0] + \iint \frac{\rho_0(\mathbf{r}_1)\rho_0(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 + E_{\text{ncl}}[\rho_0] \quad (1.15)$$

$E_{\text{ncl}}[\rho_0]$  comprises all the non-classical contributions to the electron-electron interaction, such as the self-interaction correction, the exchange and the correlation contributions.

The Hohenberg-Kohn functional  $F_{\text{HK}}[\rho_0]$  plays a crucial role in DFT. In fact, if the explicit form of  $T[\rho_0]$  and  $E_{\text{ncl}}[\rho_0]$  were known, we could determine the ground-state energy of any system simply by inserting the ground-state density  $\rho_0(\mathbf{r})$  into the energy functional,

$$E[\rho_0] = \int V_{\text{Ne}}(\mathbf{r})\rho_0(\mathbf{r})d\mathbf{r} + F_{\text{HK}}[\rho_0] \quad (1.16)$$

Unfortunately an explicit practical form of  $T[\rho_0]$  and  $E_{\text{ncl}}[\rho_0]$  remains unknown and still represents a very active area of research (see Sec. 1.3).

The statement that the energy functional  $E[\rho]$  assumes a minimum value when the arbitrary density  $\rho(\mathbf{r})$  corresponds to the correct ground-state electron density  $\rho_0(\mathbf{r})$  is the heart of the second Hohenberg-Kohn theorem:<sup>8</sup>

“...  $E[\rho]$  assumes its minimum value for the correct  $\rho_0(\mathbf{r})$ , if the admissible functions are restricted by the condition

$$N[\rho] \equiv \int \rho_0(\mathbf{r})d\mathbf{r} = N ”$$



The proof of this theorem relies on the variational principle. In particular, we have seen from the first Hohenberg-Kohn theorem that the electron density  $\bar{\rho}(\mathbf{r})$  corresponds to the external potential  $\tilde{V}_{\text{ext}}(\mathbf{r})$ , hence,

$$\bar{\rho}(\mathbf{r}) \Rightarrow \tilde{V}_{\text{ext}}(\mathbf{r}) \Rightarrow \tilde{H} \Rightarrow \tilde{\Psi} \quad (1.17)$$

$\tilde{\Psi}$  can be used as trial function for the Hamiltonian  $H$  generated by the true external potential  $V_{\text{ext}}(\mathbf{r})$ ,

$$\langle \tilde{\Psi} | H | \tilde{\Psi} \rangle = T[\bar{\rho}] + V_{ee}[\bar{\rho}] + \int \bar{\rho}(\mathbf{r}) V_{\text{ext}} d\mathbf{r} = E[\bar{\rho}] \quad (1.18)$$

$E[\bar{\rho}]$  represents an upper bound of the true ground state energy,

$$E[\bar{\rho}] \geq E_0[\rho_0] = \langle \Psi_0 | H | \Psi_0 \rangle \quad (1.19)$$

thus proving the hypothesis.

Now, the first Hohenberg-Kohn theorem demonstrates the existence of a one-to-one mapping between the ground-state electron density  $\rho_0(\mathbf{r})$  and the external potential  $V_{\text{ext}}(\mathbf{r})$ , and hence also between the ground-state electron density  $\rho_0(\mathbf{r})$  and the ground-state wave function  $\Psi_0$ . Furthermore, the second Hohenberg-Kohn theorem states that the ground-state energy  $E_0$  is a functional of the ground-state electron density  $\rho_0(\mathbf{r})$  and that it can be obtained through a variational scheme.

One problem arising at this point is that, while the ground-state density can be easily obtained from the ground-state wave function by integration [see Eq.(1.1)], the opposite is clearly not true. In fact, from the same electron density we can derive an infinite number of different wave functions. The methodology that allows us to “search” among the different wave functions for the ground-state wave function is the subject of the following section.

### 1.1.1 The Levy Constrained Search

The issue of adapting the variational principle,

$$E_0 = \min_{\Psi \rightarrow N} \langle \Psi | T + V_{ee} + V_{Ne} | \Psi \rangle \quad (1.20)$$

to DFT has been investigated by Levy,<sup>9,10</sup> by Lieb<sup>11</sup> and by Levy and Perdew.<sup>12</sup> The resulting approach is known as the “Levy constrained search” and consists of two main steps. In the first step the electron density  $\check{\rho}(\mathbf{r})$  is chosen and a search for the energy minimum is conducted over all the wave functions  $\check{\Psi}$  that give  $\check{\rho}(\mathbf{r})$  by quadrature,

$$\check{E}_{\min} = \min_{\check{\Psi} \rightarrow \check{\rho}} \langle \check{\Psi} | T + V_{ee} + V_{Ne} | \check{\Psi} \rangle \quad (1.21)$$

In the second step the search is extended to all the electron densities  $\rho(\mathbf{r})$  which integrate to the number of electrons of the system, i.e. to all the N-representable electron densities ( $\check{\rho}(\mathbf{r}) \in \rho(\mathbf{r})$ ),

$$E_0 = \min_{\rho \rightarrow N} \left[ \min_{\Psi \rightarrow \rho} \langle \Psi | T + V_{ee} + V_{Ne} | \Psi \rangle \right] \quad (1.22)$$

Since the external potential is fixed by the electron density, it can be excluded from the first step of the search,

$$E_0 = \min_{\rho \rightarrow N} \left[ \min_{\Psi \rightarrow \rho} \langle \Psi | T + V_{ee} | \Psi \rangle + \int \rho(\mathbf{r}) V_{Ne}(\mathbf{r}) d\mathbf{r} \right] \quad (1.23)$$

Hence, in analogy with Eq.(1.14) we can define a universal functional of the type,

$$F[\rho] = \min_{\Psi \rightarrow \rho} \langle \Psi | T + V_{ee} | \Psi \rangle \quad (1.24)$$

which corresponds to the Hohenberg-Kohn functional  $F_{HK}[\rho_0]$  exclusively when the electron density  $\rho$  corresponds to the ground-state electron density  $\rho_0$ . In terms of the universal functional  $F[\rho]$  Eq.(1.23) becomes,

$$E_0 = \min_{\rho \rightarrow N} \left[ F[\rho] + \int \rho(\mathbf{r}) V_{Ne}(\mathbf{r}) d\mathbf{r} \right] \quad (1.25)$$

which represents the DFT counterpart of the variational principle in Eq.(1.20).

Unfortunately, the implementation of the Levy constrained search in any practical approach is impossible because of two main obstacles. The first is related to the fact that for a given electron density  $\rho(\mathbf{r})$  the lowest energy is determined upon a search over all of the wave functions that give  $\rho(\mathbf{r})$  by quadrature. Within a finite basis set this implies a full configuration-interaction (CI) calculation constrained to give a particular density, plus a scan density. This approach becomes simply impossible for a complete basis set. The second problem is represented by the explicit form of the universal functional  $F[\rho]$ . The legitimacy of the Levy constrained search is bound to the use of the exact expression for  $F[\rho]$ . As it has been underlined, this exact expression is known but impractical, hence it is important to find a suitable approximation for  $F[\rho]$ .

## 1.2 The Kohn-Sham Approach

In the second historic paper for the beginning of DFT, Kohn and Sham<sup>13</sup> revealed a simple way to circumvent the problems created by the infinite number of wave functions obtainable from the electron density  $\rho_0$  and from the unknown form of the universal functional  $F[\rho]$ .

The strategy is based on the idea that the N-electron system can be described by a set of exact equations for N *non*-interacting electrons. These N one-electron equations are known as “Kohn-Sham” (KS) equations,

$$\left[ -\frac{1}{2}\nabla_i^2 + V_{\text{eff}}(\mathbf{r}) \right] \psi_i = \epsilon_i \psi_i \quad (1.26)$$

$\psi_i$  are the lowest eigenstates, with relative eigenvalues  $\epsilon_i$ , which build the ground-state wave function  $\Psi = (N!)^{-1/2} |\psi_1 \psi_2 \cdots \psi_N\rangle$  and are referred to as KS or-

bitals. The KS orbitals are determined under the constraint that the corresponding wave function  $\Psi$  gives by quadrature the exact ground-state electron density of the system of  $N$  fully interacting particles,

$$\rho_0(\mathbf{r}) = \sum_i^N \sum_{\sigma} |\psi_i(\mathbf{r}, \sigma)|^2 \quad (1.27)$$

In order to accomplish this objective the one-electron potential  $V_{\text{eff}}$  is constructed as,

$$V_{\text{eff}}(\mathbf{r}) = V_{\text{ext}}(\mathbf{r}) + \int \frac{\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r} + V_{\text{xc}}(\mathbf{r}) \quad (1.28)$$

where the external potential  $V_{\text{ext}}$  includes the nucleus-electron attraction and may also include the interaction with other external fields. The major innovation of the KS approach is related to the introduction of the term  $V_{\text{xc}}$ , known as the exchange-correlation potential,

$$V_{\text{xc}}(\mathbf{r}) = \frac{\partial (T[\rho] - T_s[\rho])}{\partial \rho(\mathbf{r})} + \frac{\partial (E_{\text{ee}}[\rho] - J[\rho])}{\partial \rho(\mathbf{r})} = \frac{\partial T_{\text{C}}[\rho]}{\partial \rho(\mathbf{r})} + \frac{\partial E_{\text{ncl}}[\rho]}{\partial \rho(\mathbf{r})} \quad (1.29)$$

$V_{\text{xc}}$  includes the correlation contribution to the kinetic energy  $T_{\text{C}}[\rho]$ , arising from treating the fully interacting system within a non-interacting framework. Moreover,  $V_{\text{xc}}$  also takes into account the non-classical portion of the electron-electron interaction energy with the term  $E_{\text{ncl}}[\rho]$ , which is analogous to the one encountered in the definition of the Hohenberg-Kohn functional  $F_{\text{HK}}[\rho_0]$  [see Eq.(1.15)].

The great advantage of the KS approach resides in the separation of the kinetic energy term in two. The first and largest contribution,

$$T_s[\rho] = \sum_i^N \left\langle \psi_i \left| -\frac{1}{2} \nabla_i^2 \right| \psi_i \right\rangle \quad (1.30)$$

is exact for  $N$  non-interacting electrons. The exact expression for the kinetic contribution to the  $E_{xc}$ ,  $T_C[\rho]$ , is unknown and very few studies have been done on the determination of its exact value. It appears, however, that the  $T_C[\rho]$  represents only a small fraction of the total kinetic energy  $T[\rho]$ .<sup>14-18</sup> Earlier DFT based models, such as the “Thomas-Fermi” approach,<sup>19,20</sup> were based on the approximation of the kinetic energy functional for the fully interacting system  $T[\rho]$ , which results in significant inaccuracy of the results.

The ensemble of Eq.(1.26), (1.28) and (1.29) indicates that the KS methodology is formally exact since for its derivation no approximations have been used. This implies that if the correct expression for the exchange-correlation potential  $V_{xc}[\rho]$  were known, the KS approach would lead directly to the exact density and to the exact energy.

### 1.3 Approximate Exchange-Correlation Functionals

Needless to say, the drawback of the KS method is precisely that a practical exact form of the exchange-correlation energy functional  $E_{xc}[\rho]$  still represents the holy grail for most researchers involved in DFT. Furthermore, the obscurity enclosing the exact  $E_{xc}[\rho]$  makes impossible a systematic approach for generating improved approximations.

The path usually followed in the pursuit of the best approximate exchange correlation functional is dictated by two main constraints. The first entails a direct validation through standard testing procedure, such as the “Gaussian- $n$ ” method introduced by Pople and coworkers.<sup>21-24</sup> In particular, the G1<sup>21,22</sup> and G2<sup>23</sup> databases are based on the calculation of thermochemical data. The second constraint,

although less stringent, is represented by the reproduction of the physical characteristics of what is known as the “exchange-correlation hole”  $h_{xc}(\mathbf{x}_1, \mathbf{x}_2)$ .

### 1.3.1 The Exchange-Correlation Hole

In order to comprehend thoroughly the significance of the exchange-correlation hole, this digression will start with the introduction of the pair density,

$$\rho_2(\mathbf{x}_1, \mathbf{x}_2) = \frac{N(N-1)}{2} \int \cdots \int |\Psi(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N)|^2 d\mathbf{x}_3 \cdots d\mathbf{x}_N \quad (1.31)$$

normalized to the number of distinct pairs of electrons. In analogy to the probabilistic interpretation of the electron density, the pair density  $\rho_2(\mathbf{x}_1, \mathbf{x}_2)$  represents the probability of finding two electrons with arbitrary spins  $\sigma_1$  and  $\sigma_2$  within two volume elements  $d\mathbf{r}_1$  and  $d\mathbf{r}_2$ .

For charged fermions, such as electrons, the probability of finding one of two electrons is highly dependent on the position of the other, in other words their motion is correlated. In fact, the nature of fermions hinders the approach of two electrons with the same spin, while the fact that they are also charged particles implies that they repel each other through a Coulomb-type potential. The former type of non-classical interaction is defined as “exchange” or “Fermi” correlation, while the latter is known as “Coulomb” correlation or simply as electron correlation. The effect of these two types of correlation can be expressed through the pair density defined as,

$$\rho_2(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{2} \rho(\mathbf{x}_1) \rho(\mathbf{x}_2) [1 + f(\mathbf{x}_1, \mathbf{x}_2)] \quad (1.32)$$

where  $f(\mathbf{x}_1, \mathbf{x}_2)$  is known as a “pair correlation function” and incorporates the effects of Fermi and Coulomb correlation in the interaction between the two electrons. When  $f(\mathbf{x}_1, \mathbf{x}_2)$  is set to zero, Eq.(1.32) reflects a completely uncorrelated

motion for the two electrons. In addition, it is important to mention that in the latter case Eq.(1.32) would also incorporate an unphysical self-interaction due to the fact that the electron densities  $\rho(\mathbf{x}_1)$  and  $\rho(\mathbf{x}_2)$  are each normalized to the number of electrons  $N$ .

The pair correlation function  $f(\mathbf{x}_1, \mathbf{x}_2)$  satisfies the fundamental relation,

$$\int \rho(\mathbf{x}_2) f(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2 = -1 \quad (1.33)$$

known as the “sum rule”. The interpretation of this relation becomes more clear when related to the function,

$$h_{xc}(\mathbf{x}_1, \mathbf{x}_2) = \rho(\mathbf{x}_2) f(\mathbf{x}_1, \mathbf{x}_2) \quad (1.34)$$

which represents the “exchange-correlation hole”. The term “hole” function is intended to represent metaphorically the depletion of electron density at  $\mathbf{x}_2$  due to the non-classical electron-electron interactions. When Eq.(1.33) is expressed in terms of an exchange-correlation hole,

$$\int h_{xc}(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2 = -1 \quad (1.35)$$

we understand that the exchange-correlation hole integrates exactly to minus the unit charge. The sum rule is only one of the physical characteristics of the hole function which should also be satisfied by an approximate exchange-correlation functional.

The total exchange-correlation hole  $h_{xc}(\mathbf{x}_1, \mathbf{x}_2)$  may be considered as the sum of the hole functions,

$$h_{xc}(\mathbf{x}_1, \mathbf{x}_2) = h_x^{\uparrow\uparrow}(\mathbf{x}_1, \mathbf{x}_2) + h_x^{\downarrow\downarrow}(\mathbf{x}_1, \mathbf{x}_2) + h_c^{\uparrow\downarrow}(\mathbf{x}_1, \mathbf{x}_2) + h_c^{\downarrow\uparrow}(\mathbf{x}_1, \mathbf{x}_2) \quad (1.36)$$

which represent the contribution of the Fermi ( $\uparrow\uparrow$  and  $\downarrow\downarrow$ ) and of the Coulomb correlation ( $\uparrow\downarrow$  and  $\downarrow\uparrow$ ). Although only the total hole has a real physical meaning,

this separation is useful in order to identify and reproduce the specific characteristics of the total hole function, which is the objective in modeling all approximated exchange-correlation functionals. In particular, only to outline some of the fundamental characteristics, the Fermi hole represents the largest contribution to the total hole. As the total hole integrates to minus the unit charge,

$$\int h_x(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2 = -1 \quad (1.37)$$

and it is negative everywhere,

$$h_x(\mathbf{x}_1, \mathbf{x}_2) < 0 \quad (1.38)$$

Whereas, according to Eq.(1.35), the Coulomb hole satisfies the relation,

$$\int h_c(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2 = 0 \quad (1.39)$$

and its shape is determined by the fact that it is generated by the Coulomb  $r_{12}^{-1}$  interaction. Hence, it presents a cusp at  $\mathbf{x}_1$  and it is negative in proximity of  $\mathbf{x}_1$ , but, in accord with Eq.(1.39), is positive elsewhere.

A more detailed description of the exchange-correlation hole would be outside the purpose of this work. For a comprehensive review the articles by Perdew and Burke<sup>25</sup> and by Adamo and coworkers<sup>26</sup> are suggested.

### 1.3.2 The Adiabatic Connection

According to Eq.(1.29), the exchange-correlation potential is made up not only by the non-classical electron-electron interaction potential  $V_{\text{ncl}}$ , but also by another contribution coming from the difference between the kinetic energy potential of the fully interacting system  $T[\rho]$  and the kinetic energy potential of the KS non-interacting system  $T_s[\rho]$ . In order to determine the best possible approximation for



the exact exchange-correlation functional, the kinetic energy contribution has to be included. The exchange-correlation hole function  $h_{xc}$  as such does not contain any information about the kinetic energy, which, in fact, is not a quantity related to the pair density.

A simple way to account for both the kinetic energy contribution and the non-classical component of the electron-electron interaction energy is given by the “adiabatic connection”.<sup>27–33</sup> In accord with Eq.(1.29), the exchange-correlation functional can be expressed as,

$$E_{xc}[\rho] = (T[\rho] - T_s[\rho]) + (E_{ee}[\rho] - J[\rho]) \quad (1.40)$$

It is also possible to imagine that the variation between a non-interacting system to a fully-interacting system takes place progressively. In analogy with the thermodynamic concept of an adiabatic process, we could imagine that the transformation takes place through an adiabatic path. Within this scenario, the universal functional in Eq.(1.24) can be expressed in terms of a coupling strength parameter  $\lambda$  varying from 0 to 1,

$$F_\lambda[\rho] = \min_{\Psi \rightarrow \rho} \langle \Psi | T + \lambda V_{ee} | \Psi \rangle \quad (1.41)$$

When  $\lambda = 1$ , Eq.(1.41) reflects the picture of a fully interacting system,

$$F_1[\rho] = T[\rho] + E_{ee}[\rho] \quad (1.42)$$

while for  $\lambda = 0$  we have a non-interacting system,

$$F_0[\rho] = T_s[\rho] \quad (1.43)$$

In agreement with Eq.(1.40), (1.41) and (1.42), we can define,

$$E_{xc}[\rho] = \int_0^1 \frac{\partial F_\lambda[\rho]}{\partial \lambda} d\lambda - J[\rho] \quad (1.44)$$

From Eq.(1.41), we obtain,

$$\frac{\partial F_\lambda[\rho]}{\partial \lambda} = \min_{\Psi \rightarrow \rho} \langle \Psi | V_{ee} | \Psi \rangle \quad (1.45)$$

Hence, we can express Eq.(1.44) in terms of the pair density,

$$E_{xc}[\rho] = \iint \frac{\bar{\rho}_2(\mathbf{x}_1, \mathbf{x}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 - J[\rho] \quad (1.46)$$

with  $\bar{\rho}_2(\mathbf{x}_1, \mathbf{x}_2)$  defined as the coupling-strength averaged pair density,

$$\bar{\rho}_2(\mathbf{x}_1, \mathbf{x}_2) = \int_0^1 \rho_2(\mathbf{x}_1, \mathbf{x}_2) d\lambda \quad (1.47)$$

In agreement with Eq.(1.32) and (1.34),  $E_{xc}[\rho]$  can be expressed by,

$$E_{xc}[\rho] = \frac{1}{2} \iint \frac{\rho(\mathbf{x}_1) \bar{h}_{xc}(\mathbf{x}_1, \mathbf{x}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 \quad (1.48)$$

where  $\bar{h}_{xc}(\mathbf{x}_1, \mathbf{x}_2)$  is the coupling-strength integrated exchange-correlation hole,

$$\bar{h}_{xc}(\mathbf{x}_1, \mathbf{x}_2) = \int_0^1 h_{xc}(\mathbf{x}_1, \mathbf{x}_2) d\lambda \quad (1.49)$$

The integration process in Eq.(1.49) is responsible for the transfer of the kinetic energy component information into the exchange-correlation hole function.

### 1.3.3 The Local Density Approximation

One of the guidelines to follow in order to obtain a good approximation for the exchange-correlation functional  $E_{xc}[\rho]$  is that the model exchange-correlation hole should mimic as closely as possible the physical characteristics of the coupling-strength integrated hole  $\bar{h}_{xc}$ , rather than of  $h_{xc}$ . The key of the success of the “local density approximation” (LDA) for  $E_{xc}[\rho]$  lies primarily in that.

The LDA, or “local spin-density approximation” (LSD) in the unrestricted case, represents the foundation of all modern approximate exchange-correlation functionals,

$$E_{xc}^{LDA}[\rho] = \int \rho(\mathbf{r}) \epsilon_{xc}(\rho(\mathbf{r})) d\mathbf{r} \quad (1.50)$$

$$E_{xc}^{LSD}[\rho_\alpha, \rho_\beta] = \int \rho(\mathbf{r}) \epsilon_{xc}(\rho_\alpha(\mathbf{r}), \rho_\beta(\mathbf{r})) d\mathbf{r} \quad (1.51)$$

It is, in fact, the only functional for which a nearly exact explicit form is known. The approximate  $E_{xc}[\rho]$  is based on a uniform electron gas model, where the term  $\epsilon_{xc}(\rho(\mathbf{r}))$  represents the exchange-correlation energy per particle of the uniform electron gas of density  $\rho(\mathbf{r})$ . The  $\epsilon_{xc}(\rho(\mathbf{r}))$  can be considered as the sum of exchange and correlation energy contributions,

$$\epsilon_{xc}(\rho(\mathbf{r})) = \epsilon_x(\rho(\mathbf{r})) + \epsilon_c(\rho(\mathbf{r})) \quad (1.52)$$

The exchange only term, commonly referred to as “Slater exchange” (S), is known exactly,<sup>34,35</sup>

$$\epsilon_x^S(\rho(\mathbf{r})) = -\frac{3}{4} \left( \frac{3\rho(\mathbf{r})}{\pi} \right)^{\frac{1}{3}} \quad (1.53)$$

while there is not such an exact expression for the correlation contribution  $\epsilon_c(\rho(\mathbf{r}))$ . Two of the most successful approximations for this term have been obtained by Vosko, Wilk and Nusair.<sup>36</sup> The first is based on the random phase approximation (RPA) and the second on an interpolation scheme of the highly accurate results of a uniform electron gas quantum Monte Carlo simulation by Ceperley and Alder.<sup>37</sup>

The idea behind the LDA is that the exchange-correlation energy of a uniform electron gas of density  $\rho(\mathbf{r})$  at the position  $\mathbf{r}$  represents a good approximation for the exchange-correlation energy of a real system with electron density  $\rho(\mathbf{r})$  at

position  $\mathbf{r}$ . Although, the integration over all  $\mathbf{r}$  in Eq.(1.50) reproduces the situation better for a system with slowly varying density, such as simple metals, than the highly anisotropic density of atoms and molecules. Hence, in the assessment of the potentials of the LDA against the reference experimental values for atomization energies, an average deviation of  $35.7 \text{ kcal mol}^{-1}$  has been determined.<sup>38</sup> This result represents a considerable improvement if compared to the HF model, which in the same test scores a disappointing  $85.9 \text{ kcal mol}^{-1}$ , however, with regard to bond energies, we are still very far from an acceptable level of chemical accuracy †.

### 1.3.4 The Generalized Gradient Approximation

In order to correct the erroneous behavior of the LDA, it has been found that the expression for  $E_{xc}[\rho]$  needs to include some information on the homogeneity of the electron density in the system. Several different functionals containing the gradient of the electron density  $\nabla\rho(\mathbf{r})$  as a homogeneity parameter have been developed. The explicit form of these functionals derives from the specific constraints imposed on the gradient expansion with the purpose of making the model hole function reflect the physical characteristics of the correct exchange-correlation hole  $\bar{h}_{xc}$ . All of these functionals come from a model known as the “generalized gradient approximation” (GGA) and can be collectively indicated by the expression,

$$E_{xc}^{GGA}[\rho; \nabla\rho] = \int f(\rho; \nabla\rho) d\mathbf{r} \quad (1.54)$$

---

†According to the G2 method, chemical accuracy is defined as an average absolute error of  $\pm 0.1 \text{ eV}$  or  $\pm 2 \text{ kcal mol}^{-1}$

Usually, exchange and correlation contributions are considered and developed separately,

$$E_{xc}^{GGA}[\rho; \nabla\rho] = E_x^{GGA}[\rho; \nabla\rho] + E_c^{GGA}[\rho; \nabla\rho] \quad (1.55)$$

Once the origin of the LDA problems was identified in the exchange energy term,<sup>39-44</sup> particular attention has been focused on the development of the exchange only component,

$$E_x^{GGA}[\rho; \nabla\rho] = -\frac{3}{4} \left(\frac{3}{\pi}\right)^{1/3} \int F(s(\mathbf{r})) \rho^{4/3}(\mathbf{r}) d\mathbf{r} \quad (1.56)$$

where  $F$  is a function of the “reduced density gradient”  $s$ ,

$$s(\mathbf{r}) = \frac{|\nabla\rho(\mathbf{r})|}{\rho^{4/3}(\mathbf{r})} \quad (1.57)$$

It is important to note that if the value  $s = 0$  ( $F(s) = 1$ ), corresponding to the homogeneous electron gas model, is inserted in Eq.(1.56), we obtain the expression for the Slater exchange.

Within the GGA framework, numerous different types of expressions for  $F(s)$  have been developed<sup>26</sup> and, among the corresponding exchange functionals, I would like to mention the work of Becke for the B88 (or B),<sup>45</sup> of Perdew<sup>46</sup> and of Burke and coworkers for the PW91,<sup>47</sup> of Hamprecht and coworkers for the HCTH,<sup>48</sup> of Becke for the B86,<sup>49</sup> of Perdew and Yue for the PW86<sup>50</sup> and finally the work of Perdew and coworkers for the PBE.<sup>51</sup>

As for the correlation component  $E_c^{GGA}$ , various approximations are also available and among them the most successful have been the correlation counterpart of the PW86 exchange, named P86,<sup>52</sup> the correlation counterpart of the PW91 exchange<sup>53</sup> and the LYP correlation component from Lee and coworkers.<sup>54</sup>

The gradient corrected exchange-correlation functional represents a very significant improvement over the LDA and the major cause of this is the correction

of the exchange-only component. In the determination of the atomization energies, the exchange-only B88 functional in combination with the VWN correlation component lowers the average absolute deviation to 4.4 kcal mol<sup>-1</sup>, from 35.7 kcal mol<sup>-1</sup> calculated for SVWN.<sup>38</sup> However, the effect of the introduction of a correlation term  $E_c^{\text{GGA}}$  is rather modest. In fact, the average absolute deviation determined for a “pure” GGA functional, such as BLYP, in the atomization energies test is 5.6 kcal mol<sup>-1</sup>.<sup>38</sup>

### 1.3.5 Hybrid Functionals

The key to improving the performance of the approximate exchange-correlation functionals lies mainly in the determination of a suitable exchange-only component. It is known that within the HF framework, the exchange term can be computed exactly as,

$$E_x^{\text{HF}}[\{\chi_{i,\sigma}\}] = -\frac{1}{2} \sum_{\sigma=\alpha,\beta} \sum_{i,j} \iint \chi_{i,\sigma}^*(\mathbf{x}_1) \chi_{j,\sigma}(\mathbf{x}_1) \mathbf{r}_{12}^{-1} \chi_{i,\sigma}(\mathbf{x}_2) \chi_{j,\sigma}^*(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \quad (1.58)$$

where  $\chi_{i,\sigma}$  are the HF spin orbitals. In principle the exact exchange  $E_x^{\text{exact}}[\rho]$ , calculated according to Eq.(1.58) but with the KS spin orbitals  $\psi_{i,\sigma}$  instead, could take the place of the approximate exchange only term in the approximate exchange-correlation functional  $E_{xc}[\rho]$ . However, this approach does not work for molecular systems simply because the *non*-local nature of the exact exchange combined with a local correlation hole fails to reproduce the characteristics of the correct  $h_{xc}$ .

A good compromise is represented by the inclusion of only an opportune fraction of the exact exchange.<sup>55,56</sup> This approach finds its theoretical justification in the adiabatic connection discussed in Sec. 1.3.2. Accordingly, we have seen that

$E_{xc}[\rho]$  can be expressed as,

$$E_{xc}[\rho] = \int_0^1 E_{ncl}^\lambda[\rho] d\lambda \quad (1.59)$$

where  $\lambda$  was the coupling strength parameter. For  $\lambda = 1$  the electron system is fully interacting through the Coulomb potential  $V_{ee}$ . However, for  $\lambda = 0$  no interaction is included and the electrons are treated as *non*-charged fermions. Hence, only the exact exchange term is included.

In the hypothesis that  $E_{ncl}^\lambda$  is a linear function of the parameter  $\lambda$ , a straightforward solution is given by,

$$E_{xc}[\rho] = \frac{1}{2} E_x^{\text{exact}}[\rho] + \frac{1}{2} E_{xc}^{\text{LDA}}[\rho] \quad (1.60)$$

This approach is known as “half-and-half” (HH)<sup>55</sup> and its performance against the G2 thermochemical databank is comparable to that of the GGA functionals. Even better results are obtained by adding more flexibility through the introduction of fitting parameters,<sup>56</sup>

$$E_{xc}^{\text{B3PW91}} = E_{xc}^{\text{LSD}} + a (E_x^{\text{exact}} - E_x^{\text{LSD}}) + b \Delta E_x^{\text{B}} + c \Delta E_c^{\text{PW91}} \quad (1.61)$$

This functional, known as B3PW91, reduces the average absolute deviation against G2 thermochemical reference data to a level near the order of chemical accuracy (2-3 kcal mol<sup>-1</sup>).

However, an even lower average absolute deviation<sup>57</sup> is obtained with the B3LYP “combination” introduced by Stevens and coworkers.<sup>58</sup> In fact, it turns out that this functional is now one of the most successful functionals in use.<sup>59</sup> B3LYP is analogous to B3PW91 but instead of the  $E_c^{\text{PW91}}$  it uses the LYP correlation component,

$$E_{xc}^{\text{B3LYP}} = E_{xc}^{\text{LSD}} + a (E_x^{\text{exact}} - E_x^{\text{LSD}}) + b \Delta E_x^{\text{B}} + c (E_c^{\text{LYP}} - E_c^{\text{LSD}}) \quad (1.62)$$

The values of the  $\alpha$ ,  $b$  and  $c$  coefficients, derived for  $E_{xc}^{B3PW91}$  and used also in  $E_{xc}^{B3LYP}$ , are obtained by least squares fitting to the G1 thermochemical data.  $\alpha$  represents the fraction of the uniform electron gas exchange to be replaced with the exact exchange and is set to 0.20, while,  $b = 0.72$  and  $c = 0.81$  fix the portion of B88 exchange and PW91, or LYP, correlation respectively.

The incorporation into the hybrid functionals of the aforementioned adjustable parameters, has allowed the development of some *ad hoc* functionals. Noteworthy, the  $B3LYP_{GGA}^{0.05}$  and the  $B96_{GGA}^{0.04}$ , which include only 5% and 4% of the exact exchange respectively, developed specifically for NMR shielding calculations.<sup>60</sup>

The inclusion of 20 to 25% of exact exchange has been evaluated as appropriate on the basis of a theoretical justification in the works by Perdew and coworkers,<sup>61</sup> by Ernzerhof,<sup>62</sup> and by Ernzerhof and coworkers,<sup>44</sup> which introduced the parameter-free PBE0<sup>‡</sup> hybrid functional,

$$E_{xc}^{PBE0} = E_{xc}^{PBE} + 0.25 (E_x^{exact} - E_x^{PBE}) \quad (1.63)$$

This functional shows very good results in the calculation of a broad set of properties<sup>63–65</sup> and it is particularly accurate in determining NMR shielding parameters.<sup>66,67</sup>

### 1.3.6 Meta-GGA Functionals

The effort to ameliorate the performance of the GGA functionals takes a different path in the development of the functionals based on the “meta-generalized gradient approximation” or meta-GGAs.<sup>68–74</sup> Within the meta-GGA framework, the exchange-correlation energy functional includes information also on the kinetic

<sup>‡</sup>Also known as PBE1PBE



energy density of the KS orbitals,

$$\tau_{\uparrow,\downarrow} = \frac{1}{2} \sum_i^{\text{Nocc}} |\nabla \psi_{i,\uparrow,\downarrow}|^2 \quad (1.64)$$

besides the dependence on the gradient of the density already considered in GGAs.

The  $E_x^{\text{mGGA}}$  term can be written in the general spin restricted case as,

$$E_x^{\text{mGGA}}[\rho] = -\frac{3}{4} \left(\frac{3}{\pi}\right)^{1/3} \int f(\rho, \nabla\rho, \tau) \rho^{4/3}(\mathbf{r}) d\mathbf{r} \quad (1.65)$$

The dependence on the kinetic energy density  $\tau$  is also an ingredient of the correlation energy,<sup>69</sup> hence an expression analogous to Eq.(1.65) can be written for  $E_c^{\text{mGGA}}[\rho]$  as well. Among the available meta-GGA correlation energy functionals, I would like to highlight the LAP functionals<sup>75-80</sup> and specifically LAP1,<sup>75,76</sup> LAP3<sup>79</sup> and  $\tau 1$ .<sup>80</sup>

The addition of the dependence on the Laplacian of the density ameliorates considerably the accuracy; for instance, the average absolute deviation in the G2 test is lowered to 3.6 kcal mol<sup>-1</sup> for the meta-GGA from Perdew and coworkers<sup>70</sup> compared to 7.85 kcal mol<sup>-1</sup> of the PBE GGA. However, the shortcoming of the meta-GGA approach compared to the hybrid functional scheme, is that the calculation of the kinetic energy dependence of the exchange-correlation functional makes the SCF procedure computationally more demanding.

### 1.3.7 Optimized Effective Potential

Another strategy developed in order to introduce the exact-exchange energy functional  $E_x^{\text{exact}}$  in the KS equation is represented by the “optimized effective potential” (OEP) approach.<sup>81,82</sup> The OEPs  $V_\sigma^0$ , where  $\sigma = \alpha, \beta$ , are defined as

spin-polarized potentials which minimize the total energy expression,

$$E_{\text{tot}} = \sum_{i,\sigma} f_{i,\sigma} \left\langle \psi_{i,\sigma}^0 \left| -\frac{1}{2} \nabla^2 \right| \psi_{i,\sigma}^0 \right\rangle + \int V_{\text{ext}}(\mathbf{r}) \rho^0(\mathbf{r}) d\mathbf{r} + \frac{1}{2} \iint \frac{\rho^0(\mathbf{r}_1) \rho^0(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 + E_x^{\text{HF}}[\{\psi_{i,\sigma}^0\}] \quad (1.66)$$

$$\left( \text{where: } \rho^0(\mathbf{r}) = \sum_{i,\sigma} f_{i,\sigma} \psi_{i,\sigma}^{0*}(\mathbf{r}) \psi_{i,\sigma}^0(\mathbf{r}) \right) \quad (1.67)$$

by determining the optimized eigenfunctions  $\psi_{i,\sigma}^0$ . The minimization condition is expressed by,

$$\frac{\partial E_{\text{tot}}}{\partial V_{\sigma}^0(\mathbf{r})} = \sum_{i,\sigma} \int \frac{\partial E_{\text{tot}}}{\partial \psi_{i,\sigma}^0(\mathbf{r}')} \frac{\partial \psi_{i,\sigma}^0(\mathbf{r}')}{\partial V_{\sigma}^0(\mathbf{r})} d\mathbf{r}' + \text{const.} = 0 \quad (1.68)$$

where the optimized orbitals  $\psi_{i,\sigma}^0$  are the eigenfunctions of,

$$\left[ -\frac{1}{2} \nabla^2 + V_{\sigma}^0(\mathbf{r}) \right] \psi_{i,\sigma}^0(\mathbf{r}) = \epsilon_{i,\sigma}^0 \psi_{i,\sigma}^0(\mathbf{r}) \quad (1.69)$$

In the exchange-only case,

$$V_{\sigma}^0(\mathbf{r}) = V_{\text{ext}}(\mathbf{r}) + \int \frac{\rho^0(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_2 + V_{x\sigma}^0(\mathbf{r}) \quad (1.70)$$

Substituting the expression in Eq.(1.66) for the total energy into the minimization condition in Eq.(1.68), we obtain,

$$\frac{\partial E_{\text{tot}}}{\partial \psi_{i,\sigma}^0(\mathbf{r}')} = f_{i,\sigma} \left\{ -\frac{1}{2} \nabla^2 \psi_{i,\sigma}^{0*}(\mathbf{r}') + [V_{\text{ext}}(\mathbf{r}') + V_{\text{Coul}}(\mathbf{r}') + v_{i,\sigma}(\mathbf{r}')] \right\} \psi_{i,\sigma}^{0*}(\mathbf{r}') = f_{i,\sigma} [\epsilon_{i,\sigma}^0 - V_{x\sigma}^0(\mathbf{r}') + v_{i,\sigma}(\mathbf{r}')] \psi_{i,\sigma}^{0*}(\mathbf{r}') \quad (1.71)$$

where,

$$v_{i,\sigma}(\mathbf{r}) = \frac{1}{f_{i,\sigma} \psi_{i,\sigma}^{0*}(\mathbf{r})} \frac{\partial E_x^{\text{HF}}}{\partial \psi_{i,\sigma}^0(\mathbf{r})} \quad (1.72)$$

Moreover, still from Eq.(1.68),

$$\frac{\partial \psi_{i,\sigma'}^0(\mathbf{r}')}{\partial V_{\sigma}^0(\mathbf{r})} = -G_{i,\sigma}^0(\mathbf{r}', \mathbf{r}) \psi_{i,\sigma}^0(\mathbf{r}) \delta_{\sigma,\sigma'} \quad (1.73)$$

where  $G_{i,\sigma}^0$  represents the Green's function for non-interacting electrons moving in the potential  $V_{\sigma}^0(\mathbf{r})$ . In view of the expressions given by Eq.(1.71) and (1.73), the OEP equations are,

$$\sum_i f_{i,\sigma} \int [V_{x\sigma}^0(\mathbf{r}') - v_{i,\sigma}(\mathbf{r}')] \times G_{i,\sigma}^0(\mathbf{r}', \mathbf{r}) \psi_{i,\sigma}^{0*}(\mathbf{r}') \psi_{i,\sigma}^0(\mathbf{r}) d\mathbf{r}' + c. = 0 \quad (1.74)$$

Since the exact analytical expression for the OEP  $V_{x\sigma}^0(\mathbf{r})$  is unknown, Eq.(1.74) has to be solved numerically and therefore its application is limited to rather simple atomic systems.

An approximate analytical form for the OEP has been proposed by Krieger and coworkers,<sup>83,84</sup> namely the KLI-OEP,

$$V_{x\sigma}^{\text{KLI}}(\mathbf{r}) = V_{x\sigma}^{\text{S}}(\mathbf{r}) + \sum_i \frac{\rho_{i,\sigma}(\mathbf{r})}{\rho_{\sigma}(\mathbf{r})} (\bar{V}_{xi,\sigma}^{\text{KLI}} - \bar{v}_{i,\sigma}) \quad (1.75)$$

where  $V_{x\sigma}^{\text{S}}$  represents the Slater exchange potential,<sup>85</sup> while  $\bar{V}_{xi,\sigma}^{\text{KLI}}$  and  $\bar{v}_{i,\sigma}$  correspond to the average values of  $V_{xi,\sigma}^{\text{KLI}}$  and  $v_{i,\sigma}$  respectively.

The KLI potential has been proven to reflect the most important characteristics of the exact OEP<sup>83</sup> and is the most frequently implemented OEP based approach. Nonetheless, even within the KLI framework, the nature of the equations entails a considerable computational effort, restraining the applications to small molecular systems.

### 1.3.8 Asymptotically Corrected Functionals

This class of functionals has been developed in order to reproduce correctly properties related to the description of excited states, for instance response properties due to the introduction of an electromagnetic field perturbation.

The description of the more diffuse Rydberg orbitals, as well as the energy of the highest occupied molecular orbital (HOMO)  $\epsilon_H$ , depends on the behavior of the exchange-correlation potential in the region far from the nucleus, also known as the asymptotic region or as the “large  $\mathbf{r}$ ” region.<sup>86,87</sup> The exact exchange-correlation potential has the asymptotic form,<sup>88,89</sup>

$$V_{xc}(\mathbf{r}) \Rightarrow -\frac{1}{\mathbf{r}} + I + \epsilon_H \quad (1.76)$$

where  $I$  corresponds to the lowest ionization energy. Moreover, the exact exchange-correlation potential follows Koopmans’ theorem ( $I = -\epsilon_H$ ), hence in the asymptotic limit,

$$V_{xc}^\infty(\mathbf{r}) \equiv \lim_{\mathbf{r} \rightarrow \infty} V_{xc}(\mathbf{r}) = \lim_{\mathbf{r} \rightarrow \infty} -\frac{1}{\mathbf{r}} = 0 \quad (1.77)$$

The problem encountered in using approximate LDA and GGA functionals arises from the fact that the asymptotic behavior of their corresponding potential is much different than the one expressed by Eq.(1.77). They also vanish at infinity but exponentially, hence too fast. The same cannot be said for the hybrid potential and for the OEP, in which the characteristics of the exact HF exchange are respectively partially and fully respected.

As a result of the exponential decay, the LDA and GGA functionals lead to a less attractive potential in the asymptotic region, hence to a wrong estimation of the HOMO energy. In fact,  $-\epsilon_H$  is significantly smaller than  $I$  for most of the commonly used approximate functionals.<sup>87</sup>

The accurate description of the low lying states depends instead mainly on another property of the exact exchange-correlation potential, known as the “derivative discontinuity”.<sup>90</sup> This property requires the correct exchange-correlation potential  $\frac{\partial E_{xc}}{\partial \rho(\mathbf{r})}$  to “jump” discontinuously by the constant  $(I - A)$ , where  $A$  corresponds to the electron affinity, as the number of electrons varies by an integer

N.<sup>90</sup> LDAs and GGAs are defined as “derivative discontinuity free” functionals, since the corresponding potentials are smooth with respect to the variation of the number of electrons. Nonetheless, their corresponding charge density is still a good estimate of the exact charge density since their potential parallels the exact potential in the energetically significant regions.

A considerable problem in correcting the asymptotic behavior of the approximate functionals is represented by the fact that the modification of the asymptotic behavior only leads to a deterioration of the energies of the low-lying states. The reason is that an exchange-correlation potential not showing derivative discontinuity cannot present the correct asymptotic behavior and at the same time give the right energy.<sup>87</sup> For example, the asymptotically corrected LB94 functional,<sup>91</sup> in which the asymptotic behavior of the LDA has been corrected without taking into account the derivative discontinuity, shows some problems in the determination of excitation energies to the low-lying states and of other related properties<sup>87,92</sup>

The solution is attained by correcting the asymptotic behavior exclusively in the asymptotic region, leaving the characteristics of the approximate potential unaffected at distances closer to the nucleus. In the HCTH(AC)<sup>89</sup> functional, for example, the potential generated by HCTH is preserved in energetically important regions while it is corrected exclusively in the asymptotic regions. Another example is represented by the asymptotically corrected LDA (AC-LDA) developed by Casida and Salahub.<sup>87</sup> In the AC-LDA the asymptotic behavior of the LDA potential is corrected by a “shift and splice” approach, which consists in shifting the  $V_{xc}^{LDA}$  towards the correct potential by a quantity given by,

$$\Delta_{xc} = \epsilon_H^{LDA} + I_{\Delta SCF}^{LDA} \quad (1.78)$$

only in the energetically significant regions. The term  $I_{\Delta SCF}^{LDA}$  represents the LDA

ionization energy calculated with the  $\Delta$ SCF approach<sup>§</sup>. In the asymptotic regions the LDA is spliced onto the LB94 taking the switchover point where the two potentials cross,

$$V_{xc}^{AC-LDA}(\mathbf{r}) = \text{Max} [V_{xc}^{LDA}(\mathbf{r}) - \Delta, V_{xc}^{LB94}(\mathbf{r})] \quad (1.79)$$

The two functionals HCTH(AC) and AC-LDA have led to considerable improvements in the determination of excited state energies and related properties.<sup>87,89</sup>

---

<sup>§</sup>Direct difference between the total energies of the N-electron system and of the (N - 1)-electron system respectively.

## Chapter 2

### The Time Domain

The topic addressed by this chapter is the extension of the concepts of time-independent DFT, seen in Ch. 1, to the time domain. Time-dependent density-functional theory (TDDFT)<sup>93-97</sup> represents a very efficient methodology to deal with time-dependent systems. Its application has been found quite successful in the determination of linear response properties, i.e. properties deriving from the application of a small time-dependent perturbation.<sup>98-101</sup> Some examples of linear response properties are given by polarizabilities, dielectric functions, excitation energies and, as it will be shown in the forthcoming chapters, also NMR paramagnetic shieldings.

#### 2.1 The Time-Dependent Kohn-Sham Equations

The main advantage of TDDFT is that it allows one to deal with a time-dependent fully interacting many-body problem in terms of a non-interacting particle system. In fact, the solution of the complete time-dependent Schrödinger equation is

computationally extremely expensive, hence the resulting wave function has been determined only for very small systems.<sup>102</sup>

The formulation of the KS equations in the time domain requires the proof that, for a given initial state, the time-dependent density  $\rho(\mathbf{r}, t)$  is a unique functional of the external potential  $v_{\text{ext}}(\mathbf{r}, t)$ . This postulate represents exactly the time-dependent counterpart of the first Hohenberg-Kohn theorem. The Runge-Gross theorem<sup>103</sup> covers this function establishing a one-to-one mapping between a time-dependent potential  $v(\mathbf{r}, t)$  and the density  $\rho(\mathbf{r}, t)$ . It states,

*“Two densities  $\rho(\mathbf{r}, t)$  and  $\rho'(\mathbf{r}, t)$  evolving from a common initial state  $|\Phi_0\rangle$  under the influence of two potentials  $v(\mathbf{r}, t)$  and  $v'(\mathbf{r}, t)$  are always different provided that the potentials differ by more than a purely time-dependent ( $\mathbf{r}$ -independent) function,*

$$v(\mathbf{r}, t) \neq v'(\mathbf{r}, t) + C(t) \quad (2.1)$$

It can also be proven that the density  $\rho(\mathbf{r}, t)$  of the fully interacting particle system can be also reproduced, under some assumptions, by an external potential with the characteristics of the KS potential  $V_{\text{eff}}(\mathbf{r}, t)$  in an initial state corresponding to a Slater determinant  $\Psi_0$ , constructed from the time-dependent KS orbitals  $\psi_i(\mathbf{r}, t)$ , rather than in an initial state corresponding to the exact unperturbed fully interacting state  $\Phi_0$ .<sup>104</sup>

Accordingly, the time-dependent KS equations are given by,

$$\left[ i \frac{\partial}{\partial t} + \frac{1}{2} \nabla^2 - V_{\text{eff}}(\mathbf{r}, t) \right] \psi_i(\mathbf{r}, t) = 0 \quad (2.2)$$

while the time-dependent density is obtained from the time-dependent KS orbitals



by,

$$\rho(\mathbf{r}, t) = \sum_i^N |\psi_i(\mathbf{r}, t)|^2 \quad (2.3)$$

The KS potential  $V_{\text{eff}}(\mathbf{r}, t)$  in Eq.(2.2) can be expressed analogously to the time-independent case as,

$$V_{\text{eff}}(\mathbf{r}, t) = V_{\text{ext}}(\mathbf{r}, t) + \int \frac{\rho(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + V_{\text{xc}}(\mathbf{r}, t) \quad (2.4)$$

although it cannot be defined as a functional derivative of the time-dependent density, or in other words, there is no functional of the  $v$ -representable density  $\rho(\mathbf{r}, t)$  that has the KS potential as its derivative.

Nonetheless, it is possible to define a particular functional  $A[\rho]$ , known as the Keldysh action,<sup>104,105</sup> for which,

$$A[\rho] = A_s[\rho] - A_{\text{xc}}[\rho] - \frac{1}{2} \int_{t_0}^{t_1} \int \int \frac{\rho(\mathbf{r}, t)\rho(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|} dt d\mathbf{r} d\mathbf{r}' \quad (2.5)$$

It is possible to recognize in  $A[\rho]$  the role of a time-dependent universal functional, where  $A_s[\rho]$  represents the action functional for the non-interacting system, and  $A_{\text{xc}}[\rho]$  corresponds to the exchange-correlation part,

$$V_{\text{xc}}(\mathbf{r}, t) = \frac{\partial A_{\text{xc}}}{\partial \rho(\mathbf{r}, t)} \quad (2.6)$$

The explicit form of the exchange-correlation action is not known, however in the limit of an external potential varying slowly in time, it is possible to establish a connection between time-independent exchange-correlation functionals and exchange-correlation action,

$$A_{\text{xc}} = \int_{t_0}^{t_1} E_{\text{xc}}[\rho_t] dt \quad (2.7)$$

where  $\rho_t$  is the electron-density evaluated at the time  $t$ . The relation in Eq.(2.7) is known as the “adiabatic approximation” and it is particularly useful in practical

calculations due to the lack of independently developed approximate expressions for  $V_{xc}(\mathbf{r}, t)$ . Accordingly, we can write,

$$V_{xc}(\mathbf{r}, t) = \frac{\partial A_{xc}}{\partial \rho(\mathbf{r}, t)} \approx \frac{\partial E_{xc}}{\partial \rho_t(\mathbf{r})} = V_{xc}(\mathbf{r}) \quad (2.8)$$

Among all the available approximations for exchange-correlation potentials used in time-dependent calculations, the most popular is the “adiabatic local density approximation” (ALDA),<sup>106,107</sup>

$$V_{xc}^{ALDA}[\rho](\mathbf{r}, t) = V_{xc}^{LDA}[\rho_t](\mathbf{r}) \quad (2.9)$$

where the LDA potential  $V_{xc}^{LDA}$  is described in detail in Sec.( 1.3.3).

## 2.2 Time-Dependent Linear Response Theory

As it has been mentioned at the beginning of this chapter, the most successful applications of TDDFT have been in the linear response domain. The linear response corresponds to the dynamical reaction of a system, originally in the ground state  $|\Phi_0\rangle$ , to a small external perturbation  $\delta v_{\text{appl}}(\mathbf{r}, \omega)$ . The linear response of the density can be given in the frequency domain by,

$$\delta \rho(\mathbf{r}_1, \omega) = \int \tilde{\chi}[\rho_0](\mathbf{r}_1 \mathbf{r}_2, \omega) \delta v_{\text{appl}}(\mathbf{r}_2, \omega) d\mathbf{r}_2 \quad (2.10)$$

where  $\tilde{\chi}$  is known as the “generalized susceptibility”, or as “exact density-density response function”,

$$\tilde{\chi}(\mathbf{r}_1 \mathbf{r}_2, \omega) = \int_{-\infty}^{+\infty} \chi(\mathbf{r}_1 \mathbf{r}_2, t) e^{i\omega t} dt \quad (2.11)$$

$\tilde{\chi}$  can be explicitly defined, following a second quantization notation, by the sum-over-states (SOS) expression,<sup>108</sup>

$$\tilde{\chi}(\omega) = \lim_{\eta \rightarrow 0^+} \sum_{I \neq 0} \left\{ \frac{\langle \Phi_0 | \hat{a}_{k\sigma}^\dagger \hat{a}_{i\sigma} | \Phi_I \rangle \langle \Phi_I | \hat{a}_{j\tau}^\dagger \hat{a}_{l\tau} | \Phi_0 \rangle}{\omega - (E_I - E_0) + i\eta} - \frac{\langle \Phi_0 | \hat{a}_{j\tau}^\dagger \hat{a}_{l\tau} | \Phi_I \rangle \langle \Phi_I | \hat{a}_{k\sigma}^\dagger \hat{a}_{i\sigma} | \Phi_0 \rangle}{\omega + (E_I - E_0) + i\eta} \right\} \quad (2.12)$$

where  $\{\Phi_I\}$  represents a complete set of eigenstates, with corresponding eigenvalues  $E_I$ , of the time-dependent fully interacting unperturbed Hamiltonian  $H$  and where  $\eta$  is a positive infinitesimal. The definition for  $\tilde{\chi}(\omega)$  in Eq.(2.12) shows that the poles of the exact response function correspond to the excitation energy of the system.<sup>109</sup>

SOS expressions can be also used in the definition of linear response properties, such as the dynamic polarizability  $\alpha(\omega)$ .<sup>99</sup> In this particular case, the perturbation, represented by the electric field  $\mathbf{E}^*$ ,

$$\delta v_{\text{appl}}(t) = \hat{z} E_z(t) \quad (2.13)$$

can induce a time-dependent dipole moment,

$$\delta \mu_x(t) = \int_{-\infty}^{+\infty} \alpha_{xz}(t-t') E_z(t') dt' \quad (2.14)$$

where  $\alpha_{xz}$  indicates the  $xz$ -component of the dynamic polarizability tensor. The total dipole moment in the presence of an electric field perturbation can be expressed as,

$$\mu_x(t) = \mu_x + \delta \mu_x(t) + \dots \quad (2.15)$$

\*Following the derivation scheme adopted by Casida,<sup>99</sup> for simplicity of notation, only the  $(x, z)$  components will be considered.

hence,

$$\alpha_{xz}(t) = \frac{\delta\mu_x(t)}{\partial E_z(t)} \quad (2.16)$$

where  $\delta\mu_x(t)$  represents the linear response of the dipole moment. According to Eq.(2.10),  $\delta\mu_x$  can be expressed, in the frequency domain, as,

$$\delta\mu_x(\omega) = - \iint \mathbf{x}^\dagger(\mathbf{r}_1) \tilde{\chi}(\mathbf{r}_1\mathbf{r}_2, \omega) \mathbf{z}(\mathbf{r}_2) E_z(\omega) d\mathbf{r}_1 d\mathbf{r}_2 \quad (2.17)$$

Comparing Eq.(2.17) to Eq.(2.14), it follows that,

$$\alpha_{xz}(\omega) = - \iint \mathbf{x}^\dagger(\mathbf{r}_1) \tilde{\chi}(\mathbf{r}_1\mathbf{r}_2, \omega) \mathbf{z}(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad (2.18)$$

From Eq.(2.12) it is possible to obtain the SOS expression for the dynamic polarizability,

$$\alpha_{xz}(\omega) = \sum_I \frac{2(E_I - E_0) \langle \Phi_0 | \hat{x} | \Phi_I \rangle \langle \Phi_I | \hat{z} | \Phi_0 \rangle}{(E_I - E_0)^2 - \omega^2} \quad (2.19)$$

where the poles are represented by the excitation energies  $(E_I - E_0)$ .

### 2.2.1 Density-Functional Response Theory

The change in electron density  $\delta\rho(\mathbf{r}_1, \omega)$ , determined for the fully interacting system by a (small) perturbation is equal to the change in electron density associated with the corresponding non-interacting system. Accordingly, it is possible to rewrite Eq.(2.10) in terms of a perturbation to the KS potential  $V_{\text{eff}}(\mathbf{r}, \omega)$  as,

$$\delta\rho(\mathbf{r}_1, \omega) = \int \chi_s[\rho_0](\mathbf{r}_1\mathbf{r}_2, \omega) \delta V_{\text{eff}}(\mathbf{r}_2, \omega) d\mathbf{r}_2 \quad (2.20)$$

where  $\chi_s$  stands for the generalized susceptibility for the non-interacting KS system. In view of the definition of  $V_{\text{eff}}$  given in Eq.(2.4), we can write,

$$\begin{aligned} \delta V_{\text{eff}}(\mathbf{r}_1, \omega) &= \delta V_{\text{appl}}(\mathbf{r}_1, \omega) + \delta V_{\text{SCF}}(\mathbf{r}_1, \omega) = \\ & \delta V_{\text{appl}}(\mathbf{r}_1, \omega) + \int \left( \frac{1}{\mathbf{r}_{12}} + f_{\text{xc}}(\mathbf{r}_1\mathbf{r}_2, \omega) \right) \delta\rho(\mathbf{r}_2, \omega) d\mathbf{r}_2 \end{aligned} \quad (2.21)$$

where  $\delta V_{\text{SCF}}$  represents the linear response of the SCF field to the applied perturbation. In Eq.(2.21), the term  $f_{\text{xc}}$  is known as the “exchange-correlation kernel” and it is given by,

$$f_{\text{xc}}(\mathbf{r}_1\mathbf{r}_2, \omega) = \frac{\delta V_{\text{xc}}(\mathbf{r}_1, \omega)}{\delta \rho(\mathbf{r}_2, \omega)} \quad (2.22)$$

In the case of a non-interacting KS system, the SOS expression for the generalized susceptibility is expressed in terms of KS orbitals  $\psi_{k(j)\tau}$ ,

$$\chi_s(\mathbf{r}_1\mathbf{r}_2, \omega) = \sum_{ik\tau, jl\sigma} (f_{i\tau} - f_{k\tau}) \frac{\psi_{i\tau}(\mathbf{r}_1)\psi_{k\tau}^*(\mathbf{r}_1)\psi_{l\sigma}^*(\mathbf{r}_2)\psi_{j\sigma}(\mathbf{r}_2)}{\omega - (\epsilon_{k\tau} - \epsilon_{i\tau})} \quad (2.23)$$

where  $\epsilon_{i(k)\tau}$  and  $f_{i(k)\tau}$  are respectively the eigenvalues and occupation numbers<sup>†</sup>. Accordingly, we can write the response of the density matrix  $\delta P(\omega)$  as,

$$\begin{aligned} \delta P_{ik\tau}(\omega) &= \frac{f_{k\tau} - f_{i\tau}}{\omega - (\epsilon_{i\tau} - \epsilon_{k\tau})} \delta V_{ik\tau}^{\text{eff}}(\omega) \\ &= \frac{f_{k\tau} - f_{i\tau}}{\omega - (\epsilon_{i\tau} - \epsilon_{k\tau})} \left[ \delta V_{ik\tau}^{\text{appl}}(\omega) + \delta V_{ik\tau}^{\text{SCF}}(\omega) \right] \end{aligned} \quad (2.24)$$

The response of the density matrix is non-zero if  $f_{i\tau} \neq f_{k\tau}$ . In particular, if  $f_{k\tau} < f_{i\tau}$  the corresponding terms of  $\delta P(\omega)$  are defined as “particle hole” (*ph*) elements, while the elements corresponding to  $f_{k\tau} > f_{i\tau}$  are defined as “hole particle” (*hp*) elements.<sup>99</sup>

From the definition of the linear response of the SCF field, Eq.(2.21), we can see that  $\delta V_{\text{SCF}}(\omega)$  depends on the response of the density matrix. The relation between  $\delta V_{\text{SCF}}(\omega)$  and  $\delta P(\omega)$  can be expressed also through the coupling matrix  $K$  as,

$$\delta V_{\text{SCF}}(\omega) = K(\omega)\delta P(\omega) \quad (2.25)$$

<sup>†</sup>The Latin characters  $i, k, j$  and  $l$  are used as indices for the radial part of the KS orbitals, while the Greek characters  $\sigma$  and  $\tau$  specify the spin

The terms of the K matrix are determined as,

$$K_{ik\tau,jl\sigma}(\omega) = \frac{\delta V_{ik\tau}^{\text{SCF}}(\omega)}{\delta P_{jl\sigma}(\omega)} \quad (2.26)$$

In view of the definition given for  $\delta V^{\text{SCF}}(\omega)$ , the latter expression can be written also in a more explicit form as,

$$K_{ik\tau,jl\sigma}(\omega) = [\psi_{i\tau}\psi_{k\tau}^*|\psi_{j\sigma}\psi_{l\sigma}^*] + \iint \psi_{i\tau}^*(\mathbf{r}_1)\psi_{k\tau}(\mathbf{r}_1) \frac{\partial^2 \Lambda_{xc}}{\partial \rho(\mathbf{r}_1, \omega) \partial \rho(\mathbf{r}_2, \omega)} \psi_{j\sigma}(\mathbf{r}_2)\psi_{l\sigma}^*(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad (2.27)$$

The first integral is simply an exchange integral, while the second involves the second derivative of the exchange-correlation component of the Keldysh action, as described in Eq.(2.6). Going back to Eq.(2.24), we can define the linear response generic property in terms of the coupling matrix as,

$$\delta V_{ik\tau}^{\text{appl}}(\omega) = \sum_{jl\sigma} \left[ \delta_{\tau,\sigma} \delta_{i,j} \delta_{k,l} \frac{\omega - (\epsilon_{j\sigma} - \epsilon_{l\sigma})}{f_{l\sigma} - f_{j\sigma}} - K_{ik\tau,jl\sigma} \right] \delta P_{jl\sigma}(\omega) \quad (2.28)$$

According to the aforementioned definition of *ph* and *hp*-elements of the response of the density matrix  $\delta P(\omega)$ , we can separate Eq.(2.28) in two expressions, corresponding respectively to the *ph*,

$$\delta V_{ik\tau}^{\text{appl}}(\omega) = \sum_{jl\sigma}^{f_{l\sigma} < f_{j\sigma}} \left[ \delta_{\tau,\sigma} \delta_{i,j} \delta_{k,l} \frac{\omega - (\epsilon_{j\sigma} - \epsilon_{l\sigma})}{f_{l\sigma} - f_{j\sigma}} - K_{ik\tau,jl\sigma} \right] \delta P_{jl\sigma}(\omega) - \sum_{jl\sigma}^{f_{l\sigma} < f_{j\sigma}} K_{ik\tau,lj\sigma} \delta P_{lj\sigma}(\omega) \quad (2.29)$$

and to the *hp* contribution,

$$\delta V_{ik\tau}^{\text{appl}}(\omega) = \sum_{jl\sigma}^{f_{l\sigma} > f_{j\sigma}} \left[ \delta_{\tau,\sigma} \delta_{i,j} \delta_{k,l} \frac{\omega - (\epsilon_{l\sigma} - \epsilon_{j\sigma})}{f_{j\sigma} - f_{l\sigma}} - K_{ki\tau,lj\sigma} \right] \delta P_{lj\sigma}(\omega) - \sum_{jl\sigma}^{f_{l\sigma} > f_{j\sigma}} K_{ki\tau,jl\sigma} \delta P_{jl\sigma}(\omega) \quad (2.30)$$

Within the ALDA framework, and in general for all the local exchange-correlation energy functionals  $E_{xc}[\rho_t]$ , the K matrix is real when the KS orbitals are real and,

$$K_{ik\tau,jl\sigma} = K_{ik\tau,lj\sigma} = K_{ki\tau,lj\sigma} = K_{ki\tau,jl\sigma} \quad (2.31)$$

The same relation does not hold in case of hybrid functionals, due to the non-local exchange contribution. This feature, as we will see shortly, complicates considerably the solution of the equations since K shows in this case a significantly reduced symmetry,

$$K_{ik\tau,jl\sigma} = K_{ki\tau,lj\sigma}^* \text{ but } K_{ik\tau,jl\sigma} \neq K_{ik\tau,lj\sigma} \quad (2.32)$$

In a classic TDLDA approach, for instance, we can take advantage of the expression in Eq.(2.31) and combine the two equations for the *ph* and *hp*-contributions to the response property [see Eq.(2.29) and (2.30)] in one matrix equation,

$$\left\{ \begin{array}{c|c} A(\omega) & B(\omega) \\ \hline B(\omega) & A(\omega) \end{array} \right| - \omega \left| \begin{array}{c|c} C & 0 \\ \hline 0 & -C \end{array} \right\} \begin{pmatrix} \partial P(\omega) \\ \partial P^*(\omega) \end{pmatrix} = \begin{pmatrix} \partial V_{\text{appl}}(\omega) \\ \partial V_{\text{appl}}^*(\omega) \end{pmatrix} \quad (2.33)$$

where,

$$A_{ik\tau,jl\sigma}(\omega) = \delta_{\tau,\sigma} \delta_{i,j} \delta_{k,l} \frac{\epsilon_{j\tau} - \epsilon_{l\tau}}{f_{j\tau} - f_{l\tau}} - K_{ik\tau,jl\sigma}(\omega) \quad (2.34)$$

$$B_{ik\tau,jl\sigma}(\omega) = -K_{ik\tau,lj\sigma}(\omega) \quad (2.35)$$

$$C_{ik\tau,jl\sigma}(\omega) = \frac{\delta_{\tau,\sigma} \delta_{i,j} \delta_{k,l}}{f_{j\tau} - f_{l\tau}} \quad (2.36)$$

## 2.3 Working Equations for Real and Imaginary Perturbations

The matrix equation in Eq.(2.33) represents the most general formulation of the time-dependent linear response problem. In fact, up until now, we did not specify

with what type of perturbation we are dealing with, or if the perturbation and the corresponding response were real, as is the case of an electric field perturbation, or if they were imaginary, as is the case of a magnetic field perturbation.

In this section, we will derive from Eq.(2.33) two slightly different equations that apply respectively in the case of a generic real perturbation  $\delta V_R(\omega)$  and in the case of a generic imaginary perturbation  $\delta V_I(\omega)$ .

In order to proceed to the identification of the real and imaginary part it is necessary to carry out a matrix diagonalization on Eq.(2.33), which leads to,

$$\left\{ \begin{array}{c} \left| \begin{array}{cc} A(\omega) + B(\omega) & 0 \\ 0 & A(\omega) - B(\omega) \end{array} \right| - \omega \left| \begin{array}{cc} 0 & -1 \\ -1 & 0 \end{array} \right| \end{array} \right\} \begin{array}{c} \left| \begin{array}{c} \text{Re } \delta P(\omega) \\ -i\text{Im } \delta P(\omega) \end{array} \right| \\ \left| \begin{array}{c} \text{Re } \delta V_{\text{appl}}(\omega) \\ -i\text{Im } \delta V_{\text{appl}}(\omega) \end{array} \right| \end{array} = \quad (2.37)$$

Note that we have considered the occupation number  $f_{k\tau}$  to be unitary, while  $f_{l\tau}$  is zero, thus  $C = 1$ .

### 2.3.1 Linear Response from a Real Perturbation

In the case of a real perturbation  $\delta V_R(\omega) \equiv \text{Re } \delta V_{\text{appl}}(\omega)$ , while  $-i\text{Im } \delta V_{\text{appl}}(\omega) = 0$ , hence,

$$\left| \begin{array}{cc} A(\omega) + B(\omega) & \omega 1 \\ \omega 1 & A(\omega) - B(\omega) \end{array} \right| \begin{pmatrix} \text{Re } \delta P(\omega) \\ -i\text{Im } \delta P(\omega) \end{pmatrix} = \begin{pmatrix} \delta V_R(\omega) \\ 0 \end{pmatrix} \quad (2.38)$$

The corresponding system of equations is,

$$\begin{cases} (A + B)(\text{Re } \delta P(\omega)) + \omega 1(-i\text{Im } \delta P(\omega)) = \delta V_R(\omega) \\ \omega 1(\text{Re } \delta P(\omega)) + (A - B)(-i\text{Im } \delta P(\omega)) = 0 \end{cases} \quad (2.39)$$



which can be solved by the elimination of the imaginary term ( $-i\text{Im } \partial P(\omega)$ ), leading to,

$$\left[ (A + B) - \omega^2 \mathbf{1} (A - B)^{-1} \right] \text{Re } \delta P(\omega) = \delta V_R(\omega) \quad (2.40)$$

which is equivalent to,

$$(A - B)^{-1/2} \left[ (A - B)^{1/2} (A + B) (A - B)^{1/2} - \omega^2 \mathbf{1} \right] (A - B)^{-1/2} \times \\ \text{Re } \delta P(\omega) = \delta V_R(\omega) \quad (2.41)$$

For simplicity of notation, we can indicate,

$$S = (A - B) \quad (2.42)$$

Combining Eq.(2.41) with Eq.(2.42), we obtain the expression for the real part of the response of the density matrix,

$$\text{Re } \partial P(\omega) = S^{1/2} [\Omega(\omega) - \omega^2 \mathbf{1}]^{-1} S^{1/2} \delta V_R(\omega) \quad (2.43)$$

where,

$$\Omega(\omega) = S^{1/2} (A + B) S^{1/2} \quad (2.44)$$

In the specific case of an electric field perturbation [see Eq.(2.13)], the response of the electric dipole is also real and, in terms of the linear response of the density matrix, is given by,

$$\delta \mu_x(\omega) = 2x^\dagger \text{Re } \delta P(\omega) \quad (2.45)$$

Substituting the expression for  $\text{Re } \partial P(\omega)$  in Eq.(2.43), we obtain,

$$\delta \mu_x(\omega) = 2x^\dagger S^{1/2} [\Omega(\omega) - \omega^2 \mathbf{1}]^{-1} S^{1/2} z E_z(\omega) \quad (2.46)$$

Thus, according to the definition in Eq.(2.16), the dynamic polarizability is given by,

$$\alpha_{xz}(\omega) = 2x^\dagger S^{1/2} [\Omega(\omega) - \omega^2 1]^{-1} S^{1/2} z \quad (2.47)$$

The excitation energies, representing the poles of the dynamic polarizability, as suggested by the SOS expression for  $\alpha_{xz}$  in Eq.(2.19), are obtained by solving the pseudo-eigenvalue equation,

$$\boxed{\Omega(\omega)F = \omega^2 F} \quad (2.48)$$

### 2.3.2 Linear Response from an Imaginary Perturbation

The system of matrix-equations we have to solve in case of a generic imaginary perturbation ( $\delta V_1(\omega) \equiv -i\text{Im} \delta V_{\text{appl}}(\omega)$ ) is given by,

$$\begin{cases} (A + B) (\text{Re} \delta P(\omega)) + \omega 1 (-i\text{Im} \delta P(\omega)) = 0 \\ \omega 1 (\text{Re} \delta P(\omega)) + (A - B) (-i\text{Im} \delta P(\omega)) = \delta V_1(\omega) \end{cases} \quad (2.49)$$

This time we will solve the system as a function of the imaginary response of the density matrix instead of the real response, hence,

$$\left[ (A - B) - \omega^2 1 (A + B)^{-1} \right] (-i\text{Im} \delta P(\omega)) = \delta V_1(\omega) \quad (2.50)$$

Following the same notation as we did in the previous case, we define the quantity,

$$S' = (A + B) \quad (2.51)$$

and accordingly we can write,

$$(-i\text{Im} \delta P(\omega)) = S'^{1/2} [\Omega'(\omega) - \omega^2 1]^{-1} S'^{1/2} \delta V_1(\omega) \quad (2.52)$$

where,

$$\Omega'(\omega) = (A + B)^{1/2} (A - B) (A + B)^{1/2} \quad (2.53)$$

One example of a linear response property which could be described by Eq.(2.52) is the NMR paramagnetic shielding component of the NMR shielding tensor  $\sigma$ .

In the exact theory, the NMR shielding constant for a nucleus N can be expressed by an SOS equation analogous to Eq.(2.19) shown for the dynamic polarizability. In fact, magnetic and electric perturbations are considered as equivalent and the two properties are both described by the general expression,

$$\alpha(\omega) = \langle \Phi_0 | \hat{O}_{AB} | \Phi_0 \rangle + \left\{ \sum_I \frac{2\omega_I \text{Re} \langle \Phi_0 | \hat{O}_A | \Phi_I \rangle \langle \Phi_I | \hat{O}_B | \Phi_0 \rangle}{\omega^2 - \omega_I^2} \right\} \quad (2.54)$$

where  $\alpha(\omega)$  is an arbitrary molecular property and where  $\hat{O}_{AB}$ ,  $\hat{O}_A$  and  $\hat{O}_B$  are respectively second and first-order perturbing operators. The difference between the equation describing the dynamic polarizability and the NMR shielding constant arises because of the different nature of the electric field (real) and magnetic field (imaginary) perturbations.<sup>110</sup> In particular, in the case of the dynamic polarizability we have that  $\hat{O}_{AB} = 0$  and  $\hat{O}_A = \hat{O}_B = \mathbf{r}$  yielding the SOS formulation in Eq.(2.19). On the other hand, in order to derive the general formulation for the NMR shielding tensor, it is sufficient for now to indicate simply  $\hat{O}_{AB}$  as  $h^{(11)}$ ,  $\hat{O}_A$  as  $h^{(10)}$  and  $\hat{O}_B$  as  $h^{(01)}$  †. According to this notation, we can separate the

†The operators  $h^{(11)}$ ,  $h^{(10)}$ , and  $h^{(01)}$  will be defined, respectively, in Eq.(3.41) on page 94, Eq.(3.36) on page 93, and Eq.(3.40) on page 94.

expression in Eq.(2.54) as,

$$\sigma^d = \langle \Phi_0 | h^{(11)} | \Phi_0 \rangle \quad (2.55)$$

$$\sigma^p = \left\{ \sum_I \frac{2\omega_I \text{Re} \langle \Phi_0 | h^{(10)} | \Phi_I \rangle \langle \Phi_I | h^{(01)} | \Phi_0 \rangle}{\omega^2 - \omega_I^2} \right\} \quad (2.56)$$

which correspond to the diamagnetic  $\sigma^d$  and the paramagnetic  $\sigma^p$  components to the shielding tensor. In the static limit ( $\omega \Rightarrow 0$ ),  $\sigma_{xz}^p(N)$  is given by,

$$\sigma_{xz}^p(N) = -2 \sum_{I \neq 0} \frac{\text{Re} \langle \Phi_0 | h_x^{10} | \Phi_I \rangle \langle \Phi_I | h_z^{01} | \Phi_0 \rangle}{E_I - E_0} \quad (2.57)$$

which represents the magnetic field “counterpart” of the SOS expression in Eq.(2.19) for the dynamic polarizability.

The SOS relation in Eq.(2.56) shows that the paramagnetic shielding also has poles at the excitation energies. Hence, following the arguments used in the case of the electric field perturbation, we could consider determining the excitation energies from the pseudo-eigenvalue equation, comparable to Eq.(2.48), derived from Eq.(2.52),

$$\Omega'(\omega)F' = \omega^2 F' \quad (2.58)$$

However similar, the two pseudo-eigenvalue equations are not equivalent since the real and imaginary responses of the density are not the same. Furthermore, only the derivation of Eq.(2.48) is theoretically justified at this level. In order to bring back the symmetry between magnetic and electric perturbations we have to introduce an approximation, which will be the subject of the next section.

### 2.3.3 The Tamm-Dancoff Approximation

In Sec. 2.2, it has been shown how the equation expressing the perturbing field  $\delta V_{\text{appl}}(\omega)$ , Eq.(2.28), can be separated into two different equations giving the *ph*-contribution [Eq.(2.29)] and the *hp*-contribution [Eq.(2.30)] respectively. The matrix-equation system in Eq.(2.33) has been obtained from the combination of these two equations.

In order to simplify the TDDFT calculation, Hirata and Head-Gordon<sup>111</sup> implemented an approximation known as the ‘‘Tamm-Dancoff approximation’’ (TDA),<sup>112</sup> which consists in setting to zero the *hp*-contribution to  $\delta V_{\text{appl}}(\omega)$ . This results in the elimination of the  $B(\omega)$  matrix in Eq.(2.33), and in the consequent reduction of the computational effort by a factor of approximately two. The good results obtained with the TDDFT/TDA approach in the calculation of excitation energies<sup>111,113</sup> assure that the role of the  $B$  matrix is negligible.

The matrix-equation system which is solved in the full TDDFT approach can also be expressed by the following matrix notation, equivalent to the shorthand form in Eq.(2.48)<sup>§</sup>,

$$\begin{vmatrix} A & B \\ B & A \end{vmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} = \omega \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} \quad (2.59)$$

and, within the TDA, when the matrix  $B$  is set to zero, it becomes,

$$AX = \omega X \quad (2.60)$$

where the definition of the  $A$  matrix remains the one given in Eq.(2.34). Accordingly, the expression for the dynamic polarizability becomes,

$$\alpha_{xz}(\omega) = 2x^\dagger A^{1/2} [A^2 - \omega^2 1]^{-1} A^{1/2} z \quad (2.61)$$

<sup>§</sup>In Eq. 2.48  $F = S^{-1/2}(X + Y)$

Moreover, if we consider the application of the TDA even in the case of a magnetic field perturbation we obtain that the expression for the paramagnetic component of the shielding tensor is given by,

$$\sigma_{xz}^p(\mathbf{N}) = -2 [\mathbf{h}_x^{10}]^\dagger \mathbf{A}^{1/2} [\mathbf{A}^2 - \omega^2 \mathbf{1}]^{-1} \mathbf{A}^{1/2} \mathbf{h}_z^{01} \quad (2.62)$$

The comparison of Eq.(2.61) to Eq.(2.62) shows how the TDA restores the lost symmetry between the electric field and magnetic field perturbations. This conclusion represents the physical basis for the calculation of  $\sigma_{xz}^p(\mathbf{N})$  through the SOS-DFPT approach (see Sec. 3.3.1), and it is the first step toward the derivation of the “Loc.3” approximation.

## Chapter 3

# Theory of the NMR Shielding Tensor

The NMR shielding tensor is a second-order molecular property and it can be expressed as the static second derivative of the ground state energy with respect to the magnetic field,  $\mathbf{B}$ , and the nuclear magnetic moments,  $\mu_K$ ,

$$\sigma_{i,j}^{(K)} = \lim_{\mu_K, \mathbf{B}_i \rightarrow 0} \left( \frac{\partial^2 E}{\partial \mu_{K,j} \partial \mathbf{B}_i} \right) \quad \{j(i) = x, y, z\} \quad (3.1)$$

where K indicates the nuclei. In principle, if we knew the exact unperturbed wave function  $\Psi_0$ , we could carry out the calculation of this derivative through the perturbation theory formalism. Accordingly, in the presence of a magnetic field perturbation, the nonrelativistic electronic Hamiltonian contains only terms linear and quadratic in the magnetic field strength  $\mathbf{B}$  (see Sec. 3.1), therefore it corresponds to the general form,

$$H = H_0 + \lambda V_{10} + \theta V_{01} + \lambda^2 V_{20} + \theta^2 V_{02} + \lambda \theta V_{11} \quad (3.2)$$

where  $\lambda$  and  $\theta$  are two perturbation parameters and  $V_{nm}$  represent the perturbation operators. The energy corresponding to the “mixed” perturbation  $V_{11}$  is given by,

$$E_{11} = \langle \Psi_0 | V_{11} | \Psi_0 \rangle + \langle \Psi_0 | V_{01} | \Psi_{10} \rangle \equiv \langle \Psi_0 | V_{11} | \Psi_0 \rangle + \langle \Psi_0 | V_{10} | \Psi_{01} \rangle \quad (3.3)$$

Moreover, according to perturbation theory,

$$\begin{aligned} E_{11} &= \frac{\partial^2 E}{\partial \lambda \partial \theta} = \left\langle \Psi_0 \left| \frac{\partial H_0^2}{\partial \lambda \partial \theta} \right| \Psi_0 \right\rangle - 2 \sum_{n \neq 0} \frac{\langle \Psi_0 | \frac{\partial H_0}{\partial \lambda} | \Psi_n \rangle \langle \Psi_n | \frac{\partial H_0}{\partial \theta} | \Psi_0 \rangle}{E_n - E_0} \\ &= \langle \Psi_0 | V_{11} | \Psi_0 \rangle + \langle \Psi_0 | V_{01} | \Psi_{10} \rangle \end{aligned} \quad (3.4)$$

Unfortunately, Eq.(3.4) does not hold in case  $\Psi_0$  and  $\Psi_n$  are not the exact eigenstates of the unperturbed Hamiltonian  $H_0$ .

The “mixed” derivative can be calculated numerically from the energy corresponding to an approximate wave function  $\Phi$  of the Hamiltonian in Eq.(3.2) constructed for various values of the perturbation parameters  $\lambda$  and  $\theta$ . This approach is known as “Finite Field” (FF) or “Finite Perturbation” (FP) theory.<sup>114–116</sup> The FP approach is not usually recommended to treat magnetic field perturbations. In addition to the accuracy problems encountered for numerical derivation, the SCF calculation becomes rather complex since the operator describing the interaction with the magnetic field is purely imaginary.<sup>115–117</sup> Only a few examples of the application of FP based methodologies to magnetic field perturbations have been reported in the literature.<sup>118–120</sup>

We can circumvent the problems brought in by the implementation of the FP theory, if we consider that the use of Eq.(3.4) can be accepted if the approximate wave function of the unperturbed state,  $\Phi$ , is such that the Hellmann-Feynman theorem holds,

$$E_{01} = \frac{\partial E}{\partial \lambda} = \left\langle \Phi \left| \frac{\partial H_0}{\partial \lambda} \right| \Phi \right\rangle \quad (3.5)$$



The latter condition entails that,

$$\left\langle \frac{\partial \Phi}{\partial \lambda} \left| H_0 \right| \Phi \right\rangle = \left\langle \Phi \left| H_0 \right| \frac{\partial \Phi}{\partial \lambda} \right\rangle = 0 \quad (3.6)$$

where  $H_0$  is the unperturbed Hamiltonian and  $\lambda$  is the perturbation parameter.

The relation shown in Eq.(3.6) corresponds to the Brillouin condition,<sup>121</sup> hence we can choose as approximate wave function  $\Phi$  a single Slater determinant, or any other wave function that satisfies the Brillouin condition and calculate the “mixed” derivative within the perturbation theory framework. This approach is known as the “Stationary Perturbation Theory”.

In the case the chosen approximate wave function is a single Slater determinant, we talk about the “coupled Hartree-Fock” (CHF) method,<sup>114–117,122</sup> otherwise, if a linear combination of Slater determinants is used, we refer to “coupled multi-configurational SCF”.<sup>123–126</sup>

Later on in this chapter we will delineate the fundamentals of the CHF approach (see Sec. 3.2), however now we shall describe the structure of the electronic Hamiltonian in a magnetic field.

### 3.1 The Electronic Hamiltonian in a Magnetic Field

The effect of the introduction of a magnetic field,  $\mathbf{B}$ , in an atomic or molecular system represented by the electronic Hamiltonian,  $H$ , is best described by means of the “minimal coupling”.<sup>114,115</sup> According to this principle, in the presence of a magnetic field perturbation, the kinetic momentum operator,  $\pi$ , replaces the momentum operator,  $\mathbf{p}$ , in the kinetic energy operator,  $T$ , of the one-electron Hamiltonian,

$$T = \frac{1}{2} \mathbf{p}^2 \rightarrow T(\mathbf{B}) = \frac{1}{2} \pi^2 \quad (3.7)$$

For a magnetic field,  $\mathbf{B}$ , with vector, or magnetic, potential,  $\mathbf{A}$ , the kinetic momentum operator,  $\boldsymbol{\pi}$ , can be defined as,

$$\boldsymbol{\pi} = \mathbf{p} + \frac{1}{c}\mathbf{A} \quad (3.8)$$

where the magnetic field,  $\mathbf{B}$ , is related to the magnetic potential,  $\mathbf{A}$ , by,

$$\mathbf{B} = \nabla \times \mathbf{A} \quad (3.9)$$

The kinetic energy operator in the presence of a magnetic field is given by the expansion of the term on the right in Eq.(3.7), which gives,

$$T(\mathbf{B}) = \frac{1}{2}\mathbf{p}^2 + \frac{1}{c}\mathbf{A} \cdot \mathbf{p} + \frac{1}{2c^2}\mathbf{A}^2 \quad (3.10)$$

In order to derive Eq.(3.10) the ‘‘Coulomb gauge’’ has been chosen as gauge origin for the magnetic potential  $\mathbf{A}$  ( $\nabla \cdot \mathbf{A} = 0$ ). This choice is by far the most frequently used in non-relativistic calculations.<sup>116</sup>

Another term that appears in the electronic Hamiltonian in the presence of a magnetic field perturbation is the term that describes the interaction between the magnetic field,  $\mathbf{B}$ , and the electron spin,  $\mathbf{s}$ . The magnetic moment of the electron is given by,

$$\mathbf{m} = -g\mu_B\mathbf{s} = -\mathbf{s} \quad (3.11)$$

since the electron  $g$ -factor ( $g$ ) is approximately equal to 2 and the Bohr magneton ( $\mu_B$ ) is equal to,

$$\mu_B = \frac{e\hbar}{2m_e} = \frac{1}{2} \quad (3.12)$$

Hence, the complete structure of the molecular, or atomic, electronic Hamiltonian in the presence of a magnetic field is given by,

$$H(\mathbf{B}) = \frac{1}{2} \sum_i \mathbf{p}_i^2 + \frac{1}{c} \sum_i \mathbf{A}_i \cdot \mathbf{p}_i + \frac{1}{2c^2} \sum_i \mathbf{A}_i^2 + V_{ee} + V_{Ne} - \sum_i \mathbf{m}_i \cdot \mathbf{B} \quad (3.13)$$

where  $\mathbf{A}_i$  is the vector potential at the position  $\mathbf{r}_i$  of the electron  $i$ . Still in Eq.(3.13),  $V_{ee}$  and  $V_{Ne}$  represent the electron-electron repulsion and the electron-nucleus attraction respectively and the last term on the right corresponds to the interaction between the electron magnetic moments and the magnetic field.

According to Eq.(3.1) and (3.4), the NMR shielding tensor,  $\sigma^K$ , can be calculated, under the condition expressed by Eq.(3.6), through double-perturbation theory. The perturbation parameters are represented by the external magnetic field,  $\mathbf{B}$ , and by the nuclear magnetic moments,  $\boldsymbol{\mu}_K$ . Hence, the perturbation operators,  $V_{01}$ ,  $V_{10}$  and  $V_{11}$ , are obtained by differentiating the molecular Hamiltonian respectively by the nuclear magnetic moments ( $V_{01}$ ), the external magnetic field ( $V_{10}$ ) and by both perturbations parameters ( $V_{11}$ ). Therefore, we should also consider the contribution to the electronic Hamiltonian in Eq.(3.13) from the magnetic field,  $\mathbf{B}^\mu$ , generated by the nuclear magnetic moments  $\boldsymbol{\mu}_K$ . The total magnetic field is given by,

$$\mathbf{B}^{\text{tot}} = \mathbf{B} + \mathbf{B}^\mu \quad (3.14)$$

Accordingly, the total magnetic potential is given by,

$$\mathbf{A}^{\text{tot}} = \mathbf{A} + \mathbf{A}^\mu \quad (3.15)$$

The magnetic potential  $\mathbf{A}$  associated with the external magnetic field  $\mathbf{B}$  is expressed by,

$$\mathbf{A} = \frac{1}{2} \mathbf{B} \times (\mathbf{r} - \mathbf{R}_O) \quad (3.16)$$

where  $\mathbf{R}_O$  indicates the position of the origin  $O$  of the vector potential, known as the “gauge origin”. Moreover, the magnetic potential  $\mathbf{A}^\mu$  associated with the magnetic field generated by the nuclear magnetic moments  $\boldsymbol{\mu}_K$  is given by,

$$\mathbf{A}^\mu = \sum_K^M \frac{\boldsymbol{\mu}_K \times (\mathbf{r} - \mathbf{R}_K)}{|\mathbf{r} - \mathbf{R}_K|^3} \quad (3.17)$$

where  $M$  is the number of nuclei in the system and  $\mathbf{R}_K$  indicates the position of the nucleus  $K$ . Now we can rewrite the electronic Hamiltonian in Eq.(3.13) in terms of the two sources of magnetic field perturbation,  $\mathbf{B}$  and  $\mathbf{B}^\mu$ , as,

$$H(\mathbf{B}) = \frac{1}{2} \sum_i \mathbf{p}_i^2 + \frac{1}{c} \sum_i (\mathbf{A}_i + \mathbf{A}_i^\mu) \cdot \mathbf{p}_i + \frac{1}{2c^2} \sum_i (\mathbf{A}_i + \mathbf{A}_i^\mu)^2 + V_{ee} + V_{Ne} - \sum_i \mathbf{m}_i \cdot \mathbf{B} \quad (3.18)$$

Apart from the interaction of the magnetic field with the electron spin, all of the magnetic terms of the electronic Hamiltonian in Eq.(3.18) depend explicitly on the magnetic potential  $\mathbf{A}$  and not on the magnetic field  $\mathbf{B}$ , hence  $H(\mathbf{B})$  is not uniquely defined. In fact, from the observation of the definitions for  $\mathbf{A}$  and  $\mathbf{A}^\mu$ , we notice that, while  $\mathbf{A}_K^\mu$  has a preferential gauge origin at the position of the nucleus  $K$  [see Eq.(3.17)], the choice of the gauge origin  $O$  for  $\mathbf{A}$  is arbitrary [see Eq.(3.16)]. Furthermore, according to Eq.(3.9),  $\mathbf{B}$  is uniquely determined if  $\mathbf{A}$  is known, although the opposite is not true, since the magnetic field  $\mathbf{B}$  is totally independent of the choice of gauge origin, but  $\mathbf{A}$  is not. This is the basis for what is known as the “gauge origin problem” which will be the subject of the next subsection.

### 3.1.1 The Gauge Origin Problem

The requirement, expressed in Eq.(3.9), is satisfied by any magnetic potential  $\mathbf{A}$  determined up to the gradient of an arbitrary scalar function  $\xi$ ,

$$\mathbf{A}' = \mathbf{A} + \text{grad } \xi \quad (3.19)$$

In fact,

$$\text{rot grad } \xi = \nabla \times \nabla \xi = 0 \quad (3.20)$$

The scalar function,  $\xi$ , is usually referred to as the “gauge function”.<sup>115,116</sup> The magnetic potential  $\mathbf{A}$ , associated with the magnetic field  $\mathbf{B}$ , has been defined, suitably with the choice of the “Coulomb gauge”, in Eq.(3.16). Therefore, according to Eq.(3.19) and (3.20), the gauge function  $\xi$  can be chosen as,

$$\xi = \frac{1}{2} \mathbf{B} \times (\mathbf{R}_0 - \mathbf{R}) \cdot \mathbf{r} \quad (3.21)$$

which allows what is known as a “gauge transformation”.

Observables, such as the shielding constant, are dependent only on the magnetic field  $\mathbf{B}$  and therefore they are completely independent of the choice of the gauge origin. Hence, only the exact solution of the Schrödinger equation is also gauge invariant. However, in the case of an approximate solution, unless a “natural” gauge origin can be identified such as the position of the nucleus in an atomic calculation, it is necessary to introduce some adjustments.

The most common modification consists in adapting the orbital functions with corrective phase factors. The approximate wave functions would then take up the form,

$$\tilde{\phi}_{\kappa,j}(\mathbf{B}, \mathbf{r}) = \exp \left[ -\frac{i}{2c} \mathbf{B} \times (\mathbf{R} - \mathbf{R}_0) \cdot \mathbf{r} \right] \phi_{\kappa,j}(\mathbf{r}) \quad (3.22)$$

where  $j$  indicates the different orbital functions  $\phi$  centered at the  $\kappa$ -th nucleus. The idea is to shift the gauge dependency from the true gauge origin  $\mathbf{R}_0$  to another “common” gauge origin  $\mathbf{R}$  as a function of which the Hamiltonian is defined. By integration of this type of orbital functions, the dependency upon the arbitrarily chosen gauge origin  $\mathbf{R}$  will appear to disappear and consequently the approximate solution of the Schrödinger equation will appear to be gauge independent.

If the orbital function  $\phi$  in Eq.(3.22) corresponds to an atomic orbital (AO), then the  $\tilde{\phi}_{\kappa,j}$  are identified as “gauge-invariant atomic orbitals” (GIAOs), also

known as London atomic orbitals.<sup>127–129</sup> Within this framework, the true gauge origin  $\mathbf{R}_0$  for each GIAO  $\tilde{\phi}_{K,j}$  corresponds to the position of the nucleus  $K$ . The only shortcoming of the use of GIAOs has been represented by the fact that the calculation of integrals involving this type of functions was much more complicated than for the original basis functions. However, new integration techniques have been recently introduced and the aforementioned difficulties have been finally resolved.<sup>129–131</sup>

As an alternative, the phase factors, of the type shown in Eq.(3.22), could be joined to entities called “localized molecular orbitals” (LMOs),<sup>132</sup> instead of to AOs. This is the idea behind the “individual gauge for localized orbitals” (IGLOs) approach.<sup>122, 133, 134</sup> The localization process involves a unitary transformation of the MOs,

$$\phi_k^{\text{MO}} \rightarrow \psi_k^{\text{LMO}} = \sum_j U_{kj} \phi_j^{\text{MO}} \quad (3.23)$$

As a result, each MO becomes associated to something that often resembles a particular bond or lone-pair of the molecule. At this point an optimal gauge origin  $\mathbf{R}_k$  is chosen for each LMO, usually corresponding to the centroid of charge of the LMO, with respect to which  $\mathbf{r}$  is measured. In analogy with Eq.(3.22) in the GIAO approach, for the IGLO method we have,

$$\phi_k^{\text{MO}} = e^{i\Lambda_k} \psi_k^{\text{LMO}} \quad (3.24)$$

where  $e^{i\Lambda_k}$  is the phase-factor and  $\Lambda_k$  is the gauge function,

$$\Lambda_k(\mathbf{r}) = \frac{1}{2c} (\mathbf{R}_k \times \mathbf{B}) \cdot \mathbf{r} \quad (3.25)$$

The main difference between the GIAO and IGLO is that in the GIAO approach all the AOs, within each MO, have different phase-factors, while in the

IGLO approach only different LMOs have different phase-factors. The IGLO working equations will be shown within the sum-over-states density-functional perturbation theory (SOS-DFPT) formalism in Sec. 3.3.1.

In conclusion, for completeness' sake, I will just mention two approaches that are less popular than GIAO and IGLO known as “localized orbital local origin” (LORG) and “second order LORG” (SOLO).<sup>135–137</sup> These methodologies are strikingly similar to the IGLO approach and the very subtle differences have been analyzed in an article by Kutzelnigg.<sup>115</sup>

## 3.2 The Coupled Hartree-Fock (CHF) Method

Now that we have derived the structure of the electronic Hamiltonian in a magnetic field we can proceed to the description of the “coupled Hartree-Fock” (CHF) method, one of the most popular theories for the calculation of the NMR shielding tensor.

In the Hartree-Fock (HF) approximation the wave function is represented by a single, variationally optimized, Slater determinant,  $\Phi^{\text{HF}}$ , of a set of occupied one-electron spin-orbitals, (i.e.  $\chi_i(\mathbf{x}) = \phi_i(\mathbf{r})\tau(\omega)$ ),

$$\Phi^{\text{HF}} = (N!)^{-\frac{1}{2}} |\chi_1 \chi_2 \chi_3 \cdots \chi_N\rangle \quad (3.26)$$

where  $N$  is the number of electrons of the system. The spin-orbitals  $\chi_i$  are eigenfunctions of the Fock operator,  $f$ ,

$$f = \mathfrak{h} + v^{\text{HF}} = -\frac{\mathbf{p}^2}{2} - \sum_{i,M} \frac{Z_M}{r_{iM}} + \left[ \sum_i J_i + \sum_i K_i \right] \quad (3.27)$$

where  $\mathfrak{h}$  is the one-electron core-hamiltonian operator, which includes the kinetic energy operator and the nuclear attraction operator. The operator indicated with

$v^{\text{HF}}$  is an effective one-electron potential operator called the “HF potential”.<sup>121</sup> The operators  $J_i$  and  $K_i$  indicate respectively the “Coulomb operator” and the “exchange operator”. The Coulomb operator,

$$J_i(\mathbf{r}_1) = \int \chi_i^*(\mathbf{x}_2) r_{12}^{-1} \chi_i(\mathbf{x}_2) d\mathbf{r}_2 d\sigma_2 \quad (3.28)$$

represents the average local potential experienced at  $\mathbf{r}_1$  due to the presence of the electron described by  $\chi_i$ . The exchange operator,

$$K_i(\mathbf{r}_1) \chi_j(\mathbf{x}_1) = \left[ \int \chi_i^*(\mathbf{x}_2) r_{12}^{-1} \chi_j(\mathbf{x}_2) d\mathbf{r}_2 d\sigma_2 \right] \chi_i(\mathbf{x}_1) \quad (3.29)$$

does not have a “classical” interpretation. In fact, contrary to the Coulomb operator, it is a non-local operator in the sense that it is not possible to identify a simple potential  $K_i(\mathbf{r}_1)$  uniquely defined at  $\mathbf{r}_1$  and the effect of  $K_i(\mathbf{r}_1)$  can only be expressed through its action on a spin orbital  $\chi_j(\mathbf{x}_1)$ .<sup>121</sup>

The HF equation is given by,

$$f |\chi_i\rangle = \epsilon_i |\chi_i\rangle \quad (3.30)$$

where the eigenvalues,  $\epsilon_i$ , represent, within the restricted HF (RHF) framework, the energies of the doubly-occupied orbitals.

According to the conclusions drawn at the beginning of the chapter, the effect of the introduction of a homogeneous magnetic field of strength  $\mathbf{B}$  can be accounted for by double perturbation theory. However, instead of expanding in terms of powers of  $\mathbf{B}$ , it is more convenient to multiply  $\mathbf{B}$  by a parameter  $\lambda$  and express all the terms of the expansion in powers of  $\lambda$ . At the end  $\lambda$  can be eliminated from the resulting equations due to a linear independence argument. Hence, the Fock operator becomes,

$$f(\mathbf{B}) = f^{(0)} + i\lambda f^{(10)} + \lambda^2 f^{(20)} \quad (3.31)$$



where the unperturbed operator  $f_0$  is the one defined in Eq.(3.27). We can also expand the spin orbitals  $\chi_i$  as,

$$\chi_i(\mathbf{B}) = \chi_i^{(0)} + i\lambda\chi_i^{(10)} + \lambda^2\chi_i^{(20)} \quad (3.32)$$

and the total energy  $E$  as,

$$E(\mathbf{B}) = E^{(0)} + i\lambda E^{(10)} + \lambda^2 E^{(20)} \quad (3.33)$$

where  $E_0$  is the total energy of the unperturbed state,

$$E_0 = \sum_i \epsilon_i - \frac{1}{2} \sum_{i \neq j} (J_{ij} - K_{ij}) \quad (3.34)$$

The operator  $f_{10}$  in Eq.(3.31) is linear in  $\mathbf{B}$  and corresponds to,

$$f^{(10)} = h^{(10)} + v_{\text{HF}}^{(10)} \quad (3.35)$$

The first order one-electron core-hamiltonian  $h^{(10)}$  can be expressed, according to Eq.(3.13) and (3.16) by,

$$h^{(10)} = \frac{1}{2ci} [\mathbf{B} \times (\mathbf{r} - \mathbf{R}_O)] \cdot \mathbf{p} = \frac{1}{2ci} \mathbf{B} \cdot \mathbf{l}_O \quad (3.36)$$

where  $\mathbf{l}_O$  represents the angular momentum with centroid in the gauge origin  $\mathbf{R}_O$ ; the factor  $i$  has been introduced in order to make all the operators in Eq.(3.31) real.<sup>115</sup>

Also in Eq.(3.35), the operator  $v_{\text{HF}}^{(10)}$  represents the linear response of the HF potential and is obtained by introducing the expansion for the spin orbital [Eq.(3.32)] into the definition of both the Coulomb [Eq.(3.28)] and the exchange [Eq.(3.29)] operators. While the first order correction for the Coulomb operator,  $J_i^1$ , is zero,

the first order correction for the exchange operator is given by,

$$K_i^1(\mathbf{r}_1)\chi_j(\mathbf{x}_1) = \left[ \int \chi_i^{*(0)}(\mathbf{x}_2) r_{12}^{-1} \chi_j(\mathbf{x}_2) d\mathbf{x}_2 \right] \chi_i^{(10)}(\mathbf{x}_1) - \left[ \int \chi_i^{*(10)}(\mathbf{x}_2) r_{12}^{-1} \chi_j(\mathbf{x}_2) d\mathbf{x}_2 \right] \chi_i^{(0)}(\mathbf{x}_1) \quad (3.37)$$

Now that all the terms of the first order correction to the Fock operator have been described, we can introduce the HF equation, in terms of spatial orbitals  $\phi$  ( $\chi(\mathbf{r}, \omega) = \phi(\mathbf{r})\sigma(\omega)$ ), to be solved in the case of a magnetic field perturbation,

$$f^{(0)}\phi_i^{(10)} + f^{(10)}\phi_i^{(0)} = \epsilon_i^{(0)}\phi_i^{(10)} \quad (3.38)$$

These equations are defined as “coupled” HF equations, since the perturbed HF orbitals  $\phi_i^{(10)}$  are coupled with each other by the linear response of the HF potential  $v_{10}^{\text{HF}}$ . The  $\phi_j^{(10)}$  are uniquely determined under the requirement that the orbitals form the basis of an orthonormal set for all values of the perturbation parameter and that,

$$\left\langle \phi_i^{(0)} \left| \phi_j^{(10)} \right. \right\rangle = 0 \quad (3.39)$$

According to Eq.(3.1) and (3.4), the calculation of the shielding tensor  $\sigma^K$  requires the determination of a “mixed” perturbation energy term  $E_{11}$ . Therefore we have to define the first order perturbations due to the magnetic field generated by the nuclear magnetic moments. Consistently with Eq.(3.17) and (3.18), we can write,

$$h^{(01)} = -\frac{i}{c} \frac{\boldsymbol{\mu}_K \times (\mathbf{r} - \mathbf{R}_K)}{|\mathbf{r} - \mathbf{R}_K|^3} \cdot \mathbf{p} \quad (3.40)$$

Finally the “mixed” perturbation operator is given by,

$$h^{(11)} = \frac{1}{2c^2} \frac{[\mathbf{B} \times (\mathbf{r} - \mathbf{R}_O)] \cdot [\boldsymbol{\mu}_K \times (\mathbf{r} - \mathbf{R}_K)]}{|\mathbf{r} - \mathbf{R}_K|^3} \quad (3.41)$$

and the expression obtained for the “mixed” perturbation energy  $E^{(11)}$ , or shielding tensor, is therefore given by,

$$E^{(11)} = 2 \left[ \sum_i^N \langle \phi_i^{(0)} | h^{(11)} | \phi_i^{(0)} \rangle - 2 \langle \phi_i^{(0)} | h^{(01)} | \phi_i^{(10)} \rangle \right] \quad (3.42)$$

The results obtained for the NMR shielding tensor with the CHF approach can reach a fairly good level of accuracy in case of simple molecules with rather localized electrons. A comprehensive review of CHF results, obtained within the IGLO framework, has been presented by Kutzelnigg and coworkers.<sup>122</sup>

In case the molecular system under scrutiny presents delocalized bonds and (or) nuclei with lone pairs, the inclusion of dynamical correlation effects becomes crucial. Among the higher correlated methods, for which an NMR shielding calculation procedure has been developed, the most popular are definitely the “coupled-cluster” (CC) theory,<sup>138–141</sup> in particular the “coupled-cluster singles and doubles augmented by a perturbative correction for triple excitations” (CCSD(T)),<sup>142</sup> and the Möller-Plesset (MP) theory.<sup>140, 141, 143–145</sup> A comprehensive portrayal of the aforementioned methodologies, among others, is given in a review article by Helgaker and coworkers.<sup>116</sup>

### 3.3 From CHF to Uncoupled Density-Functional Theory (UDFT)

The DFT counterpart of the CHF equations [see Eq.(3.38)] is represented by,

$$\left[ -\frac{1}{2} \sum_i \nabla_i^2 + v_{\text{eff}}^{(0)} \right] \phi_i^{(10)} + \left[ h_i^{(10)} + v_{\text{eff}}^{(10)} \right] \phi_i^{(0)} = \epsilon_i^{(0)} \phi_i^{(10)} \quad (3.43)$$

where  $v_{\text{eff}}^{(0)}$  is the unperturbed KS effective potential and  $h^{(10)}$  is the first-order perturbation operator, analogous to the one defined in Eq.(3.36).  $v_{\text{eff}}^{(10)}$  represents the linear response of the KS effective potential to the magnetic field perturbation. A general expression for the linear response of the KS potential  $v_{\text{eff}}$  to a small time-dependent perturbation  $\delta u(\mathbf{r}; t)$  [see also Eq.(2.21) in Sec.( 2.2.1)] is given, in terms of the associated Fourier components, by,

$$\delta v_{\text{eff}}(\mathbf{r}; \omega) = \delta u(\mathbf{r}; \omega) + \int d\mathbf{r}' \frac{\delta \rho(\mathbf{r}'; \omega)}{|\mathbf{r} - \mathbf{r}'|} + \delta v_{\text{xc}}(\mathbf{r}; \omega) \quad (3.44)$$

where  $\delta \rho(\mathbf{r}; \omega)$  is the density response and  $\delta v_{\text{xc}}(\mathbf{r}; \omega)$  the exchange-correlation potential response.

In the case of a magnetic field perturbation, both the linear responses of the electron density  $\rho^{(10)}(\mathbf{r})$  and of the local, or gradient-corrected, exchange-correlation potentials  $v_{\text{xc}}^{(10)}[\rho, \nabla \rho]$  vanish. Hence, in accordance with Eq.(3.44), the linear response of the KS effective potential  $v_{\text{eff}}^{(10)}$  in Eq.(3.43) also vanishes, leading to what is referred to as the ‘‘uncoupled’’ DFT (UDFT) approach.<sup>146–148</sup>

This major difference between DFT and HF, derives from the fact that the features characterizing the commonly used local and gradient-corrected exchange-correlation functionals are aimed to define the system in the absence of an external magnetic field. However, in the presence of a magnetic field perturbation the Hohenberg-Kohn theorem does not hold and the  $v_{\text{xc}}$  is not only dependent on the electron density  $\rho(\mathbf{r})$  but also on the current density  $\mathbf{j}(\mathbf{r})$  introduced by the magnetic field  $\mathbf{B}$ ,

$$E_{\text{xc}}[\rho(\mathbf{r})] \xrightarrow{\mathbf{B}} E_{\text{xc}}[\rho(\mathbf{r}), \mathbf{j}(\mathbf{r})] \quad (3.45)$$

The current density  $\mathbf{j}(\mathbf{r})$  is defined as,

$$\mathbf{j}(\mathbf{r}) = \mathbf{j}_p(\mathbf{r}) + \mathbf{j}_d(\mathbf{r}) = \frac{i}{2} [\phi^* \nabla \phi - \phi \nabla \phi^*] + \frac{1}{c} \rho(\mathbf{r}) \mathbf{A}(\mathbf{r}) \quad (3.46)$$

where,  $\mathbf{j}_p(\mathbf{r})$  is known as the “paramagnetic current density” and  $\mathbf{j}_d(\mathbf{r})$  as the “diamagnetic current density”.

The already challenging task of developing new and improved exchange-correlation functionals for the unperturbed state appears to be much more involved in the case of the presence a magnetic field perturbation. An attempt has been made within the “current density functional theory” (CDFT) framework<sup>149–151</sup> by Vignale and coworkers,<sup>152</sup> although their functional has been the object of rather strong criticism.<sup>150</sup> Another current density dependent functional has been developed by Becke.<sup>153</sup>

The aforementioned approaches are still far from routine application. Thus the great majority of DFT calculations in the presence of a magnetic field perturbation are performed with exchange-correlation functionals developed for the unperturbed state under the assumption that the contribution of the current density can be considered as negligible,<sup>150</sup>

$$E_{xc}[\rho(\mathbf{r}), \mathbf{j}(\mathbf{r})] \approx E_{xc}[\rho(\mathbf{r})] \quad (3.47)$$

Therefore, the general DFT-based framework on which the NMR shielding calculations are nowadays performed is the UDFT approach. The accuracy of the NMR shielding parameters depends mainly on the particular exchange-correlation functional adopted in the calculation.

Having set the first-order response of the KS potential to zero in Eq.(3.43), we can rewrite the uncoupled KS (UKS) one-electron equations as,

$$\left( h_i^{(0)} + h_i^{(10)} \right) \phi_i(\mathbf{r}; \mathbf{B}) = \epsilon_i^{(0)} \phi_i(\mathbf{r}; \mathbf{B}) \quad (3.48)$$

where  $h_i^{(0)}$  is the unperturbed one-electron hamiltonian, while  $h_i^{(10)}$  is the first-order perturbation operator, given in Eq.(3.36). The wave function in the presence

of the magnetic field  $\phi_i(\mathbf{r}; \mathbf{B})$  can be expressed as,

$$\begin{aligned}\phi_i(\mathbf{r}; \mathbf{B}) &= \phi_i^{(0)}(\mathbf{r}) + i\mathbf{B}\phi_i^{(10)}(\mathbf{r}; \mathbf{B}) \\ &= \phi_i^{(0)}(\mathbf{r}) + i\mathbf{B} \sum_b C_{bi} \phi_b^{(0)}(\mathbf{r})\end{aligned}\quad (3.49)$$

where the coefficients  $C_{bi}$  represent the first-order orbital changes. Unlike the CHF equations, the UKS equations, with local or gradient-corrected exchange-correlation functionals, do not need to be solved iteratively. The solutions are of the form,

$$C_{bi} = \frac{\langle \phi_b^{(0)} | \mathbf{h}^{(10)} | \phi_i^{(0)} \rangle}{\epsilon_b - \epsilon_i} \quad (3.50)$$

where the denominator represents the difference between the energies of the unperturbed KS virtual,  $\epsilon_b$ , and occupied,  $\epsilon_i$ , orbitals.

According to double-perturbation theory, the nuclear shielding tensor can be expressed by Eq.(3.42). Hence, if we substitute the solutions obtained from the UKS equations into Eq.(3.42), we obtain,

$$\begin{aligned}\sigma_K &= 2 \sum_i^N \langle \phi_i^{(0)} | \mathbf{h}^{(11)} | \phi_i^{(0)} \rangle - \\ &4 \sum_i^N \sum_{b=N+1}^{N'} \frac{\langle \phi_i^{(0)} | \mathbf{h}^{(10)} | \phi_b^{(0)} \rangle \langle \phi_b^{(0)} | \mathbf{h}^{(01)} | \phi_i^{(0)} \rangle}{\epsilon_b - \epsilon_i}\end{aligned}\quad (3.51)$$

The UDFT methodology has been implemented in numerous codes and nearly all of the successful exchange-correlation functional for the unperturbed state have been tested for NMR properties. Particularly remarkable results have been obtained with the GGA functionals, as HCTH,<sup>60</sup> and with the hybrid-functionals, PBE0,<sup>66,67</sup> B3LYP<sup>48</sup> and B3LYP<sub>GGA</sub><sup>0.05</sup>,<sup>154</sup> where the 0.05 stands for the value of the exact-exchange coefficient.

Generally, the GGA functionals outperform the hybrid-functionals for first and second-row nuclei with low lying excited states,<sup>154</sup> however both types of functionals give NMR shieldings that are too deshielded. Significant improvements have been obtained recently with the use of the OEP, especially within the “multiplicative Kohn-Sham” (MKS) framework developed by Wilson and coworkers.<sup>155</sup> Another recent implementation of the OEP for NMR calculation is represented by the work of Patchkovskii and coworkers<sup>156</sup> where the self-interaction-corrected-VWN (SIC-VWN) functional is used to determine the OEP defined according to the Krieger-Li-Iafrate (KLI) approximation,<sup>83,84</sup> (see Sec. 1.3.7). This methodology, commonly known as SIC-KLI-OEP, produces results that are less accurate than the ones provided by the MKS and the reason behind this is most probably related to the use of the LDA.

However, one of the most successful methodologies for the NMR shielding parameters calculation remains the SOS-DFPT<sup>4</sup> which will be the subject of the next subsection.

### 3.3.1 The Sum-Over-States Density-Functional Perturbation Theory

The success of the SOS-DFPT is mainly related to the fact that it allows one to obtain results that are considerably more accurate than UKS for most of the first and second-row nuclei, with the exact same computational cost. This aspect represents a significant advantage for the NMR study of rather big molecular systems, for instance of 100 to 200 atoms.

The SOS-DFPT approach is based on the UKS equations shown in the previous section. Therefore, the determination of the NMR shielding tensor involves

the use of Eq.(3.51). In order to simplify the description of the working equations, the shielding tensor will be described as the sum of two contributions,

$$\sigma_{\mathbf{k}} = \sigma_{\mathbf{k}}^{\text{d}} + \sigma_{\mathbf{k}}^{\text{p}} \quad (3.52)$$

where  $\sigma_{\mathbf{k}}^{\text{d}}$  is known as the “diamagnetic” shielding term and  $\sigma_{\mathbf{k}}^{\text{p}}$  as the “paramagnetic” shielding term. These two contributions, if considered isolated one from the other, have no physical meaning since only their sum  $\sigma_{\mathbf{k}}$  is an observable.

Within the IGLO framework, see Sec. 3.1.1, the expression for the diamagnetic contribution to the shielding tensor is given by,

$$\sigma_{\mathbf{k}}^{\text{d}} = 2 \sum_{\mathbf{k}} \langle \psi_{\mathbf{k}}^{\text{LMO}} | \mathbf{h}^{11} | \psi_{\mathbf{k}}^{\text{LMO}} \rangle \quad (3.53)$$

$\mathbf{h}^{11}$  is given in Eq.(3.41), where the gauge origin  $\mathbf{R}_{\text{O}}$  is the centroid of the  $\mathbf{k}$ -th LMO. The paramagnetic contribution can be further divided into two terms,

$$\sigma_{\mathbf{k}}^{\text{p}} = \sigma_{\mathbf{k}}^{\text{p}0} + \sigma_{\mathbf{k}}^{\text{p}1} \quad (3.54)$$

where  $\sigma_{\mathbf{k}}^{\text{p}0}$  and  $\sigma_{\mathbf{k}}^{\text{p}1}$  account for the contribution of the occupied-occupied LMOs and for the occupied LMOs-virtual MOs respectively. These terms derive directly from the expansion of the first-order KS orbitals into,

$$\psi_{\mathbf{k}}^{10}(\mathbf{r}, \mathbf{B}) = \sum_j^{N_{\text{occ}}} \psi_j^0 C_{jk} + \sum_a^{M_{\text{virt}}} \psi_a^0 C_{ak} \quad (3.55)$$

Within the IGLO framework the first-order orbital variation coefficients,  $C_{jk}$  and



$C_{ak}$ , are defined as,

$$C_{jk} = -\frac{1}{2c} \langle \psi_j^{\text{LMO}} | \Lambda_j - \Lambda_k | \psi_k^{\text{LMO}} \rangle \quad (3.56)$$

$$C_{ak} = \sum_m^{\text{N}_{\text{occ}}} \left\{ \frac{\sum_n^{\text{N}_{\text{occ}}} -\frac{1}{2c} \left[ \langle \phi_a^{\text{MO}} | l_{nO} | \psi_n^{\text{LMO}} \rangle + \langle \phi_a^{\text{MO}} | (\Lambda_n - \Lambda_j) | \psi_j^{\text{LMO}} \rangle \langle \psi_j^{\text{LMO}} | h^0 | \psi_n^{\text{LMO}} \rangle \right] U_{nm}}{\epsilon_m - \epsilon_a} \right\} \times U_{mk} \quad (3.57)$$

where the indices  $k$ ,  $n$  and  $j$  refer to LMOs and  $m$  to MOs;  $h^0$  represents the unperturbed Kohn-Sham hamiltonian, while  $\Lambda_{j((k),(n))}$  represents the gauge function as defined in Eq.(3.25).  $l_{nO}$  is the orbital angular momentum operator, which derives from the definition of  $h^{10}$  shown in Eq.(3.36). Accordingly, the two components of  $\sigma_k^p$  (i.e.  $\sigma_k^{p0}$  and  $\sigma_k^{p1}$ ) can be written as,

$$\sigma_k^{p0} = -\frac{2}{c} \sum_k^{\text{N}_{\text{occ}}} \sum_j^{\text{N}_{\text{occ}}} \langle \psi_k^{\text{LMO}} | h^{01} | \psi_j^{\text{LMO}} \rangle \langle \psi_j^{\text{LMO}} | \Lambda_j - \Lambda_k | \psi_k^{\text{LMO}} \rangle \quad (3.58)$$

$$\sigma_k^{p1} = -4 \sum_k^{\text{N}_{\text{occ}}} \sum_a^{\text{M}_{\text{virt}}} \langle \psi_k^{\text{LMO}} | h^{01} | \phi_a^{\text{MO}} \rangle C_{ak} \quad (3.59)$$

where  $h^{01}$  is given in Eq.(3.40).

In the SOS-DFPT UKS formalism the excitation energy in the denominator of Eq.(3.56) is calculated as the energy difference between the ground-state Slater determinant  $\Phi = |\phi_1, \phi_2, \dots, \phi_k, \phi_l, \dots, \phi_N\rangle$  and the excited-state Slater determinant generated by replacing the KS occupied orbital  $\phi_k$  by the KS virtual orbital  $\phi_a$ ,  $\Phi_k^a = |\phi_1, \phi_2, \dots, \phi_a, \phi_l, \dots, \phi_N\rangle$ . It is important to underline that in SOS-DFPT we are dealing only with “singlet-singlet” excitations.

The NMR shieldings obtained with the SOS-DFPT UKS approximations are overly deshielded for the first and second-row elements with strong correlation.<sup>146</sup>

The solution to this problem has been found by adding a correction to the excitation energy which takes into account the change in the exchange-correlation energy connected with the formation of a hole in the KS orbital  $\phi_k$  and the consequent occupation of the KS virtual orbital  $\phi_a$ ,<sup>4</sup>

$$\boxed{-\Delta E_{k \rightarrow a} = \epsilon_k - \epsilon_a - \Delta E_{k \rightarrow a}^{xc}} \quad (3.60)$$

Accordingly, the following represents the first expression which has been proposed as the exchange-correlation correction term,

$$\Delta E_{k \rightarrow a}^{xc} = \int \frac{\rho_k(\mathbf{r})}{\rho^\uparrow(\mathbf{r})} [\epsilon_{xc}(\mathbf{r}) - v_{xc}(\mathbf{r})] \rho_a(\mathbf{r}) d\mathbf{r} \quad (3.61)$$

where  $\epsilon_{xc}(\mathbf{r})$  represents the exchange-correlation energy density. In case we consider an exchange-only correction within the LDA, we obtain what is known as “local approximation 1” or more concisely Loc.1,

$$\Delta E_{k \rightarrow a}^x = \frac{1}{3} C_x \int \rho^\uparrow(\mathbf{r})^{-\frac{2}{3}} \rho_k(\mathbf{r}) \rho_a(\mathbf{r}) d\mathbf{r} \quad (3.62)$$

where

$$C_x = \frac{3}{2} \left( \frac{3}{4\pi} \right)^{\frac{1}{3}} \quad (3.63)$$

On the same premises, another local approximation has been developed and has been named “local approximation 2” or Loc.2,

$$\Delta E_{k \rightarrow a}^{xc} = - \int \frac{\partial \rho_{xc}[\rho^\uparrow; \rho^\downarrow]}{\partial \rho^\uparrow(\mathbf{r})} \rho_k(\mathbf{r}) \rho_a(\mathbf{r}) d\mathbf{r} \quad (3.64)$$

The exchange-only Loc.2 expression, analogous to Eq.(3.62), is given by,

$$\Delta E_{k \rightarrow a}^x = \frac{4}{9} C_x \int \rho^\uparrow(\mathbf{r})^{-\frac{2}{3}} \rho_k(\mathbf{r}) \rho_a(\mathbf{r}) d\mathbf{r} \quad (3.65)$$

we can notice that the two exchange-only formulations of Loc.1 and Loc.2 approximations, Eq.(3.62) and (3.65), differ only by a coefficient of  $\frac{4}{3}$ .

The scaling of the excitation energies in Eq.(3.60) represents the key of the SOS-DFPT success but it has also been the subject of a rather harsh controversy. The criticisms are based on the fact that the scaling of the energy does not have any clear physical foundation and is therefore to be considered as arbitrary.<sup>60, 116, 156</sup>

In the next subsection it will be shown how in the effort to improve the description of the energy of the excited states, appearing in Eq.(3.56), we also uncovered the physical basis of the SOS-DFPT energy-scaling formula.<sup>157</sup>

### 3.3.2 The Loc.3 Approximation

In the previous chapter we have seen how the lost symmetry between electric field and magnetic field perturbations is restored by the introduction of the TDA (see Sec. 2.3.3). In fact, we have obtained an SOS expression for the paramagnetic component of the shielding tensor [Eq.(2.62)], perfectly symmetric to the expression for the dynamic polarizability [Eq.(2.61)]. In the static limit Eq.(2.62) becomes,

$$\sigma_{xz}^p(K) = -2 [\mathbf{h}_x^{10}]^\dagger \mathbf{A}^{-1} \mathbf{h}_z^{01} \quad (3.66)$$

where  $K$  is the magnetic nuclei index. From the definition of the  $\mathbf{A}$  matrix in Eq.(2.34) and from Eq.(2.60), we can write,

$$\sigma_{xz}^p(K) = -2 \sum_I \sum_{k,j}^{N_{occ}} \sum_{a,b}^{N'_{vac}} \langle \psi_k | \mathbf{h}_x^{10} | \psi_a \rangle X_{ka}^I \omega_I^{-1} X_{jb}^I \langle \psi_j | \mathbf{h}_z^{01} | \psi_b \rangle \quad (3.67)$$

where,

$$\omega_I = \sum_{k,j}^{N_{occ}} \sum_{a,b}^{N'_{vac}} (\epsilon_a - \epsilon_k) + K_{ka\sigma,jb\tau} \quad (3.68)$$

$X^I$  are the normalized eigenvectors and  $I$  is the excitation index. The indices  $k, j$  run over the occupied orbitals ( $f_{k,j} = 1$ ), while the indices  $a, b$  run over the vir-

tual orbitals ( $f_{a,b} = 0$ ). Eq.(3.67) represents the singles configuration interaction (CIS) version of the SOS-DFPT expression. It can be noted that the correction terms  $\Delta E_{k \rightarrow a}^{xc}$ , added to the excitation energy in the Loc.1 [Eq.(3.62)] and Loc.2 approximations [Eq.(3.65)], are replaced in the expression for  $\omega_1$  by the terms of the coupling matrix  $K$  [Eq.(2.27)].

In order to facilitate the comparison of Eq.(3.67) to conventional SOS-DFPT, we introduced the so called “two level model” (2LM) approximation. Within this framework each transition is approximated by the contribution of a single excitation from an occupied orbital  $\psi_k$  to an unoccupied orbital  $\psi_a$  to form the corresponding singlet state. Accordingly, the singlet excitation energy  $\Delta E_{k \rightarrow a}$ , appearing at the denominator in the SOS expression, is obtained by solving the pseudoeigenvalue equation  $AX = \omega X$ , which, from the definition of the  $A$  matrix in Eq.(2.34), is given by,

$$\begin{bmatrix} (\epsilon_{a,\sigma} - \epsilon_{k,\sigma}) + K^{\sigma=\tau} & +K^{\sigma \neq \tau} \\ +K^{\sigma \neq \tau} & (\epsilon_{a,\sigma} - \epsilon_{k,\sigma}) + K^{\sigma=\tau} \end{bmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \omega_s \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (3.69)$$

where  $a$  and  $k$  are MO indices, while  $\sigma$  and  $\tau$  are spin indices. Therefore,

$$\omega_s = \Delta E_{k \rightarrow a} = \epsilon_a - \epsilon_k + 2K_{ka} + \iint \rho_k(\mathbf{r}) \left( \frac{\partial v_{xc}^{LDA,\uparrow}(\mathbf{r})}{\partial \rho^\uparrow(\mathbf{r}')} + \frac{\partial v_{xc}^{LDA,\uparrow}(\mathbf{r})}{\partial \rho^\downarrow(\mathbf{r}')} \right) \rho_a(\mathbf{r}') d\mathbf{r} d\mathbf{r}' \quad (3.70)$$

From the definition of the coupling matrix  $K$  given in Eq.(2.27), it results that  $K_{ka}$  corresponds to the exact exchange integral  $(ak||ka)$ . Moreover, since Eq.(3.70) is defined within the adiabatic approximation, the LDA  $v_{xc}$  is used to evaluate the exchange-correlation kernel [see Eq.(2.22)]. The terms added in Eq.(3.70) to the KS orbitals energy difference  $(\epsilon_a - \epsilon_k)$  define the “Loc.3” correction.<sup>157</sup>

Table 3.I: Corrections to the excitation energy in SOS-DFPT.

Type of Correction	$\Delta E_{k \rightarrow a}^{xc}$
UKS	no correction
Loc. 1	$-\int \rho_k(\mathbf{r}) \frac{\delta \epsilon_{xc}^{LDA}(\mathbf{r})}{\delta \rho_{\uparrow}(\mathbf{r})} \rho_a(\mathbf{r}) d\mathbf{r}$
Loc. 2	$-\int \rho_k(\mathbf{r}) \frac{\delta v_{xc}^{\uparrow, LDA}(\mathbf{r})}{\delta \rho_{\uparrow}(\mathbf{r})} \rho_a(\mathbf{r}) d\mathbf{r}$
Loc. 3	$+ 2K_{ka} + \int \rho_k(\mathbf{r}) \left( \frac{\delta v_{xc}^{\uparrow, LDA}(\mathbf{r})}{\delta \rho_{\uparrow}(\mathbf{r})} + \frac{\delta v_{xc}^{\uparrow, LDA}(\mathbf{r})}{\delta \rho_{\downarrow}(\mathbf{r})} \right) \rho_a(\mathbf{r}) d\mathbf{r}$

In Tab. 3.I, all the SOS-DFPT approximations are shown together for comparison. We can see that all the local approximations contain different exchange and exchange-correlation integrals, although only Loc.3 has a strong physical root. As it will be shown in the section dedicated to the results, for singly bonded nuclei, whose NMR shielding is equally dependent on lower and higher excitations, the contribution given by all of these corrections is very similar. However, when the contribution from the lower excitations becomes more important, such as for unsaturated or delocalized systems, the contribution of the exchange-correlation corrections also becomes more important. Hence, in these cases we are able to differentiate the performance of the different SOS-DFPT approximations.

## **Part II**

# **Computational Methodology**

## Chapter 4

# Solution of the KS Equations: the *deMon-KS* Program

The solution of the integro-differential KS equations [see Eq.(1.26)] is the central part of each DFT program. The most widely implemented technique used to accomplish this task consists in defining the KS orbitals as a linear combination of atomic orbitals (LCAO). This scheme has been proposed for the first time in 1951 by C.C.J. Roothaan within the HF framework.<sup>158</sup> He used the atomic orbitals of the hydrogen atom as functions in the linear combination, whereas the atomic functions currently considered are rather different. The Slater-type orbitals,

$$\psi_{1s}^{\text{STO}}(\zeta, \mathbf{r}) = \left(\frac{\zeta^3}{\pi}\right)^{1/2} e^{-\zeta r} \quad (4.1)$$

where  $\zeta$  is the Slater orbital exponent, and the Gaussian-type orbitals,

$$\psi_{1s}^{\text{GTO}}(\alpha, \mathbf{r}) = \left(\frac{2\alpha}{\pi}\right)^{3/4} e^{-\alpha r^2} \quad (4.2)$$

where  $\alpha$  is the Gaussian orbital exponent, are now the functions most commonly employed\*. Both alternatives present advantages and drawbacks. The STOs reflect more closely the shape of the atomic orbitals, however their integration is rather involved. On the other hand, the GTOs have the wrong behavior at short distances (small  $\mathbf{r}$ ) and at long distances from the nucleus (large  $\mathbf{r}$ ). Nonetheless, they can be integrated much more easily than STOs, since, for example, the product of two GTOs centered on different atoms is, apart from a constant, another GTO centered on another atom.<sup>121</sup> The problem of choosing which function to implement can be bypassed using contracted Gaussian functions (CGF),

$$\phi_{\mu}^{\text{CGF}}(\mathbf{r}) = \sum_{p=1}^L d_{p\mu} \phi_p^{\text{GTO}}(\alpha_{p\mu}, \mathbf{r}) \quad (4.3)$$

where  $d_{p\mu}$  is a contraction coefficient which can be determined so that the CGF reflects more closely the STO radial behavior.

The program *deMon-KS*,<sup>159-161</sup> which has been used to perform all the SCF calculations presented in this work<sup>†</sup>, makes use of CGF to define the KS orbitals,

$$\psi_i^{\text{KS}}(\mathbf{r}) = \sum_{\nu} C_{\nu,i} \phi_{\nu}^{\text{CGF}}(\mathbf{r}) \quad (4.4)$$

$C_{\nu,i}$  represents the contraction coefficients which are determined in the SCF routine. Introducing Eq.(4.4) into the KS equation [see Eq.(1.26)], we obtain,

$$\left[ -\frac{1}{2} \nabla_i^2 + V_{\text{eff}}(\mathbf{r}_i) \right] \sum_{\nu} C_{\nu,i} \phi_{\nu}^{\text{CGF}}(\mathbf{r}_i) = \epsilon_i \sum_{\nu} C_{\nu,i} \phi_{\nu}^{\text{CGF}}(\mathbf{r}_i) \quad (4.5)$$

\*The STOs and GTOs corresponding to orbitals with higher angular momentum are obtained multiplying by a polynomial in the components of  $\mathbf{r}$  the exponential  $e^{-\zeta r}$  and  $e^{-\alpha r^2}$  respectively.

<sup>†</sup>*deMon-KS* version 3.5



By multiplying from the left the first term by an arbitrary function  $\phi_\mu^*$  and by integrating the equation over all space, we obtain †,

$$\sum_{\nu}^{N_{\nu}} C_{\nu,i} \int d\mathbf{r}_i \phi_{\mu}^*(\mathbf{r}_i) \left[ -\frac{1}{2} \nabla_i^2 + V_{eff}(\mathbf{r}_i) \right] \phi_{\nu}(\mathbf{r}_i) = \epsilon_i \sum_{\nu}^{N_{\nu}} C_{\nu,i} \int d\mathbf{r}_i \phi_{\mu}^*(\mathbf{r}_i) \phi_{\nu}(\mathbf{r}_i) \quad (4.6)$$

Thus, within the LCAO framework, we can solve the KS equations as a system of linear equations. The matrix form of the Eq.(4.6) results by identifying the terms,

$$F_{\mu\nu}^{KS} = \int d\mathbf{r}_i \phi_{\mu}^*(\mathbf{r}_i) \left[ -\frac{1}{2} \nabla_i^2 + V_{eff}(\mathbf{r}_i) \right] \phi_{\nu}(\mathbf{r}_i) \quad (4.7)$$

as Fock matrix elements and the terms,

$$S_{\mu} = \int d\mathbf{r}_i \phi_{\mu}^*(\mathbf{r}_i) \phi_{\nu}(\mathbf{r}_i) \quad (4.8)$$

as overlap matrix elements. Accordingly, we obtain a Roothaan-like matrix equation,

$$\mathbf{F}^{KS} \mathbf{C} = \mathbf{S} \mathbf{C} \boldsymbol{\epsilon} \quad (4.9)$$

where  $\mathbf{C}$  is the matrix containing the coefficient vectors and the  $\boldsymbol{\epsilon}$  is the eigenvalues diagonal matrix.

At this point, the solution of the KS matrix equation entails the same procedure followed in the HF method. Once the system under study is defined, the first step involves the calculation of the integrals making up the Fock matrix  $\mathbf{F}^{KS}$ . According to the definition of  $V_{eff}$  in Eq.(1.28), from Eq.(4.6) we can define three different integrals. The first is a one-electron (two-center) integral, which accounts

†From now on we will omit the superscript CGF and we will consider  $\phi_{\mu} \equiv \phi_{\mu}^{CGF}$ .

for the kinetic energy and for the attraction between electron  $i$  and the nuclei,

$$H_{\mu\nu}^{\text{core}} = \int d\mathbf{r}_i \phi_{\mu}^*(\mathbf{r}_i) \left[ -\frac{1}{2} \nabla_i^2 - \sum_A^M \frac{Z_A}{r_{iA}} \right] \phi_{\nu}(\mathbf{r}_i) \quad (4.10)$$

These terms are the elements of the so called core matrix  $\mathbf{H}^{\text{core}}$ . The calculation of these terms is straightforward and computationally inexpensive<sup>§</sup>, especially because it is performed only once at the beginning of the first SCF cycle.

The bottle-neck of the calculation is the evaluation of the two-electron (four-center) Coulomb integral,

$$J_{\mu\nu} = \sum_{\lambda}^{N_{\nu}} \sum_{\sigma}^{N_{\nu}} C_{\lambda,j} C_{\sigma,j} \iint d\mathbf{r}_i d\mathbf{r}_j \phi_{\mu}^*(\mathbf{r}_i) \phi_{\nu}(\mathbf{r}_i) \frac{1}{r_{ij}} \phi_{\lambda}^*(\mathbf{r}_j) \phi_{\sigma}(\mathbf{r}_j) \equiv \iint d\mathbf{r}_i d\mathbf{r}_j \phi_{\mu}^*(\mathbf{r}_i) \phi_{\nu}(\mathbf{r}_i) \frac{\rho(\mathbf{r}_j)}{r_{ij}} \quad (4.11)$$

and of the exchange-correlation integral,

$$V_{\mu\nu}^{\text{xc}} = \int d\mathbf{r}_i \phi_{\mu}^*(\mathbf{r}_i) V_{\text{xc}} \phi_{\nu}(\mathbf{r}_i) \quad (4.12)$$

The techniques implemented in *deMon-KS* to decrease the computational cost required by these calculations are described in the two following sections.

## 4.1 The Coulomb Integral

The evaluation of the Coulomb integral represents one of the highest hurdles, in the SCF routine. As we can observe from Eq.(4.11), it scales as  $O(N_{\nu}^4)$ , where  $N_{\nu}$  is the number of basis functions. One way to decrease the computational cost

<sup>§</sup>It scales as  $O(N_{\nu}^2)$ , where  $N_{\nu}$  is the number of basis functions.

is represented by expanding the electron density  $\tilde{\rho}(\mathbf{r})$  in a basis set of auxiliary functions,<sup>162,163</sup>

$$\tilde{\rho}(\mathbf{r}) = \sum_k^L c_k f_k(\mathbf{r}) \quad (4.13)$$

In particular, in *deMon-KS* the auxiliary functions are also contracted Gaussian functions.

The expansion coefficients are determined through the minimization of the Coulomb self-repulsion of the residual density,<sup>163</sup>

$$\iint d\mathbf{r}_i d\mathbf{r}_j [\rho(\mathbf{r}_i) - \tilde{\rho}(\mathbf{r}_i)] \frac{1}{r_{ij}} [\rho(\mathbf{r}_i) - \tilde{\rho}(\mathbf{r}_i)] \quad (4.14)$$

under the constraint,

$$\int d\mathbf{r} \tilde{\rho}(\mathbf{r}) = N \quad (4.15)$$

Hence, the true density  $\rho(\mathbf{r})$  can be approximated accurately by the fitted density  $\tilde{\rho}(\mathbf{r})$ . Accordingly, the Coulomb integral, defined in Eq.(4.11), can be rewritten as,

$$J_{\mu\nu} \approx \tilde{J}_{\mu\nu} = \sum_k^L c_k \iint d\mathbf{r}_i d\mathbf{r}_j \phi_\mu^*(\mathbf{r}_i) \phi_\nu(\mathbf{r}_i) \frac{f_k(\mathbf{r}_j)}{r_{ij}} \quad (4.16)$$

This procedure reduces the computational cost by approximately one order of magnitude. In fact, from  $O(N_\nu^4)$  we reached  $O(N_\nu^2 L)$  where  $L$  is the number of auxiliary basis functions [see Eq.(4.13)], which is generally the same as the number of the orbital basis functions.

The integral in Eq.(4.16) is evaluated analytically.<sup>160</sup>

## 4.2 The Exchange-Correlation Integral

On the other hand, the exchange-correlation integral [see Eq.(4.12)] cannot be evaluated analytically due to the complexity of the mathematics involved. Hence,

a numerical integration has to be carried out.

In order to reduce the computational cost, in *deMon-KS* the exchange-correlation, like the electron density, is fitted,

$$V_{xc} \approx \tilde{V}_{xc} = \sum_h^M a_h g_h(\mathbf{r}) \quad (4.17)$$

where  $g_h$  are the auxiliary basis functions and  $a_h$  are the coefficients. The procedure for the determination of the coefficients requires the minimization of the numerical integral,

$$\sum_p^P [\tilde{V}_{xc}(\mathbf{r}_p) - V_{xc}(\mathbf{r}_p)]^2 W(p) \quad (4.18)$$

The index  $p$  runs over the number of points of a grid built according to the algorithm of Becke.<sup>164</sup>  $W(p)$  is a weight factor proportional to the space associated to each grid point.<sup>160</sup> Once the expansion coefficients  $a_h$  are determined, the exchange-correlation potential  $V_{xc}$  can be rightfully expressed through Eq.(4.17). Hence, the integral in Eq.(4.12) becomes a three-center overlap integral, which can now be solved analytically according to the formulas of Obara and Saika.<sup>165</sup>

### **4.3 Response Module: Calculation of the Coupling Matrix Elements**

The response calculations have been performed using the *dynaRho* module (version 3.2)<sup>166</sup> of the *deMon* suite of programs. The time-dependent working equations, presented in Ch. 2, have been implemented using the multicenter orbital and auxiliary basis function method described by Casida.<sup>99</sup> The auxiliary function method is required for the calculation of the coupling matrix elements [see

Eq.(2.27)]. These are structured as the sum of two contributions, a Coulomb term and an exchange-correlation term,

$$K_{ij\tau,kl\sigma} = K_{ij\tau,kl\sigma}^{\text{Coul}} + K_{ij\tau,kl\sigma}^{\text{xc}} = \frac{\delta V_{ij\tau}^{\text{Coul}}}{\delta P_{kl\sigma}} + \frac{\delta V_{ij\tau}^{\text{xc}}}{\delta P_{kl\sigma}} \quad (4.19)$$

where  $P_{kl\sigma}$  are terms of the density matrix. The evaluation of  $K_{ij\tau,kl\sigma}^{\text{Coul}}$  requires the calculation of the derivative of the Coulomb potential, which matrix elements are expressed in terms of three-center integrals, thanks to the charge-density fitting basis  $f_I(\mathbf{r})$  [see Eq.(4.13)],

$$V^{\text{Coul}}(\mathbf{r}_k) = \sum_I^L c_I \int d\mathbf{r}_l \frac{f_I(\mathbf{r}_l)}{r_{kl}} \quad (4.20)$$

Therefore, the resulting exchange integral  $K_{ij\tau,kl\sigma}^{\text{Coul}}$  is also expressed in terms of a three-center integral.

In relation to the exchange-correlation term, the exchange-correlation potential  $V_{\text{xc}}$  is also expanded in terms of the auxiliary basis as shown in Eq.(4.17). Hence, the expansion coefficients  $\alpha_h$  are determined through the solution of the numerical integral in Eq.(4.18), which results in,

$$\alpha_h = \sum_u^M S_{u,h}^{-1} \left[ \sum_q^Q W_q g_u^*(\mathbf{r}) V_{\text{xc}} \right] = \sum_u^M S_{u,h}^{-1} \{g_u | V_{\text{xc}}\} \quad (4.21)$$

where the sum runs over the  $Q$  grid points.  $W_q$  are weight factors and  $S_{h,u}$  are elements of the overlap matrix,

$$S_{u,h} = [g_u | g_h] \quad (4.22)$$

In terms of auxiliary basis, the exchange-correlation kernel can be rewritten as,

$$K_{ij\tau,kl\sigma}^{\text{xc}} = \sum_{h,u}^M \langle \phi_i | g_h | \phi_j \rangle S_{u,h}^{-1} \left\{ g_u \left| \int d\mathbf{r}' \frac{\delta V_{ij\tau}^{\text{xc}}(\mathbf{r}, \omega)}{\delta \rho_{\sigma}(\mathbf{r}', \omega)} \phi_k(\mathbf{r}') \phi_l^*(\mathbf{r}') \right. \right\} \quad (4.23)$$

The integration of the exchange-correlation kernel can be rather complicated except in the ALDA case. In fact,  $f_{xc}^{ALDA}$  can be approximated by the functional derivative of the time-independent LDA exchange-correlation potential with respect to the density,

$$f_{xc}^{ALDA}(\mathbf{r}, \mathbf{r}', \omega) \approx \delta(\mathbf{r} - \mathbf{r}') \frac{\delta^2 [\rho \epsilon_{xc}(\rho)]}{\delta \rho^2} \quad (4.24)$$

where  $\epsilon_{xc}$  represents the xc energy per particle of a homogeneous electron gas (see Sec. 1.3.3). Thus, due to the delta function, the determination of the integral becomes trivial,

$$\int d\mathbf{r}' \frac{\delta V_{\tau}^{xc,LDA}(\mathbf{r}, \omega)}{\delta \rho_{\sigma}(\mathbf{r}', \omega)} \phi_{\mathbf{k}}(\mathbf{r}') \phi_{\mathbf{l}}^*(\mathbf{r}') = \frac{\delta V_{\tau}^{xc,LDA}(\mathbf{r}, \omega)}{\delta \rho_{\sigma}(\mathbf{r}, \omega)} \phi_{\mathbf{k}}(\mathbf{r}) \phi_{\mathbf{l}}^*(\mathbf{r}) \quad (4.25)$$

All the response calculations presented in this work have been performed within the ALDA formalism.

### 4.3.1 Implementation of the TDA in the *dynaRho* Program

The development of Loc.3 is based on the recognition that the introduction of the TDA restores the symmetry between the electric field and the magnetic field perturbations. According to the working equation shown in Sec. 2.3.3 and 3.3.2, some minor modifications to the time-dependent module are needed to implement the TDA. Essentially, the contribution of the B matrix to Eq.(2.59), on page 81, has to be set to zero. Consequently, the excitations are described by the general expression  $A F_I = \omega_I F_I$ , where  $\omega_I$  indicates the excitation energies and  $F_I$  represents a set of normalized eigenvectors. The A matrix is defined in Eq.(2.34). The full TDA SOS expression, shown in Eq.(3.67), represents a more accurate approximation for the paramagnetic shielding. However, the computational costs

### 4.3 Response Module: Calculation of the Coupling Matrix Elements 115

associated with it are prohibitive for routine application. The introduction of the 2LM, therefore, is necessary to derive an approximation analogous to the classic SOS-DFPT local corrections (i.e. Loc.3), which can be used to improve the calculation of the NMR shielding in systems larger than the ones accessible with a full TDA (CIS) approach. The 2LM routines had been already implemented in the *dynaRho* program by Casida.<sup>167</sup>

# Chapter 5

## Computational Details

In this chapter will be discussed the selection of the various alternative parameters made for the calculations presented in this work. These parameters include primarily the exchange-correlation functional, the auxiliary and orbital basis sets and the integration grid. Furthermore, some specifications on the threshold values will be given.

### 5.1 Exchange-Correlation Functionals

The geometry optimization of the small organic molecules, of the pyridine and water clusters and of the enzyme active-site model have been performed with the PLAP3 functional.<sup>79</sup> PLAP3 originates from the synchronization of the LAP3 correlation functional, which includes parallel spin correlation, with the GGA exchange counterpart PW86.<sup>50</sup> The 3 indicates the number of parameters used in the correlation functional. This choice has been based on the fact that the PLAP3 functional has shown significant ability in the description of hydrogen-bond net-



works<sup>168</sup> and, more in general, of weakly interacting systems.<sup>169</sup> Moreover, a test has been performed to evaluate the performance of the PLAP3 and BLAP3\* functionals in determining the bond lengths of 17 small organic molecules containing different types of nitrogen atoms, because the major role of the geometry optimizations in this work was to refine the molecular environment especially around the nitrogen atoms. Tab. 5.I shows the results. While for the NH bond the performance of the two functionals is extremely close and not even the choice of basis-set makes much of a difference, for the CN single bond BLAP3 in combination with the TZVP basis-set (see Sec. 5.2 for details) seems to be slightly better. Nonetheless, the CN triple bonds are described better by PLAP3 in combination with the TZVP basis-set. The overall performance between BLAP3 and PLAP3 is fairly close, however considering also the tests published in the literature, PLAP3 seemed to be the best choice.

Table 5.I: Absolute average deviation ( $\text{\AA}$ ) from experimental references of bond lengths calculated with different  $V_{xc}$  and different basis-sets.

	DZVP		TZVP	
	BLAP3	PLAP3	BLAP3	PLAP3
N-H	0.0105	0.0113	0.0106	0.0109
C-N Single	0.0118	0.0171	0.0067	0.0103
C=N Double <sup>†</sup>	0.0060	0.0103	0.0037	0.0059
CN Triple	0.0045	0.0060	0.0030	0.0015
NN	0.0183	0.0255	0.0107	0.0167
All bonds	0.0102	0.0140	0.0069	0.0091

All the optimizations have been performed with the Broyden-Fletcher-Goldfarb-

\*LAP3 in combination with the B88 exchange

<sup>†</sup>or Aromatic

Shanno (BFGS) algorithm.<sup>170</sup> The gradient threshold has been set to  $1 \times 10^{-5}$ , while the step has been kept at 0.05 Bohrs.

As for the SCF calculations, preceding the NMR shielding calculation, the exchange-correlation potential of choice has been the LDA, with Slater exchange (S) and Volko-Wilk-Nusair (VWN) correlation.<sup>36</sup> The decision to use the SVWN functional for the NMR shielding calculations was based on a consistency argument. The exchange-only (i.e. Loc.1 and Loc.2) and the exchange-correlation (i.e. Loc.3) local corrections in SOS-DFPT are all defined within the LDA framework. Hence, it would not be physically rigorous to define the KS orbitals and their eigenvalues in term of a GGA or hybrid functional, while the correction of the excitation energy is determined within the LDA. Nonetheless, in order to evaluate the performance of the LDA within SOS-DFPT, we performed some test calculations with other functionals.

## 5.2 Basis-Sets

The orbital basis set chosen to perform the geometry optimization of the small organic systems and of the pyridine and water clusters is a triple-zeta with polarization functions (TZVP). In particular, the scheme (4s3p1d), with contraction pattern (7111/411/1\*), has been used for heavy atoms (e.g. C, N, O, F), while (2s1p), with contraction pattern (41/1\*), has been used for hydrogen. For the geometry optimization of the model of the active-site of the serine protease enzyme, two types of orbital basis sets have been selected. An STO-3G quality basis set has been used in the first phase of the optimization in order to speed up the minimum search. Whenever a stable structure had been found, then a second optimization

with the TZVP basis was performed to refine it. For the NMR calculations the orbital basis used is the IGLO-III,<sup>122</sup> consisting of (11s7p2d), with contraction (5111111/2111111/11), for heavy atoms and (4s2p), with contraction (4111/11) for hydrogen. This loosely contracted and highly diffuse basis set has been developed to work within the IGLO framework for the solution of the gauge origin problem.

As for auxiliary basis sets, a (5,2;5,2) scheme has been used for heavy atoms, while the (5,1;5,1) scheme has been used for hydrogen. According to the nomenclature ( $k_1, k_2; l_1, l_2$ ),  $k_1$  ( $l_1$ ) refers to the number of type s gaussian functions used for the fitting of the density (exchange-correlation potential), while  $k_2$  ( $l_2$ ) refers to the number of type s, p and d gaussian functions constrained to the same exponent, used for the fitting of the density (exchange-correlation potential).

The coefficients of all auxiliary and orbital basis sets have been optimized specifically for DFT calculations<sup>171</sup> and have been taken from the *deMon* basis database.

### 5.3 Integration Grid

As it has been described in Ch. 4, the programs *deMon-KS* and *deMon-dynaRho* make use of numerical grids to reduce the computational cost of the four-center Coulomb integrals and to solve the integrals involving the exchange-correlation potential. The numerical grid is generated through Becke's algorithm.<sup>164</sup> The chosen radial grid consists of 64 points per atom (i.e. 64 radial shells per atom), while the angular grid is defined as FINE (i.e. 26 angular points per radial shell). The benchmark NMR calculations tests have been performed with an EXTRAFINE

angular grid (194 points per radial shell). The NONRANDOM option has been chosen, which forces the angular points of the radial shells to line up.

In all calculations, at the SCF convergence, a numerical extra iteration is done without the fitting of the  $V_{xc}$ . This step is mostly important in the SOS-DFPT calculation,<sup>4</sup> but it also can improve the determination of the geometry and of the total energy.



## **Part III**

# **Results and Discussion**



## Chapter 6

# Strengths and Limitations of the Loc.3 Correction

In Sec. 3.3.2 we have seen how the Loc.3 correction has been derived formally. Most importantly, it has been shown that, with the introduction of the 2LM approximation, the Loc.3 expression embodies the physical basis of SOS-DFPT. Actually, as we can see in Tab. 3.I, all of the local corrections in SOS-DFPT contain only exchange and exchange-correlation integrals.

In this chapter I will present the results of two series of benchmark tests performed to establish the capabilities of the Loc.3 approximation in the calculation of the NMR shielding constant of different nuclei. In the first series, NMR shielding constants for  $^{13}\text{C}$ ,  $^{14,15}\text{N}$  and  $^{17}\text{O}$  are evaluated with SOS-DFPT only. Consequently, the behavior of Loc.3 will be compared exclusively to that of the UKS, Loc.1 and Loc.2 approximations. In the second series of tests, the performance of Loc.3 will be assessed against some of the most recent and reliable DFT-based methodologies for NMR shielding calculation. The results of these

calculations prove that not only does the Loc.3 correction play an essential role in characterizing the physical foundations of SOS-DFPT, but that it also represents a considerable improvement over Loc.1 and Loc.2 \*.

## 6.1 Loc.3 vs. Classic SOS-DFPT

In Tab. A.I on page 238 are shown the results of the calculation of the  $^{13}\text{C}$  isotropic shielding constants obtained with all the different SOS-DFPT approximations. The results are compared to high quality *ab initio* calculations and experimental absolute shieldings. The geometries of the test molecules are the same as those of the reference values.<sup>141,142,175</sup> The LDA exchange-correlation functional has been used here not only in the evaluation of the local corrections but also throughout the SCF calculation. These results confirm the superiority of the local corrections over UKS. Nonetheless, it is important to underline that there is not much difference in the performance of the four approximations for the shielding of aliphatic carbon atoms, for which the average deviation goes from 7.8 ppm for UKS down to 5.7 ppm for Loc.3. The quality of the results begins to differ when we are dealing with unsaturated carbon atoms. Very good results are obtained with the local corrections, and especially with Loc.3, for the  $^{13}\text{C}$  shielding of cyanides and isocyanides. However, a considerable problem is represented by the highly deshielded carbonyl groups. Actually, most of the maximum deviation values are registered for formaldehyde and acetic aldehyde. As a matter of fact, the NMR shielding of the carbonyl carbon is also a big challenge for the GIAO-DFT and

---

\*The large majority of the results presented in this chapter have been published also in the two articles, E. Fadda and coworkers<sup>157</sup> and E. Fadda and coworkers.<sup>172</sup>

GIAO-MP2 methodologies.<sup>176</sup>

In the case of CO we notice that Loc.3 fails compared to Loc.1 and Loc.2. We think that the reason for this is the breakdown of the 2LM approximation. In fact, in order to describe the excited state of CO more than two orbitals are needed, since the coupling between the x- and y-components of the  $\pi$  orbital generate a spatial multiplet ( $\Sigma^+ + \Sigma^- + \Delta$ ).<sup>172</sup> An overall analysis of the LDA results is shown in Tab. 6.I.

Table 6.I: Statistical analysis of the SOS-DFPT (SVWN) and high quality *ab initio*  $^{13}\text{C}$  absolute shieldings, relative to experimental references. See Tab. A.I for details.

	SVWN				Ab Initio
	UKS	Loc.1	Loc.2	Loc.3	
Average Abs. Dev.	15.6	12.0	11.2	11.5	5.6
Max. Deviation	35.3*	29.9 <sup>†</sup>	26.1 <sup>‡</sup>	24.4 <sup>‡</sup>	10.5*
Intercept	-27.0	-19.7	-17.5	-14.1	5.7
Slope	1.1255	1.0866	1.0749	1.0525	0.9984
Corr. Coeff.	0.9981	0.9970	0.9965	0.9923	0.9995

The LDA functional has been declared by many<sup>60,156,173</sup> as a rather mediocre choice for NMR parameter calculations. In our opinion, this statement is somewhat excessive. In fact, all of the aforementioned studies evaluate their calculations on the shift scale, where the standard reference ( $^{13}\text{C}$  absolute shielding of TMS or  $\text{CH}_4$ ) is also calculated. It turns out that the LDA (and HF) give the best performance in the calculation of the standard references, when compared to GGAs (BPW91 and BLYP) and to hybrid functionals (B3PW91 and B3LYP).<sup>173</sup> Thus, the absolute shieldings obtained with these more sophisticated function-

\*HCHO  
<sup>†</sup>CH<sub>3</sub>CHO  
<sup>‡</sup>CF<sub>4</sub>



als are “corrected” in the transition from the absolute to the shift scale.<sup>177</sup> As an example, we tested the PW91-PW91 exchange-correlation functional in the calculation of the <sup>13</sup>C isotropic shielding constants of the same molecules as in Tab. A.I. The results obtained are shown in Tab. A.II. As it has been pointed out recently, the use of a GGA, or any other functional but the LDA, within SOS-DFPT is not physically rigorous.<sup>172,178</sup> All the local corrections are in fact dependent on the functional derivative of the  $V_{xc}$ , for which the only practical analytical form is obtained with the LDA. Therefore, the use of a GGA influences only the KS orbitals and their energy, while the  $\Delta E_{xc}$  correction is defined within the LDA. However, from this test we can derive useful information on the role of different  $V_{xc}$ .

Table 6.II: Statistical analysis of the SOS-DFPT (PW91-PW91) and high quality *ab initio* <sup>13</sup>C absolute shieldings, relative to experimental references. See Tab. A.II for details.

	PW91-PW91				Ab Initio
	UKS	Loc.1	Loc.2	Loc.3	
Average Abs. Dev.	15.9	10.7	10.1	10.1	5.6
Max. Deviation	27.3*	23.1 <sup>†</sup>	22.7 <sup>†</sup>	21.4 <sup>†</sup>	10.5 <sup>‡</sup>
Intercept	-21.2	-14.4	-12.2	-10.0	5.7
Slope	1.0780	1.0423	1.0311	1.0171	0.9984
Corr. Coeff.	0.9985	0.9971	0.9965	0.9943	0.9995

As we can determine from the statistical analysis shown in Tab. 6.II, there is not a significant difference in the performance of the PW91-PW91 compared to the SVWN. This should not be very surprising, in view of the fact that the quality of the calculated NMR shieldings is largely influenced by how well the

\*CH<sub>3</sub>COCH<sub>3</sub>

<sup>†</sup>CF<sub>4</sub>

<sup>‡</sup>HCHO

paramagnetic term is evaluated. The latter depends closely upon the quality of the description of the excited states and especially of the valence excited states.<sup>179</sup> As it has been determined by several TDDFT studies,<sup>180-182</sup> the ability in predicting accurate excitation energies of LDA, GGA and hybrid functionals are fairly similar. Problems with the LDA functional arise only when the excitation energy is approximated simply by the difference of the KS orbitals eigenvalues, as it is done in UKS.

Table 6.III: Statistical analysis of the SOS-DFPT (ACLDA) and high quality *ab initio* <sup>13</sup>C absolute shieldings, relative to experimental references. See Tab. A.III for details.

	ACLDA				Ab Initio
	UKS	Loc.1	Loc.2	Loc.3	
Average Abs. Dev.	18.2	15.4	13.7	14.5	5.6
Max. Deviation	47.3*	38.7*	36.0*	34.1*	10.5 <sup>†</sup>
Intercept	-30.2	-22.9	-20.7	-18.3	5.7
Slope	1.1307	1.0927	1.0811	1.0671	0.9984
Corr. Coeff.	0.9945	0.9928	0.9921	0.9898	0.9995

The proof that the accuracy of the NMR shielding is highly related to the ability of the functional in describing the lower excitations, is given by the results obtained with the asymptotically-corrected LDA (ACLDA).<sup>87</sup> These are shown, for <sup>13</sup>C, in Tab. A.III with the statistical analysis in Tab. 6.III. As it has been discussed in Sec. 1.3.8, the correction of the asymptotic behavior allows the right description of the high lying Rydberg states. Therefore, the fact that the quality of the results is not appreciably influenced by this correction, indicates that only the low lying states, below half of the ionization threshold, contribute significantly to

\*CH<sub>3</sub>CHO

<sup>†</sup>HCHO

the paramagnetic shielding value.<sup>60, 157, 172, 179</sup>

In order to summarize the results obtained from SOS-DFPT calculations of the  $^{13}\text{C}$  shielding, we can examine Fig. 6.1 on page 137. This shows the performance of all the SOS-DFPT approximations and of the *ab initio* calculations against the experimental results. The colours indicate the different functionals used in the SOS-DFPT calculation. We note that, for high values of the absolute shielding, it is rather difficult to recognize any of the SOS-DFPT symbols. In fact, this portion of the graph corresponds to aliphatic carbon atoms, for which all the SOS-DFPT local corrections are negligible. For the lower-intermediate values of the absolute shielding, corresponding to unsaturated carbons, the difference in the performance of the four SOS-DFPT approximations becomes more visible. We note that the Loc.2 and Loc.3 results appear to be less deshielded than the Loc.1 and UKS. The most significant problems appear to be in the left-hand corner of the graph. This portion, in fact, corresponds to the carbonyl carbon nuclei, for which we have significant deviations. Here the difference with the *ab initio* calculations is quite remarkable.

The results obtained in the SOS-DFPT calculation of the  $^{14,15}\text{N}$  shielding with the LDA functional are shown in Tab. A.IV. In contrast to the previous case, here there is a significant difference in the performance of the four SOS-DFPT approximations. The local corrections are a considerable improvement over UKS, especially for the unsaturated nitrogen nuclei. Moreover, as we can see from the statistical analysis shown in Tab. 6.IV on page 128, among the local corrections Loc.3 gives the absolute best performance. The only difficult case for Loc.3 is represented by the external nitrogen nucleus of diazomethane, which shows the maximum deviation. As a matter of fact, the NMR shielding of this nucleus rep-

resents a rather difficult case for all non-perturbative approaches, due to the large correlation effects. In order to achieve quantitative accuracy it is necessary to make use of highly correlated methods, such as CCSD with a perturbative correction for connected triple excitations (CCSD(T)).<sup>139,142</sup> A similar case is given by the external nitrogen nucleus of N<sub>2</sub>O. However, here Loc.2 and Loc.3 give a much more accurate result than the *ab initio* reference calculation (MP2 level).

The influence of the functional has been assessed also in this case by the use of the PW91-PW91 in addition to SVWN. The <sup>14,15</sup>N shielding values are shown in Tab. A.V, while the statistical analysis is shown in Tab. 6.V on page 129. Similarly to what we have seen for the <sup>13</sup>C shielding case, the performances of SVWN and of PW91-PW91 are very close. The only remarkable changes are the improved accuracy in the evaluation of the <sup>14,15</sup>N shielding of HCN counterbalanced by the loss of accuracy in the case of N<sub>2</sub>.

Table 6.IV: Statistical analysis of the SOS-DFPT (SVWN) and high quality *ab initio* <sup>14,15</sup>N absolute shieldings, relative to experimental references. See Tab. A.IV for details.

	SVWN				Ab Initio
	UKS	Loc.1	Loc.2	Loc.3	
Average Abs. Dev.	19.3	10.7	9.0	6.3	14.4
Max. Deviation	33.8*	20.1 <sup>†</sup>	17.3 <sup>‡</sup>	24.2 <sup>‡</sup>	35.6 <sup>§</sup>
Intercept	-19.8	-7.7	-4.0	2.6	14.2
Slope	1.0522	1.0042	0.9892	0.9726	1.0165
Corr. Coeff.	0.9973	0.9966	0.9962	0.9972	0.9961

The influence of the asymptotic correction in the calculation of the <sup>14,15</sup>N

\*HCN

<sup>†</sup>CH<sub>3</sub>CN

<sup>‡</sup>CH<sub>2</sub>NN

<sup>§</sup>NNO

shielding seems to be more significant than in the case of the  $^{13}\text{C}$  shielding. The results obtained with the ACLDA functional (see Tab. A.VI for the calculation results and Tab. 6.VI for the statistical analysis) are considerably less accurate compared to the ones obtained with the LDA and the PW91-PW91 functionals. The single-bonded nitrogen in  $\text{NH}_3$  appears to be the least influenced, while the accuracy of the shielding of all the other nitrogen atoms, which are double and triple-bonded, is heavily compromised. The worse cases are given by the  $^{14,15}\text{N}$  shielding of  $\text{N}_2$  and  $\text{CH}_3\text{CN}$ . For these systems the “best” absolute average deviations, calculated as Loc.3 vs. Exp., are 28.3 and 27.6 ppm respectively. On the other hand, with the LDA functional we obtained an absolute average deviation (Loc.3 vs. Exp.) of 3.6 and 6.0 ppm respectively.

Table 6.V: Statistical analysis of the SOS-DFPT (PW91-PW91) and high quality *ab initio*  $^{14,15}\text{N}$  absolute shieldings, relative to experimental references. See Tab. A.V for details.

	PW91-PW91				Ab Initio
	UKS	Loc.1	Loc.2	Loc.3	
Average Abs. Dev.	21.8	11.0	9.3	7.0	14.4
Max. Deviation	37.7*	27.8*	24.6*	17.3*	35.6 <sup>†</sup>
Intercept	-22.5	-10.6	-6.9	-0.3	14.2
Slope	1.0601	1.0117	0.9968	0.9789	1.0165
Corr. Coeff.	0.9978	0.9971	0.9967	0.9971	0.9961

All the calculated  $^{14,15}\text{N}$  shieldings are shown in Fig. 6.2 on page 138 for comparison. The analysis of the results, founded on the statistical data, is confirmed here graphically. We can clearly recognize the difference in the performance of the four SOS-DFPT approximations. Loc.3 tends to be the closest SOS-DFPT

\* $\text{CH}_3\text{CN}$

<sup>†</sup> $\text{NNO}$

approximation to the bisecting line, especially when calculated with the SVWN functional. Moreover, for difficult nuclei, such as CH<sub>3</sub>CN, CH<sub>2</sub>NN and NNO, Loc.3 shows a much higher accuracy than the *ab initio* reference.

Table 6.VI: Statistical analysis of the SOS-DFPT (ACLDA) and high quality *ab initio* <sup>14,15</sup>N absolute shieldings, relative to experimental references. See Tab. A.VI for details.

	ACLDA				
	UKS	Loc.1	Loc.2	Loc.3	Ab Initio
Average Abs. Dev.	33.9	21.2	17.6	13.6	14.4
Max. Deviation	59.7*	42.5 <sup>†</sup>	38.3 <sup>†</sup>	28.4 <sup>‡</sup>	35.6 <sup>§</sup>
Intercept	-34.9	-21.6	-17.8	-3.7	14.2
Slope	1.0895	1.0389	1.0181	0.9713	1.0165
Corr. Coeff.	0.9947	0.9949	0.9949	0.9887	0.9961

Oxygen is the last nucleus to be examined in this section and it represents also the most difficult case. The error margins here are considerably broader than for the <sup>13</sup>C and <sup>14,15</sup>N shielding, due to the larger NMR scale. As an example, we can observe the absolute error associated with the experimental reference shown in Tab. A.VII together with the SOS-DFPT and *ab initio* results. We can appreciate the level of uncertainty also in the results of the statistical analysis shown in Tab. 6.VII. Even though the absolute average deviation is fairly high for all the SOS-DFPT approximations, compared to the deviations evaluated for the *ab initio* results, we note the improvement gained with Loc.3. In fact, the absolute average deviation is reduced and the maximum deviation, corresponding to OF<sub>2</sub> for all the SOS-DFPT approximations, is almost cut in half, when compared to the second

\*N<sub>2</sub>

<sup>†</sup>CH<sub>3</sub>CN

<sup>‡</sup>CH<sub>2</sub>NN

<sup>§</sup>NNO

“best” maximum deviation coming from Loc.2.

Table 6.VII: Statistical analysis of the SOS-DFPT (SVWN) and high quality *ab initio*  $^{17}\text{O}$  absolute shieldings, relative to experimental references. See Tab. A.VII for details.

	SVWN				Ab Initio
	UKS	Loc.1	Loc.2	Loc.3	
Average Abs. Dev.	79.7	52.2	43.9	33.0	16.8
Max. Deviation	202.3*	130.0*	108.6*	56.1*	43.0 <sup>†</sup>
Intercept	-77.0	-50.7	-42.7	-27.1	-14.6
Slope	1.1923	1.1071	1.0818	1.0430	1.0193
Corr. Coeff.	0.9980	0.9984	0.9985	0.9981	0.9989

Table 6.VIII: Statistical analysis of the SOS-DFPT (PW91-PW91) and high quality *ab initio*  $^{17}\text{O}$  absolute shieldings, relative to experimental references. See Tab. A.VIII for details.

	PW91-PW91				Ab Initio
	UKS	Loc.1	Loc.2	Loc.3	
Average Abs. Dev.	67.3	41.9	34.2	22.3	16.8
Max. Deviation	164.1*	98.6*	79.0*	34.4*	43.0 <sup>†</sup>
Intercept	-65.6	-41.2	-33.8	-21.5	-14.6
Slope	1.1199	1.0445	1.0218	0.9845	1.0193
Corr. Coeff.	0.9975	0.9978	0.9978	0.9991	0.9989

Some of the oxygen nuclei analyzed are more sensitive to the change of functional than nitrogen or carbon. From the data obtained with the PW91-PW91 functional, shown in Tab. A.VIII, we notice a significant improvement in the accuracy of the  $^{17}\text{O}$  shielding of  $\text{CH}_3\text{CHO}$ ,  $\text{HCHO}$  and  $\text{OF}_2$ . This improvement is reflected also in the statistical analysis shown in Tab. 6.VIII. However, the  $^{17}\text{O}$

\* $\text{OF}_2$

<sup>†</sup> $\text{HCHO}$

shielding of the other oxygen nuclei is influenced rather little by the use of the GGA. In fact, if we take into consideration the width of the  $^{17}\text{O}$  NMR shielding scale, a variation of 5 ppm can be definitely considered negligible.

As for the asymptotic correction, we note a significant effect in  $\text{CH}_3\text{CHO}$  and  $\text{CH}_3\text{COCH}_3$ , where it causes a down-shift (i.e.  $^{17}\text{O}$  shieldings too deshielded). On the other hand, in the case of  $\text{OF}_2$  the AC causes a very strong up-shift of over 100 ppm for all the SOS-DFPT approximations (see Tab. A.IX for the calculation results and Tab. 6.IX for the statistical analysis). The correction in the  $^{17}\text{O}$  shielding of  $\text{OF}_2$  makes HCHO the most difficult case for UKS, Loc.1 and Loc.2. Nonetheless,  $\text{OF}_2$  remains the most difficult case for Loc.3, since the corresponding  $^{17}\text{O}$  shielding value is much too shielded compared to the experimental reference.

Table 6.IX: Statistical analysis of the SOS-DFPT (ACLDA) and high quality *ab initio*  $^{17}\text{O}$  absolute shieldings, relative to experimental references. See Tab. A.IX for details.

	ACLDA				Ab Initio
	UKS	Loc.1	Loc.2	Loc.3	
Average Abs. Dev.	49.0	32.3	32.3	39.4	16.0
Max. Deviation	131.1*	76.4*	59.8*	68.5 <sup>†</sup>	43.0*
Intercept	-48.3	-25.9	-19.1	-8.8	-14.6
Slope	1.0513	0.9851	0.9649	0.9382	1.0193
Corr. Coeff.	0.9949	0.9954	0.9956	0.9919	0.9989

Fig. 6.3 on page 139 summarizes all the results obtained in the calculation of the  $^{17}\text{O}$  shielding with SOS-DFPT. The *ab initio* reference values are also shown for comparison. This diagram has been divided in two sections in order to show in detail both the down-field and up-field regions of the spectrum. In the down-

\*HCHO

<sup>†</sup> $\text{OF}_2$



field portion, we can recognize the significant difficulty suffered by SOS-DFPT in the evaluation of the  $^{17}\text{O}$  shielding of  $\text{OF}_2$  and  $\text{HCHO}$ . In these two cases, as it has been underlined previously, the type of functional makes a considerable difference, while for the other systems in the up-field region we see that this choice is less relevant.

## 6.2 Loc.3 vs. other DFT-based Methodologies

The good performance shown by the Loc.3 approximation in comparison to the other “classic” SOS-DFPT corrections has encouraged us to test it against some of the more recent and promising DFT-based methods for NMR shielding calculation.<sup>172</sup> Among these we chose the multiplicative Kohn-Sham (MKS) method of Wilson and Tozer,<sup>155</sup> the  $\text{B3LYP}_{\text{GGA}}^{0.05}$  functional of Wilson and coworkers,<sup>60</sup> the PBE0 functional used in the calculation of Adamo and Barone<sup>66</sup> and the self-interaction-corrected VWN (SIC-VWN) functional of Patchkovskii and coworkers.<sup>156</sup> Furthermore, we compare the Loc.3 results against two corrected LDA functionals, which have been developed very recently for NMR calculations.<sup>179</sup> The correction to  $E_{\text{xc}}^{\text{LDA}}$  consists in an additional exchange gradient contribution adjusted by two empirically determined parameters (i.e. KT1 functional). In the second functional, namely KT2, the LDA exchange and correlation terms are fine-tuned through the introduction of least-squares fitted parameters.

For this test we calculated the  $^{13}\text{C}$ ,  $^{14,15}\text{N}$  and  $^{17}\text{O}$  NMR shielding on the same molecules which have been analyzed in the previous section. Moreover, the Loc.3 calculations presented here have been all obtained with the LDA functional. The main reasons being that, firstly, as it has been proven in the previous tests, the

accuracy of the LDA functional, within the SOS-DFPT framework, is comparable in most cases to that obtained with the GGA PW91-PW91. Moreover, the choice of the LDA functional for the SCF routine preceding the SOS-DFPT calculation is more physically rigorous (see Sec. 5.1 on page 116).

The first nucleus to be considered is  $^{13}\text{C}$ . The results are shown in Tab. A.X and A.XI on pages 247 and 248 respectively. The statistical analysis is shown in Tab. 6.X.

Table 6.X: Statistical analysis of the  $^{13}\text{C}$  absolute shieldings calculated with the SOS-DFPT Loc.3 (LDA), other DFT-based and high quality *ab initio* methods, relative to experimental references. See Tab. A.X and A.XI for details.

	Ave. Abs. Dev.	Max. Dev.	Intercept	Slope	Correl.
<b>Loc.3</b>	<b>11.5</b>	<b>24.4*</b>	<b>-14.1</b>	<b>1.0525</b>	<b>0.9923</b>
MKS	4.8	10.3 <sup>†</sup>	-6.4	1.0169	0.9990
KT1	4.5	9.8 <sup>‡</sup>	7.0	0.9739	0.9990
KT2	3.8	6.8 <sup>‡</sup>	5.1	0.9775	0.9994
B3LYP <sup>00.5</sup>	2.0	4.1 <sup>‡</sup>	1.0	0.9887	0.9995
PBE0	3.4	7.3 <sup>§</sup>	-5.2	1.0238	0.9994
SIC-VWN	6.8	19.4 <sup>‡</sup>	-9.7	1.0368	0.9979
<i>Ab initio</i>	5.6	10.5 <sup>‡</sup>	5.7	0.9984	0.9995

The MKS data considered here are those obtained with the electron density calculated with the B97-1 functional,<sup>48</sup> a reparametrized version of Becke's B97 functional.<sup>184</sup> The performance of the Loc.3 approximation is definitely the weakest among these different approaches. Its problems are related mainly to the  $^{13}\text{C}$  shielding of carbonyl carbons, of CO and of CF<sub>4</sub>, which shows the maximum de-

\*CF<sub>4</sub>

†HCHO

‡CO

§CH<sub>2</sub>CCH<sub>2</sub>

viation. All the other approaches also have problems with carbonyl systems. In fact, both the maximum deviations of MKS(B97-1) and of SIC-VWN correspond to HCHO. Moreover, the KT1, KT2 and B3LYP<sup>0.05</sup> functionals show difficulties with CO, although to a much lesser extent than Loc.3.

The picture changes dramatically moving to the <sup>14,15</sup>N nucleus. The results are shown in in Tab. A.XII and A.XIII, while the statistical analysis is on Tab. 6.XI.

Table 6.XI: Statistical analysis of the <sup>14,15</sup>N absolute shieldings calculated with the SOS-DFPT Loc.3 (LDA), other DFT-based and high quality *ab initio* methods, relative to experimental references. See Tab. A.XII and A.XIII for details.

	Ave. Abs. Dev.	Max. Dev.	Intercept	Slope	Correl.
<b>Loc.3</b>	<b>6.3</b>	<b>24.2*</b>	<b>2.6</b>	<b>0.9726</b>	<b>0.9972</b>
MKS	4.8	11.5*	-1.3	0.9848	0.9990
KT1	6.5	20.7*	6.9	0.9724	0.9991
KT2	2.7	10.6*	2.9	0.9863	0.9997
B3LYP <sup>00.5</sup>	6.5	28.4*	6.9	0.9571	0.9979
PBE0	8.8	15.3 <sup>†</sup>	-13.8	1.0504	0.9999
SIC-VWN	7.6	13.9 <sup>‡</sup>	-5.7	1.0032	0.9982
<i>Ab initio</i>	14.4	35.6 <sup>‡</sup>	14.2	1.0165	0.9961

Here we see that Loc.3 represents a very competitive approach compared to the others. For instance, the external nitrogen in CH<sub>2</sub>N<sub>2</sub> represent the most difficult case for most of the functional tested. Nonetheless, Loc.3 gives a better evaluation of its <sup>14,15</sup>N shielding than B3LYP<sup>0.05</sup>. Furthermore, as it has been underlined also in the previous section, Loc.3 performs very well in the case of cyanides, isocyanides and in the case of N<sub>2</sub>O even when compared to these more

\*CH<sub>2</sub>NN

†CH<sub>3</sub>CN

‡NNO

sophisticated functionals. The latter represents the most difficult case for the SIC-VWN functional, while the weakness of PBE0 is represented by cyanide groups.

Also in the case of the  $^{17}\text{O}$  shielding the Loc.3 approximation performs rather well. As we can see from the results (see Tab. A.XIV and A.XV) and from the statistical analysis (see Tab. 6.XII), Loc.3 shows a level of accuracy comparable to PBE0. The KT1 and KT2 functionals have considerable problems with  $\text{OF}_2$  and  $\text{H}_2\text{O}$ . These large deviations influence negatively an otherwise very good behavior. It is important to underline that the disappointing results of the statistical analysis of the SIC-VWN results are mainly due to the extremely big deviation recorded for  $\text{OF}_2$ .

Table 6.XII: Statistical analysis of the  $^{17}\text{O}$  absolute shieldings calculated with the SOS-DFPT Loc.3 (LDA), other DFT-based and high quality *ab initio* methods, relative to experimental references. See Tab. A.XIV and A.XV for details.

	Ave. Abs. Dev.	Max. Dev.	Intercept	Slope	Correl.
<b>Loc.3</b>	<b>33.0</b>	<b>56.1*</b>	<b>-27.1</b>	<b>1.0430</b>	<b>0.9981</b>
MKS	19.6	31.9*	-14.9	0.9872	0.9989
KT1	22.3	43.6*	-22.2	1.0089	0.9994
KT2	26.5	60.9*	-26.2	1.0155	0.9987
B3LYP <sup>00.5</sup>	19.6	29.4*	-10.4	0.9792	0.9983
PBE0	33.2	47.2 <sup>†</sup>	-35.2	1.0289	0.9999
SIC-VWN	53.6	224.7*	-38.6	1.1220	0.9774
<i>Ab initio</i>	16.8	43.0 <sup>†</sup>	14.6	1.0193	0.9989

\* $\text{OF}_2$

<sup>†</sup> $\text{HCHO}$

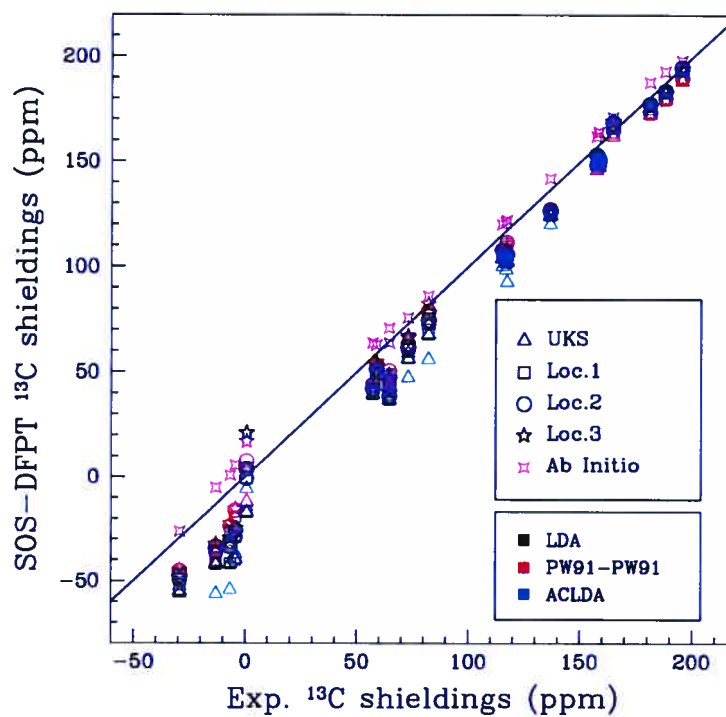


Figure 6.1:  $^{13}\text{C}$  NMR shieldings from SOS-DFPT and *ab initio* calculations compared to experimental results. The color of the symbols depends upon the functional chosen

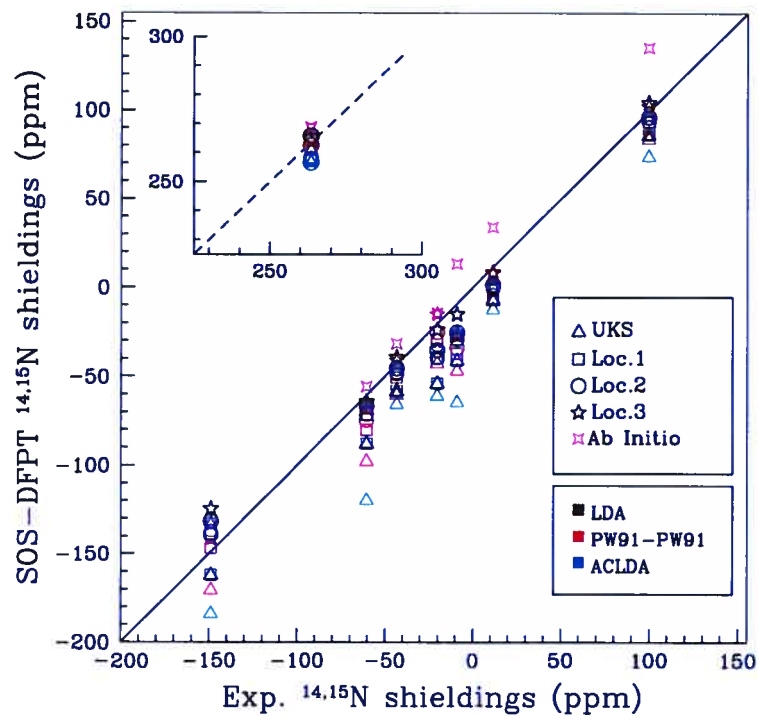


Figure 6.2:  $^{14,15}\text{N}$  NMR shieldings from SOS-DFPT and *ab initio* calculations compared to experimental results. The color of the symbols depends upon the functional chosen. The results obtained for  $\text{NH}_3$  are shown on the left-top corner of the graph.

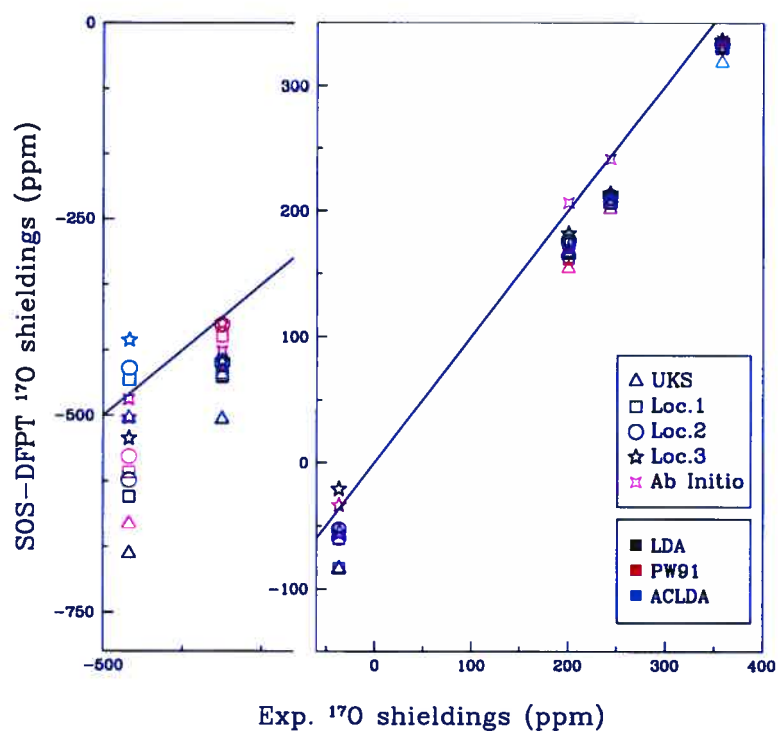


Figure 6.3:  $^{17}\text{O}$  NMR shieldings from SOS-DFPT and *ab initio* calculations compared to experimental results. The color of the symbols depends upon the functional chosen. The left portion of the diagram shows the down-field region of the spectrum, where are indicated the results for  $\text{OF}_2$  and  $\text{HCHO}$ . Here the small ticks mark 100 ppm. The bisecting line is shown in both regions of the diagram.

## Chapter 7

# Nitrogen Shielding in Different Molecular Structures

The most significant result obtained from the tests shown in the previous chapter is definitely the good level of accuracy achieved with the Loc.3 approximation in the calculation of the absolute shielding of the nitrogen nucleus. The molecular systems considered were not the easiest ones to deal with. In fact, we have considered only two molecules with a single-bonded nitrogen atom (i.e.  $\text{NH}_3$  and  $\text{CH}_3\text{NH}_2$ ), while all the others involved multiple bonds and different degrees of electron delocalization. The former are considerably easier to handle and the accuracy attained by the four SOS-DFPT approximations is very similar. Meanwhile, in the latter case the inclusion of electron correlation becomes essential for the quality of the NMR shielding results.<sup>142,185</sup> Nonetheless, we have seen how the Loc.3 approximation in combination with the LDA functional is capable of evaluating the NMR shielding with a degree of accuracy comparable not only to other more sophisticated DFT based methods, but also to the high quality *ab initio*



references.

The molecules tested for the  $^{14,15}\text{N}$  shielding span a wide range of the nitrogen NMR spectrum, from +265 ppm for  $\text{NH}_3$  down to -125 ppm for  $\text{CH}_2\text{NN}$ . In this spectrum each region belongs to a very different type of nitrogen nucleus. Unfortunately, due to the limited availability of absolute shielding experimental references, we had to restrain our calculations to only eight different nitrogen nuclei \*. The data obtained are sufficient to give an idea of the quality of the results attainable for these nitrogen nuclei. However, due to a strong correlation between the nitrogen type and the difficulty of the shielding calculation, we cannot rightfully draw any general conclusion regarding the capabilities of Loc.3 across the  $^{14,15}\text{N}$  NMR spectrum.

The solution to this problem is to test the calculation results against experiments on the chemical shift scale instead of on the absolute scale. In fact, since the great majority of the experimental results are registered directly against a standard reference, virtually all the experimental NMR shielding studies are expressed in terms of the chemical shift. Moving to this unit of measure allows us to take advantage of a very substantial amount of experimental reference data. The only difficult aspect concerning the use of this collection of experimental references is that the majority of NMR shielding experiments are conducted in the liquid phase and at room temperature.

Meanwhile, the experimental condition which can be reproduced best by the calculations is the zero-pressure, zero-temperature limit. The challenge here is to find a way to reasonably account for solvent effects and temperature dependency.

Several methods have been developed in order to include the presence of the

---

\* $\text{NH}_3$  and  $\text{CH}_3\text{NH}_2$ , HNC, HCN and  $\text{CH}_3\text{CN}$ ,  $\text{N}_2$ ,  $\text{CH}_2\text{NN}$ ,  $\text{CH}_2\text{N}\underline{\text{N}}$ ,  $\underline{\text{NNO}}$ ,  $\underline{\text{NNO}}$

solvent in a DFT or *ab initio* calculation. These can be generally classified into two categories. If the method does not account explicitly for direct solute-solvent interactions, then it is defined as a *continuum* model.<sup>186,187</sup> On the other hand, if these interactions are considered but the bulk effect is disregarded, the method is defined as a *supermolecular* approach.<sup>188</sup> A few other methods, whose characterization lies between these two categories, have been also developed. Among these, the QM/MM approach has been particularly successful. This method allows one to treat the solute with a quantum mechanical (QM) approach, while the solvent molecules are described with molecular mechanics (MM).<sup>189,190</sup>

The introduction of the solvent effect through a continuum model is appropriate in case the solvent is aprotic and has a low dielectric constant (e.g. cyclohexane, n-alkanes and carbon tetrachloride). Here, only a correction of the order of a few ppm maximum is needed.<sup>191</sup> However, large changes in the NMR shielding are observed when there is a direct contact between the solute and the solvent, either through hydrogen bonds or through dipole-dipole interactions. A recent study on the <sup>14,15</sup>N NMR shielding of tetrazines and tetrazoles points out that the inclusion of the solvent through a continuum model is not sufficient to attain a high level of accuracy due to the lack of direct solute-solvent interactions.<sup>192</sup> On the other hand, the computational cost of a full supermolecular approach would have been very expensive considering the number of molecules which we had in mind to study. The best compromise is represented by a QM/MM treatment. However the software required is still under development in our group.

Under these premises, one of the possibilities left is represented by not including the solvent effect explicitly but to consider it as a constant within a particular group of molecules. In fact, according to the Kamlet-Taft solvatochromic scale

system<sup>191</sup> the NMR shielding of the nucleus N in the solvent X is given,

$$\sigma_N(X) = \sigma_N(\text{cyclohexane}) + a_N\alpha_X + b_N\beta_X + s_N(\pi_X^* + d_N\delta_X) \quad (7.1)$$

The first contribution is represented by the NMR shielding of the nucleus N in cyclohexane, while the other three terms involve four solvatochromic coefficients † (i.e.  $a_N$ ,  $b_N$ ,  $s_N$  and  $d_N$ ) and parameters (i.e.  $\alpha_X$ ,  $\beta_X$ ,  $\pi_X^*$  and  $\delta_X$ ). The  $\alpha_X$  parameter represents the hydrogen-donor strength of the solvent X,  $\beta_X$  is the hydrogen-bond acceptor strength of the solvent X,  $\pi^*$  takes into account the polarity/polarizability characteristics, while  $\delta$  represents a correction for “superpolarizability” of aromatic compounds. According to this method each type of molecule (solvent or solute) is characterized by different solvatochromic parameters. Alternatively, one can state that nuclei that belong to similar molecular systems will be effected in a similar way by the solvent. This conclusion has been proven in the case of  $^{14,15}\text{N}$  NMR shielding on the basis of experimental data. In fact, Witkowski and coworkers<sup>193</sup> were able to determine that the range of solvent effect on nitrogen shielding is very similar for similar nitrogen types.

In this chapter I will present the results obtained from the calculation of the  $^{14,15}\text{N}$  shielding of 132 nuclei. These results have been arranged and analyzed within 9 different classes which include, amines, hydrazines, amides, cyanides, isocyanides, azoles, azines, azine N-oxides and nitrates. The experimental references have been chosen mainly as a function of the nature of the solvent. The preferred references come from aprotic solvents with the lowest dielectric constant. Needless to say that this selection is limited by the availability of experimental data. This systematic approach allows to maintain the solvent effect contribution

†Determined via least square fit between different molecular properties of the solvent X.

constant, hence to assess clearly the limits and strengths of the different SOS-DFPT approaches, and in particular of Loc.3, for each type of nitrogen nucleus.

The calculation results will be presented on the same scale as the experimental references. The scale adopted, namely the chemical shielding scale, differs from the chemical shift scale only in the sign. In fact, in the chemical shift scale the direction of the “shielding” and “deshielding” of a particular signal is exactly the opposite as the one in the absolute scale. However, the definition of the chemical shielding scale,

$$\sigma_{\text{N}} = \sigma_{\text{N}}^{\text{Abs. (Calc.)}} - \sigma_{\text{N}}^{\text{Abs. (Std.)}} \quad (7.2)$$

allows us to maintain the same ordering of direction and sign used in the absolute scale.<sup>194</sup> The use of this convention is very common whenever the absolute shielding of the standard reference is negative. In <sup>14,15</sup>N NMR, for example, one of the most used standards is nitromethane. The most accurately known value of the absolute shielding of nitromethane is -135.8 ppm (liq. 298 to 303 K).<sup>183</sup> We chose to refer the calculation results to the latter for two main reasons. First and foremost, the use of a fixed standard reference, instead of a calculated one, allows one to avoid the “scaling” problems discussed in the previous chapter (see Sec. 6.1 at page 123). The second reason is related to the negligible temperature dependence of CH<sub>3</sub>NO<sub>2</sub> <sup>14,15</sup>N shielding (only 0.0045 ppm K<sup>-1</sup>).<sup>193</sup> As for the temperature dependence of the NMR shielding of the studied molecules †, accurate values of ro-vibrational corrections have been calculated only for a very few systems.<sup>195,196</sup> Additionally, for system such as N<sub>2</sub> for example, there is significant discrepancy between the results reported in the literature.<sup>195</sup> All the same,

†The temperature dependence is calculated by summation over rovibrationally averaged shieldings weighted by Boltzmann factors.

these corrections are considerably less important than the solvent effect, which we expect to be accountable for the largest fraction of the deviation from the experimental references.

The span of the  $^{14,15}\text{N}$  NMR spectrum analyzed in the present study is shown in Fig. 7.1 on page 146. We can see that each nitrogen atom type occupies a particular region of the spectrum. Only the results obtained with Loc.3 and UKS are shown here in order to avoid confusion in the interpretation. We can see from the comparison of these two approaches that for some systems the accuracy attained is very similar, while for others the Loc.3 behaves significantly better. In the following sections, I will describe in detail the performance of all four SOS-DFPT approximation for each nitrogen type<sup>§</sup>.

## 7.1 Amines

The amine-type nitrogen is an  $\text{sp}^3$  hybrid, single bonded to a carbon or a hydrogen atom. As it is indicated in Fig. 7.1, the  $^{14,15}\text{N}$  shielding of amines occupies mainly the up-field region of the nitrogen NMR spectrum, approximately between 300 and 420 ppm. This spectrum section is shown enlarged in Fig. 7.2 on page 148. Here the results for Loc.1 and Loc.2 are also indicated and the linear regression line associated to each SOS-DFPT correction replaces the bisecting line in Fig. 7.1. The performance shown by the UKS and by the four local corrections is very similar (see statistical analysis in Tab. 7.I on page 147). This indicates that the exchange-only and exchange-correlation corrections are very small and

<sup>§</sup>Part of the results shown in this chapter are published in the article: E.Fadda, M.E. Casida and D.R. Salahub, *J. Phys. Chem. A*, **107**, 9924 (2003).

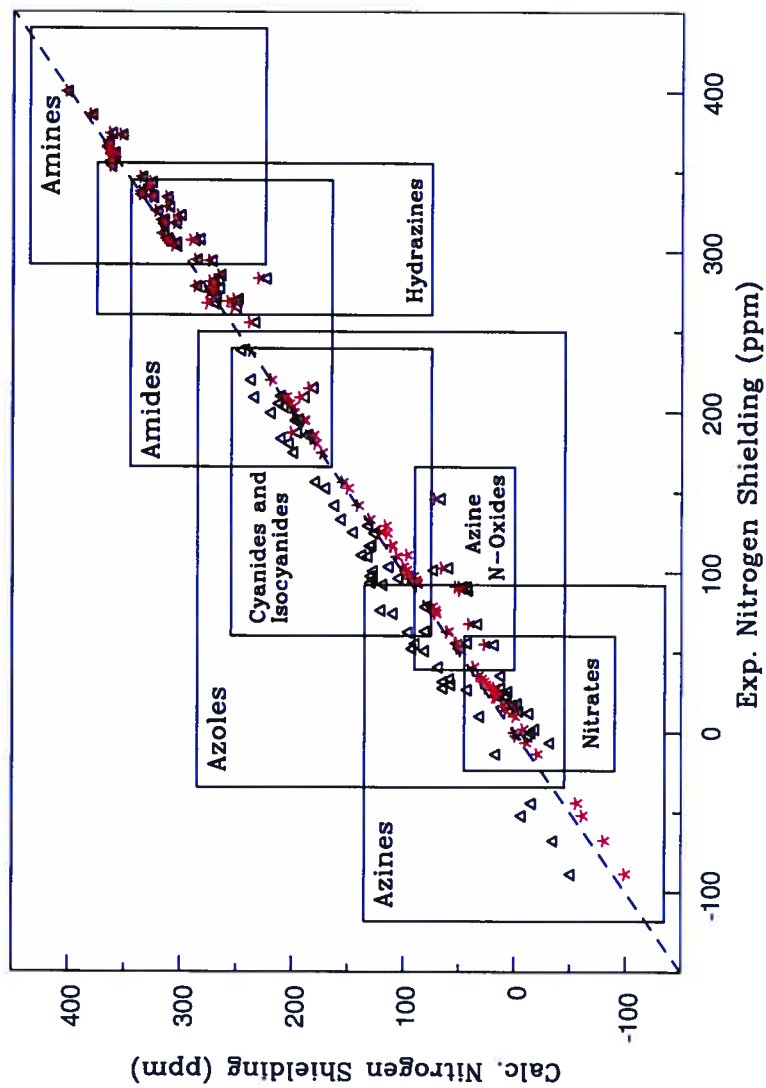


Figure 7.1: Portion of the  $^{14,15}\text{N}$  NMR shielding spectrum (vs.  $\text{CH}_3\text{NO}_2$ ) analyzed in this study. The Loc.3 results, indicated with red \*, and the UKS results, indicated with black  $\Delta$ , are shown for comparison. The sections associated with the different atom types are also indicated.

do not contribute significantly in the evaluation of the paramagnetic component. Nonetheless, the accuracy attained is sufficient to recognize some of the most important trends in the NMR shielding of amines. For example, the  $\beta$ -effect represents the downfield shift of the  $^{14,15}\text{N}$  shielding signal due to the substitution of the H atoms of the alkyl group in the  $\beta$ -position with respect to the nitrogen.<sup>194</sup> In order to assess if the calculation could reflect this phenomenon we studied three amines, shown in Fig. 7.3 on page 149, which have a different number of alkyl groups bonded to the  $\beta$ -carbon.

The results obtained with the Loc.3 approximation for *n*-butyl-amine, *s*-butyl-amine and *t*-butyl-amine are respectively 362.3, 336.1 and 315.8 ppm. The experimental results reveal the same deshielding trend with the values of 353.4 ppm for *n*-butyl-amine, 337.5 for *s*-butyl-amine and 317.3 for *t*-butyl-amine.

Table 7.I: Statistical analysis of the calculated  $^{15}\text{N}$  chemical shifts (ppm) in amines relative to neat liquid nitromethane

	Amines			
	<i>UKS</i>	<i>Loc.1</i>	<i>Loc.2</i>	<i>Loc.3</i>
Average. Abs. Dev.	5.9	5.4	5.3	5.4
Max. Deviation	21.2*	19.9*	19.5*	19.4*
Intercept	5.1	6.4	7.3	13.2
Slope	0.9737	0.9737	0.9723	0.9551
Correlation Coeff.	0.9498	0.9509	0.9518	0.9493

The accuracy of the calculation is higher for primary amines ( $\text{NH}_2\text{R}$ ), while it decreases with the substitution of two ( $\text{NHR}_2$ ) or of all three H atoms on the nitrogen nucleus ( $\text{NR}_3$ ). The average deviation determined for primary amines (2.9 ppm with Loc.3) is in fact four times larger for secondary amines (11.8 ppm

\* $\text{N}(\text{CH}_3)_3$

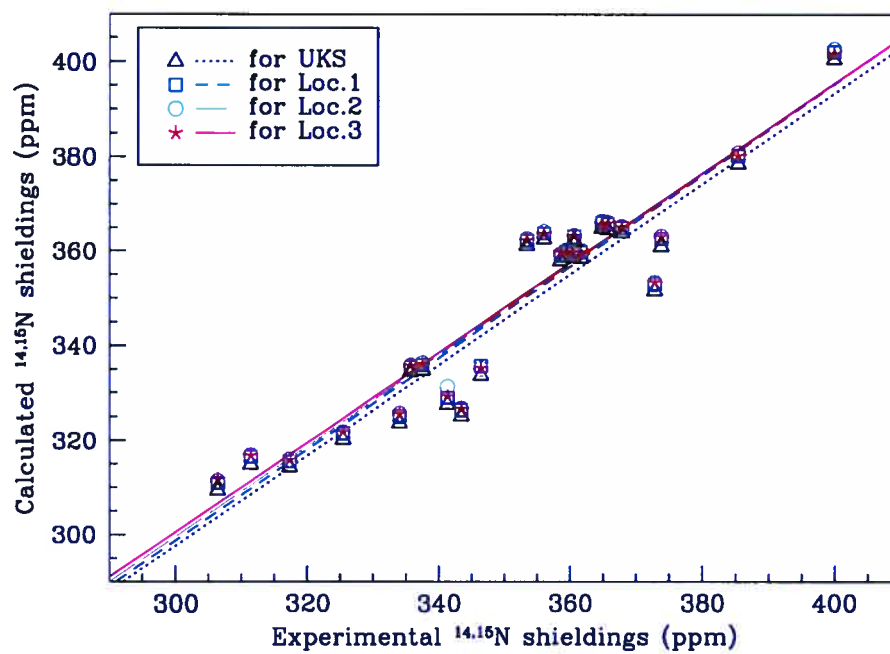


Figure 7.2:  $^{14,15}\text{N}$  NMR shielding in various amines. The different SOS-DFPT approximations are represented by the symbols listed on the legend. The linear regression line, relative to each SOS-DFPT approach, is also shown.



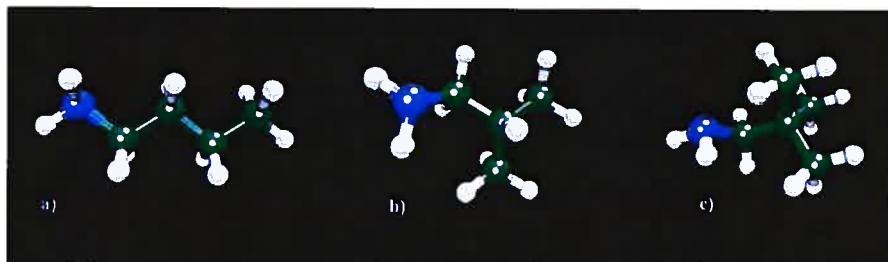


Figure 7.3: Three system studied for the evaluation of the  $\beta$ -effect. a) *n*-butylamine, b) *s*-butylamine and c) *t*-butylamine.

with Loc.3). Moreover, the maximum deviation for all four SOS-DFPT approximations corresponds to the  $^{14,15}\text{N}$  shielding of  $\text{N}(\text{CH}_3)_3$ . The results obtained for these and for the other amines studied are shown in Tab. A.XVI.

In conclusion, in the case of primary amines we are able to describe trends involving signal shifts of the order of 10 ppm minimum with any of the four SOS-DFPT approximations. We cannot rightfully interpret smaller induced shifts because of the solvent effect. In fact, we have to consider that the majority of the results are compared to experimental data taken in neat liquid, where the solute is also the solvent. In this matrix, shifts of less than 10 ppm can not be confidently assigned on the experimental level, since they may be solvent-induced.<sup>194</sup>

## 7.2 Hydrazines

The nitrogen nucleus in hydrazines is still an  $\text{sp}^3$  hybrid, therefore we expect the calculation results to be of the same quality as those seen previously for amines.

This is in fact the case; nonetheless, we can observe from the statistical analysis in Tab. 7.II a small decrease in accuracy, most probably due to the presence of the second nitrogen nucleus in the  $\alpha$ -position. In analogy with what we have seen for amines, the calculations approximate fairly well the experimental data when the H atoms are not replaced by alkyl groups. The average absolute deviation evaluated for  $-\text{NH}_2$  groups is in fact only 4.4 ppm with Loc.3, while for  $-\text{NHR}$  and  $-\text{NR}_2$  groups it is 19.5 and 19.9 ppm respectively. All the calculation results are shown in Tab. A.XVII.

Table 7.II: Statistical analysis of the calculated  $^{15}\text{N}$  chemical shifts (ppm) in hydrazines relative to neat liquid nitromethane

	Hydrazines			
	<i>UKS</i>	<i>Loc.1</i>	<i>Loc.2</i>	<i>Loc.3</i>
Average Abs. Dev.	15.3	13.6	13.2	12.9
Max. Deviation	25.6*	24.0*	23.5*	22.7*
Intercept	-36.7	-36.4	-36.8	-31.0
Slope	1.0693	1.0737	1.0769	1.0585
Correlation Coeff.	0.9182	0.9186	0.9184	0.9215

Another analogy with the previous case is that the four SOS-DFPT approximations show a very similar behavior. Here as well the exchange-only and exchange-correlation corrections are too small to contribute significantly.

### 7.3 Amides, Ureas and Guanidines

The polar character of the amido moiety is higher than in the two previous cases. Therefore the NMR shielding of the nitrogen nucleus, an  $\text{sp}^2$  hybrid, tends to be



more influenced by solvent effects. The role played by the local corrections is relatively more substantial than for amines and hydrazines, although it is not decisive to improve the quality of the results. As we can see from the statistical analysis in Tab. 7.III, the average absolute deviation attained by the local correction is only 1.0 ppm larger than UKS for Loc.1 and 1.9 ppm larger than UKS for Loc.3. Both of these values are considerably smaller than the solvent-induced shift for amides, which can reach roughly 8 ppm.<sup>193</sup> We can see from the data shown in Tab. A.XVIII that the calculations cannot reproduce trends involving shifts of the order of the estimated solvent-induced shift. For example, the effect of the substitution of the amido hydrogens with alkyl group in the calculation results is indecipherable [see the data corresponding to the series,  $\text{HCONH}_2 \rightarrow \text{HCONHCH}_3 \rightarrow \text{HCON}(\text{CH}_3)_2$ ].

Table 7.III: Statistical analysis of the calculated  $^{15}\text{N}$  chemical shifts (ppm) in amides, ureas and guanidines relative to neat liquid nitromethane

Amides, Ureas and Guanidines				
	<i>UKS</i>	<i>Loc.1</i>	<i>Loc.2</i>	<i>Loc.3</i>
Average Abs. Dev.	14.9	13.9	13.6	13.0
Max. Deviation	60.2*	56.4*	55.1*	53.1*
Intercept	-14.1	-9.0	-7.3	-3.2
Slope	1.0045	0.9938	0.9902	0.9793
Correlation Coeff.	0.9203	0.9198	0.9196	0.9210

Another example of undetected signal shifts is given by the substitution of the H atom bonded to the carbonyl carbon with an alkyl group, e.g. formamide to acetamide [ $\text{HCONH}_2 \rightarrow \text{}^3\text{HCCONH}_2$ ]. However, more significant signal shifts can be easily reproduced by the calculations. The substitution of the H atom bonded to

\* $(\text{CH}_3)_2\text{NC}(\text{NH})\text{OCH}_3$

the carbonyl carbon in formamide [H-CONH<sub>2</sub>] by a phenyl group [Ph-CONH<sub>2</sub>], or by an ethoxy group [CH<sub>3</sub>CH<sub>2</sub>O-CONH<sub>2</sub>], or even by an another -NH<sub>2</sub> group to form urea [H<sub>2</sub>N-CONH<sub>2</sub>], is qualitatively and quantitatively reproduced by all four SOS-DFPT approximations.

The most difficult cases are represented by cyclic amides, which structures are shown in Fig. B.1 on page 264, and by guanidines. Cyclic amides present the additional problem of being involved in a tautomeric equilibrium with their enol counterpart. Although the most stable structure is the amide tautomer, the enol's contribution to the NMR signal can be still significant. On the other hand, the difficulties encountered in the calculation of the <sup>14,15</sup>N shielding of guanidines are related to their very distinct basic character. The maximum deviation for all four SOS-DFPT approximations corresponds to the =NH group. However, when the H atom of =NH is replaced by a methyl group, this deviation is considerably reduced.

## 7.4 Cyanides and Isocyanides

The <sup>14,15</sup>N shielding calculation of various cyanides and isocyanides represents one of the most interesting cases. The type of nitrogen nucleus present in both of these systems is the same, a linear sp hybrid. Nonetheless, the levels of accuracy attained in the two cases are diametrically opposed. In order to appreciate the different degree of difficulty involved in the SOS-DFPT calculation of cyanides (R-CN) and isocyanides (R-NC), we can observe the distribution of the results against the experimental references in Fig. 7.4 on page 154. The left portion of the graph shows the results obtained for cyanides. A good agreement is observed

only for nitriles (i.e. alkyl cyanides), where the Loc.3 approximation improves significantly the accuracy of the calculations (see also the statistical analysis in Tab. 7.IV and the results in Tab. A.XIX). The solvent effect in the latter case does not play a significant part since the experimental references have been determined in cyclohexane. Hence, this is one of the cases where we can really see the different potentials of the four SOS-DFPT approximations.

Table 7.IV: Statistical analysis of the calculated  $^{15}\text{N}$  chemical shifts (ppm) in cyanides relative to neat liquid nitromethane

Cyanides				
	<i>UKS</i>	<i>Loc.1</i>	<i>Loc.2</i>	<i>Loc.3</i>
Average Abs. Dev.	41.4	33.2	30.5	22.5
Max. Deviation	55.8*	50.1*	48.2*	43.5*
Intercept	-92.3	-95.7	-96.5	-157.1
Slope	1.3973	1.4882	1.5148	2.0503
Correlation Coeff.	0.4890	0.4753	0.4712	0.5638

The picture changes dramatically when we consider the group of aryl cyanides. Unfortunately in this case even the solvent, or phase, in which the experiments have been conducted, changes. Therefore, it is difficult to establish the origin of the discrepancy between the calculation and the reference. The  $^{14,15}\text{N}$  shielding of cyanides has been determined to be very sensitive to interactions with the solvent. Not considering the influence of covalent or H-bond interactions, which involvement cause shifts of the order of 100 ppm,<sup>193</sup> a strong dipole-dipole interaction can induce shifts of up to 26 ppm.<sup>194,197</sup> This sensitivity is undoubtedly enhanced in aryl cyanides where the electrons are highly delocalized, hence more easily polarizable. However, the solvent effect can account only for part of the deviation

\**p*-CN-Ph-CN

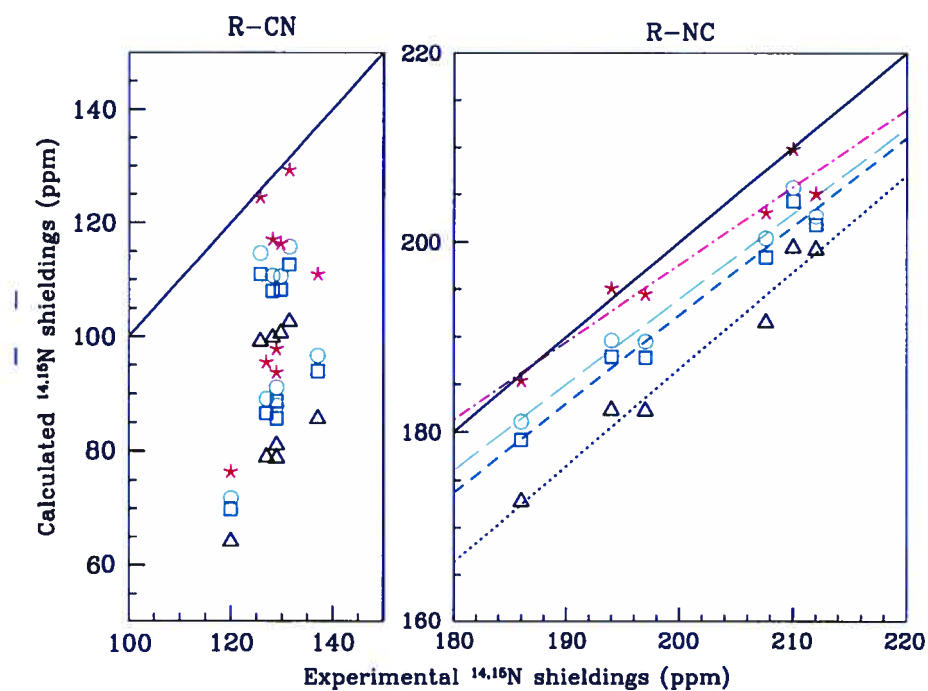


Figure 7.4: Distribution of the calculated results on cyanides (left part) and isocyanides (right part) against experimental references. The symbols,  $\Delta$  (black),  $\square$  (blue),  $\circ$  (green) and  $\star$  (red), represent the results for UKS, Loc.1, Loc.2 and Loc.3 respectively. The black continuous line on both parts represents the bisecting line. The linear regression lines are shown only on the right part.

evaluated for aryl cyanides. The remainder must be due to a fundamental difficulty of the SOS-DFPT approach, defined within the LDA functional, in dealing with such highly delocalized systems. The local character of the  $V_{xc}$  hinders the correct description of the charge distribution along the structure, with an obvious failure in the evaluation of the NMR shielding of the nuclei composing the systems. As a proof of this argument, we obtain a considerable improvement with the use of GGA PD91-PD91, instead of LDA. The corresponding calculation results are on average 10 ppm more shielded (88.3 ppm for UKS, 96.1 ppm for Loc.1, 98.5 ppm for Loc.2 and 104.8 ppm for Loc.3). However, a more sophisticated functional is required to reach a better level of accuracy.

Table 7.V: Statistical analysis of the calculated  $^{15}\text{N}$  chemical shifts (ppm) in isocyanides relative to neat liquid nitromethane

<b>Isocyanides</b>				
	<i>UKS</i>	<i>Loc.1</i>	<i>Loc.2</i>	<i>Loc.3</i>
Average Abs. Dev.	13.2	7.9	6.3	2.6
Max. Deviation	16.1*	10.2 <sup>†</sup>	9.3 <sup>†</sup>	6.9 <sup>†</sup>
Intercept	-15.5	6.4	12.7	33.3
Slope	1.0113	0.9290	0.9058	0.8231
Correlation Coeff.	0.9822	0.9839	0.9822	0.9633

In the case of isocyanides the agreement with the experimental references is definitely improved. In the right portion of the graph in Fig. 7.4 the results are shown together with the regression lines for each of the four SOS-DFPT approximations. As we have seen in the case of cyanides, Loc.3 enhances the level of accuracy, with the essential difference that here the improvement comes together

\* $\text{CF}_3\text{NC}$

<sup>†</sup> $\text{Ph-CH}_2\text{NC}$

with a better correlation. This significant improvement obtained with Loc.3 confirms the results obtained earlier in the absolute shielding scale. The statistical analysis in Tab. 7.V clearly shows that best value of the average absolute deviation is obtained with Loc.3. At the same time, it shows also that the worse results in the linear regression analysis are obtained with Loc.3. From a close examination of the results, shown in Tab. A.XX, we can see that the Loc.3 approximation does consistently better than all the other local corrections. Therefore, the results of the linear regression analysis can be attributed simply to the shortage of data and to the narrowness of the spectral region under study.

## 7.5 Azoles

The behavior of the SOS-DFPT approximations, and in particular of the Loc.3 correction, in relation to the azole systems is of major importance to this work. In fact, the imidazole ring, which is part of the side chain of the histidine residue, plays a crucial role in the catalytic mechanism of the serine proteases. The  $^{14,15}\text{N}$  shielding of its two nitrogen nuclei will be used as a probe to study the catalytic mechanism governing this class of enzymes. Hence, it is most important to delineate the boundaries of the information which can be derived from the calculations. Only the knowledge of these limits can allow us to draw relevant and reliable conclusions.

First and foremost it is important to underline that two different types of nitrogen nuclei can be part of an azole ring, namely the pyrrole-type and the pyridine-type. They are both  $\text{sp}^2$  hybrids, however, in the pyrrole-type nitrogen two electrons take part in the delocalization, while in the pyridine-type only one electron



is shared. This structural difference is reflected also in the NMR characteristics of these two nitrogen nuclei. A typical example is given by the strength and direction of the solvent-induced shifts. The effect of the solvent polarity results in a deshielding in pyrrole-type nitrogens, while a comparable shift, but opposite in sign, is observed for pyridine-type nitrogens. The dipole-dipole interaction effect on the NMR shielding of azoles, estimated with the Kamlet-Taft method [Eq.(7.1)], is in the range of -6 ppm for pyrrole-type and up to +8 ppm for the pyridine-type nitrogen.<sup>198</sup> More significant is the effect of H-bonding, which is particularly interesting from the enzyme catalysis point of view. If we exclude protonation, a H-bond involving directly the pyrrole-type nitrogen as a donor can cause a deshielding in the order of -10 ppm.<sup>199</sup> Meanwhile, the participation of a pyridine-type nitrogen as an H-bond acceptor results in a shielding between +15 ppm and +3 ppm depending on the position of the N nucleus in the azole ring.<sup>198</sup> Keeping in mind the impact of the solvent effect, we move on to the analysis of the results.

The distribution of the calculated  $^{14,15}\text{N}$  shielding against the experimental references is shown in Fig. 7.5. From this graph we can clearly see that the SOS-DFPT methodology evaluates with much higher accuracy the pyrrole-type nuclei than the pyridine-type. The structures of all the azoles examined in this section are shown in Fig. B.2 on page 265.

The statistical analysis for the pyrrole-type nuclei, shown in Tab. 7.VI, indicates a fairly large average absolute deviation of the calculated results. Among all the systems seen so far, a larger deviation have been evaluated only for arylcyanides. However, the good correlation obtained in this case assures a reliable interpretation of the experimental data. Part of this large deviation is solvent induced

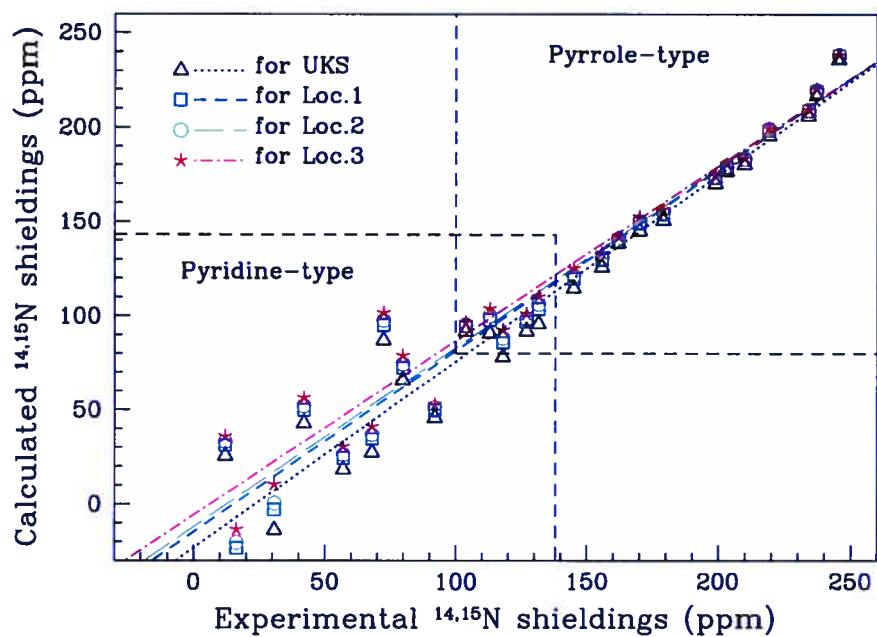


Figure 7.5: Distribution of the calculated  $^{14,15}\text{N}$  shieldings in azoles against the experimental references. The linear regression lines have been derived from the analysis of both pyrrole- and pyridine-type  $^{14,15}\text{N}$  shieldings.

and in particular related to the effect of the dipole-dipole interactions aforementioned. In fact, the majority of the experimental references used (see Tab. A.XXI) have been determined in DMSO, a highly polar solvent with a polarity index value, relative to H<sub>2</sub>O, of 0.444.<sup>200</sup>

Table 7.VI: Statistical analysis of the calculated pyrrole-type <sup>15</sup>N chemical shifts (ppm) in azoles relative to neat liquid nitromethane

Azoles (pyrrole-type N)				
	<i>UKS</i>	<i>Loc.1</i>	<i>Loc.2</i>	<i>Loc.3</i>
Average Abs. Dev.	22.0	19.4	18.5	16.8
Max. Deviation	30.2*	26.4 <sup>†</sup>	25.2 <sup>†</sup>	21.4 <sup>‡</sup>
Intercept	-28.0	-23.0	-21.5	-16.2
Slope	1.0311	1.0195	1.0157	0.9929
Correlation Coeff.	0.9913	0.9923	0.9926	0.9936

The choice of functional has also influences the results. The azoles are aromatic systems with 6  $\pi$  electrons delocalized on 5 centers. One characteristic of the NMR of aromatic compounds is the deshielding induced by the formation of a strong ring current in the presence of an external magnetic field. In particular, the aromatic character, proportional to the strength of the ring current,<sup>202</sup> has been found to be larger for pyrrole than for benzene, furan and thiophene.<sup>203</sup> Not having considered any current-dependent functionals, we cannot describe correctly such a type of system. Nonetheless, we can verify if and to what extent a non-local functional is more apt to give a better description. We tested the PD91-PD91 exchange-correlation functional in the case of pyrrole. The latter represent a good trial system in that its experimental reference has been determined in cyclohex-

\*1-Me-1,2,3-triazole

<sup>†</sup>1-NH<sub>2</sub>-1,2,3-triazole

<sup>‡</sup>Me-pyrrole

ane, hence the solvent effect is reduced to the minimum. The results indicate an improvement of the performance of all the SOS-DFPT approximations of only a +3 ppm on average. The effect of the ring current is evidently rather intense and out of our grasp. In order to achieve a good level of accuracy for this type of molecules the only alternative consist in multiconfigurational methods (MCSCF) as it has been demonstrated by the calculations of Jaszúński and coworkers.<sup>204</sup>

Pertaining to the different level of accuracy shown by the four SOS-DFPT approximations, the addition of any of the local corrections greatly improves the results. In the particular case of the pyrrole-type nitrogen nucleus, a noticeable distinction among the three local corrections can be only seen whenever there are two or more nitrogen atoms within the ring. The Loc.3 approximation performs particularly well in the case of adjacent nitrogen nuclei (i.e. in pyrazoles or triazoles). It has been known that these nuclei are difficult to evaluate. In fact, the correct approximation of their shielding values requires the introduction of correlation effects.<sup>201</sup> The Loc.3 correction is the only local correction in SOS-DFPT which includes a correlation contribution [see Eq.(3.70) in Sec. 3.3.2].

The calculation of the  $^{14,15}\text{N}$  shieldings of pyridine-type atoms suffers from the aforementioned problems (i.e. solvent effects and ring current effect), but in a much more unpredictable way. The linear regression analysis carried out for pyridine-type nuclei, showed in Tab. 7.VII, reports a very inconsistent interpretation of the experimental results. This can be clearly seen also in Fig. 7.5, where a scattered pattern characterizes the region of the spectrum occupied by the pyridine-type  $^{14,15}\text{N}$  shieldings. The average absolute deviation is also larger than in the case of the pyrrole-type nitrogen. The Loc.3 correction provides for a small improvement of the overall performance. As for the pyrrole-type nitrogen,

its major contribution is given in the calculation of the  $^{14,15}\text{N}$  shielding of adjacent nuclei within the ring.

Table 7.VII: Statistical analysis of the calculated pyridine-type  $^{15}\text{N}$  chemical shifts (ppm) in azoles relative to neat liquid nitromethane

Azoles (pyridine-type N)				
	<i>UKS</i>	<i>Loc.1</i>	<i>Loc.2</i>	<i>Loc.3</i>
Average Abs. Dev.	29.2	25.2	24.5	22.5
Max. Deviation.	49.0*	42.5 <sup>†</sup>	41.3 <sup>†</sup>	39.6 <sup>†</sup>
Intercept	-16.6	-8.9	-6.4	-0.3
Slope	0.8796	0.8657	0.8611	0.8313
Correlation Coeff.	0.8674	0.8712	0.8719	0.8776

## 7.6 Azines

The azines are aromatic systems characterized by the presence of one or more nitrogen atoms fused in the ring. The only type of nitrogen atom present in azines is the pyridine-type (see Fig. B.3 on page 266). Therefore, we would expect to find the calculation results to share common features with those for the same type of nucleus in azoles. However, from the statistical analysis shown in Tab. 7.VIII we notice some remarkable differences. The average absolute deviation, for example, is slightly larger and the same can be said for the maximum deviation. Moreover, the correlation coefficient is much higher, reflecting a more consistent interpretation of the experimental data. The distribution of the calculated results against the experimental reference is shown in Fig. 7.6 on page 163.

\* 1-Me-1,2,3-triazole

<sup>†</sup> 2-Me-benzopyrazole

Table 7.VIII: Statistical analysis of the calculated  $^{15}\text{N}$  chemical shifts (ppm) in azines relative to neat liquid nitromethane

	Azines			
	<i>UKS</i>	<i>Loc.1</i>	<i>Loc.2</i>	<i>Loc.3</i>
Average Abs. Dev.	52.8	39.9	36.6	30.6
Max. Deviation	74.9*	59.4*	54.6*	45.4*
Intercept	-60.4	-45.9	-41.4	-32.5
Slope	1.2239	1.1363	1.1103	1.0427
Correlation Coeff.	0.9921	0.9918	0.9915	0.9900

The reason behind such differences in the calculations performance could be twofold. The first argument pertains to the solvent effect. In the present case the experimental references available are all referred to apolar solvents, i.e. cyclohexane and  $\text{CCl}_4$ , with the exception of three cases (see results in Tab. A.XXII). Therefore, here the effect of the dipole-dipole interactions is not relevant. Within the approximation that the impact of the ring current on the shielding value is of the same order of magnitude as for azoles, the average deviation evaluated for azines is understandably larger. In fact, the direction of the solvent induced shift (i.e. shielding) is opposite to the paramagnetic contribution introduced by the ring current (i.e. deshielding). Therefore, it turns out that the cancellation of errors seen in the case of the pyridine-type atoms in azoles, cannot occur here.

As it has been mentioned before, the most striking discrepancy emerging from the comparison of the results obtained for the pyridine-type nitrogen here in azines and in the azole group is the different degree of correlation between calculations and experimental references. The cause of the erratic behavior of the SOS-DFPT results shown for azoles can be only attributed to an irregular influence of the

---

\*1,2,4-triazine

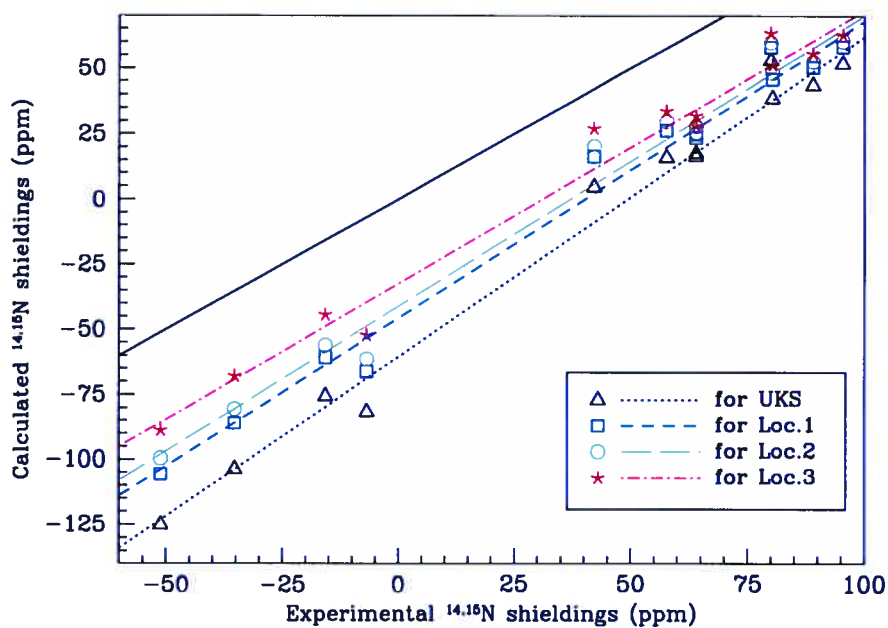


Figure 7.6: Distribution of the calculated  $^{14,15}\text{N}$  shieldings in azines against the experimental references. The black continuous line indicates the bisecting line.

solvent effect. As it has been predicted by the Kamlet-Taft system, the solvent-induced shift does not depend only on the solvent, which determines the value of  $\pi^*$ , but also on the position of the nitrogen atom within the ring, through the value of the coefficient  $s$ . The latter can attain values between +8 and +2 for pyridine-type nitrogen atoms within the same ring.<sup>198</sup> The impact of the dipole-dipole interactions on the  $^{14,15}\text{N}$  shielding of azines is of the same order of magnitude, even a little more marked in 1,2-diazines.<sup>205</sup> However, it does not influence the results because almost all the experimental references used here have been evaluated in apolar solvents.

The minimization or the complete elimination of solvent effects is essential in order to evaluate reliably the level of accuracy attainable by the calculations in this type of nitrogen nuclei. Besides the asymmetrical influence of the dipole-dipole interactions discussed above, the most significant solvent-induced shift is brought in by the formation of H-bonds between solute and solvent. If we exclude protonation, the impact of an H-bonding solvent on the  $^{14,15}\text{N}$  shielding of azines can reach +30ppm.<sup>205,206</sup> In an attempt to assess the ability of SOS-DFPT in evaluating the dependence of the  $^{14,15}\text{N}$  shielding upon the extent of the H-bond network, I have carried out several tests on a pyridine molecule surrounded by a different number of water molecules. The aim of the calculation was to see if the SOS-DFPT results tend to the experimental value of the  $^{14,15}\text{N}$  chemical shift in water at infinite dilution. The tests shown here involve systems containing between 1 and 27 water molecules. The clusters have been built in two steps. Initially a full geometry optimization has been performed with molecular mechanics in order to bring the system close to a local minimum. The MM3(1999)<sup>207-214</sup> force field as implemented in version 3.8 of the TINKER<sup>©215</sup> software package has been used



to carry out this part of the calculation. The final structure of the clusters has been obtained after a few cycles of DFT full geometry optimization (see Ch. 5 for details). This step is crucial in order to further relax the strained interactions. For the smallest clusters the following SCF and SOS-DFPT calculations could be easily performed. However, for the largest ones, the full supermolecular approach was too computationally expensive due to the large basis set needed. In fact, although we applied a locally-dense basis approach, the IGLO-III basis set had to be used for all atoms directly connected to the pyridine nitrogen and for the ones adjacent to the atoms directly connected to the pyridine nitrogen. This scheme gives raise to a basis set superposition error (BSSE) of only 2 ppm on average, well within the value of the deviations seen as yet. A further reduction of the IGLO-III atoms results in much larger BSSE, which indicates an inadequate basis set scheme.<sup>216</sup> Therefore, in order to get around the problem, for clusters containing more than 41 atoms, the water molecules farthest from the pyridine nitrogen have been substituted by point charges calculated with the Singh-Kollmann method.<sup>217</sup> The results are shown in Fig. 7.7.

The calculated data have been analyzed through a least-squares logarithmic fitting. For each SOS-DFPT approximation we obtained a function of the form,

$$y = b \ln x + a \quad (7.3)$$

where the coefficients can be found from least-square fitting<sup>¶</sup>. The resulting equations are shown also in Fig. 7.7. Unfortunately, the Loc.3 approximation can not

$$b = \frac{n \sum_{i=1}^n (y_i \ln x_i) - \sum_{i=1}^n y_i \sum_{i=1}^n \ln x_i}{n \sum_{i=1}^n (\ln x_i)^2 - (\sum_{i=1}^n \ln x_i)^2} \quad a = \frac{\sum_{i=1}^n y_i - b \sum_{i=1}^n (\ln x_i)}{n}$$

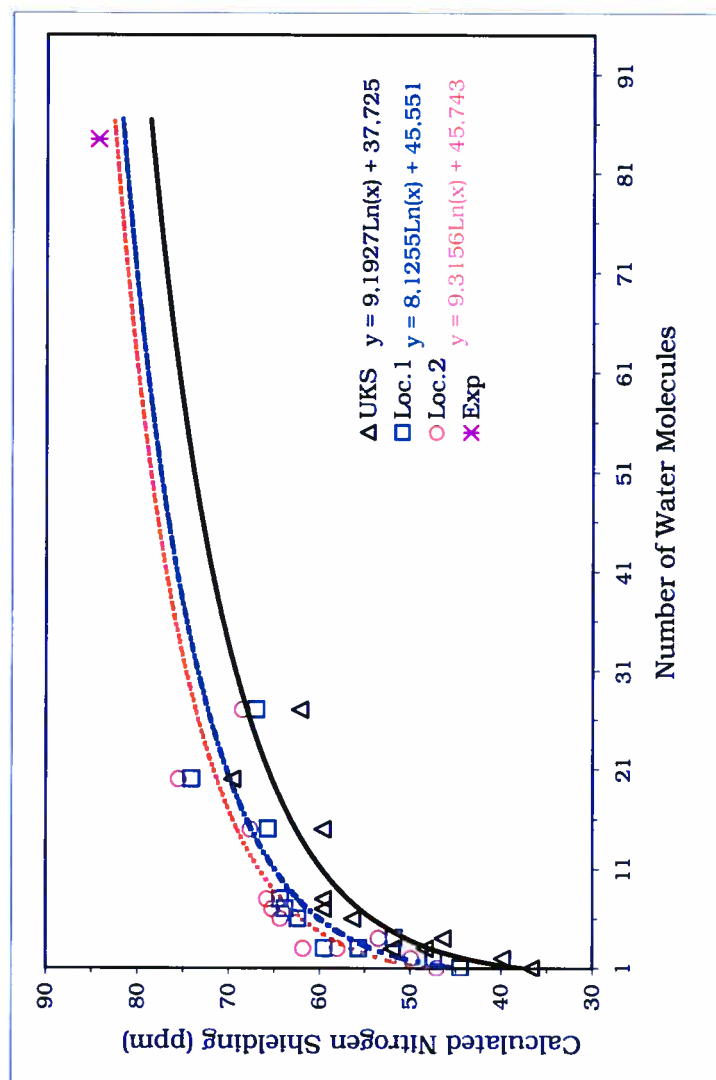


Figure 7.7: Calculated  $^{14,15}\text{N}$  shieldings of a pyridine molecule in different water clusters. The  $^{14,15}\text{N}$  shieldings are expressed as a function of the number of water molecules in the cluster.

be tested for these systems, since the time-dependent calculation of clusters consisting of more than 5 water molecules is too computationally expensive. However, from the previous experience obtained with other azines, we do expect Loc.3 to behave in a similar and possibly more accurate fashion than the other local corrections. The results obtained are very encouraging. The SOS-DFPT method is capable of accurately predicting the impact on the  $^{14,15}\text{N}$  shielding of an increasingly complex H-bond network. Not surprisingly, the highest accuracy is predicted with the Loc.1 and Loc.2 approximations. The most important implication of these results is that in order to evaluate correctly  $^{14,15}\text{N}$  the shielding of a molecule in a highly polar protic solvent we need to consider very large solvent clusters within a supermolecular approach. Similar conclusions had been drawn for the  $^1\text{H}$  and  $^{17}\text{O}$  shieldings from tests on smaller water clusters.<sup>218</sup>

## 7.7 Azine N-Oxides

This particular group of molecules has been studied in order to evaluate the response of the  $^{14,15}\text{N}$  shielding in azines to oxidation. Experimentally, in fact, the  $^{14,15}\text{N}$  shielding of azines and the  $^{14,15}\text{N}$  shielding of azine N-oxides are highly correlated. The only difference is that the formation of the N-O bond induces a significant shielding effect.<sup>193</sup> The results and the statistical analysis are shown in Tab. A.XXIII and Tab. 7.IX respectively.

Table 7.IX: Statistical analysis of the calculated  $^{15}\text{N}$  chemical shifts (ppm) in azine N-oxides relative to neat liquid nitromethane

Azine N-oxides				
	<i>UKS</i>	<i>Loc.1</i>	<i>Loc.2</i>	<i>Loc.3</i>
Average Abs. Dev.	46.5	42.6	41.1	39.4
Max. Deviation	80.1*	77.2*	74.3*	74.4*
Intercept	-1.9	3.9	4.1	8.2
Slope	0.5052	0.4835	0.4982	0.4714
Correlation Coeff.	0.9295	0.9268	0.9374	0.9311

The range of uncertainty is of similar weight as in the case of azines, with the exception of the very large deviation found for 2-OMe-pyridine N-oxide, which also represents the maximum deviation. This difficult case is largely responsible for the poor results obtained in the linear regression analysis. It is clear from the graph in Fig. 7.8 on page 169 that, apart from this value, all the others run along lines more parallel to the bisecting line.

The entity of the shielding induced by the oxidation is well reproduced by the calculations. For example, the oxidation of pyridine induces a shielding of +18.7 ppm (calculated as the difference of the experimental  $^{14,15}\text{N}$  shieldings of pyridine and pyridine N-oxide in cyclohexane). The corresponding calculated value is +15 ppm with the Loc.3 correction. The correlation between the data calculated for azines and for azine N-oxides is rather good. The correlation coefficients obtained for UKS, Loc.1, Loc.2 and Loc.3 are respectively 0.9047, 0.9082, 0.9023 and 0.9016. The level of correlation between the corresponding experimental data, shown in Tab. A.XXII and Tab. A.XXIII, is only 0.8381. The reason for this is most certainly the difference between the solvents in which the experiments have

\*2-OMe-pyridine N-oxide

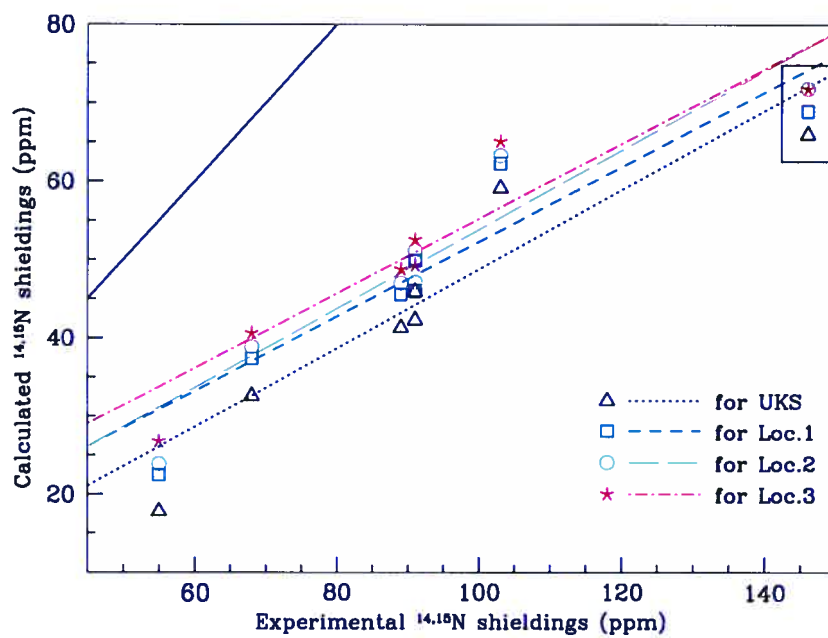


Figure 7.8: Distribution of the calculated <sup>14,15</sup>N shieldings in various azine N-oxides against the experimental references. The points corresponding to the 2-Ome-pyridine N-oxide are indicated within a box in the top right corner of the graph. The continuous black line corresponds to the bisecting line

been conducted.

## 7.8 Nitrates

The nitrate group exerts a strong electron withdrawing effect on the molecule to which it is bonded. The concentration of the electron density around the nitrogen atom, makes its  $^{14,15}\text{N}$  shielding very sensitive to the solvent's polarity.<sup>193</sup> Unfortunately we were not able to find experimental references carried out in solvents with a low dielectric constant. Hence, in analyzing the results we should take into account that the range of solvent induced shifts can oscillate between 7 and 9 ppm for solvents such as DMSO, DMF and  $\text{Me}_2\text{CO}$ .<sup>194</sup> The results and their statistical analysis are shown in Tab. A.XXIV and Tab. 7.X respectively.

Table 7.X: Statistical analysis of the calculated  $^{15}\text{N}$  chemical shifts (ppm) in nitrates relative to neat liquid nitromethane

Nitrates				
	<i>UKS</i>	<i>Loc.1</i>	<i>Loc.2</i>	<i>Loc.3</i>
Average Abs. Dev.	18.3	12.5	10.7	7.3
Max. Deviation	26.0*	19.7 <sup>†</sup>	17.8 <sup>†</sup>	14.8 <sup>†</sup>
Intercept	-19.1	-11.6	-9.2	-4.2
Slope	1.0544	0.9269	0.8877	0.7716
Correlation Coeff.	0.9392	0.9221	0.9139	0.8977

The nitrogen atom in nitrates represents the most clear example of how the accuracy of the calculations is improved by the addition of the local corrections. Among those, the *Loc.3* gives the best performance, as it can be seen also from Fig. 7.9. Nonetheless, the linear regression analysis results apparently states oth-

\* $\text{CH}_2\text{CHCH}_2\text{NO}_2$

<sup>†</sup> $\text{PhNO}_2$

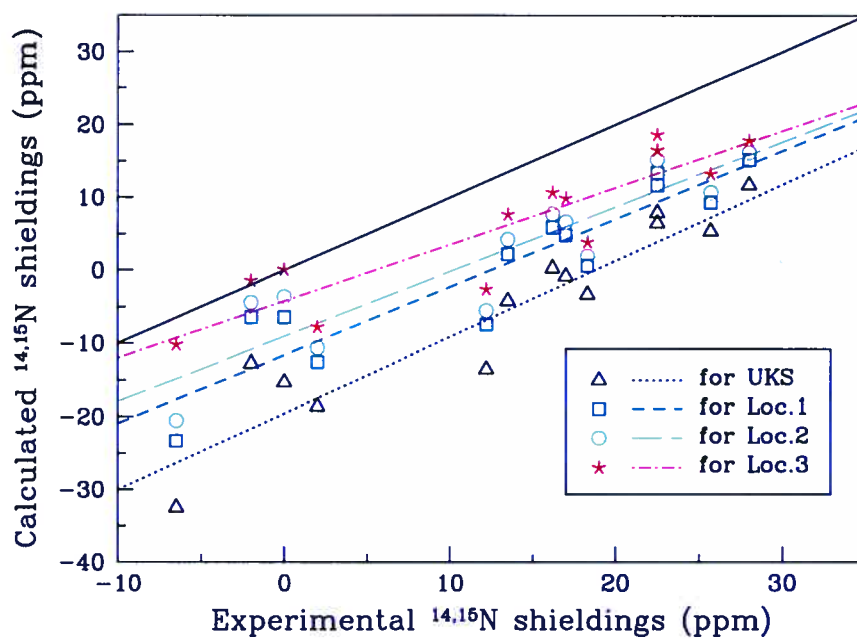


Figure 7.9: Distribution of the calculated  $^{14,15}\text{N}$  shieldings in nitrates against experimental references. The black continuous line indicates the bisecting line.

erwise. The correlation coefficient and the slope values obtained for Loc.3 are in fact very low. The reason for this is that these parameters have to take into account also the results obtained for the three most difficult cases, which are nitrobenzene ( $\text{PhNO}_2$ ), 2,4-dinitro-pyrrole and 4,5-dinitro-imidazole. These three systems are the only cases where the absolute deviation, calculated for Loc.3, exceeds 10 ppm. Furthermore, the maximum deviation for all the local approximation corresponds to nitrobenzene. The correlation coefficient and the slope values are considerably different if we do not consider them in the analysis, while the intercept remains of the same order of magnitude. Specifically, the correlation coefficient, the slope and the intercepts values for Loc.3 become 0.9551, 0.8803 and -3.4 ppm respectively. A similar improvement is also registered for Loc.2, while for Loc.1 and UKS the results are almost unaffected.



## Chapter 8

# The Catalytic Mechanism of Serine Proteases

The serine proteases constitute a family\* of proteolytic enzymes characterized by a nucleophilic serine (Ser) residue at the active site. This class of peptidases is present in all living organisms and is responsible for a wide variety of physiological functions, such as protein digestion, blood coagulation and immune response.<sup>220</sup> Four clans can be identified within the serine protease family: chymotrypsin, subtilisin, carboxypeptidase Y, and Clp protease.<sup>219</sup> These enzymes all share a characteristic structural feature defined as the “catalytic triad”,<sup>221</sup> which consists of three residues: aspartate (Asp), histidine (His), and serine (Ser) interconnected by hydrogen bonds (H-bonds). The location of these residues in the acyl enzyme structure of a pancreatic elastase (chymotrypsin family), together with a schematic representation of their H-bond network, is shown in Fig. 8.1 on

---

\*The “family” denomination is assigned on the basis of statistically significant similarities in amino acid sequence.

page 175. Recently, a fourth residue, a Ser in the chymotrypsin family, has been recognized as a functional part of the catalytic triad structure. The carbonyl group of its peptide chain forms an H-bond with the C<sub>ε1</sub>-H bond of the catalytic His.<sup>5,222</sup> The recurrence of this element among all the serine proteases indicates its possible mechanistic relevance.

The catalytic function of serine proteases has been the subject of extensive study for over fifty years. Many experimental and theoretical methods have been used to reveal all the details within this process. The most significant advancements in the identification of the reaction steps have been obtained through X-Ray crystallography<sup>223–228</sup> and NMR.<sup>229,230</sup> However, although the general aspects of the reaction path are now well understood and documented, many questions about the formation of the intermediate structures remain unanswered. This chapter provides a detailed description of these problems and their analysis with DFT and SOS-DFPT calculations.

The comparison between the different SOS-DFPT approximations, in the matter of the <sup>14,15</sup>N NMR shielding, has shown us the strengths and the limits of each of them. One of the conclusions derived is that while for saturated nitrogen atoms there is no substantial difference between the SOS-DFPT approximations, for unsaturated systems the local corrections perform significantly better than UKS. Among those, Loc.3 showed the highest accuracy. The nitrogen nucleus of interest here is a pyrrole-type nitrogen in an imidazole ring (the His side-chain is in fact a 4-substituted imidazole). In Sec. 7.5 we were able to assess that Loc.3 gave an average absolute deviation for the pyrrole-type nitrogen of ~17 ppm. Further, the correlation coefficient demonstrated a good interpretation of the experimental data. Under these premises, we started an investigation of the various alternative

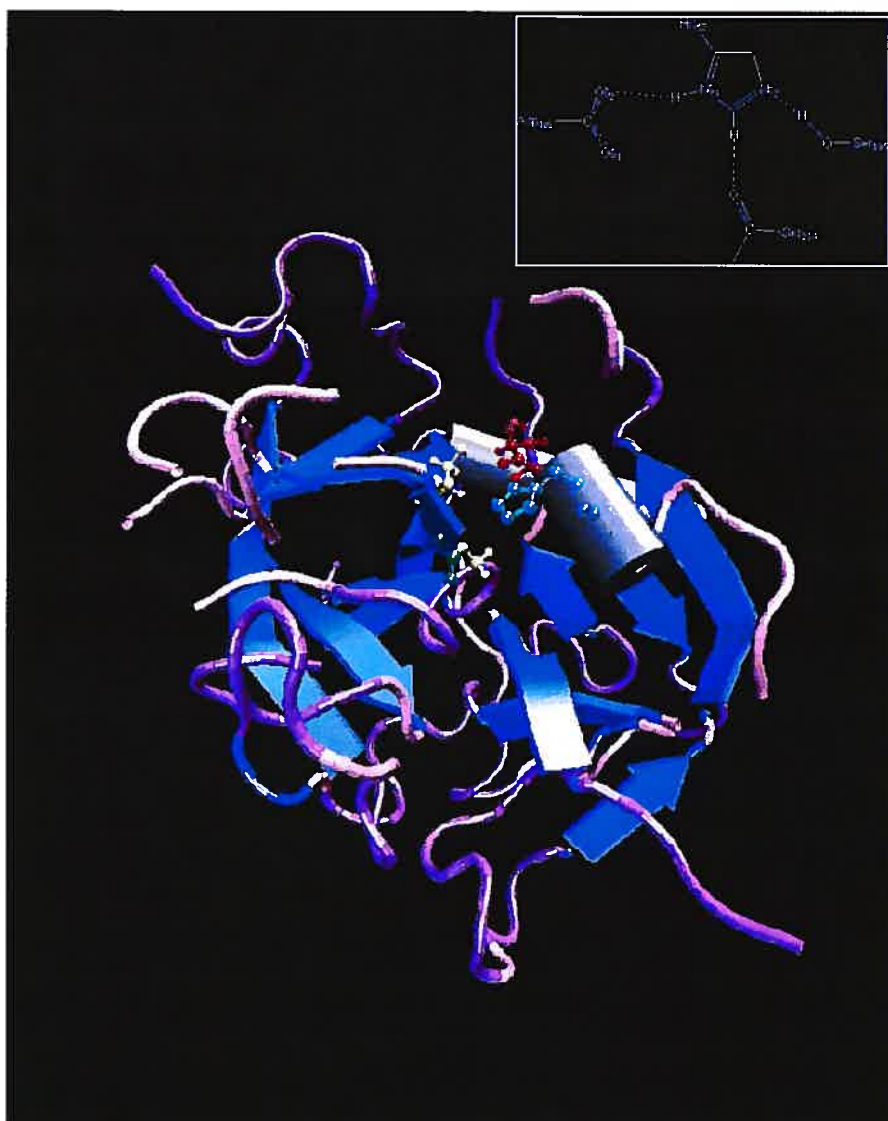


Figure 8.1: Porcine pancreatic elastase acyl enzyme (PDB code 1GVK). The catalytic triad is represented schematically on the top-right corner of the picture. Its position within the enzyme is also indicated (His in cyan, Asp in red and Ser in yellow).

catalytic mechanisms proposed in the literature for the serine proteases through the comparison between calculated and experimental  $^{14,15}\text{N}$  NMR shieldings.

## 8.1 The Active Site

The active site is the part of the enzyme containing the ensemble of peptide residues recruited to carry out the reaction. The active site of serine proteases can be divided into two sub-regions. The first is occupied by the aforementioned catalytic triad, otherwise known as the “charge-relay” system,<sup>221</sup> and the other is defined as the “oxyanion hole”. While each of these two parts has a specific duty, their functions are deeply intertwined. In the enzymes belonging to the chymotrypsin family, the residues constituting the catalytic triad are numbered: Asp<sub>102</sub>, His<sub>57</sub> and Ser<sub>195</sub>. In other clans the residue numbers are different, although their nature remains the same. An image of the catalytic triad in a chymotrypsin-like protease (PDB code 1GVK) is shown in Fig. 8.2<sup>†</sup>. Besides the “classic” members of the charge-relay system, two other components are shown in the figure, namely Ser<sub>214</sub>, and a water molecule (H<sub>2</sub>O<sub>305</sub>). We mentioned previously that the catalytic role of Ser<sub>214</sub> is related to the H-bond that it forms with the C<sub>e1</sub>-H bond of His<sub>57</sub>. The identification of this interaction dates back from the  $^1\text{H}$  NMR work of Markley in 1978.<sup>231</sup> Nonetheless, its definitive assignment is only relatively recent,<sup>5,222</sup> as is the recognition of the C-H...O interaction as a real H-bond.<sup>232</sup> A theoretical study on the energy involved in the C-H...O bond in amino acids affirms that the C-H group is a potent H-bond donor. The binding energy for the equilibrium separation (e.g. 3.339 Å for the alanine side-chain, -CH<sub>3</sub>, as proton

<sup>†</sup>The positions of the H atoms were not resolved in the X-ray structure.

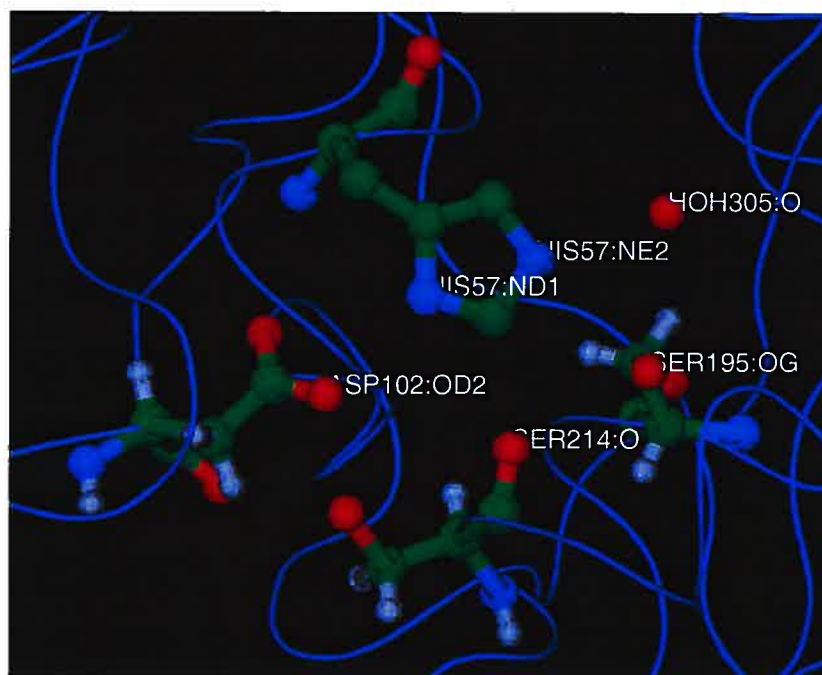


Figure 8.2: Active site of porcine pancreatic elastase acyl enzyme (PDB code 1GVK). Besides the residues composing the catalytic triad, Ser<sub>214</sub> and the water molecule H<sub>2</sub>O<sub>305</sub> are also shown.

donor) is  $\sim 2$  kcal/mol.<sup>233</sup> Furthermore, if we consider the same H-bond type but with an alkyl ammonium donor ( $\text{NR}_3^+-\text{CH}_3 \cdots \text{O}$ ) the interaction energy increases considerably.<sup>234</sup> The distance between the heavy atoms in the  $\text{C}_{\epsilon 1}-\text{H} \cdots \text{O}$  bond has been found experimentally to be between 3.28 and 2.88 Å for a wide variety of serine proteases.<sup>222</sup> Therefore, the energy involved in this bond should be in the same order of magnitude as the one calculated for the model systems.

The determination of which tautomer of the His side chain ( $\text{N}_{\delta 1}$  or the  $\text{N}_{\epsilon 2}$ ) is present in the catalytic triad of the resting enzyme has also been controversial for many years. The predominance of the more unstable tautomer (i.e.  $\text{N}_{\delta 1}$ ), and

therefore of the Asp in its ionized form, has been shown definitely by  $^{15}\text{N}$  NMR studies<sup>235,236</sup> and confirmed by molecular dynamics calculations.<sup>237,238</sup> Still, the role of the  $\text{N}_{\delta 1}\text{-H}\cdots\text{O}_{\delta 2}$  bond in the catalytic mechanism remains difficult to define. In fact, it is not yet certain if this strong H-bond remains unchanged throughout the whole catalytic process, as maintained by some,<sup>220,239</sup> or if it breaks in favor of another arrangement.<sup>5,230</sup>

The water molecule, shown in Fig. 8.2, has not been recognized to have a specific function like the other residues. Nonetheless, according to the X-ray structure of 1GVK<sup>240</sup> from which the model depicted has been derived, it appears that  $\text{H}_2\text{O}_{305}$  forms a strong H-bond with the  $\text{N}_{\epsilon 2}$  of His<sub>57</sub>. The distance between the  $\text{H}_2\text{O}$  oxygen and the  $\text{N}_{\epsilon 2}$  is in fact only 2.65 Å.

The oxyanion hole represents a pocket of positive charge and its role is to stabilize the oxyanion formed during the proteolytic reaction. In the chymotrypsin family it is formed by the amino groups of the backbone of Ser<sub>195</sub> and Gly<sub>193</sub>. In other serine protease clans it can contain other residues since the nature of the peptide side chain is not involved in the interaction.

## 8.2 The Generally Accepted Catalytic Mechanism

The production of such a large amount of experimental and theoretical studies on serine proteases has revealed that this family of proteolytic enzymes works through a very elegant catalytic mechanism. The main steps of the reaction are supported by irrefutable experimental proof and thus the salient features of the process are widely accepted within the scientific community. However, some intermediate stages occur so fast, or are too unstable, that they cannot be seen

experimentally and therefore are still a matter of debate.

It is generally accepted that the catalytic reaction of serine proteases develops in two phases. The qualitative reaction profile, based on the quantitative free energy study of West and coworkers,<sup>241</sup> and the structures involved are shown in Fig. 8.3 on page 180. The first phase is known as the acylation step or aminolysis. Following the formation of the Michaelis complex, a non-covalent complex between enzyme and substrate, the neutral His<sub>57</sub> acts as a general base. It accepts the Ser<sub>195</sub> hydroxyl proton, hence allowing the formation of an imidazolium ring. The positive charge of the His<sub>57</sub> side chain is balanced by the negative Asp<sub>102</sub> through their H-bond interaction. Meanwhile, the alkoxide ion of Ser<sub>195</sub> attacks the carboxylic carbon of the substrate causing the formation of an oxyanion. The latter is stabilized by the electrophilic amino groups of the oxyanion hole. The intermediate structure formed is known as the tetrahedral intermediate 1 (TI-1) because of the characteristic tetrahedral sp<sup>3</sup> carbon formed. Its presence has never been proven experimentally and it has been supposed only on the basis of solution chemistry. The lifetime of the tetrahedral intermediate has been determined through theoretical studies to be in the order of picoseconds.<sup>239</sup> However, it is known that its stability strictly depends on the pK<sub>a</sub> value of the leaving group. In fact, if the leaving group is optimal, its lifetime is so short that this intermediate state can overlap with the transition state.<sup>242-244</sup> The only method that allows the formation of stable tetrahedral intermediates is the use of particular inhibitors [e.g. difluorophosphates (DFP) and phenylmethanesulphonylfluorides (PMSF)]. These molecules block the catalytic reaction by forming stable adducts at the transition state/tetrahedral intermediate stage. Several serine proteases constrained in such complexes have been characterized through X-ray diffraction and NMR.

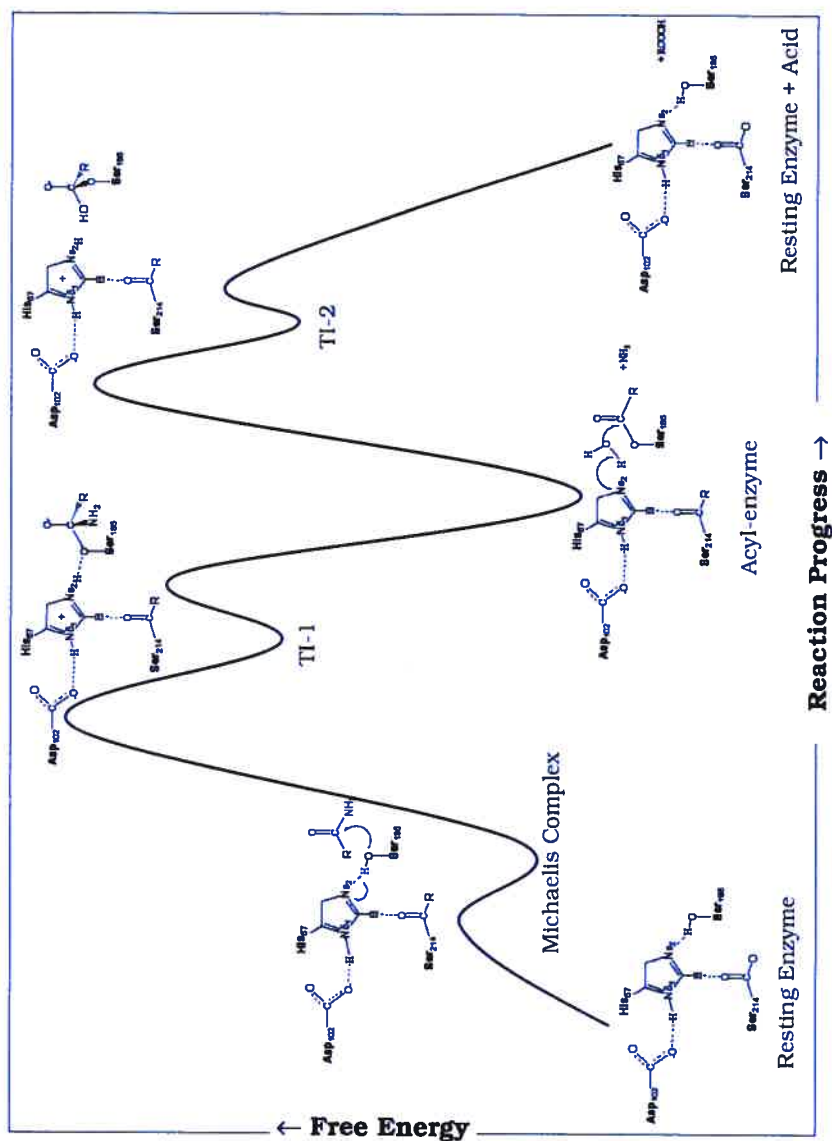


Figure 8.3: Qualitative reaction profile of the generally accepted catalytic mechanism of serine proteases. The information on the relative stability of each reaction stage has been based on the quantitative diagram in the work of West and coworkers.<sup>241</sup>



For the breakdown of TI-1 the charged His<sub>57</sub> acts as a general acid. N<sub>ε2</sub> releases its proton to the leaving group, in this case the amino group -NH<sub>2</sub>, which forms a free molecule of NH<sub>3</sub> and the acyl-enzyme complex. Contrary to TI-1, the acyl-enzyme complex is a stable intermediate and it has been observed experimentally.<sup>224–226</sup> This stage represents the end of the aminolysis and the beginning of the second part of the reaction known as the deacylation step, or hydrolysis. As the latter definition indicates, a water molecule plays the part of the nucleophile. The process is activated again by the His<sub>57</sub> N<sub>ε2</sub>, which acts as a general base and deprotonates the water molecule. The hydroxyl group so formed, attacks the carbonyl carbon of the acyl group. As a result, a second tetrahedral intermediate (TI-2) takes shape. The decomposition of TI-2 is perfectly analogous to the decomposition of TI-1. The charged His<sub>57</sub> acts as a general acid and protonates the oxygen in the Ser<sub>195</sub> alcoxide, which here corresponds to the leaving group. The products are a free acid and the resting enzyme.

The development of the catalytic mechanism, as described above, is based on the assumption that the H-bond between Asp<sub>102</sub> and His<sub>57</sub> endures almost unaltered throughout the reaction. To the advocates of this formulation, the characteristics of this H-bond are the key to the catalytic function of serine proteases. A controversial theory developed upon the discovery of unusually strong H-bonds in the gas-phase,<sup>245–247</sup> tries to explain the great stabilizing effect that this interaction exerts on the transition state/tetrahedral intermediate. According to this theory, the “normal” H-bond between Asp<sub>102</sub> and His<sub>57</sub> in the resting enzyme becomes a “low barrier hydrogen bond” (LBHB) when His<sub>57</sub> becomes protonated.<sup>248–251</sup> The stabilization due to this interaction is estimated to be between 10 and 24 kcal/mol on the basis of gas-phase experiments on model systems. The presence of an

LBHB in serine proteases was deduced from  $^1\text{H}$  NMR chemical shift data. In fact, LBHBs are characterized by an extremely lowfield  $^1\text{H}$  chemical shift, i.e.  $>16$  ppm.<sup>248</sup> Upon protonation the  $^1\text{H}$  chemical shift in the  $(\text{Asp}_{102})\text{-O}_{\delta 2} \cdots \text{H-N}_{\delta 1}\text{-(His}_{57})$  H-bond shifts from 13-15 ppm to 17-18 ppm. Besides the  $^1\text{H}$  chemical shift, other indicators of the existence of an LBHB are: matching  $\text{pK}_a$  values for the donor and the acceptor, complete seclusion from the solvent and obviously a short distance between the two heavy atoms (i.e.  $< 2.6 \text{ \AA}$ ). Although the latter condition applies,<sup>227</sup> the active site of serine proteases can hardly be described as secluded from the solvent.<sup>252</sup> Furthermore,  $^{15}\text{N}$  NMR chemical shift and coupling constant studies have proven that the proton is 85% localized on  $\text{N}_{\delta 1}$ ,<sup>252,253</sup> invalidating the matching  $\text{pK}_a$  values condition. Selective modifications of the residues in the catalytic triad have also proven that the energy of the supposed LBHB is definitely  $< 6$  kcal/mol, more like the energy of a “normal” H-bond.<sup>254,255</sup> These and other shortcomings<sup>239,256</sup> indicate the inconsistency of the LBHB theory.

Even within the hypothesis of an LBHB stabilizing the transition state of the serine proteases catalysis, the generally accepted catalytic mechanism has some flaws, of which the most evident was first recognized in 1970.<sup>257-259</sup> It pertains to the ability of  $\text{His}_{57}$  in catalyzing the formation of the tetrahedral intermediate and its destruction without any modification of the structure of the charge relay system and without the intervention of any external molecule or residue. In other words, there is no apparent reason why  $\text{His}_{57}$  would abstract a proton from the  $\text{Ser}_{195}$  side chain and then give it selectively to the amino group, the latter being a much worse leaving group ( $\text{pK}_b \approx 16$ ) than the alcoxide  $\text{Ser}_{195}\text{-O}$  (i.e.  $\text{pK}_b \approx 1$ ).

### 8.3 The “Ring Flip” Mechanism

A variation of the generally accepted mechanism has been proposed by Ash and coworkers<sup>5</sup> following the detection of the H-bond between C<sub>ε1</sub>-H and the backbone of Ser<sub>214</sub>.<sup>222</sup> It involves a 180° rotation of the imidazolium ring in His<sub>57</sub> around the C<sub>β</sub>-C<sub>γ</sub> bond subsequent to the acquisition of the proton from the Ser<sub>195</sub> hydroxyl group. As it is shown in Fig. 8.4 on page 184, after the formation of the Michaelis complex, the general base catalysis of His<sub>57</sub> takes place as in the conventional mechanism. N<sub>ε2</sub> acquires the hydroxyl proton and subsequently the alkoxide attacks the carbonyl carbon of the substrate. Plausibly, at the transition state, N<sub>ε2</sub>-H should be closer to the alkoxide group from which it has extracted the proton shortly before. Therefore, if the TI-1 structure remained the same, the preferred leaving group would most likely be the alkoxide, with the consequent inversion of the reaction. However, according to the “ring flip” mechanism, after the protonation, the imidazolium side chain rotates. The H-bond between N<sub>δ1</sub>-H and Asp<sub>102</sub> is broken. N<sub>δ1</sub> is now in a more favorable position, δ2, to donate the proton to the amino leaving group. N<sub>ε2</sub> becomes N<sub>ε1</sub> and forms an H-bond with the backbone of Ser<sub>214</sub>.

The flipped structure can activate more easily the water molecule for the nucleophilic attack of the carbonyl group in the acyl-enzyme intermediate (see Fig. 8.5 on page 186). Once TI-2 is obtained, His<sub>57</sub> can perform another rotation, bringing the catalytic triad to the original conformation. The latter is now more favorable for the protonation of the alkoxide of Ser<sub>195</sub> and the reinstatement of the resting enzyme.

A perfectly flipped conformation of the charge relay system has not been isolated yet. The closest structure to it has been identified by X-ray diffraction and

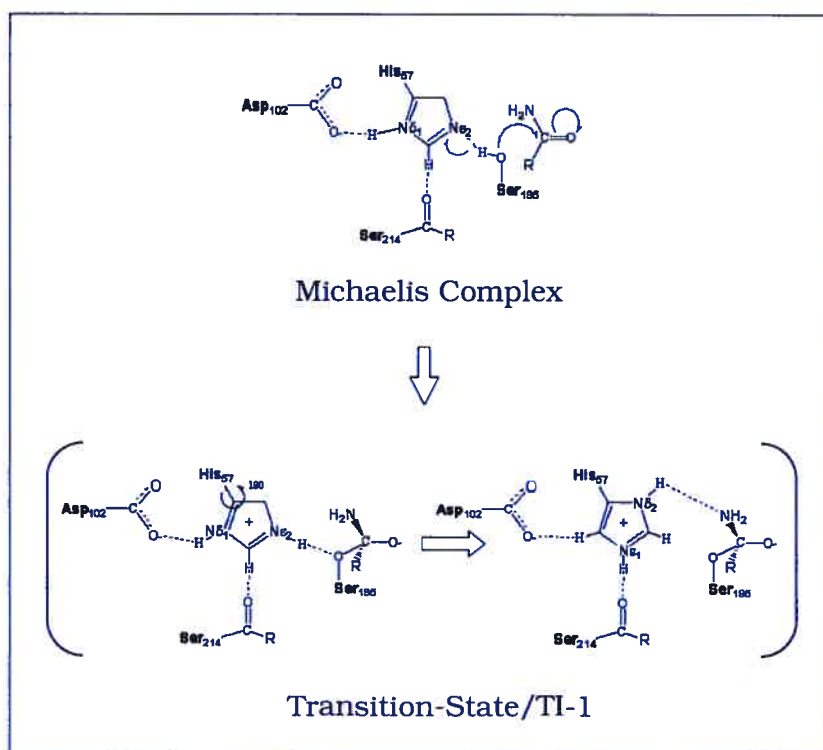


Figure 8.4: Formation of the TI-1 during the aminolysis according to the “ring flip” mechanism.

it corresponds to an almost completely rotated imidazole ring ( $164^\circ$ ) in the serine protease subtilisin BPN<sup>228</sup> in 50% DMS. This study confirms the hypothesis that N<sub>ε2</sub> methylated serine proteases and serine proteases in high concentration of organic co-solvents could cause the flipping of the neutral imidazole ring in order to perform the general base catalysis.<sup>241,260–263</sup> In these cases the enzyme is obviously destabilized and the most common result is the inversion of the catalytic path, i.e. a peptide synthesis instead of hydrolysis. The negative impact on the catalysis of a flipped ring in His<sub>57</sub> in the resting enzyme has been also confirmed by molecular dynamics calculations.<sup>238</sup> According to this study, the final structure obtained from the trajectory of the flipped rotamer is not catalytically productive.

On the other hand, the “ring flip” mechanism does not involve the flipping of a neutral imidazole. The resting enzyme is much more stable with His<sub>57</sub> in the N<sub>δ1</sub>-N<sub>ε2</sub> rotamer than with a flipped ring, simply because the latter cannot form the same number of H-bonds with the other residues of the catalytic triad. The flipping of the ring is predicted to take place exclusively when the imidazole is protonated, as it is in the tetrahedral intermediates 1 and 2. Only in this condition can the two rotamers form the same number of H-bonds and therefore their energy is comparable.<sup>5</sup> As mentioned, the lifetime of intermediates such as TI-1 or TI-2 is very short, making their structural characterization through the available experimental techniques nearly impossible. Nonetheless, strong arguments in favor of the His imidazole rotation come from <sup>14,15</sup>N NMR chemical shift data.<sup>235</sup>

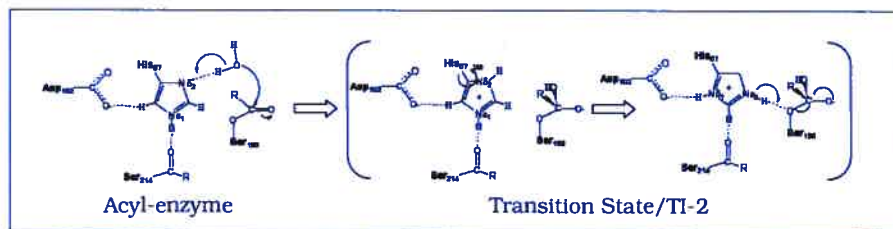


Figure 8.5: Formation of the TI-2 from the flipped acyl-enzyme during the hydrolysis.

## 8.4 Construction of a Model System for the Tetrahedral Intermediate

One of the challenges of this project has been the successful modeling of the TI-1 structure. The extremely short lifetime of this intermediate is the main cause for the absence of any direct experimental characterization of its structure. Therefore, all the available information have been derived from the TI-like inhibited complexes and from solution chemistry. In addition a significant obstacle is represented by the size of the enzyme. Serine proteases contain generally more than 2000 atoms, a prohibitive number to handle with DFT.

The design of the model system has been conceived keeping in mind the computational cost, while at the same time trying to maintain a close correspondence between the model structure and the experimental information. The area of interest concerns mainly the active site, and specifically the residues in the charge-relay system. Hence, our working set of residues has been reduced to Asp<sub>102</sub>, His<sub>57</sub>, Ser<sub>214</sub> and Ser<sub>195</sub>. Moreover, two water molecules have also been considered. One is the H<sub>2</sub>O<sub>305</sub> found in the 1GVK structure and shown in Fig. 8.2 on

page 177, which has been included for its close proximity to  $N_{\epsilon 2}$ . Meanwhile, the other replaces the  $-NH_2$  backbone group of the second residue forming the oxyanion hole (i.e. Gly<sub>193</sub> in the chymotrypsin family or Asn<sub>155</sub> in the subtilisin family). The presence of this latter group is fundamental to stabilize the open charge on the carboxylic oxygen created upon the formation of the tetrahedral carbon in the substrate. The substrate is indeed the last group included in the working model. A simple formamide molecule has been chosen to replace the peptide chain in order to reduce the size of the system, and also since we could not include in the active site model any part of the substrate binding sites.

The final objective is to determine which one of the two possible mechanisms predicts the “correct” reaction intermediate. This problem can be approached through the comparison between the  $^{14,15}N$  NMR calculations on the TI-1 predicted by the two mechanisms and the experimental  $^{14,15}N$  NMR shielding data determined on a low pH TI-1-like inhibited complexes.<sup>235</sup>

Two prototypes of TI-1 have been modeled, one according to the generally accepted mechanism (see Fig. 8.6 on page 188) and the other according to the “ring flip” mechanism (see Fig. 8.7 on page 188). The geometry of each structure has been obtained through a constrained optimization<sup>†</sup>, where the positions of the atoms that are normally part of the protein backbone have been kept fixed. The imidazolium atoms  $N_{\epsilon 1}$  in Fig. 8.7 and the  $C_{\epsilon 1}$  in Fig. 8.6 have been constrained also. The need for these additional constraints derives from the fact that the absence of the proteic environment, surrounding the active site, enhances the mobility of the His ring. The strength of the H-bonds is not sufficient to hold together the model system throughout the optimization. An additional control is

---

<sup>†</sup>RMS gradient 0.001

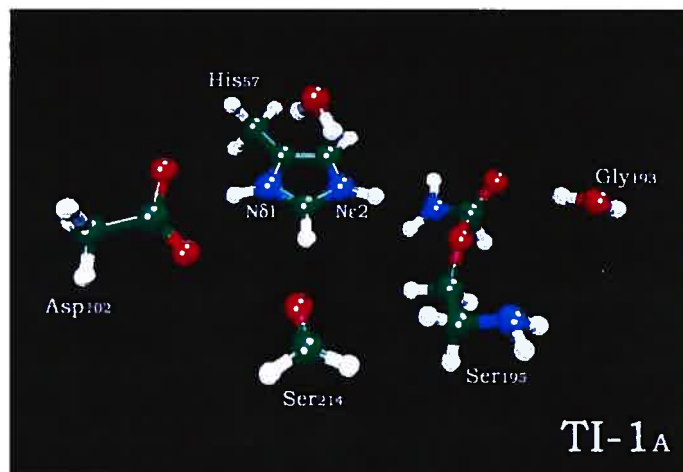


Figure 8.6: Model of the tetrahedral intermediate 1 (TI-1<sub>A</sub>) built according to the generally accepted mechanism.

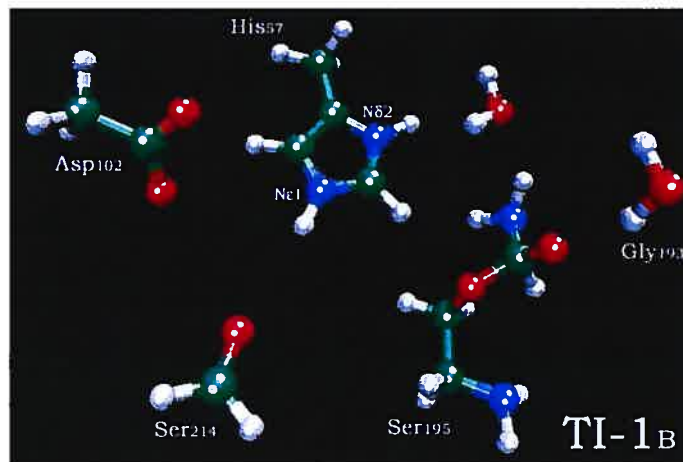


Figure 8.7: Model of the tetrahedral intermediate 1 (TI-1<sub>B</sub>) built according to the “ring flip” mechanism.



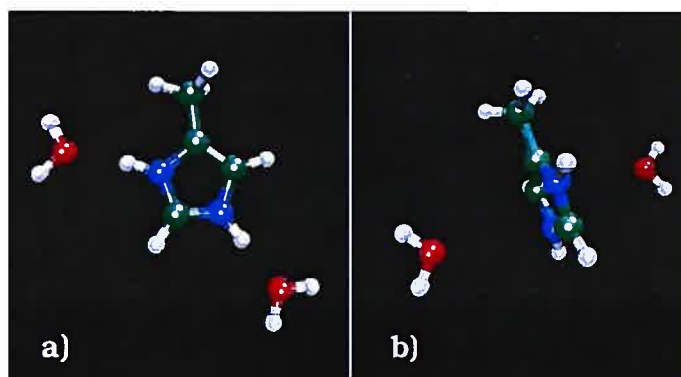


Figure 8.8: The two models indicated as a) and b) have been constructed in order to evaluate approximately the energy required for the breaking of the two strong H-bonds in the active site of serine protease.

applied to the H-bond length between the  $N_{\delta 1}$  and the  $Asp_{102}$  in TI-1<sub>A</sub>. In fact, as mentioned, the design of the generally accepted catalytic mechanism is based on the endurance of this interaction throughout the reaction path. The strength of this interaction has been considered key to the catalytic function of the enzyme. Therefore, the distance between  $N_{\delta 1}$  and the carboxylic oxygen of  $Asp_{102}$  has been kept fixed to the experimental value of  $2.63865 \text{ \AA}$ <sup>227</sup> throughout the optimization. If this constraint is released the imidazolium ring performs a small rotation around the  $C_{\beta}$ - $C_{\gamma}$  bond<sup>§</sup>, which is sufficient to break the  $O_{\delta 2} \cdots H-N_{\delta 1}$  H-bond.<sup>264</sup> The corresponding structure has been analyzed as an example for the “moving histidine” theory, which predicts that small rotations of the  $His_{57}$  imidazolium ring are responsible for the progress of the catalytic mechanism (see Sec. 8.7).

The energetic balance is a crucial element in the evaluation of the “ring flip” reaction path. The rotation of the imidazolium ring requires the rupture of two strong H-bonds: the one between  $His_{57}-N_{\delta 1}$  and  $Asp_{102}-O_{\delta 2}$ , and the other be-

<sup>§</sup>The dihedral angle between  $Asp_{102}-O_{\delta 2}$   $His_{57}-C_{\gamma}$   $His_{57}-C_{\epsilon 1}$  and  $His_{57}-H$  is  $19.56^{\circ}$ .

tween His<sub>57</sub>-N<sub>ε2</sub> and Ser<sub>195</sub>-O-R (see Fig. 8.6). The energy needed to perform this rotation has been evaluated through a very simple model, which involves a 5-Me-imidazolium, as a surrogate for His<sub>57</sub>, and two water molecules H-bonded respectively to the N<sub>δ1</sub> and the N<sub>ε2</sub> (see structure a) on Fig. 8.8). A second structure has been constructed with the 5-Me-imidazolium ring perpendicular to the two water molecules, as shown on Fig. 8.8 b). The total energy difference between these two models is 15.9317 Kcal mol<sup>-1</sup><sup>¶</sup>. This estimate is in agreement with the value of ~15 Kcal mol<sup>-1</sup> predicted by Bachovchin<sup>230</sup> for the protein. It is reasonable to presume that the free energy barrier would be lower due to a positive entropic contribution introduced by the free rotation of the ring.

According to the Ash and coworkers, the formation of a TI-1 involving a flipped His<sub>57</sub> is more favorable energetically.<sup>5</sup> Therefore, the cost of breaking the H-bonds can be explained in terms of a higher stabilization of the reaction intermediates. The energetic analysis of the two TI-1 models studied confirms that the flipped conformer (i.e. TI-1<sub>B</sub>) is more stable than TI-1<sub>A</sub> by 17.1076 kcal mol<sup>-1</sup>.

## 8.5 $^{14,15}\text{N}$ NMR Shielding in His<sub>57</sub>

Nitrogen NMR is considered one of the most powerful experimental tools available for the study of protein structure and function. This is especially true in the case of serine proteases, where, as it has been underlined previously, the mechanistic function is strictly related to the nature of the two nitrogen atoms in the

<sup>¶</sup>PLAP3 functional, FINE GRID with 64 angular points per shell, TZVP basis set (see Ch. 5 for further details).

catalytic His side chain. According to the definition introduced by Witanowski and coworkers,<sup>265</sup> the pyrrole-type and the pyridine-type nitrogens in the neutral imidazole can be indicated as  $\alpha$  and  $\beta$ -type, respectively. In the imidazolium ring both nitrogens are in the same protonation state and are simply indicated with  $\alpha^+$ . This classification is shown in Fig. 8.9.

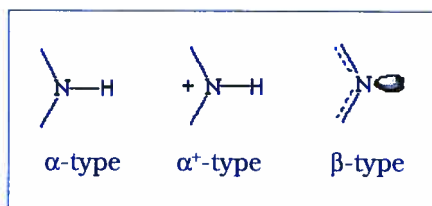


Figure 8.9: Types of nitrogen nuclei of the catalytic His involved in the different steps of the serine protease reaction path.

The  $\alpha$  and  $\beta$ -type nitrogens can be easily distinguished by the position of their NMR shielding signals. For instance, in Tab. 8.I are shown the  $^{14,15}\text{N}$  NMR calculated and experimental shieldings of imidazole, N(Me)-imidazole and of the imidazole ring in the His side-chain. The neutral imidazole has  $C_{2v}$  symmetry hence the two tautomers are equally probable. The tautomeric equilibrium is so fast that only the average shielding value can be measured experimentally. The calculated counterpart shown is also the average of the shieldings obtained for the  $\alpha$  and  $\beta$ -type nitrogens. Here, the Loc.3 prediction is 19.4 ppm too deshielded, however it is not possible to see which of the nitrogen types is the most difficult to describe. In order to do that, the results for the N-methylated imidazole are also shown. In this case, the calculated NMR shielding of both nitrogen types is underestimated by  $\sim 13$  ppm. The deviation from the experiment is in fact larger for an  $\alpha$ -type nitrogen bonded to a methyl group than for the same type of nitrogen

but bonded to an H atom (see Ch. 7). The equivalent level of accuracy attained for  $\alpha$  and  $\beta$ -nitrogens allows one to obtain a very good estimate of the difference between their shieldings. (i.e., 109.2 ppm from Loc.3 and 109.0 ppm from Exp.).

Table 8.I:  $^{14,15}\text{N}$  shieldings (ppm from neat liquid  $\text{MeNO}_2$ ) of  $\alpha$ ,  $\alpha^+$  and  $\beta$  nitrogen types in imidazole, N(Me)-imidazole and in His.

	N-type	SOS-DFPT (Loc.3)	Exp.*	Solvent
Imidazole	$\alpha(\beta)$	157.8 <sup>†</sup>	177.2	$\text{H}_2\text{O}$
	$\alpha^+$	201.9	208.2	$\text{H}_2\text{O}$
N <sub><math>\delta</math>1</sub> -Me-Imidazole	$\alpha$	205.2	218 $\pm$ 1	$\text{CCl}_4$ <sup>‡</sup>
	$\beta$	96.0	109 $\pm$ 3	$\text{CCl}_4$ <sup>‡</sup>
	$\alpha_{\text{N}\delta 1}^+$	181.3	208.7 <sup>§</sup>	$\text{H}_2\text{O}$
	$\alpha_{\text{N}\epsilon 2}^+$	203.5	207.6 <sup>§</sup>	$\text{H}_2\text{O}$
His	$\alpha_{\text{N}\delta 1}$	211.6	212.9 <sup>§</sup>	$\text{EtOH}$ <sup>¶</sup>
	$\beta_{\text{N}\epsilon 2}$	91.9	136.2 <sup>§</sup>	$\text{EtOH}$ <sup>¶</sup>
	$\alpha_{\text{N}\delta 1}^+$	193.1	203.4 <sup>§</sup>	$\text{EtOH}$ <sup>¶</sup>
	$\alpha_{\text{N}\epsilon 2}^+$	208.8	206.0 <sup>§</sup>	$\text{EtOH}$ <sup>¶</sup>

The proficiency of the calculations has been tested also for a model of the His amino acid, which has been obtained from His<sub>57</sub> of the 1GVK PDB structure where H atoms have been added to the missing valences. This model of His reflects closely part of the system in the active site and thus it represents a very good case for the benchmark test of the calculations. At the same time, however, it also presents considerable difficulties. Unlike the N(Me)-imidazole, both tau-

\* Experimental data from Witanowski and coworkers<sup>193</sup> unless otherwise specified

<sup>†</sup> 215.2 ppm for N <sub>$\delta$ 1</sub> and 100.4 ppm for N <sub>$\epsilon$ 1</sub>

<sup>‡</sup> Experimental data from Witanowski and coworkers<sup>194</sup>

<sup>§</sup> Experimental data from Bachovchin,<sup>235</sup> +4.6 ppm correction to change ref. from 1M  $\text{HNO}_3$  to  $\text{CH}_3\text{NO}_2$ <sup>183</sup>

<sup>¶</sup> Experiment performed with EtOH 80% in vol. and at -60°C in order to slow down the tautomeric equilibrium.

tomers can play a role, even though the reference values have been obtained in an experiment conducted at low temperature (i.e. -60°C). Furthermore, there is a significant solvent effect to take into account. While the  $\alpha$ -type nitrogen seems to be less affected by these problems, showing a deviation of only 1.3 ppm downfield, the  $\beta$ -nitrogen is too deshielded by 44.3 ppm. Considering the strong shielding effect induced by protic solvents on the pyridine-type nitrogen, as we have seen in Sec. 7.6 for azines, this error can very well be due to the solvent effect.

The role played by the His backbone on the NMR shielding of the nitrogens in the side-chain has been evaluated considering 5-Me-imidazole as a model for His. The results, shown in Tab. 8.II, indicate that this considerable reduction of the system's size is responsible for a small deviation of the  $\alpha$  and  $\beta$ -nitrogen shieldings.

Table 8.II: <sup>14,15</sup>N shieldings (ppm from neat liquid MeNO<sub>2</sub>) of  $\alpha$ ,  $\alpha^+$  and  $\beta$  nitrogen types in 5-Me-imidazole.

	N-type	SOS-DFPT (Loc.3)	Exp.(His)*	Solvent†
5-Me-Imidazole	$\alpha_{N\delta 1}$	206.9	212.9	EtOH
	$\beta_{N\epsilon 2}$	95.3	136.2	EtOH
	$\alpha^+_{N\delta 1}$	192.7	203.4	EtOH
	$\alpha^+_{N\epsilon 2}$	198.1	206.0	EtOH

The  $N_{\delta 1}$  ( $\alpha$ -type) and the  $N_{\epsilon 2}$  ( $\beta$ -type) signals are displaced, respectively, only 4.7 ppm downfield and 3.4 ppm upfield, compared to the results obtained for the previous His model. These shifts are minimal compared to the reduction of the computational cost. Therefore, the level of accuracy in the calculation of the two types of nitrogens can be considered as only marginally affected by the model's

\* Experimental data from Bachovchin<sup>235</sup>

† Experiment performed with EtOH 80% in vol. and at -60°C.

modification.

It remains to discuss the nitrogen shielding in the protonated imidazole, i.e. the  $\alpha^+$ -type. As we can see from both Tab. 8.I and 8.II, the agreement with the experiment is rather good. For 5-Me-imidazolium the average deviation is -7.9 ppm. The importance of this result is related to the fact that this nitrogen-type is present in all the intermediate states of the catalytic process and, as we will see shortly, its correct reproduction by the calculations is a key element in the study of the reaction mechanisms.

## 8.6 H-bond Induced Shift

The  $^{14,15}\text{N}$  shieldings of the imidazole ring are particularly sensitive to H-bonding. While the order of magnitude of the induced shifts is similar for  $\alpha$ ,  $\beta$  and  $\alpha^+$ -nitrogens (i.e.  $\sim 10$  ppm), its direction depends on the nitrogen type.<sup>266</sup> Upon formation of a H-bond with a carboxylate cation, the  $\alpha$  and  $\alpha^+$ -type nitrogen shieldings experience a downfield displacement of -10 ppm. Meanwhile, the formation of an H-bond to a neutral -OH group causes the the  $\beta$ -type  $^{14,15}\text{N}$  signal to undergo a shielding of +10 ppm.<sup>230</sup> The conclusive proof of the predominance of the  $\text{N}_{\delta 1}$  tautomer on the resting enzyme of serine proteases was based precisely on the different effect that the H-bonding has on the NMR shielding of  $\alpha$  and  $\beta$ -nitrogens.<sup>235</sup> It is therefore crucial to assess the capability of the calculation in reproducing this behavior correctly. In order to do this, three different systems have been studied, each of them focusing on a particular nitrogen type.

The first system consists of an imidazole ring with each of the  $\text{N}_{\delta 1}$  and  $\text{N}_{\epsilon 2}$  H-bonded to a methanol molecule (see structure a) on Fig. 8.10 on page 195).

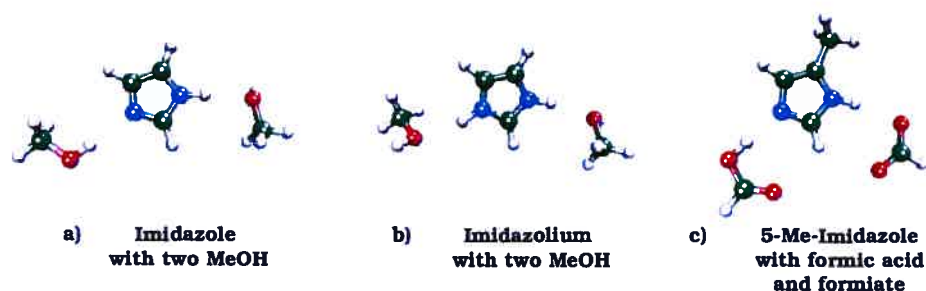


Figure 8.10: Three model systems built to analyze the H-bond induced shifts on the imidazole  $^{14,15}\text{N}$  shieldings.

Through this model we can assess the effects on the  $\alpha$  and  $\beta$ -nitrogens. The second system includes an imidazolium ring with two MeOH molecules and considers the H-bond effect on the shielding of the  $\alpha^+$ -nitrogen (see structure b) on Fig. 8.10). The third system, indicated as c) on Fig. 8.10, differs from the others in that  $\text{N}_{\delta 1}$  forms an H-bond with a charged group, i.e. a formiate. This model mimics the H-bond that takes place in the catalytic triad of the resting serine protease between  $\text{His}_{57}\text{-N}_{\delta 1}$  and  $\text{Asp}_{102}\text{-O}_{\delta 2}$ , where the formiate plays the role of the ionic  $\text{Asp}_{102}$ . Meanwhile, the substitution of methanol with formic acid in the H-bond with  $\text{N}_{\epsilon 2}$  is aimed at determining the impact of the strength of the interaction on the  $^{14,15}\text{N}$  shielding of the  $\beta$ -type nitrogen. The results are shown in Tab. 8.III on page 196. The H-bond induced shifts ( $\Delta\sigma$ ) have been evaluated relative to the isolated molecules, for which the results are shown in Tab. 8.I for imidazole and imidazolium and in Tab. 8.II for 5-Me-imidazole.

The calculations slightly overestimate the H-bond effect on the  $\alpha$  and  $\alpha^+$ -type nitrogens when H-bonded to methanol. In fact, the H-bonds in the a) and b) systems are considerably weaker than the  $\text{N}_{\delta 1}\text{-H}\cdots\text{O}$  H-bond in the c) complex for which the experimental estimated shift is 8-10 ppm. The H-bond induced shift

on the  $\beta$ -type nitrogen in a) is rather small, and well within the level of uncertainty determined for this type of nitrogen.

Table 8.III: H-bond effect on the  $^{14,15}\text{N}$  shielding of the  $\alpha$  and  $\beta$  nitrogen types in imidazole and 5-Me-imidazole and of the  $\alpha^+$  nitrogen type in imidazolium. The systems studied are represented in Fig. 8.10.

	N-type	SOS-DFPT (Loc.3)*	$\Delta\sigma^\dagger$
System a) <sup>‡</sup>	$\alpha$	204.9	-10.3
	$\beta$	107.0	+6.6
System b) <sup>§</sup>	$\alpha^+$	189.7	-12.2
System c) <sup>¶</sup>	$\alpha$	167.1	-39.8
	$\beta$	124.5	+29.2

The effect of the formation of a strong H-bond between the  $\text{N}_{\delta 1}$  and the oxygen of the formate group in c) is significantly amplified by the calculation. For both  $\alpha$  and  $\beta$ -type nitrogens the induced shifts are overestimated. The largest effect is given by the very strong downfield shift of the  $\alpha$ -type  $^{14,15}\text{N}$  shielding (i.e. -39.8 ppm). This could be related to the modification of the acidity constants in the gas-phase.<sup>267</sup> The absence of solvent can significantly decrease the strength of an acid. In this particular case, a higher  $\text{pK}_a$  for the formic acid implies the formation of a much stronger H-bond with the 5-Me-imidazole than in the one in the liquid-phase.

It is known that the strength of an H-bond interaction depends highly on their

\*The chemical shieldings are expressed in ppm units relative to neat liquid  $\text{MeNO}_2$

<sup>‡</sup>Deviation from the NMR shielding calculated for the isolated molecule (see Tab. 8.I for details).

<sup>‡</sup>Distances between heavy atoms:  $\text{N} \cdots \text{H}-\text{O}$  2.9526 Å,  $\text{N}-\text{H} \cdots \text{O}$  2.9987 Å

<sup>§</sup>Distances  $\text{N}-\text{O}$  2.7789 Å,  $\text{O}-\text{H}$  1.7480 Å; angle  $\text{O}\hat{\text{H}}\text{N}$  178.87°

<sup>¶</sup>Distances  $\text{N}_{\delta 1}-\text{O}$  2.6840 Å,  $\text{O}-\text{H}$  1.6270 Å, angle  $\text{N}_{\delta 1}\hat{\text{H}}\text{O}$  171.70°; distances  $\text{N}_{\epsilon 2}-\text{O}$  2.6263 Å,  $\text{N}_{\epsilon 2}-\text{H}$  1.5842 Å, angle  $\text{N}_{\epsilon 2}\hat{\text{H}}\text{O}$  167.80°



structural parameters, such as interatomic distances and bond angles.<sup>268</sup> Therefore, the dependence of the  $N_{\delta 1}$  shielding upon the varying  $N_{\delta 1}$ -H distance has been analyzed as a key indicator of the H-bond strength. This test involves the optimization of the H-bond network connecting a 5-Me-imidazole and a 5-Me-imidazolium to a formiate, through  $N_{\delta 1}$ , and to a methanol through  $N_{\epsilon 2}$  (see Fig. 8.11 on page 198). In the case of the 5-Me-imidazole we compare the calculations with experimental shieldings obtained on similar systems.<sup>269</sup> On the other hand, for the 5-Me-imidazolium no experimental references were available. The steepest descent optimization path, involving the transfer of the proton from the imidazolium  $N_{\delta 1}$  to the formiate oxygen, has been followed by the shielding calculations for  $N_{\delta 1}$ . Due to limitations in the calculation time, the NMR calculations in this test have been performed with the Loc.2 approximation. Nonetheless, in view of the results obtained in Ch. 6 and 7, we can consider the Loc.2 calculations indicative of the trends attainable with Loc.3. The results are shown in Fig. 8.11 on page 198.

The  $N_{\delta 1}$  shielding shows a very high sensitivity to changes in the  $N_{\delta 1}$ -H distance. In fact, the  $N_{\delta 1}$  signal undergoes a progressive deshielding from 194.6 ppm, corresponding to an  $N_{\delta 1}$ -H distance of 1.0170 Å in the neutral 5-Me-imidazole, to 110.0 ppm, corresponding to an  $N_{\delta 1}$ -H distance of 1.5128 Å in the positively charged 5-Me-imidazolium. The results for the 5-Me-imidazole are in good agreement with the trends observed in the experimental solid state NMR  $^{14,15}\text{N}$  chemical shifts of similar complexes.<sup>269</sup> The deviations from the experiment oscillate between -10 and -20 ppm. This is a fairly reasonable result, especially if we consider that the experimental uncertainty on the N-H distance is on average 0.02 Å, which entails a  $^{14,15}\text{N}$  shielding downshift of ~10 ppm.

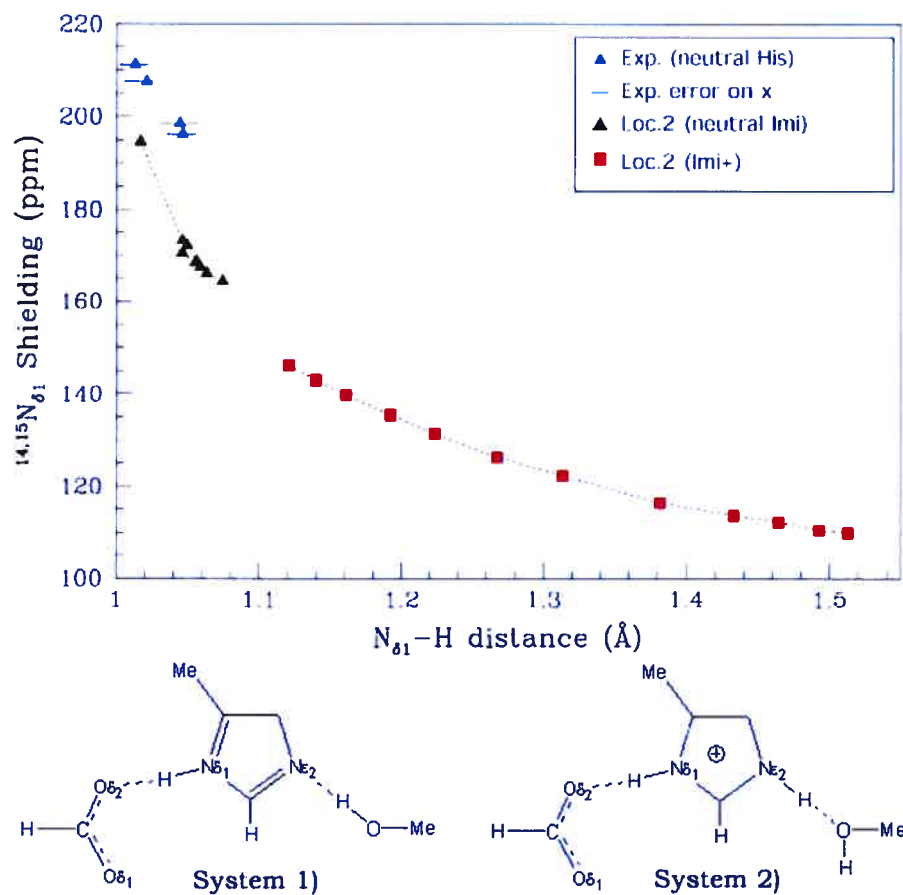


Figure 8.11:  $\text{N}_{\delta_1}$  NMR shielding dependence upon the varying  $\text{N}_{\delta_1}\text{-H}$  distance. The NMR calculations have been performed on the two systems, indicated as 1) and 2), shown schematically below the diagram.

The variation of the N<sub>δ1</sub>-O distance influences the <sup>14,15</sup>N shielding value indirectly. In fact, for short and strong H-bonds (i.e. N···O < 2.7 Å) the modification of the N···O distance influences the length of the covalent N-H bond.<sup>247,270,271</sup>

Given the uncertainty associated with the methodology and the simplicity of the model used to mimic the complex proteic environment, these results can rightfully be regarded as positive. Further, the most important aspect of this test pertains to the direction of the induced shifts. In fact, the correct reproduction of the direction allows us to interpret through the calculation on suitable model systems the formation and the breaking of the H-bonds seen experimentally, even if the absolute value of the calculated induced shifts is overestimated. As we can see in Tab. 8.III and in Fig. 8.11, the direction of the induced shifts is perfectly reproduced by the calculations.

## 8.7 <sup>14,15</sup>N Shieldings in TI-1

The results obtained in the tests shown in Sec. 8.5 and 8.6 are crucial for the study of the catalytic mechanism of serine proteases. Through the first test we were able to assess how accurately the Loc.3 approximation reflects the experimental data, while the second test was intended to reveal the ability of the calculations to reproduce the H-bond induced shifts. Also, considering the outcome of the calculations on azoles and azines discussed in Ch. 7, it is not surprising that the most difficult nucleus to describe with SOS-DFPT is the β-type nitrogen. The deviation of the calculated results for both His models (see Tab. 8.I and 8.II) is over -40 ppm. Even though the H-bonding induced shift is clearly reproduced, the margin of error is so large that no reliable study can be conducted on the NMR

shielding of this type of nucleus.

Very encouraging results are obtained, however, in the case of  $\alpha$  and  $\alpha^+$ -type nitrogens. The deviations calculated for the His model (i.e. 5-Me-imidazole) are -6.0 ppm and -9.3 ppm for  $\alpha$  and  $\alpha^+$ -types respectively. The H-bonding effect is also well reproduced, with the intensity of the induced shifts reflecting the strength of the interaction. Therefore, the Loc.3 approximation can be rightfully used to analyze the experimental results pertaining to  $\alpha$  and  $\alpha^+$ -type nitrogens in serine proteases.

The validation of one of the two catalytic mechanisms represents the perfect framework to employ Loc.3. In fact, as we have seen in Fig. 8.4, the ring flip occurs after the general base catalysis, i.e. after the protonation of the imidazole ring. In other words, the only steps along the reaction path involving structural differences are the tetrahedral intermediates and the acyl enzyme. While the latter includes a neutral imidazole, hence with  $\alpha$  and  $\beta$ -type nitrogens, the TI-1 and TI-2 contain a protonated His with only  $\alpha^+$ -type nitrogens.

The NMR experimental data supporting the “ring flip” mechanism have been obtained from a TI-like inhibited serine protease at low pH (see in Fig. 8.12 the active site structure of an analogous complex). Bachovchin<sup>235</sup> reported that by lowering the pH of DIFP- and PMSF-inhibited  $\alpha$ -lytic proteases complexes, both  $N_{\delta 1}$  and  $N_{\epsilon 2}$  show a signal characteristic of an  $\alpha^+$ -type nucleus not engaged in H-bonds<sup>||</sup>. In the resting  $\alpha$ -lytic protease the drop of the pH causes a downfield shift of the  $N_{\delta 1}$  signal from 204.0 to 196.2 ppm. These data are perfectly consistent with the conversion of an  $\alpha$ -type nitrogen into an  $\alpha^+$ -type nitrogen, both

<sup>||</sup>The diisopropylfluorophosphate (DIFP) and phenylmethanesulphonyl fluoride (PMSF) inhibitors bind the enzyme in a TI-like configuration (see also Sec. 8.2).

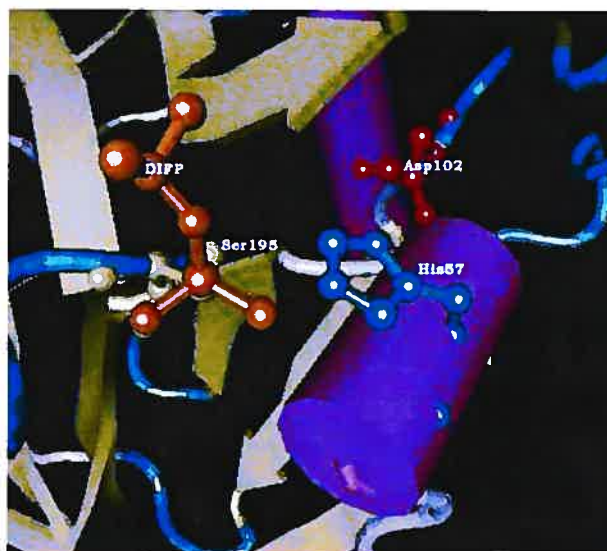


Figure 8.12: Active site of  $\gamma$ -chymotrypsin complexed with DIFP (PDB identifier 1GMH). His<sub>57</sub> in cyan, Asp<sub>102</sub> in red, Ser<sub>195</sub> in yellow and DIFP in orange.

H-bonded to the Asp<sub>102</sub>. However, the lowering of the pH of the inhibited complexes causes the N<sub>δ1</sub> NMR signal to shift from 204.3 to 202.6, for the PMSF-complex, and from 210.5 to 203.1 ppm, for the and DIFP-complex. At that time these results were simply interpreted as a sign of a “moving His”. According to this theory, upon protonation the H-bond between N<sub>δ1</sub> of His<sub>57</sub> and O<sub>δ2</sub> of Asp<sub>102</sub> breaks and the His ring is then free to rotate a few degrees around the C<sub>β</sub>-C<sub>γ</sub> bond. This behavior is completely reversed when the pH is increased.<sup>230,235</sup> The hypothesis of a complete rotation, i.e. flipping, of the imidazolium ring was not considered.

The calculations conducted on a model of TI-1 demonstrate that these NMR experimental data are consistent also with a flipped tetrahedral intermediate. The formation of this structure, shown schematically in Fig. 8.4, requires, in fact, the

rupture of the  $\text{N}_{\delta 1}\text{-H}\cdots\text{O}_{\delta 2}$  H-bond.

### 8.7.1 Loc.3 $^{14,15}\text{N}$ Shieldings of the TI-1 Model Systems

The NMR calculations have been performed on more compact versions of the structures TI-1<sub>A</sub> and TI-1<sub>B</sub> shown, respectively, in Fig. 8.6 and 8.7 on page 188. These have been obtained by eliminating the groups not involved in direct contacts with His<sub>57</sub>, which therefore do not influence the  $^{14,15}\text{N}$  shieldings. In the case of the TI-1<sub>A</sub>, this implies the elimination of the Ser<sub>214</sub> surrogate, of the oxyanion hole (i.e.  $\text{-NH}_2$  of Ser<sub>195</sub> and  $\text{H}_2\text{O}$  replacing  $\text{-NH}_2$  of Gly<sub>193</sub>) and of the water molecule. Tests have shown that the latter does not play an essential role in the NMR shielding of the  $\text{N}_{\epsilon 2}$ . Furthermore, because of the elimination of the oxyanion hole, the charged oxygen on the tetrahedral carbon had to be replaced by a hydrogen. Meanwhile, with regards to the NMR model of the “ring flip” intermediate, TI-1<sub>B</sub>, the reduction scheme followed is the same as for TI-1<sub>A</sub>, except that the Ser<sub>214</sub> substitute remains, due to the fact that it interacts with  $\text{N}_{\epsilon 1}$  through a weak H-bond. Both reduced models are shown in Fig. 8.13.

The results are shown in Tab. 8.IV on page 204. The main difference between the two set of calculations is in the shieldings of  $\text{N}_{\delta 1}$  in TI-1<sub>A</sub> and  $\text{N}_{\delta 2}$  in TI-1<sub>B</sub>. The former signal is very deshielded because of the presence of the strong H-bond with the carboxylic group of Asp<sub>102</sub>. The  $\text{N}_{\delta 1}\text{-H}$  distance is 1.1778 Å, which corresponds, according to the results shown in Fig. 8.11, to a significantly deshielded signal. As aforementioned, the acidity of the formic acid is underestimated in gas-phase causing the formation of a very strong H-bond with the imidazolium ring. Meanwhile, the Asp<sub>102</sub> in serine proteases has been proven to retain its strong acid character (i.e.  $\text{pK}_a=3$ ).<sup>230,272–274</sup> Hence, the absence of the rest of the protein

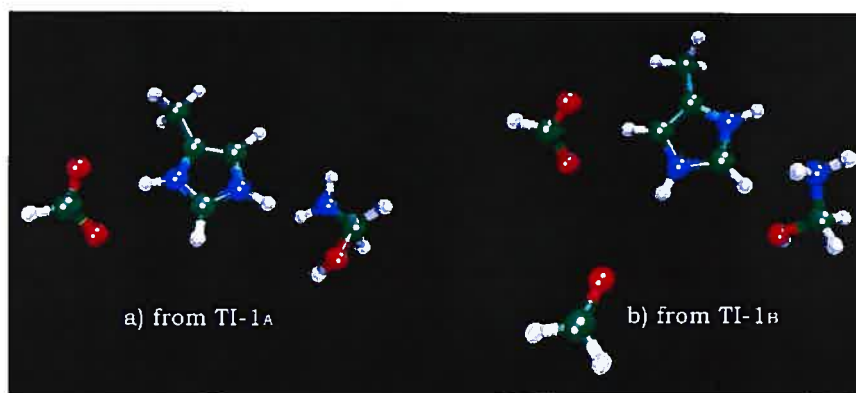


Figure 8.13: Reduced version of the TI-1<sub>A</sub> and TI-1<sub>B</sub> models used to perform the NMR calculations.

in the model can be the cause of a slight overestimation of the strength of this H-bond. Nevertheless, it is important to underline that the occurrence of an unusually strong H-bond between Asp<sub>102</sub> and His<sub>57</sub> is not observed experimentally either.<sup>230,235</sup>

The N<sub>δ2</sub> shielding of the “ring flip” TI-1<sub>B</sub> model corresponds to a free α<sup>+</sup>-type nitrogen non bonded to any strong H-bond acceptor. The result obtained for N<sub>ε1</sub> reflects a weak interaction with the Ser<sub>214</sub> carboxylic oxygen. As we have been able to assess through the test on the H-bond induced shifts in Sec. 8.6, Loc.3 tends to overestimate slightly the effect of weak H-bonds, compared to experiment. Therefore we expect this value to be overly deshielded by 5-10 ppm. The removal of the Ser<sub>214</sub> substitute causes the N<sub>ε1</sub> shielding to move upfield to 207.3 ppm.

The hypothesis of a “moving” His,<sup>235</sup> has also been analyzed (i.e. TI-1<sub>A'</sub>). As we have seen in Sec. 8.4, the removal of the constraint holding Asp<sub>102</sub>-O<sub>δ2</sub> at the experimental distance during the optimization, causes a small rotation of the

imidazolium ring relative to the Asp<sub>102</sub>. This movement is sufficient to break the H-bond between the His<sub>57</sub>-N<sub>δ1</sub> and the Asp<sub>102</sub>-O<sub>δ2</sub>. The N<sub>δ1</sub>-H distance in TI-1<sub>A'</sub> is contracted to 1.0142 Å. Consequently, the N<sub>δ1</sub> signal shifts upfield to 194.0 ppm, while the N<sub>ε2</sub> shielding is 213.5 ppm. The N<sub>δ1</sub> shielding here is influenced significantly less than in TI-1<sub>A</sub> by the interaction with the ionic Asp<sub>102</sub>, and it represents a case very similar to the N<sub>ε1</sub> shielding in the flipped rotamer TI-1<sub>B</sub>.

Table 8.IV: Experimental <sup>14,15</sup>N shieldings (ppm from neat liquid MeNO<sub>2</sub>) of the α-lytic protease inhibited complexes in a full protonation state compared to the Loc.3 calculations on models of the TI-1.

Experiment	N-type	<sup>14,15</sup> N σ (ppm)*	Δσ
α-lytic protease + PMSF	α <sup>+</sup> <sub>N<sub>δ1</sub></sub>	202.6	3.5
	α <sup>+</sup> <sub>N<sub>ε2</sub></sub>	206.1	
α-lytic protease + DIFP	α <sup>+</sup> <sub>N<sub>δ1</sub></sub>	203.1	3.6
	α <sup>+</sup> <sub>N<sub>ε2</sub></sub>	206.7	
Model Systems	N-type	<sup>14,15</sup> N σ (ppm)*	Δσ
TI-1 <sub>A</sub> (no flip)	α <sup>+</sup> <sub>N<sub>δ1</sub></sub>	140.8	58.6
	α <sup>+</sup> <sub>N<sub>ε2</sub></sub>	199.4	
TI-1 <sub>A'</sub> ("moving" His)	α <sup>+</sup> <sub>N<sub>δ1</sub></sub>	194.0	19.4
	α <sup>+</sup> <sub>N<sub>ε2</sub></sub>	213.4	
TI-1 <sub>B</sub> (ring flip)	α <sup>+</sup> <sub>N<sub>δ2</sub></sub>	203.9	7.0
	α <sup>+</sup> <sub>N<sub>ε1</sub></sub>	196.9	

Within the accuracy limits of the methodology, it is clear that the calculations on the model systems support the breaking of the Asp<sub>102</sub> ··· H-N<sub>δ1</sub> H-bond and the consequent rotation of the imidazolium ring. However, while we cannot state with certainty if this rotation is complete or partial, the occurrence of a "ring flip"

\* Experimental Shieldings from Bachovchin,<sup>235</sup> +4.6 ppm correction to change ref. from 1M HNO<sub>3</sub> to CH<sub>3</sub>NO<sub>2</sub><sup>183</sup>

† Resting enzyme



in the low pH structures of DIFP and PMSF-inhibited proteases is supported by the calculations. Moreover, the NMR shieldings corresponding to the flipped TI-1<sub>B</sub> structure show the best agreement with the experimental references.

## Chapter 9

# Conclusions and Perspectives

The focus of this thesis is the development and implementation of a new SOS-DFPT-based methodology for the calculation of NMR shieldings. The definition of the Loc.3 approximation represents the primary result of this project. As we have seen in Ch. 2, Loc.3 has its roots in TDDFT and has been developed within the framework of the TDA. The subsequent introduction of the 2LM clearly demonstrates the connection between Loc.3 and the “classic” local corrections, i.e. Loc.1 and Loc.2. The fundamental result is that Loc.3 represents the missing physical basis for the SOS-DFPT approach. In fact, the *ad hoc* correction has been viewed by many as a convenient semi-empirical technique to adjust the excitation energies, calculated as the difference of the KS excited and occupied MOs. Therefore, the substantiation of the physical integrity of SOS-DFPT opens the door for the development of other types of  $\Delta E_{xc}$  corrections that can increase the level of accuracy of the NMR calculations.

The reliability of Loc.3 as an independently derived approximation was another crucial step. At first, its capability in determining the NMR absolute shield-

ings has been assessed against the standard UKS, Loc.1 and Loc.2 approaches. The nuclei tested are  $^{13}\text{C}$ ,  $^{14,15}\text{N}$  and  $^{17}\text{O}$ . This choice was based on the availability of reliable experimental data and on the width of the NMR scale of each one of the nuclei. For instance, tests on  $^1\text{H}$  are not indicative of the quality of a method, just because the NMR scale spans only 10-20 ppm; hence the relative errors are too small. The NMR calculations have been performed with three types of functional: the LDA (i.e. S exchange and VWN correlation), the asymptotically corrected LDA (ACLDA) and the PW91-PW91 GGA. The use of a GGA for the evaluation of the KS orbitals and of their eigenvalues does not significantly influence the quality of the SOS-DFPT results. This is in agreement with the observation that the quality of NMR shielding calculations depends mainly on how accurately the energy of the valence states is evaluated. Indeed, according to several TDDFT studies this level of accuracy is similar for LDAs, GGAs and for hybrid functionals.<sup>180-182</sup> Moreover, the role of the AC is found to be negligible, in agreement with the most recent studies in the literature.<sup>60, 157, 172, 179</sup>

The results obtained for the  $^{13}\text{C}$  shieldings show that the local corrections play a significant role only in the case of unsaturated nuclei. Further, their contribution becomes crucial when dealing with double or triple bonds. Loc.3 performs particularly well in the case of cyanides and isocyanides. Considerable problems are encountered in the case of carbonyl groups. Loc.3 fails completely in the description of CO, predicting a  $^{13}\text{C}$  signal more deshielded than the one obtained with UKS, probably due to the breakdown of the 2LM. Preliminary tests performed with a  $\Delta E_{xc}$  correction defined within the TDA only (i.e. CIS) predict for the  $^{13}\text{C}$  shielding of CO the same level of accuracy obtained with the Loc.1 and Loc.2.<sup>275</sup> The implementation of the full TDA approach has not been completed yet, how-

ever, its computational cost exceeds the requirements for routine applications. The results obtained for  $^{17}\text{O}$  are affected by large errors. Nonetheless Loc.3 shows a significant improvement for some difficult systems, such as  $\text{OF}_2$  and  $\text{H}_2\text{O}_2$ . The real success for Loc.3 is represented by the nitrogen nucleus. Difficult cases such as  $\text{CH}_2\text{NN}$ ,  $\text{HNC}$  and  $\text{HCN}$  are predicted much more accurately than with any other SOS-DFPT approximation.

These results are generally confirmed in the second set of tests, where Loc.3 is compared to the most recent and promising DFT-based methodologies for NMR calculations. For  $^{13}\text{C}$  shieldings Loc.3 represents the least accurate approach. The most difficult cases include  $^{13}\text{C}$  nuclei in carbonyl groups, in  $\text{CO}$  and in  $\text{CF}_4$ . Although Loc.3 shows a small improvement when compared to other SOS-DFPT approximations, its margins of error, evaluated as absolute deviations, are always considerably larger than those predicted by all the functionals analyzed here. The situation changes slightly for the  $^{17}\text{O}$  shielding. It is known that the NMR of this nucleus is fairly difficult to describe correctly. In fact, from the statistical analysis shown in Tab. 6.XII on pag. 136, we notice a general decline of the quality of the results. The Loc.3 performs rather competitively here. Its performance shows the same level of accuracy as GIAO-PBE0. The results obtained for the  $^{14,15}\text{N}$  nucleus are even more positive. Here Loc.3 proves not only to represent a significant improvement over the other SOS-DFPT approaches, but also to be very competitive compared to the more sophisticated functionals tested.

The considerable improvement of the accuracy of the  $^{14,15}\text{N}$  shielding obtained on the absolute scale represents the basis from which we undertook a more challenging step. The quality of the results obtained in comparison to high level *ab initio* and experimental data on the absolute scale is equivalent if we switch to

a relative scale, i.e. a chemical shift scale\*. The straight majority of the experimental data available are expressed as chemical shifts. Hence, access to these data allows us to test Loc.3 for a much wider range of nitrogen nuclei. This approach is not only useful from a theoretical point of view, i.e. as a validation of the methodology. Indeed, the knowledge of the margin of error enables the use of the calculations as a reliable tool to assist in the interpretation of experimental spectra. Proper evaluation of the margins of error requires consideration of several effects deriving from the fact that the experimental conditions are definitely not represented by a zero pressure and zero temperature state. Among the most significant effects we have the gas-to-liquid shift, the ro-vibrational correction and the solvent-induced shift. The latter can be considered the most important for  $^{14,15}\text{N}$  NMR. The approaches we could use to account for the presence of the solvent consist of the continuum model, or the inclusion of explicit solvent molecules. The use of continuum models is not very successful for NMR shielding calculations, since it does not account for direct solvent-solute interaction. A supermolecular approach, on the other hand, is not feasible from the point of view of computational cost. We have performed only one test in which we monitored the variation of the pyridine  $^{14,15}\text{N}$  shielding as a function of the number of explicit water molecules added to the system. The results show that over 20 molecules of water are needed to reasonably approximate the experimental shielding value (see Sec. 7.1). Our approach is based on the Kamlet-Taft or solvatochromic theory, according to which solvents with similar characteristics induce comparable shifts to the NMR nucleus shielding of the solute. As it has been demonstrated by several experimental studies,<sup>193</sup> the opposite argument is also true, meaning that

\*or chemical shielding scale, in function of the sign of the standard reference (see Ch. 7).

a particular solvent induces comparable shifts on the NMR shielding of similar nuclei. In other words, if we analyze the  $^{14,15}\text{N}$  shieldings within separate groups of similar nitrogen-types, we can consider the effect of the solvent as an additive constant. In order to minimize the solvent effect we compared the calculated results to experimental references determined in aprotic solvents with the lowest possible dielectric constant. This choice was limited by the availability of the reference data.

The groups of nitrogen atoms that we tested cover a rather large portion of the  $^{14,15}\text{N}$  NMR spectrum (see Fig. 7.1 on Ch. 7), including amines, hydrazines, cyanides and isocyanides, amides, ureas and guanidines, azoles, azines, azine N-oxides and finally nitrates. The highest accuracy has been obtained for  $\text{sp}^3$ -hybrid nitrogens, especially when bonded to H atoms, as in primary amines. The addition of alkyl groups or of other amine groups to form, respectively, secondary (tertiary) amines and hydrazines, increases the average error by  $\sim 10$  ppm. This behavior is likely due to the fact that the calculation of the paramagnetic term for alkyl-amines is the result of the sum of equivalent contributions but with opposite signs.<sup>157</sup> This requires a very accurate evaluation of all the excitation energies. For primary amines this cancellation has been proven to be not so effective.

As for  $\text{sp}^2$ -hybrid nuclei, the results for amides and ureas are rather positive, despite the increased polarity compared to amines. The average deviation is of the order of  $\sim 10$  ppm. Much larger errors are shown by guanidines and by cyclic amides, or lactams. These difficult cases can be explained by simple chemistry. Guanidines are among the strongest bases known, hence the interaction with the solvent must play a significant role. On the other hand, lactams are involved in a tautomeric equilibrium with their enol counterparts. Therefore, although the equi-

librium is shifted towards the lactam, the experimental reference can be influenced by the presence of the other tautomer.

The effect of adding local corrections to UKS is very small for  $sp^3$  nitrogens and only slightly larger for amides, ureas and guanidines. The nitrogen-type in these systems has in fact a partial double bond character and the local corrections are known to play a decisive role in the accuracy of the NMR shielding of unsaturated nuclei. A perfect example of the improvement of the results that can be achieved with the local corrections is given by the  $sp$ -hybrid nitrogens of cyanides and isocyanides. Loc.3 here increases significantly the level of accuracy, especially for nitriles (i.e. alkyl-cyanides) and for isocyanides, for which the calculated absolute average deviation is only 2.6 ppm. This confirms the findings of the previous tests on the absolute scale. However, this set also includes aryl-cyanides, which represent one of the most difficult systems among the ones we studied. The best estimate of the NMR shielding of aryl-cyanides is obtained with Loc.3, which predicts deviations of the order of 30 ppm. The magnitude of this error suggests that the problem does not lie exclusively in the solvent effect. The major contribution in fact comes from the fundamental difficulty of SOS-DFPT in describing highly delocalized systems. The presence of an external magnetic field causes the formation of paramagnetic ring currents in aromatic systems and their non-local character can hardly be described by SOS-DFPT.

The pyridine-type nitrogen in azoles, azines and in azine N-oxide is also largely influenced by the ring current effect. Their average absolute errors are, respectively, 22.5, 30.6 and 39.4 ppm. However, the pyrrole-type nitrogen is less affected and its average absolute deviation is 16.8 ppm. Loc.3 plays a decisive role in reducing the error in the approximation of the NMR shielding of these systems,

especially in the case of rings with more than one nitrogen nucleus.

The largest difference in the performance of the local corrections can be seen for nitrates. Here again, Loc.3 shows the best agreement to the experimental references, with an average deviation of 7.3 ppm.

The extensive testing both on the absolute and on the relative NMR scale has proved that for unsaturated  $^{14,15}\text{N}$  nuclei, and especially for the most difficult ones, Loc.3 represents a clear improvement over the “classic” local approximations. Nonetheless, some cases such as nitrogen nuclei in highly delocalized systems, remain out of the grasp of SOS-DFPT. In the absence of a current-dependent functional, the determination of the excitation energies through a non-local potential (e.g. OEP) could reduce the divergence of the calculation results.

With the Loc.3 approximation it is reasonable, however, to approach the analysis of a system involving pyrrole-type and protonated pyrrole-type nitrogen nuclei, also known as  $\alpha$  and  $\alpha^+$ -nitrogens, respectively. The study, presented as the last part of this thesis, pertains to the catalytic mechanism of the serine proteases. In Ch. 8 the two most recent versions of the reaction mechanism have been described in detail. Their comparison indicates that structural differences occur only in the intermediate steps of the reaction, i.e. transition states and tetrahedral intermediates (i.e. TI-1 and TI-2). These states are so unstable that the corresponding structures have never been isolated or characterized experimentally. The existence of tetrahedral intermediates and their structural features have been derived on the basis of solution chemistry. Moreover, a class of inhibitors binds the serine proteases in a TI-like structure, which has been studied extensively by X-ray diffraction and by NMR. The study presented here is based on the comparison between these experimental  $^{14,15}\text{N}$  NMR data and the results obtained from Loc.3



calculations on suitable model systems. The SOS-DFPT analysis is able to confirm that the formation of the tetrahedral intermediate occurs with the break of the Asp<sub>102</sub>-O<sub>δ2</sub> ··· H-N<sub>δ1</sub>-His<sub>57</sub> H-bond. This event most probably triggers the rotation of the imidazolium ring. We tested Loc.3 in the case of two different degrees of rotation, a complete rotation of 180° and a small rotation of the order of ~20°. Due to the margin of error established for the methodology and the simplicity of the model systems, we cannot state with certainty if the formation of a tetrahedral intermediate involves a complete or partial rotation of the imidazolium ring. Both sets of results are in agreement with the experimental reference. The most important conclusion of this analysis is that the “ring flip” mechanism is supported by the calculations, and the corresponding Loc.3 <sup>14,15</sup>N shieldings show the best agreement with the experimental data.

Looking to the future, a step further in the study of the catalytic mechanism of serine protease could be made by including dynamics in the calculation. Through an *ab initio* or DFT-MD approach it could be possible to study quantitatively the effect of the breaking of the H-bond between Asp<sub>102</sub> and His<sub>57</sub> and the occurrence of the ring flip. The shortcoming of these methods is that they can be applied only to reasonably small systems. This means that a large part of the enzyme would have to be discarded. The inclusion of important side chains, such as the ones forming the binding site for the substrate, can be extremely important in order to see how the catalysis is initiated. It is known in fact that residues of the binding site are connected through an H-bond network directly to the catalytic triad.<sup>220</sup> MD calculations on the whole enzyme, based on empirical force fields, have failed to recognize a complete rotation of the imidazolium ring during the formation of a tetrahedral intermediate.<sup>239</sup> However, according to the authors, the

occurrence of a full rotation can require a longer simulation timescale, hence it cannot be excluded in principle.

Snapshots of the trajectory could be used to construct new model systems for NMR calculations. The “classic” local corrections of SOS-DFPT, and in particular Loc.2, offer a very good compromise between accuracy and computational cost. As we have seen, Loc.3 shows a clear improvement over Loc.1 and Loc.2, however, its use is significantly more expensive. The bottleneck of the Loc.3 calculation is represented by the solution of the exchange-correlation kernel integral. The implementation of this formula in the TD module was never intended for large basis sets such as the ones required for NMR shielding calculations. Therefore, a necessary step to make Loc.3 a more practical tool for routine application on large systems is a more efficient programming of the integrals.

From the results of the benchmark tests we gathered that the accuracy of the calculations depends greatly on the atom type and on its chemical environment. While for saturated atoms the accuracy of the results rests mainly in the quality of the approximate exchange-correlation functional used, for unsaturated and delocalized systems, however, the introduction of  $\Delta E_{xc}$  corrections to the excitation energy becomes necessary. Here we have seen that the use of more sophisticated GGAs and OEPs, instead of the LDA, is crucial. Nonetheless, for the most difficult systems, an increase in accuracy can be achieved introducing a Loc.3-type correction to the GGA or OEP excitation energy.

The use of functionals other than the ALDA for the approximation of the kernel would not influence the accuracy of the NMR results. The future development of SOS-DFPT should be addressed to the design of new exchange-correlation functionals, focused on the correct description of the valence states, a crucial as-

pect for the accuracy of NMR shielding calculations. Such functionals would eliminate the necessity of a  $\Delta E_{xc}$  correction, reducing considerably the computational cost, and opening the way to the use of SOS-DFPT on large molecular systems.

## Bibliography

- [1] P.B. Karadakov and K. Morokuma, *Chem. Phys. Lett.*, **317**, 589 (2000)
- [2] P.A. Molina, R.S. Sikorski and J.H. Jensen, *Theor. Chem. Acc.*, **109**, 100 (2003)
- [3] P.A. Molina and J.H. Jensen, *J. Phys. Chem. B*, **107**, 6226 (2003)
- [4] V.G. Malkin, O.L. Malkina, M.E. Casida and D.R. Salahub, *J. Am. Chem. Soc.*, **116**, 5898 (1994)
- [5] E.K. Ash, J.L. Sudmeier, R.M. Day, M. Vincent, E.V. Torchilin, K. Coffman Haddad, E.M. Bradshaw, D.G. Sanford and W.W. Bachovchin, *Proc. Natl. Acad. Sci. USA*, **97**, 10371 (2000)
- [6] W. Koch and M.C. Holthausen, *A Chemist's Guide to Density Functional Theory*, Wiley-VCH Verlag GmbH, Weinheim, Germany (2000)
- [7] R.G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*; Oxford Science Publications, New York, (1989)
- [8] P. Hohenberg and W. Kohn, *Phys. Rev.*, **136**, B864 (1964)
- [9] M. Levy, *Proc. Natl. Acad. Sci. USA*, **76**, 6062 (1979)

- [10] M. Levy, *Phys. Rev. A*, **26**, 1200 (1982)
- [11] E.H. Lieb, in *Physics as Natural Philosophy*, p.111, A. Shimony and H. Feshbach Editors, MIT, Cambridge (1982)
- [12] M. Levy and J.P. Perdew, The Constrained Search Formulation in Density-Functional Theory, in *Density-Functional Methods in Physics*, p 11, R.M. Dreizler and J. da Providencia Editors, Plenum, New York (1985)
- [13] W. Kohn and L.J. Sham, *Phys. Rev.*, **140**, A1133 (1965)
- [14] C.-O. Almbladh and A.C. Pedroza, *Phys. Rev. A*, **29**, 2322 (1984)
- [15] A. Savin, H. Stoll and H. Preuss, *Theor. Chim. Acta*, **70**, 407 (1986)
- [16] R.C. Morrison and Q. Zhao, *Phys. Rev. A*, **51**, 1980 (1995)
- [17] P. Süle, *Chem. Phys. Lett.*, **259**, 69 (1996)
- [18] O.V. Gritsenko, R. van Leeuwen and E.J. Baerends, *J. Chem. Phys.*, **104**, 8535 (1996)
- [19] E. Fermi, *Rend. Accad. Lincei*, **6**, 602 (1927)
- [20] L.H. Thomas, *Proc. Camb. Phil. Soc.*, **23**, 542 (1927); Reprinted in March 1975
- [21] J.A. Pople, M. Head-Gordon, D.J. Fox, K. Raghavachari and L.A. Curtiss, *J. Chem. Phys.*, **90**, 5622 (1989)
- [22] L.A. Curtiss, C. Jones, G.W. Trucks, K. Raghavachari and J.A. Pople, *J. Chem. Phys.*, **93**, 2537 (1990)

- [23] L.A. Curtiss, K. Raghavachari, G.W. Trucks and J.A. Pople, *J. Chem. Phys.*, **94**, 7221 (1991)
- [24] L.A. Curtiss, K. Raghavachari, P.C. Redfern and J.A. Pople, *J. Chem. Phys.*, **109**, 42 (1998)
- [25] J.P. Perdew and K. Burke, *Int. J. Quantum Chem.*, **57**, 309 (1996)
- [26] C. Adamo, A. di Matteo and V. Barone, *Adv. Quantum Chem.*, **36**, 45 (1999)
- [27] J. Harris and J.E. Jones, *J. Phys. F*, **4**, 1170 (1974)
- [28] O. Gunnarsson and B.I. Lundqvist, *Phys. Rev. B*, **13**, 4274 (1976)
- [29] D.C. Langreth and J.P. Perdew, *Phys. Rev. B*, **15**, 2884 (1977)
- [30] R.A. Harris, *J. Chem. Phys.*, **81**, 2403 (1984)
- [31] A.D. Becke, *J. Chem. Phys.*, **88**, 1053 (1988)
- [32] A.D. Becke, Exchange-Correlation Approximations in Density-Functional Theory, in *Modern Electronic Structure Theory, Part II*, D.R. Yarkony Editor, World Scientific, Singapore (1995)
- [33] R.O. Jones and O. Gunnarsson, *Rev. Mod. Phys.*, **61**, 689 (1989)
- [34] F. Bloch, *Z. Physik*, **57**, 545 (1929)
- [35] P.A.M. Dirac, *Proc. Camb. Phil. Soc.*, **26**, 376 (1930)
- [36] S.J. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.*, **58**, 1200 (1980)
- [37] D.M. Ceperley and B.J. Alder, *Phys. Rev. Lett.*, **45**, 566 (1980)

- [38] B.G. Johnson, P.M.W. Gill and J.A. Pople, *J. Chem. Phys.*, **98**, 5612 (1993)
- [39] B.Y. Tong and L.J. Sham, *Phys. Rev.*, **144**, 1 (1966)
- [40] R.O. Jones and O. Gunnarsson, *Phys. Rev. Lett.*, **55**, 107 (1985)
- [41] O. Gunnarsson and R.O. Jones, *Phys. Rev. B*, **31**, 7588 (1985)
- [42] J.P. Perdew, *Phys. Rev. Lett.*, **55**, 1665 (1985)
- [43] Y. Wang and A.W. Overhauser, *Phys. Rev. B*, **34**, 6839 (1986)
- [44] M. Ernzerhof, J.P. Perdew and K. Burke, *Int. J. Quantum Chem.*, **64**, 285 (1997)
- [45] A.D. Becke, *Phys. Rev. A*, **38**, 3098 (1988)
- [46] J.P. Perdew, Unified Theory of Exchange and Correlation Beyond the Local Density Approximation, in *Electronic Structure of Solids*, P. Ziesche and H. Eschrig Editors, Akademie Verlag, Berlin (1991)
- [47] K. Burke, J.P. Perdew and Y. Wang, Derivation of a Generalized Gradient Approximation: The PW91 Density Functional, in *Electronic Density Functional Theory. Recent Progress and New Directions*, J.F. Dobson, G. Vignale and M.P. Das Editors, Plenum Press, New York (1998)
- [48] F.A. Hamprecht, A.J. Cohen, D.J. Tozer and N.C. Handy, *J. Chem. Phys.*, **109**, 6264 (1998)
- [49] A.D. Becke, *J. Chem. Phys.*, **84**, 4524 (1986)
- [50] J.P. Perdew and W. Yue, *Phys. Rev. B*, **33**, 8800 (1986)

- [51] J.P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, **77**, 3865 (1996);  
Erratum, *Phys. Rev. Lett.*, **78**, 1396 (1997)
- [52] J.P. Perdew, *Phys. Rev. B*, **33**, 8822 (1986)
- [53] J.P. Perdew and Y. Wang, *Phys. Rev. B*, **45**, 13244 (1992)
- [54] C. Lee, W. Yang and R.G. Parr, *Phys. Rev. B*, **37**, 785 (1988)
- [55] A.D. Becke, *J. Chem. Phys.*, **98**, 1372 (1993)
- [56] A.D. Becke, *J. Chem. Phys.*, **98**, 5648 (1993)
- [57] L.A. Curtiss, K. Raghavachari, P.C. Redfern and J.A. Pople, *J. Chem. Phys.*, **106**, 1063 (1997)
- [58] P.J. Stevens, J.F. Devlin, C.F. Chabalowski and M.J. Frisch, *J. Phys. Chem.*, **98**, 11623 (1994)
- [59] K. Raghavachari, *Theor. Chem. Acc.*, **103**, 361 (2000)
- [60] P.J. Wilson, R.D. Amos and N.C. Handy, *Mol. Phys.*, **97**, 757 (1999)
- [61] J.P. Perdew, M. Ernzerhof and K. Burke, *J. Chem. Phys.*, **105**, 9982 (1996)
- [62] M. Ernzerhof, *Chem. Phys. Lett.*, **263**, 499 (1996)
- [63] C. Adamo and V. Barone, *J. Chem. Phys.*, **110**, 6158 (1999)
- [64] C. Adamo, G. E. Scuseria and V. Barone, *J. Chem. Phys.*, **111**, 2889 (1999)
- [65] C. Adamo, M. Cossi, G. Scalmani and V. Barone, *Chem. Phys. Lett.*, **307**, 265 (1999)



- [66] C. Adamo and V. Barone, *Chem. Phys. Lett.*, **298**, 113 (1998)
- [67] C. Adamo, M. Cossi and V. Barone, *J. Mol. Struct. (Theochem)*, **493**, 145 (1999)
- [68] A.D. Becke and M.R. Roussel, *Phys. Rev. A*, **39**, 3761 (1998)
- [69] A.D. Becke, *J. Chem. Phys.*, **109**, 2092 (1998)
- [70] J.P. Perdew, S. Kurth, A. Zupan and P. Blaha, *Phys. Rev. Lett.*, **82**, 2544 (1999)
- [71] T.V. Voorhis and G.E. Scuseria, *Mol. Phys.*, **92**, 601 (1997)
- [72] T.V. Voorhis and G.E. Scuseria, *J. Chem. Phys.*, **109**, 400 (1998)
- [73] M. Filatov and W. Thiel, *Phys. Rev. A*, **57**, 189 (1998)
- [74] A.D. Becke, *J. Chem. Phys.*, **112**, 4020 (2000)
- [75] E.I. Proynov, A. Vela and D.R. Salahub, *Chem. Phys. Lett.*, **230**, 419 (1994)
- [76] E.I. Proynov, A. Vela and D.R. Salahub, *Chem. Phys. Lett.*, **234**, 462E (1995)
- [77] E.I. Proynov and D.R. Salahub, *Phys. Rev. B*, **49**, 7874 (1994)
- [78] E.I. Proynov and D.R. Salahub, *Phys. Rev. B*, **57**, 12616E (1998)
- [79] E.I. Proynov, S. Sirois and D.R. Salahub, *Int. J. Quantum Chem*, **64**, 427 (1997)
- [80] E.I. Proynov, H. Chermette and D.R. Salahub, *J. Chem. Phys.*, **113**, 10013 (2000)

- [81] R.T. Sharp and G.K. Horton, *Phys. Rev.*, **30**, 317 (1953)
- [82] J.D. Talman and W.F. Shadwick, *Phys. Rev. A*, **14**, 36 (1976)
- [83] G.B. Krieger, Y. Li and G.J. Iafrate, *Phys. Rev. A*, **45**, 101 (1992)
- [84] G.B. Krieger, Y. Li and G.J. Iafrate, *Phys. Rev. A*, **46**, 5453 (1992)
- [85] J.C. Slater, *Phys. Rev.*, **81**, 385 (1951)
- [86] M.E. Casida, C. Jamorski, K.C. Casida and D.R. Salahub, *J. Chem. Phys.*, **108**, 4439 (1998)
- [87] M.E. Casida and D.R. Salahub, *J. Chem. Phys.*, **113**, 8918 (2000)
- [88] C.O. Almbladh and U. von Barth, *Phys. Rev. B*, **56**, 16021 (1997)
- [89] D.J. Tozer and N.C. Handy, *J. Chem. Phys.*, **109**, 10180 (1998)
- [90] J.P. Perdew, R.G. Parr, M. Levy and J.L. Balduz Jr, *Phys. Rev. Lett.*, **49**, 1691 (1982)
- [91] R. van Leeuwen and E.J. Baerends, *Phys. Rev. A*, **49**, 2421 (1994)
- [92] R. Neumann, R.H. Nobes and N.C. Handy, *Mol. Phys.*, **87**, 1 (1996)
- [93] V. Peuckert, *J. Phys.*, **C11**, 4945 (1978)
- [94] E.K.U. Gross and W. Kohn, *Adv. Quant. Chem.*, **21**, 255 (1990)
- [95] E.K.U. Gross, C.A. Ullrich and U.J. Grossman, *Density Functional Theory*, E.K.U. Gross and R.M. Dreizler Editors, NATO ASI Series B337 (1995)

- [96] E.K.U. Gross, J.F. Dobson and M. Petersilka, *Density Functional Theory*, R.F. Nalewajski Editor, Springer (1996)
- [97] G.D. Mahan and K.R. Subbaswamy, *Local Density Theory of Polarizability*, Plenum Press, New York (1990)
- [98] P. Mlynarski and D.R. Salahub, *Phys. Rev. B*, **43**, 1399 (1990)
- [99] M.E. Casida, Time-Dependent Density-Functional Response Theory for Molecules, in *Recent Advances in Density Functional Methods, Part I*, D.P. Chong Editor, World Scientific, Singapore (1995)
- [100] C. Jamorski, M.E. Casida and D.R. Salahub, *J. Chem. Phys.*, **104**, 5134 (1996)
- [101] K. Burke and E.K.U. Gross, A Guided Tour of Time-Dependent Density-Functional Theory, in *Density Functionals: Theory and Applications*, D. Joubert Editor, Springer, Berlin (1998)
- [102] J.S. Parker, L.R. Moore, D. Dundas and K.T. Taylor, *J. Phys.*, **B33**, L691 (2000)
- [103] E. Runge and E.K.U. Gross, *Phys. Rev. Lett.*, **52**, 997 (1984)
- [104] R. van Leeuwen, *J. Modern Phys. B*, **15**, 1969 (2001)
- [105] R. van Leeuwen, *Phys. Rev. Lett.*, **80**, 1280 (1998)
- [106] T. Ando, *Z. Phys. B*, **26**, 263 (1977)
- [107] T. Ando, *Solid State Commun.*, **21**, 133 (1977)

- [108] H. Lehmann, *Nuovo Cimento*, **11**, 432 (1954)
- [109] M. Petersilka, U.J. Gossmann and E.K.U. Gross, *Phys. Rev. Lett.*, **76**, 1212 (1996)
- [110] M. Jaszuński, A. Rizzo and D.L. Yeager, *Chem. Phys.*, **136**, 385 (1989), and references therein.
- [111] S. Hirata and M. Head-Gordon, *Chem. Phys. Lett.*, **314**, 291 (1999)
- [112] A.L. Fetter and J.D. Walecka, *Quantum Theory of Many-Particle Systems*, Mc-Graw-Hill, New York (1971)
- [113] M.E. Casida, F. Gutierrez, J. Guan, F.-X. Gadea, D.R. Salahub and J.-P. Daudey, *J. Chem. Phys.*, **113**, 7062 (2000)
- [114] R. McWeeny, *Methods of Molecular Quantum Mechanics*, 2nd edn.; Academic Press, London, (1989)
- [115] W. Kutzelnigg, *J. Mol. Struct. (THEOCHEM)*, **202**, 11 (1989)
- [116] T. Helgaker, M. Jaszuński and K. Ruud, *Chem. Rev.*, **99**, 293 (1999)
- [117] V.G. Malkin, O.L. Malkina, L.A. Eriksson and D.R. Salahub, The Calculation of NMR and EPR Spectroscopy Parameters Using Density Functional Theory, in *Modern Density Functional Theory: A Tool for Chemistry*, Theoretical and Computational Chemistry, Vol.2, p.273, J.M. Seminario and P. Politzer Editors, Elsevier Science B.V., (1985)
- [118] I. Carmichael, *J. Phys. Chem.*, **97**, 1789 (1993)

- [119] O.L. Malkina, D.R. Salahub and V.G. Malkin, *J. Chem. Phys.*, **105**, 8793 (1996)
- [120] V.G. Malkin, O.L. Malkina and D.R. Salahub, *Chem. Phys. Lett.*, **221**, 91 (1994)
- [121] A. Szabo and N.S. Ostlund, *Modern Quantum Chemistry. Introduction to Advanced Electronic Structure Theory*, Dover Publication Inc., Mineola, New York (1996)
- [122] W. Kutzelnigg, U. Fleischer and M. Schindler, The IGLO-Method: Ab-initio Calculations and Interpretation of NMR Chemical Shifts and Magnetic Susceptibilities, in *NMR Basic Principles and Progress*, Vol.13, p.165, Springer-Verlag, Heidelberg (1990)
- [123] J. Olsen and P. Jørgensen, *J. Chem. Phys.*, **82**, 3235 (1985)
- [124] T. Helgaker and J. Almlöf, *Int. J. Quantum Chem.*, **26**, 275 (1984)
- [125] K. Ruud and T. Helgaker, *Chem. Phys. Lett.*, **264**, 17 (1997)
- [126] C. van Wüllen and W. Kutzelnigg, *J. Chem. Phys.*, **104**, 2330 (1996)
- [127] F. London, *J. Phys. Radium*, **8**, 397 (1937)
- [128] R. Ditchfield, *Mol. Phys.*, **27**, 789 (1974)
- [129] K. Wolinski, J.F. Hinton, and P. Pulay, *J. Am. Chem. Soc.*, **112**, 8251 (1990)
- [130] G. Schreckenbach and T. Ziegler, *J. Phys. Chem.*, **99**, 606 (1995)
- [131] G. Schreckenbach and T. Ziegler, *Int. J. Quantum Chem.*, **60**, 753 (1996)

- [132] S. Foster and S.F. Boys, *Rev. Mod. Phys.* **32**, 296, 303, 305 (1960)
- [133] W. Kutzelnigg, *Isr. J. Chem.*, **19**, 193 (1980)
- [134] M. Schindler and W. Kutzelnigg, *J. Chem. Phys.*, **76**, 1919 (1982)
- [135] A.E. Hansen and T.D. Bouman, *J. Chem. Phys.*, **82**, 5035 (1985)
- [136] A.E. Hansen and T.D. Bouman, in *Nuclear Magnetic Shielding*, J.A. Tossel Editor, Kluwer, Dordrecht (1993), p.117
- [137] A.E. Hansen and M. Bilde, Shielding Calculations: LORG and SOLO Approaches, in *Encyclopedia of NMR*, Vol. 7, 4292 (1996)
- [138] J. Gauss and J.F. Stanton, *J. Chem. Phys.*, **102**, 251 (1995)
- [139] J. Gauss and J.F. Stanton, *J. Chem. Phys.*, **103**, 3561 (1995)
- [140] S.M. Cybulski and D.M. Bishop, *Chem. Phys. Lett.*, **250**, 471 (1996)
- [141] S.M. Cybulski and D.M. Bishop, *J. Chem. Phys.*, **106**, 4082 (1997)
- [142] J. Gauss and J.F. Stanton, *J. Chem. Phys.*, **104**, 2574 (1996)
- [143] M. Kollwitz and J. Gauss, *Chem. Phys. Lett.*, **260**, 639 (1996)
- [144] M. Kollwitz, M. Häser and J. Gauss, *J. Chem. Phys.*, **108**, 8295 (1998)
- [145] J. Gauss, *Chem. Phys. Lett.*, **229**, 198 (1994)
- [146] V.G. Malkin, O.L. Malkina and D.R. Salahub, *Chem. Phys. Lett.*, **204**, 80 (1993)
- [147] K. Friedrich, G. Seifert and G. Grossmann, *Z. Phys. D*, **17**, 45 (1990)

- [148] W. Bieger, G. Seifert, H. Eschrig and G. Grossmann, *Chem. Phys. Lett.*, **204**, 80 (1993)
- [149] C. van Wüllen, *J. Chem. Phys.*, **102**, 2806 (1995)
- [150] A.M. Lee, N.C. Handy and S.M. Colwell, *J. Chem. Phys.*, **103**, 10095 (1995)
- [151] G. Vignale and M. Rasolt, *Phys. Rev. B*, **38**, 10685 (1988)
- [152] G. Vignale, M. Rasolt and D.J.W. Geldart, *Adv. Quantum Chem.*, **21**, 235 (1990)
- [153] A.D. Becke, *Can. J. Chem.*, **74**, 995 (1996)
- [154] P.J. Wilson, R.D. Amos and N.C. Handy, *Chem. Phys. Lett.*, **312**, 475 (1999)
- [155] P.J. Wilson and D.J. Tozer, *Chem. Phys. Lett.*, **337**, 341 (2001)
- [156] S. Patchkovskii, J. Autschbach and T. Ziegler, *J. Chem. Phys.*, **115**, 26 (2001)
- [157] E. Fadda, M.E. Casida and D.R. Salahub, *Int. J. Quantum Chem.*, **91**, 67 (2003)
- [158] C.C.J. Roothaan, *Rev. Mod. Phys.*, **23**, 69 (1951)
- [159] A. St-Amant and D.R. Salahub, *Chem. Phys. Lett.*, **169**, 387 (1990)
- [160] A. St-Amant, Ph.D. Thesis, Universit de Montral (1992)

- [161] *deMon-KS* version 3.5, M.E. Casida; C. Daul; A. Goursot; A. Koester; Pettersson L.G.M.; Proynov E.; St-Amant A.; Salahub D.R.; Chretien S.; Duarte H.; Godbout N.; Guan J.; Jamorski C.; Leboeuf M.; Malkin V.G.; Malkina O.L.; Nyberg M.; Pedocchi L.; Sim F.; Vela A. *deMon Software* **1998**
- [162] H. Sambe and R.H. Felton, *J. Chem. Phys.*, **72**, 1122 (1975)
- [163] B.I. Dunlap, J.W.D. Connolly and J.R. Sabin, *J. Chem Phys.*, **71**, 3396 (1979)
- [164] A.D. Becke, *J. Chem. Phys.*, **88**, 2547 (1988)
- [165] S. Obara and A. Saika, *J. Chem. Phys.*, **84**, 3963 (1986)
- [166] M.E. Casida, C. Jamorski, E. Fadda, J. Guan, S. Hamel, and D.R. Salahub, *deMon-DynaRho* version 3.2, University of Montreal and Universit Joseph-Fourier, copyright 2002
- [167] M.E. Casida, F. Gutierrez, J. Guan, F.-X. Gadea, D.R. Salahub and J.-P. Dauday, *J. Chem. Phys.*, **113**, 7062 (2000)
- [168] S. Sirois, E.I. Proynov, D.T. Nguyen and D.R. Salahub, *J. Chem. Phys.*, **107**, 6770 (1997)
- [169] H.A. Duarte, E.I. Proynov and D.R. Salahub, *J. Chem. Phys.*, **109**, 26 (1998)
- [170] see H.B. Schlegel, in *Ab Initio Methods in Quantum Chemistry I* K.P. Lawley Editor, Wiley, New York, 1987



- [171] N. Godbout, D.R. Salahub, J. Andzelm and E. Wimmer, *Can. J. Chem.*, **70**, 560 (1992)
- [172] E. Fadda, M.E. Casida and D.R. Salahub, *J. Chem. Phys.*, **118**, 6758 (2003)
- [173] J.R. Cheeseman, G.W. Trucks, T.A. Keith and M.J. Frisch, *J. Chem. Phys.*, **104**, 5497 (1996)
- [174] A.K. Jameson and C.J. Jameson, *Chem. Phys. Lett.* **134**, 461 (1987)
- [175] J. Gauss, *J. Chem. Phys.* **99**, 3629 (1993)
- [176] K.B. Wiberg, *J. Comput. Chem.*, **20**, 1299 (1999)
- [177] see data at page 202 of W. Koch and M.C. Holthausen, *A Chemist's Guide to Density Functional Theory*, Wiley-VCH Verlag GmbH, Weinheim, Germany (2000)
- [178] G. Magyarfalvi and P. Pulay, *J. Chem. Phys.*, **119**, 1350 (2003)
- [179] T.W. Keal and D.J. Tozer, *J. Chem. Phys.*, **119**, 3015 (2003)
- [180] R. Bauernschmitt and R. Ahlrichs, *Chem. Phys. Lett.*, **256**, 454 (1996)
- [181] M.E. Casida, K.C. Casida and D.R. Salahub, *Int. J. Quantum Chem.*, **70**, 933 (1998)
- [182] R.E. Stratmann, G.E. Scuseria and M.J. Frisch, *J. Chem. Phys.*, **109**, 8218 (1998)
- [183] C.J. Jameson, A.K. Jameson, D. Oppusunggu, S. Wille, P.M. Burrell and J. Mason, *J. Chem. Phys.*, **74**, 81 (1981)

- [184] A.D. Becke, *J. Chem. Phys.*, **107**, 8554 (1997)
- [185] H. Fukui, T. Baba, H. Matsuda and K. Miura, *J. Chem. Phys.*, **100**, 6608 (1994)
- [186] J. Tomasi and M. Persico, *Chem. Rev.*, **94**, 2027 (1994)
- [187] C.J. Cramer and D.J. Truhlar, *Chem. Rev.*, **99**, 2161 (1999)
- [188] M. Orozco and F.J. Luque, *Chem. Rev.*, **100**, 4187 (2000)
- [189] *Combined Quantum Mechanical and Molecular Mechanics Methods*, J. Gao and M.A. Thompson Editors; ACS Symposium Series 712, American Chemical Society, Washington D.C. (1998)
- [190] J. Tomasi and C.S. Pomelli in *Encyclopedia of Computational Chemistry*, P.v.R. Schleyer Editor, Wiley, New York, p.53 (1999)
- [191] M.J. Kamlet, J.L.M. Abboud and R.W. Taft in *Progress in Physical Organic Chemistry*, R.W. Taft Editor, **13**, 435 (1981)
- [192] M.N. Manalo, A.C. de Dios and R. Cammi, *J. Phys. Chem. A*, **104**, 9600 (2000)
- [193] M. Witanowski, L. Stefaniak and G.A. Webb, in *Annual Reports on NMR Spectroscopy* G.A. Webb Editor, Academic Press, London, **25**, 1 (1993)
- [194] M. Witanowski, L. Stefaniak and G.A. Webb, in *Annual Reports on NMR Spectroscopy* G.A. Webb Editor, Academic Press, London, **7**, 117 (1977)
- [195] D. Sundholm, J. Gauss, A. Schäfer, *J. Chem. Phys.*, **105**, 11051 (1996)

- [196] B. Crompt, T. Carrington Jr., D.R. Salahub, O.L. Malkina and V.G. Malkin, *J. Chem. Phys.*, **110**, 7153 (1999)
- [197] M. Witanowski, W. Sicinska and G.A. Webb, *Magn. Reson. Chem.*, **27**, 380 (1989)
- [198] M. Witanowski, Z. Biedrzycka, W. Sicinska and Z. Grabowski, *J. Magn. Reson.*, **131**, 54 (1998)
- [199] M. Witanowski, W. Sicinska, Z. Grabowski and G.A. Webb, *J. Magn. Reson. Series A*, **104**, 310 (1993)
- [200] C. Reichardt, *Solvents and Solvent Effects in Organic Chemistry*, VCH Publishers, 2nd Ed. (1988)
- [201] M. Schindler, *Magn. Reson. Chem.*, **26**, 394 (1988)
- [202] J.A. Elvidge and L.M. Jackman, *J. Chem. Soc.*, 859 (1961)
- [203] J. Jusélius and D. Sundholm, *Phys. Chem. Chem. Phys.*, **1**, 3429 (1999)
- [204] M. Jaszuński, K.V. Mikkelsen, A. Rizzo, M. Witanowski, *J. Phys. Chem. A*, **104**, 1466 (2000)
- [205] M. Witanowski, W. Sicinska, S. Biernat and G.A. Webb, *J. Magn. Reson.*, **91**, 289 (1991)
- [206] M. Witanowski, J. Sitkowski, S. Biernat, B. Kamiński, B.T. Hamdi and G.A. Webb, *Magn. Reson. Chem.*, **23**, 748 (1985)
- [207] N.L. Allinger, Y.H. Yuh, and J.H. Lii, *J. Amer. Chem. Soc.*, **111**, 8551 (1989)

- [208] J.H. Lii and N.L. Allinger, *J. Amer. Chem. Soc.*, **111**, 8566 (1989)
- [209] J.H. Lii and N.L. Allinger, *J. Amer. Chem. Soc.*, **111**, 8576 (1989)
- [210] N.L. Allinger, H.J. Geise, W. Pyckhout, L.A. Paquette and J.C. Gallucci, *J. Am. Chem. Soc.*, **111**, 1106 (1989)
- [211] N.L. Allinger, F. Li and L. Yan, *J. Comp. Chem.*, **11**, 848 (1990)
- [212] N.L. Allinger, F. Li, L. Yan and J.C. Tai, *J. Comp. Chem.*, **11**, 868 (1990)
- [213] J.H. Lii and N.L. Allinger, *J. Phys. Org. Chem.*, **7**, 591 (1994)
- [214] J.H. Lii and N.L. Allinger, *J. Comp. Chem.*, **19**, 1001 (1998)
- [215] <http://dasher.wustl.edu/tinker/>
- [216] E.R. Davidson and D. Feller, *Chem. Rev.*, **86**, 681 (1986)
- [217] U.C. Singh and P.A. Kollman, *J. Comput. Chem.* **5**, 129 (1984)
- [218] V.G. Malkin, O.L. Malkina, G. Steinebrunner and H. Huber, *Chem. Eur. J.*, **2**, 452 (1996)
- [219] G. Dodson and A. Wlodawr, *Trends Biochem. Sci.*, **23**, 347 (1998)
- [220] L. Hedstrom, *Chem. Rev.*, **102**, 4501 (2002)
- [221] D.M. Blow, *Trends Biochem. Sci.*, **22**, 405 (1997)
- [222] Z.S. Derewenda, U. Derewenda and P.M. Kobos, *J. Mol. Biol.*, **241**, 83 (1994)
- [223] D.M. Blow, J.J. Birktoft and B.S. Harley, *Nature (London)*, **221**, 337 (1969)

- [224] M.M. Dixon and B.W. Matthews, *Biochemistry*, **28**, 7033 (1989)
- [225] M.M. Dixon, R.G. Brennan and B.W. Matthews, *Int. J. Biol. Macromol.*, **13**, 89 (1991)
- [226] M. Harel, C.T. Su, F. Frolow, I. Silman and J.L. Sussman, *Biochemistry*, **30**, 5217 (1991)
- [227] P. Kuhn, M. Knapp, S.M. Soltis, G. Ganshaw, M. Thoene and R. Bott, *Biochemistry*, **37**, 13446 (1998)
- [228] R.D. Kidd, P. Sears, D.H. Huang, K. Witte, C.H. Wong and G.K. Farber, *Protein Sci.*, **8**, 410 (1999)
- [229] J.L. Markley, D.E. Neves, W.M. Westler, I.B. Ibanez, M.A. Porubcan and M.W. Baillargeon, *Dev. Biochem.*, **10**, 31 (1980)
- [230] W.W. Bachovchin, *Magn. Reson. Chem.*, **39**, S199 (2001)
- [231] J.L. Markley, *Biochemistry*, **17**, 4648 (1978)
- [232] S. Scheiner, *Hydrogen Bonding: A Theoretical Perspective*, Oxford University Press, New York (1997)
- [233] S. Scheiner, T. Kar and Y. Gu, *J. Biol. Chem.*, **276**, 9832 (2001)
- [234] C.E. Cannizzaro and K.N. Houk, *J. Am. Chem. Soc.*, **124**, 7163 (2002)
- [235] W.W. Bachovchin, *Biochemistry*, **25**, 7751 (1986)
- [236] S. Farr-Jones, W.Y.L. Wong, W.G. Gutheil and W.W. Bachovchin, *J. Am. Chem. Soc.*, **115**, 6813 (1993)

- [237] F.C. Lightstone, Y-J. Zheng and T.C. Bruice, *J. Am. Chem. Soc.*, **120**, 5611 (1998)
- [238] E.Y. Lau and T.C. Bruice, *Biophys. J.*, **77**, 85 (1999)
- [239] M. Topf, P. Várnai and G. Richards, *J. Am. Chem. Soc.*, **124**, 14780 (2002)
- [240] G. Katona, R.C. Wilmouth, P.A. Wright, G.I. Berglund, J. Hajdu, R. Neutze and C.J. Schofield, *J. Biol. Chem.*, **277**, 21962 (2002)
- [241] J. Blair West, W.J. Hennen, J.L. Lalonde, J.A. Bibbs, Z. Zhong, E.F. Meyer Jr. and C.H. Wong, *J. Am. Chem. Soc.*, **112**, 5313 (1990)
- [242] S. Ba-Saif, A.K. Luthra and A. Williams, *J. Am. Chem. Soc.*, **111**, 2647 (1989)
- [243] D. Stefanidis, S. Cho, S. Dhe-Paganon and W.P.J. Jencks, *J. Am. Chem. Soc.*, **115**, 1650 (1993)
- [244] A.C. Hengge and R.A. Hess, *J. Am. Chem. Soc.*, **116**, 11256 (1994)
- [245] D.J. Millen, A.C. Legon and O. Schrems, *J. Chem. Soc. Faraday Trans. 2*, **75**, 592 (1979)
- [246] A.C. Legon, D.J. Millen and S.C. Rogers, *Proc. R. Soc. London Ser. A*, **370**, 213 (1980)
- [247] F. Hibbert and J. Emsley, *Adv. Phys. Org. Chem.*, **26**, 255 (1990)
- [248] P.A. Frey, S.A. Whitt and J.B. Tobin, *Science*, **264**, 1927 (1994)
- [249] W.W. Cleland and M.M. Kreevoy, *Science*, **264**, 1887 (1994)

- [250] J.A. Gerlt and P.G. Gassman, *Biochemistry*, **32**, 11943 (1993)
- [251] J.A. Gerlt, M.M. Kreevoy, W. Cleland and P.A. Frey, *Chem. Biol.*, **4**, 259 (1997)
- [252] E.L. Ash, J.L. Sudmeier, E.C. De Fabo and W.W. Bachovchin, *Science*, **278**, 1128 (1997)
- [253] C.J. Halkides, Y.Q. Wu and C.J. Murray, *Biochemistry*, **35**, 15941 (1996)
- [254] A. Warshel, G. Naray-Szabo, F. Sussman and J.K. Hwang, *Biochemistry*, **28**, 3629 (1989)
- [255] J.R. Stratton, J.G. Pelton and J.F. Kirsh, *Biochemistry*, **40**, 10411 (2001)
- [256] A. Warshel, A. Papazyan and P.A. Kollman, *Science*, **269**, 102 (1995)
- [257] J.H. Wang, *Proc. Natl. Acad. Sci. USA*, **66**, 874 (1970)
- [258] L. Polgar, *Acta Biochim. Biophys.*, **7**, 29 (1972)
- [259] A.C. Satterthwait and W.P. Jenks, *J. Am. Chem. Soc.*, **96**, 7018 (1974)
- [260] P.J.F. Henderson, *Biochem. J.*, **124**, 13 (1971)
- [261] L.D. Byers and D.E. Koshland, *Bioorg. Chem.*, **7**, 15 (1978)
- [262] J.D. Scholten, J.L. Hogg and F.M. Raushel, *J. Am. Chem. Soc.*, **110**, 8246 (1988)
- [263] J. Rebeck Jr., *Struct. Chem.*, **1**, 129 (1989)
- [264] A. Khan, *J. Phys. Chem. B*, **104**, 11268 (2000)

- [265] M. Witanowski, L. Stefaniak, H. Januszewski, Z. Grabowski and G.A. Webb, *Tetrahedron*, **28**, 637 (1972)
- [266] I.I. Schuster and J.D. Roberts, *J. Org. Chem.*, **44**, 3864 (1979)
- [267] H. Park and K.M. Merz Jr., *J. Am. Chem. Soc.*, **125**, 901 (2003)
- [268] P. Schuster and P. Wolschann, *Mh. Chem.*, **130**, 947 (1999)
- [269] X-J. Song, C.M. Riënstra and A.E. McDermott, *Magn. Reson. Chem.*, **39**, S30 (2001)
- [270] M. Ichikawa, *Acta Crystallogr., Sect. B*, **34**, 2074 (1978)
- [271] M. Ichikawa, *Chem. Phys. Lett.*, **79**, 583 (1978)
- [272] J.L. Markley and I.B. Ibanez, *Biochemistry*, **17**, 4627 (1978)
- [273] A.A. Kossiakoff and S.A. Spencer, *Nature (London)*, **288**, 414 (1980)
- [274] A.A. Kossiakoff and S.A. Spencer, *Biochemistry*, **20**, 6462 (1981)
- [275] E. Fadda, M.E. Casida and D.R. Salahub, *unpublished results*



# **Appendix A**

## **Tables**

All the tables showing the shielding calculation results are reported in this section. I have preferred to detach them from the main block of the work since they give useful but not essential information for the understanding of the subjects.

Table A.I:  $^{13}\text{C}$  shielding constant (ppm) calculated with the four SOS-DFPT approximations and the SVWN  $v_{xc}$

Molecule	SVWN				Ab Initio*	Expt. <sup>†</sup>
	UKS	Loc.1	Loc.2	Loc.3		
CH <sub>4</sub>	193.3	194.4	194.8	194.3	198.6 <sup>‡</sup>	195.1
CH <sub>3</sub> CH <sub>3</sub>	176.2	177.0	177.3	176.8	188.0	180.9
<u>C</u> H <sub>3</sub> CHO	151.2	152.5	153.0	153.1	162.8	157.2
<u>C</u> H <sub>3</sub> COCH <sub>3</sub>	149.0	150.0	150.4	150.4	164.5	158.0
<u>C</u> H <sub>3</sub> CN	182.9	183.5	183.7	183.0	193.6	187.7
CH <sub>3</sub> OH	124.0	125.6	126.1	125.8	142.2	136.6
CH <sub>3</sub> NH <sub>2</sub>	149.3	150.6	151.0	150.4	164.9	158.3
CH <sub>3</sub> F	101.9	103.7	104.2	103.5	121.8	116.8
CH <sub>2</sub> CH <sub>2</sub>	41.1	46.0	47.6	46.2	71.2	64.5
<u>C</u> H <sub>2</sub> CCH <sub>2</sub>	104.1	106.7	107.6	107.5	120.9	115.2
CH <sub>2</sub> <u>C</u> CH <sub>2</sub>	-55.0	-49.5	-47.7	-47.8	-26.0	-29.3
CHCH	102.4	104.9	105.7	105.2	122.6 <sup>‡</sup>	117.2
C <sub>6</sub> H <sub>6</sub>	39.5	41.4	42.1	41.9	64.0	57.2
HCN	67.7	73.0	74.6	78.4	86.3 <sup>§</sup>	82.1
CH <sub>3</sub> <u>C</u> N	56.4	60.4	61.7	66.3	76.1	73.8
HNC	-0.5	13.9	18.2	28.9	28.5 <sup>‡</sup>	...
CH <sub>2</sub> NN	165.8	167.8	168.4	169.0	171.9 <sup>§</sup>	164.5 <sup>§</sup>
HCHO	-39.7	-29.0	-25.7	-24.8	6.1 <sup>‡</sup>	-4.4±3 <sup>¶</sup>
CH <sub>3</sub> <u>C</u> HO	-41.5	-33.6	-31.2	-29.5	1.2	-6.7
CH <sub>3</sub> <u>C</u> OCH <sub>3</sub>	-41.8	-36.0	-34.2	-32.3	-5.8	-13.1
CO	-16.9	-0.8	3.9	21.1	5.6 <sup>§</sup>	0.6 <sup>§</sup>
CO <sub>2</sub>	49.0	50.9	51.5	53.1	63.5	58.8
CF <sub>4</sub>	36.8	38.0	38.4	40.1	64.4	64.5

\*MBPT(2) from Gauss<sup>175</sup> unless otherwise specified.

<sup>†</sup>Experimental absolute shielding values from Jameson and Jameson<sup>174</sup> unless otherwise specified.

<sup>‡</sup>L-CCD from Cybulski and Bishop<sup>141</sup>

<sup>§</sup>CCSD(T) and/or Exp. from Gauss and Stanton<sup>142</sup>

<sup>¶</sup>Exp. from Wilson and Tozer<sup>155</sup>

Table A.II:  $^{13}\text{C}$  shielding constant (ppm) calculated with the four SOS-DFPT approximations. The PW91  $v_x$  is used in combination with the PW91  $v_c$ .

Molecule	PW91-PW91				<i>Ab Initio</i> *	<i>Expt.</i> <sup>†</sup>
	<i>UKS</i>	<i>Loc.1</i>	<i>Loc.2</i>	<i>Loc.3</i>		
CH <sub>4</sub>	188.6	189.8	190.1	189.8	198.6 <sup>‡</sup>	195.1
CH <sub>3</sub> CH <sub>3</sub>	172.4	173.2	173.5	173.0	188.0	180.9
<u>CH</u> <sub>3</sub> CHO	145.9	147.4	147.9	148.0	162.8	157.2
<u>CH</u> <sub>3</sub> COCH <sub>3</sub>	148.3	149.2	149.6	149.5	164.5	158.0
<u>CH</u> <sub>3</sub> CN	179.4	180.1	180.3	179.7	193.6	187.7
CH <sub>3</sub> OH	125.2	126.7	127.2	127.1	142.2	136.6
CH <sub>3</sub> NH <sub>2</sub>	147.3	148.7	149.1	148.9	164.9	158.3
CH <sub>3</sub> F	102.4	104.1	104.6	104.0	121.8	116.8
CH <sub>2</sub> CH <sub>2</sub>	44.7	49.2	50.7	49.1	71.2	64.5
<u>CH</u> <sub>2</sub> CCH <sub>2</sub>	103.6	106.2	107.1	107.8	120.9	115.2
<u>CH</u> <sub>2</sub> <u>C</u> CH <sub>2</sub>	-52.1	-46.6	-44.8	-44.4	-26.0	-29.3
CHCH	108.4	110.7	111.5	110.9	122.6 <sup>‡</sup>	117.2
C <sub>6</sub> H <sub>6</sub>	41.8	43.8	44.4	44.2	64.0	57.2
HCN	72.5	77.5	79.0	82.5	86.3 <sup>§</sup>	82.1
CH <sub>3</sub> <u>C</u> N	59.1	62.5	63.6	67.5	76.1	73.8
HNC	7.9	21.2	25.2	35.0	28.5 <sup>‡</sup>	...
CH <sub>2</sub> NN	162.2	164.3	164.9	165.5	171.9 <sup>§</sup>	164.5 <sup>§</sup>
HCHO	-28.4	-18.8	-15.8	-15.1	6.1 <sup>‡</sup>	-4.4±3 <sup>¶</sup>
CH <sub>3</sub> CHO	-33.0	-25.7	-23.4	-22.0	1.2	-6.7
CH <sub>3</sub> <u>C</u> COCH <sub>3</sub>	-40.4	-34.9	-33.1	-32.0	-5.8	-13.1
CO	-11.6	3.5	8.0	17.3	5.6 <sup>§</sup>	0.6 <sup>§</sup>
CO <sub>2</sub>	50.6	52.7	53.4	55.0	63.5	58.8
CF <sub>4</sub>	40.2	41.4	41.8	43.1	64.4	64.5

\*MBPT(2) from Gauss<sup>175</sup> unless otherwise specified.

<sup>†</sup>Experimental absolute shielding values from Jameson and Jameson<sup>174</sup> unless otherwise specified.

<sup>‡</sup>L-CCD from Cybulski and Bishop<sup>141</sup>

<sup>§</sup>CCSD(T) and/or Exp. from Gauss and Stanton<sup>142</sup>

<sup>¶</sup>Exp. from Wilson and Tozer<sup>155</sup>

Table A.III:  $^{13}\text{C}$  shielding constant (ppm) calculated with the four SOS-DFPT approximations and the ACLDA  $v_{xc}$

Molecule	ACLDA				<i>Ab Initio</i> *	<i>Expt.</i> <sup>†</sup>
	<i>UKS</i>	<i>Loc.1</i>	<i>Loc.2</i>	<i>Loc.3</i>		
CH <sub>4</sub>	193.4	194.6	195.0	194.9	198.6 <sup>‡</sup>	195.1
CH <sub>3</sub> CH <sub>3</sub>	174.6	175.5	175.8	175.2	188.0	180.9
CH <sub>3</sub> CHO	148.0	149.5	149.9	150.1	162.8	157.2
CH <sub>3</sub> COCH <sub>3</sub>	148.9	150.0	150.3	150.3	164.5	158.0
CH <sub>3</sub> CN	182.1	182.7	182.9	182.1	193.6	187.7
CH <sub>3</sub> OH	119.9	121.6	122.1	122.0	142.2	136.6
CH <sub>3</sub> NH <sub>2</sub>	147.5	148.9	149.4	148.8	164.9	158.3
CH <sub>3</sub> F	98.4	100.3	100.9	100.5	121.8	116.8
CH <sub>2</sub> CH <sub>2</sub>	45.2	49.7	51.2	49.8	71.2	64.5
CH <sub>2</sub> CCH <sub>2</sub>	100.1	102.8	103.6	103.6	120.9	115.2
CH <sub>2</sub> CCH <sub>2</sub>	-55.9	-50.4	-48.6	-48.8	-26.0	-29.3
CHCH	92.4	95.2	96.1	95.6	122.6 <sup>‡</sup>	117.2
C <sub>6</sub> H <sub>6</sub>	40.1	42.0	42.7	42.5	64.0	57.2
HCN	55.9	61.5	63.3	67.5	86.3 <sup>§</sup>	82.1
CH <sub>3</sub> CN	47.3	51.8	53.2	58.3	76.1	73.8
HNC	-23.9	-7.3	-2.5	10.3	28.5 <sup>‡</sup>	...
CH <sub>2</sub> NN	165.5	167.5	168.1	170.3	171.9 <sup>§</sup>	164.5 <sup>§</sup>
HCHO	-37.0	-27.0	-23.9	-23.0	6.1 <sup>‡</sup>	-4.4±3 <sup>¶</sup>
CH <sub>3</sub> CHO	-54.0	-45.4	-42.7	-40.8	1.2	-6.7
CH <sub>3</sub> COCH <sub>3</sub>	-56.1	-49.5	-47.5	-45.3	-5.8	-13.1
CO	-5.8	9.0	13.2	22.0	5.6 <sup>§</sup>	0.6 <sup>§</sup>
CO <sub>2</sub>	48.0	50.5	50.6	52.1	63.5	58.8
CF <sub>4</sub>	38.4	39.6	39.9	40.9	64.4	64.5

\*MBPT(2) from Gauss<sup>175</sup> unless otherwise specified.

<sup>†</sup>Experimental absolute shielding values from Jameson and Jameson<sup>174</sup> unless otherwise specified.

<sup>‡</sup>L-CCD from Cybulski and Bishop<sup>141</sup>

<sup>§</sup>CCSD(T) and/or Exp. from Gauss and Stanton<sup>142</sup>

<sup>¶</sup>Exp. from Wilson and Tozer<sup>155</sup>

Table A.IV:  $^{14,15}\text{N}$  shielding constant (ppm) calculated with the four SOS-DFPT approximations and the SVWN  $v_{xc}$ .

Molecule	SVWN				<i>Ab Initio</i>	<i>Expt.*</i>
	<i>UKS</i>	<i>Loc.1</i>	<i>Loc.2</i>	<i>Loc.3</i>		
$\text{NH}_3$	263.9	265.4	265.8	264.6	268.8 <sup>†</sup>	263.5 <sup>¶</sup>
$\text{CH}_3\text{NH}_2$	262.2	264.5	265.1	264.1	261.2 <sup>‡</sup>	...
$\text{HNC}$	87.1	96.0	98.7	105.2	105.5 <sup>†</sup>	...
$\text{HCN}$	-54.2	-39.7	-35.2	-23.8	-14.4 <sup>†</sup>	-20.4 <sup>§</sup>
$\text{CH}_3\text{CN}$	-41.3	-29.2	-25.3	-15.1	13.2 <sup>‡</sup>	-9.1 <sup>¶</sup>
$\text{N}_2$	-88.3	-71.4	-66.2	-64.1	-55.7 <sup>†</sup>	-60.5 <sup>¶</sup>
$\text{CH}_2\text{NN}$	-58.4	-48.7	-45.6	-39.2	-31.6 <sup>§</sup>	-43.4 <sup>§</sup>
$\text{CH}_2\text{N}\underline{\text{N}}$	-162.2	-138.9	-131.7	-124.8	-142.4 <sup>§</sup>	-149.0 <sup>§</sup>
$\text{NNO}$	85.6	93.6	96.2	104.3	135.1 <sup>‡</sup>	99.5 <sup>‡</sup>
$\text{N}\underline{\text{N}}\text{O}$	-6.8	-0.7	1.3	8.2	33.8 <sup>‡</sup>	11.3 <sup>‡</sup>

\* Experimental absolute shieldings

<sup>†</sup> L-CCD from Cybulski and Bishop<sup>141</sup>

<sup>‡</sup> MBPT(2) from Gauss<sup>175</sup> unless otherwise specified.

<sup>§</sup> CCSD(T) and/or Exp. from Gauss and Stanton<sup>142</sup>

<sup>¶</sup> From Jameson and coworkers<sup>183</sup>

Table A.V:  $^{14,15}\text{N}$  shielding constant (ppm) calculated with the four SOS-DFPT approximations. The PW91  $v_x$  is used in combination with the PW91  $v_c$ .

Molecule	PW91-PW91				<i>Ab Initio</i>	<i>Expt.*</i>
	<i>UKS</i>	<i>Loc.1</i>	<i>Loc.2</i>	<i>Loc.3</i>		
$\text{NH}_3$	260.5	262.0	262.5	261.4	268.8 <sup>†</sup>	263.5 <sup>¶</sup>
$\text{CH}_3\text{NH}_2$	236.2	238.0	238.5	237.7	261.2 <sup>‡</sup>	...
$\text{HNC}$	93.2	101.8	104.5	110.7	105.5 <sup>†</sup>	...
$\text{HCN}$	-42.7	-29.3	-25.1	-15.0	-14.4 <sup>†</sup>	-20.4 <sup>§</sup>
$\text{CH}_3\text{CN}$	-46.8	-36.9	-33.7	-26.4	13.2 <sup>‡</sup>	-9.1 <sup>¶</sup>
$\text{N}_2$	-97.9	-80.2	-74.7	-67.4	-55.7 <sup>†</sup>	-60.5 <sup>¶</sup>
$\text{CH}_2\text{NN}$	-60.3	-50.5	-47.4	-41.1	-31.6 <sup>§</sup>	-43.4 <sup>§</sup>
$\text{CH}_2\text{NN}$	-170.6	-146.9	-139.6	-133.4	-142.4 <sup>§</sup>	-149.0 <sup>§</sup>
$\text{NNO}$	83.4	91.5	94.0	102.1	135.1 <sup>‡</sup>	99.5 <sup>‡</sup>
$\text{NNO}$	-7.9	-1.7	0.3	7.6	33.8 <sup>‡</sup>	11.3 <sup>‡</sup>

\*Experimental absolute shielding values

<sup>†</sup>L-CCD from Cybulski and Bishop<sup>141</sup>

<sup>‡</sup>MBPT(2) from Gauss<sup>175</sup> unless otherwise specified.

<sup>§</sup>CCSD(T) and/or Exp. from Gauss and Stanton<sup>142</sup>

<sup>¶</sup>From Jameson and coworkers<sup>183</sup>

Table A.VI:  $^{15}\text{N}$  shielding constant (ppm) calculated with the four SOS-DFPT approximation and the AC-LDA  $v_{\text{xc}}$

Molecule	AC-LDA				<i>Ab Initio</i>	<i>Expt.*</i>
	<i>UKS</i>	<i>Loc.1</i>	<i>Loc.2</i>	<i>Loc.3</i>		
$\text{NH}_3$	256.4	258.1	256.6	257.7	268.8 <sup>†</sup>	263.5 <sup>¶</sup>
$\text{CH}_3\text{NH}_2$	265.8	267.5	268.0	267.5	261.2 <sup>‡</sup>	...
$\text{HNC}$	78.5	88.9	92.0	99.8	105.5 <sup>†</sup>	...
$\text{HCN}$	-61.3	-46.1	-41.4	-29.3	-14.4 <sup>†</sup>	-20.4 <sup>§</sup>
$\text{CH}_3\text{CN}$	-64.9	-51.6	-47.4	-36.7	13.2 <sup>‡</sup>	-9.1 <sup>¶</sup>
$\text{N}_2$	-120.2	-101.0	-95.1	-88.8	-55.7 <sup>†</sup>	-60.5 <sup>¶</sup>
$\text{CH}_2\text{NN}$	-65.9	-55.6	-52.4	-38.2	-31.6 <sup>§</sup>	-43.4 <sup>§</sup>
$\text{CH}_2\text{N}\underline{\text{N}}$	-184.2	-159.0	-151.3	-120.6	-142.4 <sup>§</sup>	-149.0 <sup>§</sup>
$\underline{\text{N}}\text{NO}$	75.3	83.8	86.5	103.5	135.1 <sup>†</sup>	99.5 <sup>‡</sup>
$\text{N}\underline{\text{N}}\text{O}$	-12.6	-6.3	-4.3	12.2	33.8 <sup>‡</sup>	11.3 <sup>‡</sup>

\*Experimental absolute shieldings

<sup>†</sup>L-CCD from Cybulski and Bishop<sup>141</sup>

<sup>‡</sup>MBPT(2) from Gauss<sup>175</sup> unless otherwise specified.

<sup>§</sup>CCSD(T) and/or Exp. from Gauss and Stanton<sup>142</sup>

<sup>¶</sup>From Jameson and coworkers<sup>183</sup>

Table A.VII:  $^{17}\text{O}$  shielding constant calculated with the four SOS-DFPT approximations and the SVWN  $v_{xc}$

Molecule	SVWN				<i>Ab Initio</i>	<i>Expt.*</i>
	<i>UKS</i>	<i>Loc.1</i>	<i>Loc.2</i>	<i>Loc.3</i>		
CH <sub>3</sub> CHO	-432.9	-394.3	-382.4	-381.1	-291.7 <sup>†</sup>	...
CH <sub>3</sub> COCH <sub>3</sub>	-371.0	-341.2	-331.8	-325.4	-279.8 <sup>†</sup>	...
CH <sub>3</sub> OH	323.1	325.1	325.7	326.1	350.6 <sup>†</sup>	...
HCHO	-504.9	-449.4	-432.5	-430.3	-418.0 <sup>‡</sup>	-375±100 <sup>§</sup>
CO	-83.9	-60.1	-53.0	-20.9	-57.4 <sup>‡</sup>	-36.7±17.2 <sup>†</sup>
CO <sub>2</sub>	205.9	210.0	211.3	213.3	241.0 <sup>†</sup>	243.4 <sup>§</sup>
H <sub>2</sub> O	330.1	332.3	333.1	335.6	335.4 <sup>‡</sup>	357.6±17.2 <sup>†</sup>
H <sub>2</sub> O <sub>2</sub>	99.0	112.1	116.2	135.6	133.9 <sup>‡</sup>	...
N <sub>2</sub> O	166.9	173.9	176.2	181.8	206.2 <sup>‡</sup>	200.5 <sup>¶</sup>
OF <sub>2</sub>	-675.4	-603.1	-581.7	-529.2	-479.8 <sup>†</sup>	-473.1 <sup>§</sup>

\* Experimental absolute shieldings

<sup>†</sup> CCSD(T) calculation from Gauss and Stanton<sup>139</sup>

<sup>‡</sup> L-CCD calculation from Cybulski and Bishop<sup>141</sup>

<sup>§</sup> From Wilson and coworkers<sup>60</sup>

<sup>¶</sup> MBPT(2) calculation from Gauss<sup>175</sup>



Table A. VIII:  $^{17}\text{O}$  shielding constant calculated with the four SOS-DFPT approximations and the PW91-PW91  $v_{xc}$

Molecule	PW91-PW91				<i>Ab Initio</i>	<i>Expt.*</i>
	<i>UKS</i>	<i>Loc.1</i>	<i>Loc.2</i>	<i>Loc.3</i>		
CH <sub>3</sub> CHO	-399.1	-362.6	-351.3	-350.5	-291.7 <sup>†</sup>	...
CH <sub>3</sub> COCH <sub>3</sub>	-358.0	-330.9	-322.3	-320.6	-279.8 <sup>†</sup>	...
CH <sub>3</sub> OH	329.0	330.9	331.6	331.9	350.6 <sup>†</sup>	...
HCHO	-448.0	-399.4	-384.5	-382.9	-418.0 <sup>‡</sup>	-375±100 <sup>§</sup>
CO	-84.8	-60.6	-53.4	-33.7	-57.4 <sup>‡</sup>	-36.7±17.2 <sup>†</sup>
CO <sub>2</sub>	201.3	205.6	206.9	209.1	241.0 <sup>†</sup>	243.4 <sup>§</sup>
H <sub>2</sub> O	327.7	329.9	330.7	329.3	335.4 <sup>‡</sup>	357.6±17.2 <sup>†</sup>
H <sub>2</sub> O <sub>2</sub>	107.3	118.9	122.5	139.0	133.9 <sup>‡</sup>	...
N <sub>2</sub> O	154.2	161.8	164.2	170.6	206.2 <sup>†</sup>	200.5 <sup>¶</sup>
OF <sub>2</sub>	-637.2	-571.7	-552.1	-503.2	-479.8 <sup>†</sup>	-473.1 <sup>§</sup>

\* Experimental absolute shieldings

<sup>†</sup> CCSD(T) calculation from Gauss and Stanton<sup>139</sup>

<sup>‡</sup> L-CCD calculation from Cybulski and Bishop<sup>141</sup>

<sup>§</sup> From Wilson and coworkers<sup>60</sup>

<sup>¶</sup> MBPT(2) calculation from Gauss<sup>175</sup>

Table A.IX:  $^{17}\text{O}$  shielding constant calculated with the four SOS-DFPT approximations and the AC-LDA  $v_{xc}$

Molecule	AC-LDA				<i>Ab Initio</i>	<i>Expt.*</i>
	<i>UKS</i>	<i>Loc.1</i>	<i>Loc.2</i>	<i>Loc.3</i>		
CH <sub>3</sub> CHO	-490.4	-446.8	-433.4	-431.9	-291.7 <sup>†</sup>	...
CH <sub>3</sub> COCH <sub>3</sub>	-420.8	-386.8	-376.2	-368.3	-279.8 <sup>†</sup>	...
CH <sub>3</sub> OH	326.0	328.0	328.7	329.3	350.6 <sup>†</sup>	...
HCHO	-506.1	-451.4	-434.8	-432.4	-418.0 <sup>‡</sup>	-375±100 <sup>§</sup>
CO	-56.3	-35.1	-28.7	-11.0	-57.4 <sup>‡</sup>	-36.7±17.2 <sup>†</sup>
CO <sub>2</sub>	207.3	211.4	212.7	214.7	241.0 <sup>†</sup>	243.4 <sup>§</sup>
H <sub>2</sub> O	317.8	320.4	321.2	320.0	335.4 <sup>‡</sup>	357.6±17.2 <sup>†</sup>
H <sub>2</sub> O <sub>2</sub>	95.6	108.6	112.7	132.8	133.9 <sup>‡</sup>	...
N <sub>2</sub> O	164.3	172.3	174.8	182.3	206.2 <sup>†</sup>	200.5 <sup>¶</sup>
OF <sub>2</sub>	-504.5	-455.1	-440.1	-404.6	-479.8 <sup>†</sup>	-473.1 <sup>§</sup>

\* Experimental absolute shieldings

<sup>†</sup> CCSD(T) calculation from Gauss and Stanton<sup>139</sup>

<sup>‡</sup> L-CCD calculation from Cybulski and Bishop<sup>141</sup>

<sup>§</sup> From Wilson and coworkers<sup>60</sup>

<sup>¶</sup> MBPT(2) calculation from Gauss<sup>175</sup>

Table A.X:  $^{13}\text{C}$  shielding constants (ppm) calculated with the SOS-DFPT Loc.3 approximation (LDA) and with other DFT-based methodologies. High quality *ab initio* and experimental data are shown as reference

Molecule	Loc.3	MKS*	(B3LYP <sup>0.05</sup> <sub>GGA</sub> ) <sup>†</sup>	PBE0 <sup>‡</sup>	SIC-VWN <sup>§</sup>	Ab Initio <sup>¶</sup>	Expt. <sup>  </sup>
CH <sub>4</sub>	194.3	189.9	193.1	194.0	191.5	198.6	195.1
CH <sub>3</sub> CH <sub>3</sub>	176.8	...	...	179.7	177.7	188.0	180.9
CH <sub>3</sub> CHO	153.1	...	...	...	152.5	162.8	157.2
CH <sub>3</sub> COCH <sub>3</sub>	150.4	...	...	157.0	153.9	164.5	158.0
CH <sub>3</sub> CN	183.0	...	...	187.7	183.8	193.6	187.7
CH <sub>3</sub> OH	125.8	...	...	136.5	...	142.2	136.6
CH <sub>3</sub> NH <sub>2</sub>	150.4	...	...	157.1	153.9	164.9	158.3
CH <sub>3</sub> F	103.5	...	...	116.5	107.1	121.8	116.8
CH <sub>2</sub> CH <sub>2</sub>	46.2	54.8	...	58.4	58.0	71.2	64.5
CH <sub>2</sub> CCH <sub>2</sub>	107.5	...	...	112.5	109.2	120.9	115.2
CH <sub>2</sub> CCH <sub>2</sub>	-47.8	...	...	-36.6	-31.1	-26.0	-29.3
CHCH	105.2	112.1	114.2	114.0	113.7	122.6**	117.2
C <sub>6</sub> H <sub>6</sub>	41.9	...	...	55.3	54.9	64.0	57.2
HCN	78.4	78.0	81.3	76.6	74.4	86.3 <sup>††</sup>	82.1
CH <sub>3</sub> CN	66.3	...	...	68.2	67.4	76.1	73.8
HNC	28.9	...	22.2	...	...	28.5**	...
CH <sub>2</sub> NN	169.0	163.4	166.5	...	166.5	171.9 <sup>††</sup>	164.5 <sup>††</sup>
HCHO	-24.8	-14.7	-5.8	-11.1	-23.8	6.1**	-4.4±3*
CH <sub>3</sub> CHO	-29.5	...	...	...	-14.4	1.2	-6.7
CH <sub>3</sub> COCH <sub>3</sub>	-32.3	...	...	-11.1	-27.4	-5.8	-13.1
CO	21.1	-2.7	4.7	-7.8	-11.6	5.6 <sup>††</sup>	0.6 <sup>††</sup>
CO <sub>2</sub>	53.1	57.9	59.7	56.8	61.0	63.5	58.8
CF <sub>4</sub>	40.1	...	...	59.2	48.1	64.4	64.5

\*MKS(B97-1) from Wilson and Tozer<sup>155</sup>

<sup>†</sup>From Wilson and coworkers<sup>60</sup>

<sup>‡</sup>From Adamo and Barone<sup>66</sup>

<sup>§</sup>From Patchkovskii and coworkers<sup>156</sup>

<sup>¶</sup>MBPT(2) calculation or exp. from Gauss<sup>175</sup> unless specified otherwise

<sup>||</sup>Experimental absolute shieldings from Jameson and Jameson<sup>174</sup> unless specified otherwise

\*\*L-CCD calculation or exp. from Cybulski and Bishop<sup>141</sup>

<sup>††</sup>CCSD(T) calculation or exp. from Gauss and Stanton<sup>142</sup>

Table A.XI:  $^{13}\text{C}$  shielding constants (ppm) calculated with the SOS-DFPT Loc.3 approximation (LDA) and with the KT1 and KT2 functionals. High quality *ab initio* and experimental data are shown as reference

Molecule	<i>Loc.3</i>	<i>KT1</i> *	<i>KT2</i> *	<i>Ab Initio</i> <sup>†</sup>	<i>Expt.</i> <sup>‡</sup>
CH <sub>4</sub>	194.3	196.4	195.2	198.6	195.1
C <sub>2</sub> H <sub>4</sub>	47.6	64.3	63.2	71.2	64.5
CHCH	105.2	120.5	120.4	122.6 <sup>§</sup>	117.2
HCN	78.6	87.2	86.0	86.3	82.1
CH <sub>2</sub> NN	169.0	170.1	167.4	171.9	164.5 <sup>§</sup>
CO	21.1	10.4	7.4	5.6	0.6 <sup>§</sup>
CO <sub>2</sub>	53.1	65.0	63.7	63.5 <sup>¶</sup>	58.8

\*Results from Keal and Tozer<sup>179</sup>

<sup>†</sup>CCSD(T) results from Gauss and Stanton<sup>142</sup> unless otherwise specified.

<sup>‡</sup>Experimental absolute shieldings from Jameson and Jameson<sup>174</sup> unless specified otherwise

<sup>§</sup>L-CCD calculation or exp. from Cybulski and Bishop<sup>141</sup>

<sup>¶</sup>MBPT(2) calculation or exp. from Gauss<sup>175</sup>

Table A.XII:  $^{15}\text{N}$  shielding constant (ppm) calculated with the Loc.3 (LDA) and with other DFT-based methodologies. High quality *ab initio* and experimental data are shown as reference

Molecule	Loc.3	MKS*	(B3LYP <sub>GGA</sub> <sup>0.05</sup> ) <sup>†</sup>	PBE0 <sup>‡</sup>	SIC-VWN <sup>§</sup>	Ab Initio	Expt. <sup>¶</sup>
NH <sub>3</sub>	264.6	261.6	265.0	263.1	258.4	268.8 <sup>  </sup>	263.5 <sup>††</sup>
CH <sub>3</sub> NH <sub>2</sub>	264.1	...	...	244.0	243.4	261.2 <sup>**</sup>	...
HNC	105.2	...	105.8	...	...	105.5 <sup>  </sup>	...
HCN	-23.8	-27.0	-21.1	-34.9	-23.5	-14.4 <sup>  </sup>	-20.4 <sup>††</sup>
CH <sub>3</sub> CN	-15.1	...	...	-24.4	-13.7	13.2 <sup>**</sup>	-9.1 <sup>‡‡</sup>
N <sub>2</sub>	-64.1	-65.1	-55.8	-76.8	-73.8	-55.7 <sup>  </sup>	-60.5 <sup>‡‡</sup>
CH <sub>2</sub> NN	-39.2	-46.8	-38.2	...	-36.7	-31.6 <sup>††</sup>	-43.4 <sup>††</sup>
CH <sub>2</sub> NN	-124.8	-137.5	-120.6	...	-161.9	-142.4 <sup>††</sup>	-149.0 <sup>‡‡</sup>
NNO	104.3	98.9	103.5	...	85.6	135.1 <sup>**</sup>	99.5 <sup>**</sup>
NNO	8.2	6.1	12.2	...	12.5	33.8 <sup>**</sup>	11.3 <sup>**</sup>

\*MKS(B97-1) from Wilson and Tozer<sup>155</sup>

†From Wilson and coworkers<sup>60</sup>

‡From Adamo and Barone<sup>66</sup>

§From Patchkovskii and coworkers<sup>156</sup>

¶Experimental absolute shielding values

||L-CCD calculation or exp. from Cybulski and Bishop<sup>141</sup>

\*\*MBPT(2) calculation or exp. from Gauss<sup>175</sup>

††CCSD(T) calculation or exp. from Gauss and Stanton<sup>139</sup>

‡‡From Jameson and coworkers<sup>183</sup>

Table A.XIII:  $^{14,15}\text{N}$  shielding constants (ppm) calculated with the SOS-DFPT Loc.3 approximation (LDA) and with the KT1 and KT2 functionals. High quality *ab initio* and experimental data are shown as reference

Molecule	<i>Loc.3</i>	<i>KT1*</i>	<i>KT2*</i>	<i>Ab Initio</i>	<i>Expt.</i>
$\text{N}_2$	-64.1	-55.8	-59.7	-55.7 <sup>†</sup>	-60.5 <sup>¶</sup>
$\text{NH}_3$	264.6	265.9	264.5	268.8 <sup>†</sup>	263.5 <sup>§</sup>
$\text{HCN}$	-23.8	-18.6	-19.4	-14.4 <sup>†</sup>	-20.4 <sup>§</sup>
$\underline{\text{N}}\text{NO}$	104.3	106.8	102.1	135.1 <sup>‡</sup>	99.5 <sup>‡</sup>
$\text{N}\underline{\text{N}}\text{O}$	8.2	14.2	12.2	33.8 <sup>‡</sup>	11.3 <sup>‡</sup>
$\text{CH}_2\underline{\text{N}}\text{N}$	-39.2	-37.5	-41.7	-31.6 <sup>§</sup>	-43.4 <sup>§</sup>
$\text{CH}_2\underline{\text{N}}\underline{\text{N}}$	-124.8	-128.3	-138.4	-142.4 <sup>§</sup>	-149.0 <sup>¶</sup>

\*From Keal and Tozer<sup>179</sup>

<sup>†</sup>L-CCD calculation or exp. from Cybulsky and Bishop<sup>141</sup>

<sup>‡</sup>MBPT(2) calculation or exp. from Gauss<sup>175</sup>

<sup>§</sup>CCSD(T) calculation or exp. from Gauss and Stanton<sup>139</sup>

<sup>¶</sup>From Jameson and coworkers<sup>183</sup>

Table A.XIV:  $^{17}\text{O}$  shielding constant calculated with the SOS-DFPT Loc.3 approximation (LDA) and with other DFT-based methodologies. High quality *ab initio* and experimental data are shown as reference

Molecule	Loc.3	MKS*	(B3LYP <sup>0.05</sup> <sub>GGA</sub> ) <sup>†</sup>	PBE0 <sup>‡</sup>	SIC-VWN <sup>§</sup>	Ab Initio	Expt. <sup>¶</sup>
CH <sub>3</sub> CHO	-381.1	...	...	...	-273.0	-291.7 <sup>  </sup>	...
CH <sub>3</sub> COCH <sub>3</sub>	-325.4	...	...	-330.2	-259.3	-279.8 <sup>  </sup>	...
CH <sub>3</sub> OH	326.1	...	...	334.7	...	350.6 <sup>  </sup>	...
HCHO	-430.3	-360.6	-346.4	-422.2	-335.3	-418.0 <sup>**</sup>	-375.0 <sup>†</sup>
CO	-20.9	-50.8	-44.4	-70.0	-43.1	-57.4 <sup>**</sup>	-36.7 <sup>‡‡</sup>
CO <sub>2</sub>	213.3	224.2	223.2	220.0	227.6	241.0 <sup>  </sup>	243.4 <sup>†</sup>
H <sub>2</sub> O	335.6	330.2	334.4	328.9	324.0	335.4 <sup>**</sup>	357.6 <sup>‡‡</sup>
H <sub>2</sub> O <sub>2</sub>	135.6	...	119.0	...	89.3	133.9 <sup>**</sup>	...
N <sub>2</sub> O	181.8	190.1	192.0	...	199.4	206.2 <sup>  </sup>	200.5 <sup>††</sup>
OF <sub>2</sub>	-529.2	-505.0	-502.5	...	-697.0	-479.8 <sup>  </sup>	-473.1 <sup>†</sup>

\*MKS(B97-1) from Wilson and Tozer<sup>155</sup>

†From Wilson and coworkers<sup>60</sup>

‡From Adamo and Barone<sup>66</sup>

§From Patchkovskii and coworkers<sup>156</sup>

¶Experimental absolute shielding values

||CCSD(T) calculation from Gauss and Stanton<sup>139</sup>

\*\*L-CCD calculation or exp. from Cybulski and Bishop<sup>141</sup>

††MBPT(2) calculation or exp. from Gauss<sup>175</sup>

‡‡Exp.  $\pm 17.2$  from Gauss and Stanton<sup>139</sup>

Table A.XV:  $^{17}\text{O}$  shielding constants (ppm) calculated with the SOS-DFPT Loc.3 approximation (LDA) and with the KT1 and KT2 functionals. High quality *ab initio* and experimental data are shown as reference

Molecule	<i>Loc.3</i>	<i>KT1*</i>	<i>KT2*</i>	<i>Ab Initio</i>	<i>Expt.</i>
H <sub>2</sub> O	335.6	330.7	329.6	335.4*	357.6 <sup>†</sup>
CO	-20.9	-56.1	-57.1	-57.4*	-36.7 <sup>†</sup>
H <sub>2</sub> CO	-430.3	-383.8	-379.6	-418.0*	-375.0 <sup>‡</sup>
N <sub>2</sub> O	181.8	184.1	177.5	206.2 <sup>§</sup>	200.5 <sup>¶</sup>
CO <sub>2</sub>	213.3	224.5	221.6	241.0 <sup>§</sup>	233.4 <sup>‡</sup>
OF <sub>2</sub>	-529.2	-516.7	-534.0	-479.8 <sup>§</sup>	-473.1 <sup>‡</sup>

\*L-CCD calculation or exp. from Cybulski and Bishop<sup>141</sup>

<sup>†</sup>Exp.  $\pm 17.2$  from Gauss and Stanton<sup>139</sup>

<sup>‡</sup>From Wilson and coworkers<sup>60</sup>

<sup>§</sup>CCSD(T) calculation from Gauss and Stanton<sup>139</sup>

<sup>¶</sup>MBPT(2) calculation or exp. from Gauss<sup>175</sup>



Table A.XVI: Calculated and experimental  $^{15}\text{N}$  chemical shifts (ppm) in amines relative to neat liquid nitromethane

Molecule	UKS	Loc.1	Loc.2	Loc.3	Expt.*	Solvent
$\text{NH}_3$	400.5	402.1	402.6	401.6	399.9	gas phase
$\text{MeNH}_2$	378.5	380.2	380.8	380.2	385.4	gas phase
$\text{PrNH}_2$	362.6	363.8	364.1	363.6	356.0 <sup>†</sup>	neat liquid
$\text{BuNH}_2$	361.2	362.2	362.6	362.3	353.4 <sup>†</sup>	neat liquid
<i>i</i> - $\text{BuNH}_2$	358.8	359.9	360.2	359.7	359.4 <sup>†</sup>	neat liquid
<i>s</i> - $\text{BuNH}_2$	334.8	335.9	336.3	336.1	337.5 <sup>†</sup>	neat liquid
<i>t</i> - $\text{BuNH}_2$	314.3	315.5	315.9	315.8	317.3 <sup>†</sup>	neat liquid
$\text{NH}_2(\text{CH}_2)_2\text{NH}_2$	364.9	366.0	366.3	365.6	364.8 <sup>‡</sup>	neat liquid
$\text{NH}_2(\text{CH}_2)_3\text{NH}_2$	358.6	359.4	359.7	359.7	360.3 <sup>‡</sup>	neat liquid
$\text{HO}(\text{CH}_2)_2\text{NH}_2$	364.6	365.7	366.1	365.9	365.7 <sup>‡</sup>	neat liquid
$\text{MeO}(\text{CH}_2)_2\text{NH}_2$	363.9	364.9	365.2	365.0	367.8 <sup>‡</sup>	neat liquid
$\text{CH}_2\text{CHCH}_2\text{NH}_2$	358.5	359.7	360.2	359.7	361.6 <sup>‡</sup>	neat liquid
$\text{PhCH}_2\text{NH}_2$	357.9	359.0	359.3	359.7	358.5 <sup>‡</sup>	neat liquid
$\text{Ph}(\text{CH}_2)_2\text{NH}_2$	362.3	363.1	363.3	363.2	360.6 <sup>‡</sup>	neat liquid
$\text{PhNH}_2$	320.2	321.4	321.8	321.8	325.4 <sup>‡</sup>	neat liquid
1- $\text{NH}_2$ -pyrrole	306.3	307.9	308.5	311.7	306.4 <sup>†</sup>	DMSO
1- $\text{NH}_2$ -imidazole	314.9	316.4	316.8	316.7	311.4 <sup>†</sup>	DMSO
$\text{C}_6\text{H}_{12}\text{NH}_2$	334.6	335.5	335.8	335.8	335.7 <sup>†</sup>	neat liquid
$\text{Me}_2\text{NH}$	361.0	362.6	363.1	363.1	373.8	gas phase
$\text{Et}_2\text{NH}$	323.7	325.1	325.6	325.6	334.0 <sup>‡</sup>	neat liquid
$\text{Pr}_2\text{NH}$	327.6	328.9	331.3	329.2	341.3 <sup>‡</sup>	neat liquid
$\text{C}_4\text{H}_9\text{N}^{\S}$	325.2	326.3	326.7	326.6	343.4 <sup>‡</sup>	neat liquid
$\text{C}_4\text{H}_{10}\text{N}_2^{\P}$	333.7	335.7	335.1	335.3	346.4 <sup>†</sup>	$\text{CDCl}_3$
$\text{Me}_3\text{N}$	351.6	352.9	353.3	353.4	372.8	gas phase

\*Experimental shifts from Witanowski and coworkers<sup>194</sup> unless otherwise specified

<sup>†</sup>Experimental shift from Witanowski and coworkers<sup>193</sup>

<sup>‡</sup>No correction for bulk magnetic susceptibility

<sup>§</sup>Tetrahydropyrrole or Pyrrolidine

<sup>¶</sup>Tetrahydropyrazine

Table A.XVII: Calculated and experimental  $^{15}\text{N}$  chemical shifts (ppm) in hydrazines relative to neat liquid nitromethane

Molecule*	UKS	Loc.1	Loc.2	Loc.3	Expt. <sup>†</sup>	Solvent
$\text{H}_2\text{NNH}_2$	325.1	327.4	328.2	327.0	334.8	neat liquid
$\text{MeH}\underline{\text{NN}}\text{H}_2$	309.5	311.6	312.2	312.0	328.0	neat liquid
$\text{MeH}\underline{\text{NN}}\text{H}_2$	303.4	305.3	305.9	305.1	305.5	neat liquid
$\text{Me}_2\underline{\text{NN}}\text{H}_2$	399.6	301.2	301.7	302.3	322.7	neat liquid
$\text{Me}_2\underline{\text{NN}}\text{H}_2$	270.7	272.5	273.1	273.7	281.4	neat liquid
$\text{Me}_2\underline{\text{NN}}\text{HMe}$	282.1	283.7	284.2	285.0	307.7	neat liquid
$\text{Me}_2\underline{\text{NN}}\text{HMe}$	263.3	265.1	265.5	265.9	285.3	neat liquid
$\text{PhH}\underline{\text{NN}}\text{H}_2$	271.8	273.3	273.7	274.3	294.2	neat liquid
$\text{PhH}\underline{\text{NN}}\text{H}_2$	315.5	316.6	317.0	317.6	319.5	neat liquid

\* In case of non equivalent nitrogen atoms, the nucleus considered is underlined

† Experimental shifts from Witanowski and coworkers<sup>194</sup>

Table A.XVIII: Calculated and experimental  $^{15}\text{N}$  chemical shifts (ppm) in amides, ureas, guanidines and related structures relative to neat liquid nitromethane

Molecule*	UKS	Loc.1	Loc.2	Loc.3	Expt. <sup>†</sup>	Solvent
HCONH <sub>2</sub>	269.1	273.1	274.3	276.4	267.7	neat liquid
(E)-HCONHMe	267.9	270.7	271.6	272.0	270.2	neat liquid
HCONMe <sub>2</sub>	265.2	267.7	268.5	269.1	277.0	neat liquid
MeCONH <sub>2</sub>	271.6	274.3	275.2	275.8	274.7 <sup>‡</sup>	Me <sub>2</sub> CO
(E)-MeCONHMe	270.2	272.7	273.4	274.5	277.8 <sup>‡</sup>	DMSO
(Z)-MeCONHMe	270.0	272.5	273.2	273.8	276.0 <sup>‡</sup>	DMSO
MeCONMe <sub>2</sub>	265.3	267.4	268.1	268.9	282.1 <sup>‡</sup>	Me <sub>2</sub> CO
PhCONH <sub>2</sub>	285.0	286.3	286.7	287.0	278.4 <sup>‡</sup>	DMSO
EtOCONH <sub>2</sub>	311.1	312.5	313.0	313.2	308.2	neat liquid
EtOCONHEt	284.6	285.9	286.3	286.4	294.9	neat liquid
H <sub>2</sub> NCONH <sub>2</sub>	303.4	305.5	306.1	307.2	303.6	H <sub>2</sub> O
Me <sub>2</sub> NCONMe <sub>2</sub>	303.9	304.8	305.1	305.2	317.8	neat liquid
o-NMe-pyrrolidone	249.9	252.3	253.0	257.7	268.9 <sup>‡</sup>	DMSO
o-NEt-pyrrolidone	232.9	234.9	235.5	238.5	255.6 <sup>‡</sup>	DMSO
o-NMe-piperidone	248.0	249.7	250.2	252.2	270.3 <sup>‡</sup>	CHCl <sub>3</sub>
o-pyridone	188.7	191.1	191.9	193.1	209.2	Me <sub>2</sub> CO, 1:3(v/v)
o-NMe-pyridone	180.9	183.0	183.7	184.9	215.0	neat liquid
ε-caprolactam	249.4	251.3	252.0	253.6	265.3	CDCl <sub>3</sub>
H <sub>2</sub> NC(NH)NH <sub>2</sub>	286.7	289.4	290.2	290.8	307.0	0.2M in H <sub>2</sub> O <sup>§</sup>
Me <sub>2</sub> <u>NC</u> (NMe)NMe <sub>2</sub>	311.1	311.9	312.2	312.3	334.0	neat liquid
Me <sub>2</sub> NC( <u>N</u> Me)NMe <sub>2</sub>	195.3	198.3	199.3	200.5	187.0	neat liquid
Me <sub>2</sub> <u>NC</u> (NH)OMe	314.5	315.7	316.1	316.4	320.0	neat liquid
Me <sub>2</sub> NC( <u>NH</u> )OMe	223.1	226.9	228.2	230.2	283.3	neat liquid

\* In case of non equivalent nitrogen atoms, the nucleus considered is underlined

<sup>†</sup> Experimental shifts from Witanowski and coworkers<sup>194</sup> unless otherwise specified

<sup>‡</sup> Experimental shift from Witanowski and coworkers<sup>193</sup>

<sup>§</sup> pH=7-11

Table A.XIX: Calculated and experimental  $^{15}\text{N}$  chemical shifts (ppm) in cyanides relative to neat liquid nitromethane

Molecule	UKS	Loc.1	Loc.2	Loc.3	Expt.*	Solvent
MeCN	99.1	111.0	114.7	124.5	125.81	0.25M Cyclohexane
EtCN	99.8	108.0	110.7	117.1	128.22	0.25M Cyclohexane
<i>i</i> -PrCN	100.6	108.2	110.6	116.4	129.8	0.25M Cyclohexane
<i>t</i> -BuCN	102.6	112.7	115.8	129.3	131.51	0.25M Cyclohexane
Ph-CN	76.8	85.0	87.6	94.0	123.8	$\text{CDCl}_3$
<i>p</i> -CN-Ph-Me	78.9	86.6	89.1	95.5	127.0 <sup>†</sup>	solid state (MASS)
<i>p</i> -CN-Ph-CN	64.2	69.9	71.8	76.5	120.0 <sup>†</sup>	solid state (MASS)
<i>p</i> -CN-Ph-N(Me) <sub>2</sub>	81.0	88.7	91.1	97.8	129.0 <sup>†</sup>	solid state (MASS)
<i>p</i> -CN-Ph-OMe	78.8	85.7	87.9	93.7	129.0 <sup>†</sup>	solid state (MASS)
HOOCCH <sub>2</sub> CN	85.6	93.9	96.6	111.0	137.1	$\text{CDCl}_3$

\*Experimental shifts from Witanowski and coworkers<sup>193</sup>

<sup>†</sup>Uncorrected for bulk-susceptibilities

Table A.XX: Calculated and experimental  $^{15}\text{N}$  chemical shifts (ppm) in isocyanides relative to neat liquid nitromethane

Molecule	UKS	Loc.1	Loc.2	Loc.3	Expt.*	Solvent
<i>i</i> -PrNC	182.3	187.9	189.6	195.1	194.0	0.1M CH <sub>2</sub> Cl <sub>2</sub>
BuNC	199.4	204.3	205.7	209.8	210.0	0.1M CH <sub>2</sub> Cl <sub>2</sub>
<i>t</i> -BuNC	172.7	179.2	181.1	185.4	186.0	0.1M CH <sub>2</sub> Cl <sub>2</sub>
C <sub>6</sub> H <sub>12</sub> NC	182.2	187.8	189.5	194.5	197.0	0.1M CH <sub>2</sub> Cl <sub>2</sub>
CF <sub>3</sub> NC	191.5	198.4	200.4	203.1	207.6	neat liquid
Ph-CH <sub>2</sub> NC	199.2	201.8	202.7	205.1	212.0	0.1M CH <sub>2</sub> Cl <sub>2</sub>

\*Experimental shifts from Witanowski and coworkers.<sup>193</sup> Uncorrected for bulk susceptibility

Table A.XXI: Calculated and experimental  $^{15}\text{N}$  chemical shifts (ppm) in azoles relative to neat liquid nitromethane

Molecule	UKS	Loc.1	Loc.2	Loc.3	Expt. <sup>†</sup>	Solvent
Pyrrole-type nitrogen						
pyrrole	218.1	220.7	221.6	220.3	238.77 <sup>‡</sup>	cyclohexane
1-Me-pyrrole	209.7	211.7	212.4	212.6	234.0	$\text{CCl}_4$
1-NH <sub>2</sub> -pyrrole	188.9	191.3	192.1	192.9	209.8 <sup>§¶</sup>	DMSO
Indole	236.0	237.5	238.0	238.2	245.5 <sup>§</sup>	DMSO
1-Me-pyrazole	153.7	156.7	157.7	159.4	179	$\text{CCl}_4$
1-Me-indazole <sup>  </sup>	176.5	178.5	179.1	180.2	203.0 <sup>§</sup>	DMSO
2-Me-indazole	138.6	140.4	140.9	141.8	162 <sup>§</sup>	DMSO
1-Me-imidazole	199.9	202.2	203.0	202.0	219 <sup>§</sup>	DMSO
1-NH <sub>2</sub> -imidazole	178.3	180.9	181.8	181.4	198.7 <sup>§¶</sup>	DMSO
1-Me-benzimidazole	216.8	218.6	219.1	219.9	237	$\text{CCl}_4$
1-Me-1,2,4-triazole	145.1	148.8	150.0	152.5	170	$\text{CCl}_4$
1-NH <sub>2</sub> -1,2,4-triazole	125.8	129.3	130.5	133.1	155.7 <sup>§¶</sup>	DMSO
1-Me-1,2,3-triazole	114.8	119.5	121.1	125.1	145.0 <sup>§¶</sup>	$\text{CDCl}_3$
2-Me-1,2,3-benzotriazole	91.8	94.3	95.1	96.4	103.8 <sup>§</sup>	DMSO

*Continued on next page . . .*

<i>... Continued from last page</i>						
Molecule	UKS	Loc.1	Loc.2	Loc.3	Expt. <sup>†</sup>	Solvent
Pyridine-type nitrogen						
1-Me-pyrazole	31.0	38.6	40.0	45.2	68	CCl <sub>4</sub>
1-Me-indazole	18.4	24.0	25.8	30.1	57.0 <sup>§</sup>	DMSO
2-Me-indazole	45.9	49.5	50.7	52.4	92 <sup>§</sup>	DMSO
1-Me-imidazole	85.0	92.1	94.3	95.8	118 <sup>§</sup>	DMSO
1-NH <sub>2</sub> -imidazole	81.8	87.4	89.2	88.9	72.4 <sup>§¶</sup>	DMSO
1-Me-benzimidazole	92.1	96.8	98.3	101.0	127	CCl <sub>4</sub>
1-Me-1,2,4-triazole	43.0	49.6	51.7	56.1	42	CCl <sub>4</sub>
1-Me-1,2,4-triazole	90.8	97.7	99.9	103.7	113	CCl <sub>4</sub>
1-NH <sub>2</sub> -1,2,4-triazole	65.8	72.2	74.2	78.7	79.8 <sup>§¶</sup>	DMSO
1-NH <sub>2</sub> -1,2,4-triazole	95.8	103.5	106.0	110.2	131.6 <sup>§¶</sup>	DMSO
1-Me-1,2,3-triazole	-32.7	-23.5	-20.6	-13.7	16.3 <sup>§¶</sup>	CDCl <sub>3</sub>
1-Me-1,2,3-triazole	-13.8	-3.3	0.0	10.2	30.7 <sup>§¶</sup>	CDCl <sub>3</sub>
2-Me-1,2,3-benzotriazole	25.5	30.8	32.6	35.7	12.0 <sup>§</sup>	DMSO

\*In case of non equivalent nitrogen atoms, the nucleus considered is underlined

<sup>†</sup>Experimental shifts from Witanowski and coworkers<sup>194</sup> unless otherwise specified

<sup>‡</sup>Experimental shift from Witanowski and coworkers<sup>199</sup>

<sup>§</sup>Experimental shift from Witanowski and coworkers<sup>193</sup>

<sup>¶</sup>Uncorrected for bulk susceptibility

<sup>||</sup>or 1-Me-benzopyrazole

Table A.XXII: Calculated and experimental  $^{15}\text{N}$  chemical shifts (ppm) in azines relative to neat liquid nitromethane

Molecule	UKS	Loc.1	Loc.2	Loc.3	Expt.*	Solvent
pyridine	51.8	26.2	29.6	33.7	57.7	cyclohexane
2-OMe-pyridine	61.7	68.8	71.0	74.5	109.0 <sup>†</sup>	neat liquid
3-OMe-pyridine	16.3	24.9	27.7	31.9	64.0 <sup>†</sup>	Me <sub>2</sub> CO 1:3
4-OMe-pyridine	43.4	50.2	52.4	55.5	89.0 <sup>†</sup>	neat liquid
1,2-diazine	-103.7	-86.0	-80.5	-67.9	-35.3	cyclohexane
1,3-diazine	38.5	45.7	48.1	51.1	80.3	cyclohexane
1,4-diazine	4.6	16.3	20.1	26.9	42.2	cyclohexane
1,3,5-triazine	51.6	58.2	60.3	62.7	95.4	cyclohexane
1,2,4-triazine: N <sub>1</sub>	-125.1	-105.6	-99.6	-88.8	-51.2	cyclohexane
1,2,4-triazine: N <sub>2</sub>	-81.7	-66.2	-61.4	-52.2	-6.8	cyclohexane
1,2,4-triazine: N <sub>4</sub>	52.8	58.0	59.8	63.4	79.9	cyclohexane
1,2,4,5-tetrazine	-75.7	-60.7	-56.2	-44.4	-15.7 <sup>‡</sup>	cyclohexane
Quinoline	17.4	23.5	25.5	28.0	64.0 <sup>†</sup>	CCl <sub>4</sub>

\*Experimental shifts from Witanowski and coworkers<sup>205</sup> unless otherwise specified.

<sup>†</sup>Experimental shift from Witanowski and coworkers<sup>193</sup>

<sup>‡</sup>Experimental shift from Witanowski and coworkers, *J. Magn. Reson.*, **124**, 127 (1997)



Table A.XXIII: Calculated and experimental  $^{15}\text{N}$  chemical shifts (ppm) in azine N-oxides relative to neat liquid nitromethane

Molecule	UKS	Loc.1	Loc.2	Loc.3	Expt.*	Solvent
pyridine N-oxide	41.3	45.5	46.9	48.7	76.43 <sup>†</sup>	cyclohexane <sup>‡</sup>
2-OMe-pyridine N-oxide	65.9	68.8	71.7	71.7	146.0	Me <sub>2</sub> CO
3-OMe-pyridine N-oxide	42.3	45.9	47.0	49.2	91	Me <sub>2</sub> CO
4-OMe-pyridine N-oxide	59.1	62.2	63.2	65.0	103.0	Me <sub>2</sub> CO
pyridazine N-oxide	17.8	22.4	23.8	26.7	55.0	Me <sub>2</sub> CO
pyrimidine N-oxide	45.9	49.8	51.1	52.5	91.0	Me <sub>2</sub> CO
pyrazine N-oxide	32.6	37.3	38.8	40.6	68.0	Me <sub>2</sub> CO

\*Experimental shifts from Witanowski and coworkers<sup>194</sup> unless otherwise specified.

<sup>†</sup>Experimental shift from Witanowski and coworkers<sup>193</sup>

<sup>‡</sup>0.008M

Table A.XXIV: Calculated and experimental  $^{15}\text{N}$  chemical shifts (ppm) in nitrates relative to neat liquid nitromethane

Molecule	UKS	Loc.1	Loc.2	Loc.3	Expt.*	Solvent
$\text{MeNO}_2$	-15.3	-6.5	-3.7	0.1	0.0	neat liquid
$\text{PhNO}_2$	-13.5	-7.5	-5.6	-2.6	12.2	0.3M $\text{CCl}_4$
2- $\text{NO}_2$ -pyrrole	7.9	13.3	15.1	18.6	22.5	$\text{Me}_2\text{CO}$
1-Me-2- $\text{NO}_2$ -pyrrole	6.5	11.6	13.2	16.5	22.5	$\text{Me}_2\text{CO}$
3- $\text{NO}_2$ -pyrrole	-4.2	2.2	4.2	7.7	13.5	$\text{Me}_2\text{CO}$
2,4- $\text{NO}_2$ -pyrrole: 2- $\text{NO}_2$	5.4	9.3	10.6	13.2	25.7	$\text{Me}_2\text{CO}$
2,4- $\text{NO}_2$ -pyrrole: 4- $\text{NO}_2$	-3.3	0.6	1.9	3.9	18.3	$\text{Me}_2\text{CO}$
4- $\text{NO}_2$ -imidazole	0.3	5.9	7.7	10.7	16.2	DMSO
1-Me-4- $\text{NO}_2$ -imidazole	-0.8	4.8	6.6	9.9	17.0	DMSO
4,5- $\text{NO}_2$ -imidazole	11.7	15.1	16.1	17.8	28.0	DMSO
$\text{CH}_2\text{CHCH}_2\text{NO}_2$	-32.5	-23.4	-20.6	-10.2	-6.5	neat liquid
$\text{CH}_3\text{CHCHNO}_2$	-18.6	-12.6	-10.7	-7.8	2.0	DMF
$\text{CH}_3\text{CHC}(\text{CH}_3)\text{NO}_2$	-12.7	-6.5	-4.5	-1.4	-2.0	DMF

\* Experimental shifts from Witanowski and coworkers<sup>194</sup>



## **Appendix B**

### **Figures**



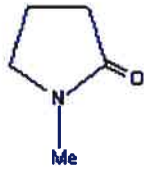
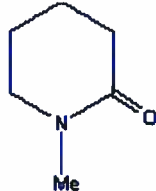
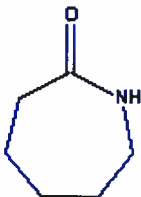
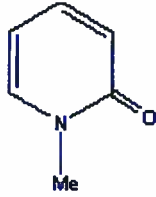
	1-Me- <i>o</i> -pyrrolidone		1-Me- <i>o</i> -piperidone
	$\epsilon$ -caprolactam		1-Me- <i>o</i> -pyridone

Figure B.1: Structure of some of the cyclic amides examined in Tab. A.XVIII.


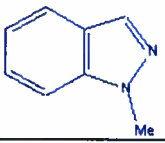
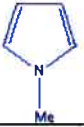
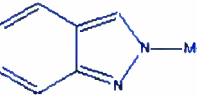

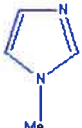


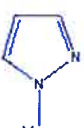
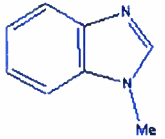
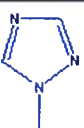

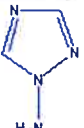
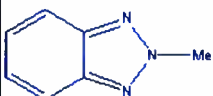
	pyrrole		1-Me-indazole
	1-Me-pyrrole		2-Me-indazole
	1-NH <sub>2</sub> -pyrrole		1-Me-imidazole
	indole		1-NH <sub>2</sub> -imidazole
	1-Me-pyrazole		1-Me-benzimidazole
	1-Me-1,2,4-triazole		1-Me-1,2,3-triazole
	1-NH <sub>2</sub> -1,2,4-triazole		2-Me-1,2,3-benzotriazole

Figure B.2: Structure of the azoles examined in Tab. A.XXI.



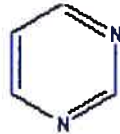
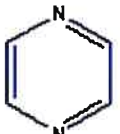

	pyridine		1,2-diazine (pyridazine)
	1,3-diazine (pyrimidine)		1,4-diazine (pyrazine)
	quinoline		

Figure B.3: Structure of some of the azines examined in Tab. A.XXII.

