AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

Thèse en vue de l'obtention du grade de
Docteur de l'Université de Lorraine
Mention Chimie

Development and Application of Methods Based on Extremely Localized Molecular Orbitals

Benjamin MEYER

École Doctorale SESAMES

Collegium Sciences et Technologies

Laboratoire Structure et Réactivité des Systèmes Moléculaires Complexes

Thèse présentée et soutenue publiquement le 10 Octobre 2016

Composition du Jury

Rapporteurs :		
Arianna Fornili	Research Fellow	University of London
Simon Grabowsky	Professor	University of Bremen
Examinateurs :		
Julia Contreras-Garcia	Chargé de recherche	Université Pierre et Marie Curie
Benoit Guillot	Professeur	Université de Lorraine
Directeur de thèse		
Manuel Ruiz-Lopez	Directeur de recherche	Université de Lorraine
Co-directeur de thèse		
Alessandro Genoni	Chargé de recherche	Université de Lorraine

À Justine

À ma famille

Acknowledgements/Remerciements

At first, I would like to express my gratitude to the jury members who have accepted to evaluate this manuscript. I am especially grateful to Prof. Arianna Fornili and Simon Grabowsky for having reviewed this work.

Je tiens à remercier très chaleureusement les docteurs Manuel Ruiz-Lopez et Alessandro Genoni pour leur accompagnement et leur soutien. J'ai vraiment eu la chance de travailler avec deux directeurs de thèse très complémentaires qui m'ont appris énormément de choses durant ces 3 années. Ce manuscrit est le fruit de leur constante disponibilité à répondre à mes nombreuses sollicitations. Merci à tous les deux pour votre gentillesse, votre patience et pour ces beaux moments partagés.

Je souhaite remercier tous les membres permanents et non-permanents du Laboratoire SRSMC pour m'avoir permis de travailler dans une ambiance chaleureuse. Il m'est malheureusement impossible d'être exhaustif, mais j'aimerais particulièrement remercier le professeur Jean-Louis Rivail avec qui ce fut toujours un plaisir de discuter lors de ses visites au laboratoire ainsi que le professeur Xavier Assfeld pour tous ces formidables moments passés ensemble. Mes pensées vont également vers les professeurs Gérald Monard, Claude Millot, Ugo Ancarani, Claude Dal Cappelo, Jean-Bernard Regnouf de Vains ainsi que les docteurs Mariachiara Pastore Franscesca Ingrosso, Alexandrine Lambert, Marilia Martins-Costa, Nadia Canhilo, Andreea Pasc, Antonio Monari, Sébastien Lebègue, Christophe Chipot, François Dehez, Mounir Tarek, Fabien Pascale, Arnaud Leclerc...

J'aimerais également remercier tous membres du laboratoire CITHEFOR pour leur accueil et leur disponibilité durant mes trois années de monitorat à la faculté de pharmacie. Je tiens tout particulièrement à remercier les professeurs Pierre Leroy et Igor Clarot ainsi que les docteurs Ariane Boudier, Caroline Gaucher et Marianne Parent pour leur sympathie et tous les excellents moments passés à les écouter discuter de galénique, culture cellulaire ou d'expérimentations animales. De plus, ce fut un réel plaisir de collaborer avec cette équipe lors de projets en marge de mon travail principal qui sera exposé dans ce manuscrit.

J'aimerais aussi exprimer mes sincères reconnaissances à un bon nombre de personnes qui m'auront soutenu tant humainement que scientifiquement : Philippe Gros, Marco Marazzi, Julian Garrec, Oleksandr Loboda, Yann Cornaton, Julien Eng, Tim Krah, Alex Domingo, Pierre-François Loos, Yohann Moreau, Philippe Carbonnière, Maura Casciola, Marina Kasimova, Audrey, Pablo, Grace, Renaud, Guillaume, Jérôme, Margaux, les membres de l'association phi science...

Bien évidemment, j'ai une très grosse pensée envers mes plus proches compagnons. Je tiens à remercier très chaleureusement Antoine Marion et Ilke Ugur, Daniel Bonhenry, Thibaud Etienne et Hugo Gattuso pour leur complicité, leur aide, leur soutient... mais surtout pour leur amitié!

J'ai une pensée toute particulière pour Séverine Bonenberger qui m'a bien aidé pour toutes les démarches administratives mais surtout qui m'a soutenu moralement durant ces trois années.

Je tiens aussi à remercier le professeur Vincent Robert pour son encadrement lors de mon stage de master ainsi que tout le personnel du laboratoire de chimie quantique de Strasbourg: Chantal, Sylvie, Paola, Christophe, Emmanuel, Étienne et Roberto

L'accomplissement de cette thèse n'aurait jamais été possible sans le soutient infaillible de mes amis et de ma famille.

Pour commencer, j'aimerais remercier de tout mon cœur la personne avec qui je partage ma vie depuis plusieurs années, Justine. Merci mon amour pour ton soutient sans faille dans tous les moments difficiles que j'ai passé durant ces trois années. Je tiens également à te remercier d'avoir concédé la poursuite de mon sport en dépit du peu de temps que nous passions ensemble. D'une manière générale, tout simplement merci d'avoir accepté de poursuivre ce bout de chemin, toi et moi, malgré l'éloignement.

J'aimerais aussi remercier mes parents Sylvie et Claude, ainsi que Corinne et Yves. Merci Maman pour ta gentillesse, pour ton soutient dans les bons comme dans les mauvais moments et merci pour tout ce que tu as fait pour moi avant et pendant cette thèse. Merci Papa pour ta joie de vivre, ton optimisme et ton dynamisme que tu m'as transmis. J'aimerai également exprimer mes remerciements à toute ma famille et belle famille: Papy René, Mamie Françoise, Christine et Pierre, Sandrine et Philippe, Magali, Fabien, Timéo et Antonin, Philippe, Renate, Martine, Audrey et Jérémy, Bernard, Chantal, Olivier et sa famille, Monique et Mélanie, les parents, grands-parents ainsi que la sœur de Justine, Annick et Jean-Pierre, Clothilde, Simone, Jaques et Mamie Anna.

Je pense très fort à mon oncle Marcel et à ma grand-mère Yvonne qui ne sont plus là et dont la présence me manque beaucoup.

Je tiens à remercier tous les membres du Vélo Club de Dorlisheim et tous mes partenaires d'entraînements. Pour tous les bons moments que nous avons passés ensemble, merci à mes trois supers cousins Mathias, Thibaut et Quentin. Je tiens également tout particulièrement à remercier ce dernier pour avoir concédé de jouer avec un coéquipier éloigné et très peu présent.

Durant ces trois années j'ai partagé bon nombre de beaux moments au côté de mes amis notamment lors de sorties trails! Je souhaite vivement remercier : Mathilde, Jerem et le petit Gauthier, Joséphine et Thierry, Anthony et Julie, Franz et Julie. Je tiens également à remercier tout particulièrement Alex et Léa, avec qui je me suis lié d'amitié durant cette thèse, pour leur accueil lors de mes derniers mois de thèse ainsi que pour tous les moments (sportifs ou autour d'un bon dîner) conviviaux passés à leur côté.

Enfin je tenais à exprimer mes profonds regrets pour toutes les personnes que j'ai omis de citer dans ces quelques lignes. Donc à toutes celles et ceux que j'ai oublié j'aimerai tout de même dire : MERCI !!!

Contents

Acknowledgements/Remerciements iii

Introduction 1

I The transferability of the extremely localized molecular orbitals

1 Linear Scaling Methods 9
1.1 Introduction . . . 11
1.2 The Divide and Conquer method and the Molecular Tailoring Approach . . . 14
1.3 The Fragment Molecular Orbital method . . . 23
1.4 The Additive Fuzzy Density Fragmentation approaches . . . 25

2 Localized Molecular Orbitals 31
2.1 Introduction . . . 33
2.2 Unitary Transformation Methods . . . 34
2.3 Extremely Localized Molecular Orbitals: the Stoll technique . . . 37
2.4 Rotation . . . 42

3 Model molecule approximation and ELMOs transferability 49
3.1 Introduction . . . 51
3.2 Methods . . . 52
3.2.1 General strategy . . . 52
3.2.2 Description of the target system . . . 52
3.2.3 Model molecules approximation . . . 53
3.2.4 Computational methods . . . 54
3.2.5 Comparison of the obtained electron densities . . . 56
3.3 Results and discussion . . . 58
3.3.1 Accuracy of the different model molecule approximations . . . 58
3.3.2 Effects of the molecular orbitals localization . . . 64
3.3.3 Effects of basis sets . . . 67
3.4 Conclusions . . . 72

4 A comparison with the pseudoatoms transferability **75**
4.1 Introduction . . . 77
4.1.1 The Independent Atom Model . . . 78
4.1.2 The Hansen and Coppens multipole model . . . 79
4.1.3 The pseudoatoms . . . 80
4.1.4 Beyond atomistic models . . . 83
4.2 Computational Details . . . 84
4.2.1 Methods . . . 84
4.2.2 Comparison of the electron densities . . . 86
4.3 Results and discussion . . . 87
4.3.1 Topological properties at the bond critical points . . . 87
4.3.2 Net atomic charges . . . 96
4.3.3 Similarity Indexes . . . 98
4.4 Conclusions . . . 101

Conclusions of Part I, future directions **103**

II Assessing the capabilities of the X-ray constrained wave function methods

5 Experimental Wave Function Methods **111**
5.1 Introduction . . . 112
5.1.1 Motivations for deriving experimental constrained wave functions . . . 112
5.1.2 Experimentally constrained wave function methods . . . 113
5.1.3 The Jayatilaka approach and beyond . . . 116
5.2 Basic assumptions of the X-ray constrained wave function method . . . 117
5.3 The X-ray constrained Hartree-Fock method . . . 119
5.4 The X-ray constrained ELMO technique . . . 122
5.5 Details about the X-ray constrained calculations . . . 125
5.6 ELMO based Valence Bond method . . . 127
5.7 Conclusion . . . 130

6 Effects of the *a priori* localization on the X-Ray constrained wavefunction **131**
6.1 Motivations . . . 132
6.2 Methods . . . 132
6.2.1 A case study: the L-Alanine . . . 132
6.2.2 Computational details . . . 134
6.2.3 Comparison of the obtained charge densities . . . 134
6.3 Results and Discussion . . . 134
6.3.1 Performance and convergence of the methods . . . 134
6.3.2 Effects of the localization on the topological properties . . . 139
6.4 Conclusions . . . 144

7 Extraction of electron correlation effects from X-ray diffraction data **145**
7.1 Introduction . . . 147
7.2 Methods . . . 148
7.2.1 General Strategy . . . 148
7.2.2 Computational details . . . 148
7.3 Results and discussion . . . 149
7.3.1 Comparison between RHF and CCSD structure factors . . . 149
7.3.2 Effects on some topological properties . . . 151
7.3.3 Similarity indexes . . . 154
7.3.4 Attachment and detachment densities . . . 156
7.4 Conclusions . . . 158

8 A theoretical study of the Biscarbonyl[14]annulene charge density **161**
8.1 The context of the study . . . 162
8.2 The Biscarbonyl[14]annulene . . . 163
8.3 Methodology . . . 165
8.4 Results and discussion . . . 167
8.4.1 Unconstrained ELMO calculations . . . 167
8.4.2 Unconstrained and X-ray constrained "ELMO-VB" calculations. . . . 168
8.4.3 Determination of the Wiberg Bond Index . . . 173
8.5 Conclusions . . . 175

Conclusions of Part II, future directions **177**

Appendix **183**

A Model molecule approximation and ELMOs transferability **183**

B A comparison with the pseudoatoms transferability **191**

C Extraction of electron correlation effects on the electron density from X-Ray diffraction data **203**

D A theoretical study of the Biscarbonyl[14]annulene **207**

List of Figures **211**

List of Tables **214**

List of Publications **221**

Bibliography **222**

Résumé **234**

Introduction

The research described in this thesis aimed at achieving two main goals: 1) evaluating in detail the transferability of the so-called Extremely Localized Molecular Orbitals (ELMOs) in order to afterwards develop new ELMO-based linear-scaling techniques; 2) assessing the capabilities of the X-ray constrained wave function approaches, especially (but not exclusively) when the ELMOs are used in that context.

Concerning the first point, it is worth noting that, nowadays, since computational chemistry methods are becoming more and more important in a large number of fields, the size of the systems to be examined is also increasing. This obviously stimulates the development of novel computational strategies to investigate very large systems at a reasonable computational cost. To overcome this problem, many research groups have developed linear scaling techniques that are approximate quantum-mechanical methods whose computational cost scales linearly with the increasing size of the system. Among them, a prominent position is occupied by those strategies that exploit the transferability principle, namely the observation that molecular systems are composed by recurrent functional units that generally keep their features when they are in a similar chemical environment. In this context, we can imagine a simple procedure through which it is possible to take advantage of the transferability of molecular orbitals strictly localized on small molecular subunits to recover wave functions and electron densities of large systems. Unfortunately, the molecular orbitals traditionally used in quantum chemistry are completely delocalized on the system in exam and this prevents their transfer from a molecule to another. This problem can be solved only considering molecular orbitals extremely localized on small molecular fragments: the ELMOs. As we will see, in the next chapters, these orbitals are variationally determined under the constraint of expanding them on subsets of basis functions (local basis sets) associated with pre-determined molecular fragments. Due to their strict localization, these orbitals are in principle transferable from molecule to molecule and our final goal is to construct databanks of ELMOs that will enable to recover almost instantaneously approximate wave functions and electron densities of macromolecules at a very low computational cost.

The first part of the manuscript focuses on this topic. In particular, after a brief overview of the most popular linear scaling methods (Chapter 1), in the second Chapter the concept of Localized Molecular Orbital will be fully examined. At first, some traditional Localized Molecular Orbitals (LMOs) methods consisting in unitary transformations of the canonical Hartree-Fock Molecular Orbitals will be presented. Afterwards, we will introduce the Stoll technique that can be considered as the first reliable generalisation of the Hartree-Fock equations to obtain Extremely Localized Molecular Orbitals and that is the strategy that we have adopted to compute ELMOs throughout the thesis. Finally, in the last section of the Chapter, the strategy exploited for the rotation transfer of the ELMOs will be discussed in detail.

In the third Chapter we will describe our real investigations on the ELMOs transferability. In particular, the most suitable level of approximation for the construction of model molecules, the effects of using transferred ELMOs and the effects of changing the atomic basis-set in the reconstruction of macromolecular electron densities will be shown and discussed.

Finally, in Chapter 4, the transferability of the ELMOs will be compared to the one of the pseudoatoms, which are aspherical atomic electron densities largely used in crystallography to refine crystallographic structure and to reconstruct electron distributions of very large macromolecules.

In the second part of the thesis, the focus will shift toward the X-ray constrained wave function approach which can be considered as an alternative way to investigate the electronic structure of many-electron systems, although it is not very well-know in the theoretical chemistry community yet. This method basically consists in finding single Slater determinants that not only minimize the electronic energy of the systems under exam, but that also reproduce sets of experimental structure factors amplitudes within a desired accuracy. The technique, initially proposed by Jayatilaka in the framework of the Hartree-Fock machinery, has been recently extended to the theory of the Extremely Localized Molecular Orbitals and it is in this last context that the X-ray constrained wave function approach will be mainly analysed and discussed in the last four chapters of the manuscript.

Initially, in Chapter 5, after a historical overview of some experimentally constrained wave function/density matrix techniques, we will present the main assumptions and

equations of the original X-ray constrained Hartree-Fock strategy. Afterwards, the theory of the recent X-ray constrained ELMO technique and of a prototype of X-ray constrained ELMO-based Valence Bond strategy will be also shown. In Chapter 6, we will present the investigations conducted in order to evaluate the effects of introducing a strict *a priori* localization on the electronic structure in X-ray constrained wave function calculations, while, in Chapter 7, we will try to prove one of the hypotheses associated with the Jayatilaka approach since its first appearance, namely the fact that the X-ray constrained wave function is intrinsically able to capture the electron correlation effects on the electron densities.

Finally, in Chapter 8 we will report our theoretical studies conducted on the *syn*-1,6:8,13-Biscarbonyl[14] annulene (BCA). In particular, we will describe and discuss both the traditional unconstrained ELMO calculations and the X-ray constrained ELMO-VB computations that have been performed to investigate the partial rupture of the aromatic character of the BCA molecule at high pressure, which has been recently observed through the analysis of accurate X-ray diffraction measurements.

At the end of this manuscript, we will present some general conclusions as well as some future directions of research concerning the use of the extremely localized molecular orbitals (ELMOs) in the field of theoretical chemistry.

Introduction

Les recherches menées dans le cadre de cette thèse avaient un double objectif. Premièrement, le développement d'une nouvelle méthode de chimie quantique à croissance linéaire basée sur le concept d'Orbitales Moléculaires Extrêmement Localisées (ELMOs) et adaptée à l'étude de très gros systèmes moléculaires. Deuxièmement, il s'agit d'évaluer le potentiel des méthodes de calcul utilisant de fonctions d'ondes contraintes et leur capacité à reproduire des données de diffraction aux rayons-X. En ce qui concerne le premier objectif, notre approche se base sur le principe de transférabilité, à savoir l'observation que les systèmes moléculaires sont composés par des unités fonctionnelles récurrentes qui conservent leurs caractéristiques lorsqu'elles se trouvent dans un même environnement chimique. Malheureusement, les orbitales moléculaires traditionnellement employées en chimie théorique dans des modèles de particule indépendante (Hartree-Fock, Kohn-Sham) sont complètement délocalisées sur le système étudié et, par conséquent, ne peuvent pas être transférées d'une molécule à une autre. Ce problème peut être résolu en ayant recours à des orbitales moléculaires déterminées de manière variationnelle sous la contrainte d'être exprimées à partir des fonctions de base centrées sur des atomes de fragments présélectionnés : les ELMOs. En fait, puisqu'elles sont strictement localisées, ces orbitales sont en principe transférables d'une molécule à une autre. L'objectif à terme est d'exploiter cette transférabilité en construisant une base de données d'ELMOs permettant de calculer quasiment instantanément, de manière approximative, des fonctions d'ondes et des densités électroniques de macromolécules.

La première partie de cette thèse est consacrée à ce sujet. En particulier, après une brève présentation des méthodes à croissance linéaire les plus connues (Chapitre 1), dans le deuxième chapitre, nous examinerons le concept des orbitales moléculaires localisées. Pour commencer, nous présenterons certaines méthodes de localisation qui utilisent des transformations unitaires des orbitales canoniques d'Hartree-Fock. Ensuite, nous introduirons la technique de Stoll qui peut être considérée comme la première généralisation des équations d'Hartree-Fock pour obtenir des orbitales moléculaires extrêmement localisées. Cette stratégie est celle que nous avons adoptée tout au long de cette thèse pour générer les ELMOs. Enfin, dans la dernière partie de ce chapitre, nous reviendrons en détail sur la méthode que nous avons utilisée pour le transfert des ELMOs.

Dans le troisième chapitre, nous décrirons nos études sur la transférabilité des ELMOs. Plus particulièrement, nous chercherons le niveau d'approximation le plus approprié pour la construction des molécules modèles. De plus, nous étudierons les effets de

la localisation et les effets du changement de base dans la reconstruction de densités électroniques de macromolécules.

A la fin de cette première partie, dans le Chapitre 4, la transférabilité des ELMOs sera comparée avec celle des pseudo-atomes, qui sont des densités électroniques atomiques asphériques largement répandues dans le domaine de la cristallographie pour affiner des structures cristallographiques et reconstruire des densités électroniques de très grandes molécules.

Dans la seconde partie de cette thèse, nous nous concentrerons sur l'approche de la fonction d'onde contrainte et sa capacité à reproduire des données de diffraction aux rayons-X. Cette approche peut être considérée comme une méthode alternative aux calculs *ab initio* pour l'étude de systèmes moléculaires. Elle repose essentiellement sur la recherche d'un déterminant de Slater qui doit non seulement minimiser l'énergie électronique du système, mais également reproduire au mieux un jeu d'amplitudes de facteurs de structure expérimentaux. Cette technique, initialement développée par Jayatilaka dans le cadre de la théorie Hartree-Fock, a ensuite été étendue à la théorie des orbitales moléculaires extrêmement localisées. C'est dans ce contexte que nous analyserons et discuterons le comportement de ces méthodes dans les quatre derniers chapitres de ce manuscrit.

Dans un premier temps, dans le Chapitre 5, après une présentation historique de quelques approches utilisant des fonctions d'ondes contraintes, nous présenterons les principales approximations et les équations de la méthode Hartree-Fock contrainte. Puis, nous décrirons la technique des ELMOs contraintes ainsi que d'un prototype de méthode Valence Bond « expérimentale » basée sur les ELMOs. Dans le Chapitre 6, nous présenterons nos études sur l'évaluation des effets de localisation stricte de la structure électronique dans les calculs de fonction d'onde contrainte. Dans le Chapitre 7, nous essaierons de prouver l'une des premières hypothèses énoncées par Jayatilaka, à savoir que sa méthode capture de manière intrinsèque des effets de la corrélation sur la densité électronique.

Finalement, dans le Chapitre 8, nous allons reporter nos études théoriques effectuées sur le syn-1,6:8,13-Biscarbonyl[14] annulène (BCA). Nous reviendrons notamment en détail sur notre stratégie mise en place pour étudier la rupture partielle de l'aromaticité du BCA à haute pression qui a été observé par des mesures précises de diffraction aux rayons-X.

Ce manuscrit s'achève par une conclusion générale sur les recherches menées dans le

cadre de cette thèse et quelques perspectives pour l'utilisation des orbitales moléculaires extrêmement localisées, ELMOs, dans des domaines propres à la chimie théorique.

Part I

The transferability of the extremely localized molecular orbitals

Chapter 1

Linear Scaling Methods

Résumé

De nos jours, la chimie théorique est devenue un outil indispensable dans de nombreux domaines scientifiques et techniques pour la prédiction de structures et de propriétés moléculaires, la conception de nouveaux médicaments ou l'étude de matériaux innovants. La taille des systèmes étudiés est par conséquent en augmentation constante. Développer et appliquer de nouvelles stratégies pour l'étude de très grands systèmes tout en conservant un temps de calcul raisonnable est donc un objectif prépondérant de la chimie théorique actuelle.

Avec l'amélioration constante des algorithmes et l'augmentation de la puissance des ordinateurs, il est désormais possible d'utiliser des techniques de calcul *ab initio* (Hartree-Fock ou Théorie de la Fonctionnelle de la Densité) pour des systèmes allant jusqu'à plusieurs centaines d'atomes. En revanche, un traitement quantique complet sur une macromolécule comme une protéine reste un défi. Afin de surmonter ce problème, de nombreux groupes de recherche ont développé des approches dites à croissance linéaire, pour lesquelles, le temps de calcul est linéairement proportionnel à la taille du système étudié.

Dans ce premier chapitre, qui est dédié aux méthodes à croissance linéaire, nous avons distingué trois grandes familles de méthodes. La première adopte la philosophie "Divide & Conquer" et se base sur la partition d'une grande molécule en petit fragments. Ces sous-unités sont ensuite caractérisées par des Hamiltoniens locaux qui vont permettre de traiter séparément la structure électronique de chaque fragment avant de reconstituer la densité électronique du système global. Une autre grande famille est représentée par les techniques d'interaction de fragments. Dans ce groupe d'approches, nous nous attarderons plus bas sur la méthode FMO (Fragment Molecular Orbital). Finalement, la troisième famille importante est constituée par des méthodes où l'on utilise le principe

de transférabilité. Elles se basent sur l'observation que les systèmes moléculaires sont souvent constitués de groupes fonctionnels récurrents et que ceux-ci conservent leurs caractéristiques d'une molécule à une autre si ces molécules se trouvent dans un même environnement chimique. Nous reviendrons tout particulièrement sur les approches de type LEGO initialement introduites par Walker et Mezey.

1.1 Introduction

In recent years, computational chemistry has become an indispensable tool in a large number of scientific and technological fields (e.g. prediction of molecular structures and properties, design of pharmaceutical drugs and novel materials, etc). As a consequence, the size of the systems which have to be examined is continuously increasing, thus generating further theoretical and computational needs. In fact, developing and applying novel computational strategies to investigate very large systems at a reasonable computational cost, has become a major aim in computational chemistry. Accomplishing this task will enable to get deeper insights into fundamental problems and phenomena involving macromolecules or extended disordered systems, as found for instance in biochemistry, biotechnology or nanotechnology.

As well known, in modern quantum chemistry, the Hartree-Fock (HF) (*1*–*5*) and Density Functional Theory (DFT) (*6*) methods are considered as the basic approaches. The former consists in finding the single Slater determinant that minimizes the energy of the system under investigation and it usually represents the starting point for more elaborated many-determinant techniques that try to approach the exact solution of the Schrödinger equation. DFT is a valid and popular alternative to HF method that, relies on the Hohenberg and Kohn theorem (*7*) and, in principle, has the great advantage of using only the three-dimensional electron density as main physical entity instead of the more complicated multi-variable wave function. Nevertheless, the exact functional relation between the ground-state electron density and the ground-state energy of an electronic system is unknown. Therefore, over the years, by exploiting the Kohn-Sham approach (*6*), a multitude of approximate DFT method have been developed. These approximate methods currently offer a good compromise between accuracy and computational cost and, therefore, nowadays they have become standard tools.

However, the rapid growth of the computational cost as a function of the system size, even at the Hartree-Fock and DFT levels, remains one of the main difficulties associated with the application of quantum chemical methods. For instance, the HF and DFT approaches scale as $O(M^3)$, where M is a relative measure of the size of the system in exam. This means that when we study a molecule 10 times larger compared to a reference one, the computational effort increases by a factor of 1000. Of course, this scaling becomes

even more dramatic when using methods which directly deal with the electron correlation, such as the MP2 or CCSD techniques, for which the logout time increases as $O(M^5)$ and $O(M^6)$, respectively (see Figure 1.1 for a schematic comparison of the $O(M)$, $O(M^3)$ and $O(M^5)$ scalings).

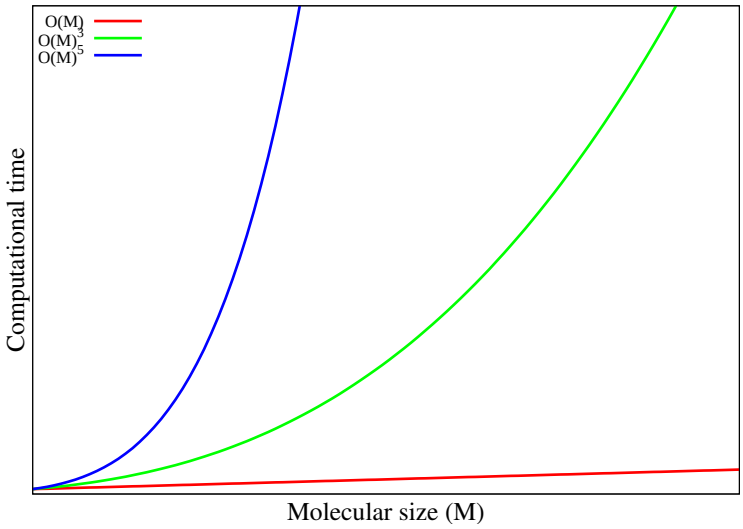

FIGURE 1.1: Schematic comparison of different scaling behaviors

Fortunately, the continuous improvement of algorithms and the advances in computer technologies have allowed electronic structure calculations on larger and larger systems. Thus, *ab initio* calculations at HF or DFT levels are nowadays affordable for molecules constituted by hundreds of atoms (*8*–*12*) using a reasonable size of basis sets. However, a full quantum mechanical treatment of larger systems remains a big challenge and, to overcome this drawback, many research groups have been involved in the development of linear scaling $O(M)$ techniques, for which the computational cost scales linearly with the increasing size of the system (*13*–*16*). Nowadays, through these new strategies it is possible to carry out calculations on systems composed by thousands of atoms, such as very large proteins and large portions of nucleic acids.

One of the most well-known families of linear-scaling strategies is represented by the so-called "Divide & Conquer" (DC) philosophy which is currently implemented in the framework of DFT, HF and semiempirical approaches (*17*–*27*). Initially proposed by Yang

in 1991 (*17*), it is based on the partition of large molecules into small overlapping fragments. The subunits are characterized by local Hamiltonians that are projections of the global one on each subsystem and the electronic problem is therefore solved for each fragment. Afterwards, the global electron density of the molecule is properly reconstituted. In the same context, we can also mention the Molecular Tailoring Approach (MTA), that can be seen as a fragment-based linear scaling technique developed for *ab initio* calculations at HF or MP2 levels. In an analogous way to the Divide and Conquer approach, at first, small subunit density matrices are obtained and, at a later stage, they are properly recombined to obtain the global density matrix.

Another important family of linear-scaling methods is represented by the fragment interaction techniques. Although these approaches are based on a partition into different subunits as well, they differ from the previous ones by the use of the theory of molecular interactions to write the total energy of the systems, which becomes a sum of monomer energies and intermolecular interaction energies corresponding to dimers and larger molecular assemblies. In this group of methods it is worth mentioning the MFCC (molecular fractionation with conjugated caps) approach(*28*, *29*) originally devised by Zhang and Zhang and more recently improved by Li and co-workers through an energy-corrected formalism (EC-MFCC technique)(*30*). However, the prototypes of this family of techniques are probably the fragment molecular orbital (FMO) (*31*–*37*) and the kernel energy methods (KEM) (*38*–*41*). The FMO technique includes both the effects of the electrostatic field of the whole system and the exchange interactions with the other fragments in all the subunits calculations. Its use has rapidly grown in a few years now, FMO it is recognized as a useful technique to investigate the electronic structure of very large and complex molecules and molecular clusters. In KEM, one subdivides large molecules into small fragments capped with hydrogen atoms. The method relies on an equation that fully considers self- and two-body interaction terms, and it has turned out to be an extremely useful strategy for the study of macromolecules, ranging from small proteins to graphene and extended aromatics, even in the presence of external electric fields. The reliability of the kernel energy method has been confirmed by recent investigations (*42*, *43*) showing that this technique can accurately reproduce electron densities, electrostatic potentials, QTAIM (quantum theory of atoms in molecules (*44*)) localization and delocalization matrices, and QTAIM charges of very large molecules.

Finally, a third prominent family of the $O(M)$ techniques, is represented by methods

that exploit the transferability principle. The latter is based on the observation that molecular systems are composed by recurrent functional units that generally conserve their features when they are in a similar chemical environment. In this group, a first important and pioneering example is the molecular electron density LEGO assembler (MEDLA) technique proposed by Walker and Mezey (45–49). In fact, using a database of fuzzy electron densities associated with small molecular fragments, this method has enabled to successfully reconstruct charge distributions of very large systems at *ab initio* level at reduced computation efforts. Subsequently, this strategy has been extended by Exner and Mezey leading to the adjustable density matrix assembler (ADMA) approach (50–54). This procedure uses a library of fragment density matrices, which allow fast calculations of electron densities and electrostatic potentials of very large systems. Another method, closely related the strategy proposed by Mezey and co-workers is the TAE (Transferable Atom Equivalents) (55) technique proposed by Breneman. The TAE method is based on the retrieval of a set of appropriate atomic electron density representations from a library, followed by a self-consistent assembly process in which each atom is slightly adjusted to its new environment.

After this non exhaustive overview of linear scaling methods, in the following sections we will briefly discuss some specific techniques. In particular, we will consider the Divide & Conquer strategy, the Molecular Tailoring Approach, the Fragment Molecular Orbital method and the approaches based on the Additive Fuzzy Density Fragmentation, namely the MEDLA and ADMA techniques.

1.2 The Divide and Conquer method and the Molecular Tailoring Approach

In 1637, the famous French philosopher René Descartes published "Le Discours de la méthode" and established four different precepts to follow a scientific approach. One of them is: "Diviser chacune des difficultés que j'examinerais, en autant de parcelles qu'il se pourrait et qu'il serait requis pour les mieux résoudre. - Divide each of the difficulties under examination into as many parts as possible and as might be necessary for its defining adequate solution." This precept could be seen as the statement of the Divide & Conquer approach (DC). The significance of this idea is widespread in all the domains of science, especially in computer science where the DC method has been recognized as a

fundamental technique for devising low-scaling efficient algorithms. Following this procedure, a tricky problem is divided into two or more smaller ones, the solutions of which are them recombined to obtain the answer to the original question.

The DC philosophy has been incorporated by Yang into the field of electronic structure calculations (*17*, *18*). In the original version, the method uses the density as the basic variable, divides a system into subsystems and determines the density for each subsystem to build up an approximate overall electron distribution. If the original paper in 1991 only treated the simple N_2 molecule, Yang has afterwards improved the DC method by introducing the concept of buffer zones in order to obtain better representations of the subunits density and to decrease the truncation error. The applicability of this method was then extended to other larger systems conserving a reasonable accuracy (*19*, *20*). Another important step in the breakthrough of the DC strategy was accomplished by Yang and Lee (*21*) extending this method to the one-electron density matrix, which is especially interesting since the density matrix could be obtained through Hartree-Fock or semiempirical calculations. Nevertheless, the extension to the HF theory has been limited by the requirement of the non-local HF exchange computation. At a later stage, the density matrix based DC approach was considered in the framework of the semiempirical calculations by York, Lee and Yang (*22*–*24*) and independently also by Dixon and Merz (*25*) to treat biological macromolecules with solvent effects. In particular, the latter introduced inner and outer buffer regions in order to reduce errors in the energy calculations (*26*).

In addition to the methods mentioned above, several groups all over the world have suggested different variants and improvements of the DC philosophy, which will not be discussed here. However, for the sake of completeness, in the next paragraphs we will focus our attention on the equations at the basis of the original formulation of the DC method.

To this purpose, let us consider a closed shell system of $2N$ electrons in an external field $\upsilon(\mathbf{r})$ which, in case of a molecular system, is the nuclear potential acting on the electrons. As well known, we can write the total energy of the system in terms of the total electron density $\rho(\mathbf{r})$ as

$$E[\rho] = T_s[\rho] + \int_{\mathbb{R}^3} \upsilon(\mathbf{r})\rho(\mathbf{r})\,\mathrm{d}\mathbf{r} + E_{xc}[\rho] + \frac{1}{2}\int_{\mathbb{R}^3} \frac{\rho(\mathbf{r})\rho(\mathbf{r'})}{|\mathbf{r}-\mathbf{r'}|}\,\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r'} \tag{1.1}$$

where $T_s[\rho]$ is the Kohn-Sham kinetic energy of a noninteracting electron gas in its ground

state with density ρ and $E_{xc}[\rho]$ is the exchange-correlation energy. The total electron density is expressed in terms of a set of orbitals

$$\rho(\mathbf{r}) = 2\sum_{i}^{N} \varphi_i^*(\mathbf{r})\varphi_i(\mathbf{r}) \tag{1.2}$$

which allows the determination of the kinetic energy as

$$T_s[\rho] = 2\sum_{i}^{N} -\frac{1}{2}\int_{\mathbb{R}^3} \varphi_i^*(\mathbf{r})\nabla_r^2\varphi_i(\mathbf{r})\,\mathrm{d}\mathbf{r} \tag{1.3}$$

Then, the minimization of the energy functional $E[\rho]$ with respect to the electron density is achieved solving the Kohn-Sham equations (6):

$$\hat{H}_{KS}\varphi_i(\mathbf{r}) = [-\frac{1}{2}\nabla^2 + V_{eff}(\mathbf{r})]\varphi_i(\mathbf{r}) = \epsilon_i\varphi_i(\mathbf{r}) \tag{1.4}$$

where $V_{eff}(\mathbf{r})$ is the Kohn-Sham effective local potential,

$$V_{eff}(\mathbf{r}) = \upsilon(\mathbf{r}) + \int_{\mathbb{R}^3} \frac{\rho(\mathbf{r'})}{|\mathbf{r}-\mathbf{r'}|}\mathrm{d}\mathbf{r'} + V_{xc}(\mathbf{r}) \tag{1.5}$$

with

$$V_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[\rho]}{\delta\rho(\mathbf{r})} \tag{1.6}$$

Now, the total energy of the system expressed in (1.1) can be simply rewritten in terms of the Kohn-Sham eigenvalues:

$$E[\rho] = 2\sum_{i}^{N} \epsilon_i + Q[\rho] \tag{1.7}$$

where $Q[\rho]$ is an electron density functional expressed as:

$$Q[\rho]) = \int_{\mathbb{R}^3} \rho\left[\frac{-\phi(\mathbf{r})}{2 - V_{xc}(\mathbf{r})}\right]\mathrm{d}\mathbf{r} + E_{xc}[\rho] \tag{1.8}$$

and

$$\phi(\mathbf{r}) = \int_{\mathbb{R}^3} \frac{\rho(\mathbf{r'})}{|\mathbf{r}-\mathbf{r'}|}\mathrm{d}\mathbf{r'} \tag{1.9}$$

is the electrostatic potential due to the electrons.

Given the previous conventional DFT formulation, the aim of the DC approach is to bypass the Kohn-Sham equations (1.4) and to compute the electron density without using

the complete set of N orbitals as in equation (1.2). To this purpose, it is possible to rewrite equation (1.2) as:

$$\rho(\mathbf{r}) = 2\langle \mathbf{r}|\eta(\epsilon_F - \hat{H}_{KS})|\mathbf{r}\rangle \tag{1.10}$$

where $\eta(x)$ is the Heaviside step function, namely

$$\eta(x) = \begin{cases} 1 & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x \leq 0 \end{cases} \tag{1.11}$$

and ϵ_F can be any value between the highest occupied and the lowest unoccupied eigenvalues. Now, if the system is subdivided into subsystems in the physical space by means of the following partition functions

$$\sum_{\alpha} p^{\alpha}(\mathbf{r}) = 1 \tag{1.12}$$

where $p^{\alpha}(\mathbf{r})$ is a positive weighting function for the α-th subunits, the total density can be expressed as

$$\rho(\mathbf{r}) = 2\sum_{\alpha} p^{\alpha}(\mathbf{r})\langle \mathbf{r}|\eta(\epsilon_F - \hat{H}_{KS})|\mathbf{r}\rangle = \sum_{\alpha} \rho^{\alpha}(\mathbf{r}) \tag{1.13}$$

The partition of the electron density in (1.13) permits to introduce local approximations to $\hat{H}_{KS}$, namely it allows the introduction of different approximations to $\hat{H}_{KS}$ for the different subunits. Thus, for the α-th subsystem we have:

$$\tilde{\rho}^{\alpha}(\mathbf{r}) = 2p^{\alpha}(\mathbf{r})\langle \mathbf{r}|f_{\beta}(\epsilon_F - \hat{H}^{\alpha}_{KS})|\mathbf{r}\rangle \tag{1.14}$$

where $\hat{H}^{\alpha}_{KS}$ is the modified Kohn-Sham Hamiltonian for the α-th subunit and $f_{\beta}(x)$ is the Fermi function:

$$f_{\beta}(x) = \frac{1}{1 + e^{-\beta x}} \tag{1.15}$$

which is a convenient choice to make unique the value of ϵ_F. Here, $\beta = \frac{1}{kT}$, where k is the Boltzmann constant and T is the absolute temperature.

We now let $\hat{H}^{\alpha}_{KS}$ be the projection of the original Kohn-Sham Hamiltonian $\hat{H}_{KS}$ to the space spanned by the nonorthogonal basis functions $\{\chi^{\alpha}_{\mu}\}$ localized on the subunit α:

$$\hat{H}^{\alpha}_{KS} = \sum_{i} |\varphi^{\alpha}_{i}\rangle \epsilon^{\alpha}_{i} \langle \varphi^{\alpha}_{i}| \tag{1.16}$$

where the $\hat{H}^{\alpha}_{KS}$ operator is projected in terms of its eigenvalues ϵ^{α}_{i} and its eigenfunctions φ^{α}_{i} that are linear combinations of the basis functions $\{\chi^{\alpha}_{\mu}\}$:

$$\varphi^{\alpha}_{i}(\mathbf{r}) = \sum_{\mu} C^{\alpha}_{\mu i}\chi^{\alpha}_{\mu}(\mathbf{r}) \tag{1.17}$$

The linear coefficients are solutions of the following eigenvalue equation:

$$\left(\mathbf{H}^{\alpha}_{KS} - \epsilon^{\alpha}_{i}\mathbf{S}^{\alpha}\right)\mathbf{C}^{\alpha}_{i} = 0 \tag{1.18}$$

with

$$\left(\mathbf{S}^{\alpha}\right)_{\mu\nu} = \langle\chi^{\alpha}_{\mu}|\chi^{\alpha}_{\nu}\rangle \quad \text{and} \quad \left(\mathbf{H}^{\alpha}_{KS}\right)_{\mu\nu} = \langle\chi^{\alpha}_{\mu}|\hat{H}^{\alpha}_{KS}|\chi^{\alpha}_{\nu}\rangle$$

Using the spectral resolution (1.16) of $\hat{H}^{\alpha}_{KS}$, it is possible to determine the subspace density ρ^{α} by equation (1.14). Then, exploiting equation (1.13), we can obtain the expression for the direct calculation of the total electron density:

$$\tilde{\rho}(\mathbf{r}) = 2\sum_{\alpha} p^{\alpha}(\mathbf{r}) \sum_{i} f_{\beta}(\epsilon_{F} - \epsilon^{\alpha}_{i})|\varphi^{\alpha}_{i}(\mathbf{r})|^{2} \tag{1.19}$$

where the value of ϵ_F is determined by the normalization constraint

$$2N = \int_{\mathbb{R}^3} \tilde{\rho}(\mathbf{r})\,\mathrm{d}\mathbf{r} = 2\sum_{\alpha}\sum_{i} f_{\beta}(\epsilon_{F} - \epsilon^{\alpha}_{i})\langle\varphi^{\alpha}_{i}|p^{\alpha}|\varphi^{\alpha}_{i}\rangle \tag{1.20}$$

For the sake of completeness, it is worth noting that, in order to obtain a unique solution of ϵ_F for a given N, it is necessary to keep finite the value of β, so that the right-hand side of equation (1.20) is a continuous monoatomic function of ϵ_F. Actually, in the DC technique, T has no real physical meaning and its choice is not critical, although extremely high temperature could lead to significant occupations of high-energy molecular orbitals, which could also cause SCF convergence problems. Now, the total electronic energy of the system can be rewritten like this

$$E = 2\int_{\mathbb{R}^3} \mathrm{d}\mathbf{r}\langle\mathbf{r}|\hat{H}_{KS}\,\eta(\epsilon_{F} - \hat{H}_{KS})|\mathbf{r}\rangle + Q[\rho] \tag{1.21}$$

Using the same approximation adopted in (1.14), and introducing the approximate electron density (1.19) to evaluate the functional Q, we can finally write the following approximate expression for the electronic energy of the system:

$$\tilde{E} = 2\sum_{\alpha}\sum_{i} f_{\beta}(\epsilon_F - \epsilon_i^{\alpha})\langle\varphi_i^{\alpha}|p^{\alpha}|\varphi_i^{\alpha}\rangle + Q[\tilde{\rho}] \tag{1.22}$$

To summarize, the procedure to compute $\tilde{\rho}(\mathbf{r})$ in the DC approach consists in the following steps:

- selecting a partition function $p^{\alpha}(\mathbf{r})$ and a local basis set $\{\chi_{\mu}^{\alpha}\}$ for each subunit α;
- computing the $\mathbf{H}_{KS}^{\alpha}$ and $\mathbf{S}^{\alpha}$ matrices for each subsystem;
- solving equation (1.18) to obtain φ_i^{α} and ϵ_i^{α} for each subunit;
- determining ϵ_F by solving equation (1.20);
- computing $\tilde{\rho}$ using equation (1.19).

Even if the computational advantage of this approach is obvious, it is important to note that what makes the method linear in terms of computational cost is the use of a set of localized basis functions.

Until now, we have considered only the original formulation of the Divide & Conquer approach. In the next lines, the extension of the method to the density matrix will be discussed. In this case, the electron density for a $2N$ electrons closed-shell systems can be written as

$$\rho(\mathbf{r}) = \sum_{\mu,\nu=1} \chi_{\mu}^{*}(\mathbf{r})\mathbf{D}_{\mu\nu}\chi_{\nu}(\mathbf{r}) \tag{1.23}$$

where the density matrix $\mathbf{D}_{\mu\nu}$ is expressed in functions of the linear coefficients that expand the Kohn-Sham orbitals $\varphi_i(\mathbf{r})$ in the spanned space by the atomic orbitals $\chi_{\mu}(\mathbf{r})$:

$$\mathbf{D}_{\mu\nu} = 2\sum_{i=1}^{N} C_{\mu i}C_{\nu i} \tag{1.24}$$

It is important to precise that here we are discussing the method in the framework of Density Functional Theory, but this approach can be also extended to the Hartree-Fock and semiempirical techniques. Then, for each subsystem α, it is possible to define a partition

matrix $\mathbf{P}^{\alpha}_{\mu\nu}$ that satisfies the following normalisation condition:

$$\sum_{\alpha} \mathbf{P}^{\alpha}_{\mu\nu} = 1 \tag{1.25}$$

This is the simplest way to construct such matrices.

$$\mathbf{P}^{\alpha}_{\mu\nu} = \begin{cases} 1 & \text{if} \quad \chi_\mu \in \alpha \quad \text{and} \quad \chi_\nu \in \alpha \\ \frac{1}{2} & \text{if} \quad \chi_\mu \in \alpha \quad \text{and} \quad \chi_\nu \notin \alpha \quad \text{or} \quad \chi_\mu \notin \alpha \quad \text{and} \quad \chi_\nu \in \alpha \\ 0 & \text{if} \quad \chi_\mu \notin \alpha \quad \text{and} \quad \chi_\nu \notin \alpha \end{cases} \tag{1.26}$$

As a consequence, the density matrix can be divided into subsystems contributions like this:

$$\mathbf{D}_{\mu\nu} = \sum_{\alpha} \mathbf{P}^{\alpha}_{\mu\nu} \mathbf{D}_{\mu\nu} = \sum_{\alpha} \mathbf{D}^{\alpha}_{\mu\nu} \tag{1.27}$$

which is comparable to equation (1.13). Afterwards, introducing the approximation taken into account for the electron density approach, we obtain:

$$\tilde{\mathbf{D}}_{\mu\nu} = 2 \sum_{\alpha} \mathbf{D}^{\alpha}_{\mu\nu} \sum_{i} f_\beta(\epsilon_F - \epsilon^{\alpha}_i) C^{\alpha}_{\mu i} C^{\alpha}_{\nu i} \tag{1.28}$$

which is analogous to equation (1.19). To approximate the density matrix of a subsystem, a set of local eigenvectors is used also in this case and this is the origin of the linear scaling behaviour of the computational effort. In fact, the set of local eigenvectors for a subsystem is finite and independent of the size of the whole system. The Fermi energy is determined again exploiting the following normalisation condition:

$$2N = \sum_{\mu\nu} \tilde{\mathbf{D}}_{\mu\nu} \mathbf{S}_{\mu\nu} = 2 \sum_{\mu\nu} \mathbf{S}_{\mu\nu} \sum_{\alpha} \mathbf{D}^{\alpha}_{\mu\nu} \sum_{i} f_\beta(\epsilon_F - \epsilon^{\alpha}_i) C^{\alpha}_{\mu i} C^{\alpha}_{\nu i} \tag{1.29}$$

It is worth noting that the main novelty in the density matrix formulation of the DC technique is that the division of the molecule is accomplished in the space of the atomic orbitals, while the approximation of each subunit by means of a set of local basis set is unchanged. There are two principle advantages in using the density matrix approach compared to the electron density case. The first one is that the computation of the integrals associated with the partition functions are not necessary. Due to the fact that three-dimensional numerical integrations are very time consuming, this makes this new

approach much more efficient. The second advantage is that the density matrix formulation can also be extended to other computational strategies such as the Hartree-Fock or semiempirical methods. However, the main drawback of the subdivision in the atomic orbitals space is that it is less localized in the physical space, in particular when diffuse functions are used.

It is important to point out that, in the initial formulation of the density matrix DC method proposed by Yang and Lee (*21*), the subdivision of the global system into small subunits was very simplistic. As already mentioned in the introduction of this Chapter, Merz and coworkers have proposed afterwards more sophisticated schemes. Especially, they have tested different ways to subdivide a system (*25*, *26*) introducing buffer regions (Fig 1.2). They suggested to subdivide a subsystem into a core region R^{α}_{core}, which is distinct from the other subsystem cores, and two different buffer regions $R^{\alpha}_{buff_1}$ and $R^{\alpha}_{buff_2}$, which overlap with the cores of adjacent subsystems. The R^{α}_{core} contains only high quality density matrix information because it is insulated from the subsystem boundary. The inner buffer layer $R^{\alpha}_{buff_1}$ provides information of varying degree of reliability, while the outer buffer layer $R^{\alpha}_{buff_2}$ is only considered for insulation purpose. The contribution of these regions to the global density matrix can be summarize in the following definition of the partition matrix:

$$\mathbf{P}^{\alpha}_{\mu\nu} = \begin{cases} 0 & \text{if} \quad \chi_\mu \notin R^\alpha \quad \text{or} \quad \chi_\nu \notin R^\alpha \\ 0 & \text{if} \quad \chi_\mu \in R^{\alpha}_{buff_2} \quad \text{or} \quad \chi_\nu \in R^{\alpha}_{buff_2} \\ 0 & \text{if} \quad \chi_\mu \in R^{\alpha}_{buff_1} \quad \text{and} \quad \chi_\nu \in R^{\alpha}_{buff_2} \\ \frac{1}{n_{\mu\nu}} & otherwise \end{cases} \tag{1.30}$$

where $n_{\mu\nu}$ is the total number of fragments that overlap to make a non-zero contribution to the global density matrix element $\mathbf{D}^{\alpha}_{\mu\nu}$. Therefore, any density matrix element is only considered if χ_μ and χ_ν are localized within the same core region R^{α}_{core}, or if they couple the core region and the first buffer layer. This last type of contribution is necessary in order to take into account the chemical bonding across subunits boundaries.

Finally, for the sake of completeness, let us consider another DC-based method, namely the Molecular Tailoring Approach (MTA) (*56*–*61*), which was independently developed by Gadre and coworkers in 1994. The MTA is a fragment-based linear scaling technique which was initially proposed for the evaluation of molecular properties and later for geometry optimizations using a parallel implementation. As the different DC approaches

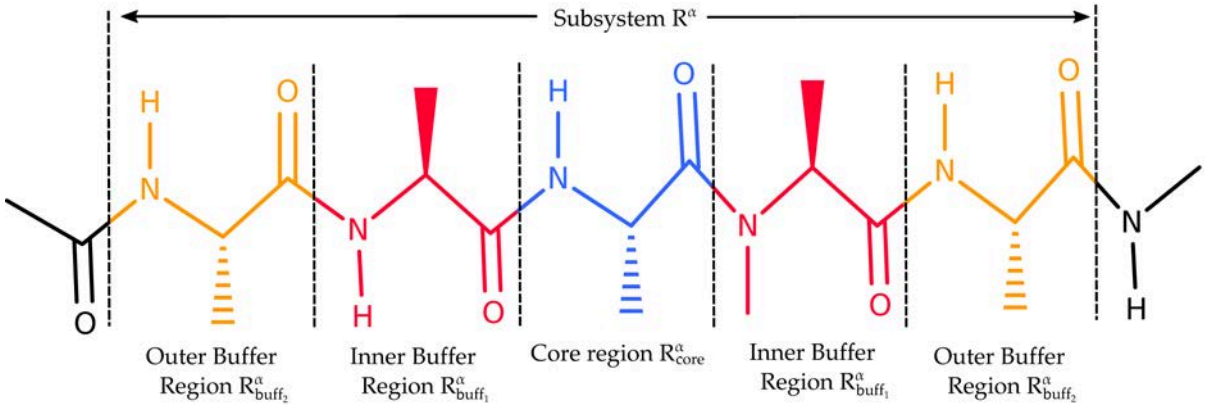

FIGURE 1.2: Partition of a subsytem into core and buffer regions

described above, the MTA subdivides a large molecule into a set of small, overlapping fragments. Nevertheless, here the covalent bonds are cut and dummy hydrogen atoms are added at appropriate positions to satisfy the valence requirements. The important steps in the MTA algorithm are:

- the partitioning of the target molecule under consideration into smaller fragments, using automatic or manual fragmentation;
- the set up of cardinality-based equations to afterwards estimate properties (e.g., energy and gradient) on the target molecule from the ones of the parent molecules;
- the *ab initio* calculations performed to determine the density matrix, the energy and the gradient for each individual fragment;
- the recombination of the results for the different fragments in accordance with cardinality-based equations to obtain the properties of the target molecule.

The cardinality-based equations are based on the inclusion-exclusion principle. The desired molecular property, P, of the target molecule is obtained by patching those of the individual fragments:

$$P = \sum P^{F_i} - \sum P^{F_i \cap F_j} + ... + (-1)^{k-1} \sum P^{F_i \cap F_j \cap ... F_k} ... \tag{1.31}$$

where P^{F_i} is the contribution to the molecular property from the i-th fragment, $P^{F_i \cap F_j}$ stands for the molecular property resulting from the overlap between the i-th and the j-th subunits, and so on... Initially, the MTA was developed for determining the density

matrix of large molecules. It was later extended to perform geometry optimizations based on an estimation of energy and gradient from appropriate fragment contributions.

1.3 The Fragment Molecular Orbital method

Another well-established linear scaling philosophy is the fragmentation interaction approach, which is based on the theory of molecular interactions. Among all the methods of this group, the most important and widespread is undoubtedly the Fragment Molecular Orbital (FMO) strategy originally proposed by Kitaura and co-workers in 1999 (*31*–*37*). The technique is based on the following equation:

$$E = \sum_i E_i + \sum_{i>j} \Delta E_{ij} + \sum_{i>j>k} \Delta E_{ijk} + ... \tag{1.32}$$

in which the global energy of the system is given by the sum of monomer contributions E_i and intermolecular interactions energies obtained from dimers (ΔE_{ij}) and possibly larger subunits ($\Delta E_{ijk},\ \Delta E_{ijkl}$).

Therefore, the main idea underling the FMO strategy consists in dividing a molecule into N fragments and, afterwards, in performing calculations on the monomers, dimers, trimers, etc. At the final stage, the total energy of the system is obtained exploiting equation (1.32). This approach displays two main advantages:

- the avoidance of MOs computation of the whole system which significantly reduces the computational cost for large molecules;
- the easiness of using parallel processing, since the calculation on fragments and fragment-pairs can be carried out independently.

In the following paragraphs, the main strategy and equations of the FMO method will be briefly introduced. To accomplish this task, let us consider the simple propanol molecule.

At first, the system in exam is properly subdivided into fragments among which the electrons are properly distributed. However, it is important to point out that the bond electron-pairs are conserved and, therefore, for the propanol molecule we have:

- one CH_3 fragment of 8 electrons;

- two CH_2 fragments of 8 electrons;
- one OH fragment of 10 electrons.

Afterwards, the following Hamiltonian is introduced for each subunit:

$$H_\alpha = \sum_i^{n_\alpha} \left(-\frac{1}{2}\nabla_i^2 - \sum_A^{all} \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{\beta\neq\alpha}^{N} \int \frac{\rho_\beta(\mathbf{r'})}{|\mathbf{r}_i - \mathbf{r'}|}\, \mathrm{d}\mathbf{r'} \right) + \sum_{j>i}^{n_\alpha} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{1.33}$$

where n_α is the number of electrons in the α-th fragment, N is the number of fragments in the molecule, Z_A is the nuclear charge of the atom A and $\rho_\beta(\mathbf{r'})$ is the electron distribution of the fragment β. The electron distributions are then iteratively obtained for all the fragments. It is worthwhile to note that the fragment Hamiltonian H_α includes the electrostatic potential from the electrons in the surrounding (N-1) subunits in addition to the nuclear attraction from all the nuclei. Likewise, the fragment pair Hamiltonian can be written as

$$H_{\alpha\beta} = \sum_i^{n_\alpha+n_\beta} \left(-\frac{1}{2}\nabla_i^2 - \sum_A^{all} \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{\gamma\neq\alpha,\beta}^{N} \int \frac{\rho_\gamma(\mathbf{r'})}{|\mathbf{r}_i - \mathbf{r'}|}\, \mathrm{d}\mathbf{r'} \right) + \sum_{j>i}^{n_\alpha+n_\beta} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{1.34}$$

which also takes into account the electrostatic potential term from the electrons in the surrounding (N-2) subunits. It is important to precise that $\rho_\gamma(\mathbf{r'})$ is the electron distribution obtained from the monomer calculation and is not varied in the pair computation.
If we solve the following Schrödinger equations at a suitable level of (for instance, at Hartree-Fock level under the restriction that the orbitals are localized on the proper fragments or fragment-pairs):

$$H_\alpha \Psi_\alpha = E'_\alpha \Psi_\alpha \tag{1.35}$$

and

$$H_{\alpha\beta} \Psi_{\alpha\beta} = E'_{\alpha\beta} \Psi_{\alpha\beta} \tag{1.36}$$

we can have access to the electronic energy E'_α of the fragment α and to the energy $E'_{\alpha\beta}$ for the fragment pair $\alpha\ \beta$. Afterwards, the total energy of the system can be written as

$$E = \sum_{\alpha>\beta}^{N} E'_{\alpha\beta} - (N-2) \sum_{\alpha}^{N} E'_\alpha + \sum_{A>B}^{all} \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|} \tag{1.37}$$

where the last term is the nuclear repulsion energy. Furthermore, even the total electron density ρ of the molecule can be deduced from the fragment ρ_α and the fragment pair $\rho_{\alpha\beta}$

densities as

$$\rho(\mathbf{r}) = \sum_{\alpha>\beta}^{N} \rho(\mathbf{r})_{\alpha\beta} - (N-2) \sum_{\alpha}^{N} \rho(\mathbf{r})_{\alpha} \tag{1.38}$$

In conclusion, the computational procedure of the FMO method can be resumed in this way:

- subdividing the target molecule into subsystems and assigning electrons to each fragment;
- calculating the initial electron density associated with each fragment;
- constructing the fragment Hamiltonians using the given electron densities and solving the corresponding Schrödinger equations to obtain the monomer energies $\{E_{\alpha}^{'}\}$ and the electron distributions $\{\rho_{\alpha}\}$;
- checking if the electron densities of all the fragments are identical to the previous ones using a specific criterion (if the procedure has not converged, it is necessary to return to the third step with new density distributions.);
- building up the fragment pair Hamiltonians by means of the converged densities and solving the associated Schrödinger equations for all fragment pairs to obtain the dimer energies $\{E_{\alpha\beta}^{'}\}$;
- computing the total energy using the monomer and dimer energies $\{E_{\alpha}^{'}\}$ and $\{E_{\alpha\beta}^{'}\}$.

Finally it is worth pointing out that among all the "fragmentation interaction approaches", the FMO technique has been supported by a continuous method development (e.g. extension to the MP2, Coupled-Cluster, DFT, MCSCF approaches) and by a series of applications that confirm its usefulness to study practical problems associated with large systems, including molecular clusters, proteins, DNA, enzymes, ionic liquids, molecular crystals, zeolites and nanowires.

1.4 The Additive Fuzzy Density Fragmentation approaches

Besides the linear scaling techniques considered above, there is another class of methods that result from a fuzzy representation of the electron density, namely the Additive Fuzzy

Density Fragmentation (AFDF) principle (*62*, *63*). This principle is at the basis of two related linear scaling approaches developed by Mezey and coworkers: the MEDLA (Molecular Electron Density Loge (or Lego) Assembler) method (*45*–*49*) and the ADMA (Adjustable Density Matrix Assembler) method (*50*–*54*). In the next section, we will manly review the Additive Fuzzy Density Fragmentation principle.

The *ab initio* electron density $\rho(\mathbf{r}, K)$ of a molecule with a nuclear configuration K can be expressed in terms of the atomic orbitals $\{\chi_\mu\}_{\mu=1}^{NAO}$ used for the expansion of the molecular wave function. If the basis functions and the density matrix explicitly depend on the nuclear configuration K, we obtain:

$$\rho(\mathbf{r}, K) = \sum_{\mu,\nu=1} \chi_\mu^*(\mathbf{r}, K)\mathbf{D}_{\mu\nu}(K)\chi_\nu(\mathbf{r}, K) \tag{1.39}$$

In the AFDF procedure, the set of nuclei of the system are subdivided into m mutually exclusive groups denoted by

$$f_1, f_2, ..., f_h, ..., f_m$$

in order to create m fuzzy additive density fragments

$$F_1, F_2, ..., F_h, ..., F_m$$

which correspond to fragment density functions designated as

$$\rho^1(\mathbf{r}, K), \rho^2(\mathbf{r}, K), ..., \rho^h(\mathbf{r}, K), ..., \rho^m(\mathbf{r}, K)$$

Then, introducing a formal membership function $m_h(\mu)$ which indicates if the atomic orbital χ_μ is centred on a nucleus of the set f_h:

$$m_h(\mu) = \begin{cases} 1 & \text{if} \quad \chi_\mu \text{ is centered on one of the nuclei of set } f_h \\ 0 & \text{otherwise} \end{cases} \tag{1.40}$$

the elements $\mathbf{D}^h_{\mu\nu}(K)$ of the density matrix associated with the h-th fragment F_h can be written as

$$\mathbf{D}^h_{\mu\nu}(K) = \left[m_h(\mu)w_{\mu\nu} + m_h(\nu)w_{\nu\mu}\right]\mathbf{D}_{\mu\nu}(K) \tag{1.41}$$

where $w_{\mu\nu}$ and $w_{\nu\mu}$ are weighting factors respecting the following conditions:

$$\begin{cases} w_{\mu\nu} + w_{\nu\mu} = 1 \\ w_{\mu\nu} \quad \text{and} \quad w_{\nu\mu} > 0 \end{cases} \tag{1.42}$$

The simplest version of this scheme corresponds to

$$w_{\mu\nu} = w_{\nu\mu} = \frac{1}{2}$$

and gives us:

$$\mathbf{D}^h_{\mu\nu}(K) = \frac{1}{2}\left[m_h(\mu) + m_h(\nu)\right]\mathbf{D}_{\mu\nu}(K) \tag{1.43}$$

According to the general fuzzy fragmentation scheme, it is important to notice that, for each index pair (μ, ν), the elements $\mathbf{D}^h_{\mu\nu}(K)$ of the fragment density matrices $\mathbf{D}^h(K)$ are additive:

$$\mathbf{D}_{\mu\nu}(K) = \sum_{h=1}^{m} \mathbf{D}^h_{\mu\nu}(K) \tag{1.44}$$

and the sum of fragment density matrices is equal to the density matrix $\mathbf{D}(K)$ of the complete molecule:

$$\mathbf{D}(K) = \sum_{h=1}^{m} \mathbf{D}^h(K) \tag{1.45}$$

In the same way, using the fragment density matrices, the fuzzy electron density of the h-th density fragment can be written as:

$$\rho^h(\mathbf{r}, K) = \sum_{\mu,\nu=1} \chi^*_\mu(\mathbf{r}, K)\mathbf{D}^h_{\mu\nu}(K)\chi_\nu(\mathbf{r}, K) \tag{1.46}$$

According to equation (1.39), the electron density of the system is linear in the density matrix elements and, consequently, following the additivity properties in (1.44) and (1.45), the fuzzy density fragments are additive and their sum is the global molecular electron density:

$$\rho(\mathbf{r}, K) = \sum_{h=1}^{m} \rho^h(\mathbf{r}, K) \tag{1.47}$$

The above fuzzy fragmentation scheme can be exploited to study the shape of local moieties and the interactions of various parts of small molecules. However, the most useful application of the AFDF philosophy is the construction of approximate electron densities

for large systems. This can be performed using smaller model molecules, containing various parts of the original target system. For each nuclear set f_h of the large system M, a small model molecule M_h is created, where M_h contains the same nuclear set f_h with an equivalent local arrangement and surroundings as the target molecule M. A fuzzy density fragmentation is then accomplished for the small model molecule M_h to get the fuzzy density fragments $\rho^h(\mathbf{r}, K)$ corresponding to the nuclear configuration K. If the procedure is repeated for each nuclear family f_h of M, the fuzzy fragments

$$\rho^1(\mathbf{r}, K), \rho^2(\mathbf{r}, K), ..., \rho^h(\mathbf{r}, K), ..., \rho^m(\mathbf{r}, K)$$

obtained from the set of m model molecules

$$M_1, M_2, ..., M_h, ..., M_m$$

can be combined to construct the electron density $\rho(\mathbf{r}, K)$ of the target system.

As mentioned above, the first implemented version of the AFDF approach was the MEDLA method. This procedure is based on the construction of a database of precalculated numerical electron density fragments and on the numerical reconstruction of molecular charge distributions. Some detailed tests (*45*, *46*) have shown that the MEDLA method reproduces the results of conventional 6-31G(d,p) *ab initio* quality electron densities. In particular, *ab initio* quality charge distributions of several proteins have been computed (crambin, bovine insuline, gene-5 protein, etc.) However, the MEDLA method exhibits also some drawbacks like the requirement of a numerical database and some problems associated with the grid alignment of the combined density data.

To tackle these deficiencies, the more sophisticated ADMA (Adjustable Density Matrix Assembler) has been developed. This technique accomplishes the combination of fuzzy electron density fragments constructing a molecular density matrix from fuzzy fragment density matrices. This matrix representation displays several advantages. In fact, the electron density representation in terms of Density Matrices avoids the grid alignment problems and, furthermore, interesting one-electron properties can be easily computed, for instance the molecular electrostatic potential.

In this first chapter, we have introduced some basic ideas at the basis of the linear scaling quantum chemical methods. We have mainly considered the Divide & Conquer

strategy, the Fragment Molecular Orbital method and the Additive Fuzzy Density Fragmentation approaches. However, in this context, following the linear scaling philosophy, an alternative is represented by the possibility of defining molecular orbitals strictly localized on small molecular subunits and, therefore, easily transferable to other molecules containing the same fragment. Unfortunately, the molecular orbitals traditionally obtained and used in quantum chemistry are generally delocalized over the system in exam and, consequently, not easily transferable from molecule to molecule. This problem will be discussed in details in the next chapter, where we will also introduce the concept of Extremely Localized Molecular Orbitals (ELMOs) that, as we will see, are characterized by the absence of tails beyond the main localization region and, for this reason, are the most suitable localized orbitals to develop new orbital based linear scaling strategies.

Chapter 2

Localized Molecular Orbitals

Résumé

Dans le chapitre précédent, nous avons souligné le rôle important du concept de localisation dans le développement de nouvelles méthodes à croissance linéaire. En effet, si l'on suit le raisonnement de l'approche LEGO introduite par Mezey et basée sur le transfert de fragments de densité électronique pour construire une distribution de charge d'une macromolécule, on peut imaginer une procédure similaire impliquant un transfert d'orbitales strictement localisées sur de petits fragments moléculaires qui nous permettrait de reconstruire la fonction d'onde ou la densité électronique de très grands systèmes. Malheureusement, les orbitales moléculaires habituellement utilisées en chimie théorique sont totalement délocalisées sur le système étudié et, par conséquent, ne peuvent pas être transférées d'une molécule à une autre.

Pour remédier à ce problème, nous devons nous tourner vers des orbitales localisées. Les techniques de localisation utilisent des transformations unitaires des orbitales Hartree-Fock comme décrit dans la première partie de ce chapitre. Les orbitales ainsi localisées conservent toutefois des "queues" hors de leurs régions de localisation et de ce fait ne peuvent pas être transférées d'une molécule à une autre. Notons que la suppression de ces queues engendre une augmentation importante de l'énergie du système.

Pour aller plus loin, nous devons avoir recours à des techniques de localisation dites *a priori*, qui divisent le système étudié en fragments et dont la fonction d'onde globale est une somme de contributions locales provenant de chacun des fragments. C'est ce type de partition que l'on considère pour définir les Orbitales Moléculaires Extrêmement Localisées (ELMOs). Ces orbitales sont obtenues de manière variationnelle sous la contrainte d'être exclusivement exprimées sur des fonctions de base centrées sur des atomes de fragments présélectionnés. Dans ce chapitre, nous détaillons l'approche de Stoll utilisée pour générer les ELMOs ainsi que la technique de rotation des orbitales employée

pour le transfert des ELMOs d'une molécule à une autre.

2.1 Introduction

In the previous chapter, we have briefly mentioned the important role of the localization in the development of novel linear scaling methods. Indeed, a molecular system can be considered as an addition of recurrent local units such as atoms, bonds or functional groups. Following the concepts introduced by Mezey *et al.* (45–49) in the Molecular Electron Density Lego Assembler method, which is based on the transfer of numerical electron density fragments to construct molecular charge distributions of macromolecules, it is possible to conceive a procedure involving molecular orbitals strictly localized on small molecular subunits to recover wave functions and electron densities of large systems. Unfortunately, the molecular orbitals routinely used in quantum chemistry are spread out all over the whole system in exam and this prevents their transfer from a molecule to another.

In order to preserve the traditional chemical picture of molecules (e.g., the traditional Lewis picture) in theoretical chemistry, a lot of methods have been proposed to localize molecular orbitals using unitary transformations of the traditional Hatree-Fock molecular orbitals. In particular, in these *a posteriori* methods, the localized orbitals are obtained though minimization or maximization procedures of suitable properties. For example, the spatial extension of the orbitals is minimized in the approach proposed by Boys (64, 65), whereas in the Edmiston-Ruedenberg technique (66, 67) the self-repulsion energy is maximized. In another approach devised by Pipek and Mezey (68, 69), a functional correlated with the Mulliken population analysis is minimized. Even if these *a posteriori* defined localized molecular orbitals are centred on small subunits, they conserve small orthogonalization tails beyond their localization region. For this reason, these orbitals can not be transferred from a molecule to another because, although the coefficients associated with these tails are quite small, their deletion entails a dramatic increase of the molecular energy of the system.

To avoid this problem, we should turn to the *a priori* localization techniques, which subdivide the system into fragments and determine the global wave function of a molecule as a sum of local contributions coming from each subunit. The resulting tail-free Extremely Localized Molecular Orbitals (ELMOs) are variationally obtained under the restraint of expanding them only using subsets of basis functions centred on atoms associated with the preselected fragments.

In this chapter, after a brief non exhaustive overview of some traditional Localized Molecular Orbitals (LMOs) methods that exploit unitary transformations, we will present the *a priori* strategy proposed by Stoll and coworkers (*70*) in 1980, a strategy that represents the first generalization of the Hartree-Fock equations to compute Extremely Localized Molecular Orbitals. Finally, in the last section of this chapter, we will also discuss in detail a technique that has been initially proposed by Philipp and Friesner (*71*), for the rotation of Strictly Localized Molecular Orbitals and that we have afterwards used for the transfer of ELMOs from a molecule to another.

2.2 Unitary Transformation Methods

In the Hartree-Fock approximation, the wave function that describes the system under exam is a single Slater determinant constructed with NMO molecular orbitals:

$$\{|\varphi_i\rangle, i = 1, ..., NMO\} \tag{2.1}$$

As it is well known, applying any unitary transformation to this set of orbitals, the wave function remains invariant and, therefore, the energy does not change. The objective of the traditional localization approaches is to get a new set of orbitals that satisfy a given localization criterion. Historically, the first localization procedure was proposed by Foster and Boys in 1960 (*64*, *65*). In this case, the transformation minimizes the orbital self-extension, which is represented by the following functional J:

$$J_{Boys}\{|\varphi_i\rangle\} = \sum_{i=1}^{NMO} \int \varphi_i^*(\mathbf{r}_1)\varphi_i^*(\mathbf{r}_1)\left(\mathbf{r}_1 - \mathbf{r}_2\right)^2 \varphi_i(\mathbf{r}_2)\varphi_i(\mathbf{r}_2)\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \tag{2.2}$$

and which is equivalent to the formulation that maximizes the sum of the squares of the distances between orbitals centroids,

$$J_{Boys}\{|\varphi_i\rangle\} = \sum_{i>j=1}^{NMO} \left[\int d(\mathbf{r})\mathbf{r}\varphi_i^*(\mathbf{r})\varphi_i(\mathbf{r}) - \int d(\mathbf{r})\varphi_j^*(\mathbf{r})\varphi_j(\mathbf{r})\right]^2 \tag{2.3}$$

This localization method introduced by Boys and Foster requires only the one-electron integrals over the occupied molecular orbitals. Therefore, due to the straightforward calculation of these integrals, this procedure only scales as $O(N^3)$. However, although the

Boys procedure is quite simple to implement, some problems occur when one tries to localize orbitals associated with double bonds.

Another method proposed by Edmiston and Ruedenberg (66, 67) determines a transformation that maximizes the self-repulsion energy defined by the functional

$$J_{ER}\{|\varphi_i\rangle\} = \sum_{i=1}^{NMO} \int \varphi_i^*(\mathbf{r}_1)\varphi_i^*(\mathbf{r}_1) \left(\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}\right) \varphi_i(\mathbf{r}_2)\varphi_i(\mathbf{r}_2)\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2 \tag{2.4}$$

This procedure was suitable in early applications because the double bonds were particularly well described, preserving the separation between σ and π orbitals. However, from a computational point of view, the Edmiston-Ruedenberg method has some deficiencies. In particular, this technique needs the two-electron integrals over the occupied canonical molecular orbitals. For the computation of these integrals, a transformation must be performed from the atomic orbitals to the molecular orbitals basis. This transformation is especially time consuming for large systems and scales as $O(N^5)$ according to the number of basis functions. The Edmiston-Ruedenberg approach was then extended by Von Niessen (72) in 1971 minimizing the sum of the following charge density overlap functional:

$$J_{VonNiessen}\{|\varphi_i\rangle\} = \sum_{i=1}^{NMO} \int \varphi_i^*\varphi_i^* \, |\delta(\mathbf{r}_1 - \mathbf{r}_2)|\varphi_i\varphi_i \, d\mathbf{r}_1 \, d\mathbf{r}_2 \tag{2.5}$$

In the same period, Magnasco and Perico suggested a localization procedure (73) based on the local orbital populations $\{P_i\}$ for each MO:

$$P_i = \sum_{\mu\in\Gamma_t} \sum_{\nu\in\Gamma_t'} C_{\mu i} S_{\mu\nu} C_{\nu i} \tag{2.6}$$

where $C_{\mu i}$ and $C_{\nu i}$ are the expansion coefficient of the molecular orbitals in the Linear Combination of Atomic Orbitals (LCAO) approximation, $S_{\mu\nu}$ is the overlap between the atomic orbitals and Γ_t and Γ_t' are properly chosen according to the nature of the localized orbitals. The populations $\{P_i\}$ allow to define the functional

$$J_{MP}\{|\varphi_i\rangle\} = \sum_{i=1}^{NMO} P_i \tag{2.7}$$

which should be maximized in order to obtain a set of uniformly localized molecular orbitals. The main advantage of this localization criterion is that it only requires the knowledge of the molecular orbital coefficients ant the overlap integrals between the atomic orbitals.

In analogy with the technique of Magnasco and Perisco, some years later, Pipek and Mezey proposed another strategy (*68*, *69*) to determine localized molecular orbitals using local orbital populations. This procedure aims at localizing the Mulliken atomic charge distributions defining localization quantities $\{d_i\}$ that measure the number of atoms over which the molecular orbital $\{|\varphi_i\rangle\}$ extend. The quantity d_i is expressed as

$$d_i = \left\{ \sum_{A=1}^{n} (Q_A^i)^2 \right\}^{-1} \tag{2.8}$$

where A is the label for the different atoms of the molecule, n is the number of atoms and Q_A^i is the gross atomic Mulliken population of the orbital $|\varphi_i\rangle$ when the atom A is taken into account. The quantity d_i is a relative good approximation to estimate the number of atoms over which the orbital is localized and should take any value between 0 and n:

$$0 < d_i < n \tag{2.9}$$

In order to obtain localized molecular orbitals, the value of d_i for each MO should be minimized. Hence, the Pipek-Mezey procedure seeks a set of orbitals that maximize a functional expressed like this:

$$J_{PM}\{|\varphi_i\rangle\} = \sum_{i=1}^{NMO} d_i^{-1} \tag{2.10}$$

This technique is suitable to separate core and valence orbitals and it also supports well the σ-π separation. The resulting orbitals are similar to those obtained by means of the Edmiston-Ruedenberg method but the algorithmic complexity scales only as $O(N^3)$, namely it is fast as the Boys strategy. The only drawback of the strategy is its strong dependence on the quality of the basis-sets used for the calculations.

Even if all the *a posteriori* unitary transformation procedures mentioned above give localized molecular orbitals having the same general features, the large variety of criteria

denote some degree of arbitrariness. Moreover, the obtained orbitals exhibit orthogonalization tails beyond the localization regions, and hence, they are not strictly localized. In the next sections, we will discuss how to overcome this drawback considering the possibility of determining Extremely Localized Molecular Orbitals, (namely, Molecular Orbitals Strictly Localized on small molecular fragments) directly by means of a variational procedure.

2.3 Extremely Localized Molecular Orbitals: the Stoll technique

In this section we will briefly review the approach proposed by Stoll and coworkers (*70*) in 1980 for the direct determination of Extremely Localized Molecular Orbitals (ELMOs). This strategy is stricly related to the earlier group function method devised by McWeeny (*74*–*76*) and it can be considered one of the many theoretical approaches that have been developed over the years in order to decompose the global electronic wave function into functions describing smaller subsets of electrons (*70*, *77*–*86*). As we have just seen (section 2.2), in most cases, the LMOs are obtained through unitary transformations of the canonical Molecular Orbitals. These *a posteriori* methods provide LMOs that are more or less localized on some fragments but that have the so-called orthogonalization tails extending over the whole system (see Figure 2.1). To easily transfer these LMOs from one molecule to another, this tails must be deleted. Unfortunately, the deletion of these tails produce a large increase of the molecular electronic energy. Moreover, Sundberg *et al.* (*87*) have well demonstrated that through a preliminary variational deorthogonalization of the LMOs before truncation does not entail a spectacular increase in localization and it does not lead to the disappearance of the tails.

Following the original idea of Mc Weeny, another possibility of obtaining MOs strictly localized on molecular fragments consists in the direct variational determination of Molecular Orbitals using for each fragment (e.g., atoms and bond) a separate local basis set, constituted by the only basis functions centred on the atoms belonging to the fragment. This is the simple strategy proposed by Stoll and co-workers to prevent *a priori* the presence of tails in LMOs.

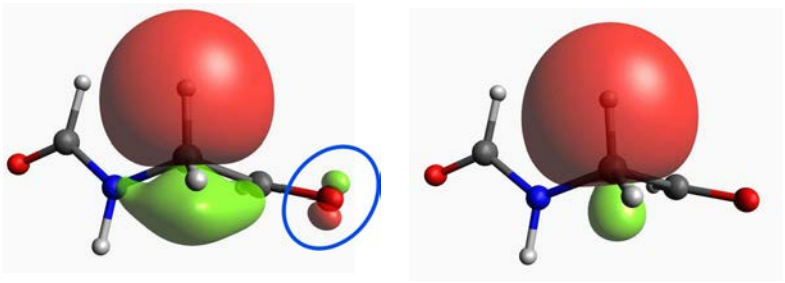

FIGURE 2.1: Isosurfaces for an LMO (left, obtained through the Boys localization technique) and for an ELMO (right, obtained through the Stoll technique) describing one of the C-H bonds of the N-ethanalmethanamide molecule (Isovalue set equal to $0.03e/bohr^3$).

To introduce this method, let us consider a 2N-electrons closed-shell molecule. According to the traditional concept of core, electrons, lone-pairs, covalent bonds and functional groups, we introduce a localization scheme that subdivides the system under exam into a set of overlapping fragments as illustrated in Figure (2.2). Each fragment or subunit j is automatically associated with a local basis-set $\beta_j = \{|\chi_{j\mu}\rangle\}_{\mu=1}^{M_j}$ that consists of the only basis functions centred on the atoms belonging to the fragment. Consequently the LMOs of the different subunits are expended only in the corresponding local basis-sets and, for instance, the generic α-th LMO of the j-th fragment can be simply written as follow:

$$|\varphi_{j\alpha}\rangle = \sum_{\mu=1}^{M_j} C_{j\mu,j\alpha}|\chi_{j\mu}\rangle \tag{2.11}$$

Furthermore, the global matrix of the LMOs coefficients has a particular block-structure as the one reported in Figure (2.2).

This means that the Stoll's LMOs are orbitals constrained *a priori* to be strictly localized on the corresponding fragments and, for this reason, from now on they will be indicated as Extremely Localized Molecular Orbitals, or, more simply ELMOs.

Following Stoll, the wave function that describes the global system is a normalized single Slater determinant constructed with the ELMOs:

$$|\psi_{ELMO}\rangle = \frac{1}{\sqrt{(2N)!det[\mathbf{S}]}}\hat{A}\left[\varphi_{11}\bar{\varphi}_{11}...\varphi_{1n_1}\bar{\varphi}_{1n_1}...\varphi_{f1}\bar{\varphi}_{f1}...\varphi_{fn_f}\bar{\varphi}_{fn_f}\right] \tag{2.12}$$

where $\hat{A}$ is the antisymmetrizer, n_i is the number of occupied ELMOs for the i-th fragment and f is the total number of fragments. $\varphi_{i\alpha}$ is a spin-orbital with spatial part $\varphi_{i\alpha}$ and spin

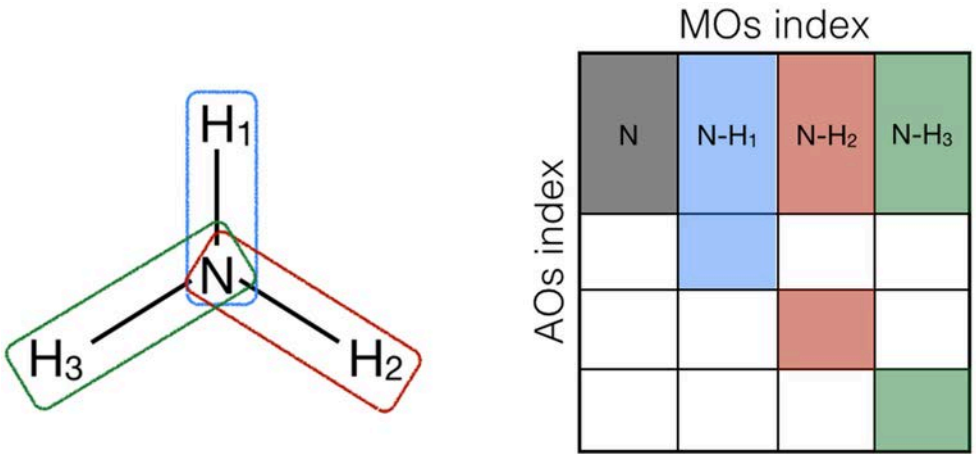

FIGURE 2.2: Localization scheme (left) and block-structured matrix of the LMO coefficients (right) for the ammonia molecule. In the localization scheme, the three overlapping bond fragments N-H1, N-H2 and N-H3 are explicitly shown, while, for the sake of clarity, the atomic fragment N, which describes the core and lone-pair electrons of the nitrogen atom, is not depicted.

part α, whereas $\bar{\varphi}_{i\alpha}$ is a spin-orbital with a spatial part $\varphi_{i\alpha}$ and a spin β. Finally, $det[\mathbf{S}]$ is the determinant of the overlap matrix between the occupied-ELMOs, which is due to the non orthogonal character of the ELMOs, arising from the fact that the predefined overlapping fragments (see Figure 2.2) share part of their local basis-sets.

Now, exploiting the usual definition of the one-electron operator part $\hat{h}$ of the standard Hamiltonian $\hat{H}$ and of the Fock operator $\hat{F}$, the total electronic energy of the closed-shell system can be expressed as

$$E = \langle \psi_{ELMO} | \hat{H} | \psi_{ELMO} \rangle = \sum_{i=1}^{f} \sum_{\alpha=1}^{n_i} \langle \varphi_{i\alpha} | \hat{h} + \hat{F} | \tilde{\varphi}_{i\alpha} \rangle \tag{2.13}$$

where the reciprocal orbitals $|\tilde{\varphi}_{i\alpha}\rangle$ can be defined using the following relation:

$$|\tilde{\varphi}_{i\alpha}\rangle = \sum_{j=1}^{f} \sum_{\beta=1}^{n_j} \left[\mathbf{S}^{-1}\right]_{j\beta,i\alpha} |\varphi_{j\beta}\rangle \tag{2.14}$$

where $\left[\mathbf{S}^{-1}\right]_{j\beta,i\alpha}$ is the inverse overlap matrix of the occupied ELMOs.

The goal is to determine the set of occupied ELMOs which minimize the energy. To accomplish this task, let us consider the variations of the total electronic energy E due to

the arbitrary variation of the occupied ELMOs $|\varphi_{j\beta}\rangle$:

$$\delta_{(j\beta)}E = 2\left(\langle\delta\varphi_{j\beta}|\hat{F}|\tilde{\varphi}_{j\beta}\rangle + \sum_{i\alpha}\langle\varphi_{i\alpha}|\hat{F}|\delta_{(j\beta)}\tilde{\varphi}_{i\alpha}\rangle\right) \tag{2.15}$$

Now, exploiting the relation

$$\delta\mathbf{S}^{-1} = -\mathbf{S}^{-1}\cdot\delta\mathbf{S}\cdot\mathbf{S}^{-1} \tag{2.16}$$

it can be easily shown that the variation of the reciprocal orbital $|\delta_{(j\beta)}\tilde{\varphi}_{i\alpha}\rangle$ can be expressed as

$$|\delta_{(j\beta)}\tilde{\varphi}_{i\alpha}\rangle = (1-\hat{\rho})|\delta\varphi_{j\beta}\rangle\cdot\mathbf{S}^{-1}_{j\beta,i\alpha} - |\tilde{\varphi}_{j\beta}\rangle\langle\delta\varphi_{j\beta}|\tilde{\varphi}_{i\alpha}\rangle \tag{2.17}$$

where the density operator $\hat{\rho}$ depends on all the occupied ELMOs and is given by:

$$\hat{\rho} = \sum_{k=1}^{f}\sum_{\gamma}^{n_k}|\tilde{\varphi}_{k\gamma}\rangle\langle\varphi_{k\gamma}| = \sum_{k=1}^{f}\sum_{\gamma}^{n_k}|\varphi_{k\gamma}\rangle\langle\tilde{\varphi}_{k\gamma}| \tag{2.18}$$

After some algebraic manipulations, inserting (2.17) in (2.15), we can get:

$$\delta_{(j\beta)}E = 4\langle\delta\varphi_{j\beta}|(1-\hat{\rho})\hat{F}|\tilde{\varphi}_{j\beta}\rangle \tag{2.19}$$

Since the lowest energy is achieved if $\delta_{(j\beta)}E$ vanishes for all j, β, namely if

$$\langle\delta\varphi_{j\beta}|(1-\hat{\rho})\hat{F}|\tilde{\varphi}_{j\beta}\rangle = 0 \quad \forall j,\beta \tag{2.20}$$

given the arbitrariness of the variation $\delta\varphi_{j\beta}$, the ELMOs that minimize the energy of the system are the ones satisfying the following equation:

$$(1-\hat{\rho})\hat{F}|\tilde{\varphi}_{j\beta}\rangle = 0 \quad \forall j,\beta \tag{2.21}$$

Introducing for each fragment j a partial density operator constructed with the only ELMOs localized on the subunit:

$$\hat{\rho}^{(j)} = \sum_{\gamma=1}^{n_j}|\tilde{\varphi}_{j\gamma}\rangle\langle\varphi_{j\gamma}| \tag{2.22}$$

the generic dual orbital can be written as

$$|\tilde{\varphi}_{k\gamma}\rangle = (1-\hat{\rho}+\hat{\rho}_k)|\varphi_{k\gamma}\rangle \tag{2.23}$$

and equation (2.21) becomes

$$(1-\hat{\rho})\hat{F}(1-\hat{\rho}+\hat{\rho}_j)|\varphi_{j\beta}\rangle = 0 \quad \forall j,\beta \tag{2.24}$$

Now, adding to both the hand-sides of equation (2.24) the following term

$$\hat{\rho}_j^{\dagger}\hat{F}(1-\hat{\rho}+\hat{\rho}_j)|\varphi_{j\beta}\rangle,$$

equation (2.24) becomes:

$$(1-\hat{\rho}+\hat{\rho}_j^{\dagger})\hat{F}(1-\hat{\rho}+\hat{\rho}_j) = \hat{\rho}_j^{\dagger}\hat{F}(1-\hat{\rho}+\hat{\rho}_j)|\varphi_{j\beta}\rangle \quad \forall j,\beta \tag{2.25}$$

Defining the following hermitian operator:

$$\hat{F}^{(j)} = (1-\hat{\rho}+\hat{\rho}_j^{\dagger})\hat{F}(1-\hat{\rho}+\hat{\rho}_j) \tag{2.26}$$

we eventually obtain

$$\hat{\rho}^{(j)}|\tilde{\varphi}_{j\beta}\rangle = \hat{\rho}_j^{\dagger}\hat{F}(1-\hat{\rho}+\hat{\rho}_j)|\varphi_{j\beta}\rangle \tag{2.27}$$

that, exploiting the definition (2.22), is equivalent to:

$$\hat{\rho}^{(j)}|\tilde{\varphi}_{j\beta}\rangle = \sum_{\gamma=1}^{n_j}|\varphi_{j\gamma}\rangle\langle\tilde{\varphi}_{j\gamma}|\hat{F}|\tilde{\varphi}_{j\beta}\rangle \tag{2.28}$$

and, consequently, equation (2.28) can be simply rewritten like this:

$$\hat{\rho}^{(j)}|\tilde{\varphi}_{j\beta}\rangle = \sum_{\gamma=1}^{n_j}\langle\varphi_{j\gamma}|\hat{F}|\varphi_{j\beta}\rangle|\varphi_{j\gamma}\rangle = \sum_{\gamma=1}^{n_j}\epsilon_{j\beta}^{(j)}|\varphi_{j\gamma}\rangle \quad \forall j,\beta \tag{2.29}$$

that is formally similar to the non-canonical Hartree-Fock equations, with the only difference that, in this case, the modified Fock operator $\hat{F}^{(j)}$ depends only on the fragment j. [1]

Now, in analogy with the Hartree-Fock equations, we apply a unitary transformation to the occupied MOs of the systems, but here, the unitary transformation mixes among themselves only the occupied ELMOs of each subunit. Given the invariance of $\hat{F}^{(j)}$ to that transformation, it is easy to show that, for each fragment, equation (2.29) simply becomes:

[1] However, it is constructed using all the occupied Molecular Orbitals of all the fragments. In this way all the equations associated with the different fragments are coupled.

$$\hat{F}^{(j)}|\varphi_{j\beta}\rangle = \epsilon_{j\beta}|\varphi_{j\beta}\rangle \quad \forall j, \beta \tag{2.30}$$

that is the canonical equation to determine Molecular Orbitals extremely localized on the generic fragment j.

For the sake of completeness, it is worth mentioning that, as already observed by Stoll (*70*) and by Smits and Altona (*81*), the ELMOs non-orthogonality may lead to convergence instabilities in the resolution of equations (2.30). This is the reason why Fornili and co-workers (*85*), following an original idea proposed by Stoll in his seminal paper, have successfully implemented an algorithm that allows the determination of Extremely Localized Molecular Orbitals simply minimizing the E energy of the ELMO wave function with respect to the coefficients of the MOs. This has been performed through a quasi-Newton procedure where an approximate Hessian is computed analytically only at the first iteration and is afterwards updated by means of the Broyden-Fletcher-Goldfarb-Shamo formula (*88*).

2.4 Rotation

In the previous section we have briefly shown the essential part of the Stoll strategy for the ELMO computation. These tail-free orbitals are particularly suitable to be transferred from a molecule to another and, therefore, we can imagine the construction of a database of ELMOs that cover all the possible functional units of a particular class of molecules (e.g., amino acids and proteins) in order to instantaneously obtain the wave function or the electron density of very large systems. To achieve this goal, the selected ELMOs must be transferred, or specifically rotated, from small model molecule geometries, where they have been initially determined, to the geometry of the target system. The rotation of the ELMOs can be performed by means of the Philipp and Friesner approach (*71*) that will be broadly discussed in this section.

At first, it is important to point out that the ELMOs are expressed as linear combinations of Cartesian Gaussian basis functions. Moreover, it is also worth noting that the method does not take into account the modifications in the overlap between basis functions caused by the variations of bond lengths or angles in the new geometry. Thus, all the transferred ELMOs should be renormalized after their rotation.

In order to determine an appropriate rotation matrix for the ELMOs coefficients, it is very important to choose a suitable reference frame both for the model and for the target molecule. As it was proposed by Philipp and Friesner, to assure the uniqueness of the rotation, we have to select three atoms for defining these two reference frames. To this purpose, let us consider the three most common situations:

- in case of a one-atom ELMO (e.g., ELMO describing core and lone pair electrons), the triad is determined by the atom in exam and two other atoms, usually the ones directly linked to it;
- in case of a bond-type ELMO, (which is the most common situation), along with the two atoms involved in the bond, it is necessary to select a third atom that is bonded to one of the other two and that properly characterizes the local dissymetry of the bond in exam (*89*);
- in the three-center ELMO case (e.g., peptide bond, carboxylate group, phenyl ring subunits, guanidine group), the three atoms are automatically selected. This case corresponds to the most delocalized orbital treated in the transfer procedure.

In order to briefly present the strategy for the of ELMO rotation proposed by Philipp and Friesner, first of all, let us indicate the triads of atoms in the model and in the target system as (A_1, A_2, A_3) and $(A_1^{'}, A_2^{'}, A_3^{'})$, respectively. These triads enable the definition of two associated reference frames that are given by the triad vectors $(\mathbf{a}, \mathbf{c}, \mathbf{d})$ and $(\mathbf{a}^{'}, \mathbf{c}^{'}, \mathbf{d}^{'})$ where $\mathbf{a}$ is the the position vector of A_2 relative to A_1 ($\mathbf{a}^{'}$ is the the position vector of $A_2^{'}$ relative to $A_1^{'}$) and with

$$\begin{cases} \mathbf{c} = \mathbf{a} \times \mathbf{b} & \text{and} \quad \mathbf{c}^{'} = \mathbf{a}^{'} \times \mathbf{b}^{'} \\ \mathbf{d} = \mathbf{c} \times \mathbf{a} & \text{and} \quad \mathbf{d}^{'} = \mathbf{c}^{'} \times \mathbf{a}^{'} \end{cases} \tag{2.31}$$

with $\mathbf{b}$ as the position vector of A_3 relative to A_1 (see Figure 2.3).

To obtain the ELMO coefficients associated with the transformation from $(\mathbf{a}, \mathbf{c}, \mathbf{d})$ to $(\mathbf{a}^{'}, \mathbf{c}^{'}, \mathbf{d}^{'})$, we have to determine a rotation matrix $\mathbf{P}$ which is composed by two consecutive transformations (see Figure 2.3):

- a rotation from the reference frame $(\mathbf{a}, \mathbf{c}, \mathbf{d})$ to the reference orthonormal frame $(\hat{x}, \hat{y}, \hat{z})$ which is associated to the matrix $\mathbf{P}_1$ (from now on Transformation 1);
- a rotation from $(\hat{x}, \hat{y}, \hat{z})$ to $(\mathbf{a}^{'}, \mathbf{c}^{'}, \mathbf{d}^{'})$ which corresponds to the matrix $\mathbf{P}_2$ (from now on Transformation 2).

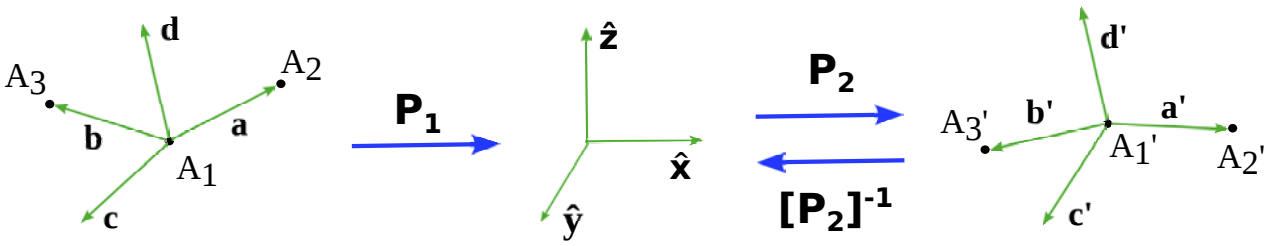

FIGURE 2.3: Definition of the reference frames for the rotation from geometry of the model molecule to the geometry of the target system.

Now, since each generic vector **v** can be expressed in the frame $(\hat{x}, \hat{y}, \hat{z})$ like this:

$$\mathbf{v} = v_x\hat{x} + v_y\hat{y} + v_z\hat{z} \tag{2.32}$$

The matrix $\mathbf{P}_1$ associated with Transformation 1 is given by the directional cosine matrix

$$\mathbf{P}_1 = \begin{pmatrix} \frac{d_x}{|\mathbf{d}|} & \frac{d_y}{|\mathbf{d}|} & \frac{d_z}{|\mathbf{d}|} \\ \frac{c_x}{|\mathbf{c}|} & \frac{c_y}{|\mathbf{c}|} & \frac{c_z}{|\mathbf{c}|} \\ \frac{a_x}{|\mathbf{a}|} & \frac{a_y}{|\mathbf{a}|} & \frac{a_z}{|\mathbf{a}|} \end{pmatrix} \tag{2.33}$$

while the matrix $\mathbf{P}_2$, which is the inverse of the matrix associated with the transformation from $(\mathbf{a}', \mathbf{c}', \mathbf{d}')$ to $(\hat{x}, \hat{y}, \hat{z})$, can be expressed like this:

$$\mathbf{P}_2 = \begin{pmatrix} \frac{d'_x}{|\mathbf{d}'|} & \frac{c'_x}{|\mathbf{c}'|} & \frac{a'_x}{|\mathbf{a}'|} \\ \frac{d'_y}{|\mathbf{d}'|} & \frac{c'_y}{|\mathbf{c}'|} & \frac{a'_y}{|\mathbf{a}'|} \\ \frac{d'_z}{|\mathbf{d}'|} & \frac{c'_z}{|\mathbf{c}'|} & \frac{a'_z}{|\mathbf{a}'|} \end{pmatrix} \tag{2.34}$$

Afterwards, we can express the rotation matrix **P** that aligns the reference frame of the model system $(\mathbf{a}, \mathbf{c}, \mathbf{d})$ to the one of the target molecule $(\mathbf{a}', \mathbf{c}', \mathbf{d}')$ as

$$\mathbf{P} = \mathbf{P}_2\mathbf{P}_1 = \begin{pmatrix} P_{xx} & P_{xy} & P_{xz} \\ P_{yx} & P_{yy} & P_{yz} \\ P_{zx} & P_{zy} & P_{zz} \end{pmatrix} \tag{2.35}$$

Using the transformation matrix **P**, it is now possible to define the rotation matrices for different types of basis functions. It is evident that the s functions are not affected, because they are invariant to rotations owing to their spherical symmetry. Hence, the matrix that

transforms the coefficients of the s functions is given by

$$\mathbf{S} = S_{11} = \mathbf{I} = 1 \tag{2.36}$$

In order to transform the p functions, we can directly apply the matrix **P** defined above (2.57). The rotation matrices for the functions with angular momentum number greater than 1 can be deduced from the matrix **P**.

For example, taking into account the normalization factors, the rotated Cartesian Gaussian basis function d'_{yz} can be expressed like

$$d'_{yz} = \frac{p'_y p'_z}{s'} \tag{2.37}$$

where, as explained above, the rotated basis function s' is simply the original s function, while the rotated Cartesian Gaussian function p'_y is given by

$$p'_y = P_{xy}p_x + P_{yy}p_y + P_{zy}p_z \tag{2.38}$$

and the rotated function p'_z by

$$p'_z = P_{xz}p_x + P_{yz}p_y + P_{zz}p_z \tag{2.39}$$

with p_x, p_y and p_z as the starting Cartesian Gaussian p functions and with $\mathbf{P}_{ij}$ as the element of the rotation matrix **P**.

Then, substituting equations (2.38) and (2.39) into equation (2.37), we get:

$$\begin{aligned} d'_{yz} = {} & P_{xy}P_{xz}\frac{p_x p_x}{s} + P_{yy}P_{yz}\frac{p_y p_y}{s} + P_{zy}P_{zz}\frac{p_z p_z}{s} \\ & + [P_{xy}P_{yz} + P_{xz}P_{yy}]\frac{p_x p_y}{s} + [P_{xy}P_{zz} + P_{xz}P_{zy}]\frac{p_x p_z}{s} + [P_{yy}P_{zz} + P_{yz}P_{zy}]\frac{p_y p_z}{s} \end{aligned} \tag{2.40}$$

And now, considering that for the three Cartesian Gaussian basis functions d_{x^2}, d_{y^2} and d_{z^2}:

$$\frac{p_k p_k}{s} = \sqrt{3}d_{k^2} \quad \text{with} \quad (k = x, y, z) \tag{2.41}$$

and that for d_{xy}, d_{xz} and d_{yz}:

$$\frac{p_k p_k}{s} = d_{hk} \quad \text{with} \quad (h, k = x, y, z \quad \text{and} \quad h \neq k) \tag{2.42}$$

equation (2.41) becomes

$$d'_{yz} = \sqrt{3}P_{xy}P_{xz}d_{x^2} + \sqrt{3}P_{yy}P_{yz}d_{y^2} + \sqrt{3}P_{zy}P_{zz}d_{z^2} + [P_{xy}P_{yz} + P_{xz}P_{yy}]\, d_{xy} + [P_{xy}P_{zz} + P_{xz}P_{zy}]\, d_{xz} + [P_{yy}P_{zz} + P_{yz}P_{zy}]\, d_{yz} \quad (2.43)$$

where the coefficients multiplying the six starting Cartesian basis functions correspond to the elements of the last column of the rotation matrix for the d functions. Following the same strategy for d_{x^2}, d_{y^2}, d_{z^2}, d_{xy} and d_{xz}, we obtain the following 6×6 **D** matrix:

$$\mathbf{D} = \begin{pmatrix} P_{xx}^2 & P_{xy}^2 & P_{xz}^2 & \sqrt{3}P_{xx}P_{xy} & \sqrt{3}P_{xx}P_{xz} & \sqrt{3}P_{xy}P_{xz} \\ P_{yx}^2 & P_{yy}^2 & P_{yz}^2 & \sqrt{3}P_{yx}P_{yy} & \sqrt{3}P_{yx}P_{yz} & \sqrt{3}P_{yy}P_{yz} \\ P_{zx}^2 & P_{zy}^2 & P_{zz}^2 & \sqrt{3}P_{zx}P_{zy} & \sqrt{3}P_{zx}P_{zz} & \sqrt{3}P_{zy}P_{zz} \\ \frac{2}{\sqrt{3}}P_{xx}P_{yx} & \frac{2}{\sqrt{3}}P_{xy}P_{yy} & \frac{2}{\sqrt{3}}P_{xz}P_{yz} & P_{xx}P_{yy} + P_{xy}P_{yx} & P_{xx}P_{yz} + P_{xz}P_{yx} & P_{xy}P_{yz} + P_{xz}P_{yy} \\ \frac{2}{\sqrt{3}}P_{xx}P_{zx} & \frac{2}{\sqrt{3}}P_{xy}P_{zy} & \frac{2}{\sqrt{3}}P_{xz}P_{zz} & P_{xx}P_{zy} + P_{xy}P_{zx} & P_{xx}P_{zz} + P_{xz}P_{zx} & P_{xy}P_{zz} + P_{xz}P_{zy} \\ \frac{2}{\sqrt{3}}P_{yx}P_{zx} & \frac{2}{\sqrt{3}}P_{yy}P_{zy} & \frac{2}{\sqrt{3}}P_{yz}P_{zz} & P_{yx}P_{zy} + P_{yy}P_{zx} & P_{yx}P_{zz} + P_{yz}P_{zx} & P_{yy}P_{zz} + P_{yz}P_{zy} \end{pmatrix} \quad (2.44)$$

Afterwards, we can simply exploit the rotation matrices (**S**, **P**, **D** and **F**, *etc.*) for the different type of basis functions (s, p, d and f, *etc.*), in order to construct the global rotation matrix **R** for the ELMO coefficients. In fact, since during a rotation only the coefficients corresponding to basis functions on the same atom and of the same type are combined among each other, the global rotation matrix **R** has a block-diagonal structure, with each block being the rotation matrix for a particular subset of basis functions (see Figure 2.4).

To conclude, it is worth noting that a transformation matrix $\mathbf{R}^{i\alpha}$ is built up for each occupied ELMO $|\varphi_{i\alpha}\rangle$. All of these matrices are then assembled in a rank-three tensor that is applied to the coefficient matrix **C** of the starting ELMOs, (namely, the ones corresponding to the model molecules), to obtain the coefficients $\mathbf{C}'$ of the transferred (rotated) ELMOs:

$$\mathbf{C}' = \mathfrak{R}\mathbf{C} \quad (2.45)$$

In this chapter, we have initially presented a non-exhaustive overview of some traditional "unitary transformation methods" to obtain Localized Molecular Orbitals. Afterwards, the main sections of this chapter have been dedicated to present the Stoll approach to

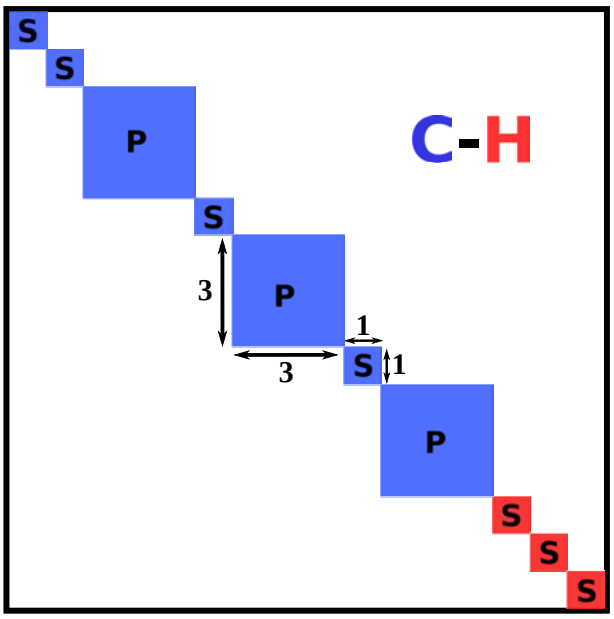

FIGURE 2.4: Block-diagonal structure of the rotation matrix **R** for an ELMO localized on a C-H bond when the split-valence basis set 6-311G is used. The blue blocks transform the carbon atom basis functions (and the related coefficients), while the red ones rotate the atomic orbitals (and the corresponding coefficients) of the hydrogen atom. The **S**, and **P** blocks are 1×1 and 3×3 submatrices, respectively, while the overall rotation matrix R has dimension 16×16.

obtain Extremely Localized Molecular Orbitals and the ELMOs rotation technique introduced by Philipp and Friesner.

Since the Extremely Localized Molecular Orbitals are in principle transferable from molecule to molecule, we have envisaged the construction of libraries of ELMOs that cover all the possible molecular fragments of the twenty natural amino acids. Our final goal is to have databanks of Extremely Localized Molecular Orbitals that will enable to reconstruct almost instantaneously approximate wave functions and electron densities of polypeptides and large proteins.

However, before starting the construction of these databases, a further assessment of the ELMO transferability was necessary. For these reasons, preliminary studies have been performed. They will be describe in detail in the next chapter together with the main results that we have obtained.

Chapter 3

Model molecule approximation and ELMOs transferability

Résumé

Dans ce chapitre, nous présentons une étude de faisabilité sur la construction d'une base de données ELMOs. Nos deux principaux objectifs sont d'examiner la transférabilité des ELMOs et de déterminer le niveau d'approximation approprié pour choisir les molécules modèles sur lesquelles seront définies les ELMOs stockées dans la base de données.

Dans ce but, nous avons testé trois niveaux d'approximation différents. Il a été démontré que l'approximation que nous avons appelé "du groupe fonctionnel le plus proche" (NFGA) donne les meilleurs résultats, tant d'un point de vue énergétique qu'en terme de reconstruction de la densité. De plus, en utilisant ce niveau d'approximation, nos résultats ont montré que les ELMOs sont réellement transférables de manière fiable. Cependant, les molécules modèles utilisées dans l'approximation NFGA sont parfois caractérisées par des interactions intramoléculaires indésirables car non présentes dans la molécule cible. Il nous semble évident que l'approximation du groupe fonctionnel le plus proche ne peut être considéré que comme une approximation à l'ordre zéro. Pour cette raison, nous envisageons dans un travail ultérieur, la possibilité de relaxer les ELMOs après leur transfert afin de tenir compte de l'environnement chimique dans la molécule cible.

En outre, nous avons observé que les densités de charge obtenues en utilisant notre approche ELMO-NFGA sont très similaires à celles obtenues par une méthode Hartree-Fock (HF). Plus particulièrement, nous avons montré que les différences entre les densités ELMO-NFGA et HF sont comparables à celles observées entre des distributions de charges HF et DFT. Pour finir, nous avons également remarqué que les effets de changement de bases d'orbitales atomiques produisent des modifications sur la densité du même

ordre de grandeur pour chacune des méthodes utilisées (ELMO-NFGA, HF or DFT).

3.1 Introduction

In this chapter, we will present the construction of new ELMOs libraries with particular attention to the study of the ELMOs transferability (*90*). As already mentioned, our main goal in this work is to construct libraries of ELMOs from small model molecules that cover all the possible functional units of the twenty natural amino acids so that we can recover almost instantaneously the wave function or electron density of proteins and polypeptides. These libraries may be considered as an alternative to the widely used databanks of pseudoatoms, which have been developed by different research groups in the framework of the popular multipole model approaches of crystallography in order to refine the crystallographic structures and to reconstruct the electron distributions of macromolecules. (*91*–*117*). A comparison of the ELMOs and pseudoatoms transferabiltity will be discussed in the next chapter.

First of all, let us mention that a database of ELMOs has been already developed by Sironi and co-workers (*118*) and has been applied to obtain the electron distribution of several polypeptides. Nevertheless, this library displays some shortcomings. On the one hand, it is limited to the STO-4G basis set and, one the other hand, the associated transfer protocol gives final geometries of the target polypeptides that are slightly different from the input ones. Even if these issues are not a serious drawback for the global reconstruction of the charge density, they obviously make the database unsuitable for our final purpose of refining crystallographic structures. The preliminary step for the construction of ELMOs libraries consists in performing ELMO calculations on suitable model molecules. The choice of these model systems obviously has a crucial influence on the quality of the databases and of the target molecule description. This point was not discussed in the work of Sironi *et al.*.

At first, we will describe in detail the general strategy of our approach: choice of the target system, the model molecules approximations, the computational methods adopted. Afterwards, we will discuss the results that have allowed us to determine the most suitable level of approximation for the construction of the model molecules. Besides, the effects of using transferred extremely localized molecular orbitals and the effects of changing the atomic basis-set in the reconstruction of macromolecular electron densities will be analysed.

3.2 Methods

3.2.1 General strategy

To assess the feasibility of constructing ELMOs databases, we have performed our calculations on a crystallographic structure of the Leu-enkephalin pentapeptide (see subsection 3.2.2). In order to investigate the ELMOs transferability and the accuracy of different model molecules approximations, which will be discussed more in detail in subsection (3.2.3), we have initially chosen as a reference the wave function and the electron density resulting from a variational ELMO calculation, namely a computation with ELMOs directly determined and optimized on the target system. Using this wave function as a reference, we have performed different comparisons with charge distributions obtained through the transfer of ELMOs previously computed on the basis of the adopted model molecule approximation.

Afterwards, we have investigated the effects of the molecular orbitals localization on the reconstruction of the charge distribution. To accomplish this task, considering different basis sets, we have performed transferred-ELMO calculations using the most accurate model molecule approximation, namely, as we will see, the Nearest Functional Group Approximation. The obtained electron densities have been then compared to those resulting from Hartree-Fock and DFT/B3LYP computations. We have also investigated the effect of the different basis-set quality on the transferability of ELMOs.

3.2.2 Description of the target system

All our calculations have been performed on a system composed of the Leu-enkephalin pentapeptide (Tyr-Gly-Gly-Phe-Leu) and three interacting water molecules (*119*). The geometry used for our investigations has been determined through an X-ray diffraction experiment conducted at 100 K with a resolution of 1.15 $Å^{-1}$. The hydrogen atoms positions, including those of the three water molecules, have been initially determined on the basis of experimental electron density peaks and then refined against experimental data. In this study, these positions were then modified to take into account the positional bias due to the X-ray diffraction method, which actually locates the hydrogen atoms bonding electrons rather than hydrogen atoms nuclei. This is the reason why, the positions of the hydrogen atoms have been properly optimized by elongating the X-H bond lengths to

match the average values obtained from neutron diffraction experiments (*120*).

The chosen polypeptide is a suitable target system for our purposes. Indeed, the considerable variety of functional groups, the presence of charged fragments and the necessity of reproducing intra-molecular and inter-molecular interactions represent major difficulties that allowed us to test the capabilities of our method.

3.2.3 Model molecules approximation

As mentioned above, the first step for the construction of ELMOs databases consists in performing ELMO calculations on a variety of small model molecules and store the obtained orbitals in a library. Obviously, the choice of these model systems has a crucial influence on the quality of the future datasets and, therefore, on the accuracy of the target molecule description. This is the reason why we have decided to test three different levels of approximation, which are schematically depicted in Figure 3.1.

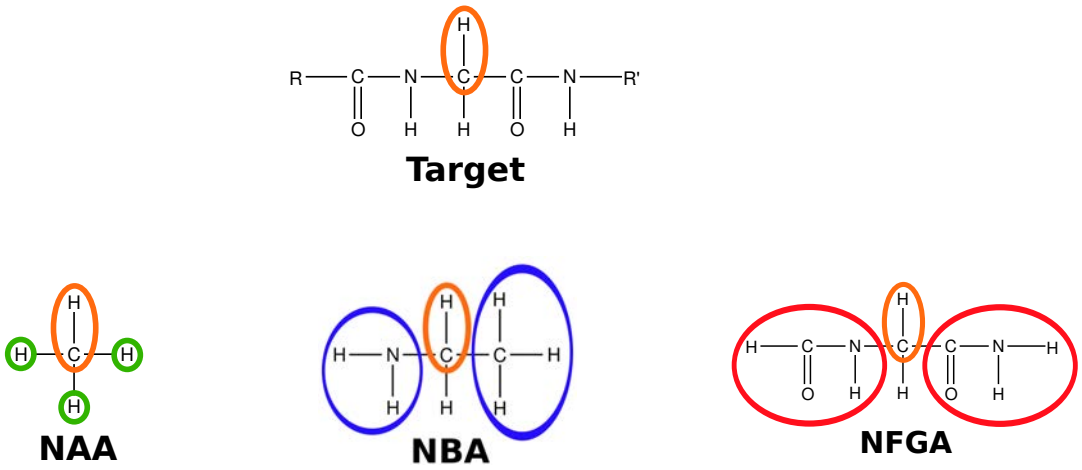

FIGURE 3.1: Model molecules approximations. The fragment in exam is framed in orange.

- the simplest Nearest Atom Approximation (NAA), which consists in considering the fragment of interest capped by hydrogen atoms. If the fragment is centred on only one atom, the first neighbour atoms of the fragment in exam are also taken into account for the construction of the model molecules;
- the intermediate Nearest Bond Approximation (NBA), which assumes that the model system is constructed considering the fragment in exam and its nearest neighbour

bonds properly capped with hydrogen atoms. In the case of bonds is involved in aromatic rings, the complete ring is taken into account. Note that for atomic fragments, NAA and NBA provide identical model molecules.

- the more elaborated Nearest Functional Group Approximation (NFGA), according to which the model molecule is constructed considering the investigated subunit and its nearest neighbour functional groups properly capped with hydrogen atoms. In the case of an asymmetric carbon, all its substituents are taken into account.

This three different levels of approximation were also extended to the cases in which the subunit of interest is involved in an intermolecular interaction (e.g., hydrogen bond), as illustrated in Figure 3.2.

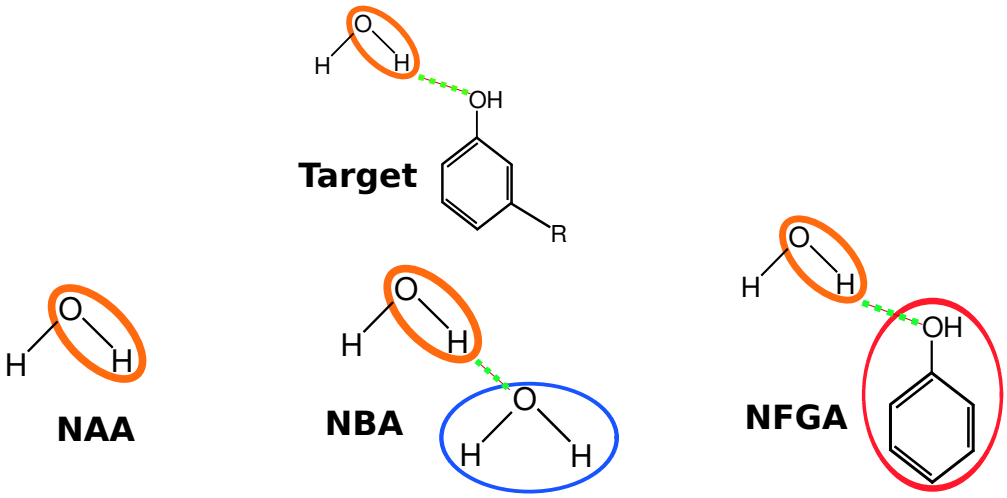

FIGURE 3.2: Extension of the model molecules approximations in the case of subunits involved in intermolecular interactions. The fragment in exam is framed in orange.

3.2.4 Computational methods

As already mentioned above, the first part of our studies consisted in assessing the transferability of the ELMOs comparing the different model molecules approximations. To achieve this goal, the variational and the transferred ELMO calculations have been carried out using a 6-31G basis set leading to a total number of 436 basis functions which is close to the current limitations of the available ELMO program (note that the program

was initially developed to perform variational ELMO calculations on small systems in order to obtain extremely localized molecular orbitals to be exported to larger molecules). Therefore, the use of larger basis sets was not considered in this case.

However, at a later stage, we have used larger and more flexible basis sets (6-311G, 6-31G(d,p) and 6-311G(d,p)) to investigate the effects of the molecular orbitals localization in the reconstruction of electron densities. In particular, we have compared the charge distributions obtained by transferring ELMOs determined on properly designed model molecules (see the Nearest Functional Group Approximation) and we have compared them to those resulting from calculations carried out at Hartree-Fock and B3LYP levels. To limit the computational effort, we have decided to perform all our preliminary computations without introducing diffuse basis functions. Nevertheless, given the nature of the system under exam, we expect that their use would not provide significantly different results from those obtained with the above mentioned basis sets.

According to the three different approximations, three sets of model molecules have been created for the calculations of the ELMOs to be transferred to the Leu-enkephalin target system. In particular, we have considered 23 different model molecules for the NAA, 31 for the NBA and 25 for NFGA. For all these model molecules, a geometry optimization at B3LYP/6-311G(d,p) level has been performed. Afterwards, the different sets of ELMOs for the transfer have been obtained carrying out variational ELMO calculations on each model molecules with the basis set 6-31G for all the approximations and, in addition, with the 6-311G, 6-31G(d,p) and 6-311G(d,p) basis sets for the NFGA.

It is worth noting that, for all the variational and transferred ELMO calculations performed in our investigations, we have almost always adopted localization schemes corresponding to the Lewis structure of the molecules. Therefore, we have used atomic subunits to describe the core and the lone-pair electrons and diatomic fragments to characterize the bond electrons pairs. The only three exceptions in our case study are represented by the following three-atom fragments which are necessary to take into account the two resonant Lewis structures of the system (see figure 3.3):

- the subunit O-C-O that treats the σ end the π electrons of the carboxylate group;
- the fragment N-C-O to describe the eight electrons of the peptide groups consisting of two σ bonds, the CO π bond and the delocalized lone pair of the nitrogen atom;

- the subunit C-C-C to describe each of the delocalized π electron pairs of the phenyl rings.

All the ELMO computations have been carried out using version 8 of the GAMESS-UK quantum chemistry package (*121*) that has been properly modified to introduce the ELMO technique, while the ELMOs have been transferred by exploiting an in-house code that implements the rotation strategy previously described (see section 2.2.5). Finally, all the other calculations (Hartree-Fock and DFT) have been performed through the Gaussian09 (*122*) suite of programs.

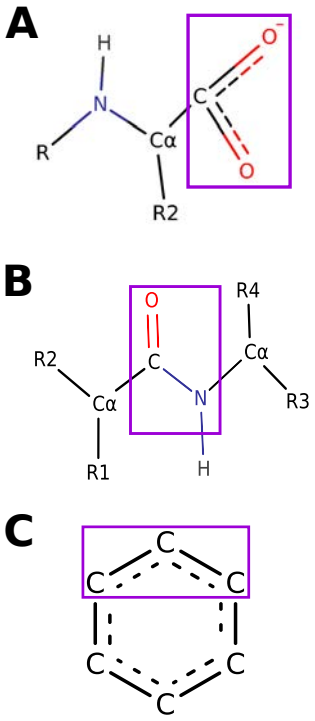

FIGURE 3.3: Three-atom fragments used to describe (A) the σ and π electrons of the carboxylate group, (B) the electrons involved in the peptide bonds, and (C) the delocalized π electron pairs of the phenyl rings.

3.2.5 Comparison of the obtained electron densities

In order to assess the different levels of approximations and the effects of the orbitals localization in the charge density reconstruction, it is crucial to compare the obtained electron densities. At first, we have performed these comparisons through Quantum

Theory Atoms In Molecules analyses (44). We have especially focused our attention on the values of the electron density at the Bond Critical Points (BCPs) $\rho(\mathbf{r}_b)$ and on the Laplacian of the electron density at the BCPs $\nabla^2\rho(\mathbf{r}_b)$. Moreover, we have also compared the net integrated atomic charges obtained by means of the integration of the different electron distributions over the QTAIM atomic basins. Since our target pentapeptide system is defined by an important number of atoms (86) and of bond critical points (98), we decided to present our discussion and our results of the QTAIM analyses in terms of average values. In particular, for $\rho(\mathbf{r}_b)$ and $\nabla^2\rho(\mathbf{r}_b)$ we will consider their mean absolute relative variations (MARVs) with respect to the reference data. The MARVs can be expressed like

$$MARV(X) = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{X_i - X_{i,ref}}{X_{i,ref}}\right| \tag{3.1}$$

where N is the number of considered values. Now, considering the comparison of the net integrated atomic charges, we have preferably used the mean absolute deviations (MADs) because very small charges may lead to huge relative variations. We define the MADs as

$$MAD(X) = \frac{1}{N}\sum_{i=1}^{N}|Q_i - Q_{i,ref}| \tag{3.2}$$

At a later stage, in order to get more global comparisons between electron densities, two real-space similarity indexes have been also computed: the real-space R value (123) (RSR) and the Walker-Mezey index $L(a, a')$ (46). The former is simply given by

$$RSR(\rho_x, \rho_y) = 100\frac{\sum_{i=1}^{n_p}|\rho_x(\mathbf{r}_i) - \rho_y(\mathbf{r}_i)|}{\sum_{i=1}^{n_p}|\rho_x(\mathbf{r}_i) + \rho_y(\mathbf{r}_i)|} \tag{3.3}$$

with n_p as the number of grid points. So, the complete similarity is reached if $RSR = 0$.

The latter allows a point-by-point comparison of two charge distributions within sets of points defined by the density shells

$$S(\rho_x, a, a') = \{\mathbf{r} : a \leq \rho_x \leq a'\} \tag{3.4}$$

$$S(\rho_y, a, a') = \{\mathbf{r} : a \leq \rho_y \leq a'\} \tag{3.5}$$

Following Walker and Mezey (46), the similarity index is defined as

$$L(a,a') = 100\ \frac{L^*(\rho_x,\rho_y,a,a') + L^*(\rho_y,\rho_x,a,a')}{2} \tag{3.6}$$

where

$$L^*(\rho_x,\rho_y,a,a') = 1 - \left[\sum_{\mathbf{r}\in S(\rho_x,a,a')} \frac{|\rho_x(\mathbf{r}) - \rho_y(\mathbf{r})|}{max(\rho_x(\mathbf{r}),\rho_y(\mathbf{r}))}\right] / n(S(\rho_x,a,a')) \tag{3.7}$$

and

$$L^*(\rho_y,\rho_x,a,a') = 1 - \left[\sum_{\mathbf{r}\in S(\rho_y,a,a')} \frac{|\rho_x(\mathbf{r}) - \rho_y(\mathbf{r})|}{max(\rho_x(\mathbf{r}),\rho_y(\mathbf{r}))}\right] / n(S(\rho_y,a,a')) \tag{3.8}$$

The $n(S(\rho_x,a,a'))$ and $n(S(\rho_y,a,a'))$ are the number of grid points belonging to the density shells $S(\rho_x,a,a')$ and $S(\rho_y,a,a')$ respectively. For this index, the complete similarity is obtained when $L(a,a') = 100$.

For the sake of completeness, all the similarity indexes have been computed by considering three-dimensional grids with a 0.083131 bohr step-size for each direction. All the grids have been constructed exploiting the Cubegen utility of the Gaussian09 package (122).

3.3 Results and discussion

3.3.1 Accuracy of the different model molecule approximations

As already mentioned in section (3.2.1), in order to evaluate the accuracy of the three different levels of model molecules approximations, we have compared the results obtained through a fully variational ELMO calculation to the ones resulting from transferred ELMO computations associated with the NAA, NBA and NFGA.

Initially, using the variational ELMO energy as a reference, we have determined the energy increase due to the transfer of ELMOs from the constructed model molecules to the target system. In table 3.1, it is easy to note that, when all the 163 occupied ELMOs are simultaneously transferred to the target system, the most sophisticated (NFGA) approximation entails, by far, the lowest energy increase. Indeed, the energy difference after a full transfer (ΔE_{Full}) is 58.52 kcal/mol for the NFGA, while it is 149.58 kcal/mol and

118.41 kcal/mol for the NAA and the NBA, respectively.

Moreover, in order to give more insights into the reliability of the different levels of approximations, we have also decided to compute the energy difference associated with the transfer of each single ELMO. To achieve this goal, we have frozen all the variationally obtained ELMOs, except the one in exam that was substituted by a transferred ELMO from the desired model molecule. Therefore, determining the energy of the modified wave function, we were able to evaluate the energy difference ΔE corresponding to the single-ELMO transfers. Due to the large number of occupied molecular orbitals, we have only reported in table 3.1 the average energy variations $\langle \Delta E \rangle$ that have been computed both over the complete set of ELMOs ($\langle \Delta E \rangle_{Overall}$) and over particular subsystems of them, such as, for example, the ELMOs localized on atomic fragments ($\langle \Delta E \rangle_{Atoms}$) or the ELMOs localized on bond fragments ($\langle \Delta E \rangle_{Bonds}$). Afterwards, in the table, we have also chosen to detail all the energy variations associated with each specific atom and bond type.

In table 3.1, we can clearly notice that the NFGA globally provides the lowest energy difference per ELMO. This trend is also confirmed by the average $\langle \Delta E \rangle$ values related to the transfer of ELMOs only localized on atoms $\langle \Delta E \rangle_{Atoms}$ or bonds $\langle \Delta E \rangle_{Bonds}$. Examining more in detail the atomic fragments, we can see that the three different approximations are quite similar if we take into account the carbon or the nitrogen atoms. However, the NFGA provides a much better description for the transfer of the three doubly occupied ELMOs (core and lone pairs) localized on the oxygen atoms. In fact, this is notably due to the inadequate description of the carbonyl groups in the case of the NAA and NBA. Furthermore, it is important to note that the nearest atom and the nearest bond approximations provide equivalent energy changes $\langle \Delta E \rangle_{Atoms}$ for the fragments centred on atoms since the same model molecules have been used in those situations.

Concerning the bond fragments, the nearest functional group approximation exhibits the lowest energy differences in almost all the cases. For example, the $\langle \Delta E \rangle$ values of the C-C and O-H bonds decrease significantly from 2.55 and 4.96 (NAA) to 0.61 and 1.77 kcal/mol (NFGA), respectively. Then, the only situation in which the NFGA provides a larger energy variation is represented by the C-O fragment of the phenol group (tyrosine residue of the Leu-enkephalin peptide) for which the nearest atom approximation gives the best agreement. Nevertheless, in this case, it is important to note that the difference

TABLE 3.1: Comparison of the Nearest Atom (NAA), Nearest Bond (NBA), and Nearest Functional Group (NFGA) Approximations: full and average energy variations (kcal/mol) associated with the transfer of ELMOs using the variational ELMO/6-31G calculation as reference. [a]

energy variation	NAA	NBA	NFGA
ΔE_{Full}	149.58	118.41	58.52
$\langle \Delta E \rangle_{Overall}$	0.92	0.73	0.36
$\langle \Delta E \rangle_{Atoms}$	0.20	0.20	0.04
$\langle \Delta E \rangle_{Bonds}$	1.33	1.04	0.56
$\langle \Delta E \rangle_{C}$	0.02	0.02	0.03
$\langle \Delta E \rangle_{N}$	0.05	0.05	0.00
$\langle \Delta E \rangle_{O}$	0.41	0.41	0.07
$\langle \Delta E \rangle_{C-H}$	0.70	0.63	0.44
$\langle \Delta E \rangle_{C-C}$	2.55	0.95	0.61
$\langle \Delta E \rangle_{C-N}$	5.62	5.44	0.69
$\langle \Delta E \rangle_{C-O}$	0.16	0.23	0.23
$\langle \Delta E \rangle_{N-H}$	3.93	3.40	0.36
$\langle \Delta E \rangle_{O-H}$	4.96	3.03	1.77
$\langle \Delta E \rangle_{(C-C-C)_{ar}}$	0.50	0.42	0.52
$\langle \Delta E \rangle_{O-C-O}$	0.42	0.29	0.29
$\langle \Delta E \rangle_{N-C-O}$	0.84	0.93	1.07

[a] The number of data used to compute the average values are reported in Table A.10.

between the NAA and the NFGA amounts only to 0.07 kcal/mol and, therefore, we can consider the three approximations practically equivalent.

Concluding the analysis of the energetic effects associated with the ELMOs transfer in Table 3.1, we can also note that, for the three atom fragments, the NFGA can not be always considered as the best approximation. Indeed, we have especially detected a non negligible discrepancy for the N-C-O fragment for which we obtain a difference of 0.23 kcal/mol between the NAA and the NFGA. Our investigations have revealed that this fact is actually due to the greater complexity of the NFGA model molecules that, although successful and useful in many situations, may sometimes be characterized by non-covalent interactions that are not present in the target system. For that reason, in these particular situations, undesired interactions artificially perturb the ELMOs to be transferred and,

consequently, the NAA and the NBA for which these interactions are not implicated, become superior to the NFGA, even if they do not completely take into account the chemical environment of the fragments of interest.

At a second stage, we have analysed the accuracy of the different levels of model molecules approximations comparing the topological properties of the obtained charge distributions. If we consider the electron density at the bond critical points, we can see that the three approximations reproduce quite well the references values associated with the variational ELMO wave function even if we can note (see Table 3.2) that the best description is the one provided by the NFGA.

TABLE 3.2: Comparison of the Nearest Atom (NAA), Nearest Bond (NBA), and Nearest Functional Group (NFGA) Approximations: Mean Absolute Relative Variations (%) of the values of the electron density and of its Laplacian at the bond critical points using the variational ELMO/6-31G calculation as reference.[a]

	ρ_{BCP}			$\nabla^2_{\rho_{BCP}}$		
bond	NAA	NBA	NFGA	NAA	NBA	NFGA
overall	1.67	1.62	1.18	5.34	5.14	2.99
C-H	1.32	1.12	0.93	4.29	3.84	3.36
C-C	0.83	0.61	0.43	1.03	1.20	0.89
$(C\text{-}C)_{ar}$	0.31	0.30	0.28	0.77	0.77	0.91
$(C\text{-}N)_{peptide}$	4.99	4.40	1.31	32.28	28.22	5.28
$(C\text{-}N)_{term}$	4.02	6.00	1.42	28.43	65.43	14.23
C_{α}-N	7.35	7.17	2.29	21.95	21.36	5.97
$(C\text{-}O)_{term}$	0.34	0.61	0.26	2.76	3.50	2.39
$(C\text{-}O)_{peptide}$	0.66	0.13	0.31	17.08	9.40	8.16
$(C\text{-}O)_{phenol}$	0.73	2.18	2.06	18.37	25.04	16.99
N-H	0.62	0.93	0.96	2.38	2.36	0.87
O-H	1.78	1.36	0.96	0.78	0.61	0.53

[a]The number of data used to compute the average values are reported in Table A.11.

Actually, we can confirm this trend also analysing more in detail the mean absolute relative variations of the electron density at the BCPs for the different types of bonds. The largest MARV for the NAA, NBA and NFGA amounts only to 7.35%, 7.17% and 2.29%

respectively, and, furthermore, in most of the situations, the NFGA provides the lowest discrepancies. In particular, we can observe that, for the $(C\text{-}N)_{peptide}$, $(C\text{-}N)_{term}$, C_{α}-N and O-H bonds, the average relative variations significantly decrease from 4.99%, 4.02%, 7.35% and 1.78% (NAA) to 1.31%, 1.42%, 2.29% and 0.96% (NFGA), respectively. Only in the case of the $(C\text{-}O)_{phenol}$ bond (as in the description of the energy changes), the NFGA (2.06%) provides a quite larger difference compared to the NAA (0.73%). We suppose that in the case of the NFGA model molecule is not able to properly mimic the description of the hydrogen bond interaction in which the fragment C-O is directly involved with a water molecule.

Concerning the Laplacian of the electron density at the BCPs, as expected, we have observed larger relative variations (see Table 3.2). Nevertheless, considering the sensitivity of this topological property (*124*, *125*), we can assert that the obtained MARVs are relatively small, which is a further evidence of the ELMOs transferability. Moreover, in this context, the NFGA generally provides again the best agreements with the reference electron density. This is particularly evident for the $(C\text{-}N)_{peptide}$ BCPs, for which the MARVs significantly drops from 32.21% and 28.22% (NAA and NBA, respectively) to 5.28% (NFGA).

In order to complete the comparison of the topological properties of the different electron densities, we have performed an analysis of the net atomic charges integrated over the atomic basins. In Table 3.3 we can easily observe that, for each atom type, all the three approximations provide charges that are very close to those resulting from the reference variational ELMO calculation. Furthermore, once again, the NFGA displays the best agreements, with the only exception noticed for the nitrogen atom for which the NFGA mean absolute deviation is smaller than the one associated with the NBA but larger than the one corresponding to the NAA. This is principally due to the difference of the terminal ammonium charge which is not perfectly described by the considered NFGA model molecule.

To conclude, as already mentioned in subsection (3.2.5), in order to obtain a more global quantity to evaluate the accuracy of the different model molecules approximations, we have used proper point-by-point similarity indexes to compare the NAA, NBA and NFGA electron densities to the reference charge distribution. In Table 3.4, if we focus

TABLE 3.3: Comparison of the Nearest Atom (NAA), Nearest Bond (NBA), and Nearest Functional Group (NFGA) Approximations: Mean Absolute Deviation (e) of the net integrated atomic charges using the variational ELMO/6-31G calculation as a reference. [a]

atom	NAA	NBA	NFGA
overall	0.056	0.048	0.037
C	0.075	0.056	0.043
N	0.051	0.061	0.053
O	0.042	0.039	0.035
H	0.047	0.043	0.032

[a]The number of data used to compute the average values are reported in Table A.12.

our attention on the RSR and the Walker-Mezey L(0.001,10) indexes, despite a great similarity among all the considered electron densities, we can note that the NFGA allowed the reconstruction of charge distribution that is globally more similar to the variational one. Moreover, since the Walker-Mezey index enables one to study the similarity in different electron density ranges by changing the a and the a' values (see equations 3.4-3.8), we can also observe that the greater global accuracy of the NFGA is even more evident in regions very far from the nuclei (L(0.001,0.01)) than in regions closer to the nuclei (index L(0.1,10)).

TABLE 3.4: Comparison of the Nearest Atom (NAA), Nearest Bond (NBA), and Nearest Functional Group (NFGA) Approximations: values of real-space R and Walker-Mezey similarity indexes (%) [a] with the variational ELMO/6-31G calculation used as reference.

similarity index	NAA	NBA	NFGA
RSR	0.74	0.70	0.51
L(0.001,10)	96.15	96.34	97.21
L(0.001,0.01)	94.60	94.85	96.01
L(0.01,0.1)	97.51	97.28	98.01
L(0.1,10)	98.55	98.64	98.95

[a]For the Walker-Mezey indicator, the electron densities are compared within the a and a' limits expressed in e/bohr^3.

To summarize, we can say that the results presented in this subsection have further confirmed the reliability of the ELMO transferability. In particular, using the RSR and different Walker-Mezey indexes, we have shown the great similarity of the NAA, NBA

and NFGA charge distributions with the reference electron density. Furthermore, we have also highlighted the small absolute relative deviations of the considered topological properties. Among the three different approximations taken into account, we have seen that the Nearest Functional Group Approximation is the one that better reproduces the variational ELMO charge density. However, we have discovered that, in some cases, the more sophisticated NFGA model molecules are characterized by undesired non covalent intramolecular interactions that artificially perturb the ELMOs to be transferred to the target system. Nevertheless, in spite of this drawback, we can conclude that the NFGA represents a decisive improvement compared to the NAA and the NBA, and therefore, it should be used for the construction of the future databases.

3.3.2 Effects of the molecular orbitals localization

To further assess the level of approximation introduced by using extremely localized molecular orbitals in the reconstruction of charge distributions, ELMO-NFGA electron densities have been compared to corresponding Hartree-Fock electron distributions considering four different basis sets (6-31G, 6-311G, 6-31G(d,p) and 6-311G(d,p)). Moreover, in order to have a reference, we have also compared B3LYP and HF charge densities. As in the previous subsection, we initially focused on two important topological properties: the electron density and its Laplacian at the bond critical points. In Table 3.5, we have reported the mean absolute relative variations (with respect to the Hartree-Fock references) obtained at the ELMO-NFGA and B3LYP levels for the values of the electron density at the bond critical points. We can easily see that, for all the basis sets, the transfers of ELMOs display quite small discrepancies with respect to the Hartree-Fock calculations. Moreover, these differences are of the same order of magnitude of the ones arising from the DFT computations. These observations are confirmed if we consider the MARVs computed for specific types of bond critical points. In fact, the largest discrepancies at the ELMO-NFGA level have been detected for the C_α-N bonds, with MARVs only in the amount of 6.11%, 6.25%, 7.02% and 6.47% for basis sets 6-31G, 6-311G, 6-31G(d,p) and 6-311G(d,p), respectively. Similarly, also the DFT calculations exhibit the largest MARVs for the C_α-N bonds, but slightly smaller than the ones associated with the ELMO-NFGA.

As in the previous subsection, we have obtained larger differences for the values of

TABLE 3.5: Effects of the Molecular Orbitals localization: Mean Absolute Relative Variations (%) of the values of the electron density at the bond critical points using the Hartee-Fock values as references.[a]

	6-31G		6-311G		6-31G(d,p)		6-311G(d,p)	
bond	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP
overall	1.83	1.34	1.90	1.33	2.16	2.76	1.99	2.74
C-H	1.08	0.59	1.19	0.38	1.20	2.09	1.12	2.21
C-C	0.42	0.51	1.68	2.21	1.57	4.08	2.01	5.21
$(C\text{-}C)_{ar}$	0.74	0.22	1.08	1.24	0.83	2.86	0.53	3.24
$(C\text{-}N)_{peptide}$	1.37	0.34	1.33	0.86	1.55	1.77	1.55	1.62
$(C\text{-}N)_{term}$	2.08	4.01	2.25	3.26	2.19	1.65	1.83	2.13
$C_{\alpha}\text{-}N$	6.11	4.68	6.25	4.63	7.02	5.45	6.47	5.07
$(C\text{-}O)_{term}$	1.78	1.19	1.68	1.00	2.04	1.76	2.13	1.25
$(C\text{-}O)_{peptide}$	0.95	1.15	0.94	0.90	1.16	1.79	1.24	1.16
$(C\text{-}O)_{phenol}$	0.89	0.81	0.97	1.05	0.58	0.19	0.45	0.31
N-H	1.07	0.80	0.63	0.17	1.07	1.64	0.86	1.63
O-H	1.24	0.92	0.69	0.51	1.11	1.06	1.54	0.44

[a]The number of data used to compute the average values are reported in Table A.11.

the Laplacian of the electron density at the bond critical points (see Table 3.6). Nevertheless, at least for the basis sets without polarization functions (e.g., 6-31G and 6-311G), the ELMO-NFGA and B3LYP methods provide absolute relative variations of the same order of magnitude for all the considered types of BCPs. Unfortunately, for the two other basis sets, the trends are much less clear. In fact, if in some circumstances the transfer of ELMOs gives much larger deviations that the DFT method (e.g., $(C\text{-}O)_{term}$ BCP for the 6-31G(d,p) basis set), in other cases the opposite behaviour is observed (e.g., $(C\text{-}N)_{term}$ BCP for the 6-311G(d,p) basis set).

Moreover, it is important to precise that we have obtained unusually small $\nabla^2_{\rho_{BCP}}$ values at the Hartree-Fock level for the 6-31G(d,p) and 6-311G(d,p) basis sets. Since we choose to work considering relative variations, these very small values have generated extremely high MARVs. Therefore, in our comparisons, we decided to exclude the C48-O39 BCP (only for the 6-31G(d,p) basis set), which corresponds to the carbonyl group of a peptide bond, and the C12-O5 BCP (for the basis sets 6-31G(d,p) and 6-311G(d,p)), which is associated with the C-O bond of the phenol group in the tyrosine residue. For

the former, the Laplacian values at HF and B3LYP levels are clearly too small for a covalent C=O bond critical point (-0.0036 and -0.03597 e/Å^5), while the one obtained through the ELMO-NFGA method is completely reasonable (-3.8758 e/Å^5). In the latter case, only the Hartree-Fock method provided too small Laplacian values (-0.3671 and 0.0180 e/Å^5 for the 6-31G(d,p) and 6-311G(d,p), respectively), with one of them even positive, which is not acceptable for a covalent bond. On the contrary, the DFT and the ELMO values are correctly negative both at the 6-31G(d,p) (-4.8117 and -4.6644 e/Å^5) and at the 6-311G(d,p) level (-6.7886 and -4.4229 e/Å^5).

Concerning the analysis of the net atomic charges reported in Table 3.7, we have noticed that, for each basis set, the ELMO-NFGA values are very similar to the HF ones. In fact, all the obtained MADs are lower than 0.1 e, while the DFT charges provide more significant discrepancies, especially for the nitrogen and the oxygen atoms, for which the largest MADs amount to 0.32 and 0.19 (6-311G(d,p) basis set).

For the sake of completeness, in Tables A.1 and A.4 in Appendix A, we have reported the simple mean relative variations for $\rho(\mathbf{r}_b)$ and $\nabla^2\rho(\mathbf{r}_b)$ with respect to the Hartree-Fock references. These signed variations, along with the maximum and minimum deviations (see tables A.2, A.3, A.5 and A.6 in Appendix A), have allowed us to observe that, generally, the ELMO-NFGA technique slightly overestimates the electron density at the BCPs, whereas it mostly underestimates the Laplacian at the BCPs with respect to the reference HF values. Even if less clear trends can be observed from the mean, maximum and minimum deviations for the net integrated atomic charges (see Table A.7, A.8 and A.9 in the Appendix A), we can conclude that the ELMO-NFGA method provides charges that are very close to the HF ones.

Finally, as in the previous subsection, we also compared the obtained electron densities using different similarity indexes. Our results are presented in Table 3.8 and, they clearly show that the degree of similarity between the ELMO-NFGA and the Hartree-Fock charge distributions is quite high. Furthermore, if we consider the Walker-Mezey L(a,a') indexes, we can deduce that the ELMO/HF similarity always increases closer to the nuclei, even if it is worth noting that, since this index gives the similarity in a percentage way, large absolute differences may be present in the core regions, which are indeed characterized by very high values of the electron density.

TABLE 3.6: Effects of the molecular orbitals localization: mean absolute relative variations (%) of the values of the Laplacian of the electron density at the bond critical points using the Hartee-Fock values as references.[a]

	6-31G		6-311G		6-31G(d,p)		6-311G(d,p)	
bond	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP
overall	6.44	5.42	7.62	8.05	14.90[b]	12.19[b]	10.01[c]	16.28[c]
C-H	3.18	1.31	2.33	2.87	2.86	7.75	2.00	9.02
C-C	4.85	4.28	11.17	14.51	7.95	17.94	9.58	21.27
$(C\text{-}C)_{ar}$	4.26	2.58	5.13	8.33	4.62	13.11	3.12	13.79
$(C\text{-}N)_{peptide}$	5.81	8.77	6.16	11.24	17.49	10.14	15.85	7.39
$(C\text{-}N)_{term}$	32.78	26.86	39.86	34.55	35.92	70.94	39.19	155.04
$C_{\alpha}\text{-}N$	29.63	16.54	30.16	11.17	45.47	49.20	43.43	61.44
$(C\text{-}O)_{term}$	26.59	25.60	21.84	7.51	232.43	15.34	77.76	32.57
$(C\text{-}O)_{peptide}$	14.76	29.35	15.39	5.09	108.58[d]	10.83[d]	53.36	44.54
$(C\text{-}O)_{phenol}$	58.69	10.33	86.32	93.98				
N-H	0.70	0.67	3.25	6.46	2.18	6.10	3.41	12.44
O-H	0.51	0.51	3.21	8.34	1.64	8.31	1.83	6.89

[a]The number of data used to compute the average values are reported in Table A.11.
[b]The values related to the C48-O39 and C12-O5 BCPs have not been considered.
[c]The value related to the C12-O5 BCP has not been considered.
[d]The value related to the C48-O39 BCP has not been considered.

In this subsection, we have seen that the electron densities obtained using our ELMO-NFGA technique are similar to the ones computed by means of Hartree-Fock calculations. Furthermore, the existing discrepancies between these two methods are quite comparable to differences observed between the HF and the DFT charge distributions.

3.3.3 Effects of basis sets

The effects of reconstructing an electron density through extremely localized molecular orbitals expanded in different basis sets have been also investigated. As in the previous subsections, we have accomplished this task by comparing the main topological properties associated with the examined charge distributions. In this case, instead of the mean absolute relative variations, we have preferred to use the mean deviations (MDs) of the

TABLE 3.7: Effects of the molecular orbitals localization: mean absolute deviations (e) of the net integrated atomic charges using the Hartee-Fock values as references. [a]

	6-31G		6-311G		6-31G(d,p)		6-311G(d,p)	
bond	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP
overall	0.041	0.051	0.042	0.065	0.046	0.102	0.050	0.098
C	0.054	0.063	0.055	0.079	0.064	0.121	0.077	0.121
N	0.054	0.134	0.056	0.200	0.046	0.282	0.048	0.319
O	0.045	0.106	0.043	0.140	0.046	0.163	0.039	0.185
H	0.031	0.020	0.031	0.023	0.034	0.039	0.035	0.036

[a]The number of data used to compute the average values are reported in Table A.12.

TABLE 3.8: Effects of the molecular orbitals localization: values of real-space R and Walker-Mezey similarity indexes (%). [a]

	6-31G		6-311G		6-31G(d,p)		6-311G(d,p)	
similarity	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP
RSR	0.69	0.74	0.71	0.91	0.75	0.90	0.81	1.02
L(0.001,10)	96.09	95.89	95.78	96.17	96.02	96.02	95.47	96.22
L(0.001,0.01)	94.34	94.20	93.85	94.67	94.49	94.60	93.73	94.90
L(0.01,0.1)	97.24	96.95	97.09	97.43	97.01	97.06	96.59	97.49
L(0.1,10)	98.65	98.65	98.68	98.03	98.46	98.03	98.42	97.54

[a]For the Walker-Mezey indicator, the electron densities are compared within the a and a' limits expressed in e/bohr^3.

considered properties:

$$MD(X) = \frac{1}{N}\sum_{i=1}^{N} X_i - X_{i,ref} \tag{3.9}$$

Concerning the $\rho(\mathbf{r}_b)$ and $\nabla^2\rho(\mathbf{r}_b)$ values, the results have been reported in Table 3.9. We can immediately observe that the topological properties for the 6-311G and the 6-31G electron densities are very close. On the contrary, when we introduce polarization functions, larger variations are detected. In particular, we have established that the effects due to the addition of polarization functions to a double-ζ basis-set (comparison of the 6-31G(d,p) and 6-31G electron densities) is practically equivalent to the consequences arising from the addition of polarization functions to a triple-ζ basis-set (comparison of the 6-311G(d,p) and 6-311G charge densities).

TABLE 3.9: Basis set effect: mean deviations of the values of the electron density e/Å^3 and of its Laplacian (e/Å^5) at the bond critical points when ELMO-NFGA calculations are performed with different basis sets.[a]

	$\langle X_{6-311G} - X_{6-31G} \rangle$		$\langle X_{6-31G(d,p)} - X_{6-31G} \rangle$		$\langle X_{6-311G(d,p)} - X_{6-311G} \rangle$	
bond	ρ_{BCP}	$\nabla^2\rho_{BCP}$	ρ_{BCP}	$\nabla^2\rho_{BCP}$	ρ_{BCP}	$\nabla^2\rho_{BCP}$
C-H	0.01	-0.74	0.19	-8.23	0.15	-6.23
C-C	0.04	-1.58	0.21	-7.55	0.17	-6.35
$(C\text{-}C)_{ar}$	0.01	-1.62	0.19	-7.04	0.14	-5.24
$(C\text{-}N)_{peptide}$	-0.01	2.11	0.17	2.62	0.13	0.71
$(C\text{-}N)_{term}$	0.01	0.23	0.15	-1.53	0.11	1.92
C_α-N	0.01	0.50	0.17	-3.19	0.12	-0.47
$(C\text{-}O)_{term}$	0.00	0.80	0.14	11.48	0.13	6.46
$(C\text{-}O)_{peptide}$	0.00	-0.62	0.15	13.36	0.13	8.71
$(C\text{-}O)_{phenol}$	0.00	1.96	0.14	5.08	0.12	3.36
N-H	0.01	2.35	0.21	-11.81	0.17	-11.94
O-H	0.03	-8.70	0.19	-15.67	0.16	-16.38

[a]The number of data used to compute the average values are reported in Table A.11.

Moreover, exactly the same trends were observed considering the net integrated atomic charges (see Table 3.10). In fact, while only very small variations have been detected for the 6-311G/6-31G comparison, more significant differences have been observed when we have compared polarized and unpolarized basis sets.

Finally, it is important to note that we have deduced similar conclusions during the investigations of the basis set effects for the Hartree-Fock and DFT methods (see 3.11, 3.12, 3.13 and 3.14). Actually, the deviations computed for these two approaches are quite comparable to the ELMO-NFGA ones, especially at the Hartree-Fock level.

TABLE 3.10: Basis set effect: mean deviations of the net integrated atomic charges (e) when ELMO-NFGA calculations are performed with different basis sets. [a]

atom	$\langle Q_{6-311G} - Q_{6-31G} \rangle$	$\langle Q_{6-31G(d,p)} - Q_{6-31G} \rangle$	$\langle Q_{6-311G(d,p)} - Q_{6-311G} \rangle$
C	-0.020	0.164	0.135
N	0.016	-0.384	-0.312
O	0.017	-0.246	-0.203
H	0.008	-0.005	-0.006

[a]The number of data used to compute the average values are reported in Table A.12.

TABLE 3.11: Basis-Set Effect: Mean Deviations of the values of the electron density (e/Å^3) and of its Laplacian (e/Å^5) at the Bond Critical Points when Hartree-Fock calculations are performed with different basis sets. [a]

	$\langle X_{6-311G} - X_{6-31G} \rangle$		$\langle X_{6-31G(d,p)} - X_{6-31G} \rangle$		$\langle X_{6-311G(d,p)} - X_{6-311G} \rangle$	
bond	ρ_{BCP}	$\nabla^2\rho_{BCP}$	ρ_{BCP}	$\nabla^2\rho_{BCP}$	ρ_{BCP}	$\nabla^2\rho_{BCP}$
C-H	0.01	-0.46	0.18	-7.89	0.15	-6.20
C-C	0.01	-0.81	0.19	-6.69	0.17	-5.96
$(C\text{-}C)_{ar}$	0.01	-1.40	0.19	-6.68	0.16	-5.49
$(C\text{-}N)_{peptide}$	-0.01	1.95	0.16	2.82	0.13	0.79
$(C\text{-}N)_{term}$	0.01	0.50	0.15	-0.97	0.12	1.35
C_α-N	0.01	0.45	0.15	-1.24	0.12	0.34
$(C\text{-}O)_{term}$	0.01	0.00	0.13	13.56	0.11	9.10
$(C\text{-}O)_{peptide}$	0.00	-0.41	0.14	14.65	0.12	10.07
$(C\text{-}O)_{phenol}$	0.00	1.96	0.15	5.77	0.12	4.19
N-H	0.01	2.18	0.21	-11.02	0.16	-12.07
O-H	0.04	-10.05	0.19	-15.12	0.13	-14.08

[a]The number of data used to compute the average values are reported in Table A.11.

TABLE 3.12: Basis set effect: mean deviations of the net integrated atomic charges (e) when Hartree-Fock calculations are performed with different basis sets. [a]

atom	$\langle Q_{6-311G} - Q_{6-31G} \rangle$	$\langle Q_{6-31G(d,p)} - Q_{6-31G} \rangle$	$\langle Q_{6-311G(d,p)} - Q_{6-311G} \rangle$
C	-0.024	0.159	0.132
N	0.030	-0.359	-0.304
O	0.014	-0.245	-0.207
H	0.009	-0.005	-0.003

[a]The number of data used to compute the average values are reported in Table A.12.

TABLE 3.13: Basis-Set Effect: Mean Deviations of the values of the electron density (e/Å^3) and of its Laplacian (e/Å^5) at the Bond Critical Points when B3LYP calculations are performed with different basis sets. [a]

	$\langle X_{6-311G} - X_{6-31G} \rangle$		$\langle X_{6-31G(d,p)} - X_{6-31G} \rangle$		$\langle X_{6-311G(d,p)} - X_{6-311G} \rangle$	
bond	ρ_{BCP}	$\nabla^2\rho_{BCP}$	ρ_{BCP}	$\nabla^2\rho_{BCP}$	ρ_{BCP}	$\nabla^2\rho_{BCP}$
C-H	-0.01	0.30	0.13	-5.61	0.11	-4.45
C-C	-0.02	0.40	0.11	-4.03	0.11	-3.91
$(C\text{-}C)_{ar}$	-0.02	-0.31	0.12	-4.01	0.11	-3.73
$(C\text{-}N)_{peptide}$	-0.02	2.32	0.13	-1.44	0.12	-3.25
$(C\text{-}N)_{term}$	0.00	0.17	0.12	-4.56	0.10	-3.88
C_α-N	-0.01	1.06	0.13	-5.11	0.11	-4.25
$(C\text{-}O)_{term}$	0.00	-3.16	0.12	8.58	0.11	5.00
$(C\text{-}O)_{peptide}$	0.01	-4.58	0.12	9.42	0.12	5.81
$(C\text{-}O)_{phenol}$	0.01	-1.33	0.13	1.96	0.11	1.31
N-H	0.00	4.42	0.15	-8.10	0.12	-8.57
O-H	0.04	-6.85	0.15	-10.51	0.11	-12.84

[a]The number of data used to compute the average values are reported in Table A.11.

TABLE 3.14: Basis set effect: mean deviations of the net integrated atomic charges (e) when B3LYP calculations are performed with different basis sets. [a]

atom	$\langle Q_{6-311G} - Q_{6-31G} \rangle$	$\langle Q_{6-31G(d,p)} - Q_{6-31G} \rangle$	$\langle Q_{6-311G(d,p)} - Q_{6-311G} \rangle$
C	-0.041	0.101	0.090
N	0.097	-0.211	-0.185
O	0.048	-0.188	-0.162
H	0.004	0.003	0.001

[a]The number of data used to compute the average values are reported in Table A.12.

3.4 Conclusions

In this Chapter, we have presented preliminary investigations on the feasibility of constructing libraries of extremely localized molecular orbitals (ELMOs). Our two main objectives were the study of the ELMOs transferability and the definition of a suitable model molecule approximation for the computation of the future ELMOs that will be stored in the databases.

In this context, we have tested three different levels of approximations for the model molecules and we have shown that the nearest functional group approximation (NFGA) is globally and by far the best one both from the energetic point of view and in terms of electron density reconstruction. Moreover, using this ELMO-NFGA technique, the obtained results on the charge distributions have confirmed that the ELMOs are reliably transferable from one molecule to another. However, the NFGA model molecules are sometimes characterized by undesired non-covalent intramolecular interactions (which are not present in the target system) that perturb the ELMOs to be transferred. Therefore, from these observations, it is obvious that the NFGA can be only regarded as a zero-order approximation. For this reason, we envisage the possibility of relaxing the ELMOs immediately after the transfer by means of proper linear-scaling techniques that suitably take into account the real chemical environment of the target system. Another possibility could be to generate a library using a mixture of approximations for the model molecules. For example, one could exploit the NFGA by default and switch to the NBA when undesirable non-covalent interactions are detected. A preliminary survey of non-covalent

interactions could be even done using a classical force field.

Moreover, we have also demonstrated that the charge densities obtained through our ELMO-NFGA approach are very similar to the corresponding Hartree-Fock ones. In particular, we have shown that the detected discrepancies between the ELMO-NFGA and the HF charge densities are comparable to the ones observed between B3LYP and Hartree-Fock electron distributions. Finally, we have seen that changing the basis-set produces comparable modifications on the electron densities independently of the method used for our calculations (ELMO-NFGA, HF or DFT).

Chapter 4

A comparison with the pseudoatoms transferability

Résumé

Précédemment, nous avons vu que les ELMOs sont facilement transférables d'une molécule à une autre. Dans ce chapitre, et afin de mieux évaluer la transférabilité des ELMOs, nous avons traité un cas spécifique. Concrètement, nous avons comparé les densités électroniques obtenues en transférant des pseudoatomes (densités atomiques asphériques très utilisées dans l'affinement de structures cristallographiques) et en transférant des orbitales moléculaires extrêmement localisées sur une structure cristallographique d'un pentapeptide.

Malgré la présence de différences inévitables et prévisibles, nos résultats ont montré que les deux méthodes de reconstruction fournissent des densités électroniques cohérentes. En particulier, nous avons observé que le transfert des ELMOs réussit relativement bien à reproduire les propriétés topologiques de la densité aux points critiques de liaisons non-covalentes. Cependant, à cause de leur nature très localisée, l'utilisation des ELMOs entraîne généralement une surestimation des charges associées aux différents fragments du polypeptide. A nouveau, si l'on veut surmonter cet obstacle, il sera nécessaire de développer ultérieurement une approche qui permettra de relaxer les ELMOs après leur transfert sur la molécule cible.

Les résultats présentés dans ce chapitre ont permis de démontrer que la transférabilité des ELMOs est fiable et qu'elle permet de reconstruire des densités électroniques de qualité acceptable pour un temps de calcul très réduit. Ces résultats nous encouragent donc à poursuivre nos recherches vers la construction de bases de données plus générales, ayant pour objectif final leur utilisation pour 1) raffiner des structures cristallographiques et

2) estimer de manière extrêmement rapide les propriétés physico-chimiques de macromolécules.

4.1 Introduction

In this Chapter, we will compare the ELMO method developed in our work with the usual pseudoatom approach used in crystallographic studies (*126*). To this aim, we will first make a brief introduction to crystallography and the theoretical models currently used to analyse experimental data.

Crystallography, the science of investigating crystals at an atomic scale, is based on different experimental methods, (e.g., X-ray diffraction or neutron diffraction) that permit to obtain structural information.

The crystal diffraction is an interference effect in which the angle of diffraction is connected both to the wavelength of the incident beam and to the periodicity of the crystal (Figure 4.1). The condition for X-ray diffraction to be observe is dictated by Bragg law (*127*)

$$n\ \lambda = 2d_{hkl}\ sin(\theta_{hkl}) \tag{4.1}$$

where n is an integer number known as order of diffraction, λ is the wavelength of the incident beam, d_{hkl} is the distance between two crystal planes, and θ_{hkl} is the Bragg angle between the (hkl) planes and the diffracted beam.

According to Bragg law, we can easily see that the smaller the distance d_{hkl} is, the larger $sin(\theta_{hkl})$ becomes. In other words, when the planes are very close to each other, the crystal will diffract at high-angles. For that reason, the high angles reflections correspond to high-resolution X-ray diffraction data, which allow to get more precise information about the electron density in a crystal.

Initially, X-ray diffraction enabled only the determination of the atomic positions in crystals (0.9-1.5Å). More recently, exploiting significant advancements in experimental conditions and techniques, the X-ray diffraction experiments provide more details at subatomic scale (0.6-0.9Å), such as the deformation of the electron density due to bonds and to intermolecular interactions. Nowadays, thanks to further technological improvements, we may even obtain higher-resolution data (0.3-0.6Å) that potentially can give access to quantitative descriptions of the electron distributions. However, to keep pace with the technical advancements, parallel improvements in the theoretical field are needed. In this respect, different electron distribution models have been introduced over the years and continue to be proposed. In the following, we will briefly describe two models that are widely used by the crystallography community: the basic Independent Atom Model

(IAM) and the multipole model developed by Stewart (*128*) and by Hansen and Coppens (*129*).

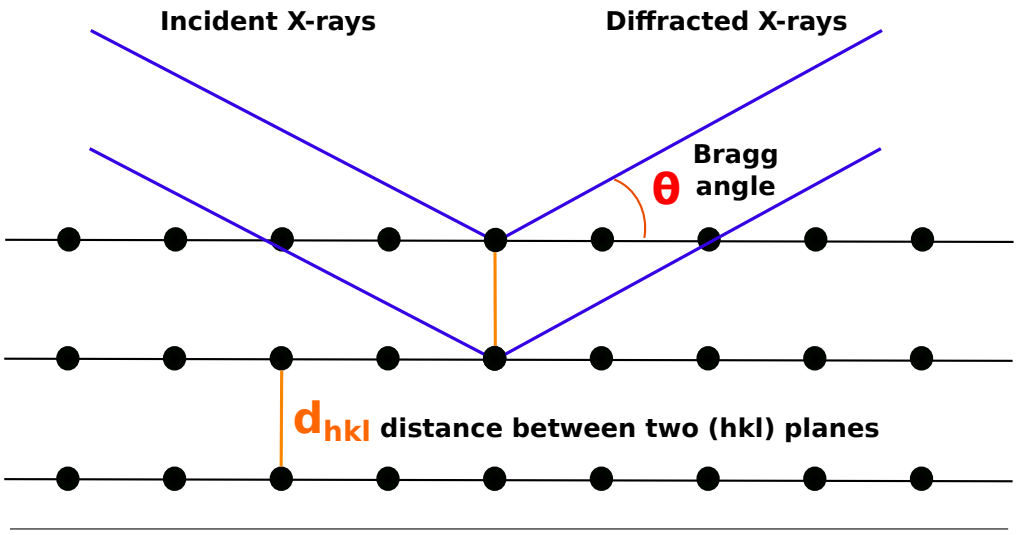

FIGURE 4.1: Schematic representation of the Bragg law.

4.1.1 The Independent Atom Model

In the independent atom model each atom is treated independently because it is assumed that the atomic electronic distribution is not affected by its chemical environment. Therefore, the molecular charge density is simply written as a sum of neutral spherical atomic distributions, namely

$$\rho(\mathbf{r}) = \sum_{i}^{N} (\rho_{\mathbf{r}} - \mathbf{R}_i) \tag{4.2}$$

where $\mathbf{R}_i$ is the position of the i-th nuclei and N is the total number of atoms.

The IAM is an extensively used approximation for the structural refinement of macromolecular crystallographic structures. Nevertheless, in the case of smaller systems, nowadays it is possible to obtain high-resolution X-ray diffraction data (d<0.7Å) whose wealth of information cannot be exploited by the independent atom model. In fact, this approach does not take into account the perturbations due to the chemical bonds and the intermolecular interactions. Therefore, in order to reproduce these deformations, more sophisticated charge density models have been introduced.

4.1.2 The Hansen and Coppens multipole model

Over the years, several aspherical atomic charge density models have been proposed by many research groups (Dawson (*130*), Demarco and Weiss (*131*), Hirshfeld (*132*)) to go beyond the independent atom model and take into account the aspherical deformation of the atomic electron densities. However in this context, among all the proposed techniques, the multipole strategy proposed by Stewart (*128*) and by Hansen and Coppens (*129*) remains by far the most largely used. Here, we will briefly illustrate the approach considering the Hansen and Coppens formalism, according to which each atomic charge distribution can be simply expressed as a sum of three terms:

$$\rho_{atom}(\mathbf{r}) = \underbrace{\rho_{core}(r)}_{core\ part} + \underbrace{\underbrace{P_{val}\kappa^3\rho_{val}(\kappa r)}_{spherical\ part:\ \rho_{val}^{sph}(\mathbf{r})} + \underbrace{\sum_{l=0}^{l_{max}}(\kappa')^3 R_l(\kappa' r)\sum_{m=0}^{l} P_{lm\pm}Y_{lm\pm}(\theta,\varphi)}_{aspherical\ part:\ \rho_{val}^{asph}(\mathbf{r})}}_{valence\ part} \tag{4.3}$$

The first ($\rho_{core}(r)$) and the second ($\rho_{val}^{sph}(\mathbf{r})$) terms are spherically averaged atomic core and valence density distributions, respectively. The former is equivalent to the one discussed in the IAM model (see equation (4.2)), while the latter is weighted by an atomic valence population P_{val} and is modulated by a radial expansion-contraction coefficient κ (expansion for $\kappa < 1$ and contraction for $\kappa > 1$). The κ formalism has been later extended by Coppens *et al.* in 1979. The deformation part of the valence electron density, namely, the third term in equation (4.3) is expressed by means of angular real spherical harmonics ($Y_{lm\pm}(\theta,\varphi)$) and through radial Slater-type functions that can be expressed like this:

$$R_l(\kappa' r) = \frac{\zeta^{n_l+3}}{(n_l+2)!}\,\kappa'^{n_l} r^{n_l}\, exp(-\zeta\kappa' r) \tag{4.4}$$

where κ' is another expansion-contraction coefficient and ζ is the Slater orbital exponent. The weight of the spherical harmonics is given by the population multipole parameters $P_{lm\pm}$ and, the more this term differs from zero, the more the angular contribution to the aspherical charge distribution is important. In equation (4.3), the sum over the l index directly depends on the level of the multipolar expansion:

- The first order is represented by a dipolar level ($l = 1$)
- The second order to a quadripolar expansion ($l = 2$)

- The third order is represented by an octupolar level $(l = 3)$
- The fourth order to a hexadecapolar expansion $(l = 4)$

A third-order multipolar development is usually used for carbon, nitrogen and oxygen atoms, while at least hexadecapolar level is commonly considered for heavier atoms (e.g., sulfur, phosphorus and transition metals). The simple bond-oriented model or a quadrupolar expansion is generally used for hydrogen atoms. In fact, this low level of expansion is enough since the X-ray diffraction is not able to collect precise information about the hydrogen localization in the molecular structure.

4.1.3 The pseudoatoms

The limited resolution of X-ray diffraction and the thermal motions of atoms complicate the crystallographic refinement of macromolecules. To overcome these drawbacks, it is necessary to start the refinement using a charge which is already similar as much as possible to the desired and expected final density. This is possible thanks to the transferability properties of the pseudoatoms models (*133*). This model is based on the observation that multipole parameters centred on atoms are reliably transferable from system to system if they are characterized by a similar chemical environment. Based on this assumption, several pseudoatoms databanks have been constructed in order to obtain electron densities and to perform accurate structural refinements of very large biomolecules.

The first pseudoatoms library was constructed in 1995 by Pichon-Pesme and co-workers to recover the electron density of a peptide backbone using multipolar parameters previously obtained from the experimental charge distributions of small peptides and amino acids (*119*). This pioneering work has led the Nancy group to the construction of the Experimental Library of Multipolar Atom Model (ELMAM), (*97*, *101*, *119*) which contains all the possible chemically unique pseudoatoms in the 20 natural amino acids. The first attempt of multipolar refinement on the LBZ octapeptide system by means of a transfer of pseudoatoms parameters has been successfully realized by Jelsch *et al.* (*134*). To enlarge the field of applications of the ELMAM database, an improved version (ELMAM2) has been developed (*95*, *135*). In fact, the databank has been extended to general functional groups usually found in organic systems, exploiting the multipolar refinements of a selected set of 54 high-resolution structures. The principal strategy to construct the ELMAM2 database was to store average electron density parameters associated with atom

types having a similar chemical environment. An automatic procedure based on the chemical connectivity recognition of the atom types, has been afterwards implemented to transfer the proper multipolar parameters to the crystal structures from the ELMAM2 generalized databank.

Taking advantage of the current power of modern computers, an alternative way in the construction of pseudoatoms databases is offered by the systematic analysis of theoretical densities that can be easily computed for small or medium size molecules. In fact, following this strategy, two other libraries have been created independently by two different research groups: the University at Buffalo Pseudoatom Databank (UBDB)(*112*, *113*, *115*, *117*) and the Invariom database (*102*–*105*, *108*).

In UBDB case, the X-ray diffraction data have been generated from theoretically calculated electron densities of isolated tripeptide molecules at the DFT/B3LYP level of theory using the 6-31G(d,p) basis set. In particular, the structure factors have been computed for reciprocal-lattice points corresponding to a pseudocubic cell with 30 Å edges. Using the theoretical structure factors, multipole model analyses have been performed for all the considered molecules and, similarly to the ELMAM2 approach, the multipolar parameters were averaged over a family of chemically similar pseudoatoms. Initially the molecular structures of the tripeptides have been obtained by geometry optimizations at the molecular mechanics level, while, later, Volkov and co-workers chose to use experimental molecular structures retrieved from the Cambridge Structural Database (CSD) in order to get a representative sample of solid-state conformations (*115*). Nowadays, the UBDB is used in a large number of applications. In particular, this library has been successfully employed to estimate the electrostatic interaction energy of small biomolecular systems and ligand-protein complexes (*111*, *112*).

In the strategy proposed by Dittrich and co-workers (*102*–*105*, *108*), the aspherical scaterring factors are based on Invarioms (invariant atoms) that are inter-molecular transfer-invariant pseudoatoms. The Invariom multipole populations have been obtained from DFT/B3LYP geometry optimizations of proper model molecule using the D95++(3df,3pd) basis set. Also in this case, the structure factors have been determined for reciprocal-lattice points corresponding to a pseudocubic cell with 30 Å edges. However, unlike the two procedures described above, here, the multipolar parameters have been obtained for

all the unique model compounds and have not been averaged. The chemical environment of each pseudoatom has been mimicked considering model molecules that include nearest neighbour atoms for single-bonded systems and next-nearest neighbour atoms for delocalized or mesomeric systems. The Invarioms library has been used in many applications, such as the calculation of electrostatic potential of macromolecules or the refinement of charge distributions from X-ray diffraction data. (*102*).

All the strategies mentioned above share the same objective and philosophy, even if each of them presents its own specific pros and cons. The main advantage of the experimental ELMAM and ELMAM2 consists in the possibility of intrinsically capture, in an average way, the influence of the chemical environment as found in the crystalline state, such as hydrogen bonds. However, the extent to which these effects can be extracted from experimental structure factors remains to be explored. Moreover, since the stored multipole parameters are determined after high-resolution X-ray diffraction data experiments, the major shortcoming associated with the experimental pseudoatoms libraries is their lack of flexibility. Actually, the possibility to add new pseudoatom types in the ELMAM dabases is directly connected to the availability of compatible sets of suitable experimental data. This drawback is overcome by the use of *ab initio* calculations on which the UBDB and Invarioms libraries are based. In these cases, the addition of a missing pseudoatom can be done by selecting appropriate CSD entries for the UBDB procedure or designing an *ad hoc* model molecule containing the pseudoatom type of interest in the case of the Invariom technique. Nevertheless, the crystal-field influence is not taken into account in these two approaches, even if, in the UBDB databank, the use of crystal geometries partially reduce this drawback.

The multipolar parameters and charge distributions from ELMAM and UBDB have been compared by Pichon-Presme *et al.* (*99*) and Volkov *et al.* (*136*). Furthermore, some similarities between the electron densities and bond critical points properties obtained from UBDB and Invarioms have been investigated (*137*) on cyclosporine A. Recently, a more global comparison between all the three existing databases has been performed by Bak *et al.* (*93*). The main conclusion is that the refinements employing pseudoatoms from the ELMAM, ELMAM2, UBDB and Invariom databanks reproduce very well the geometries optimized in *ab initio* periodic calculations and no significant discrepancies have been detected in geometries resulting from individual databases. The only large

differences have been discovered for the electrostatic properties, probably due to dissimilarities observed in the valence regions of polar atoms.

Finally, for the sake of completeness, it is important to note that a new experimental library of pseudoatoms based on supramolecular synthon fragments (SBFA) has been also recently proposed by Guru Row and co-workers (*138*, *139*). The SBFA exploits the modularity of the supramolecular synthon to obtain transferable charge densities derived from multipolar parameters on structural fragments. It has been shown that the SBFA method is generally applicable to generate charge density maps exploiting topological informations obtained from different intermolecular regions.

4.1.4 Beyond atomistic models

Although nowadays the pseudoatoms databanks seem well-established and quite powerful tools to elude the main hindrances associated with charge density studies of macromolecules, a completely new alternative way to accomplish this task has been recently envisaged: the possibility of exploiting the so-called extremely localized molecular orbitals (ELMOs).

In the previous chapter, we have presented a very detailed study to confirm that the ELMOs are indeed reliably transferable to very large systems. Since pseudoatoms transferability is widely exploited in the refinement of crystallographic structures, a comparison between the ELMOs and the pseudoatoms transferability is crucial to further assess the capabilities of the ELMOs transfer. To achieve this goal, we have carefully compared the charge distributions obtained through the transfer of extremely localized molecular orbitals and through the transfer of pseudoatoms considering the ELMAM2 and the UBDB databases. The results of this comparison will be reported and discussed in the next sections. In particular, we will focus our attention on the topological properties of the obtained electron densities both at the covalent and non covalent bond critical points. Furthermore, more global descriptors will be also considered and discussed, such as similarity indexes and net integrated charges of specific fragments.

4.2 Computational Details

4.2.1 Methods

As in the previous chapter, our target system consists in the Leu-enkephalin pentapeptide (Tyr-Gly-Gly-Phe-Leu) and three interacting water molecules (See Figure 4.2). The geometry used for all our calculations has been determined through an X-ray diffraction experiment conducted at 100 K with a resolution of 1.15 $Å^{-1}$. As already mentioned in the previous section, in the present investigations, the charge density of the polypeptide has been reconstructed both by means of the transfer of pseudoatoms and the transfer of extremely localized molecular orbitals.

Concerning the pseudoatoms, we have used the experimental databank ELMAM2 exploiting the MoPro software(*140*, *141*) (version June 2015) and the theoretical database UBDB (*112*, *115*, *117*) (version 2012) using the program LSDB have been taken into account. For each database, no "missing atom types" were detected during the transfer procedure. Because the transfer involves averaged values of atomic valence population P_{val} (see equation 4.3), at the end of the transfer procedure, small differences from the expected formal charge of the studied compound can be observed. In our study, the overall system (Leu-enkephalin and the three water molecules) is expected to be neutral with a total of 240 valence electrons. When the ELMAM2 database has been used, the transfer of the 86 pseudoatoms has led to a charge deficit of 0.0681 electron, which was subsequently corrected by uniformly adding +0.00079 e to each atomic valence populations. After the transfer of the multipolar parameters from the UBDB databank, the observed deviation was more important, with a charge deficit of 0.9699 e. In order to correct this deficiency, the nonuniform *sigmaPv* charge scaling method has been used as implemented in the LSBD software. The atomic valence populations have been shifted ranging from +0.00072 e for the hydrogen atoms of the water molecules to +0.032 e for the carbon atoms of the leucine side chain. These neutralization procedures have been performed on the whole system, without treating the poplypeptide and the water molecules independently. However, the obtained valence populations based charges of the water molecules and of the Leu-enkephalin were very small, with -0.00062 e/ +0.0019 e, and -0.0039 e / +0.0117 e respectively for ELMAM2 and UBDB, reflecting the robustness of the multipolar pseudoatoms transferability technique.

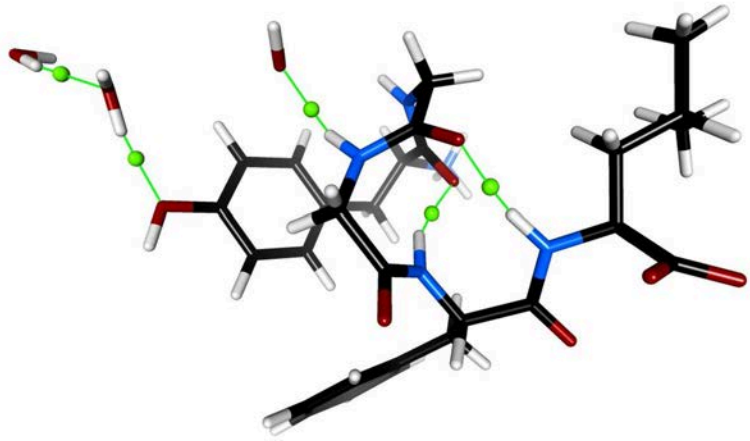

FIGURE 4.2: Leu-enkephalin pentapeptide and interacting water molecules. The five hydrogen-bond interactions and the corresponding bond critical points are explicitly shown (green spheres).

In the case of the extremely localized molecular orbitals, we have followed the strategy already described in the previous chapters. At first, the desired ELMOs have been computed on small model molecules, and at a later stage, they have been properly transferred to the target system adopting the rotation technique proposed by Philipp and Friesner (see section 2.4). In Chapter 3, we have concluded that the nearest functional group approximation (NFGA) model molecules are sometimes characterized by undesired non-covalent intramolecular interactions, which are not present in the target system and which consequently perturb the ELMOs to be transferred. This is the reason why, we have decided to consider two different model molecules approximations to determine the ELMOs; namely, the NFGA and the less sophisticated nearest bond approximation (NBA) have been used in order to get the best possible description of the target system using transferred ELMOs. All the ELMOs calculations have been performed considering six different basis sets exploiting a modified version of the GAMESS-UK quantum chemistry package (*121*). For the sake of clarity, we will only present present and discuss the results obtained at the 6-311G, 6-311G(d,p) and 6-311+G(2d,2p) levels. The analogous results obtained with the 6-31G, 6-31G(d,p) and 6-311+G(d,p) sets of basis functions are given in Appendix B.

4.2.2 Comparison of the electron densities

In order to compare the ELMO and the pseudoatoms transferability, we have examined the resulting electron densities obtained using the different methods. The analyses of the ELMO charge distribution have been carried out with the AIMAll package (version 13.11.14) (*142*), while the ones of the pseudoatoms electron densities have been performed using the MoPro software.

We have decided to mainly focus our attention on the values of the electron density $\rho(\mathbf{r}_b)$ and of its Laplacian $\nabla^2\rho(\mathbf{r}_b)$ both at the covalent and non-covalent bond critical points. Considering the large number of atoms, and consequently, of covalent BCPs, the QTAIM properties will be presented only as average values for different types of bonds occurring in the system. Furthermore, in order to obtain a more global comparison of the similarities/dissimilarities between the different considered topological properties, we have computed the following index for each couple (A, B) of methods:

$$R_{A-B}(X) = 100\frac{\sum_{i=1}^{n_{BCP}} |X_A(\mathbf{r}_i) - X_B(\mathbf{r}_i)|}{\sum_{i=1}^{n_{BCP}} |X_A(\mathbf{r}_i) + X_B(\mathbf{r}_i)|} \tag{4.5}$$

where n_{BCP} is the total number of bond critical points in exam and X is the considered topological property. The complete similarity is reached if $R_{A-B}(X) = 0$.

The features of the electron density at the non-covalent BCPs will be discussed more in details. Indeed, for five critical points corresponding to hydrogen-bond interactions, three other topological properties have been also taken into account: the kinetic energy density $G(\mathbf{r}_b)$, the potential energy density $V(\mathbf{r}_b)$ and the positive curvature of the electron distribution $\lambda_3(\mathbf{r}_b)$, which have been rigorously defined by Bader in his monograph *Atoms in Molecules: A Quantum Theory* (*44*). We have compared our results to the ones obtained through the use of three empirical exponential relations that have been proposed by Espinosa and co-workers (*143*–*146*) analysing topological and structural data from accurate electron densities studies that involve X-H$\cdots$O (X = C, N, O) hydrogen bonds:

$$G(\mathbf{r}_b) = 12(2) \times 10^3 exp\,[-2.73(9)\; d(H\cdots O)] \tag{4.6}$$

$$V(\mathbf{r}_b) = -50.0(1.1) \times 10^3 exp\,[-3.6\; d(H\cdots O)] \tag{4.7}$$

$$\lambda(\mathbf{r}_b) = 0.41(8) \times 10^3 exp\,[-2.4(1)\; d(H \cdots O)] \tag{4.8}$$

It is very important to precise that the validity of these relations is still an open problem in the charge density community. For exemple, in a very recent investigation, the reliability of the Espinosa's relation for the estimation of interactions energies between molecules in crystals has been examined and discussed by Spackman (*147*). In our present study, the empirical exponential relations (equations (4.6), (4.7) and (4.8)) have been mainly used in order to have a further reference for our comparison between the ELMOs and pseudoatoms transferability, without the aim of exploiting them to compute intermolecular interaction energies.

Moreover, we have also compared the net atomic charges obtained through the integration of the different electron distributions over the QTAIM atomic basins. Instead of our previous studies (Chapter 3), where we have decided to consider the individual atomic charges, here the charges associated with the main "functional groups" of the polypeptide backbone (i.e., the amino group, the carboxylic group, and the four peptide groups) and with the side chains (including in each case the C_α-H bond) have been taken into account.

Finally, in order to get more global comparisons between the different electron densities, two real-space similarity indexes have been also computed: the Walker-Mezey similarity indicator (*46*) (see 3.2.5) and the real-space R (RSR) value (*123*) (see 3.2.5).

4.3 Results and discussion

4.3.1 Topological properties at the bond critical points

For the comparison of the ELMOs and the pseudoatoms transferability, we have decided to focus our attention on some topological properties of the reconstructed electron densities. At first, we have analysed the charge distributions at the covalent bond critical points. It is worth noting that, both for the analysis of the electron density $\rho(\mathbf{r}_b)$ and of its Laplacian $\nabla^2\rho(\mathbf{r}_b)$, only one value has been considered for the $(C\text{-}N)_{term}$ and the $(C\text{-}O)_{phenol}$ bonds. We can easily observe in Table 4.1 that the different methods provide quite similar and reasonable charge distributions. Using the ELMAM2 results as references, we can

especially note that, within each reconstruction technique, the largest differences amount to -0.13 e/Å^3 (UBDB), -0.30 e/Å^3 (ELMO/6-311G), 0.18 e/Å^3 (ELMO/6-311G(d,p)), and -0.16 e/Å^3 (ELMO/6-311+G(2d,2p)). Moreover, concerning the ELMO transfer, it is possible to detect that, exploiting the 6-311G basis set, the values of the electron density at the BCPs are lower than the ELMAM2 ones for almost all the covalent bond types. On the contrary, the $\rho(\mathbf{r}_b)$ values at the bond critical points generally increase if larger and more flexible basis sets are employed. Furthermore, we can also observe in Table 4.1 that for polar bonds such as O-H or $(C\text{-}O)_{term}$, the ELMO values that are closest to the ELMAM2 and UBDB ones are those obtained with the 6-311+G(2d,2p) basis-set. Nevertheless, considering less polar bonds (e.g., C-H, C-C), the "unpolarized" basis-set 6-311G allows one to get values that are the closest to the ones obtained transferring pseudoatoms.

TABLE 4.1: Average values of the electron distributions at the covalent bond critical points (e/Å^3) for each bond type after the transfer of pseudoatoms and ELMOs. [a]

	pseudoatoms		ELMOs transfer		
bond	ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
C-H	1.84	1.89	1.81	1.96	2.01
C-C	1.61	1.66	1.62	1.79	1.77
$(C\text{-}C)_{ar}$	2.12	2.08	2.02	2.16	2.20
$(C\text{-}N)_{peptide}$	2.27	2.26	2.15	2.27	2.35
$(C\text{-}N)_{term}$	1.62	1.56	1.32	1.54	1.46
C_α-N	1.73	1.74	1.66	1.79	1.84
$(C\text{-}O)_{term}$	2.66	2.55	2.46	2.58	2.65
$(C\text{-}O)_{peptide}$	2.65	2.70	2.59	2.73	2.81
$(C\text{-}O)_{phenol}$	2.10	2.00	1.88	1.99	2.06
N-H	2.26	2.19	2.15	2.32	2.37
O-H	2.52	2.39	2.33	2.49	2.51

[a] The number of data used to compute the average values are reported in Table A.11.

At a second stage, we have also taken into account the Laplacian of the electron density at the covalent bond critical points (see Table 4.2). As expected, we have observed much larger variabilities and, with the respect to the reference ELMAM2 values the differences amount to 18.6 e/Å^5 (UBDB), 10.9 e/Å^5 (ELMO/6-311G), 15.5 e/Å^5 (ELMO/6-311G(d,p)), and -14.0 e/Å^5 (ELMO/6-311+G(2d,2p)). Even if it is quite difficult to discern

a real trend, we can generally note that, for almost all the bond types, the different reconstruction strategies present results that are of the same order of magnitude and, above all, the lowest discrepancies are not necessarily always between the two pseudotatoms techniques. Moreover, it is important to point out that, within all the different methods and, in particular for the ELMOs transferred, all the Laplacian values are negative, as it is expected for covalent bond interactions.

TABLE 4.2: Average values of the Laplacian of the electron density at the covalent bond critical points ($e/Å^5$) for each bond type after the transfer of pseudoatoms and ELMOs. [a]

	pseudoatoms		ELMOs transfer		
bond	ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
C-H	-17.9	-20.2	-19.7	-25.9	-28.6
C-C	-9.0	-11.3	-12.9	-19.2	-17.2
$(C\text{-}C)_{ar}$	-18.4	-17.3	-19.5	-24.8	-26.0
$(C\text{-}N)_{peptide}$	-22.1	-23.7	-20.9	-18.0	-29.8
$(C\text{-}N)_{term}$	-6.9	-9.6	-1.8	-6.4	-6.8
$C_{\alpha}\text{-}N$	-9.4	-10.1	-14.1	-16.0	-23.4
$(C\text{-}O)_{term}$	-29.8	-24.2	-21.0	-14.3	-24.7
$(C\text{-}O)_{peptide}$	-22.1	-24.5	-20.8	-12.4	-23.8
$(C\text{-}O)_{phenol}$	-17.6	-17.5	-8.2	-4.2	-14.3
N-H	-37.0	-28.5	-32.6	-44.5	-46.8
O-H	-58.0	-39.4	-47.1	-64.0	-69.0

[a] The number of data used to compute the average values are reported in Table A.11.

We have also analysed the behaviour of the electron density at the non-covalent bond critical points. At first, it is worth noting that the topological analyses of the five considered charge distributions have detected the same non-covalent interactions (five hydrogen bonds and other eight weak interactions). All of them are correctly characterized by a BCP with a positive value of the Laplacian of the electron density.

In Table 4.3, where both the hydrogen bonds and weak interactions are considered, it is easy to observe a reduced variability compared to the covalent case. Indeed, considering only the five hydrogen-bond interactions and using again the ELMAM2 values as a reference, the largest differences detected for each method are -0.070 $e/Å^3$ (UBDB), -0.058

e/Å^3 (ELMO/6-311G), -0.081 e/Å^3 (ELMO/6-311G(d,p)), and -0.088 e/Å^3 (ELMO/6-311+G(2d,2p)). Furthermore, only taking into account the eight weak interactions, a very good agreement between the pseudoatoms and the ELMO methods has been obtained. Nevertheless, in the case of the five hydrogen-bonds, for all the basis sets, the transferred ELMOs systematically slightly underestimate the values of the electron density at the non-covalent BCPs.

TABLE 4.3: Values of the electron distribution at the non-covalent bond critical points (e/Å^3) after the transfer of pseudoatoms and ELMOs.

	pseudoatoms		ELMOs transfer		
non-covalent interaction	ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
N31-H37$\cdots$Ow84	0.191	0.191	0.186	0.174	0.163
Ow81-Hw83$\cdots$O5	0.238	0.176	0.189	0.168	0.158
Ow78-Hw80$\cdots$Ow81	0.134	0.132	0.113	0.103	0.103
N58-H74$\cdots$O25	0.224	0.154	0.166	0.143	0.136
N38-H57$\cdots$O6	0.165	0.166	0.145	0.133	0.131
C11$\cdots$H35	0.024	0.028	0.030	0.030	0.029
C13$\cdots$O84	0.040	0.041	0.046	0.046	0.045
H17$\cdots$H50	0.022	0.024	0.020	0.021	0.021
H19$\cdots$C42	0.064	0.069	0.065	0.063	0.064
H20$\cdots$C46	0.017	0.018	0.023	0.023	0.022
H21$\cdots$O81	0.029	0.032	0.035	0.036	0.037
H28$\cdots$H65	0.021	0.025	0.022	0.023	0.023
H56$\cdots$O32	0.052	0.052	0.048	0.047	0.049

Then, the Laplacian of the electron density at the non-covalent BCPs has been also considered. Compared to the variability observed at the covalent bond critical points, we have detected here very small discrepancies according to the ELMAM2 values: 0.91 e/Å^5 (for UBDB), 1.72 e/Å^5 (ELMO/6-311G), 1.57 e/Å^5 (ELMO/6-311G(d,p)), and 1.17 e/Å^5 (ELMO/6-311+G(2d,2p)). Furthermore, considering the eight weak interactions, we can see in Table 4.4 that the values obtained by transferring ELMOs are very close to the ones obtained transferring pseudoatoms. However, in this case, for all the considered basis sets, the ELMO method slightly overestimates the values of the Laplacian of the electron density at the non-covalent BCPs using the ELMAM2 or the UBDB technique as references.

TABLE 4.4: Values of the Laplacian of the electron density at the non-covalent bond critical points (e/Å^5) after the transfer of pseudoatoms and ELMOs.

	pseudoatoms		ELMOs transfer		
non-covalent interaction	ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
N31-H37 $\cdots$ Ow84	1.96	2.43	3.09	2.96	2.70
Ow81-Hw83 $\cdots$ O5	1.22	1.92	2.87	2.69	2.30
Ow78-Hw80 $\cdots$ Ow81	1.42	2.33	3.14	2.99	2.59
N58-H74 $\cdots$ O25	1.13	1.41	1.87	1.74	1.63
N38-H57 $\cdots$ O6	1.43	1.87	2.50	2.36	2.19
C11 $\cdots$ H35	0.27	0.28	0.32	0.29	0.30
C13 $\cdots$ O84	0.57	0.59	0.59	0.58	0.56
H17 $\cdots$ H50	0.24	0.23	0.30	0.26	0.26
H19 $\cdots$ C42	0.64	0.63	0.76	0.69	0.75
H20 $\cdots$ C46	0.26	0.27	0.27	0.27	0.27
H21 $\cdots$ O81	0.50	0.53	0.55	0.53	0.52
H28 $\cdots$ H65	0.25	0.24	0.32	0.29	0.29
H56 $\cdots$ O32	0.67	0.62	0.71	0.67	0.70

In order to obtain a more global indication of the similarities and of the differences between the values of the topological properties associated with the electron densities, we have computed the agreement index R (see equation (4.5)) between the different methods considered (see Tables 4.5-4.8).

Considering the topological properties at the covalent bond critical points (Tables 4.5 and 4.6), we have observed similar trends both for the electron density $\rho(\mathbf{r}_b)$ and its Laplacian $\nabla^2\rho(\mathbf{r}_b)$. In fact, it is easy to note an analogous overall agreement among two main groups: the one composed by the ELMAM2, UBDB and ELMO/6-311G charge distributions and the one constituted by the ELMO/6-311G(d,p) and ELMO/6-311+(2d,2p).

Afterwards, considering the values of the topological properties at the non-covalent bond critical points in Tables 4.7 and 4.8, lower values of the agreement index have been generally observed between the different electron distributions. Moreover, in these cases, a very good agreement has been obtained between all the ELMO charge densities, with the index R always lower than or equal to 0.5%.

TABLE 4.5: Values of the agreement index R (%) between the different reconstruction methods considering as topological property the electron density at the covalent bond critical points.

	pseudoatoms		ELMO		
	ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
ELMAM2	0.0	1.3	1.6	2.1	2.7
UBDB		0.0	1.6	1.8	2.5
ELMO/6-311G			0.0	3.3	3.9
ELMO/6-311G(d,p)				0.0	0.9
ELMO/6-311+G(2d,2p)					0.0

TABLE 4.6: Values of the agreement index R (%) between the different reconstruction methods considering as topological property the Laplacian of the electron density at the covalent bond critical points.

	pseudoatoms		ELMO		
	ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
ELMAM2	0.0	7.1	7.5	16.8	17.0
UBDB		0.0	5.6	16.4	13.0
ELMO/6-311G			0.0	13.0	14.5
ELMO/6-311G(d,p)				0.0	7.0
ELMO/6-311+G(2d,2p)					0.0

As already explained in the section 4.2.2, we have also decided to also consider three other topological properties to study the detected hydrogen-bond interactions: the kinetic energy density, the potential energy density, and the positive curvature of the electron densities at the bond critical points. In Tables 4.9-4.11 we have compared our results to those obtained using the empirical relations (4.6)-(4.8) established by Espinosa.

In Table 4.9, where the kinetic energy density is considered, we can easily note that all the ELMO values are always greater than the corresponding ELMAM2 and UBDB ones. Now, according to the Abramov equation, the kinetic energy density at the bond critical points of closed-shell interactions can be expressed as follows:

$$G(\mathbf{r}_b) = \frac{3}{10}(3\pi^2)^{\frac{2}{3}}\rho^{\frac{5}{3}}(\mathbf{r}_b) + \frac{1}{6}\nabla^2\rho(\mathbf{r}_b) \tag{4.9}$$

TABLE 4.7: Values of the agreement index R (%) between the different reconstruction methods considering as topological property the electron density at the non-covalent bond critical points.

	pseudoatoms		ELMO		
	ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
ELMAM2	0.0	0.7	1.0	1.2	1.2
UBDB		0.0	0.7	0.8	0.8
ELMO/6-311G			0.0	0.3	0.5
ELMO/6-311G(d,p)				0.0	0.2
ELMO/6-311+G(2d,2p)					0.0

TABLE 4.8: Values of the agreement index R (%) between the different reconstruction methods considering as topological property the Laplacian of the electron density at the non-covalent bond critical points.

	pseudoatoms		ELMO		
	ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
ELMAM2	0.0	1.0	2.1	1.7	1.5
UBDB		0.0	1.3	0.9	0.8
ELMO/6-311G			0.0	0.4	0.5
ELMO/6-311G(d,p)				0.0	0.4
ELMO/6-311+G(2d,2p)					0.0

Considering this relation, we can infer that the trend observed in Table 4.9 is probably related to the fact that the same behaviour has been detected for the values of the Laplacian at the non-covalent BCPs (Table 4.4), and to the fact that the electron density values have not shown a large variability (Table 4.3). Furthermore, we can also note in the Table 4.9 that the ELMO charge density reconstructions generally provide the smallest differences with respect to the corresponding empirical reference values. Finally, it is worth mentioning that the exponential behaviour of the kinetic energy density in function of the hydrogen-bond length discovered by Espinosa (see equation (4.6)) is qualitatively well described by all the considered methods and especially in the case of the ELMOs (Figure 4.3). For the sake of completeness, we have reported all the other graphs that show the kinetic energy density, potential energy density, and positive curvature of the electron density at the hydrogen-bond critical points in Appendix B (Figure B.1-B.5).

TABLE 4.9: Values of the kinetic energy density at the hydrogen-bond critical points (kJ.mol.Å^{-3}) obtained through the Espinosa empirical relation and after the transfer of pseudoatoms and ELMOs.

			pseudoatoms		ELMOs transfer		
hydrogen-bond	$d(H\cdots O)$	Espinosa	ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
N31-H37$\cdots$Ow84	1.920	63.4	55.4	63.9	75.2	70.7	64.2
Ow81-Hw83$\cdots$O5	1.892	68.6	54.4	59.5	76.4	70.3	61.5
Ow78-Hw80$\cdots$Ow81	2.098	39.0	31.5	36.3	42.1	38.7	36.7
N58-H74$\cdots$O25	1.931	61.5	48.0	48.8	67.8	61.2	53.0
N38-H57$\cdots$O6	1.986	53.0	41.6	49.6	57.9	53.7	50.3

Then, considering the potential energy density (see Table 4.10), it is possible to observe that all the considered methods generally overestimate $V(\mathbf{r}_b)$ evaluated according to Espinosa's expression. However, as for the kinetic energy density, we can note that the ELMO transfer method is quite successful and, especially for the 6-311G(d,p) and the 6-311+G(2d,2p) basis sets, it allows one to obtain very good agreements with the corresponding empirical references.

TABLE 4.10: Values of the potential energy density at the hydrogen-bond critical points (kJ.mol.Å^{-3}) obtained through the Espinosa empirical relation and after the transfer of pseudoatoms and ELMOs.

		pseudoatoms	ELMOs transfer				
hydrogen-bond	$d(H\cdots O)$	Espinosa	ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
N31-H37$\cdots$Ow84	1.920	-49.8	-57.4	-61.6	-66.1	-60.8	-55.0
Ow81-Hw83$\cdots$O5	1.892	-55.2	-70.0	-55.7	-67.3	-59.2	-52.5
Ow78-Hw80$\cdots$Ow81	2.098	-26.3	-32.2	-34.2	-33.4	-29.9	-29.0
N58-H74$\cdots$O25	1.931	-47.8	-62.9	-45.3	-57.5	-49.9	-43.4
N38-H57$\cdots$O6	1.986	-39.3	-44.1	-48.3	-47.7	-43.1	-40.9

Finally, in Table 4.11, where the positive curvatures of the charge distributions at the hydrogen-bond critical points are presented, it is possible to observe that all the methods provide values that are very close to the Espinosa ones. Nevertheless, unlike the two previous considered topological properties, in this case, the ELMAM2 and the UBDB reconstruction techniques give a better agreement compared to the ELMO transfer method. Furthermore, it is also worth noting that the ELMO/6-311G electron density always provide positive curvatures that are greater than the corresponding ELMO/6-311G(d,p) and

ELMO/6-311+G(2d,2p) ones.

TABLE 4.11: Values of the positive curvature λ_3 of the electron density at the hydrogen-bond critical points (e/Å^5) obtained through the Espinosa empirical relation and after the transfer of pseudoatoms and ELMOs.

	pseudoatoms		ELMOs transfer				
hydrogen-bond	$d(H\cdots O)$	Espinosa	ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
N31-H37$\cdots$Ow84	1.920	4.1	4.0	4.2	4.9	4.6	4.2
Ow81-Hw83$\cdots$O5	1.892	4.4	4.3	4.1	5.2	4.7	4.0
Ow78-Hw80$\cdots$Ow81	2.098	2.7	2.4	2.6	2.8	2.5	2.4
N58-H74$\cdots$O25	1.931	4.0	4.0	3.5	4.6	4.1	3.5
N38-H57$\cdots$O6	1.986	3.5	3.1	3.4	3.8	3.5	3.3

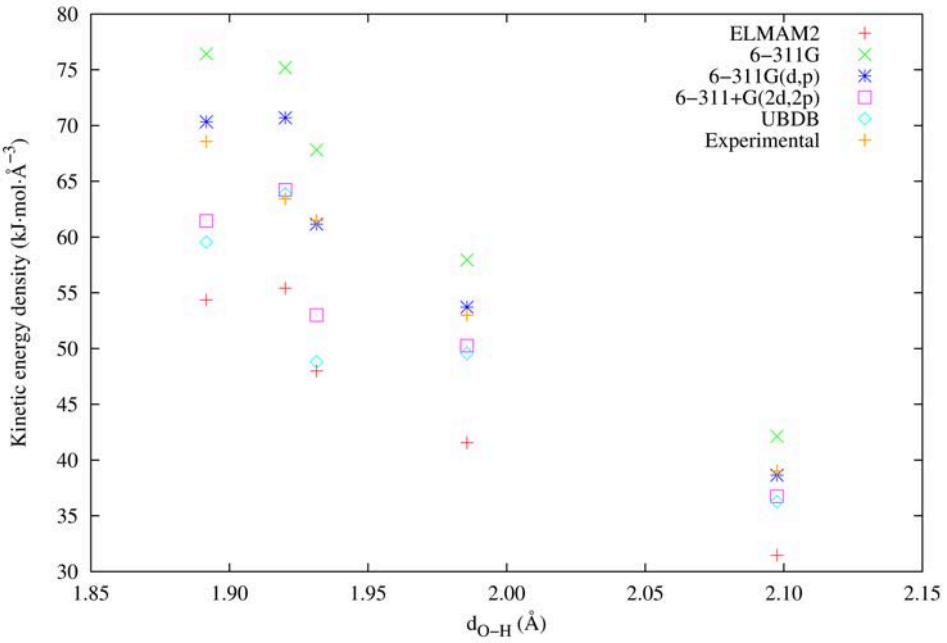

FIGURE 4.3: Kinetic energy density at the hydrogen-bond critical points (kJ.mol.Å^{-3}) in function of the experimental distance $d(H\cdots O)$ in Å. The values are obtained through the Espinosa empirical relation and after the transfer of pseudoatoms (ELMAM2 and UBDB databases) and ELMOs (6-311, 6-311G(d,p) and 6-311+G(2d,2p) basis sets).

4.3.2 Net atomic charges

To further evaluate the ELMOs and the pseudoatoms transferability, the net integrated charges determined through the integration of the different electron distributions over the QTAIM atomic basins have been also considered. In particular, we have compared the global charges for the main functional groups of the polypeptide backbone and for the five side chains. We have also taken into account the water molecules charges and the global charge of the Leu-enkephalin pentapeptide after the transfers.

In Table 4.12, we have noticed that the global electroneutrality of the polypeptide is conserved for each considered method. These results are particularly in favour of the ELMO technique since no electroneutrality constraints are imposed after the transfer of the molecular orbitals, whereas, in the case of pseudoatoms transfer, as already explained in subsection 4.2.1, the global molecular neutrality is preserved by means of a charge re-distribution over all the atoms of the system.

Table 4.12 also shows that the resulted charges always correspond to the expected electrostatic nature of the corresponding fragment for all the considered reconstruction techniques. In particular, all the hydrophobic side chains of the polypeptide display positive charges, while all the peptide bonds are negatively charged. Additionally, among all the methods, the N-terminal ammonium and the C-terminal carboxylate groups show a positive and a negative global charge, respectively.

Nevertheless, from a more detailed inspection, it seems that the ELMOs transfer approach provides a systematic overestimation of the charges. Actually, for almost all the fragments taken into account, the global charges obtained by transferring the ELMOs are more positive or more negative compared to the corresponding ELMAM2 and the UBDB values, with the two largest discrepancies observed for the two glycine $C_\alpha H_2$ moieties. These differences are probably mainly due to the strong localization of the molecular orbitals and, secondly, to the lack of charge relaxation after the transfer of the ELMOs. If the former drawback is unavoidable to guarantee the orbitals transferability, the latter could be partially overcome in the future by developing and applying an automatic procedure that allows a sort of charge re-distribution after the ELMOs transfer taking into account the environment of each orbital.

TABLE 4.12: Net QTAIM integrated charges (e) for the backbone "functional groups" the five side chains (including the $C_\alpha H_\alpha$ bonds) of the Leu-enkephalin polypeptide.

	pseudoatoms		ELMOs transfer		
fragment	ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
ammonium (N-term)	0.244	0.401	0.367	0.361	0.311
H-N-C=O (Tyr-Gly)	-0.483	-0.418	-0.489	-0.604	-0.515
H-N-C=O (Gly-Gly)	-0.448	-0.430	-0.531	-0.660	-0.557
H-N-C=O (Gly-Phe)	-0.434	-0.426	-0.594	-0.726	-0.622
H-N-C=O (Phe-Leu)	-0.435	-0.422	-0.550	-0.659	-0.570
carboxylate (C-term)	-0.695	-0.673	-0.992	-0.964	-0.972
$C_\alpha H_\alpha - R$ (Tyr)	0.321	0.395	0.697	0.687	0.739
$C_\alpha H_2$ (Gly 1)	0.391	0.257	0.526	0.655	0.547
$C_\alpha H_2$ (Gly 2)	0.386	0.250	0.588	0.725	0.622
$C_\alpha H_\alpha - R$ (Phe)	0.501	0.563	0.572	0.690	0.583
$C_\alpha H_\alpha - R$ (Leu)	0.651	0.530	0.438	0.535	0.452
water 1	0.031	0.010	-0.008	-0.005	-0.010
water 2	-0.002	-0.005	-0.002	-0.002	0.001
water 3	-0.025	-0.026	-0.015	-0.018	-0.005
global	0.004	0.006	0.007	0.015	0.005

Finally, in order to study the similarities between the obtained topological integrated charges, an evaluation of the agreement index R (see Equation (4.5)) for each couple of charge density reconstruction methods has been performed. To achieve this goal, the net integrated charges presented in Table 4.12 have been taken into account. Our results (see Table 4.13) show that the agreement is better when the comparison is conducted within either the pseudoatoms or the ELMO techniques, while it is poorer when results associated with different transfer methods are compared. This fact is consistent with trends already observed in Table 4.12, in which we see that the ELMO charges are generally larger compared to the ones associated with the pseudoatoms electron densities.

TABLE 4.13: Values of the agreement index R (%) between the different reconstruction methods considering as topological property the net QTAIM integrated charges (e) for the backbone "functional groups" and for the five side chains of the Leu-enkephalin polypeptide.

	pseudoatoms		ELMO		
	ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
ELMAM2	0.0	8.3	15.7	20.5	16.5
UBDB		0.0	16.5	21.5	18.3
ELMO/6-311G			0.0	7.4	2.3
ELMO/6-311G(d,p)				0.0	6.5
ELMO/6-311+G(2d,2p)					0.0

4.3.3 Similarity Indexes

To get a more global comparison between the different charge distributions, we have also computed two real-space similarity indexes. In particular, we have considered the Walker-Mezey similarity indicator (*46*) (see Equation 3.2.5) and the real-space R (RSR) value (*123*) (see Equation 3.2.5). At first, in Table 4.14, we can observe that, as expected, for the RSR and the global L(0.001,1000) indexes, the electron densities associated with the ELMO method are less similar than the UBDB one with respect to the reference ELMAM2 charge distribution. Furthermore, as already pointed out in section 3.2.5, namely, changing the values a and a' in the Walker-Mezey index, the comparison of the electron distributions in different regions has become possible (See Figure 4.4). We have especially noticed that in core (L(3,1000)) and in covalent (L(1,3)) regions the similarity between the ELMO and the ELMAM2 techniques remain quite significant, whereas in domains far from the nuclei (L(0.001,0.01)), the dissimilarities are more important. Nevertheless, it is worth reminding that the Walker-Mezey indicator measures the similarity in a relative sense and, since in domains far from the nuclei, the electron density values are very small, quite large relative differences between the charge distributions can be certainly observed even if these discrepancies are nearly negligible in magnitude.

Finally, for the sake of completeness, we have also determined the values of the global similarity index RSR between all the possible couples of reconstructed electron densities. From values in Table 4.15, we confirm the main trend already observed for the agreement

TABLE 4.14: Point-by-point comparison of the UBDB, ELMO/6-311G, ELMO/6-311G(d,p), and ELMO/6-311+G(2d,2p) electron densities with the ELMAM2 charge density: values of the real-space R and of the Walker-Mezey similarity L(a, a') indexes (%). [a]

		ELMOs transfer		
similarity index	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
RSR	1.3	2.3	2.0	2.0
L(0.001,1000)	91.4	83.1	83.2	82.8
L(0.001,0.01)	87.0	71.1	71.3	71.3
L(0.01,1)	93.4	88.2	88.2	87.9
L(1,3)	97.7	96.2	96.6	96.4
L(3,1000)	97.9	96.2	97.2	97.4

[a] For the Walker-Mezey indicator, the electron densities are compared within the a and a' limits expressed in $e\text{Å}^3$.

index R of the net integrated charges, namely, a larger similarity between charge distributions resulting from similar transfer methods and a reduced likeness when comparing charge densities obtained by transferring pseudoatoms to the electron distributions obtained by transferring ELMOs. In particular, we note that, compared to ELMAM2, the UBDB library provides a charge distribution that is closer to the ELMO ones. This is probably due to the fact that the structure factors used to construct the multipole-model-based UBDB databank originate from quantum-mechanical calculations as it was already explained in section 4.1.3. In addition, an expected significant global similarity between the ELMO/6-311G(d,p) and the ELMO/6-311+G(2d,2p) electron densities is also observed.

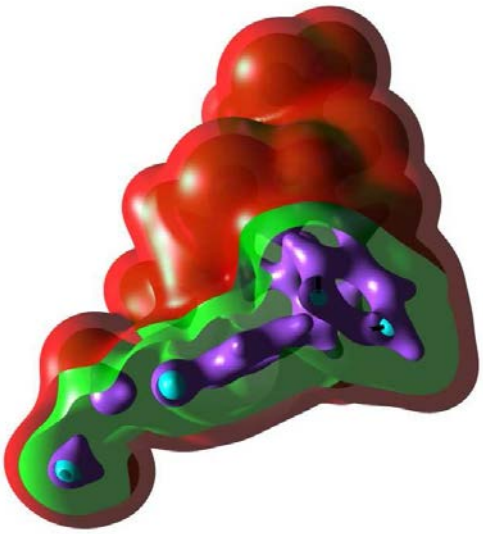

FIGURE 4.4: Sectional view of the ELMAM2 electron density isosurfaces of the Leu-enkephalin pentapeptide and interacting water molecules, corresponding to the shell limits selected for the Walter-Mezey similarity index. The 3, 1, 0.01, 0.001 $e/Å^{-3}$ isosurfaces are depicted in light blue, violet, green and red, respectively.

TABLE 4.15: Point-by-point comparison of the different electron densities: values of the real-space R similarity index (%).

	pseudoatoms		ELMO		
	ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
ELMAM2	0.0	1.3	2.3	2.0	2.0
UBDB		0.0	1.9	1.7	1.7
ELMO/6-311G			0.0	1.3	1.5
ELMO/6-311G(d,p)				0.0	0.6
ELMO/6-311+G(2d,2p)					0.0

4.4 Conclusions

In the previous Chapter, we have seen that the ELMOs can be easily transferred from one molecule to another since their are strictly localized on small molecular fragments. In this Chapter, to further assess the transferability of the ELMOs, we have compared the electron densities obtained by transferring pseudoatoms (from the experimental ELMAM2 and the theoretical UBDB databanks) and extremely localized molecular orbitals to the crystallographic structure of the Leu-enkephalin pentapeptide.

Our results show that, despite the presence of unavoidable and expected differences, all the reconstruction methods provide quite reasonable charge distributions. In particular, we have observed that the ELMOs transfer is quite successful in reproducing the topological properties at the non-covalent bond critical points. However, due to their strictly localized nature, the use of extremely localized molecular orbitals generally leads to an overestimation of the integrated charges associated with the fragments of the polypeptide. To overcome this drawback, it would be interesting to develop a proper method to relax the transferred ELMOs taking into account the environment of the fragment on which they are localized in the target system. This task will be achieved in forthcoming works.

To conclude, the results presented in this Chapter have shown that the transferability of the ELMOs is as reliable as the one of the pseudoatoms and that it enables one to obtain completely acceptable reconstructions of electron densities. Furthermore, it is worthwhile to point out that, by exploiting the ELMO-based approach, we do not reconstruct only charge distributions of very large systems but also their wave functions. These wave functions potentially give us access to properties that, up to now, cannot be obtained using only the electron density. In fact, the exact functional relation between ground state electron densities and wave functions of many-electron systems is still unknown. This is the reason why our result further encourages the construction of new libraries of extremely localized molecular orbitals with the final purpose of refining crystallographic structures and computing approximate properties of macromolecules.

Conclusions of Part I, future directions

In the first part of this manuscript we have presented the concept of extremely localized molecular orbitals and we have mainly assessed their transferability. The first chapter was an introduction to some linear scaling quantum chemical methods used to determine the electronic structure of macromolecules at a reduced computational cost. We have especially described the Divide & Conquer strategy, the Fragment Molecular Orbital method and the Additive Fuzzy Density Fragmentation approaches. In the context of the linear scaling philosophy, another possibility consists in defining molecular orbitals strictly localized on small molecular subunits and, therefore, easily transferable to other molecules containing the same fragment. However, the molecular orbitals traditionally obtained in quantum chemistry are completely delocalized over the systems in exam and, hence, they cannot be easily transferred from molecule to molecule.

This drawback has been discussed in detail in Chapter 2 where we have also given a brief overview of some traditional Localized Molecular Orbitals (LMOs) methods, which mainly consist in unitary transformations of the canonical Hartree-Fock orbitals and which aim at preserving the traditional Lewis-picture of molecules. Nevertheless, even if these *a posteriori* defined localized molecular orbitals are localized on small fragments, they conserve small orthogonalization tails beyond their localization region. For this reason, these orbitals can not be transferred from a molecule to another because the deletion of these tails entails dramatic increases in the energy of the systems. To avoid this problem, it is necessary to use *a priori* localization techniques, which subdivide the system into fragments and determine the global wave function of a molecule from local contributions from each subunit. Therefore, the main sections of Chapter 2 have been dedicated to the Stoll approach, which has been the technique exploited in our research to obtain Extremely Localized Molecular Orbitals, and to the Philipp and Friesner rotation technique, which, in this thesis, has been used to transfer all our strictly localized molecular orbitals.

Then, since the ELMOs are in principle transferable from molecule to molecule, we

have considered the possibility of constructing libraries of ELMOs that cover all the possible molecular fragments of the twenty natural amino acids. Our final goal is to reconstruct almost instantaneously approximate wave functions (and electron densities) of proteins. To this purpose, in Chapter 3, we have presented preliminary studies to investigate the ELMOs transferability and to define a suitable model molecule approximation for the computation of the ELMOs to be stored in the future databases. We have tested three different levels of approximations and we have shown that the Nearest Functional Group Approximation (NFGA) is globally and by far the best one both from the energetic point of view and in terms of electron density reconstruction. Moreover, the charge densities obtained through our ELMO-NFGA approach are very similar to the corresponding Hartree-Fock ones. In particular, we have observed that the extent of the discrepancies detected between ELMO-NFGA and HF charge densities are quite comparable to the one observed between B3LYP and Hartree-Fock electron distributions.

Nevertheless, the NFGA model molecules are sometimes characterized by undesired non-covalent intramolecular interactions (which are not present in the target system) that artificially perturb the ELMOs to be transferred. From these observations, it seems obvious that the NFGA can be only regarded as a zero-order approximation and, therefore, we envisage the possibility of relaxing the ELMOs immediately after the transfer by means of proper linear-scaling techniques that suitably take into account the real chemical environment in the target system.

At a later stage, in Chapter 4, to further assess the reliability of our approach, we have compared the transferability of the ELMOs to the one of the pseudoatoms, which are aspherical atomic density functions generally exploited in crystallography studies. We have demonstrated that, despite the presence of expected discrepancies, all the reconstruction methods provide quite reasonable charge distributions. We have especially noticed that the ELMOs transfer is quite successful in reproducing the topological properties at the non-covalent bond critical points. Nevertheless, due to the strictly localized nature of the transferred orbitals, the use of the ELMOs leads to an overestimation of the integrated charges associated with the fragments of the polypeptide. To solve this problem, as just mentioned above, we should develop a proper method to relax the transferred ELMOs taking into account the environment of the fragments on which they are localized in the target system.

The results presented in Chapter 3 and 4 encourage the construction of new libraries

of extremely localized molecular orbitals with the final purpose of computing approximate properties of macromolecules and of refining crystallographic structures of very large systems. To achieve this goal, we are currently accomplishing the construction of ELMOs databanks that cover all the possible molecular fragments of the twenty natural amino acids. This work has been initiated during the redaction of this manuscript and it is currently in progress.

At the moment, to accomplish this task, we have considered all the twenty natural amino acids in their different protonation states. This means that all the following particular cases have been taken into account:

- deprotonated aspartate;
- deprotonated histidine;
- protonated cysteine;
- protonated lysine (+1 total charge);
- histidine protonated at N-epsilon;
- histidine protonated at N-delta;
- histidine protonated at both N-epsilon and N-delta;
- protonated aspartic acid;
- protonated glutamic acid;
- deprotonated cysteine;
- deprotonated (neutral) lysine;

Furthermore, for the sake of completeness, we have also considered all the amino acids in N- and C- terminal positions. Overall, 78 different amino acids "configurations" have been taken into account for the construction of the ELMOs libraries. Afterwards, exploiting the Nearest Functional Group Approximation (see section 3.2.3) for each considered "configuration" we have constructed all the necessary model molecules (115 model molecules) for the calculation of the ELMOs to be stored in the databases. However, considering the results presented in section 3.3.1 about the performances of the NFGA, for the

peptide bonds the model molecules have been constructed exploiting only the NAA (see Figure 4.5A). The only exception occurs when the proline residue is involved in a peptide bond. In that case, the whole pirrolidine side chain is taken into account (see Figure 4.5B). For all the constructed model molecules, a geometry optimization at the B3LYP/6-311G(d,p) level has been carried out. Afterwards, using the optimized geometries, the sets of ELMOs for the libraries have been determined performing variational-ELMO calculations with different basis-set (6-31G, 6-31G(d,p), 6-311G and 6-311G(d,p)).

The next step will consist in developing a code to automatize the ELMO transfer from the databanks in construction to any target system in order to recover almost instantaneously the wave functions and the electron densities of proteins and large polypeptides. Moreover, we would like to extend this library to other families of compounds (e.g., the nucleobases to describe the DNA). Finally, as already discussed above, it will be crucial to devise new techniques to relax the ELMOs immediately after the transfer by means of proper linear-scaling strategies.

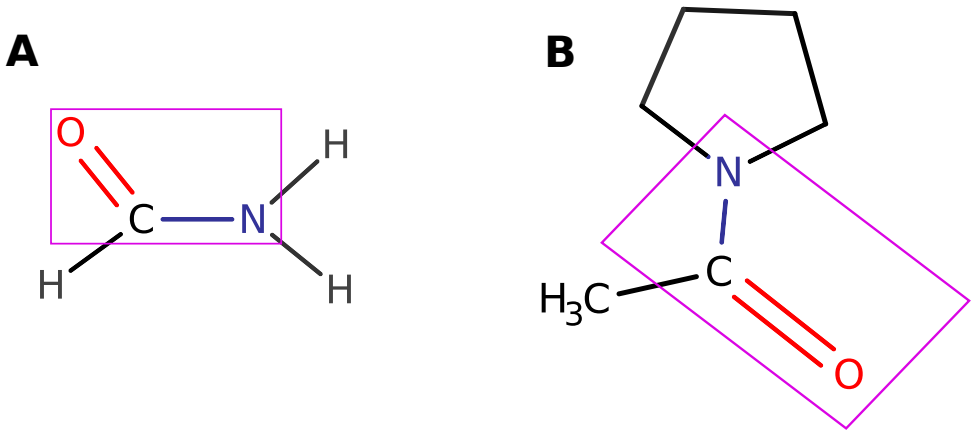

FIGURE 4.5: Model molecule for peptide bonds (A) in the general case and (B) in the presence of proline.

Conclusions de la Partie I et perspectives futures

Dans la première partie de ce manuscrit, nous avons présenté le concept d'orbitales moléculaires extrêmement localisées et nous avons essentiellement évalué leur transférabilité. Le premier chapitre a été consacré à présenter, de manière succincte quelques méthodes de chimie quantique à croissance linéaire utilisées pour déterminer la structure électronique d'une macromolécule en un temps de calcul réduit. Nous avons décrit plus particulièrement la stratégie "Divide & Conquer", la méthode des orbitales moléculaires de fragments (FMO) ou encore l'approche d'additivité des densités fragmentées. Dans le contexte de la philosophie à croissance linéaire, une autre possibilité consiste à définir des orbitales moléculaires strictement localisées sur de petits fragments de molécules afin de pourvoir les transférer à d'autres molécules contenant le même fragment. Cependant, les orbitales moléculaires habituellement utilisées en chimie théorique sont totalement délocalisées sur le système étudié et, par conséquent, ne peuvent pas être transférées d'une molécule à une autre.

Cet obstacle a été discuté en détail dans le Chapitre 2, où nous avons également présenté un bref récapitulatif de quelques méthodes habituellement utilisées pour localiser les orbitales moléculaires (LMOs). Ces méthodes reposent sur des transformations unitaires des orbitales canoniques d'Hartree-Fock et leur but est de préserver une vision chimique présente dans la structure de Lewis des molécules. Néanmoins, même si ces méthodes de localisation dîtes a posteriori permettent la localisation des orbitales, celle-ci conservent de petites "queues" en dehors de leur région de localisation et de ce fait ne peuvent pas être transférées d'une molécule à une autre. En effet, la suppression de ces "queues" engendre une grande augmentation de l'énergie des systèmes étudiés. Pour éviter ce problème, il est nécessaire d'utiliser des techniques de localisation dites *a priori*, qui divisent le système en fragments et déterminent la fonction d'onde globale de la molécule à partir de contributions locales de chacune des sous-unités. C'est pourquoi, la principale partie du Chapitre 2 a été consacrée à l'approche de Stoll, qui a été la technique utilisée dans nos travaux de recherche pour obtenir les ELMOs. De plus, nous avons détaillé la technique de rotation de Philipp et Friesner que nous avons utilisé dans cette thèse pour transférer les orbitales strictement localisées.

Ensuite, puisque les ELMOs sont en principe transférables d'une molécule à une autre, nous avons envisagé la construction de bases de données d'ELMOs qui couvriraient tous

les fragments présents dans les vingts acides aminés naturels, notre but final étant de reconstruire quasiment instantanément des fonctions d'ondes approximatives (et des densités électroniques) de protéines. Pour ce faire, dans le Chapitre 3, nous avons présenté des études préliminaires pour évaluer la transférabilité des ELMOs et tenter de définir une approximation appropriée pour la construction des molécules modèles qui vont nous permettre de générer les ELMOs qui seront stockées dans les futures bases de données. Nous avons testé trois différents niveaux d'approximations et nous avons démontré que l'approximation du groupe fonctionnel le plus proche (NFGA) donne les meilleurs résultats tant d'un point de vu énergétique qu'en termes de reconstruction de la densité. En outre, nous avons également observé que les densités de charges obtenues en utilisant notre approche ELMO-NFGA sont très similaires à celle obtenues par une méthode Hartree-Fock. Plus particulièrement, nous avons montré que les différences entre les densités ELMO-NFGA et HF sont comparables à celles observées entre des distributions de charges HF et DFT.

Cependant, les molécules modèles utilisées dans l'approximation NFGA sont parfois caractérisées par des interactions intramoléculaires indésirables car non présentes dans la molécule cible. Au vu de ces résultats, il nous semble évident que l'approximation du groupe fonctionnel le plus proche ne peut être considérée que comme une approximation d'ordre zéro. C'est pour cela que nous envisageons la possibilité de relaxer les ELMOs après leur transfert afin de tenir compte de l'environnement chimique dans la molécule cible.

Finalement, dans le Chapitre 4, afin de poursuivre l'évaluation et la fiabilité de notre approche, nous avons comparé la transférabilité des ELMOs avec celle des pseudo-atomes, qui sont des densités atomiques asphériques généralement utilisées dans le domaine de la cristallographie. Malgré la présence de différences inévitables et prévisibles, nos résultats ont montré que toutes les méthodes de reconstruction fournissent des densités électroniques cohérentes. Nous avons particulièrement observé que le transfert des ELMOs réussit relativement bien à reproduire les propriétés topologiques de la densité aux points critiques de liaisons non-covalentes. Néanmoins, à cause de leur nature très localisée, l'utilisation des ELMOs a généralement entraîné une surestimation des charges associées aux différents fragments du polypeptide. Pour régler ce problème, comme mentionné précédemment, nous allons devoir développer ultérieurement une approche qui permettra de relaxer les ELMOs après leur transfert sur la molécule cible. Les résultats présentés dans les Chapitres 3 et 4 nous encouragent à poursuive la construction de nouvelles bases de données d'orbitales moléculaires extrêmement localisées avec pour objectif final

d'affiner des structures cristallographiques de grands systèmes. Pour atteindre ce but, nous sommes en train de construire une base de données d'ELMOs qui couvrira tous les fragments possibles présents dans les vingt acides aminés. Ce travail a été initié durant la rédaction de ce manuscrit et reste en cours de développement.

Pour le moment, nous avons considéré les vingt acides aminés naturels et leurs différents états de protonation. Cela signifie que nous avons tenu compte des cas particuliers suivant :

- aspartate déprotoné;
- histidine déprotonée;
- cystéine protonée;
- lysine protonée (charge totale +1);
- histidine protonée en N-epsilon;
- histidine protonée en N-delta;
- histidine protonée en N-epsilon et en N-delta;
- acide aspartique protoné;
- acide glutamique protoné;
- cystéine déprotonée;
- lysine déprotonée (neutre);

De plus, pour être complet, nous avons également considéré tous les acides aminés en position N- et C- terminale. Au total, nous avons tenu compte de 78 "configurations" différentes d'acides aminés pour la construction de la base de données d'ELMOs. Ensuite, en utilisant l'approximation du groupe fonctionnel le plus proche pour chacune des "configurations", nous avons construit toutes les molécules modèles nécessaires au calcul des ELMOs (115 molécules modèles) qui seront stockées dans la base de données. Cependant, au vu des résultats présentés dans le Chapitre 3 sur les performances au niveau NFGA, nous avons préféré utiliser l'approximation NAA (voir Figure 4.6A) pour la molécule modèle permettant de décrire les liaisons peptidiques. La seule exception

apparaît lorsqu'un résidu proline est impliqué dans une liaison peptidique. Dans ce cas précis, toute la chaîne latérale pirrolidine est prise en compte (voir Figure 4.6B). Toutes les molécules modèles ont été optimisées au niveau B3LYP/6-311G(d,p). Ensuite, en utilisant les géométries optimisées, nous avons déterminé les ELMOs pour les bases de données en effectuant des calculs ELMO-variationnels avec différentes bases (6-31G, 6-31G(d,p), 6-311G et 6-311G(d,p)).

La prochaine étape consistera à développer un code qui automatisera le transfert des ELMOs depuis la librairie vers une molécule cible pour reconstruire quasiment instantanément sa fonction d'onde et sa densité électronique. De plus, nous aimerions étendre ces bases de données à d'autres familles de composés (par exemple les bases nucléiques pour décrire l'ADN). Enfin, il sera également crucial, comme mentionné précédemment, d'imaginer une stratégie qui permettrait la relaxation des ELMOs après leur transfert.

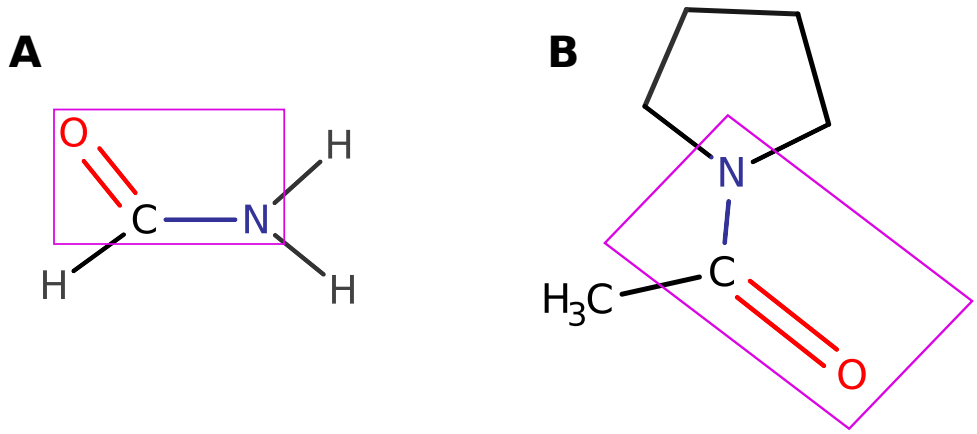

FIGURE 4.6: Molécule modèle utilisée pour décrire la liaison peptidique (A) dans un cas général et (B) en présence d'un résidu proline.

Part II

Assessing the capabilities of the X-ray constrained wave function methods

Chapter 5

Experimental Wave Function Methods

Résumé

Dans ce chapitre, après un bref résumé historique sur les méthodes de fonctions d'onde contraintes pour reproduire certaines données expérimentales, nous présentons la technique basée sur l'utilisation des données de diffraction aux rayons-X et développée par Jayatilaka. Après avoir discuté les bases théoriques de cette approche, nous exposerons les principales équations de la méthode Hartree-Fock contrainte (XC-RHF). Ensuite, nous reviendrons sur la théorie récemment développée des ELMOs contraintes expérimentalement (XC-ELMO). Finalement, nous présenterons quelques éléments qui devraient permettre le développement d'une approche Valence Bond "expérimentale" basée sur l'utilisation des ELMOs contraintes (XC-ELMO-VB).

5.1 Introduction

In quantum chemistry, *ab initio* wave function-based methods and Density Functional Theory (DFT) (*6*) are two essential approaches. In the former, one tries to find the optimal wave function that minimizes the energy of the system under investigation, usually by means of first-principles techniques that try to approach the exact solution of the Schrödinger equation. In contrast, DFT relies on the Hohenberg and Kohn theorem (*7*), according to which all ground-state properties of the physical system can be obtained from the knowledge of the electron density.

Nevertheless, although not commonly used within the quantum chemistry community, there are several alternatives to the *ab initio* and DFT techniques, as for instance the approaches based on experimentally constrained wave functions or density matrices. Theses techniques can be regarded as semi-empirical since the wave function or the density matrix of the systems is properly fitted to experimental data through well-defined procedures. In principle, there are no strict limitations in the choice of the experimental measurements to reach this goal. However, X-ray diffraction data are particularly suitable and widely because of, their abundance and, more importantly, their direct connection to the electron density. This fact explains why the experimentally constrained wave function strategies are usually known as X-ray constrained wave function approaches.

5.1.1 Motivations for deriving experimental constrained wave functions

Although still not very well known, these techniques have represented an active research field for about fifty years. The main challenge is the necessity to condense all the experimental observations into a single object of unquestionable physical meaning. In this respect, the wave function represents the best theoretical option since, according to quantum mechanics, it intrinsically contains all the information related to the physical system. Therefore, an experimentally constrained wave function determined from X-ray diffraction data would be preferable to other arbitrary chosen model objects, such as, multipole expansions (*129*).

It is also worth mentioning that the "experimental" wave functions can be extremely useful when traditional accurate quantum chemistry calculations cannot be performed.

In those cases wave functions directly obtained from experimental data could be employed to calculate accurate observables and can be considered as a very powerful technique.

Finally, another motivation is the possibility of getting novel insights into the Hohenberg and Kohn mapping between the electron density and the ground state wave function. Indeed, this goal is of paramount importance in theoretical chemistry since it would enable to perform quantum mechanical calculations using the electron density as the basic variable, which, despite many efforts and developments in the filed of DFT, cannot still be done with the current technologies.

5.1.2 Experimentally constrained wave function methods

The first attempt of extracting a wave function from experimental data has been performed by Mukherji and Karplus (*148*) in 1963. In fact, they have proposed a perturbation approach to constrain a single Slater determinant wave function using experimental values for the dipole moment and the electric field gradient of hydrofluoric acid. Other similar strategies have been subsequently exploited to perform calculations on the lithium hydride system(*149*, *150*).

However, a major contribution in this domain was the pioneering investigation conducted by Clinton and co-workers since 1969 (*151*–*157*). In their studies, original techniques to extract idempotent one-particle density matrices from theoretically generated charge density data were proposed. Then, it was noted that the single determinant wave function could also be extracted from X-ray diffraction data (*156*). In this original work, the experimental errors were taken into account using appropriate weighting factors. These methods have been continuously improved over the years. For example, steepest-descent algorithms have been introduced in order to solve the problem of fulfilling the idempotency conditions and the desired agreement with experiment data (*158*). Another extension was proposed by Frishberg and Massa to treat the case of open-shell single determinant wave functions (*159*). They obtained interesting results exploiting high-quality theoretically determined X-ray structure factors to fit the wave functions for simple atomic and molecular systems. They concluded that by fitting an accurate set of X-ray structure factors, very accurate one-electron properties can be obtained, which were

further improved using variational calculations. A similar result could not be obtained for two-electron properties. Some years later, in 1985, another significant result was obtained by Massa *et al.*,who extracted the idempotent density matrix of beryllium using single-crystal x-ray diffraction data (*160*).

More recently a variant of Clinton's approach has been proposed by Howard and coworkers (*161*), who have used simulated annealing (temperature dependent probability) methods for the determination of single determinant wave functions. They have reported applications of this procedure to larger systems through one theoretical example (methylamine) and one experimental example (formamide). In the same context, it is worth mentioning the investigations carried out by Snyder and Stevens (*162*) who have determined the experimental idempotent density matrix for the azide ion in crystalline potassium azide. Finally, let us mention another important technique, the X-ray atomic orbital method initially proposed by Tanaka (*163*, *164*) that mainly aimed at modeling the distortion of the heavy-atom atomic orbitals due to the crystal-field effects.

However, the use of a single determinant wave function to fit experimental data, as assumed in the Clinton-like techniques described above, is obviously an important limitation. To tackle this problem, Hibbs, Waller and co-workers (*165*, *166*) have recently proposed an interesting orbital-based method that describes the density in a fixed molecular orbital basis with variable orbital occupation numbers. The two main advantages of this MOON (Molecular Orbitals Occupation Numbers) approach are, firstly, that the molecular properties are easy to be interpreted and, secondly, that this technique is potentially linear scaling. Nevertheless, for the investigations of large systems, the pre-computation and the selection of the molecular orbitals could be a serious drawback.

In some approaches, the density operator is required to be derived from any antisymmetric wave function, namely, to be N-representable (instead of using the simple idempotency condition). As well known, the conditions for N-representability of the one-particle density matrix simply correspond to requiring that the density matrix eigenvalues must lie between zero and one. Among the N-representable techniques, it is worth mentioning the prominent role occupied by the methods devised by Schmider *et al.* (*167*) and by Gillet and co-workers (*168*–*170*). Actually, they have reported the first joint refinement of a local wave function model from diffraction and Compton scattering data and, moreover, they have demonstrated that Compton scattering provides a wealth of additional information

to the study of the chemical bond. Furthermore, Cassam-Chenaï has also developed a similar approach to obtain meaningful N-representable density matrices (*171*). In that case, he proposed to extract the one-particle density matrices in terms of a few number of precomputed *ab initio* wave functions and to afterwards optimize it through a fitting procedure against experimental data. In particular, this interesting idea has been applied in connection with polarized neutron diffraction data to analyse the magnetization density for the $CoCl_4^{-2}$ molecular ion in the Cs_3CoCl_5 crystal (*172*).

However, the most important turning point for this family of techniques is represented by the work of Gilbert that, in 1975, has shown that an infinite number of single Slater determinants are compatible with a desired electron density. Therefore, simply fitting the density data with N-repesentability constraints is a necessary but not sufficient criterion to guarantee the determination of physically meaningful experimentally constrained wave functions or density matrices. In this domain, the very important work of Henderson & Zimmerman should be highlighted (*173*). In fact, to overcome the previous drawback, these authors have suggested that, of all the single Slater determinants that are compatible with a given electron density, the optimal one is that which minimizes the Hartree-Fock energy. Then, modifying the initial Clinton method, they have applied their technique to the lithium hydride system using theoretical structure factors. Later, Levy and Goldstein (*174*) and also Gritsenko and Zhidomirov (*175*) have suggested that, in the case where a given wave function is not univocally determined from the data, the selected single Slater determinant is the one that minimizes the kinetic energy. This proposition was further developed by Zhao *et al.* (*176*–*178*), which have devised original strategies to extract the Kohn-Sham orbitals using electron density data obtained by means of quantum Monte-Carlo calculations or theoretical wave functions. In particular, they have observed that the Kohn-Sham orbitals for the beryllium atom, which have been derived from theoretical charge distributions, are almost identical to the Hartree-Fock ones.

Nevertheless, according to the Levy's constrained search formalism (*179*), which has shown that the exact wave function is the one that minimizes the total energy reproducing a given electron distribution, Jayatilaka proposed a more practical implementation of the Henderson & Zimmermann idea (*180*–*186*). This novel approach, which is called X-ray constrained wave function fitting, is based on the determination of a single Slater determinant that minimizes the corresponding energy and that also reproduces a given set of

structure factor amplitudes with a desired accuracy. In fact, using an external multiplier, the strategy variationally minimizes a new functional given by the energy associated with the single Slater determinant and an additional term that takes into account the statistical agreement with the experimental diffraction measurements. It is important to note that the external multiplier is manually adjusted until when the desired agreement between experiment and calculated data is reached.

5.1.3 The Jayatilaka approach and beyond

The Jayatilaka approach can be regarded as the most promising strategy among all the experimentally constrained wave function methods proposed so far. Indeed, this has been successfully used in the investigations of various chemical and physical properties, such as crystal molecular dipole moments, crystal polarizabilities and refractive indices (*187*, *188*). Moreover, Grabowsky *et al.* have successfully applied the X-ray constrained wave function approach and the Electron Localizability Indicators (ELIs) (*189*, *190*) on a series of epoxide derivatives, on α,β-unsaturated carbonyl and on hydrazone compounds. The Jayatilaka technique has also been used in the framework of the Electron Localization Functions (ELFs)(*191*) to study the molecular crystals of ammonia, urea and alloxan. Finally, the Roby bond indices derived from an X-ray wave function refinement have been exploited to investigate the ionic bonding in sulfur dioxide systems (*192*).

Recently, Genoni has proposed a novel X-ray constrained wave function fitting technique (*193*, *194*) using an *a priori* definition of extremely localized molecular orbitals (see Chapter 2). This new approach can be considered as a combination of the Jayatilaka approach and the Stoll technique, which has been described in detail in section 2.3. Some preliminary investigations have already shown that the extraction of X-ray constrained ELMOs is straightforward and that this technique can be considered as a new and potentially useful tool for the reconstruction and the analysis of electron densities from experimental diffraction data. In the following lines, we will broadly discuss the theory of the X-ray constrained wave function techniques. At first, after a short presentation of the main assumptions of the Jayatilaka approach, we will describe the fundamental equations of the original X-ray constrained Hartree-Fock (XC-HF) method. Then, we will briefly describe the theory of the recent X-ray constrained ELMO (XC-ELMO) technique

and, finally, we will present a prototype of ELMO-based Valence Bond strategy.

5.2 Basic assumptions of the X-ray constrained wave function method

Assuming to work with molecular crystals, let us consider a fictitious crystalline system in which:

- each crystal unit does not interact with the other ones;
- the global electron distribution is identical to the one of the corresponding real interacting system (see Figure 5.1).

The first assumption allows us to express the global wave function of the crystal as:

$$|\Psi_{crystal}\rangle = \prod_i |\Psi_i\rangle \tag{5.1}$$

where all the crystal-unit wave functions $|\Psi_i\rangle$ are formally identical and related to each other through the crystal symmetry operations. The selection of the functional form to describe $|\Psi_i\rangle$ is an arbitrary choice and can be regarded as another assumption of the technique since it will determine the type of X-ray constrained wave function (e.g., Hartree-Fock, ELMO, CI, MCSCF, etc, ...).

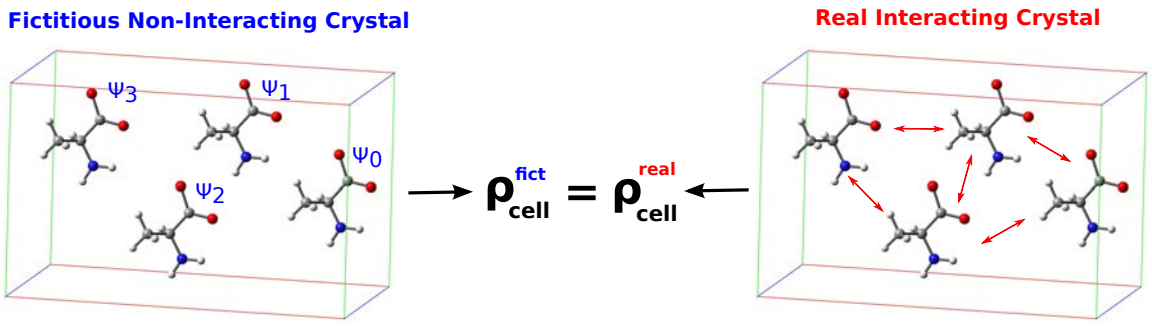

FIGURE 5.1: Schematic representation of the fictitious non-interacting crystal and of the corresponding real interacting crystal.

Now, considering the case in which all the noninteracting units are symmetry-unique portions of the crystal unit cell, the charge density of the unit-cell can be decomposed into a sum of N_m crystal-unit $\rho_i(\mathbf{r})$ that are related by unit-cell symmetry operations $\{\mathbf{R}_i, \mathbf{r}_i\}$ to the reference electron distribution $\rho_0(\mathbf{r})$:

$$\rho_{cell}(\mathbf{r}) = \sum_{i=1}^{N_m} \rho_i(\mathbf{r}) = \sum_{i=1}^{N_m} \rho_0(\mathbf{R}_i^{-1}(\mathbf{r} - \mathbf{r}_i)) \tag{5.2}$$

The charge density $\rho_0(\mathbf{r})$ is associated with the wave function $|\Psi_0\rangle$, which, following Jayatilaka, is not obtained through a simple quantum mechanical calculation, but finding those molecular orbitals that:

- minimize the energy of the single Slater determinant associated with the reference crystal-unit;
- reproduce as much as possible a set of experimental structure factor amplitudes.

That is equivalent to finding the wave function $|\Psi_0\rangle$ that minimizes the following functional:

$$J\left[\Psi_0\right] = \langle\Psi_0|\hat{H}_0|\Psi_0\rangle + \lambda(\chi^2 - \Delta) = E_0\left[\Psi_0\right] + \lambda(\chi^2 - \Delta) \tag{5.3}$$

where $\hat{H}_0$ is the Hamiltonian operator for the reference crystal unit, λ is an external multiplier associated with the experimental constraint and which is manually adjusted during the calculation, Δ is the desired agreement between theoretical and experimental data (usually $\Delta = 1$), and χ^2 is the measure of the statistical agreement between the calculated $|F_{\mathbf{h}}^{calc}|$ and the observed $|F_{\mathbf{h}}^{obs}|$ structure factor amplitudes, which can be expressed as

$$\chi^2 = \frac{1}{N_r - N_p} \sum_h \frac{(\eta|F_{\mathbf{h}}^{calc}| - |F_{\mathbf{h}}^{obs}|)^2}{\sigma_{\mathbf{h}}^2} \tag{5.4}$$

with N_r as number of considered experimental scattering data, N_p as the number of adjustable parameters (in our case, only the external multiplier λ), $\mathbf{h}$ as the triad of Miller indices that labels the reflections, $\sigma_{\mathbf{h}}$ as the error associated with each measure, and η as an $\mathbf{h}$-independent scale factor.

To conclude, it is worth mentioning that, since the functional in (5.3) takes into account the experimental constraint, the initial assumption that the global charge distribution of the fictitious noninteracting crystal is identical to the one of the real interacting system can be fulfilled. It is clear that the quality and the completeness of the experimental

X-ray diffraction data considered in our calculations determine the extent to which this assumption is respected.

5.3 The X-ray constrained Hartree-Fock method

Since the X-ray constrained Hartree-Fock (XC-HF) strategy has been very recently extended to the unrestricted case by Bucinský and co-workers (*185*), in this section, the XC-HF equations will be derived in their spin-orbital form. Let us consider an N-electron system and assume that the reference wave function $|\Psi_0\rangle$ is a single Slater determinant built up with a set of N spin orbitals:

$$|\Psi_0\rangle = \frac{1}{\sqrt{N!}}\hat{A}\left[\psi_1^0\ \psi_2^0\ \psi_3^0 \cdots \psi_N^0\right] \tag{5.5}$$

We will derive the XC-HF equations by means of the minimization of the functional (5.3) with respect to the spin orbitals, which are subject to the spin-orbital orthonormality constraint. This is the reason why, omitting any subscripts and superscripts corresponding to the reference crystal unit, the new functional to be minimized can be expressed like

$$L\left[\psi\right] = E\left[\psi\right] + \lambda(\chi^2\left[\psi\right] - \Delta) - \sum_{i,k=1}^{N} \epsilon_{ki}(\langle\psi_i|\psi_k\rangle - \delta_{ik}) \tag{5.6}$$

where $\{\epsilon_{ki}\}$ is a set of Lagrange multipliers and where ψ denotes the importance of the functional dependence on the spin orbitals.

Now, assuming that real spin orbitals are used and considering the arbitrary variation of $L[\psi]$ with respect to the occupied spin orbital $|\psi_j\rangle$, we obtain

$$\delta_{(\psi_j)}L = \delta_{(\psi_j)}E + \lambda\delta_{\{\psi_j\}}\chi^2 - 2\sum_{k=1}^{N} \epsilon_{kj}(\langle\delta\psi_j|\psi_k\rangle) \tag{5.7}$$

where the energy variation $\delta_{(\psi_j)}E$ can be written as

$$\delta_{(\psi_j)}E = 2\langle\delta\psi_j|\hat{F}|\psi_j\rangle \tag{5.8}$$

with $\hat{F}$ as the usual Fock operator for the spin-orbital case.

Then, exploiting the relation $|F_{\mathbf{h}}^{calc}| = \left[F_{\mathbf{h}}^{calc}(F_{\mathbf{h}}^{calc})^*\right]^{\frac{1}{2}}$, we can express the variation of the statistical agreement with respect to the occupied spin-orbital $\delta_{(\psi_j)}\chi^2$ like this:

$$\delta_{(\psi_j)}\chi^2 = \frac{\eta}{N_r - N_p}\sum_{\mathbf{h}}\frac{\eta|F_{\mathbf{h}}^{calc}| - |F_{\mathbf{h}}^{obs}|}{\sigma_{\mathbf{h}}^2|F_{\mathbf{h}}^{calc}|}\{(F_{\mathbf{h}}^{calc})^*\delta_{(\psi_j)}F_{\mathbf{h}}^{calc} + F_{\mathbf{h}}^{calc}\delta_{(\psi_j)}(F_{\mathbf{h}}^{calc})^*\} \tag{5.9}$$

Since the structure factors are Fourier transforms of the unit-cell electron density and given that the structure factor operator can be defined as

$$\hat{I}_{\mathbf{h}} = \sum_{j=1}^{N_m} e^{i2\pi(\mathbf{R}_j\mathbf{r}+\mathbf{r}_j)\cdot(\mathbf{Bh})} = \hat{I}_{\mathbf{h},R} + i\hat{I}_{\mathbf{h},C} \tag{5.10}$$

where $\mathbf{B}$ is the reciprocal lattice matrix and both $\hat{I}_{\mathbf{h},R}$ and $\hat{I}_{\mathbf{h},C}$ (real and the imaginary part of $\hat{I}_{\mathbf{h}}$, respectively) are Hermitian, we have:

$$F_{\mathbf{h}}^{calc} = \sum_{i=1}^{N}\langle\psi_i|\hat{I}_{\mathbf{h}}|\psi_i\rangle \tag{5.11}$$

and

$$(F_{\mathbf{h}}^{calc})^* = \sum_{i=1}^{N}\langle\psi_i|\hat{I}_{\mathbf{h}}^{\dagger}|\psi_i\rangle \tag{5.12}$$

Moreover, from equations (5.11) and (5.12), we can easily obtain

$$\delta_{(\psi_j)}F_{\mathbf{h}}^{calc} = 2\langle\delta\psi_j|\hat{I}_{\mathbf{h}}|\psi_j\rangle \tag{5.13}$$

and

$$\delta_{(\psi_j)}(F_{\mathbf{h}}^{calc})^* = \left(\delta_{(\psi_j)}F_{\mathbf{h}}^{calc}\right)^* \tag{5.14}$$

Now, exploiting the equation (5.14), the equation (5.9) can be written like this:

$$\delta_{(\psi_j)}\chi^2 = \frac{2\,\eta}{N_r - N_p}\sum_{\mathbf{h}}\frac{\eta|F_{\mathbf{h}}^{calc}| - |F_{\mathbf{h}}^{obs}|}{\sigma_{\mathbf{h}}^2|F_{\mathbf{h}}^{calc}|}\,Re\{(F_{\mathbf{h}}^{calc})^*\delta_{(\psi_j)}F_{\mathbf{h}}^{calc}\} \tag{5.15}$$

and, using the structure factor operator definition (5.10), the relation (5.13) can be expressed as

$$\delta_{(\psi_j)}F_{\mathbf{h}}^{calc} = 2\left[\langle\delta\psi_j|\hat{I}_{\mathbf{h,R}}|\psi_j\rangle + i\langle\delta\psi_j|\hat{I}_{\mathbf{h,C}}|\psi_j\rangle\right] \tag{5.16}$$

Then, it is possible to substitute (5.16) into (5.15) in order to get

$$\delta_{(\psi_j)}\chi^2 = 2\Bigg\{ \sum_{\mathbf{h}} K_{\mathbf{h}} Re\{F_{\mathbf{h}}^{calc}\}\langle\delta\psi_j|\hat{I}_{\mathbf{h,R}}|\psi_j\rangle + \sum_{h} K_{\mathbf{h}} Im\{F_{\mathbf{h}}^{calc}\}\langle\delta\psi_j|\hat{I}_{\mathbf{h,C}}|\psi_j\rangle \Bigg\} \tag{5.17}$$

with

$$K_{\mathbf{h}} = \frac{2\,\eta}{N_r - N_p}\frac{\eta|F_{\mathbf{h}}^{calc}| - |F_{\mathbf{h}}^{obs}|}{\sigma_{\mathbf{h}}^2|F_{\mathbf{h}}^{calc}|} \tag{5.18}$$

and the arbitrary variation of the functional L initially described in (5.7) becomes

$$\delta_{(\psi_j)}L = 2\Bigg\{\langle\delta\psi_j|\hat{F}|\psi_j\rangle \tag{5.19}$$

$$+\lambda\sum_{\mathbf{h}} K_{\mathbf{h}} ReF_{\mathbf{h}}^{calc}\langle\delta\psi_j|\hat{I}_{\mathbf{h,R}}|\psi_j\rangle \tag{5.20}$$

$$+\lambda\sum_{h} K_{\mathbf{h}} ImF_{\mathbf{h}}^{calc}\langle\delta\psi_j|\hat{I}_{\mathbf{h,C}}|\psi_j\rangle \tag{5.21}$$

$$-\sum_{k=1}^{N}\epsilon_{kj}\langle\delta\psi_j|\psi_k\rangle\Bigg\} \tag{5.22}$$

Since the variation $\langle\delta\psi_j|$ is arbitrary, the minimum of the functional L, which corresponds to $\delta_{(\psi_j)}L = 0$ is obtained when the following relation is fulfilled:

$$\hat{F}^{exp}|\psi_j\rangle = \sum_{k=1}^{N}\epsilon_{kj}|\psi_k\rangle \tag{5.23}$$

where the "experimental" Fock operator operator

$$\hat{F}^{exp} = F + \lambda\sum_{\mathbf{h}} K_{\mathbf{h}} Re\{F_{\mathbf{h}}^{calc}\}\hat{I}_{\mathbf{h,R}} + \lambda\sum_{h} K_{\mathbf{h}} Im\{F_{\mathbf{h}}^{calc}\}\hat{I}_{\mathbf{h,C}} \tag{5.24}$$

takes into account the constraint of the experimental structure factor amplitudes.

Finally, if we perform a unitary transformation of the spin orbitals among themselves and if we consider the invariance of $\hat{F}^{exp}$ to that transformation, we can rewrite equation (5.23) in canonical form:

$$\hat{F}^{exp}|\psi_j'\rangle = \epsilon_j'|\psi_j'\rangle \tag{5.25}$$

5.4 The X-ray constrained ELMO technique

After the introduction of the XC-HF method, we will now present a new technique which aims at extracting from X-ray diffraction data a single Slater determinant constructed with extremely localized molecular orbitals. This strategy can be regarded as a combination of the Jayatilaka approach and the Stoll technique, which has been already discussed in section 2.3. Unlike the XC-HF method, the more recent X-ray constrained ELMO (XC-ELMO) method has been developed in the restricted case. In fact, in the XC-ELMO strategy, each crystal unit is a $2N$-electron closed-shell system only described by a single Slater determinant constructed with ELMOs.

In particular, the reference crystal unit is subdivided into f fragments (e.g., atoms, bonds or functional groups) that may overlap. Each fragment is characterized by a proper number of extremely localized molecular orbitals expanded on a local basis-set which is only constituted by the basis functions centred on the atoms involved in the fragment. Therefore, if for the i-th subunit the local basis-set of M_i^0 atomic orbitals is $\{|\chi_{i\mu}^0\rangle\}_{\mu=1}^{M_i^0}$, the α-th ELMO of the fragment can be expressed like this

$$|\varphi_{i\alpha}^0\rangle = \sum_{\mu=1}^{M_i^0} C_{i\mu,i\alpha}^0 |\chi_{i\mu}^0\rangle \tag{5.26}$$

and the ELMO wave function of the reference crystal unit as

$$|\psi_{ELMO}^0\rangle = \frac{1}{\sqrt{(2N)!det[\mathbf{S}]}} \hat{A}\left[\varphi_{11}^0\bar{\varphi}_{11}^0...\varphi_{1n_1}^0\bar{\varphi}_{1n_1}^0...\varphi_{f1}\bar{\varphi}_{f1}^0...\varphi_{fn_f}^0\bar{\varphi}_{fn_f}^0\right] \tag{5.27}$$

where $\hat{A}$ is the antisymmetrizer, n_i is the number of occupied ELMOs for the i-th fragment, f is the total number of fragments, $\varphi_{i\alpha}^0$ is a spin-orbital with spatial part $\varphi_{i\alpha}^0$ and spin part α, whereas $\bar{\varphi}_{i\alpha}^0$ is a spin-orbital with a spatial part $\varphi_{i\alpha}^0$ and spin part β.

Now, omitting subscripts and superscripts corresponding to the reference crystal unit, let us consider the arbitrary variation of the functional J (5.3) associated with the arbitrary variation of the occupied ELMOs $\varphi_{j\beta}$:

$$\delta_{(\varphi_{j\beta})}J = \delta_{(\varphi_{j\beta})}E + \lambda\delta_{(\varphi_{j\beta})}\chi^2 \tag{5.28}$$

where, similarly to equation (5.9), the the variation of the statistical agreement with respect to the occupied ELMO can be written in this way

$$\delta_{(\varphi_{j\beta})}\chi^2 = \frac{\eta}{N_r - N_p}\sum_{\mathbf{h}}\frac{\eta|F_{\mathbf{h}}^{calc}| - |F_{\mathbf{h}}^{obs}|}{\sigma_{\mathbf{h}}^2|F_{\mathbf{h}}^{calc}|}\{(F_{\mathbf{h}}^{calc})^*\delta_{(\varphi_{j\beta})}F_{\mathbf{h}}^{calc} + F_{\mathbf{h}}^{calc}\delta_{(\varphi_{j\beta})}(F_{\mathbf{h}}^{calc})^*\} \tag{5.29}$$

Then, as already mentioned in (2.19), following Stoll we have

$$\delta_{(\varphi_{j\beta})}E = 4\langle\delta\varphi_{j\beta}|(1-\hat{\rho})\hat{F}|\tilde{\varphi}_{j\beta}\rangle \tag{5.30}$$

where the reciprocal orbitals $|\tilde{\varphi}_{j\beta}\rangle$ can be defined like

$$|\tilde{\varphi}_{j\beta}\rangle = \sum_{k=1}^{f}\sum_{\gamma=1}^{n_k}\left[\mathbf{S}^{-1}\right]_{k\gamma,j\beta}|\varphi_{k\gamma}\rangle \tag{5.31}$$

and the global density operator $\hat{\rho}$ depends on all the occupied ELMOs and is given by:

$$\hat{\rho} = \sum_{j=1}^{f}\sum_{\beta}^{n_j}|\tilde{\varphi}_{j\beta}\rangle\langle\varphi_{j\beta}| = \sum_{j=1}^{f}\sum_{\beta}^{n_k}|\varphi_{j\beta}\rangle\langle\tilde{\varphi}_{j\beta}| \tag{5.32}$$

Moreover, using again the structure factor operator (5.10), we obtain

$$F_{\mathbf{h}}^{calc} = 2\sum_{i=1}^{f}\sum_{\alpha=1}^{n_i}\langle\varphi_{i\alpha}|\hat{I}_{\mathbf{h}}|\tilde{\varphi}_{i\alpha}\rangle \tag{5.33}$$

and

$$(F_{\mathbf{h}}^{calc})^* = 2\sum_{i=1}^{f}\sum_{\alpha=1}^{n_i}\langle\tilde{\varphi}_{i\alpha}|\hat{I}_{\mathbf{h}}^{\dagger}|\varphi_{i\alpha}\rangle \tag{5.34}$$

Then, by means of the relation (2.17), we get

$$\delta_{(\varphi_{j\beta})}F_{\mathbf{h}}^{calc} = 2\left\{\langle\delta\varphi_{j\beta}|(1-\hat{\rho})\hat{I}_{\mathbf{h}}|\tilde{\varphi}_{j\beta}\rangle + \langle\delta\tilde{\varphi_{j\beta}}|\hat{I}_{\mathbf{h}}(1-\hat{\rho})|\varphi_{j\beta}\rangle\right\} \tag{5.35}$$

and

$$\delta_{(j\beta)}(F_{\mathbf{h}}^{calc})^* = (\delta_{(j\beta)}F_{\mathbf{h}}^{calc})^* \tag{5.36}$$

Then, as in the XC-HF case, using equation (5.36) in (5.29), the variation of the statistical agreement $\delta_{(\psi_{j\beta})}\chi^2$ becomes

$$\delta_{(\psi_{j\beta})}\chi^2 = \frac{2\,\eta}{N_r - N_p}\sum_{\mathbf{h}}\frac{\eta|F_{\mathbf{h}}^{calc}| - |F_{\mathbf{h}}^{obs}|}{\sigma_{\mathbf{h}}^2|F_{\mathbf{h}}^{calc}|}\,Re\{(F_{\mathbf{h}}^{calc})^*\delta_{(\psi_{j\beta})}F_{\mathbf{h}}^{calc}\} \tag{5.37}$$

and, moreover, decomposing the structure factor operator (5.10) into its real and imaginary parts , we can express $\delta_{(j\beta)}F_{\mathbf{h}}^{calc}$ like this

$$\delta_{(j\beta)}F_{\mathbf{h}}^{calc} = 4\Big\{\langle\delta\varphi_{j\beta}|(1-\hat{\rho})\hat{I}_{\mathbf{h},R}|\tilde{\varphi}_{j\beta}\rangle + i\,\langle\delta\varphi_{j\beta}|(1-\hat{\rho})\hat{I}_{\mathbf{h},C}|\tilde{\varphi}_{j\beta}\rangle\Big\} \tag{5.38}$$

Now, if we substitute equations (5.38) and (5.34) into equation (5.37), we can write

$$\delta_{(\psi_{j\beta})}\chi^2 = 4\left\{\sum_{\mathbf{h}}K_{\mathbf{h}}Re\{F_{\mathbf{h}}^{calc}\}\langle\delta\varphi_{j\beta}|(1-\hat{\rho})\hat{I}_{\mathbf{h},R}|\tilde{\varphi}_{j\beta}\rangle + \sum_{\mathbf{h}}K_{\mathbf{h}}Im\{F_{\mathbf{h}}^{calc}\}\langle\delta\varphi_{j\beta}|(1-\hat{\rho})\hat{I}_{\mathbf{h},C}|\tilde{\varphi}_{j\beta}\rangle\right\} \tag{5.39}$$

where $K_{\mathbf{h}}$ is equivalent to the one in the XC-HF strategy (see (5.18)).

Therefore, the arbitrary variation of the functional J given by equation (5.28) becomes:

$$\begin{aligned}\delta_{(\varphi_{j\beta})}J = 4\Big\{&\langle\delta\varphi_{j\beta}|(1-\hat{\rho})\hat{F}|\tilde{\varphi}_{j\beta}\rangle && (5.40)\\ &+\lambda\sum_{\mathbf{h}}K_{\mathbf{h}}Re\{F_{\mathbf{h}}^{calc}\}\langle\delta\varphi_{j\beta}|(1-\hat{\rho})\hat{I}_{\mathbf{h},R}|\tilde{\varphi}_{j\beta}\rangle && (5.41)\\ &+\lambda\sum_{\mathbf{h}}K_{\mathbf{h}}Im\{F_{\mathbf{h}}^{calc}\}\langle\delta\varphi_{j\beta}|(1-\hat{\rho})\hat{I}_{\mathbf{h},C}|\tilde{\varphi}_{j\beta}\rangle\Big\} && (5.42)\end{aligned}$$

Since the minimum of the functional is attained when $\delta_{(\varphi_{j\beta})}J = 0$ for all j and β, the desired X-ray constrained ELMOs are the ones that satisfy the following equation:

$$\left\{(1-\hat{\rho})\hat{F} + \lambda\sum_{\mathbf{h}}K_{\mathbf{h}}Re\{F_{\mathbf{h}}^{calc}\}(1-\hat{\rho})\hat{I}_{\mathbf{h},R} + \lambda\sum_{\mathbf{h}}K_{\mathbf{h}}Im\{F_{\mathbf{h}}^{calc}\}(1-\hat{\rho})\hat{I}_{\mathbf{h},C}\right\}|\tilde{\varphi}_{j\beta}\rangle = 0 \tag{5.43}$$

Now, considering the definition of fragment density operator (see equation (2.22)) and adding the following quantity

$$|Q\rangle = \left\{\hat{\rho}_j^{\dagger}\hat{F} + \lambda\sum_{\mathbf{h}}K_{\mathbf{h}}Re\{F_{\mathbf{h}}^{calc}\}\hat{\rho}_j^{\dagger}\hat{I}_{\mathbf{h},R} + \lambda\sum_{\mathbf{h}}K_{\mathbf{h}}Im\{F_{\mathbf{h}}^{calc}\}\hat{\rho}_j^{\dagger}\hat{I}_{\mathbf{h},C}\right\}|\tilde{\varphi}_{j\beta}\rangle \tag{5.44}$$

to both sides of equation (5.43), we obtain

$$\hat{F}^{j,exp}|\varphi_{j\beta}\rangle = \sum_{\gamma=1}^{n_j}\langle\varphi_{j\gamma}|\hat{F}^{j,exp}|\varphi_{j\beta}\rangle|\varphi_{j\gamma}\rangle \tag{5.45}$$

where $\hat{F}^{j,exp}$ is the modified "experimental" Fock operator for each fragment:

$$\hat{F}^{j,exp} = \hat{F}^{j} + \lambda\sum_{\mathbf{h}}\hat{I}^{j}_{\mathbf{h},R} + \hat{I}^{j}_{\mathbf{h},C} \tag{5.46}$$

with $\hat{F}^{j}$ as the modified Fock operator for the traditional ELMOs (5.46) and with $\hat{I}_{\mathbf{h},R}$ and $\hat{I}_{\mathbf{h},C}$ as

$$\hat{I}^{j}_{\mathbf{h},R} = K_{\mathbf{h}}\, Re\{F^{calc}_{\mathbf{h}}\}(1-\hat{\rho}+\hat{\rho}^{\dagger}_{j})\hat{I}_{\mathbf{h},R}(1-\hat{\rho}+\hat{\rho}_{j}) \tag{5.47}$$

$$\hat{I}^{j}_{\mathbf{h},C} = K_{\mathbf{h}}\, Im\{F^{calc}_{\mathbf{h}}\}(1-\hat{\rho}+\hat{\rho}^{\dagger}_{j})\hat{I}_{\mathbf{h},C}(1-\hat{\rho}+\hat{\rho}_{j}) \tag{5.48}$$

Finally, applying a unitary transformation that only mixes the occupied ELMOs of each fragment among themselves and taking into account the invariance of $\hat{F}^{j,exp}$ to that transformation, we can rewrite equation (5.45) like this

$$\hat{F}^{j,exp}|\varphi^{'}_{j\beta}\rangle = \epsilon^{'}_{j\beta}|\varphi_{j\beta}\rangle \tag{5.49}$$

and, therefore, the "experimental" ELMOs are obtained solving self-consistently for each fragment equation (5.49). It is worth noting that, although solved separately, the equations associated with the different subunits are coupled since each $\hat{F}^{j,exp}$ operator depends on the global density operator $\hat{\rho}$. In case of convergence problems associated with the intrinsic nonorthogonality of the ELMOs (*70*, *81*, *85*), an alternative strategy for the direct minimization of the functional J, which is based on a quasi-Newton procedure exploiting first- and second-order derivatives of J with respect to the XC-ELMO coefficients, has been proposed (*85*, *193*, *194*).

5.5 Details about the X-ray constrained calculations

In this section, we will discuss some important details related to the X-ray constrained wave functions calculations. At first, it is important to point out that the set of structure

factor amplitudes $\{|F_{\mathbf{h}}^{calc}|\}$ have been computed taking into account the effects of the nuclear thermal motion. In particular, as already observed by Jayatilaka and co-workers (*180*, *181*), the calculated structure factors, which are Fourier transforms of the unit-cell electron density, can be expressed using the following relation:

$$F_{\mathbf{h}}^{calc} = Tr[\mathbf{D}I_{\mathbf{h}}] \tag{5.50}$$

where **D** is the one-particle LCAO density matrix associated with the reference ELMO wave function Ψ_0 of the reference crystal unit and $I_{\mathbf{h}}$ is a matrix whose elements, in the atomic orbital basis, are given by the expression:

$$[I_{\mathbf{h}}]_{\mu\nu} = \sum_{j=1}^{N_m} e^{i2\pi\mathbf{r}_j\cdot(\mathbf{Bh})}\, T_{\mu\nu}[\mathbf{B}^{-1}\mathbf{R}_j^T\mathbf{Bh}] \int \chi_\mu(\mathbf{r})\chi_\nu(\mathbf{r}) e^{i2\pi\mathbf{R}_j\mathbf{r}\cdot(\mathbf{Bh})} d\mathbf{r} \tag{5.51}$$

where N_m is the number of equivalent positions in the unit-cell, $\{\mathbf{R}_j, \mathbf{r}_j\}$ are the unit-cell symmetry operations, **B** is the reciprocal lattice matrix and **h** is the triad of Miller indexes that label the reflection, while $T_{\mu\nu}[\mathbf{B}^{-1}\mathbf{R}_j^T\mathbf{Bh}]$ is a parameter that empirically includes the thermal motion effects on the charge density. In our case, we have exploited the model introduced by Stewart (*128*):

$$T_{\mu\nu}\left[\mathbf{h}^{'}\right] = exp\left[-2\pi\tau_{\mu\nu}(\mathbf{Bh}^{'})\cdot(\mathbf{U}^{\mu}+\mathbf{U}^{\nu})\mathbf{Bh}^{'}\right] \tag{5.52}$$

with $(\mathbf{U}^{\mu}+\mathbf{U}^{\nu})$ as the matrices containing the experimentally determined thermal vibrations parameters for the atoms on which the basis functions $\chi_\mu(\mathbf{r})$ and $\chi_\mu(\mathbf{r})$ are centred and $\tau_{\mu\nu}$ as an empirical parameter that depends on the distance of the atoms associated with the basis functions. In particular, we have that:

- $\tau_{\mu\nu} = 0.5$ if the distance between the atoms associated with the atomic orbitals $\chi_\mu(\mathbf{r})$ and $\chi_\nu(\mathbf{r})$ is lower that 2.5 bohr;
- $\tau_{\mu\nu} = 0.25$ in all the other cases.

Moreover, it is worth mentioning that, due to the experimental errors associated with the X-ray diffraction data, the statistical agreement χ^2 is not usually forced to be equal to 0, but the desired agreement Δ is generally chosen equal to 1.0, so that the final computed structure factor amplitudes are on average within one standard deviation of the

experimental measurements. However, in many cases (especially for low-quality crystallographic data) the X-ray constrained wave function calculations do not easily converge to to 1.0 and, therefore, to overcome this drawback, different criteria have been proposed (*187*, *188*, *194*) to determine when to stop the computations. However, although sound and reliable strategies have been introduced and successfully applied, this remains an open problem that will deserve further investigations in the future.

In this respect, one possibility consists in stopping the procedure when one of three conditions is reached (*187*, *188*, *194*):

$$\begin{cases} \chi^2 < 1 \\ \frac{\chi_i^2 - \chi_{i-1}^2}{\lambda_i - \lambda_{i-1}} > -0.5 \\ \left| \frac{E_{\lambda_i}^{el} - E_{\lambda=0}^{el}}{E_{\lambda=0}^{el}} \right| > 5 \times 10^{-4} \end{cases} \tag{5.53}$$

The first condition is the usual one. The other two have been proposed in order to prevent situations in which the gain in the statistical agreement χ^2 becomes minimal compared to a large modification of the external multiplier λ. In fact, the use of a very large value of λ can lead to large unphysical changes in the energy and, in those situations, the wave function can become meaningless. This is the reason why, the second criteria to stop the X-ray constrained procedure is based on the incremental ratio of χ^2 with respect to λ. Furthermore, the third threshold takes into account the absolute relative variation of the electronic energy of the system during the "fitting" procedure with respect to the initial unconstrained electronic energy (obtained at $\lambda = 0$).

5.6 ELMO based Valence Bond method

In order to evaluate the weight of different resonance structures in the description of a molecular system having a multi-reference character, Genoni has recently developed a prototype of an ELMO-based Valence Bond (VB) method (*195*) . In particular, this strategy has been extended in the framework of the X-ray constrained wave function approaches to also take into account the structure factor amplitudes measured experimentally. This new strategy can be also seen as the first attempt to propose a multi-determinant wave function *ansatz* in the framework of the Jayatilaka approach.

Also for this novel X-ray constrained "ELMO-VB" technique, we assume to work with molecular crystals, where each crystal unit does not interact with the other ones. As already described above, the global wave function of the crystal can be simply expressed as a product of formally identical and symmetry related wave functions corresponding to the different molecular units and, moreover, the global electron density of the fictitious non-interacting crystal is assumed equivalent to the global electron density of the corresponding real interacting system. Nevertheless, in this approach, each crystal-unit wave function can be written in the following form

$$|\Psi_{ELMO-VB}\rangle = \sum_i C_i |\Psi_i\rangle \tag{5.54}$$

where the functions $\{|\Psi_i\rangle\}$ in the linear combination are single Slater determinants that describes all the possible resonance structures of the system in exam. In particular, in our case, they will consist in ELMO wave functions pre-determined through unconstrained ELMO calculations exploiting localization schemes corresponding to the resonance structures.

For example, if we were interested in studying the benzene, the wave function for the reference crystal unit would be a linear combination of the ELMO wave function corresponding to the localization scheme for the resonance structure A and of the ELMO wave function associated with the localization scheme for the resonance structure B (see Figure 5.2).

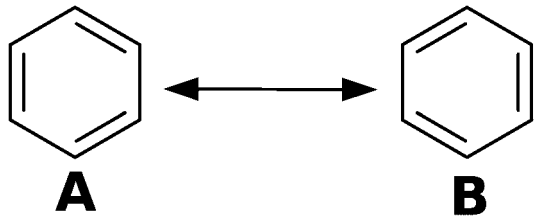

FIGURE 5.2: Resonance structures of the benzene molecule.

In the proposed method, the pre-optimized ELMOs will be kept frozen, while the coefficients $\{C_i\}$ of expression (5.54) will be kept determinant minimizing the following functional:

$$J[\mathbf{C}] = E[\mathbf{C}] + \lambda \left(\chi^2[\mathbf{C}] - \Delta\right) \tag{5.55}$$

where E is the energy of the system associated with the wave function (5.54), χ^2 is the statistical agreement between calculated and experimental structure factors already given in (5.4) and Δ is the desired agreement between calculated and experimental values. **C** indicates the functional dependence of J, E and χ^2 on the coefficient $\{C_i\}$ of expansion (5.54).

Due to the non-orthogonality of the Slater determinant $\{\Psi_i\}$, the coefficients $\{C_i\}$ do not give the real weight associated with the corresponding resonance structures. These weights are actually given by the Chirgwin-Coulson coefficients defined as:

$$K_i = |C_i|^2 + \sum_{j \neq i} C_i \, C_j \, S_{ij} \tag{5.56}$$

where $S_{ij} = \langle \Psi_i | \Psi_j \rangle$ is the overlap between the pre-determined ELMO wave functions $|\Psi_i\rangle$ and $|\Psi_j\rangle$.

In analogy with the current X-ray constrained wave function methods, in this new X-ray constrained "ELMO-VB" strategy, the external parameter λ is iteratively adjusted until a convergence criterion is fulfilled. In this case, we chose to stop the procedure when, for all the Chirgwin-Coulson coefficients, the difference between two consecutive λ-steps is lower than or equal to 2.0×10^{-4} namely:

$$|K_j^{\lambda_i} - K_j^{\lambda_{i-1}}| \leq 2.0 \times 10^{-4} \tag{5.57}$$

Finally, it is important to precise that, it is also possible to perform unconstrained "ELMO-VB" calculations. In this situation, one has to simply minimize the energy associated with the global wave function without taking into account the constraints of the experimental data, namely it is necessary to minimize the following functional in function of the coefficients $\{C_i\}$:

$$E[\mathbf{C}] = \langle \Psi | \hat{H} | \Psi \rangle = \sum_{j,i} C_i C_j \langle \Psi_i | \hat{H} | \Psi_j \rangle \tag{5.58}$$

The unconstrained and the X-ray constrained "ELMO-VB" techniques has been used to study and analyse the charge density and the difference resonance forms of the *syn*-1,6:8,13-Biscarbonyl[14] annulene. The results of this investigation will be discussed in Chapter 8.

5.7 Conclusion

To summarize, in this chapter, after an historical review of some experimentally constrained wave function methods, we have presented the X-ray constrained wave function technique initially proposed by Jayatilaka and recently extended by Genoni in the framework of the Extremely Localized Molecular Orbitals theory. Moreover, we have also briefly introduced a new ELMO based Valence Bond strategy.

As we have seen in the previous chapters, since the ELMOs are transferable from molecules to molecules, it is possible to build up ELMOs libraries in order to reconstruct the electron density and the wave function of very large systems almost instantaneously. Thus, we can also imagine to exploit extremely localized molecular orbital obtained through the XC-ELMO method to construct such databases. However, we do not know if these "experimental" ELMOs are more suitable for this purpose compared to the "theoretical" ones. In order to give more insights into this problem we have evaluated both the effects of using the ELMOs in the framework of the X-ray constrained wave function approach and the intrinsic capability of the Jayatilaka method in capturing the electron correlation effects on the electron density. These two investigations will be discussed in deeper detail in the next two chapters.

Chapter 6

Effects of the *a priori* localization on the X-Ray constrained wavefunction

Résumé

Ce chapitre est dédié à l'étude des effets de la localisation dans le cadre de la méthode de la fonction d'onde contrainte développée par Jayatilika. En effet, comme nous l'avons décrit dans le chapitre précédent, l'utilisation d'orbitales extrêmement localisées représente une contrainte significative dans la détermination de la structure électronique d'un système. Notre but dans ce chapitre est donc de déterminer comment l'utilisation des ELMOs peut affecter la précision des calculs dans le cas d'une méthode de fonction d'onde contrainte.

Pour ce faire, nous allons comparer des résultats obtenus par des approches Hartree-Fock (XC-RHF) et ELMO contraintes (XC-ELMO). Nous analyserons particulièrement l'accord statistique avec les mesures expérimentales, les tendances de convergence ou encore certaines propriétés topologiques de la densité. Comme cela est montré ci-dessous, nous avons découvert que l'utilisation de petites fonctions de bases a une grande influence sur les densités obtenues. Toutefois, si l'on utilise des fonctions de base plus étendues et plus flexibles, les deux méthodes XC-RHF et XC-ELMO tendent vers des résultats similaires. Nous avons expliqué ce comportement en considérant que l'utilisation de bases de plus en plus flexibles permet de mieux en mieux exploiter les informations contenues dans les données expérimentales et que, par conséquent, cela atténue le biais introduit par la localisation.

6.1 Motivations

In the previous chapter, some experimentally constrained wave function strategies and the X-ray constrained Hartree-Fock method proposed by Jayatilaka have been presented. Moreover, an alternative technique that enables the extraction of a single Slater determinant constructed with ELMOs from X-ray diffraction measurements has also been discussed. In particular, since the extremely localized molecular orbitals are transferable from a molecule to another, we can imagine of storing these "experimental" ELMOs (instead of the theoretical ones) in databases in order to spontaneously recover the wave function and the electron density of very large systems indirectly using experimental data. Nevertheless, the introduction of a predefined localization is a very strong empirical constraint in electronic structure calculations and, in this regard, the effects of using ELMOs in the framework of the X-ray constrained wave function fitting procedures were unknown.

Therefore, in the present chapter, we will present the studies conducted in order to assess the influence of the *a priori* localization on experimentally reconstructed wave functions (*196*). To accomplish this task, we have analysed both the convergence towards the desired statistical agreement with the collected X-ray diffraction data and the main topological properties of the resulting charge distributions.

6.2 Methods

6.2.1 A case study: the L-Alanine

In order to study the effects of the molecular orbitals localization when X-ray constrained calculations are performed, we have considered the crystallographic structure of the zwitterionic L-alanine determined by Destro and co-workers (*197*) at 23 K. Using the crystal geometry (see Figure 6.1), we have carried out single point calculations at Restricted Hartree-Fock (RHF), ELMO, XC-RHF (X-ray constrained Hartree-Fock) and XC-ELMO (X-ray constrained ELMO) levels. All these calculations have been performed considering six different basis sets of increasing size and flexibility, namely the 3-21G, 6-31G(d), cc-pVDZ, 6-311G(d,p), aug-cc-pVDZ, and 6-311++G(2d,2p) basis sets.

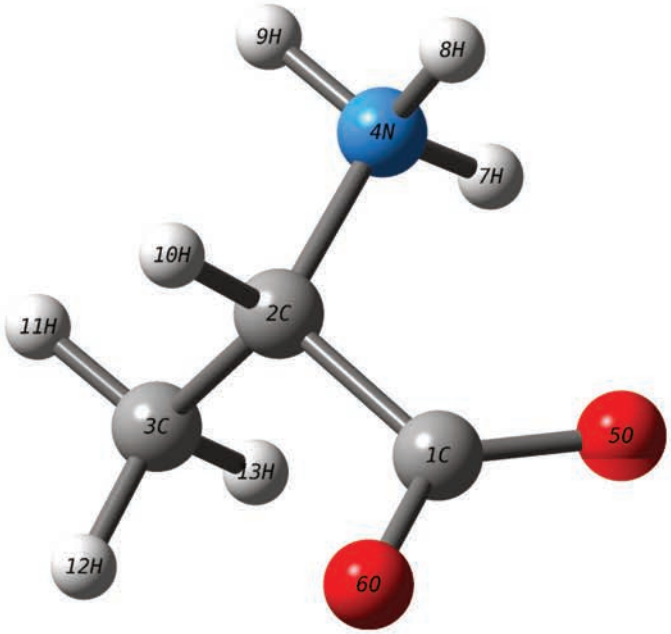

FIGURE 6.1: L-Alanine molecular geometry obtained from the X-ray diffraction experiment.

It is worth noting that, for all the ELMO and the XC-ELMO calculations performed in our investigations, we have adopted a localization scheme almost corresponding to the Lewis structure of the molecular system. Actually, we have used atomic subunits to describe the core and the lone-pair electrons and diatomic fragments to characterize the bond electron-pairs. The only exception in our study is represented by the σ and the π electrons of the carboxylate group for which the three atom fragment C1-O5-O6 (see Figure 6.1) has been considered.

Furthermore, in the case of the XC-RHF and XC-ELMO computations, we have taken into account the unit-cell parameters, the Atomic Displacement Parameters (ADPs), and the experimental structure factor amplitudes experimentally determined by Desto *et al.* in their investigation. For the X-ray diffraction data, we have only selected the reflections characterized by an amplitude greater that three times the experimental error, namely $|F_{\mathbf{h}}^{calc}| > 3\sigma_{\mathbf{h}}$. Overall, a set of 2435 structure factor amplitudes has been considered for our "experimental" wave functions calculations. It is important to precise that the diffraction measurements were unaffected by serious secondary extinction phenomena and, therefore, a particular treatment for these effects was not necessary.

6.2.2 Computational details

We have carried out all the RHF and the XC-RHF computations using the TONTO package developed by Jayatilaka and co-workers, while all the ELMO and the XC-ELMO calculations have been performed exploiting a modified version of the GAMESS-UK quantum chemistry package (*121*) to implement both the original ELMO technique and also the novel X-ray constrained ELMO strategy presented in Chapter 5 (see section 5.4). Furthermore, the topological analyses of the resulting electron densities have been carried out using the AIMAll package (version 13.11.14) (*142*).

6.2.3 Comparison of the obtained charge densities

In our study, in order to evaluate the effects of introducing an *a priori* molecular orbital localization in X-ray constrained wave function fitting procedures, we have also compared the electron densities obtained from the previously discussed calculations. To accomplish this task, we have mainly compared topological properties obtained through Quantum Theory of Atoms in Molecules analyses (*44*). We have manly focused our attention on the values of the electron density at the Bond Critical Points (BCPs) $\rho(\mathbf{r}_b)$ and on the Laplacian of the electron density at the BCPs $\nabla^2\rho(\mathbf{r}_b)$. Moreover, we have also compared the net integrated atomic charges obtained by means of the integration of the different electron distributions over the QTAIM atomic basins. In the present work, the global charges of the main fragments (i.e., the amino group, the carboxylic group, the C_α-H bond and the CH_3 group) have been taken into account.

6.3 Results and Discussion

6.3.1 Performance and convergence of the methods

At first, the agreement statistics χ^2 and the values of the total energy have been compared both for the unconstrained and experimentally constrained computations. In Table 6.1, as expected, all the energies associated with the X-ray constrained wave functions are higher than the energy arising from the corresponding unconstrained calculations. Actually, this is the consequence of adding a constraint to a variational procedure without providing another variational parameter.

TABLE 6.1: Agreement statistics χ^2 and energy values corresponding to the RHF, ELMO, XC-RHF and XC-ELMO calculations performed on the L-Alanine system.

	unconstrained calculations		X-ray constrained calculations		
method / basis-set	χ^2	energy (a.u)	λ_{max}	χ^2	energy (a.u)
3-21G					
RHF	4.43	-320.001	0.442	1.71	-319.841
ELMO	5.66	-319.850	0.580	2.29	-319.609
6-31G(d)					
RHF	2.53	-321.798	0.248	1.26	-321.747
ELMO	2.94	-321.696	0.320	1.30	-321.621
cc-pVDZ					
RHF	2.52	-321.831	0.240	1.23	-321.782
ELMO	3.02	-321.711	0.300	1.29	-321.640
6-311G(d,p)					
RHF	2.48	-321.891	0.230	1.21	-321.846
ELMO	2.94	-321.778	0.280	1.27	-321.712
aug-cc-pVDZ					
RHF	2.52	-321.863	0.230	1.20	-321.815
ELMO	2.74	-321.769	0.260	1.23	-321.712
6-311++G(2d,2p)					
RHF	2.47	-321.918	0.226	1.17	-321.873
ELMO	2.78	-321.830	0.280	1.19	-321.769

Moreover, in Table 6.1 we have can easily observe that the X-ray constrained approaches always provide a better statistical agreement with the experimental data compared to the unconstrained methods. It is also possible to note that for all the considered methods this agreement improves when larger and more flexible basis sets are used. However, even using the very large 6-311++G(2d,2p) basis-set, no calculations were able to reach desired agreement $\chi^2 = 1$. As previously explained by Genoni in other papers (*193*, *194*), the difficulties in converging toward the desired agreement is strongly associated with experimental errors in the collected measurements.

To better understand this problem, the normalized residuals $\eta|F_{\mathbf{h}}^{calc}| - |F_{\mathbf{h}}^{exp}|/\sigma_{\mathbf{h}}$ corresponding to all the experimentally constrained computations have been compared. In

Figure 6.2, where the normalized residuals as a function of scattering resolution have been plotted for the 3-21G, 6-311G(d,p) and 6-311++G(2d,2p) basis-set, we have systematically detected an outlier for all the X-ray constrained calculations at $sin\theta/\lambda \approx 1.0\text{Å}^{-1}$. In fact, this normalized residual is approximately equal to 10.8 in all the cases and, if we exclude this value in the χ^2 calculation for the 6-311++G(2d,2p) basis-set, the statistical agreement further decreases from 1.17 to 1.12 and from 1.19 to 1.14 for the XC-RHF and the XC-ELMO methods, respectively. This result further confirms that the impossibility of approaching the desired agreement ($\chi^2 = 1$) is completely related to the quality of the diffraction data used in the X-ray constrained wave functions calculations.

Moreover, observing the λ_{max} values for the constrained wave function computations in Table 6.1, we can easily remark that the procedures are earlier stopped when more flexible basis sets are used. For example, the maximum value of the external multiplier in the case of the XC-RHF calculations varied from 0.442 for the 3-21G to 0.226 using the 6-311++G(2d,2p) basis-set. However, it is worth noting that, although for large basis sets the calculations are halted for smaller values of λ, the statistical agreement is anyway lower than the one obtained with smaller sets of basis functions. This important trend can be ascribed to the fact that the information provided by the experimental data is better captured using larger and more flexible basis sets. In Table 6.1, we can also note that all the RHF calculations provide statistical agreement values lower than those corresponding to the ELMO computations. If this tendency is evident when unconstrained calculations are performed, the differences are significantly reduced in the case of X-ray constrained approaches. This is probably due to the fact that the effects introduced by the experimental data possibly compensate the quite strong approximation represented by the introduction of a predefined localization scheme. In particular, we can see that the difference between the XC-RHF and XC-ELMO statistical agreements is considerably reduced when larger basis sets are used (from 0.58 to 0.002 for the 3-21G and the 6-311++G(2d,2p) basis sets, respectively).

The same tendency has been detected in Figure 6.3, where the values of χ^2 in function of the external multiplier corresponding to X-ray constrained Hartree-Fock and X-ray constrained ELMO calculations are presented. For the sake of simplicity, in Figure 6.3 we have only presented the results obtained employing three different basis sets (3-21G, 6-311G(d,p) and 6-311++G(2d,2p)). While in the 3-21G case the XC-RHF and the XC-ELMO

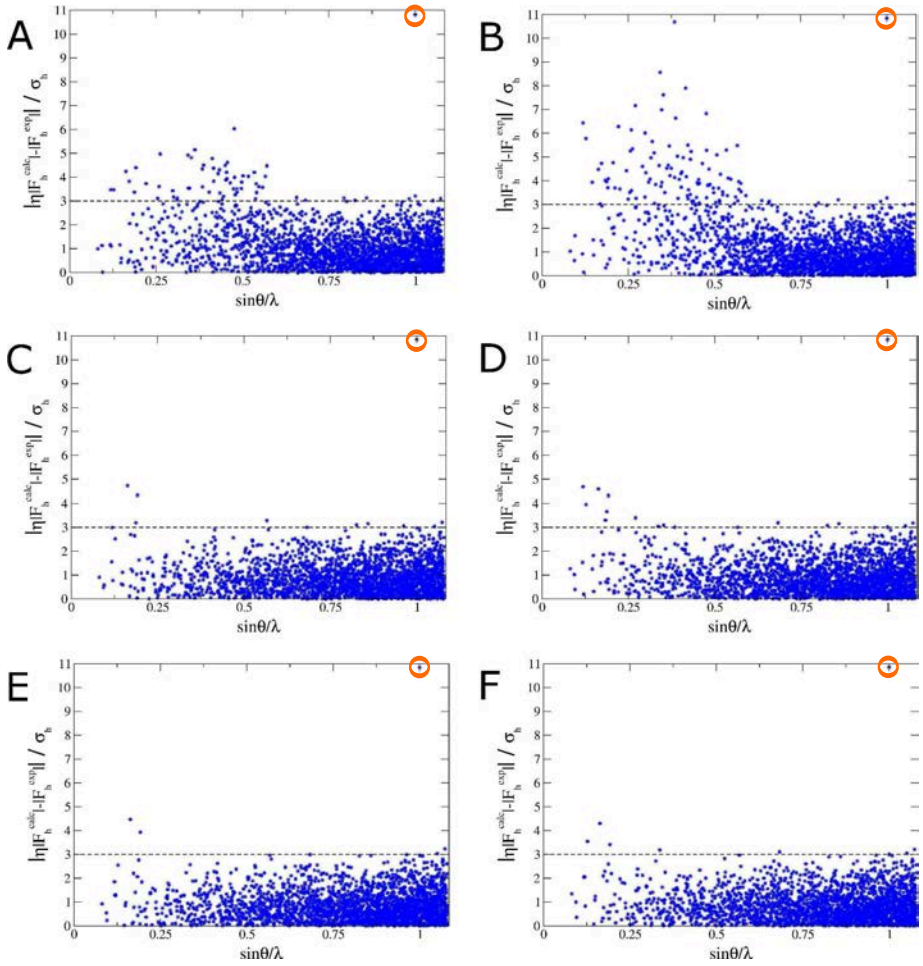

FIGURE 6.2: Normalized residuals of the structure factor amplitudes in function of the scattering resolution for the (A) XC-RHF/3-21G, (B) XC-ELMO/3-21G, (C) XC-RHF/6-311G(d,p), (D) XC-ELMO/6-311G(d,p), (E) XC-RHF/6-311++G(2d,2p), and (F) XC-ELMO/6-311++G(2d,2p) calculations. The horizontal dashed line represents the outlier limit. The systematic outliers occurring at $sin\theta/\lambda \approx 1.0 \text{Å}^{-1}$ are encircled.

calculations converge to two different values (see Figure 6.3 A), in the two other situations, the two curves seem to progressively approach the same limit (see Figure 6.3 B and C), which is not far from 1.0. This means that, when larger and more flexible basis sets are used, the two XC-RHF and XC-ELMO procedures are converging to a similar set of structure factor amplitudes.

Now, analysing more carefully the normalized residuals presented in Figure 6.2, it is possible to obtain more insights into the tendencies discussed above. Indeed, it is easy

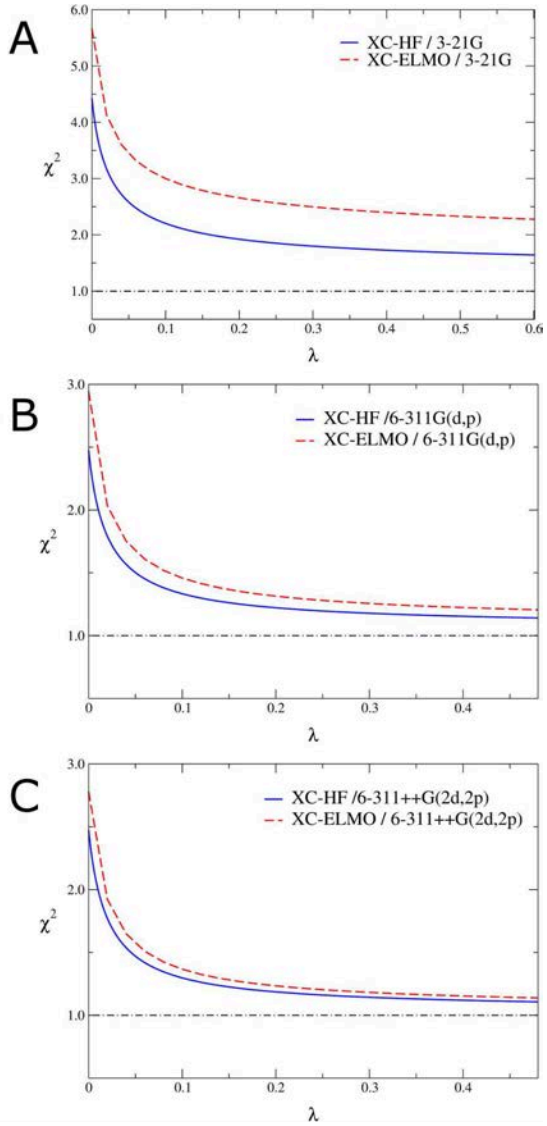

FIGURE 6.3: Variation of the χ^2 agreement statistics in function of the external multiplier λ for the XC-RHF and the XC-ELMO methods when the (A) 3-21G, (B) 6-311G(d,p), and (C) 6-311++G(2d,2p) basis-sets are considered. The solid blue and the dashed red lines represent the XC-RHF and XC-ELMO trends, respectively. The horizontal dash-and-dot line shows the desired statistical agreement.

to observe that the differences in the number of outliers (normalized residuals greater than 3.0) between the corresponding XC-RHF and XC-ELMO methods is continuously reduced when more and more flexible basis sets are used. In particular, the number of outliers for the XC-RHF computations decreases from 65 (3-21G) to 5 (6-311++G(2d,2p)), while for the XC-ELMO calculations the difference is even more important: dropping

from 116 to 8 outliers for the 3-21G and the 6-311++G(2d,2p) basis sets, respectively.

Finally, it seems clear that all the results presented and discussed in this subsection suggest that the initial bias introduced imposing a strict *a priori* localization scheme in experimentally constrained wave function calculations can be considerably reduced exploiting large sets of basis functions. Actually, larger and larger basis sets probably allow to capture more information coming from the experimental data and, therefore, they can decrease the influence of the strict localization constraints initially forced on the electronic structure.

6.3.2 Effects of the localization on the topological properties

To further study the effects of inserting an *a priori* molecular orbital localization in experimentally constrained wave function fitting procedures, we have also compared the charge densities obtained from the previously presented calculations. We have mainly focused our attention on two important topological properties currently used in QTAIM charge density analyses, namely the values of the electron density at the Bond Critical Points (BCPs) ($\rho(\mathbf{r}_b)$) and the Laplacian of the electron density at the BCPs ($\nabla^2\rho(\mathbf{r}_b)$). Furthermore, the global charges of the principal fragments (i.e., the amino group, the carboxylic group, the C_α-H bond and the CH_3 group) have been also considered. For the sake of clarity, we have only considered the electron densities computed using the 3-21G, 6-311G(d,p) and 6-311++G(2d,2p) basis sets. Since our main goal was to determine the influence of the molecular orbitals localization on the final electron distributions, we have computed the relative variation of the considered topological properties for the ELMO and the XC-ELMO procedures taking the RHF and the XC-RHF values as references like:

$$\text{U-RV(\%)} = 100 \times \frac{(X_{ELMO} - X_{RHF})}{X_{RHF}} \tag{6.1}$$

and

$$\text{XC-RV(\%)} = 100 \times \frac{(X_{XC-ELMO} - X_{XC-RHF})}{X_{XC-RHF}} \tag{6.2}$$

with $X = \rho(\mathbf{r}_b)$ or $\nabla^2\rho(\mathbf{r}_b)$.
In Table 6.2, we have reported the relative variations of the electron densities values at the BCPs. We can immediately notice that, for all the basis sets taken into account and both for the unconstrained and the XC-methods, the use of the ELMOs entails quite small

discrepancies with respect to the corresponding Hartree-Fock calculations. Indeed, the largest differences have been detected for the C1-O5 bond for the 3-21G basis-set in the case of the XC-ELMO computation, with a relative variation only in the amount of 4.22%. Moreover, concerning the X-ray constrained procedures, we can also observe that, in most cases, the relative variations decrease when larger basis sets are used. In fact, the smallest discrepancies are generally detected when the flexible 6-311++G(2d,2p) basis-set is employed. Nevertheless, it is also important to point out that an opposite trend is noticed if we consider the traditional ELMO and RHF methods. Actually, in those situations, when the modest 3-21G basis-set is used, we mainly obtain the smallest changes due to the introduction of the ELMOs. These results were not completely unexpected since similar trends were already observed for the electron densities at the BCPs in Chapter 3 (see Table 3.5).

TABLE 6.2: Unconstrained and X-ray constrained relative variations (U-RV and XC-RV, respectively) of the charge density values at the L-Alanine BCPs.

	U-RV(%)			XC-RV(%)		
BCPs	3-21G	6-311G(d,p)	6-311++G(2d,2p)	3-21G	6-311G(d,p)	6-311++G(2d,2p)
N4-H7	-0.25	0.26	0.40	2.20	-1.55	-0.86
N4-H8	-0.19	0.16	0.33	-0.05	-0.30	-0.05
N4-H9	-0.65	0.29	0.29	3.82	1.99	0.69
N4-C2	0.48	-0.01	0.99	0.66	-0.19	0.23
C2-H10	0.36	1.85	1.50	-0.57	-0.08	-0.55
C2-C3	-0.96	1.56	1.33	-0.38	0.46	0.00
C3-H11	-0.16	0.26	0.39	-0.23	0.14	-0.11
C3-H12	-0.05	0.16	0.49	1.53	0.69	-0.03
C3-H13	0.04	0.23	0.41	1.15	-0.03	-0.36
C2-C1	-0.42	3.54	2.81	-0.56	1.33	0.82
C1-O5	1.95	2.02	1.84	4.22	2.38	2.04
C1-O6	1.73	1.97	1.93	2.74	2.05	1.86

Afterwards, in Table 6.3, where the relative variations of the Laplacian of the electron density at the BCPs are presented, we can observe a larger variability that was actually expected since the Laplacian is a more sensitive descriptor of the charge distributions compared to the values of the electron density at the BCPs. In particular, the largest discrepancies are detected for the carboxylic group where the relative variations vary from

16.90% for the C1-O6 bond with the 3-21G basis-set (unconstrained ELMO) to 93.78% in the case of the C1-O5 BCP using the 6-311G(d,p) basis-set (unconstrained ELMO). Even if our previous studies were performed exploiting transferred extremely localized molecular orbitals, it is worth noting that, in Chapter 3 (see Table 3.6), the differences observed concerning the carboxilic group of the Leu-enkephalin were among the largest. Nevertheless, also for this topological property, we can note tendencies similar to the ones in Table 6.2. In fact, if we consider the values corresponding to the X-ray constrained methods, the lowest relative variations are generally obtained using the 6-311++G(2d,2p) basis-set, whereas in the case of traditional ELMO and RHF calculations, the 6-311G(d,p) and the 6-311++G(2d,2p) basis sets mainly provide a larger variability.

TABLE 6.3: Unconstrained and X-ray constrained relative variations (U-RV and XC-RV, respectively) of the Laplacian values of the charge density at the L-Alanine BCPs.

	U-RV(%)			XC-RV(%)		
BCPs	3-21G	6-311G(d,p)	6-311++G(2d,2p)	3-21G	6-311G(d,p)	6-311++G(2d,2p)
N4-H7	-0.38	0.54	0.50	2.14	0.56	0.97
N4-H8	0.03	0.37	0.38	-1.03	1.52	0.75
N4-H9	-0.94	1.56	0.74	-1.87	2.84	0.44
N4-C2	54.37	-2.26	31.29	9.28	0.21	3.49
C2-H10	0.81	3.88	2.82	0.66	2.00	0.30
C2-C3	0.03	7.88	6.66	0.13	5.50	3.62
C3-H11	-0.68	0.95	1.19	-0.18	1.31	1.04
C3-H12	-0.71	0.09	1.27	1.46	1.52	0.81
C3-H13	-0.45	0.60	1.21	1.62	0.60	-0.60
C2-C1	8.26	14.89	11.01	9.41	8.56	6.58
C1-O5	27.64	93.78	50.44	44.53	29.90	28.79
C1-O6	16.90	47.87	37.20	20.93	18.95	20.44

Afterwards, the net integrated charges determined through the integration of the different electron distributions over the QTAIM atomic basis have been also considered and reported in Table 6.4. In particular, we have compared the global charges for the main functional groups of the system. At first, we have noticed that the global electroneutrality of the L-Alanine is conserved for each considered method. However, both for the unconstrained and the experimental wave function approaches, a quite large variability

has been generally observed. In particular, the main differences appear using the 3-21G basis-set for the $C_\alpha H$ and the CH_3 fragments where the relative differences amount to 25.63% (unconstrained calculations) and 96.04% (constrained calculations), respectively. This is essentially due to the very small charge values that we have obtained. However, as for the two previous topological properties, it is also possible to note that in the case of the X-ray constrained computations, the relative variations are systematically reduced when larger and more flexible basis sets are used.

TABLE 6.4: Unconstrained and X-ray constrained relative variations (U-RV and XC-RV, respectively) of the net QTAIM integrated charges (e) for the main functional groups of the L-Alanine system.

	U-RV(%)			XC-RV(%)		
BCPs	3-21G	6-311G(d,p)	6-311++G(2d,2p)	3-21G	6-311G(d,p)	6-311++G(2d,2p)
NH_3^+	6.01	15.90	14.24	13.67	16.88	13.07
COO^-	11.04	12.26	8.89	15.81	12.29	8.49
$C_\alpha H$	25.63	12.64	8.37	9.37	6.22	4.70
CH_3	2.98	3.20	-3.36	96.04	39.92	4.94

In order to further assess if the more flexible basis sets allow to reduce the influence of the strict localization constraints initially imposed on the electronic structure in the XC-ELMO computations, we have also computed the Root Mean Squared Deviation (RMSD)

$$RMSD = \sqrt{\frac{\sum_1^{n_{BCPs}} (X_{ELMO} - X_{RHF})^2}{n_{BCPs}}} \tag{6.3}$$

and the Mean Absolute Error (MAE)

$$MAE = \frac{1}{n_{BCPs}} \sum_1^{n_{BCPs}} |X_{ELMO} - X_{RHF}| \tag{6.4}$$

of the three topological properties presented above using the RHF and the XC-RHF values as references. For the sake of simplicity, here we have only taken into account the 3-21G, the 6-311G(d,p) and the 6-311++G(2d,2p) basis sets. In Table 6.5, we can clearly see that, considering the X-ray constrained calculations, the RMSD and the MAE of all the examined properties decrease when we exploit larger basis sets with the only exception noticed for the Laplacian of the charge distribution $\nabla^2\rho(\mathbf{r}_b)$ for the 6-311++G(2d,2p) basis-set. For example, if consider the values of the electron density at the BCPs $\rho(\mathbf{r}_b)$,

the Root Mean Square Deviation decrease from 0.0067 e/Å^3 (3-21G) to 0.0045 e/Å^3 (6-311++G(2d,2p)) and the Mean Absolute Error is reduced from 0.0047 e/Å^3 (3-21G) to 0.0023 e/Å^3 (6-311++G(2d,2p)).

Furthermore, the opposite trend is generally observed if we only consider the unconstrained RHF and ELMO methods for the $\rho(\mathbf{r}_b)$ and the $\nabla^2\rho(\mathbf{r}_b)$ at the BCPs. In particular, the RMSD of the Laplacian of the electron distribution $\nabla^2\rho(\mathbf{r}_b)$ decreases from 0.1367 e/Å^5 to 0.0639 e/Å^5 for the 6-311++G(2d,2p) and the 3-21G basis sets, respectively. Nevertheless, in the same context, increasing the flexibility of the basis sets seems to reduce the RMSD and the MAE of the net integrated charges over the atomic basins.

TABLE 6.5: Root Mean Square Deviation and Mean Absolute Error of the net atomic charges (e), the electron density (e/Å^3) and its Laplacian (e/Å^5) at the bond critical points using the RHF and the XC-RHF values as references. For the sake of clarity, all the values are multiplied by 100.

	unconstrained calculations		X-ray constrained calculations	
basis-set / properties	RMSD	MAE	RMSD	MAE
3-21G				
$\rho(\mathbf{r}_b)$	0.28	0.18	0.67	0.47
$\nabla^2\rho(\mathbf{r}_b)$	6.39	3.42	9.67	5.04
$Q(\mathbf{r}_b)$	2.43	2.05	3.34	2.88
6-311G(d,p)				
$\rho(\mathbf{r}_b)$	0.43	0.29	0.43	0.29
$\nabla^2\rho(\mathbf{r}_b)$	10.85	6.32	7.07	4.65
$Q(\mathbf{r}_b)$	1.94	1.42	2.54	2.08
6-311++G(2d,2p)				
$\rho(\mathbf{r}_b)$	0.35	0.23	0.35	0.23
$\nabla^2\rho(\mathbf{r}_b)$	13.67	8.03	9.67	5.22
$Q(\mathbf{r}_b)$	1.52	1.13	2.08	1.76

6.4 Conclusions

In the previous Chapter we have presented the X-ray constrained wave function fitting approach proposed by Jayatilaka and its extension that allows the determination of molecular orbitals strictly localized on small molecular fragments from experimental X-ray diffraction data. However, the introduction of a predefined localization is a very strong empirical constraint in electronic structure calculations and we did not known how the use of the ELMOs could affect the X-ray constrained wave function calculations.

To investigate this problem, we have compared the results obtained by means of X-ray constrained Hartree-Fock computations to results of X-ray constrained ELMO calculations. In particular, we have analysed the statistical agreements with the experimental measurements, the convergence trends and some topological properties of the obtained charge densities. We have discovered that concerning the X-ray constrained calculations, using small basis sets, the *a priori* localization of the molecular orbitals has a strong influence on the final electron distributions. However, if larger and more flexible basis sets are employed, the XC-RHF and the XC-ELMO calculations tend to more similar final results. We can explain this behaviour considering that the flexibility of larger basis-sets allows to increasingly exploit the information provided by the experimental data and, consequently, to gradually mitigate the initial bias represented by the predefined localization. Therefore, when adequate basis-sets are used, the introduction of a localization scheme represents a less severe approximation and it allows us to directly include traditional chemical concepts in the X-ray constrained wave function approach.

Nevertheless, to completely understand if the XC-ELMOs can be advantageously used as electronic building blocks for the reconstruction of wave functions or electron densities of large systems, it is also necessary to investigate if and to what extent the Jayatilaka approach enables to capture electron correlation effects on the molecular charge distributions. The investigations conducted on this subject will be reported in the next chapter.

Chapter 7

Extraction of electron correlation effects from X-ray diffraction data

Résumé

Selon l'approximation cinématique de la diffraction sur un cristal, les facteurs de structure sont des transformées de Fourier de la densité électronique d'une maille cristalline. Ces réflexions obtenues par des expériences de diffraction aux rayons-X doivent donc intrinsèquement contenir des informations précieuses sur les effets de la corrélation et du champ cristallin sur la densité électronique. Dans ce chapitre, nous avons entrepris une étude pour tenter de clarifier à quel point la méthode de Jayatilaka pourrait être capable de capturer les effets de la corrélation sur la densité électronique.

Dans un premier temps, nous avons généré des facteurs de structure de référence associés à des densités électroniques obtenues par la méthode de calcul CCSD (Coupled Cluster with Single and Double excitations) qui permet d'inclure les effets de la corrélation électronique. Ensuite, nous avons déterminé dans quelle mesure la méthode de la fonction d'onde contrainte parvient à retrouver la différence entre des valeurs CCSD et RHF de certaines propriétés topologiques de la densité électronique. Dans un premier temps, nous avons montré que, pour pratiquement toutes les résolutions considérées, le taux de récupération des effets de la corrélation augmente au fur et à mesure que l'on accentue la contrainte des facteurs de structure. De plus, les plus hauts taux de récupération ont été observés lorsque seulement des réflexions à moyenne ou basse résolution ont été utilisées dans les calculs Hartree-Fock contraints. Enfin, nous avons également comparé différents indices de similarité entre les différentes densités électroniques et nous sommes parvenus aux mêmes conclusions que celles obtenues dans le cas des propriétés topologiques.

Ces résultats peuvent s'expliquer en considérant qu'un grand nombre de réflexions obtenues à haute résolution sont déjà très bien décrites au niveau Hartree-Fock. Ces nombreuses réflexions, ou facteurs de structure, réduisent la contribution des réflexions à moyenne ou basse résolution qui sont beaucoup plus affectées par les effets de la corrélation.

7.1 Introduction

The concept of electron correlation is of paramount importance in theoretical chemistry. A proper description of correlation is crucial to obtain accurate pictures of chemical bonding, spectroscopic properties and dynamics. Although the effect of electron correlation on any property is simply defined as the difference between the exact value of the considered property and the value computed using the Hartree-Fock approximation, it is extremely difficult to calculate it accurately. The main theoretical approaches are undoubtedly the post Hartree-Fock or the DFT methods. The former are based on a multi-determinant wave function ansatz that enables a systematic improvement of the Hartree-Fock description of many-electron systems but they generally are characterized by a large computational cost. The latter stem from the Hohenberg & Kohn theorem, which is in principle exact and has the great computational advantage of using only the three-dimensional electron density as basic physical entity instead of the more complicated wave function. However, as it is well known, the exact functional connecting the ground state electron density and the ground state energy remains unknown.

We have seen in Chapter 5 that many-electron systems can also be studied by the X-ray constrained wave function strategy. In particular, we have presented the Jayatilaka method which basically consists in determining the single Slater determinant that minimizes the energy of the system and, that, at the same time, reproduce as much as possible a set of collected structure factors amplitudes. Since, according to the kinematic approximation, the structure factors are Fourier transforms of the unit-cell electron density, they should intrinsically contain information about the effects of the electron correlation and of the crystal field on the molecular charge distributions which are experimentally observed. In this chapter, we present some preliminary results intended to assess to which extent the Jayatilaka approach is actually able to capture the electron correlation effects on the electron densities. This study will allow also to determine if the XC-ELMOs can be advantageously used as electronic building blocks for the reconstruction of wave functions or electron densities of large systems. Furthermore, it could open new perspectives in the development of new density functionals exploiting X-ray diffraction data, which are not generally used for this purpose.

7.2 Methods

7.2.1 General Strategy

At first, it was necessary to figure out which structure factors to consider in order to explicitly study the capability of the X-ray constrained wave function technique in capturing electron correlation effects on the electron density. To achieve our final goal, the experimental structure factors are not suitable since they intrinsically combine electron correlation effects, crystal field effects and experimental errors. The structure factors generated through periodic *ab initio* computations allow us to avoid the experimental errors but they still take into account the crystal field effects in addition to the electron correlation effects. Therefore, as constraints (and references) for our X-ray constrained wave function calculations we have decided to consider structure factors obtained as analytic Fourier transforms of electron distributions resulting from gas-phase CCSD (Coupled Cluster with Single and Double excitations) calculations.

Then, assuming that the effects of electron correlation on any property is simply defined as the difference between the exact value and the value of the considered property calculated at the Hartree-Fock level, our strategy was to determine to what extent the X-ray constrained wave function method is able to recover the difference between the CCSD and the RHF values of a given property. To achieve this goal, we have initially examined some topological properties of the charge distributions obtained at different levels of theory, particularly focusing on the values of the electron density $\rho(\mathbf{r}_b)$ and of its Laplacian $\nabla^2\rho(\mathbf{r}_b)$ at the bond critical points. Moreover, in order to perform more global comparisons and not to only consider single points of the electron distributions, similarity indexes between the different charge densities have been also computed.

7.2.2 Computational details

In our investigations, we have studied six different systems, namely N_2, CN^-, water, benzene, urea and glycine. For the sake of clarity, in the present chapter, only the results obtained for the N_2 system will be discussed.

All the molecular geometries have been optimized at CCSD level using the 6-311++G(2d,2p) basis-set. Afterwards, for each system in exam, we have carried out single point calculations at different level of theory always exploiting the 6-311++G(2d,2p) basis-set:

- Configuration Interaction with Single and Double Excitations (CISD);
- Density Functional Theory using four different functionals: BLYP, B3LYP, VSXC and B1B95;
- Restricted Hartree-Fock (RHF);

We have also performed X-ray constrained Restricted Hartree-Fock (XC-RHF) calculations, but in this case, instead of the traditional experimental structure factor amplitudes, as constraints we have used structure factor amplitudes associated with the CCSD charge densities of the molecules in exam. This allowed us to assess if this method is able to retrieve some electron correlation effects introduced at CCSD level. For all the XC-RHF computations, we have varied the external multiplier λ from 0.0 to 10.0 with a 0.5 step and, since only pure theoretical structure factors have been used as constraints, the treatment of the thermal motion has been completely neglected. Furthermore, the same weight has been assigned to all the considered reflections ($\sigma_{\mathbf{h}} = 1.0$).

The reference theoretical structure factors used as external constraints have been obtained as analytic Fourier transforms of the resulting CCSD/6-311++G(2d,2p) electron density until a resolution of 2.0 Å^{-1} for reciprocal lattice points corresponding to a pseudo-cubic unit-cell. In order to avoid all the possible interactions between neighbour molecules in the constructed pseudo-crystal, we have properly chosen the dimension of the cell ($a = b = c = 10.0$ Å). Furthermore, other than performing X-ray constrained Hartree-Fock calculations with all the obtained structure factors amplitudes, we have also carried out XC-RHF computations with six other subsets of lower maximal resolution (1.5, 1.2, 0.9, 0.7, 0.5, 0.25 Å^{-1}).

7.3 Results and discussion

7.3.1 Comparison between RHF and CCSD structure factors

At first, in order to determine which structure factors are the most affected by the correlation effects, following the procedure described in the previous section, we have computed

the structure factors amplitudes associated with the gas-phase RHF electron densities of the examined molecules. These structure factor amplitudes have been afterwards compared to the corresponding Coupled Cluster values. In particular, we have determined the absolute differences between the RHF and the CCSD structure factors amplitudes and we have afterwards plotted them as a function of the resolution $sin\theta/\lambda$. The obtained results are presented in Figure 7.1, where it is easy to observe that, as expected, the largest part of the considered reflections occurs at high-angle (0.8 Å^{-1} $\leq$ $sin\theta/\lambda$ $\leq$ 2.0 Å^{-1}). Moreover, in these regions, we can note that the differences between the corresponding CCSD and RHF structure factor amplitudes are quite small. On the contrary, the largest discrepancies are detected at low/medium angles ($sin\theta/\lambda < 0.8$ Å^{-1}) but the number of reflections in this range of resolution is considerably reduced. For the sake of completeness, in Table C.1 of Appendix C we have reported the total number of structure factors amplitudes and the number of significant CCSD/RHF discrepancies for each resolution range taken into account. As we will see in the next sections, these trends will have important consequences for the X-ray constrained wave function calculations that we have performed.

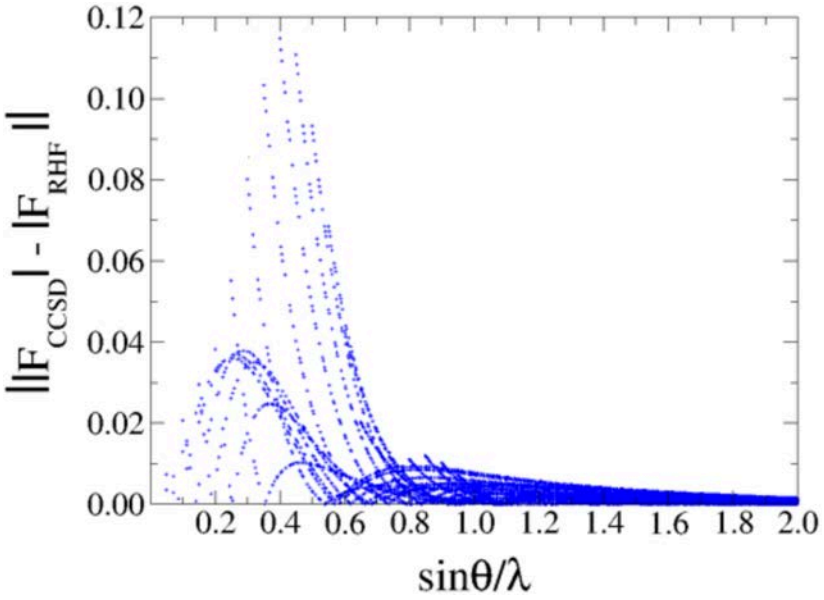

FIGURE 7.1: Absolute differences between the CCSD and the RHF structure factor amplitudes as a function of the reciprocal resolution for the N_2 molecule.

7.3.2 Effects on some topological properties

As already mentioned, the structure factors obtained through the gas-phase *ab initio* CCSD computations have been exploited as constraints in our X-ray constrained Restricted Hartree-Fock computations, using the 6-311++G(2d,2p) basis-set. In particular, we have carried out several XC-RHF calculations considering different sets of reflections characterized by decreasing resolutions (2.0, 1.5, 1.2, 0.9, 0.7, 0.5, 0.25 Å^{-1}). As explained in section 7.2.1, since the CCSD structure factors do not take into account crystal field effects and experimental errors, the XC-RHF computation should only retrieve the effects of the electron correlation on the charge density.

To start assessing the capability of the X-ray constrained wave function methods in capturing the electron correlation effects, we have analysed some topological properties of the charge distributions obtained through various calculations (see section 7.2.2). Moreover, in order to quantify the amount of recovered electron correlation in the different cases, we have considered the following index:

$$\Delta X_{CCSD-RHF} = 100 \times \frac{X_M(\mathbf{r}_b) - X_{RHF}(\mathbf{r}_b)}{X_{CCSD}(\mathbf{r}_b) - X_{RHF}(\mathbf{r}_b)} \tag{7.1}$$

where X is the considered property ($X = \rho(\mathbf{r}_b) \quad \text{and} \quad \nabla^2\rho(\mathbf{r}_b)$) and M is the method under exam.

In Table 7.1, where we have reported the percentages of the CCSD/RHF difference recovered by the different techniques in the case of electron density at the BCP of the N_2 molecule, it is worth noting that the CISD computation has enabled to recover a large part (84.15%) of the electron correlation effects introduced at CCSD level. Concerning the DFT calculations, the obtained results are quite variable, which is probably related to the less systematic nature of the used density functionals. In the case of the X-ray constrained Hartree-Fock method, for almost all the considered levels of resolution, we have observed that the amount of CCSD/RHF differences is increasingly retrieved when larger and larger values of the external multiplier λ are used. In particular, important percentages are recovered only exploiting values of λ that are much higher that the ones usually adopted in XC-RHF computations when experimental data are used. Furthermore, the largest recoveries have been detected using the subset of structure factors amplitudes with resolution $sin\theta/\lambda \leq 0.7\text{Å}^{-1}$ with a maximum value equal to 53.39 % at $\lambda = 10.0$.

TABLE 7.1: Electron density at the bond critical point: percentage of the CCSD/RHF difference recovered by the considered correlated methods for the N_2 system.

method	$(sin\theta/\lambda)_{max}(\text{Å}^{-1})$						
	2.0	1.5	1.2	0.9	0.7	0.5	0.25
CISD	84.15						
DFT/BLYP	101.81						
DFT/B3LYP	77.61						
DFT/VSXC	95.17						
DFT/B1B95	89.00						
XC-RHF (λ=0.5)	0.24	0.59	1.16	2.63	4.88	-0.62	-3.74
XC-RHF (λ=1.0)	0.49	1.17	2.29	5.14	9.35	-0.73	-5.57
XC-RHF (λ=1.5)	0.73	1.75	3.40	7.54	13.48	-0.45	-6.90
XC-RHF (λ=2.0)	0.97	2.32	4.49	9.84	17.29	0.10	-8.02
XC-RHF (λ=2.5)	1.21	2.88	5.56	12.03	20.83	0.82	-9.05
XC-RHF (λ=5.0)	2.39	5.62	10.57	21.74	35.15	5.57	-13.43
XC-RHF (λ=7.5)	3.56	8.22	15.13	29.69	45.54	10.24	-17.02
XC-RHF (λ=10.0)	4.70	10.70	19.27	36.32	53.39	15.09	-20.02

For the sake of completeness, in Table C.2 of Appendix C, we have reported the raw values of the electron density at the bond critical point. It is worth noting that our results are in agreement with previous studies performed by Gatti *et al.* (*198*) and by Boys and Wang (*199*) where it has been shown that the effect of the electron correlation on the electron density $\rho(\mathbf{r}_b)$ at the bond critical point is small ($\approx 1.0\%$). Moreover, as already observed in those studies (*198*, *199*), all the correlated methods (also including the XC-RHF strategy) generally reduce the value of the electron density at the BCP (see Table C.2 of Appendix C).

Concerning the Laplacian of the electron density at the BCP, we have generally observed similar trends to the one for the electron density at the BCP (see Table 7.2). Moreover, in this case, almost all the considered methods have retrieved a larger part of the electron correlation effects introduced with the CCSD calculation. In Table 7.2, we can easily see that the CISD charge distribution captures a large part of the correlation effects (88.30%), while, for the DFT calculations, the situation is again variable with values

going from 92.39 % (DFT/VSXC) to 121.99 % (DFT/BLYP). In the case of the XC-RHF calculations, the trend observed for the $\rho(\mathbf{r}_b)$ is confirmed since more and more important recoveries have been obtained increasing the values of the external multiplier, with the only exception for the resolution $sin\theta/\lambda \leq 0.25\text{Å}^{-1}$. Furthermore, also in the case of the Laplacian, the largest percentages have been detected when low or medium resolution subsets of reflections are used, with a maximum recovery equal to 66.21 % at $\lambda = 10.0$ for the resolution $sin\theta/\lambda \leq 0.7\text{Å}^{-1}$. For the sake of completeness, the values of the Laplacian of the electron density at the BCP have been reported in Table C.3 of Appendix C.

TABLE 7.2: Laplacian of the electron density at the bond critical point: percentage of the CCSD/RHF difference recovered by the considered correlated methods for the N_2 system.

method	$(sin\theta/\lambda)_{max}(\text{Å}^{-1})$						
	2.0	1.5	1.2	0.9	0.7	0.5	0.25
CISD	88.30						
DFT/BLYP	121.99						
DFT/B3LYP	97.71						
DFT/VSXC	92.39						
DFT/B1B95	104.53						
XC-RHF (λ=0.5)	0.35	0.82	1.58	3.67	7.03	1.88	1.41
XC-RHF (λ=1.0)	0.69	1.62	3.11	7.11	13.27	3.36	2.09
XC-RHF (λ=1.5)	1.03	2.42	4.61	10.36	18.85	4.66	2.38
XC-RHF (λ=2.0)	1.37	3.21	6.07	13.42	23.87	5.89	2.45
XC-RHF (λ=2.5)	1.71	3.98	7.49	16.31	28.41	7.05	2.38
XC-RHF (λ=5.0)	3.37	7.71	14.07	28.67	45.86	12.36	1.21
XC-RHF (λ=7.5)	5.00	11.20	19.90	38.34	57.68	17.00	-0.25
XC-RHF (λ=10.0)	6.58	14.49	25.11	46.11	66.21	21.08	-1.65

All the results described above can be easily interpreted reconsidering the expression of the χ^2 statistical agreement (see equation 5.4) presented in chapter 5 and considering Figure 7.1. In fact, in that figure we have seen that in the high-angle region, where the major part of reflections is concentrated, the differences between the corresponding CCSD and RHF structure factor amplitudes are quite small. This means that the high-angle reflections are already very well modelled at Hartree-Fock level. Nevertheless, since the high-angle reflections ($0.8\text{Å}^{-1} < sin\theta/\lambda \leq 2.0\text{Å}^{-1}$) are much more numerous

than the low/medium angle ones (see Table C.1 of Appendix C) ($sin\theta/\lambda \leq 0.8\text{Å}^{-1}$), their weight expression (5.4) becomes predominant when the most complete sets of structure factor amplitudes are considered as external constraints. As a consequence, these reflections cancel the contribution provided by the large CCSD/RHF differences occuring at low/medium angles and, thus, prevent the X-ray constrained method to capture the most important part of effects induced by electron correlation on the electron density. This is the reason why larger recoveries are obtained when only lower-resolution subsets of structure factor amplitudes are used.

Finally, for the sake of completeness, we can also mention that the XC-RHF calculations erroneously provide higher $\rho(\mathbf{r}_b)$ and lower $\nabla^2\rho(\mathbf{r}_b)$ compared to the RHF results when $sin\theta/\lambda \leq 0.25\text{Å}^{-1}$ (see Table 7.1 and 7.2) because, in this case, the selected subset of reflections is insufficient to suitably describe the electron density in the bond critical point region.

7.3.3 Similarity indexes

In order to perform more global comparisons and not to focus only the single BCPs of the systems in exam, similarity indexes between the different electron distributions have been also computed. In particular, always using the CCSD charge distribution as reference, we have carried calculations of the Real-Space (RSR) indices (*123*) which have been already defined in equation (3.3) of chapter 3. It is just worth reminding that the complete similarity between two electron densities is obtained when $RSR = 0$.

We have reported our results in Table 7.3, where, at first, it is easy to note that the CISD charge density is the most similar to CCSD one. Although to a lesser extent than the CISD method, the DFT electron densities seem to capture a significant part of the CCSD correlation effects (see, for example, the hybrid B1B95 functional where the the RSR index amounts to 2.44). Moreover, concerning the XC-RHF charge distributions, we have observed similar trends already highlighted for the topological properties at the bond critical point. In fact, we can see that the similarity with the CCSD reference always increases when larger values of λ are exploited. This trend has been detected for all the considered resolutions and, again, the best agreements have been observed when only low/medium reflections are taken into account. All these results further demonstrate

that the large number of high-angle structure factors, which are already well described at RHF level, considerably reduce the important contribution of the low/medium angle reflections (which are indeed the most affected by the electron correlation effects) in the XC-RHF calculations.

TABLE 7.3: Values of the RSR similarity index between the CCSD electron distribution and the charge densities associated with all the other considered methods. For the sake of clarity, all the RSR values are multiplied by 10.

method	$(sin\theta/\lambda)_{max}(\text{Å}^{-1})$						
	2.0	1.5	1.2	0.9	0.7	0.5	0.25
CISD	1.36						
DFT/BLYP	5.21						
DFT/B3LYP	3.06						
DFT/VSXC	3.93						
DFT/B1B95	2.44						
RHF	7.42						
XC-RHF (λ=0.5)	7.39	7.37	7.32	7.20	7.00	6.78	7.18
XC-RHF (λ=1.0)	7.37	7.32	7.32	7.01	6.65	6.32	7.05
XC-RHF (λ=1.5)	7.35	7.27	7.14	6.82	6.33	5.96	6.95
XC-RHF (λ=2.0)	7.33	7.22	7.05	6.65	6.05	5.66	6.86
XC-RHF (λ=2.5)	7.31	7.17	6.97	6.49	5.81	5.40	6.78
XC-RHF (λ=5.0)	7.20	6.94	6.57	5.80	4.87	4.54	6.45
XC-RHF (λ=7.5)	7.09	6.73	6.23	5.28	4.25	4.01	6.19
XC-RHF (λ=10.0)	6.99	6.53	5.93	4.87	3.80	3.65	5.98

It is also important to precise that, unlike the analysis of the topological properties, in this case, the best agreement with the CCSD charge density have been obtained when $sin\theta/\lambda \leq 0.5\text{Å}^{-1}$. This is probably due to the fact that the similarity indexes provide a more global comparison between two electron densities and, therefore, the lower RSR values observed at 0.5 Å^{-1} possibly results from the deletion of the reflections in the $0.5-0.7$ Å^{-1} range of resolutions, which further increases the weight of the large CCSD/RHF discrepancies observed at $sin\theta/\lambda \leq 0.5$ Å^{-1}. Nevertheless, at the same time, as it has been shown above, the electron density region close to the bond critical point is not very well described since numerous significant reflections are neglected at $sin\theta/\lambda \leq 0.5$ Å^{-1}.

Other than the RSR index, we have also exploited other similarity indicators, such as the Euclidean Carbó distance (*200*, *201*), which has been defined by Carbó and co-workers as an Euclidean distance between two molecular electron distributions $\rho_i(\mathbf{r})$ and $\rho_j(\mathbf{r})$ and which can be expressed like this:

$$D_{ij} = (Z_{ii} + Z_{jj} - 2Z_{ij})^{\frac{1}{2}} \tag{7.2}$$

where the overlap-like similarity Z_{ij} is

$$Z_{ij} = \int \rho_i(\mathbf{r})\rho_j(\mathbf{r})d\mathbf{r} \tag{7.3}$$

In this case, the complete similarity is obviously obtain when $D_{ij} = 0$, while larger values of the index indicate smaller similarities between the two considered electron densities. In Table 7.4, where the values of Carbó Euclidean distances between the reference CCSD and all the electron distributions are shown, we can easily observe similar trends to the case of the RSR index. The CISD charge density is the most similar to CCSD one and the DFT methods are able to intrinsically capture an important part of the CCSD correlation effects on the electron density. Concerning the XC-RHF electron distributions, we have noticed that the similarity with the CCSD reference always grows when larger values of λ are used and, moreover, the best agreements have been again detected when only low/medium angle reflections are taken into account. These observations further show that the large number of high-angle structure factors considerably reduce the important contribution of the low/medium angle reflections in the XC-RHF calculations. Finally, it is worth mentioning that we have observed similar trends in Table C.4 and C.5 of Appendix C, where the RMSDs and MADs between the different electron densities have been respectively reported.

7.3.4 Attachment and detachment densities

In order to qualitatively visualize the effects of the electron density reorganisation due to the electron correlation, we have exploited two tools generally used in theoretical chemistry to visualize and study the rearrangement of the electron distribution after an electron excitation, namely the detachment and the attachment densities (*202*). The detachment

TABLE 7.4: Values of Carbó Euclidean distances between the CCSD electron distribution and the charge densities associated with all the other considered methods. For the sake of clarity, all the values are multiplied by 100.

method	$(sin\theta/\lambda)_{max}(\text{Å}^{-1})$						
	2.0	1.5	1.2	0.9	0.7	0.5	0.25
CISD	0.65						
DFT/BLYP	3.30						
DFT/B3LYP	2.09						
DFT/VSXC	2.91						
DFT/B1B95	1.82						
RHF	4.07						
XC-RHF (λ=0.5)	4.05	4.04	4.01	3.95	3.84	3.72	3.91
XC-RHF (λ=1.0)	4.04	4.01	3.96	3.84	3.65	3.49	3.83
XC-RHF (λ=1.5)	4.02	3.97	3.90	3.74	3.49	3.31	3.78
XC-RHF (λ=2.0)	4.01	3.94	3.85	3.65	3.35	3.18	3.74
XC-RHF (λ=2.5)	3.99	3.91	3.80	3.56	3.23	3.07	3.70
XC-RHF (λ=5.0)	3.92	3.77	3.58	3.21	2.84	2.75	3.59
XC-RHF (λ=7.5)	3.86	3.64	3.38	2.97	2.62	2.58	3.51
XC-RHF (λ=10.0)	3.79	3.52	3.21	2.78	2.50	2.47	3.44

density is that part of the ground state charge distribution which is rearranged as attachment density after an electronic transition. In our investigation, the unconstrained Restricted Hartree-Fock wave function always corresponded to our "pseudo ground-state" (starting state), while all the correlated wave functions (i.e., the CCSD, CISD and XC-RHF wave functions) were our "pseudo excited states" (final states).

In Figure 7.2, where we have represented different isosurfaces of the detachment and attachment densities, we can easily observe that the main effect of the electron correlation is the redistribution of the charge density from the bonding and lone-pair regions to the nuclei and, this trend is obviously more evident at CCSD and CISD level. This result confirms the observations discussed in section 7.3.2, namely the fact that the correlated methods generally reduce the density at the bond critical points. Furthermore, considering the isovalues associated with the surfaces depicted in Figure 7.2, we can note that the attachment and detachment densities associated with the CCSD and CISD calculations provide

electronic reorganizations of approximatively the same order of magnitude. Nevertheless, concerning the XC-RHF method, the extent of rearrangement is smaller when the most complete set of reflections is used ($sin\theta/\lambda < 2.0\text{Å}^{-1}$ and $sin\theta/\lambda < 1.5\text{Å}^{-1}$), while, in agreement with our previous observations, it increases when only low/medium angle structure factor amplitudes are considered ($sin\theta/\lambda < 0.7\text{Å}^{-1}$ and $sin\theta/\lambda < 0.9\text{Å}^{-1}$).

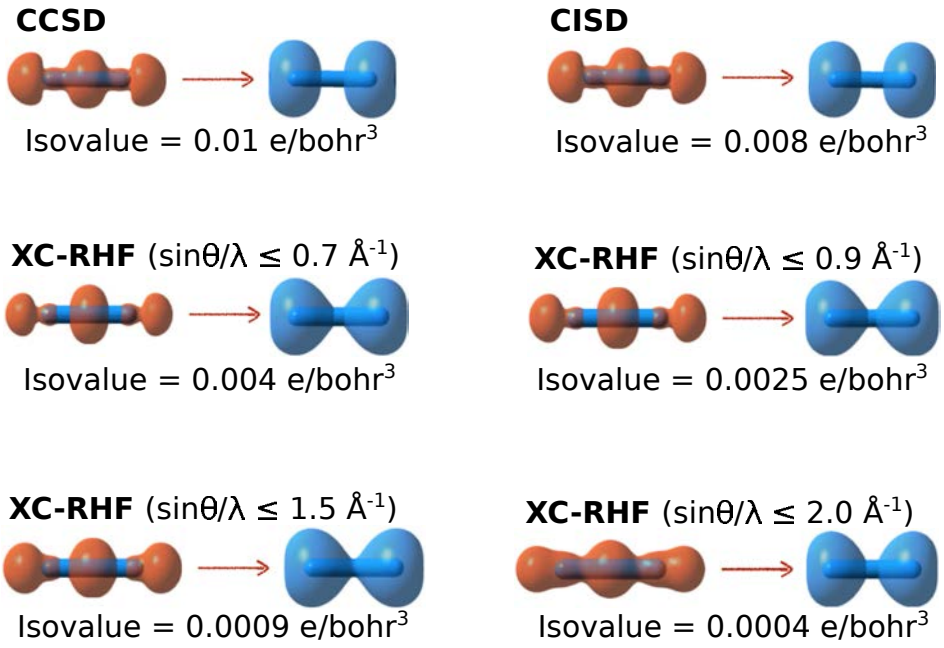

FIGURE 7.2: Representative isosurfaces of the detachment and attachment densities (in orange and blue, respectively) relative to electronic rearrangements with respect to the RHF charge distribution when different correlated methods are taken into account.

7.4 Conclusions

According to the kinematic approximation, the structure factors are Fourier transforms of the unit-cell charge distribution. Therefore, these reflections should intrinsically contain precious information about the effects of the electron correlation and of the crystal field on the electron densities. In this chapter, we have investigated if, and to what extent, the Jayatilaka approach (described in chapter 5) is actually able to intrinsically capture the electron correlation effects on the electron distributions.

To achieve this goal, we have computed reference structure factors from correlated gas-phase CCSD calculations and we have determined to what extent the X-ray constrained wave function method is able to recover the differences between the CCSD and the RHF values of two topological properties. We have discovered that, for almost all the considered levels of resolution, the amount of correlation effects is increasingly retrieved when larger and larger values of the external multiplier λ are used. Moreover, the largest recovery percentages have been detected when only low or medium-resolution subsets of reflections are exploited in the XC-RHF procedure. Then, we have analysed and compared different similarity indexes between the different electron distributions and we have detected the same trends observed in the case of the topological properties.

All these observations can be simply explained considering the large number of high-angle structure factors which are already well described at Hartree-Fock level. Indeed, they cancel the contribution of the low/medium angle reflections which are less numerous but more affected by the electron correlation effects. To overcome this drawback, the low-medium angle reflections should be properly weighted in the expression of the Jayatilaka functional that must be minimized. Furthermore, since the external multipliers λ used in our study were much greater than the ones usually considered when experimental data are exploited, it will be necessary to clarify the real meaning of this external adjustable parameter. As a consequence, we should also further investigate the meaning of X-ray constrained wave functions obtained when the contribution of the experimental data becomes predominant compared to electron energy of the system in exam.

Such studies will be crucial to completely understand if XC-ELMOs can be really exploited as "correlated" electronic building blocks for the reconstruction of wave functions or electron densities of large systems. Finally, it is worth mentioning that, in order to completely assess the capabilities of the Jayatilaka approach, we envisage an analogous study that aims at understanding whether of not this method is also able to capture the crystal field effects on molecular charge distributions.

Chapter 8

A theoretical study of the Biscarbonyl[14]annulene charge density

Résumé

Dans ce dernier chapitre, nous avons effectué une étude théorique de la densité électronique du syn-1,6:8,13-Biscarbonyl[14] annulène (BCA). Des mesures précises de diffraction aux rayons-X ont suggéré une rupture partielle du caractère aromatique de ce composé lorsque la pression augmente. Pour étudier ce phénomène, nous avons employé dans un premier temps la méthode ELMO traditionnelle et, dans un second temps, la nouvelle approche ELMO Valence Bond (ELMO-VB) avec et sans contraintes expérimentales. Dans ce cas, nous avons tenu compte de deux formes de résonance de l'annulène étudié.

Grâce à la méthode ELMO traditionnelle, nous avons démontré que le caractère aromatique de notre système est renforcé à pression ambiante. De plus, l'approche "ELMO-VB" a montré qu'une forme de résonance est favorisée par rapport à l'autre à haute pression et que, par conséquent, le caractère aromatique du BCA diminue à haute pression. D'un point de vue méthodologique, cette étude a confirmé que lorsque des données expérimentales fiables peuvent être exploitées, les techniques de la fonction d'onde contrainte peuvent être considérées comme des compléments intéressants à d'autres méthodes bien établies dans le domaine des analyses de la densité électronique dans les cristaux.

8.1 The context of the study

Aromaticity is considered as one of the most important concepts in physical organic chemistry. In fact, Kekulé was the first to introduce this concept studying benzene in 1865. One year later, Erlenmeyer proposed to design as aromatic all the compounds having a chemical reactivity similar to the benzene. Then, the 19th-century perception of oscillation between single and double bonds in benzene has been replaced by the concept of resonance between canonical structures. Hückel suggested his famous rule in 1931 stating that planar monocyclic systems with [4n+2] π-electrons are more stable than those with 4n π-electrons (*203*).

Nowadays, it is well known that aromaticity, other than affecting the structure and the reactivity of molecular systems, also influences the response of the molecule to external magnetic fields, for instance modifying the nuclear resonant frequency. However, even if, over the years, chemists have proposed several ideas based on structural, energetic or magnetic parameters (*204*–*207*), there are no universal and general criteria to quantify the aromaticity. For instance, to study the aromaticity for a class of compounds, a widely used strategy consists in varying the most relevant parameters (e.g., energetic, magnetic or structural parameters) and comparing the results obtained for the different molecular species taken into account. Another approach to investigate the aromatic character of a molecule consists in continuously modifying the molecular geometry through compression.

To reduce the aromaticity of a compound, one needs a considerable external energy in order to stabilize one of the electronic configurations over the others, thus breaking the resonance. To accomplish this task, heat, light or electrochemical potential are usually employed. However, an alternative source of external energy is represented by pressure, which is able to perturb the structure of the molecular systems at ambient conditions, in favour of otherwise inaccessible configurations.

In this context, in order to obtain more information on the aromaticity of compounds under pressure (e.g., structural, electronic and energetic changes), one relevant method is represented by the single crystal X-ray diffraction which can map the electron density distribution of a molecular crystal. However, although the electron density mapping from X-ray diffraction is nowadays a well-established technique for studies at ambient

pressure, the determinations of experimental charge distributions for molecular crystals at high pressure are not usually obtained. Some studies have reported electron density maps of simple inorganic compounds or pure elements, mainly determined through the maximum entropy method (*208*, *209*). Furthermore, models using multipoles restricted to theoretical values have been also tested against X-ray diffraction data for the propionamide and the piperazinium hydrogen oxalate systems. However, to the best of our knowledge, the only accurate charge density study conducted so far is the performed by Macchi and co-workers (*210*), who, exploiting X-ray diffraction measurements, have shown a partial rupture of the aromaticity of the *syn*-1,6:8,13-biscarbonyl[14]annulene (BCA) at high pressure.

In fact, the main drawback for exploiting X-ray diffraction at high-pressure is the pressure apparatus that drastically reduces the resolution, the completeness and the quality of the available data. To overcome this problems, one should combine high-pressure (that enables, even at ambient temperature, to significantly reduce the thermal motion) with modern synchroton sources and careful experimental strategies. Another complementary approach for investigating the high pressure forms of molecules in crystals is represented by theoretical calculations, for instance at density functional theory level with periodic boundary conditions (*211*) or, as we will show, exploiting X-ray constrained wave function methods (see chapter 6).

In this chapter, we will report the theoretical study of the charge density of the biscarbonyl[14]annulene recently investigated experimentally by Macchi et al. (*210*). In particular, exploiting the extremely localized molecular orbitals and X-ray constrained wave functions techniques, we will investigate how pressure can induce modifications on the electronic structure and, consequently, how it can reduce the aromatic character of the considered compound.

8.2 The Biscarbonyl[14]annulene

We have mainly focused our investigations on a doubly bridged annulene, namely the *syn*-1,6:8,13-Biscarbonyl[14] annulene (BCA), represented in Figure 8.1, for which high-quality crystals are available. Initially, the BCA has been studied through very detailed electron density analyses at ambient-pressure and very low temperature (19K)(*197*, *212*,

213).

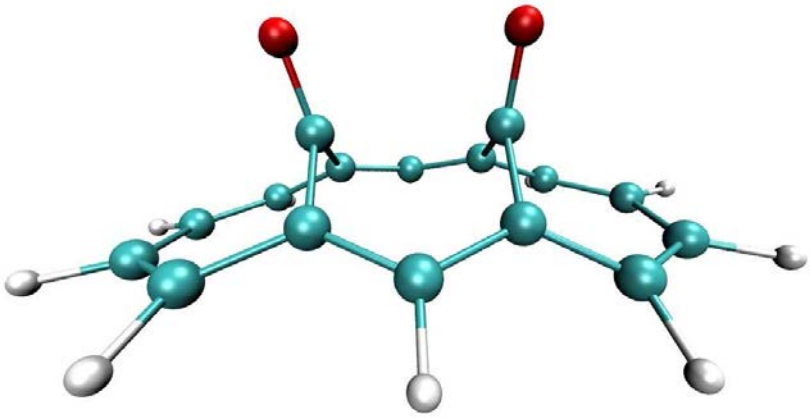

FIGURE 8.1: *syn*-1,6:8,13-Biscarbonyl[14] annulene (BCA) molecular geometry obtained from the X-ray diffraction experiment at 7.7 GPa.

More recently, single crystals of the BCA have been also analysed with multi-pressure X-ray diffraction experiments by Macchi and co-workers in order to determine experimentally the structural changes associated with different pressure conditions (210). Concerning the high pressure experiments, the most extensive data have been collected at 7.7 GPa. Therefore, for our theoretical study we will exploit both the experimental measurements obtained by Destro at ambient-pressure and the experimental measurements collected by Macchi at 7.7 GPa, which will allow us to examine how the aromatic character of the BCA is reduced at high pressure.

In fact, the Biscarbonyl[14] annulene, which consists in a ring of 14 carbon atoms, is a $4n$+2 Hückel system and, therefore, it is potentially aromatic. Nevertheless, in Figure 8.1 it is possible to note that the two carbonyl bridges disturb the planarity of the ring, decreasing the conjugation between the carbon-carbon bonds and reducing the aromatic character of the system. The C-C bond lengths can be considered as very important indicators to qualitatively determine the aromaticity of the BCA since, in principle, in an aromatic system, all the ring skeleton bonds should be equivalent (shorter than single bonds and longer that double bonds). In the case of the BCA, the X-ray diffraction measurements have shown that, at ambient-pressure, some carbon-carbon bonds are clearly identical, whereas at high pressure, the same C-C bonds are no longer equivalent (see Fig 8.2).

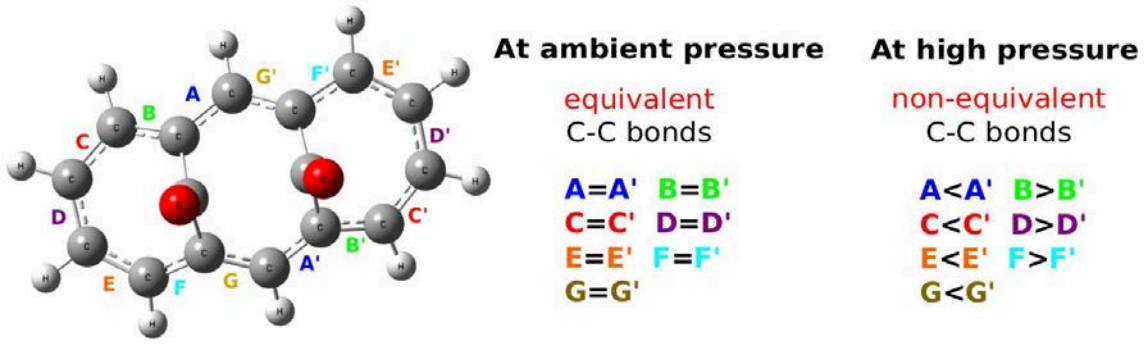

FIGURE 8.2: Schematic representation of the BCA and its C-C equivalent bond lengths at ambient-pressure and their differences at high pressure.

8.3 Methodology

In order to investigate how the aromatic character of the Biscarbonyl[14] annulene is reduced by pressure, we have carried out theoretical calculations including traditional ELMO computations and both unconstrained and X-ray constrained "ELMO-VB" calculations (for the description of the techniques see section 5.6). To accomplish this task, we have exploited the crystallographic structures of the BCA determined at ambient and at high pressures and, moreover, we have also considered the two resonance structures of the annulene system depicted in Figure 8.3.

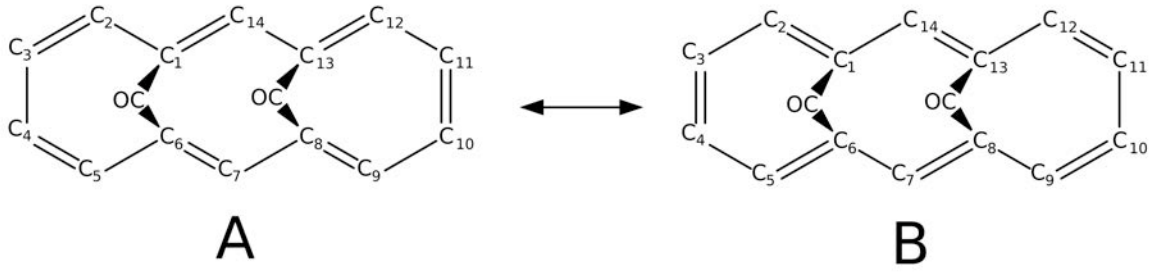

FIGURE 8.3: Resonance structures of the *syn*-1,6:8,13-Biscarbonyl[14] annulene.

For each crystallographic structure, we have performed three different ELMO calculations taking into account three different localization schemes for the π-electron system of the 14-carbon atom ring of the molecule.

- a localization scheme with a unique completely delocalized fragment (see scheme D in Figure 8.4);
- a localization pattern with the fragments for the π electrons corresponding to the carbon-carbon bonds C14-C1, C2-C3, C4-C5, C6-C7, C8-C9, C10-C11 and C12-C13 (see scheme A in Figure 8.4) and practically corresponding to the resonance structure A of the BCA depicted in Figure 8.3.
- a localization scheme with the fragments for the π electrons matching to the bonds C1-C2, C3-C4, C5-C6, C7-C8, C9-C10, C11-C12, C13-C14 (see scheme B in Figure 8.4) and practically corresponding to the resonance structure B of the BCA represented in Figure 8.3.

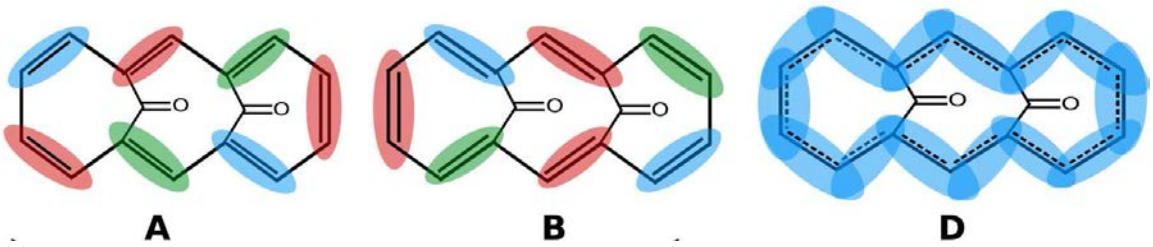

FIGURE 8.4: Schematic representation of the three different localization schemes adopted for the ELMO calculations. The ELMO configurations A and B correspond to the two resonance structures in Figure 8.3 (7 strictly localized MOs for 14 π electrons) and D is the delocalized configuration (7 delocalized MOs on the 14 carbon atoms of the ring).

It is worth mentioning that, for the remaining part of the BCA system (e.g., core, lone-pairs, σ and $\pi_{C=O}$ electrons), all the three localization patterns follow the traditional Lewis structure. As already mentioned above, unconstrained and X-ray constrained "ELMO-VB" computations have been also performed, but more details about these calculations will be given in one of following sections (section 8.4.2).

Finally, in order to further study the partial rupture of the aromatic character of the BCA at high pressure, we have determined the Wiberg bond indices (*214*) associated with the 14 carbon-carbon bonds of the annulene. To achieve this goal, we have performed CCSD (Couple Cluster with Single and Double Excitations) and XC-RHF calculations exploiting the crystallographic structures determined at ambient-pressure and at 7.7 GPa. Then, the obtained CCSD and XC-RHF wave functions have been subject to Natural Bond Orbital (NBO) analyses, from which the Wiberg bond indices have been determined.

For the sake of completeness, we specify that all our calculations have been carried out using the 6-31G, 6-31G(d,p) and cc-pVDZ basis sets.

8.4 Results and discussion

8.4.1 Unconstrained ELMO calculations

At first, unconstrained ELMO calculations have been performed for each localization scheme described in the previous section. Then, the corresponding ELMOs energies, which were obtained with the ambient-and high-pressure structures, have been compared computing the following energy differences:

$$\Delta E_X = E^{HP}_{ELMO_X} - E^{AP}_{ELMO_X} \tag{8.1}$$

where X indicates the considered localization schemes represented in Figure 8.4 (X = D, A or B) and where HP and AP stand for High-Pressure and Ambient-Pressure, respectively. The energy differences have been reported in Table 8.1 and, for the sake of clarity, they have been also depicted in a qualitative schematic diagram in Figure 8.5. Since the energy difference ΔE_D is positive, we can initially deduct that a completely delocalized electronic structure is energetically more stable at ambient-pressure than at 7.7 GPa. Moreover, according to this first result, we have also noted that the two electronic configurations represented by the A and B localization patterns provide contrasting tendencies when pressure increases. Actually, configuration A is more stable at high-pressure since the variation ΔE_A is negative, whereas the configuration B becomes significantly less stable at high-pressure.

All these results clearly demonstrate that the two resonance structures presented in Figure 8.3 do not equally contribute to the electronic description of the Biscarbonyl[14] annulene at high-pressure.

In order to further confirm our observations, we have analysed the differences at ambient and high-pressures between the ELMO energies corresponding to the localization schemes A and B:

$$\Delta E^X_{A-B} = E^X_{ELMO_A} - E^X_{ELMO_B} \tag{8.2}$$

TABLE 8.1: Differences (in kcal/mol) between the ELMO energies at high-pressure (7.7 GPa) and at ambient-pressure for the three localization schemes taken into account.

energy differences	6-31G	6-31G(d,p)	cc-pVDZ
ΔE_D	10.32	10.93	13.40
ΔE_A	-9.34	-8.75	-5.79
ΔE_B	37.14	38.98	41.87

where, in this case, X indicates the pressure (AP and HP).

TABLE 8.2: Differences (in kcal/mol) between the ELMO energies obtained with the localization schemes A and B both at ambient and high (7.7 GPa) pressures.

energy differences	6-31G	6-31G(d,p)	cc-pVDZ
ΔE^{AP}_{A-B}	-0.90	-1.03	-1.03
ΔE^{HP}_{A-B}	-47.28	-48.76	-48.69

In Table 8.2, where the differences between the ELMO energies obtained with the the localization schemes A and B are presented, we can immediately observe that the two configurations A and B of the BCA are almost equivalent from an energetic point of view at ambient-pressure (see also Figure 8.5). This means that the contributions of the two resonance structures are almost identical at that pressure. Nevertheless, the very important variations at high-pressure show that the configuration A is significantly more stable and, therefore, one resonance structure (the resonance structure A) is clearly preferred. These results further confirm that pressure directly affects the contributions of the two resonance structures of the Biscarbonyl[14] annulene and, probably, that the aromatic character of this system is partially reduced at high-pressure.

8.4.2 Unconstrained and X-ray constrained "ELMO-VB" calculations.

At a later stage, in order to further evaluate the weight of the resonance structures A and B (see Figure 8.3) in the description of the 14-electron Hückel system of the BCA at ambient and high-pressures, we have exploited a prototype of ELMO-based Valence Bond

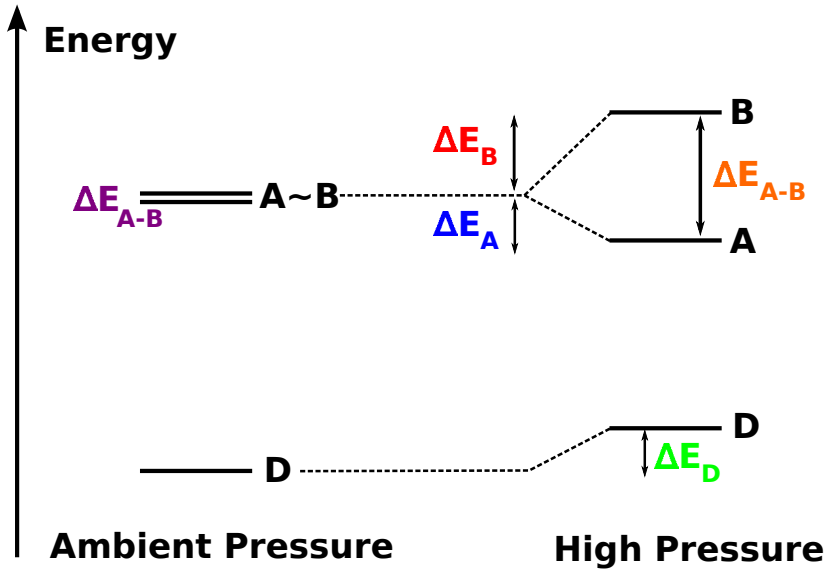

FIGURE 8.5: Diagram that qualitatively show the unconstrained ELMO energies (and the associated differences) at ambient and high pressures.

(VB) method which has been presented in section 5.6. This novel technique has been afterwards extended to the X-ray constrained wave function approach to also take into account the structure factor amplitudes collected experimentally. Following the "ELMO-VB" philosophy, the global wave function of the Biscarbonyl[14] annulene can be expressed like this:

$$|\Psi_{ELMO-VB}\rangle = C_A|\Psi_A\rangle + C_B|\Psi_B\rangle \tag{8.3}$$

where Ψ_A and Ψ_B are normalized Slater determinants constructed with ELMOs previously obtained adopting the localization schemes A and B (see Figure 8.4), respectively, and, therefore, describing the resonance structures A and B of the BCA.

As already explained in chapter 6, due to the non-orthogonality of the wave function $|\Psi_A\rangle$ and $|\Psi_B\rangle$, the real weights associated with the two resonance structures A and B are given by the Chirgwin-Coulson coefficients that in this case are simply given by:

$$K_A = |C_A|^2 + C_A\, C_B\, S_{AB} \tag{8.4}$$

and

$$K_B = |C_B|^2 + C_A\, C_B\, S_{AB} \tag{8.5}$$

where $S_{AB} = \langle\Psi_A|\Psi_B\rangle$ is the overlap between $|\Psi_A\rangle$ and $|\Psi_B\rangle$.

Furthermore, as already explained in chapter 6, for the X-ray constrained Valence Bond calculations, the external multiplier λ as been iteratively adjusted until the fulfilment of the following criterion:

$$|K_{A/B}^{\lambda_i} - K_{A/B}^{\lambda_{i-1}}| \leq 2.0 \times 10^{-4} \tag{8.6}$$

Also for these computations, the 6-31G, 6-31G(d,p) and cc-pVDZ basis sets have been employed.

In Table 8.3, we have reported the results obtained through our unconstrained "ELMO-VB" calculations. It is easy to note that, for all the basis sets, at ambient-pressure, the resonance structure A is slightly more important than the resonance structure B. However, at high-pressure, the conformation A becomes largely predominant compared to the B one. Actually, it is very important to precise that the unconstrained "ELMO-VB" computations consider only the energetic aspects and, moreover, the ELMOs used to construct the Slater determinants Ψ_A and Ψ_B are the ones obtained by means of traditional ELMO calculations. In particular, as already explained, they are simply kept frozen during the optimization procedure and only the coefficients associated to Ψ_A and Ψ_B are determined. Therefore, the large difference in the Chirgwin-Coulson coefficients detected at high-pressure with the unconstrained "ELMO-VB" method reflects the results of the simple ELMO calculations (see subsection 8.4.1) according to which, at high-pressure, structure A was about 48 kcal/mol more stable than structure B.

TABLE 8.3: Statistical agreement χ^2 and Chirgwin-Coulson coefficients obtained through unconstrained "ELMO-VB" calculations at ambient and high pressures.

	Ambient-Pressure			High-Pressure		
basis-set	χ^2	K_A	K_B	χ^2	K_A	K_B
6-31G	5.88	0.55	0.45	6.31	0.99	0.01
6-31G(d,p)	5.14	0.56	0.44	5.64	0.99	0.01
cc-pVDZ	4.97	0.56	0.44	5.68	0.99	0.01

Then, when experimental structure factor amplitudes have been used as constraints in the X-ray constrained "ELMO-VB" calculations, in the optimization of the weights associated with the two resonance structures A and B we have been able to introduce aspects related to the electron density. In Table 8.4, where the results of these computations are

reported, we can immediately notice that, at ambient-pressure, the Chirgwin-Coulson coefficients are almost equivalent. This means that Ψ_A and Ψ_B equally contribute to the description of the electronic structure of the BCA and, therefore, that at ambient-pressure the annulene under exam is essentially an aromatic compounds. On the contrary, at 7.7 GPa, the Chirgwin-Coulson coefficients K_A are systematically slightly higher that the K_B ones. These results indicate that the resonance structure A is preferred and, therefore, that the aromatic character of the biscarbonyl[14] annulene decreases at high pressure.

TABLE 8.4: Statistical agreement χ^2 and Chirgwin-Coulson coefficients obtained through X-ray constrained "ELMO-VB" calculations at ambient and high pressures.

	Ambient Pressure				High Pressure			
basis-set	λ	χ^2	K_A	K_B	λ	χ^2	K_A	K_B
6-31G	0.11	5.88	0.50	0.50	1.44	6.18	0.56	0.44
6-31G(d,p)	0.09	5.12	0.49	0.51	1.09	5.40	0.56	0.44
cc-pVDZ	0.08	4.93	0.49	0.51	0.10	5.43	0.55	0.45

For the sake of completeness, the variations of the statistical agreement χ^2 (see Figure D.1 in Appendix D) and of the Chirgwin-Coulson coefficients (see Figure 8.6) in function of the external parameter λ have been reported for all the X-ray constrained "ELMO-VB" calculations. In Figure 8.6 it is particularly evident that, when the calculations are carried out at ambient pressure the two Chirgwin-Coulson coefficients converge to the same limit for all the basis sets, ($K_A \approx K_B \approx 0.5$). An opposite tendency is observed at high pressure where the two coefficients converge to two different limits ($K_A \approx 0.55$ and $K_B \approx 0.45$).

Finally, it is also worth mentioning that the introduction of the structure factor amplitudes as constraints decreases the large difference between the Chirgwin-Coulson coefficients $K_A$andK_B obtained through unconstrained "ELMO-VB" computations, for which only the molecular geometries are considered to take into account the modifications induced by pressure. Obviously, this trend is even more relevant when larger values of the external multiplier λ are used because the contribution of the experimental measurements becomes more important. These results also underline the importance of using experimental data related to the electron density as external constraints in the X-ray constrained ELMO Valence Bond approach.

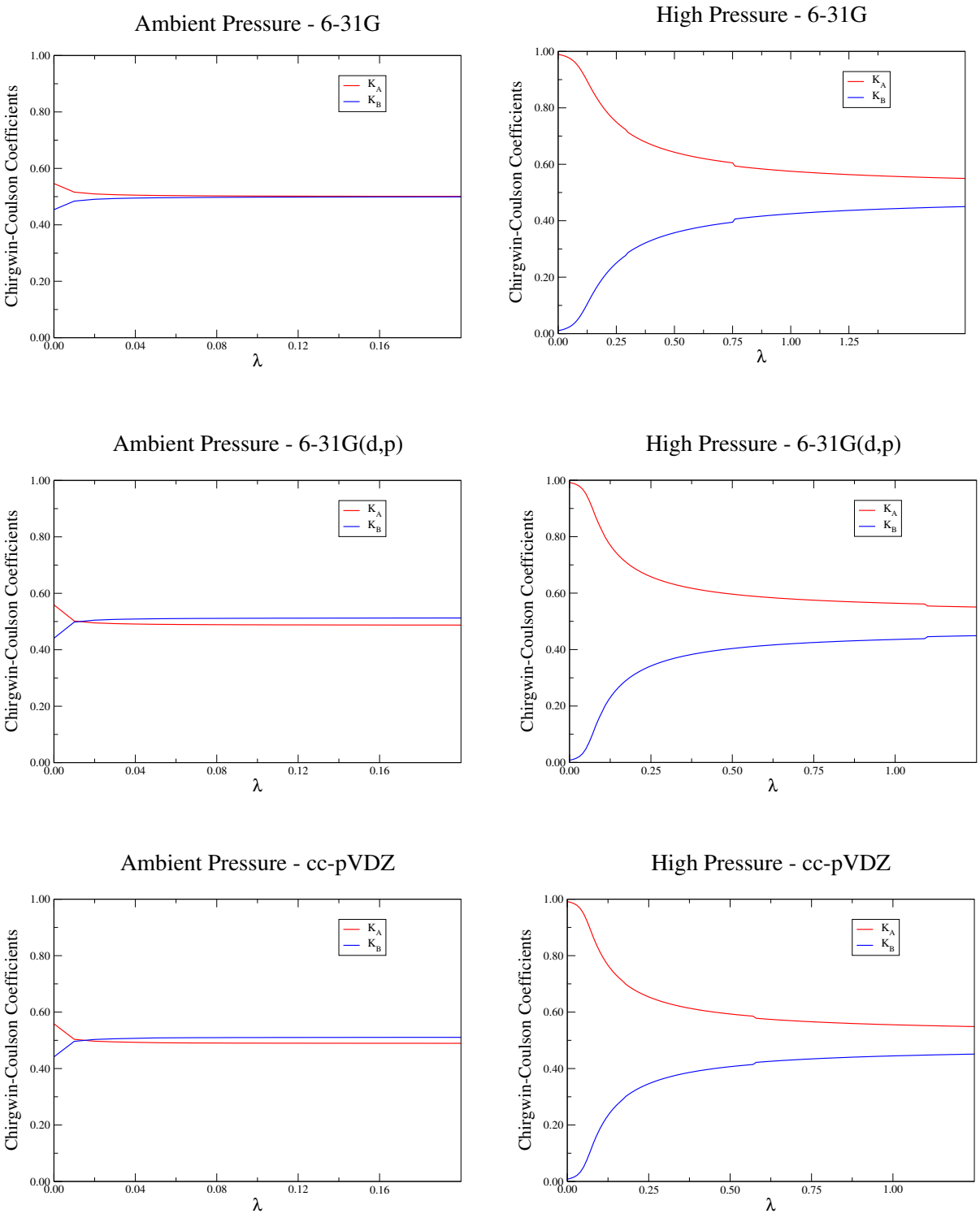

FIGURE 8.6: Variation of the Chirgwin-Coulson coefficients in function of the external multiplier λ for the X-ray constrained "ELMO-VB" calculations performed at ambient and high pressures.

8.4.3 Determination of the Wiberg Bond Index

Finally, to further investigate the effects of high-pressure on the aromaticity of the BCA, we have determined the Wiberg bond indices associated with the 14 carbon-carbon bonds of the Biscarbonyl[14] annulene. As already described, to accomplish this task, we have performed CCSD and XC-RHF calculations using the crystallographic structures determined at ambient and high pressures. Afterwards, we have performed Natural Bond Orbital (NBO) analyses considering the obtained CCSD and XC-RHF wave functions to get the Wiberg bond indices.

In Table 8.5 and 8.6 we have reported the Wiberg bond indices obtained from CCSD and XC-RHF calculations with the cc-pVDZ basis-set. Both in Table 8.5 and 8.6 we can observe that, at ambient pressure, the bonds C1-C2, C2-C3, C3-C4, C4-C5, C6-C7 and C7-C8 are practically equivalent to their pseudo-symmetry related C-C bonds, which are C8-C9, C9-C10, C10-C11, C11-C12, C12-C13, C13-C14 and C14-C1, respectively. These results confirm that the two resonance structures presented in Figure 8.3 contribute equivalently in the description of the molecular electronic structure of the BCA at ambient-pressure.

On the contrary, at high-pressure, both the CCSD and the XC-RHF calculations display Wiberg bond indices at variance with the situation observed at ambient-pressure. In fact, the pseudo-symmetry of the C-C bonds detected at ambient pressure completely disappears since the bonds C1-C2, C2-C3, C3-C4, C4-C5, C6-C7 and C7-C8 are no longer equivalent to their corresponding pseudo-symmetry related carbon-carbon bonds (C8-C9, C9-C10, C10-C11, C11-C12, C12-C13, C13-C14 and C14-C1, respectively).

Furthermore, it is also easy to note that the Wiberg bond indices of the C8-C9, C2-C3, C10-C11, C4-C5, C12-C13, C6-C7 and C14-C1 bonds increase, while the indices of their corresponding pseudo-symmetric bonds (C1-C2, C9-C10, C3-C4, C11-C12, C5-C6, C13-C14 and C7-C8 respectively) decrease. This means that the former accentuate their double bond character, whereas the latter tend to a single carbon-carbon nature. These results prove that, at high pressure, resonance structure A, which is depicted by the localization scheme A in Figure 8.4 (where the double bonds are localized on the bonds C8-C9, C2-C3, C10-C11, C4-C5, C12-C13, C6-C7 and C14-C1) is more favourable than resonance structure B. We can conclude that both the vanishing of the pseudo-symmetry and the predominance of one resonance structure over the other further confirm the reduction of the aromaticity of the Biscarbonyl[14] annulene at high-pressure.

TABLE 8.5: Wiberg indices for the C-C bonds of the BCA molecule at ambient pressure and at 7.7 GPa. The indices are obtained from CCSD calculations with the cc-pVDZ basis-set.

bonds	Ambient Pressure		High Pressure	
C1-C2 / C8-C9	1.30	1.30	1.22	1.42
C2-C3 / C9-C10	1.53	1.52	1.61	1.39
C3-C4 / C10-C11	1.32	1.33	1.24	1.46
C4-C5 / C11-C12	1.52	1.52	1.63	1.39
C5-C6 / C12-C13	1.30	1.30	1.19	1.42
C6-C7 / C13-C14	1.40	1.40	1.53	1.29
C7-C8 / C14-C1	1.40	1.41	1.28	1.51

Finally, considering our results at high pressure, it is important to note that we have detected a tendency similar to the one already observed in the previous subsection. In fact, for each couple of pseudo-symmetric bonds, the high-pressure Wiberg bond indices obtained through X-ray constrained Hartree Fock computations (see Table 8.6) are characterized by a reduced discrepancy compared to the corresponding results obtained from gas phase CCSD calculations (see Table 8.5). We can suppose that the introduction of the structure factor amplitudes as constraints attenuate the large differences of the indices observed at CCSD level. Moreover, similar trends have also been observed for the 6-31G and the 6-31G(d,p) basis sets, for which the results are given in Tables D.1-D.2 and D.3-D.4 of the Appendix D.

TABLE 8.6: Wiberg indices for the C-C bonds of the BCA molecule at ambient pressure and at 7.7 GPa. The indices are obtained from XC-RHF calculations with the cc-pVDZ basis-set.

bonds	Ambient Pressure		High Pressure	
C1-C2 / C8-C9	1.27	1.29	1.24	1.37
C2-C3 / C9-C10	1.54	1.53	1.58	1.42
C3-C4 / C10-C11	1.30	1.31	1.26	1.41
C4-C5 / C11-C12	1.54	1.53	1.59	1.42
C5-C6 / C12-C13	1.28	1.29	1.23	1.37
C6-C7 / C13-C14	1.41	1.39	1.45	1.34
C7-C8 / C14-C1	1.39	1.43	1.33	1.44

8.5 Conclusions

In this chapter we have reported theoretical investigations on the charge density of a doubly bridged annulene, namely the *syn*-1,6:8,13-Biscarbonyl[14] annulene. The X-ray diffraction measurements on this compound have shown that its aromatic character is reduced when pressure increases. In order to study this phenomenon, we have initially exploited the unconstrained ELMO technique and, at a second stage, the novel unconstrained and X-ray constrained ELMO Valence Bond approach. In particular, in both cases, we have mainly taken into account the two resonance structures of the annulene system in exam. Finally, in order to further study the partial rupture of the aromatic character of the BCA at high pressure, performing CCSD and traditional XC-RHF calculations we have also determined the Wiberg bond indices associated with the 14 carbon-carbon bonds of the annulene.

At first, the unconstrained ELMO computations have demonstrated that a completely delocalized electronic structure is energetically more stable at ambient-pressure than at 7.7 GPa. Moreover, we have shown that at ambient pressure the two configurations A and B of the BCA are almost equivalent from an energetic point of view and, consequently, that the contributions of the two resonance structures are almost identical. On the contrary, at high-pressure, we have detected that one configuration is significantly

more stable than the other one. These results have confirmed that pressure directly affect the contributions of the two resonance structures of the Biscarbonyl[14] annulene and that the aromatic character of this system is partially reduced at high-pressure.

At a later stage, by means of unconstrained and X-ray constrained "ELMO-VB" calculations, we have shown that, at ambient-pressure, the two configurations equally contribute to the description of the electronic structure of the BCA and, therefore, that the annulene is essentially an aromatic compound. At 7.7 GPa, the Chirgwin-Coulson coefficient K_A is systematically slightly higher that the K_B one and, again, these results indicate that the aromatic character of the Biscarbonyl[14] annulene decreases at high pressure. Furthermore, through Wiberg bond index analyses, we have demonstrated that, at ambient-pressure, a set of seven C-C bonds are practically equivalent to the corresponding pseudo-symmetric ones, while at high-pressure this pseudo-symmetry completely disappear. While the former result has further confirmed that the two resonance structures contribute equivalently in the description of the molecular electronic structure of the BCA at ambient-pressure, the latter outcome, highlighted again the reduction of the aromaticity of the Biscarbonyl[14] annulene at high-pressure.

Finally, analysing the results of the X-ray constrained "ELMO-VB" calculations and the Wiberg bond indices obtained through XC-RHF computations, we have noticed that when experimental data are taken into account, the large differences between the resonance structures weights reduce. These results underline the importance and the usefulness of using experimental X-ray diffraction data as external constraints in wave function-based approaches for charge density studies.

In conclusion, the results of the present study have shown that both unconstrained ELMO, unconstrained "ELMO-VB" and X-ray constrained wave function methods highlight the partial rupture of the aromaticity of the Biscarbonyl[14] annulene. From the methodological point of view this study has also confirmed that when reliable experimental data are available, X-ray constrained wave function techniques can be considered as a valid alternative/complement to other well-established methods to determine and analyse electron distribution in molecular crystals. Furthermore, the good performances of the "XC-ELMO-VB" method encourages us to develop new types of XCWF strategies based on a multideterminant ansatz for the wave functions.

Conclusions of Part II, future directions

In the second part of this manuscript, we have mainly presented our investigations on the extremely localized molecular orbitals in the framework of the X-ray constrained wave function methods. At first, in Chapter 5, we have introduced the X-ray constrained wave function technique initially developed by Jayatilaka and its recent extension proposed by Genoni to the Extremely Localized Molecular Orbitals theory. Moreover, we have also briefly introduced a new X-ray constrained ELMO-based Valence Bond strategy. Then, since in the first part of this manuscript we have demonstrated that the ELMOs are easily transferable orbitals, we can also imagine the construction of databases exploiting X-ray constrained extremely localized molecular orbitals. Nevertheless, we did not know if these "experimental" ELMOs were more suitable for this purpose compared to the "theoretical" ones. In order to answer to this question we have evaluated both the effects of using an *a priori* localization procedure in the X-ray constrained wave function approach and the intrinsic capability of the Jayatilaka method in capturing the electron correlation effects on the electron density.

In Chapter 6, where we have studied the effects of the *a priori* localization on the X-Ray constrained wave functions, we have discovered that, using small basis sets, the *a priori* localization of the molecular orbitals has a strong influence on the results of the calculations. However, if larger and more flexible basis sets are exploited, the XC-RHF and the XC-ELMO calculations tend to similar final results. We have explained this behaviour considering that the flexibility of larger basis-sets enables to capture more and more information provided by the experimental data and, consequently, to gradually reduce the initial bias represented by the predefined localization.

Then, in order to completely assess if the XC-ELMOs can be advantageously used as electronic building blocks for the reconstruction of wave functions or electron densities of large systems, we have investigated if and to what extent the Jayatilaka approach allows us to retrieve the electron correlation effects on the electron densities. In Chapter 7, where we have reported the results of this work, we have shown that, for almost all

the considered levels of resolution, the amount of correlation effects is increasingly captured when the weight of the external constraint is very large. Furthermore, the largest recoveries have been detected when low or medium-resolution subsets of reflections are taken into account in the X-ray constrained Hartree-Fock procedure. These observations can be explained considering that the very large number of high-angle structure factors are already well described at RHF level and, therefore, they cancel the contribution of the low/medium angle reflections which are less numerous but more affected by the electron correlation effects. To tackle this problem, it seems necessary to properly weight the low-medium angle reflections in the expression of the Jayatilaka functional. Furthermore, in future works, it will be necessary to clarify the real meaning of the external adjustable parameter λ, and consequently, the meaning of X-ray constrained wave functions obtained when the contribution of the experimental data becomes predominant compared to electronic energy of the system in exam. Finally, analogous studies to also understand if the XCWF methods enable to capture crystal field effects will be also fundamental in the next future.

In Chapter 8, we have reported theoretical investigations on the charge density and the electronic structure of a doubly bridged annulene: namely, the *syn*-1,6:8,13-Biscarbonyl[14] annulene (BCA). X-ray diffraction measurements on this compound have recently shown that its aromatic character reduces when pressure increases. To investigate this phenomenon, we have exploited the simple unconstrained ELMO technique and the novel unconstrained and X-ray constrained ELMO Valence Bond approaches. First of all, considering proper localization schemes in unconstrained ELMO calculations, we have demonstrated that pressure directly affects the contributions of the two resonance structures of the Biscarbonyl[14] annulene and, hence, that the aromatic character of this system is partially reduced at high-pressure. In analogous way, through unconstrained and X-ray constrained ELMO-VB calculations, we have shown that the importance of the two resonance structures is equivalent in the description of the electronic structure of the BCA at ambient pressure. Nevertheless, this is no longer the case at high pressure, where one of the resonance structures becomes predominant over the other. Overall the results of this study have shown that the unconstrained ELMO, the unconstrained and the X-ray constrained ELMO-VB methods have been able to detect the partial rupture of the aromaticity in the Biscarbonyl[14] annulene. From the methodological point of view, this investigation confirms that, when reliable experimental data are available, the X-ray constrained wave function techniques can be considered as a valid alternative/complement to other

well-established methods to determine and analyse electron distributions in molecular crystals. This is the reason why, in future works, we would also like to develop new types of X-ray Constrained Wave Function strategies based on a multideterminant ansatz for the wave function.

Finally, concerning the possibility of using XC-ELMOs as elementary building blocks to reconstruct wave functions and electron densities of large molecules, we can conclude that the use of "experimental" extremely localized molecular orbitals could be useful to obtain charge distributions closer and closer to the corresponding Hartree-Fock ones. Nevertheless, our investigations have also shown that the electron correlation effects are only partially recovered in traditional X-ray constrained wave function calculations and, of course, this will reflect on the electron densities reconstructed with XC-ELMOs, which will be only slightly correlated. However, it is worth observing that to completely assess the quality of the charge distributions resulting from the transfer of X-ray constrained ELMOs, in the next future it will be necessary to perform a further study to understand if the Jayatilaka-type approaches significantly capture or not the effects of the crystal field on the electron densities.

Conclusions de la Partie II et perspectives futures

Dans la seconde partie de cette thèse, nous avons principalement étudié les orbitales moléculaires extrêmement localisées dans le cadre des méthodes de fonctions d'ondes contraintes notamment en ce qui concerne leur capacité à reproduire des données de diffraction aux rayons-X. Premièrement, dans le Chapitre 5, nous avons introduit la technique de la fonction d'onde contrainte initialement développée par Jayatilaka ainsi que sa récente extension à la théorie des ELMOs proposée par Genoni. De plus, nous avons brièvement exposé une nouvelle stratégie Valence Bond "expérimentale" basée sur des ELMOs. Ensuite, puisque dans la première partie de ce manuscrit nous avons démontré la transférabilité des ELMOs, nous avons exploré la possibilité de construire une base de données en utilisant des orbitales moléculaires extrêmement localisées contraintes. Néanmoins, nous n'avions pas d'éléments permettant de savoir si les ELMOs "expérimentales" étaient plus appropriées que les "théoriques" pour atteindre cet objectif. Ainsi, pour tenter de répondre à cette question, nous avons évalué les effets de la procédure de localisation dans l'approche de la fonction d'onde contrainte et jaugé la capacité de l'approche de Jayatilaka à capturer des effets de la corrélation sur la densité électronique.

Dans le Chapitre 6, où nous avons étudié les effets de la localisation sur les fonctions d'ondes contraintes, nous avons découvert que l'utilisation de petites fonctions de base a une grande influence sur les densités obtenues. Toutefois, si l'on utilise des fonctions de bases plus étendues et plus flexibles, les deux méthodes XC-RHF et XC-ELMO tendent vers des résultats similaires. Nous avons expliqué ce comportement en considérant que l'utilisation de bases de plus en plus flexibles permet de mieux en mieux exploiter les informations contenues dans les données expérimentales et que, par conséquent, cela atténue le biais introduit par la localisation.

Ensuite, pour mieux comprendre si les XC-ELMOs peuvent être plus avantageuses dans la construction d'une base de données, nous avons analysé dans quelle mesure l'approche de Jayatilaka permet de capturer des effets de la corrélation sur la densité électronique. Dans le Chapitre 7, nous avons reporté nos résultats sur ces travaux et nous avons pu montrer que, pour quasiment toute les résolutions considérées, le taux de récupération des effets de la corrélation augmente au fur et à mesure que l'on accentue la contrainte des facteurs de structure. De plus, les plus hauts taux de récupération ont été observés lorsque seulement des réflexions à moyenne ou basse résolution ont été utilisées

dans les calculs Hartree-Fock contraints. Ces résultats peuvent s'expliquer en considérant qu'un grand nombre de réflexions obtenues à haute résolution sont déjà très bien décrites au niveau Hartree-Fock. Ces nombreuses réflexions, ou facteurs de structure, réduisent la contribution des réflexions à moyenne ou basse résolution qui sont beaucoup plus affectées par les effets de la corrélation. Pour surmonter ce problème, il semble nécessaire d'appliquer des poids appropriés aux réflexions de basse et moyenne résolution dans l'expression de la fonctionnelle de Jayatilaka. En outre, dans de futurs travaux, il sera crucial de clarifier la réelle signification du multiplicateur externe λ, et donc, de mieux comprendre le sens d'obtenir des fonctions d'ondes contraintes lorsque la contribution des données expérimentales devient prédominante par rapport à l'énergie électronique du système.

Dans le Chapitre 8, nous avons effectué une étude théorique de la densité électronique du syn-1,6:8,13-Biscarbonyl[14] annulène (BCA). Des mesures précises de diffraction aux rayons-X ont montré une rupture partielle du caractère aromatique de ce composé lorsque la pression augmente. Pour étudier ce phénomène, nous avons employé dans un premier temps la méthode ELMO traditionnelle et, dans un second temps, la nouvelle approche ELMO Valence Bond (ELMO-VB) avec et sans contraintes expérimentales. Dans un premier temps, en utilisant des schémas de localisation appropriés dans nos calculs ELMOs sans contrainte, nous avons démontré que la pression influe directement sur les contributions des deux structures de résonance du Biscarbonyl[14] annulène, et donc, que le caractère aromatique de ce système est réduit à haute pression. De manière analogue, par des calculs ELMO-VB sous contraintes, nous avons montré que les deux structures de résonances ont la même contribution dans la description de la structure électronique du BCA à pression ambiante. Au contraire, cela n'est plus le cas à haute pression où l'une des structures de résonance devient prédominante par rapport à l'autre. D'un point de vue méthodologique, cette étude a confirmé que les techniques de la fonction d'onde contrainte peuvent être considérées comme un complément intéressant à d'autres méthodes pour l'étude de systèmes moléculaires à l'état cristallin lorsque des données expérimentales fiables peuvent être exploitées. Dans de futurs travaux, il serait intéressant d'étendre l'approche de la fonction d'onde contrainte à une fonction d'onde multi-référence.

Pour finir, en ce qui concerne la possible utilisation d'ELMOs "expérimentales" comme briques élémentaires pour reconstruire des fonctions d'ondes de grands systèmes, nous

pouvons conclure que les XC-ELMOs peuvent se montrer utiles pour obtenir des distributions de charges plus proches de celles d'un calcul Hartree-Fock. Cependant, nos études on également montré que la méthode de la fonction d'onde contrainte n'est que partiellement capable de capturer certains effets de la corrélation sur la densité électronique. Dans des études futures, il serait néanmoins très intéressant d'évaluer la capacité de la méthode de Jayatilaka à capturer les effets de champ cristallin, ce qui serait un argument très favorable pour l'utilisation des ELMOs "expérimentales".

Appendix A

Model molecule approximation and ELMOs transferability

TABLE A.1: Effects of the Molecular Orbitals Localization: Mean Relative Variations (in percentage) of the values of the Laplacian of the electron density at the bond critical points. Hartree-Fock values are used as reference.

	6-31G		6-311G		6-31G(d,p)		6-311G(d,p)	
bond	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP
C-H	0.53	0.58	0.87	-0.12	0.78	-2.09	0.69	-2.21
C-C	0.24	0.20	1.68	-2.21	1.57	-4.08	2.01	-5.21
$(C\text{-}C)_{ar}$	0.72	0.22	1.08	-1.24	0.77	-2.86	0.36	-3.24
$(C\text{-}N)_{peptide}$	0.46	-0.31	0.48	-0.86	0.71	-1.77	0.43	-1.62
$(C\text{-}N)_{term}$	2.08	4.01	2.25	3.26	2.19	1.65	1.83	2.13
C_{α}-N	4.07	3.21	4.35	0.75	5.14	-2.51	3.91	-2.21
$(C\text{-}O)_{term}$	1.78	-1.19	1.68	-1.00	2.04	-1.76	2.13	-1.25
$(C\text{-}O)_{peptide}$	0.95	-1.15	0.94	-0.90	1.16	-1.79	1.24	-1.16
$(C\text{-}O)_{phenol}$	0.89	0.81	0.97	1.05	0.58	-0.19	0.45	0.31
N-H	0.64	0.80	0.31	0.17	0.80	-1.64	0.65	-1.63
O-H	0.77	0.92	0.08	0.51	0.57	-1.06	1.17	-0.44

TABLE A.2: Effects of the Molecular Orbitals Localization: maximum relative variations (in percentage) of the values of the electron density at the bond critical points. Hartree-Fock values are used as reference.

	6-31G		6-311G		6-31G(d,p)		6-311G(d,p)	
bond	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP
C-H	3.37	1.60	3.48	0.77	3.70	-1.10	3.45	-1.25
C-C	0.88	1.03	2.69	-1.54	2.34	-3.18	3.67	-4.35
$(C\text{-}C)_{ar}$	1.29	0.31	1.62	-1.13	1.54	-2.72	1.12	-3.40
$(C\text{-}N)_{peptide}$	2.01	0.06	2.10	-0.63	2.93	-1.43	2.60	-1.40
$(C\text{-}N)_{term}$	2.08	4.01	2.25	3.26	2.19	1.65	1.83	2.13
C_{α}-N	5.03	3.61	5.19	2.60	5.51	1.76	4.97	1.58
$(C\text{-}O)_{term}$	1.96	-1.07	1.87	-0.85	2.22	-1.65	2.31	-1.11
$(C\text{-}O)_{peptide}$	1.15	-0.94	1.28	-0.69	1.52	-1.60	1.59	-0.97
$(C\text{-}O)_{phenol}$	0.89	0.81	0.97	1.05	0.58	-0.19	0.45	0.31
N-H	2.25	1.18	1.25	0.33	2.26	-1.30	1.86	1.29
O-H	1.87	1.16	1.19	0.77	1.60	-0.78	2.22	-0.20

TABLE A.3: Effects of the Molecular Orbitals Localization: minimum relative variations (in percentage) of the values of the electron density at the bond critical points. Hartree-Fock values are used as reference.

	6-31G		6-311G		6-31G(d,p)		6-311G(d,p)	
bond	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP
C-H	-1.29	-0.05	-1.22	-0.70	-1.27	-2.86	-1.37	-2.89
C-C	-0.45	-0.61	0.69	-2.94	0.30	-5.22	0.31	-6.17
$(C\text{-}C)_{ar}$	-0.11	0.07	0.06	-1.35	-0.39	3.04	-0.75	-3.43
$(C\text{-}N)_{peptide}$	-1.00	-0.79	-0.97	-1.09	-1.11	-2.20	-1.40	-1.82
$(C\text{-}N)_{term}$	2.08	4.01	2.25	3.26	2.19	1.65	1.83	2.13
C_{α}-N	0.28	2.35	0.48	1.28	0.59	0.46	4.97	1.58
$(C\text{-}O)_{term}$	1.60	-1.30	1.49	-1.15	1.87	-1.86	2.31	-1.11
$(C\text{-}O)_{peptide}$	0.44	-1.54	0.27	-1.18	0.56	-2.11	1.59	-0.97
$(C\text{-}O)_{phenol}$	0.89	0.81	0.97	1.05	0.58	-0.19	0.45	0.31
N-H	-1.49	0.61	-1.01	-0.19	-0.95	-2.08	1.86	-1.29
O-H	-1.62	0.66	-2.10	0.28	-1.90	-1.36	2.22	-0.20

TABLE A.4: Effects of the Molecular Orbitals Localization: Mean Relative Variations (in percentage) of the values of the Laplacian of the electron density at the bond critical points. Hartree-Fock values are used as reference.

	6-31G		6-311G		6-31G(d,p)		6-311G(d,p)	
bond	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP
C-H	0.40	1.10	1.78	-2.87	1.54	-7.75	1.47	-9.02
C-C	4.85	-4.28	11.17	-14.51	7.95	-17.94	9.58	-21.27
$(C\text{-}C)_{ar}$	4.26	-2.58	5.13	-8.33	4.62	-13.11	2.90	-13.79
$(C\text{-}N)_{peptide}$	4.45	-8.77	4.13	-11.24	6.29	10.14	4.85	7.39
$(C\text{-}N)_{term}$	32.78	26.86	39.86	34.55	35.92	70.94	39.19	155.04
C_{α}-N	29.63	16.54	30.16	11.17	45.47	49.20	43.43	61.44
$(C\text{-}O)_{term}$	26.59	-25.60	21.84	-7.51	232.43	15.34	77.76	32.57
$(C\text{-}O)_{peptide}$	14.76	-29.35	15.39	-5.09	108.58[a]	3.18[a]	53.36	44.54
$(C\text{-}O)_{phenol}$	58.69	10.33	86.32	93.98				
N-H	0.65	0.24	0.16	-6.46	2.18	-6.10	-0.14	-12.44
O-H	0.51	0.25	-2.31	-2.50	1.41	-8.31	1.83	-6.89

[a]The value related to the C48-O39 BCP has not been considered.

TABLE A.5: Effects of the Molecular Orbitals Localization: maximum relative variations (in percentage) of the values of the Laplacian of the electron density at the bond critical points. Hartree-Fock values are used as reference.

	6-31G		6-311G		6-31G(d,p)		6-311G(d,p)	
bond	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP
C-H	10.33	4.32	6.81	-1.13	9.03	-6.14	5.87	-8.05
C-C	8.40	-0.69	19.29	-11.56	9.59	-16.04	14.56	-18.65
$(C\text{-}C)_{ar}$	4.09	-2.13	6.86	-7.82	7.63	-12.42	5.67	-13.02
$(C\text{-}N)_{peptide}$	11.01	-6.98	10.40	-10.43	28.55	13.32	24.10	7.99
$(C\text{-}N)_{term}$	32.78	26.86	39.86	34.55	35.92	70.94	39.19	155.04
C_{α}-N	65.56	36.87	65.36	30.61	100.53	83.24	106.07	106.47
$(C\text{-}O)_{term}$	32.76	-22.51	26.58	-5.90	375.80	24.61	100.34	41.58
$(C\text{-}O)_{peptide}$	16.94	-26.74	18.93	-2.90	140.47[a]	12.34[a]	61.54	60.92
$(C\text{-}O)_{phenol}$	58.69	10.33	86.32	93.98				
N-H	1.42	0.90	8.05	-4.93	2.68	-5.30	8.43	-10.67
O-H	1.18	0.74	3.15	20.43	6.42	-7.36	3.03	-5.94

[a]The value related to the C48-O39 BCP has not been considered.

TABLE A.6: Effects of the Molecular Orbitals Localization: minimum relative variations (in percentage) of the values of the Laplacian of the electron density at the bond critical points. Hartree-Fock values are used as reference.

	6-31G		6-311G		6-31G(d,p)		6-311G(d,p)	
bond	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP
C-H	-5.66	-1.19	-2.12	-4.33	-4.00	-9.38	-2.61	-10.25
C-C	2.68	-8.61	6.31	-18.63	5.63	-20.93	5.69	-24.44
$(C\text{-}C)_{ar}$	7.04	-3.45	2.28	-9.04	0.45	-14.05	-0.75	-14.81
$(C\text{-}N)_{peptide}$	-2.22	-11.74	-2.76	-11.64	-13.14	6.21	-12.12	6.60
$(C\text{-}N)_{term}$	32.78	26.86	39.86	34.55	35.92	70.94	39.19	155.04
C_{α}-N	10.46	6.97	9.97	1.66	10.71	30.45	0.58	36.58
$(C\text{-}O)_{term}$	20.43	-28.68	17.09	-9.12	89.06	6.06	55.17	23.57
$(C\text{-}O)_{peptide}$	12.92	-32.01	10.33	-7.86	79.73[a]	-11.47[a]	44.34	33.68
$(C\text{-}O)_{phenol}$	58.69	10.33	86.32	93.98				
N-H	-0.16	-1.16	-7.39	-9.04	1.63	-7.51	-7.40	-15.21
O-H	0.02	-0.63	-4.78	-7.20	-0.67	-9.96	0.68	-8.54

[a]The value related to the C48-O39 BCP has not been considered.

TABLE A.7: Effects of the Molecular Orbitals Localization: Mean Deviations (in e) of the net integrated atomic charges. Hartree-Fock values are used as reference.

	6-31G		6-311G		6-31G(d,p)		6-311G(d,p)	
bond	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP
C	0.010	-0.061	0.014	-0.078	0.015	-0.118	0.016	-0.120
N	0.044	0.134	0.030	0.200	0.019	0.282	0.022	0.319
O	-0.040	0.106	-0.037	0.140	-0.041	0.163	-0.033	0.185
H	-0.002	-0.001	-0.004	-0.005	-0.003	0.007	-0.007	-0.002

TABLE A.8: Effects of the Molecular Orbitals Localization: maximum deviations (in e) of the net integrated atomic charges. Hartree-Fock values are used as reference.

	6-31G		6-311G		6-31G(d,p)		6-311G(d,p)	
bond	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP
C	0.276	0.019	0.292	0.013	0.319	0.026	0.338	0.023
N	0.126	0.155	0.106	0.224	0.102	0.315	0.094	0.358
O	0.026	0.206	0.031	0.246	0.025	0.270	0.032	0.288
H	0.088	0.044	0.085	0.047	0.095	0.072	0.090	0.070

TABLE A.9: Effects of the Molecular Orbitals Localization: minimum deviations (in e) of the net integrated atomic charges. Hartree-Fock values are used as reference.

	6-31G		6-311G		6-31G(d,p)		6-311G(d,p)	
bond	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP
C	-0.143	-0.197	-0.151	-0.275	-0.172	-0.337	-0.179	-0.378
N	-0.015	0.079	-0.034	0.116	-0.030	0.180	-0.034	0.201
O	-0.074	0.050	-0.073	0.071	-0.081	0.086	-0.068	0.115
H	-0.067	-0.057	-0.071	-0.074	-0.115	-0.074	-0.084	-0.087

TABLE A.10: Number of considered molecular orbitals in the computations of the average energy variations for the comparison of the Nearest Atom (NAA), Nearest Bond (NBA), and Nearest Functional Group (NFGA) Approximations.

energy variation	number of MOs
$\langle \Delta E \rangle_{Overall}$	163
$\langle \Delta E \rangle_{Atoms}$	63
$\langle \Delta E \rangle_{Bonds}$	100
$\langle \Delta E \rangle_{C}$	28
$\langle \Delta E \rangle_{N}$	5
$\langle \Delta E \rangle_{O}$	30
$\langle \Delta E \rangle_{C-H}$	29
$\langle \Delta E \rangle_{C-C}$	13
$\langle \Delta E \rangle_{C-N}$	4
$\langle \Delta E \rangle_{C-O}$	1
$\langle \Delta E \rangle_{N-H}$	7
$\langle \Delta E \rangle_{O-H}$	8
$\langle \Delta E \rangle_{(C-C-C)_{ar}}$	16
$\langle \Delta E \rangle_{O-C-O}$	4
$\langle \Delta E \rangle_{N-C-O}$	18

TABLE A.11: Number of considered bond critical points in the computations of the mean absolute relative variations.

bond	number of bond
overall	84
C-H	29
C-C	13
$(C\text{-}C)_{ar}$	12
$(C\text{-}N)_{peptide}$	4
$(C\text{-}N)_{term}$	1
C_{α}-N	4
$(C\text{-}O)_{term}$	2
$(C\text{-}O)_{peptide}$	4
$(C\text{-}O)_{phenol}$	1
N-H	7
O-H	7

TABLE A.12: Number of considered atoms in the computations of the mean absolute variations.

bond	number of bond
overall	86
C	28
N	5
O	10
H	43

Appendix B

A comparison with the pseudoatoms transferability

TABLE B.1: Standard deviations associated with the average values of the electron distributions at the covalent bond critical points ($e/Å^3$) for each bond type after the transfer of pseudoatoms and ELMOs. (For $(C\text{-}N)_{term}$ and $(C\text{-}O)_{phenol}$ only one BCP is considered.)

	pseudoatoms		ELMOs transfer		
bond	ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
C-H	0.03	0.02	0.03	0.03	0.03
C-C	0.05	0.06	0.04	0.05	0.05
$(C\text{-}C)_{ar}$	0.03	0.03	0.02	0.02	0.02
$(C\text{-}N)_{peptide}$	0.03	0.04	0.03	0.03	0.03
$(C\text{-}N)_{term}$	0.03	0.00	0.00	0.00	0.00
C_{α}-N	0.00	0.02	0.01	0.01	0.01
$(C\text{-}O)_{term}$	0.01	0.04	0.04	0.04	0.05
$(C\text{-}O)_{peptide}$	0.05	0.02	0.02	0.02	0.02
$(C\text{-}O)_{phenol}$	0.00	0.00	0.00	0.00	0.00
N-H	0.06	0.04	0.07	0.06	0.07
O-H	0.02	0.03	0.01	0.02	0.02

TABLE B.2: Standard deviations associated with the average values of the Laplacian of the electron distributions at the covalent bond critical points ($e/Å^5$) for each bond type after the transfer of pseudoatoms and ELMOs. (For $(C\text{-}N)_{term}$ and $(C\text{-}O)_{phenol}$ only one BCP is considered.)

	pseudoatoms		ELMOs transfer		
bond	ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
C-H	1.30	0.54	0.69	0.73	1.03
C-C	0.97	0.82	0.99	1.42	1.33
$(C\text{-}C)_{ar}$	0.60	0.64	0.56	0.61	0.79
$(C\text{-}N)_{peptide}$	0.97	2.42	0.55	1.17	1.15
$(C\text{-}N)_{term}$	0.00	0.00	0.00	0.00	0.00
C_α-N	0.30	1.44	0.18	0.79	0.47
$(C\text{-}O)_{term}$	1.85	2.80	0.04	3.94	3.89
$(C\text{-}O)_{peptide}$	1.15	1.36	2.97	1.15	1.84
$(C\text{-}O)_{phenol}$	0.00	0.00	1.33	0.00	0.00
N-H	2.84	3.81	0.76	0.89	0.46
O-H	1.13	1.91	1.84	1.03	1.23

TABLE B.3: Average values of the electron distributions at the covalent bond critical points ($e/Å^3$) for each bond type after the transfer of pseudoatoms and ELMOs. (For $(C\text{-}N)_{term}$ and $(C\text{-}O)_{phenol}$ only one BCP is considered.)

	pseudoatoms		ELMOs transfer		
bond	ELMAM2	UBDB	6-31G	6-31G(d,p)	6-311+G(d,p)
C-H	1.84	1.89	1.79	1.98	1.96
C-C	1.61	1.66	1.58	1.80	1.78
$(C\text{-}C)_{ar}$	2.12	2.08	2.00	2.19	2.16
$(C\text{-}N)_{peptide}$	2.27	2.26	2.15	2.31	2.28
$(C\text{-}N)_{term}$	1.62	1.56	1.31	1.44	1.54
C_α-N	1.73	1.74	1.66	1.83	1.79
$(C\text{-}O)_{term}$	2.66	2.55	2.46	2.60	2.58
$(C\text{-}O)_{peptide}$	2.65	2.70	2.59	2.74	2.73
$(C\text{-}O)_{phenol}$	2.10	2.00	1.87	2.02	1.99
N-H	2.26	2.19	2.14	2.35	2.32
O-H	2.52	2.39	2.30	2.49	2.48

TABLE B.4: Average values of the Laplacian of the electron density at the covalent bond critical points (e/Å^5) for each bond type after the transfer of pseudoatoms and ELMOs. (For $(C\text{-}N)_{term}$ and $(C\text{-}O)_{phenol}$ only one BCP is considered.)

	pseudoatoms		ELMOs transfer		
bond	ELMAM2	UBDB	6-31G	6-31G(d,p)	6-311+G(d,p)
C-H	-17.9	-20.2	-18.8	-27.1	-25.9
C-C	-9.0	-11.3	-11.3	-18.9	-19.0
$(C\text{-}C)_{ar}$	-18.4	-17.3	-17.8	-25.0	-24.7
$(C\text{-}N)_{peptide}$	-22.1	-23.7	-22.5	-17.6	-19.2
$(C\text{-}N)_{term}$	-6.9	-9.6	-2.6	-0.6	-7.3
C_α-N	-9.4	-10.1	-14.4	-19.2	-16.3
$(C\text{-}O)_{term}$	-29.8	-24.2	-21.7	-9.9	-13.9
$(C\text{-}O)_{peptide}$	-22.1	-24.5	-19.8	-7.0	-12.5
$(C\text{-}O)_{phenol}$	-17.6	-17.5	-10.8	-5.2	-4.2
N-H	-37.0	-28.5	-35.6	-47.6	-44.5
O-H	-58.0	-39.4	-39.1	-54.2	-64.5

TABLE B.5: Standard deviations associated with the average values of the electron distributions at the covalent bond critical points (e/Å^3) for each bond type after the transfer of pseudoatoms and ELMOs. (For $(C\text{-}N)_{term}$ and $(C\text{-}O)_{phenol}$ only one BCP is considered.)

	pseudoatoms		ELMOs transfer		
bond	ELMAM2	UBDB	6-31G	6-31G(d,p)	6-311+G(d,p)
C-H	0.03	0.02	0.03	0.03	0.03
C-C	0.05	0.06	0.03	0.04	0.05
$(C\text{-}C)_{ar}$	0.03	0.03	0.02	0.02	0.02
$(C\text{-}N)_{peptide}$	0.03	0.04	0.03	0.03	0.03
$(C\text{-}N)_{term}$	0.03	0.00	0.00	0.00	0.00
C_α-N	0.00	0.02	0.01	0.01	0.01
$(C\text{-}O)_{term}$	0.01	0.04	0.04	0.04	0.04
$(C\text{-}O)_{peptide}$	0.05	0.02	0.02	0.02	0.02
$(C\text{-}O)_{phenol}$	0.00	0.00	0.00	0.00	0.00
N-H	0.06	0.04	0.07	0.06	0.06
O-H	0.02	0.03	0.01	0.02	0.02

TABLE B.6: Standard deviations associated with the average values of the Laplacian of the electron distributions at the covalent bond critical points ($e/Å^5$) for each bond type after the transfer of pseudoatoms and ELMOs. (For $(C\text{-}N)_{term}$ and $(C\text{-}O)_{phenol}$ only one BCP is considered.)

	pseudoatoms		ELMOs transfer		
bond	ELMAM2	UBDB	6-31G	6-31G(d,p)	6-311+G(d,p)
C-H	1.30	0.54	0.93	1.01	0.73
C-C	0.97	0.82	0.71	1.26	1.43
$(C\text{-}C)_{ar}$	0.60	0.64	0.52	0.65	0.61
$(C\text{-}N)_{peptide}$	0.97	2.42	0.90	1.39	0.90
$(C\text{-}N)_{term}$	0.00	0.00	0.00	0.00	0.00
C_α-N	0.30	1.44	0.29	0.84	0.77
$(C\text{-}O)_{term}$	1.85	2.80	3.69	4.56	4.32
$(C\text{-}O)_{peptide}$	1.15	1.36	1.11	1.26	1.94
$(C\text{-}O)_{phenol}$	0.00	0.00	0.00	0.00	0.00
N-H	2.84	3.81	0.79	0.66	0.81
O-H	1.13	1.91	0.35	1.42	1.02

TABLE B.7: Values of the electron distribution at the non-covalent bond critical points ($e/Å^3$) after the transfer of pseudoatoms and ELMOs.

	pseudoatoms		ELMOs transfer		
non-covalent interaction	ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
N31-H37 $\cdots$ Ow84	0.191	0.191	0.189	0.172	0.170
Ow81-Hw83 $\cdots$ O5	0.238	0.176	0.190	0.165	0.165
Ow78-Hw80 $\cdots$ Ow81	0.134	0.132	0.114	0.107	0.104
N58-H74 $\cdots$ O25	0.224	0.154	0.168	0.144	0.143
N38-H57 $\cdots$ O6	0.165	0.166	0.147	0.136	0.133
C11 $\cdots$ H35	0.024	0.028	0.029	0.029	0.030
C13 $\cdots$ O84	0.040	0.041	0.046	0.045	0.046
H17 $\cdots$ H50	0.022	0.024	0.020	0.021	0.022
H19 $\cdots$ C42	0.064	0.069	0.065	0.064	0.063
H20 $\cdots$ C46	0.017	0.018	0.022	0.022	0.023
H21 $\cdots$ O81	0.029	0.032	0.034	0.035	0.038
H28 $\cdots$ H65	0.021	0.025	0.022	0.023	0.024
H56 $\cdots$ O32	0.052	0.052	0.048	0.048	0.049

TABLE B.8: Values of the Laplacian of the electron density at the non-covalent bond critical points ($e/Å^5$) after the transfer of pseudoatoms and ELMOs.

	pseudoatoms		ELMOs transfer		
non-covalent interaction	ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
N31-H37 $\cdots$ Ow84	1.96	2.43	2.86	2.44	2.93
Ow81-Hw83 $\cdots$ O5	1.22	1.92	2.96	2.49	2.93
Ow78-Hw80 $\cdots$ Ow81	1.42	2.33	1.79	1.49	1.71
N58-H74 $\cdots$ O25	1.13	1.41	2.68	2.15	2.64
N38-H57 $\cdots$ O6	1.43	1.87	2.33	1.95	2.33
C11 $\cdots$ H35	0.27	0.28	0.35	0.34	0.29
C13 $\cdots$ O84	0.57	0.59	0.61	0.58	0.56
H17 $\cdots$ H50	0.24	0.23	0.30	0.28	0.26
H19 $\cdots$ C42	0.64	0.63	0.82	0.76	0.69
H20 $\cdots$ C46	0.26	0.27	0.29	0.28	0.27
H21 $\cdots$ O81	0.50	0.53	0.58	0.57	0.50
H28 $\cdots$ H65	0.25	0.24	0.33	0.31	0.28
H56 $\cdots$ O32	0.67	0.62	0.79	0.75	0.65

TABLE B.9: Values of the kinetic energy density at the hydrogen-bond critical points ($kJ.mol.Å^{-3}$) obtained through the Espinosa empirical relation and after the transfer of pseudoatoms and ELMOs.

			pseudoatoms		ELMOs transfer		
hydrogen-bond	$d(H \cdots O)$	Espinosa	ELMAM2	UBDB	6-31G	6-31G(d,p)	6-311+G(d,p)
N31-H37 $\cdots$ Ow84	1.920	63.4	55.4	63.9	71.4	61.1	69.6
Ow81-Hw83 $\cdots$ O5	1.892	68.6	54.4	59.5	73.4	60.8	68.6
Ow78-Hw80 $\cdots$ Ow81	2.098	39.0	31.5	36.3	40.9	34.6	38.2
N58-H74 $\cdots$ O25	1.931	61.5	48.0	48.8	64.7	51.5	60.1
N38-H57 $\cdots$ O6	1.986	53.0	41.6	49.6	55.1	46.8	53.2

TABLE B.10: Values of the potential energy density at the hydrogen-bond critical points (kJ.mol.Å^{-3}) obtained through the Espinosa empirical relation and after the transfer of pseudoatoms and ELMOs.

		pseudoatoms	ELMOs transfer				
hydrogen-bond	$d(H\cdots O)$	Espinosa	ELMAM2	UBDB	6-31G	6-31G(d,p)	6-311+G(d,p)
N31-H37$\cdots$Ow84	1.920	-49.8	-57.4	-61.6	-64.9	-55.6	-59.4
Ow81-Hw83$\cdots$O5	1.892	-55.2	-70.0	-55.7	-66.2	-53.8	-57.6
Ow78-Hw80$\cdots$Ow81	2.098	-26.3	-32.2	-34.2	-33.1	-28.6	-29.8
N58-H74$\cdots$O25	1.931	-47.8	-62.9	-45.3	-56.3	-44.4	-48.3
N38-H57$\cdots$O6	1.986	-39.3	-44.1	-48.3	-46.8	-40.4	-42.9

TABLE B.11: Values of the positive curvature λ_3 of the electron density at the hydrogen-bond critical points (e/Å^5) obtained through the Espinosa empirical relation and after the transfer of pseudoatoms and ELMOs.

		pseudoatoms	ELMOs transfer				
hydrogen-bond	$d(H\cdots O)$	Espinosa	ELMAM2	UBDB	6-31G	6-31G(d,p)	6-311+G(d,p)
N31-H37$\cdots$Ow84	1.920	4.1	4.0	4.2	4.7	4.0	4.5
Ow81-Hw83$\cdots$O5	1.892	4.4	4.3	4.1	5.0	4.3	4.5
Ow78-Hw80$\cdots$Ow81	2.098	2.7	2.4	2.6	2.8	2.3	2.5
N58-H74$\cdots$O25	1.931	4.0	4.0	3.5	4.4	3.5	4.0
N38-H57$\cdots$O6	1.986	3.5	3.1	3.4	3.7	3.1	3.5

TABLE B.12: Net QTAIM integrated charges (e) for the backbone "functional groups" the five side chains (including the $C_\alpha H_\alpha$ bonds) of the Leu-enkephalin polypeptide.

	pseudoatoms		ELMOs transfer		
fragment	ELMAM2	UBDB	6-31G	6-31G(d,p)	6-311+G(d,p)
ammonium (N-term)	0.244	0.401	0.376	0.300	0.364
H-N-C=O (Tyr-Gly)	-0.483	-0.418	-0.485	-0.614	-0.600
H-N-C=O (Gly-Gly)	-0.448	-0.430	-0.524	-0.668	-0.657
H-N-C=O (Gly-Phe)	-0.434	-0.426	-0.585	-0.719	-0.732
H-N-C=O (Phe-Leu)	-0.435	-0.422	-0.538	-0.667	-0.679
carboxylate (C-term)	-0.695	-0.673	-0.990	-0.961	-0.981
$C_\alpha H_\alpha - R$ (Tyr)	0.321	0.395	0.682	0.753	0.689
$C_\alpha H_2$ (Gly 1)	0.391	0.257	0.521	0.667	0.649
$C_\alpha H_2$ (Gly 2)	0.386	0.250	0.580	0.724	0.730
$C_\alpha H_\alpha - R$ (Phe)	0.501	0.563	0.552	0.691	0.693
$C_\alpha H_\alpha - R$ (Leu)	0.651	0.530	0.438	0.549	0.547
water 1	0.031	0.010	-0.009	-0.004	-0.007
water 2	-0.002	-0.005	-0.003	-0.003	0.002
water 3	-0.025	-0.026	-0.011	-0.013	-0.008
global	0.004	0.006	0.004	0.043	0.009

TABLE B.13: Values of the agreement index R (%) between the different reconstruction methods considering as topological property the electron density at the covalent bond critical points

	pseudoatoms		ELMO		
	ELMAM2	UBDB	6-31G	6-31G(d,p)	6-311+G(d,p)
ELMAM2	0.0	1.3	1.8	2.5	2.1
UBDB		0.0	1.9	2.2	1.7
ELMO/6-31G			0.0	4.0	3.6
ELMO/6-31G(d,p)				0.0	0.5
ELMO/6-311+G(d,p)					0.0

TABLE B.14: Values of the agreement index R (%) between the different reconstruction methods considering as topological property the Laplacian of the electron density at the covalent bond critical points

	pseudoatoms		ELMO		
	ELMAM2	UBDB	6-31G	6-31G(d,p)	6-311+G(d,p)
ELMAM2	0.0	7.1	6.2	19.8	16.6
UBDB		0.0	4.8	18.9	16.2
ELMO/6-31G			0.0	17.4	15.2
ELMO/6-31G(d,p)				0.0	4.6
ELMO/6-311+G(d,p)					0.0

TABLE B.15: Values of the agreement index R (%) between the different reconstruction methods considering as topological property the electron density at the non-covalent bond critical points

	pseudoatoms		ELMO		
	ELMAM2	UBDB	6-31G	6-31G(d,p)	6-311+G(d,p)
ELMAM2	0.0	0.7	0.9	1.1	1.3
UBDB		0.0	0.6	0.7	0.8
ELMO/6-31G			0.0	0.3	0.5
ELMO/6-31G(d,p)				0.0	0.2
ELMO/6-311+G(d,p)					0.0

TABLE B.16: Values of the agreement index R (%) between the different reconstruction methods considering as topological property the Laplacian of the electron density at the non-covalent bond critical points

	pseudoatoms		ELMO		
	ELMAM2	UBDB	6-31G	6-31G(d,p)	6-311+G(d,p)
ELMAM2	0.0	1.0	2.1	1.5	1.6
UBDB		0.0	1.3	0.7	0.9
ELMO/6-31G			0.0	0.6	0.6
ELMO/6-31G(d,p)				0.0	0.8
ELMO/6-311+G(d,p)					0.0

TABLE B.17: Values of the agreement index R (%) between the different reconstruction methods considering as topological property the net QTAIM integrated charges (e) for the backbone "functional groups" and for the five side chains of the Leu-enkephalin polypeptide.

	pseudoatoms		ELMO		
	ELMAM2	UBDB	6-31G	6-31G(d,p)	6-311+G(d,p)
ELMAM2	0.0	8.3	15.2	20.6	20.6
UBDB		0.0	16.0	22.8	21.8
ELMO/6-31G			0.0	9.2	8.0
ELMO/6-31G(d,p)				0.0	1.5
ELMO/6-311+G(d,p)					0.0

TABLE B.18: Point-by-point comparison of the UBDB, ELMO/6-31G, ELMO/6-31G(d,p), and ELMO/6-311+G(d,p) electron densities with the ELMAM2 charge density: values of the real-space R and of the Walker-Mezey similarity L(a, a') indexes (%). [a].

		ELMOs transfer		
similarity index	UBDB	6-31G	6-31G(d,p)	6-311+G(d,p)
RSR	1.3	2.3	2.1	2.0
L(0.001,1000)	91.4	82.9	83.3	82.5
L(0.001,0.01)	87.0	70.7	71.2	70.9
L(0.01,1)	93.4	88.3	88.6	87.5
L(1,3)	97.7	95.9	96.2	96.6
L(3,1000)	97.9	96.0	97.1	97.3

[a] For the Walker-Mezey indicator, the electron densities are compared within the a and a' limits expressed in e/bohr^3.

TABLE B.19: Point-by-point comparison of the different electron densities: Values of the real-space R similarity index (%).

	pseudoatoms		ELMO		
	ELMAM2	UBDB	6-31G	6-31G(d,p)	6-311+G(d,p)
ELMAM2	0.0	1.3	2.3	2.1	2.0
UBDB		0.0	2.0	1.7	1.7
ELMO/6-31G			0.0	1.4	1.5
ELMO/6-31G(d,p)				0.0	0.6
ELMO/6-311+G(d,p)					0.0

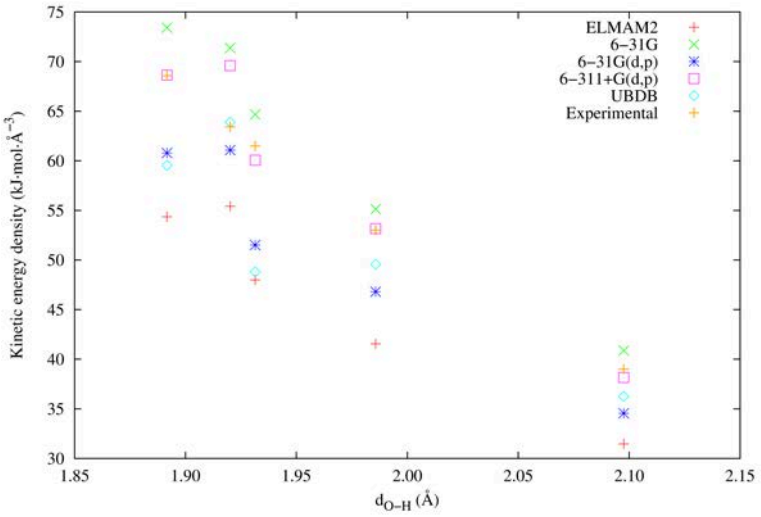

FIGURE B.1: Kinetic energy density at the hydrogen-bond critical points (kJ.mol.Å^{-3}) in function of the experimental distance $d(H\cdots O)$ in Å. The values are obtained through the Espinosa empirical relation and after the transfer of pseudoatoms (ELMAM2 and UBDB databases) and ELMOs (basis-sets 6-31G, 6-31G(d,p) and 6-311+G(d,p)).

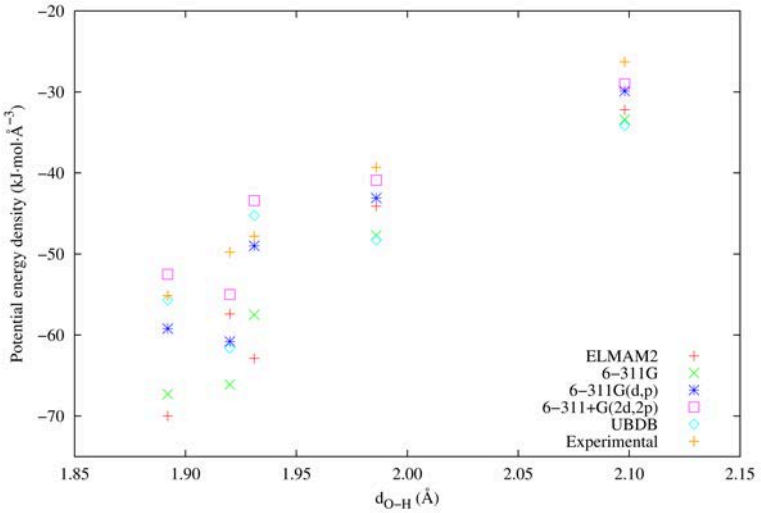

FIGURE B.2: Potential energy density at the hydrogen-bond critical points (kJ.mol.Å^{-3}) in function of the experimental distance $d(H \cdots O)$ in Å. The values are obtained through the Espinosa empirical relation and after the transfer of pseudoatoms (ELMAM2 and UBDB databases) and ELMOs (basis-sets 6-311G, 6-311G(d,p) and 6-311+G(2d,2p)).

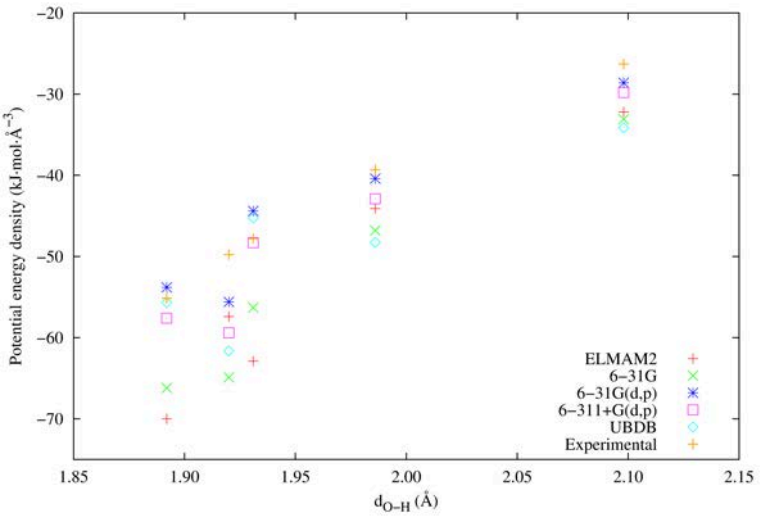

FIGURE B.3: Potential energy density at the hydrogen-bond critical points (kJ.mol.Å^{-3}) in function of the experimental distance $d(H \cdots O)$ in Å. The values are obtained through the Espinosa empirical relation and after the transfer of pseudoatoms (ELMAM2 and UBDB databases) and ELMOs (basis-sets 6-31G, 6-31G(d,p) and 6-311+G(d,p)).

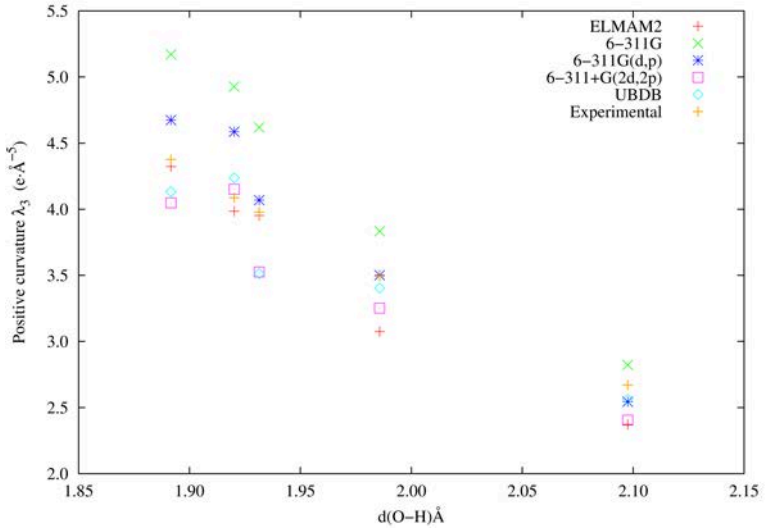

FIGURE B.4: Positive curvature λ_3 of the electron density at the hydrogen-bond critical points (e/Å^5) in function of the experimental distance $d(H \cdots O)$ in Å. The values are obtained through the Espinosa empirical relation and after the transfer of pseudoatoms (ELMAM2 and UBDB databases) and ELMOs (basis-sets 6-311G, 6-311G(d,p) and 6-311+G(2d,2p)).

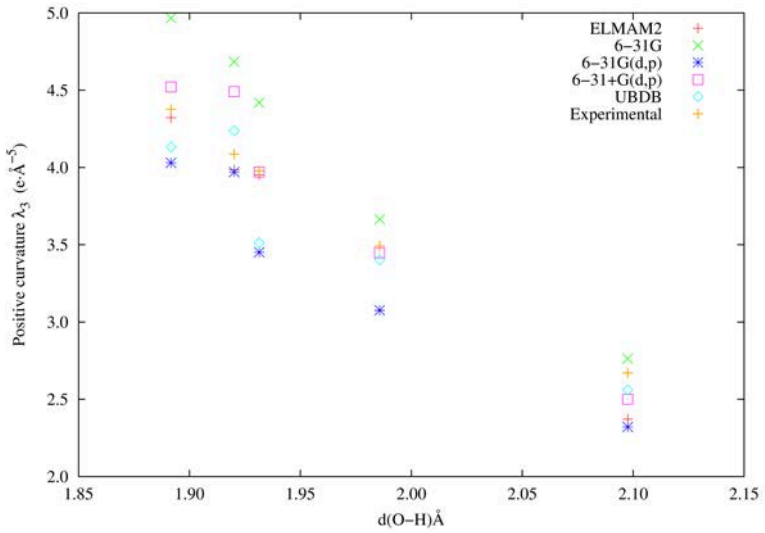

FIGURE B.5: Positive curvature λ_3 of the electron density at the hydrogen-bond critical points (e/Å^5) in function of the experimental distance $d(H \cdots O)$ in Å. The values are obtained through the Espinosa empirical relation and after the transfer of pseudoatoms (ELMAM2 and UBDB databases) and ELMOs (basis-sets 6-31G, 6-31G(d,p) and 6-311+G(d,p)).

Appendix C

Extraction of electron correlation effects on the electron density from X-Ray diffraction data

TABLE C.1: Total number of structure factor amplitudes and number of significant discrepancies (> 0.005) between the CCSD and the RHF structure factor amplitudes for each resolution range taken into account.

resolution	total number	$\lvert F_h^{CCSD} - F_h^{RHF} \rvert > 0.005$
$sin\theta/\lambda \leq 0.25\text{Å}^{-1}$	257	212
$0.25\text{Å}^{-1} < sin\theta/\lambda \leq 0.5\text{Å}^{-1}$	1827	1715
$0.5\text{Å}^{-1} < sin\theta/\lambda \leq 0.7\text{Å}^{-1}$	3672	1849
$0.7\text{Å}^{-1} < sin\theta/\lambda \leq 0.9\text{Å}^{-1}$	6446	1249
$0.9\text{Å}^{-1} < sin\theta/\lambda \leq 1.2\text{Å}^{-1}$	16686	412
$1.2\text{Å}^{-1} < sin\theta/\lambda \leq 1.5\text{Å}^{-1}$	27652	0
$1.5\text{Å}^{-1} < sin\theta/\lambda \leq 2.0\text{Å}^{-1}$	77340	0

TABLE C.2: Values of the electron density (e/Å^3)) at the bond critical point of the nitrogen molecule for all the considered methods.

method	$(sin\theta/\lambda)_{max}$(Å^{-1})						
	2.0	1.5	1.2	0.9	0.7	0.5	0.25
CCSD	4.7094						
CISD	4.7222						
DFT/BLYP	4.7079						
DFT/B3LYP	4.7275						
DFT/VSXC	4.7133						
DFT/B1B95	4.7183						
RHF	4.7901						
XC-RHF (λ=0.5)	4.7899	4.7896	4.7891	4.7879	4.7861	4.7906	4.7931
XC-RHF (λ=1.0)	4.7897	4.7891	4.7882	4.7859	4.7825	4.7906	4.7946
XC-RHF (λ=1.5)	4.7895	4.7887	4.7873	4.7840	4.7792	4.7904	4.7956
XC-RHF (λ=2.0)	4.7893	4.7882	4.7864	4.7821	4.7761	4.7900	4.7965
XC-RHF (λ=2.5)	4.7891	4.7877	4.7856	4.7804	4.7733	4.7894	4.7974
XC-RHF (λ=5.0)	4.7881	4.7855	4.7815	4.7725	4.7617	4.7856	4.8009
XC-RHF (λ=7.5)	4.7872	4.7834	4.7779	4.7661	4.7533	4.7816	4.8038
XC-RHF (λ=10.0)	4.7863	4.7814	4.7745	4.7608	4.7470	4.7779	4.8062

TABLE C.3: Values of the Laplacian of the electron density (e/Å^5)) at the bond critical point of the nitrogen molecule for all the considered methods.

method	$(sin\theta/\lambda)_{max}$(Å^{-1})						
	2.0	1.5	1.2	0.9	0.7	0.5	0.25
CCSD	-19.981						
CISD	-20.136						
DFT/BLYP	-19.689						
DFT/B3LYP	-20.011						
DFT/VSXC	-20.081						
DFT/B1B95	-19.921						
RHF	-21.306						
XC-RHF (λ=0.5)	-21.301	-21.295	-21.285	-21.257	-21.212	-21.281	-21.287
XC-RHF (λ=1.0)	-21.296	-21.284	-21.264	-21.211	-21.130	-21.261	-21.278
XC-RHF (λ=1.5)	-21.292	-21.273	-21.244	-21.168	-21.056	-21.244	-21.274
XC-RHF (λ=2.0)	-21.287	-21.263	-21.225	-21.128	-20.989	-21.228	-21.273
XC-RHF (λ=2.5)	-21.283	-21.253	-21.206	-21.089	-20.929	-21.212	-21.274
XC-RHF (λ=5.0)	-21.261	-21.203	-21.119	-20.926	-20.698	-21.142	-21.289
XC-RHF (λ=7.5)	-21.239	-21.157	-21.042	-20.798	-20.541	-21.080	-21.309
XC-RHF (λ=10.0)	-21.218	-21.114	-20.973	-20.695	-20.428	-21.026	-21.327

TABLE C.4: Root Mean Square Deviations between the CCSD electron density and the charge distributions associated with all the other considered methods. For the sake of clarity, all the values are multiplied by 10^4

method	$(sin\theta/\lambda)_{max}(\text{Å}^{-1})$						
	2.0	1.5	1.2	0.9	0.7	0.5	0.25
CISD	1.66						
DFT/BLYP	8.23						
DFT/B3LYP	5.22						
DFT/VSXC	7.25						
DFT/B1B95	4.54						
RHF	10.15						
XC-RHF (λ=0.5)	10.11	10.07	10.01	9.85	9.58	9.28	9.75
XC-RHF (λ=1.0)	10.07	9.99	9.87	9.58	9.10	8.69	9.55
XC-RHF (λ=1.5)	10.03	9.91	9.74	9.33	8.70	8.26	9.42
XC-RHF (λ=2.0)	10.00	9.84	9.61	9.10	8.36	7.93	9.32
XC-RHF (λ=2.5)	9.96	9.77	9.49	8.88	8.06	7.67	9.24
XC-RHF (λ=5.0)	9.79	9.41	8.92	8.02	7.08	6.86	8.95
XC-RHF (λ=7.5)	9.62	9.09	8.43	7.40	6.54	6.43	8.75
XC-RHF (λ=10.0)	9.46	8.78	8.00	6.93	6.23	6.16	8.59

TABLE C.5: Mean Absolute Deviations between the CCSD electron density and the charge distributions associated with all the other considered methods. For the sake of clarity, all the values are multiplied by 10^4

method	$(sin\theta/\lambda)_{max}(\text{Å}^{-1})$						
	2.0	1.5	1.2	0.9	0.7	0.5	0.25
CISD	0.24						
DFT/BLYP	0.91						
DFT/B3LYP	0.53						
DFT/VSXC	0.68						
DFT/B1B95	0.43						
RHF	1.29						
XC-RHF (λ=0.5)	1.29	1.28	1.28	1.26	1.22	1.18	1.25
XC-RHF (λ=1.0)	1.28	1.27	1.26	1.22	1.16	1.10	1.23
XC-RHF (λ=1.5)	1.28	1.27	1.24	1.19	1.10	1.04	1.21
XC-RHF (λ=2.0)	1.28	1.26	1.23	1.16	1.05	0.99	1.20
XC-RHF (λ=2.5)	1.27	1.25	1.21	1.13	1.01	0.94	1.18
XC-RHF (λ=5.0)	1.25	1.21	1.15	1.01	0.85	0.79	1.12
XC-RHF (λ=7.5)	1.24	1.17	1.09	0.92	0.74	0.70	1.08
XC-RHF (λ=10.0)	1.22	1.14	1.03	0.85	0.66	0.64	1.04

Appendix D

A theoretical study of the Biscarbonyl[14]annulene

TABLE D.1: Wiberg indices for the C-C bonds of the BCA molecule at ambient pressure and at 7.7 GPa. The indices are obtained from CCSD calculations with the 6-31G basis-set.

bonds	Ambient Pressure		High Pressure	
C1-C2 / C8-C9	1.29	1.30	1.22	1.41
C2-C3 / C9-C10	1.52	1.51	1.59	1.38
C3-C4 / C10-C11	1.32	1.32	1.24	1.44
C4-C5 / C11-C12	1.51	1.51	1.61	1.39
C5-C6 / C12-C13	1.29	1.32	1.19	1.40
C6-C7 / C13-C14	1.39	1.39	1.51	1.28
C7-C8 / C14-C1	1.39	1.40	1.28	1.49

TABLE D.2: Wiberg indices for the C-C bonds of the BCA molecule at ambient pressure and at 7.7 GPa. The indices are obtained from XC-RHF calculations with the 6-31G basis-set.

bonds	Ambient Pressure		High Pressure	
C1-C2 / C8-C9	1.25	1.30	1.25	1.32
C2-C3 / C9-C10	1.54	1.51	1.56	1.44
C3-C4 / C10-C11	1.28	1.31	1.25	1.34
C4-C5 / C11-C12	1.54	1.52	1.57	1.47
C5-C6 / C12-C13	1.27	1.29	1.21	1.30
C6-C7 / C13-C14	1.41	1.38	1.42	1.38
C7-C8 / C14-C1	1.38	1.43	1.32	1.37

TABLE D.3: Wiberg indices for the C-C bonds of the BCA molecule at ambient pressure and at 7.7 GPa. The indices are obtained from CCSD calculations with the 6-31G(d,p) basis-set.

bonds	Ambient Pressure		High Pressure	
C1-C2 / C8-C9	1.29	1.30	1.22	1.42
C2-C3 / C9-C10	1.52	1.52	1.61	1.38
C3-C4 / C10-C11	1.32	1.32	1.23	1.45
C4-C5 / C11-C12	1.52	1.52	1.63	1.39
C5-C6 / C12-C13	1.30	1.30	1.19	1.41
C6-C7 / C13-C14	1.40	1.39	1.52	1.28
C7-C8 / C14-C1	1.39	1.40	1.28	1.50

TABLE D.4: Wiberg indices for the C-C bonds of the BCA molecule at ambient pressure and at 7.7 GPa. The indices are obtained from XC-RHF calculations with the 6-31G(d,p) basis-set.

bonds	Ambient Pressure		High Pressure	
C1-C2 / C8-C9	1.26	1.29	1.23	1.36
C2-C3 / C9-C10	1.54	1.52	1.58	1.42
C3-C4 / C10-C11	1.29	1.31	1.26	1.40
C4-C5 / C11-C12	1.53	1.53	1.58	1.42
C5-C6 / C12-C13	1.28	1.29	1.23	1.36
C6-C7 / C13-C14	1.41	1.38	1.45	1.33
C7-C8 / C14-C1	1.38	1.42	1.32	1.44

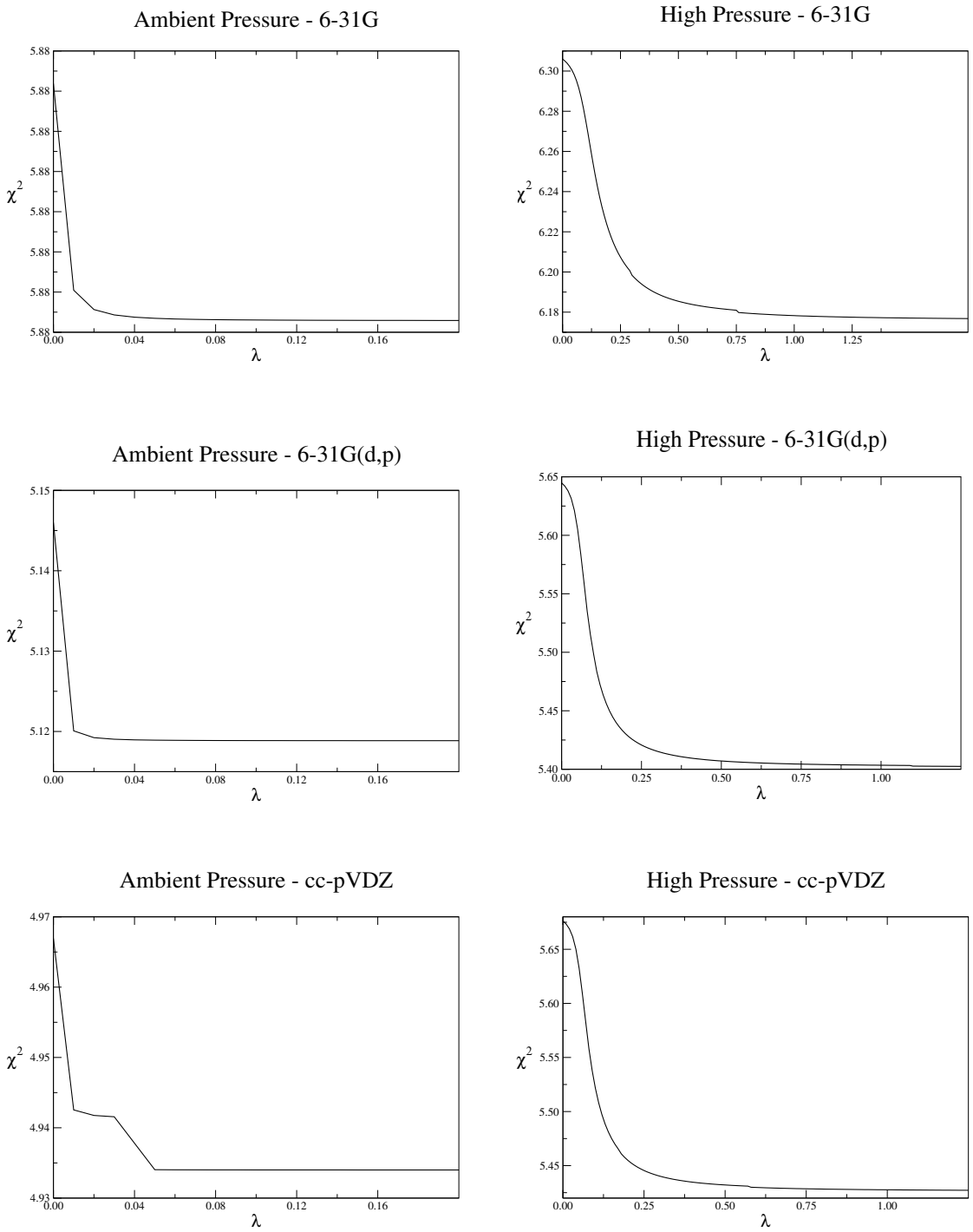

FIGURE D.1: Variation of the statistical agreement χ^2 in function of the external multiplier λ for the X-ray constrained ELMO-VB calculations performed at ambient and high pressures.

List of Figures

1.1 Schematic comparison of different scaling behaviors 12
1.2 Partition of a subsytem into core and buffer regions 22

2.1 Isosurfaces for an LMO (left, obtained through the Boys localization technique) and for an ELMO (right, obtained through the Stoll technique) describing one of the C-H bonds of the N-ethanalmethanamide molecule (Isovalue set equal to $0.03e/bohr^3$). 38
2.2 Localization scheme (left) and block-structured matrix of the LMO coefficients (right) for the ammonia molecule. In the localization scheme, the three overlapping bond fragments N-H1, N-H2 and N-H3 are explicitly shown, while, for the sake of clarity, the atomic fragment N, which describes the core and lone-pair electrons of the nitrogen atom, is not depicted. 39
2.3 Definition of the reference frames for the rotation from geometry of the model molecule to the geometry of the target system. 44
2.4 Block-diagonal structure of the rotation matrix **R** for an ELMO localized on a C-H bond when the split-valence basis set 6-311G is used. The blue blocks transform the carbon atom basis functions (and the related coefficients), while the red ones rotate the atomic orbitals (and the corresponding coefficients) of the hydrogen atom. The **S**, and **P** blocks are 1×1 and 3×3 submatrices, respectively, while the overall rotation matrix R has dimension 16×16. 47

3.1 Model molecules approximations. The fragment in exam is framed in orange. 53
3.2 Extension of the model molecules approximations in the case of subunits involved in intermolecular interactions. The fragment in exam is framed in orange. 54
3.3 Three-atom fragments used to describe (A) the σ and π electrons of the carboxylate group, (B) the electrons involved in the peptide bonds, and (C) the delocalized π electron pairs of the phenyl rings. 56

4.1 Schematic representation of the Bragg law. 78
4.2 Leu-enkephalin pentapeptide and interacting water molecules. The five hydrogen-bond interactions and the corresponding bond critical points are explicitly shown (green spheres). 85

4.3 Kinetic energy density at the hydrogen-bond critical points (kJ.mol.Å^{-3}) in function of the experimental distance $d(H \cdots O)$ in Å. The values are obtained through the Espinosa empirical relation and after the transfer of pseudoatoms (ELMAM2 and UBDB databases) and ELMOs (6-311, 6-311G(d,p) and 6-311+G(2d,2p) basis sets). 95
4.4 Sectional view of the ELMAM2 electron density isosurfaces of the Leu-enkephalin pentapeptide and interacting water molecules, corresponding to the shell limits selected for the Walter-Mezey similarity index. The 3, 1, 0.01, 0.001 e/Å^{-3} isosurfaces are depicted in light blue, violet, green and red, respectively. 100
4.5 Model molecule for peptide bonds (A) in the general case and (B) in the presence of proline. 106
4.6 Molécule modèle utilisée pour décrire la liaison peptidique (A) dans un cas général et (B) en présence d'un résidu proline. 110

5.1 Schematic representation of the fictitious non-interacting crystal and of the corresponding real interacting crystal. 117
5.2 Resonance structures of the benzene molecule. 128

6.1 L-Alanine molecular geometry obtained from the X-ray diffraction experiment. 133
6.2 Normalized residuals of the structure factor amplitudes in function of the scattering resolution for the (A) XC-RHF/3-21G, (B) XC-ELMO/3-21G, (C) XC-RHF/6-311G(d,p), (D) XC-ELMO/6-311G(d,p), (E) XC-RHF/6-311++G(2d,2p), and (F) XC-ELMO/6-311++G(2d,2p) calculations. The horizontal dashed line represents the outlier limit. The systematic outliers occurring at $sin\theta/\lambda \approx 1.0$Å^{-1} are encircled. 137
6.3 Variation of the χ^2 agreement statistics in function of the external multiplier λ for the XC-RHF and the XC-ELMO methods when the (A) 3-21G, (B) 6-311G(d,p), and (C) 6-311++G(2d,2p) basis-sets are considered. The solid blue and the dashed red lines represent the XC-RHF and XC-ELMO trends, respectively. The horizontal dash-and-dot line shows the desired statistical agreement. 138

7.1 Absolute differences between the CCSD and the RHF structure factor amplitudes as a function of the reciprocal resolution for the N_2 molecule. . . . 150
7.2 Representative isosurfaces of the detachment and attachment densities (in orange and blue, respectively) relative to electronic rearrangements with respect to the RHF charge distribution when different correlated methods are taken into account. 158

8.1 *syn*-1,6:8,13-Biscarbonyl[14] annulene (BCA) molecular geometry obtained from the X-ray diffraction experiment at 7.7 GPa. 164

8.2 Schematic representation of the BCA and its C-C equivalent bond lengths at ambient-pressure and their differences at high pressure. 165
8.3 Resonance structures of the *syn*-1,6:8,13-Biscarbonyl[14] annulene. 165
8.4 Schematic representation of the three different localization schemes adopted for the ELMO calculations. The ELMO configurations A and B correspond to the two resonance structures in Figure 8.3 (7 strictly localized MOs for 14 π electrons) and D is the delocalized configuration (7 delocalized MOs on the 14 carbon atoms of the ring). 166
8.5 Diagram that qualitatively show the unconstrained ELMO energies (and the associated differences) at ambient and high pressures. 169
8.6 Variation of the Chirgwin-Coulson coefficients in function of the external multiplier λ for the X-ray constrained "ELMO-VB" calculations performed at ambient and high pressures. 172

List of Tables

4.3 Values of the electron distribution at the non-covalent bond critical points ($e/Å^3$) after the transfer of pseudoatoms and ELMOs. 90
4.4 Values of the Laplacian of the electron density at the non-covalent bond critical points ($e/Å^5$) after the transfer of pseudoatoms and ELMOs. 91
4.5 Values of the agreement index R (%) between the different reconstruction methods considering as topological property the electron density at the covalent bond critical points. 92
4.6 Values of the agreement index R (%) between the different reconstruction methods considering as topological property the Laplacian of the electron density at the covalent bond critical points. 92
4.7 Values of the agreement index R (%) between the different reconstruction methods considering as topological property the electron density at the non-covalent bond critical points. 93
4.8 Values of the agreement index R (%) between the different reconstruction methods considering as topological property the Laplacian of the electron density at the non-covalent bond critical points. 93
4.9 Values of the kinetic energy density at the hydrogen-bond critical points ($kJ.mol.Å^{-3}$) obtained through the Espinosa empirical relation and after the transfer of pseudoatoms and ELMOs. 94
4.10 Values of the potential energy density at the hydrogen-bond critical points ($kJ.mol.Å^{-3}$) obtained through the Espinosa empirical relation and after the transfer of pseudoatoms and ELMOs. 94
4.11 Values of the positive curvature λ_3 of the electron density at the hydrogen-bond critical points ($e/Å^5$) obtained through the Espinosa empirical relation and after the transfer of pseudoatoms and ELMOs. 95
4.12 Net QTAIM integrated charges (e) for the backbone "functional groups" the five side chains (including the $C_\alpha H_\alpha$ bonds) of the Leu-enkephalin polypeptide. 97
4.13 Values of the agreement index R (%) between the different reconstruction methods considering as topological property the net QTAIM integrated charges (e) for the backbone "functional groups" and for the five side chains of the Leu-enkephalin polypeptide. 98

4.15 Point-by-point comparison of the different electron densities: values of the real-space R similarity index (%). 100

6.1 Agreement statistics χ^2 and energy values corresponding to the RHF, ELMO, XC-RHF and XC-ELMO calculations performed on the L-Alanine system. . 135
6.2 Unconstrained and X-ray constrained relative variations (U-RV and XC-RV, respectively) of the charge density values at the L-Alanine BCPs. 140
6.3 Unconstrained and X-ray constrained relative variations (U-RV and XC-RV, respectively) of the Laplacian values of the charge density at the L-Alanine BCPs. 141
6.4 Unconstrained and X-ray constrained relative variations (U-RV and XC-RV, respectively) of the net QTAIM integrated charges (e) for the main functional groups of the L-Alanine system. 142
6.5 Root Mean Square Deviation and Mean Absolute Error of the net atomic charges (e), the electron density ($e/Å^3$) and its Laplacian ($e/Å^5$) at the bond critical points using the RHF and the XC-RHF values as references. For the sake of clarity, all the values are multiplied by 100. 143

7.1 Electron density at the bond critical point: percentage of the CCSD/RHF difference recovered by the considered correlated methods for the N_2 system. 152
7.2 Laplacian of the electron density at the bond critical point: percentage of the CCSD/RHF difference recovered by the considered correlated methods for the N_2 system. 153
7.3 Values of the RSR similarity index between the CCSD electron distribution and the charge densities associated with all the other considered methods. For the sake of clarity, all the RSR values are multiplied by 10. 155
7.4 Values of Carbó Euclidean distances between the CCSD electron distribution and the charge densities associated with all the other considered methods. For the sake of clarity, all the values are multiplied by 100. 157

8.1 Differences (in kcal/mol) between the ELMO energies at high-pressure (7.7 GPa) and at ambient-pressure for the three localization schemes taken into account. 168
8.2 Differences (in kcal/mol) between the ELMO energies obtained with the localization schemes A and B both at ambient and high (7.7 GPa) pressures. 168
8.3 Statistical agreement χ^2 and Chirgwin-Coulson coefficients obtained through unconstrained "ELMO-VB" calculations at ambient and high pressures. . . . 170
8.4 Statistical agreement χ^2 and Chirgwin-Coulson coefficients obtained through X-ray constrained "ELMO-VB" calculations at ambient and high pressures. . 171
8.5 Wiberg indices for the C-C bonds of the BCA molecule at ambient pressure and at 7.7 GPa. The indices are obtained from CCSD calculations with the cc-pVDZ basis-set. 174

8.6 Wiberg indices for the C-C bonds of the BCA molecule at ambient pressure and at 7.7 GPa. The indices are obtained from XC-RHF calculations with the cc-pVDZ basis-set. 175

A.1 Effects of the Molecular Orbitals Localization: Mean Relative Variations (in percentage) of the values of the Laplacian of the electron density at the bond critical points. Hartree-Fock values are used as reference. 183
A.2 Effects of the Molecular Orbitals Localization: maximum relative variations (in percentage) of the values of the electron density at the bond critical points. Hartree-Fock values are used as reference. 184
A.3 Effects of the Molecular Orbitals Localization: minimum relative variations (in percentage) of the values of the electron density at the bond critical points. Hartree-Fock values are used as reference. 184
A.4 Effects of the Molecular Orbitals Localization: Mean Relative Variations (in percentage) of the values of the Laplacian of the electron density at the bond critical points. Hartree-Fock values are used as reference. 185
A.5 Effects of the Molecular Orbitals Localization: maximum relative variations (in percentage) of the values of the Laplacian of the electron density at the bond critical points. Hartree-Fock values are used as reference. 185
A.6 Effects of the Molecular Orbitals Localization: minimum relative variations (in percentage) of the values of the Laplacian of the electron density at the bond critical points. Hartree-Fock values are used as reference. 186
A.7 Effects of the Molecular Orbitals Localization: Mean Deviations (in e) of the net integrated atomic charges. Hartree-Fock values are used as reference. 186
A.8 Effects of the Molecular Orbitals Localization: maximum deviations (in e) of the net integrated atomic charges. Hartree-Fock values are used as reference. 187
A.9 Effects of the Molecular Orbitals Localization: minimum deviations (in e) of the net integrated atomic charges. Hartree-Fock values are used as reference. 187
A.10 Number of considered molecular orbitals in the computations of the average energy variations for the comparison of the Nearest Atom (NAA), Nearest Bond (NBA), and Nearest Functional Group (NFGA) Approximations. 188
A.11 Number of considered bond critical points in the computations of the mean absolute relative variations. 189
A.12 Number of considered atoms in the computations of the mean absolute variations. 189

B.1 Standard deviations associated with the average values of the electron distributions at the covalent bond critical points ($e/Å^3$) for each bond type after the transfer of pseudoatoms and ELMOs. (For $(C\text{-}N)_{term}$ and $(C\text{-}O)_{phenol}$ only one BCP is considered.) 191

B.2 Standard deviations associated with the average values of the Laplacian of the electron distributions at the covalent bond critical points ($e/Å^5$) for each bond type after the transfer of pseudoatoms and ELMOs. (For $(C\text{-}N)_{term}$ and $(C\text{-}O)_{phenol}$ only one BCP is considered.) . . . 192
B.3 Average values of the electron distributions at the covalent bond critical points ($e/Å^3$) for each bond type after the transfer of pseudoatoms and ELMOs. (For $(C\text{-}N)_{term}$ and $(C\text{-}O)_{phenol}$ only one BCP is considered.) . . . 192
B.4 Average values of the Laplacian of the electron density at the covalent bond critical points ($e/Å^5$) for each bond type after the transfer of pseudoatoms and ELMOs. (For $(C\text{-}N)_{term}$ and $(C\text{-}O)_{phenol}$ only one BCP is considered.) . . 193
B.5 Standard deviations associated with the average values of the electron distributions at the covalent bond critical points ($e/Å^3$) for each bond type after the transfer of pseudoatoms and ELMOs. (For $(C\text{-}N)_{term}$ and $(C\text{-}O)_{phenol}$ only one BCP is considered.) . . . 193
B.6 Standard deviations associated with the average values of the Laplacian of the electron distributions at the covalent bond critical points ($e/Å^5$) for each bond type after the transfer of pseudoatoms and ELMOs. (For $(C\text{-}N)_{term}$ and $(C\text{-}O)_{phenol}$ only one BCP is considered.) . . . 194
B.7 Values of the electron distribution at the non-covalent bond critical points ($e/Å^3$) after the transfer of pseudoatoms and ELMOs. . . . 194
B.8 Values of the Laplacian of the electron density at the non-covalent bond critical points ($e/Å^5$) after the transfer of pseudoatoms and ELMOs. . . . 195
B.9 Values of the kinetic energy density at the hydrogen-bond critical points ($kJ.mol.Å^{-3}$) obtained through the Espinosa empirical relation and after the transfer of pseudoatoms and ELMOs. . . . 195
B.10 Values of the potential energy density at the hydrogen-bond critical points ($kJ.mol.Å^{-3}$) obtained through the Espinosa empirical relation and after the transfer of pseudoatoms and ELMOs. . . . 196
B.11 Values of the positive curvature λ_3 of the electron density at the hydrogen-bond critical points ($e/Å^5$) obtained through the Espinosa empirical relation and after the transfer of pseudoatoms and ELMOs. . . . 196
B.12 Net QTAIM integrated charges (e) for the backbone "functional groups" the five side chains (including the $C_\alpha H_\alpha$ bonds) of the Leu-enkephalin polypeptide. . . . 197
B.13 Values of the agreement index R (%) between the different reconstruction methods considering as topological property the electron density at the covalent bond critical points . . . 197
B.14 Values of the agreement index R (%) between the different reconstruction methods considering as topological property the Laplacian of the electron density at the covalent bond critical points . . . 198

B.15 Values of the agreement index R (%) between the different reconstruction methods considering as topological property the electron density at the non-covalent bond critical points . . . 198
B.16 Values of the agreement index R (%) between the different reconstruction methods considering as topological property the Laplacian of the electron density at the non-covalent bond critical points . . . 198
B.17 Values of the agreement index R (%) between the different reconstruction methods considering as topological property the net QTAIM integrated charges (e) for the backbone "functional groups" and for the five side chains of the Leu-enkephalin polypeptide. . . . 199
B.19 Point-by-point comparison of the different electron densities: Values of the real-space R similarity index (%). . . . 200

C.1 Total number of structure factor amplitudes and number of significant discrepancies (> 0.005) between the CCSD and the RHF structure factor amplitudes for each resolution range taken into account. . . . 203
C.2 Values of the electron density (e/Å^3)) at the bond critical point of the nitrogen molecule for all the considered methods. . . . 204
C.3 Values of the Laplacian of the electron density (e/Å^5)) at the bond critical point of the nitrogen molecule for all the considered methods. . . . 204
C.4 Root Mean Square Deviations between the CCSD electron density and the charge distributions associated with all the other considered methods. For the sake of clarity, all the values are multiplied by 10^4 . . . 205
C.5 Mean Absolute Deviations between the CCSD electron density and the charge distributions associated with all the other considered methods. For the sake of clarity, all the values are multiplied by 10^4 . . . 205

D.1 Wiberg indices for the C-C bonds of the BCA molecule at ambient pressure and at 7.7 GPa. The indices are obtained from CCSD calculations with the 6-31G basis-set. . . . 207
D.2 Wiberg indices for the C-C bonds of the BCA molecule at ambient pressure and at 7.7 GPa. The indices are obtained from XC-RHF calculations with the 6-31G basis-set. . . . 208
D.3 Wiberg indices for the C-C bonds of the BCA molecule at ambient pressure and at 7.7 GPa. The indices are obtained from CCSD calculations with the 6-31G(d,p) basis-set. . . . 208
D.4 Wiberg indices for the C-C bonds of the BCA molecule at ambient pressure and at 7.7 GPa. The indices are obtained from XC-RHF calculations with the 6-31G(d,p) basis-set. . . . 209

List of Publications

1. A. Genoni and **B. Meyer**, *X-Ray Constrained Wave Functions: Fundamentals and Effects of the Molecular Orbitals Localization*, Adv. Quantum Chem. **73**, 333-362 (2016).

2. **B. Meyer**, B. Guillot, M. F. Ruiz-Lopez, C. Jelsch and A. Genoni, *Libraries of Extremely Localized Molecular Orbitals. 2. Comparison with the Pseudoatoms Transferability*, J. Chem. Theory Comput. **12**, 1068-1081 (2016).

3. **B. Meyer**, B. Guillot, M. F. Ruiz-Lopez and A. Genoni, *Libraries of Extremely Localized Molecular Orbitals. 1. Model Molecules Approximation and Molecular Orbitals Transferability*, J. Chem. Theory Comput. **12**, 1052-1067 (2016).

4. **B. Meyer**, A. Genoni, A. Boudier, P. Leroy and M. F. Ruiz-Lopez, *Structure and stability studies of pharmacologiaclly relevant* S*-nitrosothiols: a theoretical approach*, J. Phys. Chem. A, 120, 4191-4200 (2016).

Bibliography

1. D. R. Hartree, *Mathematical Proceedings of the Cambridge Philosophical Society* **24**, 111–132 (1928).
2. D. R. Hartree, *Mathematical Proceedings of the Cambridge Philosophical Society* **24**, 89–110 (1928).
3. D. R. Hartree, *Mathematical Proceedings of the Cambridge Philosophical Society* **24**, 426–437 (1928).
4. D. R. Hartree, *Mathematical Proceedings of the Cambridge Philosophical Society* **25**, 310–314 (1929).
5. V. Fock, *Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems* **25**, 129–148 (1930).
6. W. Kohn, L. J. Sham, *Physical Review A* **140**, 1133–1138 (1965).
7. P. Hohenberg, W. Kohn, *Physical review* **136**, 864–871 (1964).
8. C. Van Alsenoy, C.-H. Yu, A. Peeters, J. M. Martin, L. Schäfer, *The Journal of Physical Chemistry A* **102**, 2246–2251 (1998).
9. G. E. Scuseria, *The Journal of Physical Chemistry A* **103**, 4782–4790 (1999).
10. F. Sato, T. Yoshihiro, M. Era, H. Kashiwagi, *Chemical physics letters* **341**, 645–651 (2001).
11. T. Inaba, S. Tahara, N. Nisikawa, H. Kashiwagi, F. Sato, *Journal of Computational Chemistry* **26**, 987–993 (2005).
12. T. Inaba, F. Sato, *Journal of Computational Chemistry* **28**, 984–995 (2007).
13. S. Goedecker, *Reviews of Modern Physics* **71**, 1085–1123 (1999).
14. D. R. Bowler, T Miyazaki, *Reports on Progress in Physics* **75**, 1–43 (2012).
15. C. Ochsenfeld, J. Kussmann, D. S. Lambrecht, *Reviews in computational chemistry* **23**, 1–82 (2007).
16. J. Kussmann, M. Beer, C. Ochsenfeld, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **3**, 614–636 (2013).

17. W. Yang, *Physical Review A* **44**, 7823–7826 (1991).

18. W. Yang, *Physical Review Letters* **66**, 1438–1441 (1991).

19. W. Yang, *Journal of Molecular Structure (Theochem)* **255**, 461–479 (1992).

20. D. York, J. P. Lu, W. Yang, *Physical Review B* **49**, 8526–8528 (1994).

21. W. Yang, T.-S. Lee, *The Journal of Chemical Physics* **103**, 5674–5678 (1995).

22. T. Lee, D. York, W. Yang, *The Journal of Chemical Physics* **105**, 2744–2750 (1996).

23. D. York, T.-S. Lee, W. Yang, *Journal of American Chemical Society* **118**, 10940–10941 (1996).

24. D. York, T.-S. Lee, W. Yang, *Chemical Physics Letters* **263**, 297–304 (1996).

25. S. L. Dixon, K. M. Merz, *The Journal of Chemical Physics* **104**, 6643–6649 (1996).

26. S. L. Dixon, K. M. Merz, *The Journal of Chemical Physics* **107**, 879–893 (1997).

27. V. Gogonea, K. M. Merz, *The Journal of Physical Chemistry A* **103**, 5171–5188 (1999).

28. D. W. Zhang, J. Z. H. Zhang, *The Journal of Chemical Physics* **119**, 3599–3605 (2003).

29. D. W. Zhang, Y. Xiang, J. Z. H. Zhang, *The Journal of Physical Chemistry B* **107**, 12039–12041 (2003).

30. S. Li, W. Li, T. Fang, *Journal of the American Chemical Society* **127**, 7215–7226 (2005).

31. D. G. Fedorov, K. Kitaura, *The Journal of Chemical Physics* **121**, 2483–2490 (2004).

32. D. G. Fedorov, K. Kitaura, *Journal of Computational Chemistry* **28**, 222–237 (2007).

33. D. G. Fedorov, K. Kitaura, *The Journal of Physical Chemistry A* **111**, 6904–6914 (2007).

34. D. G. Fedorov, R. M. Olson, K. Kitaura, M. S. Gordon, S. Koseki, *Journal of computational chemistry* **25**, 872–880 (2004).

35. T. Nakano *et al.*, *Chemical Physics Letters* **351**, 475–480 (2002).

36. K. Kitaura, E. Ikeo, T. Asada, T. Nakano, M. Uebayasi, *Chemical Physics Letters* **313**, 701–706 (1999).

37. T. Nakano *et al.*, *Chemical Physics Letters* **318**, 614–618 (2000).

38. L. Huang, H. J. Bohorquez, C. F. Matta, L. Massa, *International Journal of Quantum Chemistry* **111**, 4150–4157 (2011).

39. L. Huang, L. Massa, J. Karle, *International Journal of Quantum Chemistry* **106**, 447–457 (2006).

40. L. Huang, L. Massa, J. Karle, *Proceedings of the National Academy of Sciences of the United States of America* **103**, 1233–1237 (2006).

41. L. Huang, L. Massa, J. Karle, *Biochemistry* **44**, 16747–16752 (2005).

42. M. J. Timm, C. F. Matta, L. Massa, L. Huang, *The Journal of Physical Chemistry A* **118**, 11304–11316 (2014).

43. L. Huang, C. Matta, L. Massa, *Structural Chemistry* **26**, 1433–1442 (2015).

44. R. F. W. Bader, *Atoms in Molecules: A Quantum Theory*, Oxford University Press (1990).

45. P. D. Walker, P. G. Mezey, *Journal of the American Chemical Society* **115**, 12423–12430 (1993).

46. P. D. Walker, P. G. Mezey, *Journal of the American Chemical Society* **116**, 12022–12032 (1994).

47. P. D. Walker, P. G. Mezey, *Canadian Journal of Chemistry* **72**, 2531–2536 (1994).

48. P. D. Walker, P. G. Mezey, *Journal of mathematical chemistry* **17**, 203–234 (1995).

49. P. D. Walker, P. G. Mezey, *Journal of computational chemistry* **16**, 1238–1249 (1995).

50. T. E. Exner, P. G. Mezey, *The Journal of Physical Chemistry A* **106**, 5504–5509 (2002).

51. T. E. Exner, P. G. Mezey, *The Journal of Physical Chemistry A* **106**, 11791–11800 (2002).

52. T. E. Exner, P. G. Mezey, *Journal of computational chemistry* **24**, 1980–1986 (2003).

53. T. E. Exner, P. G. Mezey, *The Journal of Physical Chemistry A* **108**, 4301–4309 (2004).

55. C. M. Breneman, M. Rhem, *Journal of computational chemistry* **18**, 182–197 (1997).

56. K. Babu, S. R. Gadre, *Journal of Computational Chemistry* **24**, 484–495 (2003).

57. K. Babu, V. Ganesh, S. R. Gadre, N. E. Ghermani, *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)* **111**, 255–263 (2004).

58. S. R. Gadre, V. Ganesh, *Journal of Theoretical and Computational Chemistry* **5**, 835–855 (2006).

59. A. P. Rahalkar, M. Katouda, S. R. Gadre, S. Nagase, *Journal of Computational Chemistry*, 2405–2418 (2010).

60. N. Sahu, S. R. Gadre, *Accounts of Chemical Research* **47**, 2739–2747 (2014).

61. S. R. Gadre, R. N. Shirsat, A. C. Limaye, *The Journal of Physical Chemistry* **98**, 9165–9169 (199).

62. P. G. Mezey, *Journal of mathematical chemistry* **18**, 141–168 (1995).

63. P. G. Mezey, *Structural Chemistry* **6**, 261–270 (1995).

64. J. M. Foster, S. Boys, *Reviews of Modern Physics* **32**, 300–302 (1960).

65. S. F. Boys, *Reviews of Modern Physics* **32**, 296–299 (1960).

66. C. Edmiston, K. Ruedenberg, *Reviews of Modern Physics* **35**, 457–465 (1963).

67. C. Edmiston, K. Ruedenberg, *The Journal of Chemical Physics* **43**, S97–S116 (1965).

68. J. Pipek, P. G. Mezey, *International Journal of Quantum Chemistry* **34**, 1–13 (1988).

69. J. Pipek, P. G. Mezey, *The Journal of Chemical Physics* **90**, 4916–4926 (1989).

70. H. Stoll, G. Wagenblast, H. Preuss, *Theoretica chimica acta* **57**, 169–178 (1980).

71. D. M. Philipp, R. A. Friesner, *Journal of Computational Chemistry* **20**, 1468–1494 (1999).

72. W. von Niessen, *The Journal of Chemical Physics* **56**, 4290–4297 (1972).

73. V. Magnasco, *The Journal of Chemical Physics* **47**, 971–981 (1967).

74. R. McWeeny, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **253**, 242–259 (1959).

75. R. McWeeny, *Reviews of Modern Physics* **32**, 335–369 (1960).

76. M. Klessinger, R. McWeeny, *The Journal of Chemical Physics* **42**, 3343–3354 (1965).

77. W. H. Adams, *The Journal of Chemical Physics* **34**, 89–102 (1961).

78. T. L. Gilbert, *The Journal of Chemical Physics* **60**, 3835–3844 (1974).

79. S. Huzinaga, *The Journal of Chemical Physics* **55**, 5543–5549 (1971).

80. O. Matsuoka, *The Journal of Chemical Physics* **66**, 1245–1254 (1977).

81. G. F. Smits, C. Altona, *Theoretica chimica acta* **67**, 461–475 (1985).

82. E. Francisco, A. Martín Pendás, W. H. Adams, *The Journal of Chemical Physics* **97**, 6504–6508 (1992).

83. P. Ordejón, D. A. Drabold, M. P. Grumbach, R. M. Martin, *Physical Review B* **48**, 14646–14649 (1993).

84. M. Couty, C. A. Bayse, M. B. Hall, *Theoretical Chemistry Accounts* **97**, 96–109 (1997).

85. A. Fornili, M. Sironi, M. Raimondi, *Journal of Molecular Structure: THEOCHEM* **632**, 157–172 (2003).

86. Z. Szekeres, P. R. Surján, *Chemical Physics Letters* **369**, 125–130 (2003).

87. K. R. Sundberg, J. Bicerano, W. N. Lipscomb, *The Journal of Chemical Physics* **71**, 1515–1524 (1979).

88. W. Press, B. Flannery, S. Teukolsky, W. Vetterling, *Cambridge University Press, Cambridge* (1986).

89. N. Ferré, X. Assfeld, J.-L. Rivail, *Journal of Computational Chemistry* **23**, 610–624 (2002).

90. B. Meyer, B. Guillot, M. F. Ruiz-Lopez, A. Genoni, *Journal of Chemical Theory and Computation* **12**, 1052–1067 (2016).

91. M. Ahmed, C. Jelsch, B. Guillot, C. Lecomte, S. Domagała, *Crystal Growth & Design* **13**, 315–325 (2013).

92. M. Ahmed *et al.*, *The Journal of Physical Chemistry A* **117**, 14267–14275 (2013).

93. J. M. Bąk *et al.*, *Acta Crystallographica Section A Foundations of Crystallography* **67**, 141–153 (2011).

94. Y. Bibila Mayaya Bisseyou *et al.*, *Acta Crystallographica Section B: Structural Science* **68**, 646–660 (2012).

95. S. Domagała, B. Fournier, D. Liebschner, B. Guillot, C. Jelsch, *Acta Crystallographica Section A Foundations of Crystallography* **68**, 337–351 (2012).

96. T. Koritsanszky, X. Li, P. Coppens, *Acta Crystallographica Section A Foundations of Crystallography* **60**, 638–639 (2004).

97. C. Lecomte, C. Jelsch, B. Guillot, B. Fournier, A. Lagoutte, *Journal of Synchrotron Radiation* **15**, 202–203 (2008).

98. D. Liebschner *et al.*, *The Journal of Physical Chemistry A* **115**, 12895–12904 (2011).

99. V. Pichon-Pesme, C. Jelsch, B. Guillot, C. Lecomte, *Acta Crystallographica Section A Foundations of Crystallography* **60**, 204–208 (2004).

100. A. Poulain-Paul *et al.*, *Acta Crystallographica Section A: Foundations of Crystallography* **68**, 715–728 (2012).

101. B. Zarychta, V. Pichon-Pesme, B. Guillot, C. Lecomte, C. Jelsch, *Acta Crystallographica Section A Foundations of Crystallography* **63**, 108–125 (2007).

102. B. Dittrich, C. B. Hübschle, J. J. Holstein, F. P. A. Fabbiani, *Journal of Applied Crystallography* **42**, 1110–1121 (2009).

103. B. Dittrich, C. B. Hübschle, P. Luger, M. A. Spackman, *Acta Crystallographica Section D Biological Crystallography* **62**, 1325–1335 (2006).

104. B. Dittrich *et al.*, *Acta Crystallographica Section A Foundations of Crystallography* **61**, 314–320 (2005).

105. B. Dittrich *et al.*, *Acta Crystallographica Section B Structural Science, Crystal Engineering and Materials* **69**, 91–104 (2013).

106. B. Dittrich, M. A. Spackman, *Acta Crystallographica Section A Foundations of Crystallography* **63**, 426–436 (2007).

107. B. Dittrich, M. Strumpel, M. Schäfer, M. A. Spackman, T. Koritsánszky, *Acta Crystallographica Section A Foundations of Crystallography* **62**, 217–223 (2006).

108. B. Dittrich, T. Koritsánszky, P. Luger, *Angewandte Chemie International Edition* **43**, 2718–2721 (2004).

109. B. Dittrich, P. Munshi, M. A. Spackman, *Acta Crystallographica Section C Crystal Structure Communications* **62**, 633–635 (2006).

110. B. Dittrich *et al.*, *ChemPhysChem* **16**, 412–419 (2015).

111. P. M. Dominiak, A. Volkov, A. P. Dominiak, K. N. Jarzembska, P. Coppens, *Acta Crystallographica Section D Biological Crystallography* **65**, 485–499 (2009).

112. P. M. Dominiak, A. Volkov, X. Li, M. Messerschmidt, P. Coppens, *Journal of Chemical Theory and Computation* **3**, 232–247 (2007).

113. T. Koritsanszky, A. Volkov, P. Coppens, *Acta Crystallographica Section A: Foundations of Crystallography* **58**, 464–472 (2002).

114. T. S. Koritsanszky, P. Coppens, *Chemical Reviews* **101**, 1583–1628 (2001).

115. A. Volkov, X. Li, T. Koritsanszky, P. Coppens, *The Journal of Physical Chemistry A* **108**, 4283–4300 (2004).

116. A. Volkov, Y. Abramov, P. Coppens, C. Gatti, *Acta Crystallographica Section A: Foundations of Crystallography* **56**, 332–339 (2000).

117. A. Volkov, M. Messerschmidt, P. Coppens, *Acta Crystallographica Section D Biological Crystallography* **63**, 160–170 (2007).

118. M. Sironi, M. Ghitti, A. Genoni, G. Saladino, S. Pieraccini, *Journal of Molecular Structure: THEOCHEM* **898**, 8–16 (2009).

119. V. Pichon-Pesme, C. Lecomte, H. Lachekar, *The Journal of Physical Chemistry* **99**, 6242–6250 (1995).

120. F. H. Allen, I. J. Bruno, *Acta Crystallographica Section B Structural Science* **66**, 380–386 (2010).

121. F. Martyn *et al.*, *Molecular Physics* **103**, 719–747 (2005).

122. M. J. Frisch *et al.*, *Gaussian: Wallingford* **66**, 380–386 (2009).

123. T. A. Jones, J.-Y. Zou, S. t. Cowan, M. Kjeldgaard, *Acta Crystallographica Section A: Foundations of Crystallography* **47**, 110–119 (1991).

124. V. Tognetti, L. Joubert, *The Journal of Physical Chemistry A* **115**, 5505–5515 (2011).

125. M. Messerschmidt, S. Scheins, P. Luger, *Acta Crystallographica Section B Structural Science* **61**, 115–121 (2005).

126. B. Meyer, B. Guillot, M. F. Ruiz-Lopez, C. Jelsch, A. Genoni, *Journal of Chemical Theory and Computation* **12**, 1068–1081 (2016).

127. W. H. Bragg, W. L. Bragg, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **88**, 428–438 (1913).

128. R. F. Stewart, *The Journal of Chemical Physics* **51**, 4569–4577 (1969).

129. N. K. Hansen, P. Coppens, *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography* **34**, 909–921 (1978).

130. B. Dawson, *Proceedings of the Royal Society A* **298**, 255–288 (1967).

131. J. J. DeMarco, R. J. Weiss, *Physical Review Section A* **137**, 1896–1871 (1965).

132. F. L. Hirshfeld, *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography* **32**, 239–244 (1976).

133. C. P. Brock, J. D. Dunitz, F. L. Hirshfeld, *Acta Crystallographica Section B: Structural Science* **47**, 789–797 (1991).

134. C. Jelsch, V. Pichon-Pesme, C. Lecomte, A. Aubry, *Acta Crystallographica Section D: Biological Crystallography* **54**, 1306–1318 (1998).

135. S. Domagała, P. Munshi, M. Ahmed, B. Guillot, C. Jelsch, *Acta Crystallographica Section B Structural Science* **67**, 63–78 (2011).

136. A. Volkov, T. Koritsanszky, X. Li, P. Coppens, *Acta Crystallographica Section A Foundations of Crystallography* **60**, 638–639 (2004).

137. S. K. J. Johnas, B. Dittrich, A. Meents, M. Messerschmidt, E. F. Weckert, *Acta Crystallographica Section D Biological Crystallography* **65**, 284–293 (2009).

138. V. R. Hathwar *et al.*, *The Journal of Physical Chemistry A* **115**, 12852–12863 (2011).

139. V. R. Hathwar, T. S. Thakur, T. N. G. Row, G. R. Desiraju, *Crystal Growth & Design* **11**, 616–623 (2011).

140. B. Guillot, L. Viry, R. Guillot, C. Lecomte, C. Jelsch, *Journal of Applied Crystallography* **34**, 214–223 (2001).

141. C. Jelsch, B. Guillot, A. Lagoutte, C. Lecomte, *Journal of Applied Crystallography* **38**, 38–54 (2005).

142. T. A. Keith, *TK Gristmill Softwaren, Overland Park, KS, USA* (2012).

143. E. Espinosa *et al.*, *Journal of the American Chemical Society* **118**, 2501–2502 (1996).

144. E. Espinosa, M. Souhassou, H. Lachekar, C. Lecomte, *Acta Crystallographica Section B: Structural Science* **55**, 563–572 (1999).

145. E. Espinosa, I. Alkorta, J. Elguero, E. Molins, *The Journal of Chemical Physics* **117**, 5529–5542 (2002).

146. E. Espinosa, E. Molins, C. Lecomte, *Chemical Physics Letters* **285**, 170–173 (1998).

147. M. A. Spackman, *Crystal Growth & Design* **15**, 5624–5628 (2015).

148. A. Mukherji, K. M., *The Journal of Chemical Physics* **38**, 44–48 (1963).

149. Y. Rasiel, D. R. Whitman, *The Journal of Chemical Physics* **42**, 2124–2131 (1965).

150. D. P. Chong, W. B. Brown, *The Journal of Chemical Physics* **45**, 392–395 (1966).

151. W. L. Clinton, A. J. Galli, L. J. Massa, *Physical Review* **177**, 7–13 (1969).

152. W. L. Clinton, J. Nakhleh, F. Wunderlich, *Physical Review* **177**, 1–6 (1969).

153. W. L. Clinton, G. A. Henderson, J. V. Prestia, *Physical Review* **177**, 13–18 (1973).

154. W. L. Clinton, L. G. B., *Physical Review* **177**, 19–27 (1973).

155. W. L. Clinton, L. G. B., *Physical Review* **177**, 27–33 (1973).

156. W. L. Clinton, C. A. Frishberg, L. J. Massa, P. A. Oldfield, *International Journal of Quantum Chemistry* **7**, 505–514 (1973).

157. W. L. Clinton, L. J. Massa, *Physical Review Letters* **29**, 1363–1366 (1972).

158. L. M. Pecora, *Physical Review B* **33**, 5987–5993 (1986).

159. C. Frishberg, L. J. Massa, *Physical Review B* **24**, 7018–7024 (1981).

160. L. Massa, M. Goldberg, C Frishberg, R. F. Boehme, S. J. La Placa, *Physical Review Letters* **55**, 622–625 (1985).

161. S. T. Howard, J. P. Huke, P. R. Mallinson, C. S. Frampton, *Physical Review B* **49**, 7124–7136 (1994).

162. J. A. Snyder, E. D. Stevens, *Chemical physics letters* **313**, 293–298 (1999).

163. K. Tanaka, R. Makita, S. Funahashi, T. Komori, Zaw Win, *Acta Crystallographica Section A Foundations of Crystallography* **64**, 437–449 (2008).

164. K. Tanaka, *Acta Crystallographica Section A Foundations of Crystallography* **44**, 1002–1008 (1988).

165. M. P. Waller *et al.*, *Chemistry - A European Journal* **12**, 7603–7614 (2006).

166. D. E. Hibbs, S. T. Howard, J. P. Huke, M. P. Waller, *Physical Chemistry Chemical Physics* **7**, 1772–1778 (2005).

167. H. Schmider, V. H. Smith, W. Weyrich, *The Journal of Chemical Physics* **96**, 8986–8994 (1992).

168. J.-M. Gillet, P. J. Becker, *Journal of Physics and Chemistry of Solids* **65**, 2017–2023 (2004).

169. J.-M. Gillet, *Acta Crystallographica Section A Foundations of Crystallography* **63**, 234–238 (2007).

170. J.-M. Gillet, P. J. Becker, P. Cortona, *Physical Review B* **63**, 235115–1–235115–6 (2001).

171. P. Cassam-Chenaï, *International Journal of Quantum Chemistry* **54**, 201–210 (1995).

172. P. Cassam-Chenaï, S. K. Wollf, G. S. Chandler, B. N. Figgis, *International Journal of Quantum Chemistry* **60**, 667–680 (1996).

173. G. A. Henderson, R. K. Zimmermann, *The Journal of Chemical Physics* **65**, 619–622 (1976).

174. M. Levy, J. A. Goldstein, *Physical Review B* **35**, 7887–7890 (1987).

175. O. Gritsenko, V. Zhidomirov, *Dokl. Akad. Nauk. SSSR* **293**, 1162–1166 (1987).

176. Q. Zhao, R. G. Parr, *Physical Review A* **46**, 2337–2343 (1992).

177. Q. Zhao, R. G. Parr, *The Journal of Chemical Physics* **98**, 543–548 (1993).

178. Q. Zhao, R. C. Morrison, R. G. Parr, *Pysical Review A* **50**, 2138–2142 (1994).

179. M. Levy, *Proceedings of the National Academy of Sciences* **76**, 6062–6065 (1979).

180. D. Jayatilaka, *Physical review letters* **80**, 798–801 (1998).

181. D. Jayatilaka, D. J. Grimwood, *Acta Crystallographica Section A: Foundations of Crystallography* **57**, 76–86 (2001).

182. I. Bytheway, D. J. Grimwood, D. Jayatilaka, *Acta Crystallographica Section A Foundations of Crystallography* **58**, 232–243 (2002).

183. D. J. Grimwood, I. Bytheway, D. Jayatilaka, *Journal of computational chemistry* **24**, 470–483 (2003).

184. D. J. Grimwood, D. Jayatilaka, *Acta Crystallographica Section A: Foundations of Crystallography* **57**, 87–100 (2001).

185. M. Hudák *et al.*, *Acta Crystallographica Section A Foundations of Crystallography* **66**, 78–92 (2010).

186. I. Bytheway, D. Gimwood, B. Figgis, G. Chandler, D. Jayatilaka, *Acta Crystallographica Section A Foundations of Crystallography* **58**, 244–251 (2002).

187. A. E. Whitten, D. Jayatilaka, M. A. Spackman, *The Journal of Chemical Physics* **125**, 174505–1–174505–14 (2006).

188. D. D. Hickstein, J. M. Cole, M. J. Turner, D. Jayatilaka, *The Journal of Chemical Physics* **139**, 064108–1–064108–14 (2013).

189. S. Grabowsky, D. Jayatilaka, S. Mebs, P. Luger, *Chemistry - A European Journal* **16**, 12818–12821 (2010).

190. S. Grabowsky *et al.*, *The Journal of Physical Chemistry A* **115**, 12715–12732 (2011).

191. D. Jayatilaka, D. Grimwood, *Acta Crystallographica Section A Foundations of Crystallography* **60**, 111–119 (2004).

192. S. Grabowsky *et al.*, *Angewandte Chemie International Edition* **51**, 6776–6779 (2012).

193. A. Genoni, *The Journal of Physical Chemistry Letters* **4**, 1093–1099 (2013).

194. A. Genoni, *Journal of Chemical Theory and Computation* **9**, 3004–3019 (2013).

195. A. Genoni, *Acta Cryst. A* **72**, s165 (2016).

196. A. Genoni, B. Meyer, *Advances in Quantum Chemistry* **73**, 333–362 (2016).

197. R. Destro, R. E. Marsh, R. Bianchi, *The Journal of Physical Chemistry* **92**, 966–973 (1988).

198. C. Gatti, P. J. MacDougall, R. F. W. Bader, *The Journal of Chemical Physics* **88**, 3792–3804 (1988).

199. R. J. Boyd, L.-C. Wang, *Journal of Computational Chemistry* **10**, 367–375 (1989).

200. R. Carbó, B. Calabuig, *International journal of quantum chemistry* **42**, 1681–1693 (1992).

201. R. Carbó, L. Leyda, M. Arnau, *International Journal of Quantum Chemistry* **17**, 1185–1189 (1980).

202. M. Head-Gordon, A. M. Grana, D. Maurice, C. A. White, *The Journal of Physical Chemistry* **99**, 14261–14270 (1995).

203. E Hückel, *Chemical Communications*, 71–85 (1938).

204. R. H. Mitchell, *Chemical reviews* **101**, 1301–1316 (2001).

205. J. A. N. F. Gomes, R. B. Mallion, *Chemical Reviews* **101**, 1349–1384 (2001).

206. H.-B. Bürgi, S. C. Capelli, *Helvetica chimica acta* **86**, 1625–1640 (2003).

207. T. M. Krygowski, M. K. Cyrański, *Chemical Reviews* **101**, 1385–1420 (2001).

208. J. S. Tse, D. D. Klug, S. Patchkovskii, Y. Ma, J. K. Dewhurst, *The Journal of Physical Chemistry B* **110**, 3721–3726 (2006).

209. T. Yamanaka, T. Okada, Y. Nakamoto, *Physical Review B* **80**, 094108–1–094108–8 (2009).

210. N. Casati, A. Kleppe, A. P. Jephcoat, P. Macchi, *Nature Communications* **7**, 10901–1–10901–8 (2016).

211. R. Dovesi *et al.*, *International Journal of Quantum Chemistry* **114**, 1287–1317 (2014).

212. R. Destro, M. Simonetta, *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry* **33**, 3219–3221 (1977).

213. R. Destro, F. Merati, *Acta Crystallographica Section B: Structural Science* **51**, 559–570 (1995).

214. K. B. Wiberg, *Tetrahedron* **24**, 1083–1096 (1968).

Résumé

Les recherches menées dans le cadre de cette thèse avaient un double objectif. Premièrement, le développement d'une nouvelle méthode de chimie quantique à croissance linéaire basée sur le concept d'Orbitales Moléculaires Extrêmement Localisées (ELMOs) et adaptée à l'étude de très gros systèmes moléculaires. Deuxièmement, il s'agit d'évaluer le potentiel des méthodes de calcul utilisant de fonctions d'ondes contraintes et leur capacité à reproduire des données de diffraction aux rayons-X. En ce qui concerne le premier objectif, notre approche se base sur le principe de transférabilité, à savoir l'observation que les systèmes moléculaires sont composés par des unités fonctionnelles récurrentes qui conservent leurs caractéristiques lorsqu'elles se trouvent dans un même environnement chimique. Malheureusement, les orbitales moléculaires traditionnellement employées en chimie théorique dans des modèles de particule indépendante (Hartree-Fock, Kohn-Sham) sont complètement délocalisées sur le système étudié et, par conséquent, ne peuvent pas être transférées d'une molécule à une autre. Ce problème peut être résolu en ayant recours à des orbitales moléculaires déterminées de manière variationnelle sous la contrainte d'être exprimées à partir des fonctions de base centrées sur des atomes de fragments présélectionnés : les ELMOs. En fait, puisqu'elles sont strictement localisées, ces orbitales sont en principe transférables d'une molécule à une autre. L'objectif à terme est d'exploiter cette transférabilité en construisant une base de données d'ELMOs permettant de calculer quasiment instantanément, de manière approximative, des fonctions d'ondes et des densités électroniques de macromolécules. Dans la première partie de cette thèse, nous avons évalué le degré de transférabilité des orbitales moléculaires extrêmement localisées et nous avons proposé une approximation appropriée pour les molécules modèles servant à la détermination des ELMOs qui seront stockées dans la future base de données. Nous avons également comparé la transférabilité des ELMOs avec celle de densités électroniques atomiques asphériques (pseudo-atomes) qui sont largement répandues en cristallographie pour le raffinement de structure cristallographique de grands systèmes. La seconde partie de la thèse se focalise sur les méthodes quantiques utilisant des fonctions d'ondes contraintes. Dans ces méthodes, on cherche à déterminer des fonctions d'ondes qui minimisent l'énergie électronique des systèmes étudiés, mais qui en même temps doivent reproduire un jeu d'amplitudes de facteurs de structure expérimentaux. Cette technique, initialement proposée par Jayatilaka, a récemment été étendue à la théorie des orbitales moléculaires extrêmement localisées. Dans ce contexte, nous avons tout d'abord étudié les effets d'une localisation stricte sur la structure électronique dans des calculs de la fonction d'onde contrainte. Puis, nous avons déterminé si la fonction d'onde contrainte (et la densité associée) est capable de capturer des effets de la corrélation électronique. Enfin, en utilisant une nouvelle technique dite Valence Bond "expérimentale", basée sur les ELMOs, nous avons effectué une étude théorique sur le syn-1,6:8,13-Biscarbonyl[14] annulène (BCA) pour expliquer la rupture partielle de son aromaticité à haute pression observée expérimentalement. Cette dernière étude illustre positivement la potentialité du concept d'orbitale moléculaire strictement localisée en chimie quantique, qui ouvre des perspectives très larges notamment pour l'étude statique ou dynamique de systèmes moléculaires complexes.

Abstract

The goal of the present work was dual. At first, this thesis aimed at proposing new linear scaling quantum chemistry methods based on Extremely Localized Molecular Orbitals (ELMOs) and, secondly, it focused on the assessment of the capabilities of the X-ray constrained wave function approaches. Concerning the first target, our approach is based on the transferability principle, namely the observation that molecular systems are composed by recurrent functional units that generally keep their features when they are in a similar chemical environment. In this context, it is possible to take advantage of the intrinsic transferability of molecular orbitals strictly localized on small molecular subunits to recover wave functions and electron densities of large systems. Unfortunately, the molecular orbitals traditionally used in quantum chemistry are completely delocalized on the system in exam and, therefore, are not transferable from a molecule to another. This problem can be solved only considering molecular orbitals variationally determined under the constraint of expanding them on local basis sets associated with pre-determined molecular fragments: the ELMOs. In fact, since they are strictly localized, these orbitals are in principle transferable from molecule to molecule and our final goal is to construct databanks of ELMOs that will enable to recover almost instantaneously approximate wave functions and electron densities of macromolecules at a very low computational cost. In the first part of this thesis, we have evaluated the transferability of the Extremely Localized Molecular Orbitals and we have defined a suitable model molecule approximation for the computation of the ELMOs to be stored in the future databases. We have also compared the transferability of the ELMOs to the one of the aspherical atomic electron densities (pseudoatoms), which are largely used in crystallography to refine crystallographic structures of large systems. The second part of this work focuses on the X-ray constrained wave function approach. This method consists in determining wave functions that not only minimize the electronic energy of the systems under exam, but that also reproduce sets of experimental structure factor amplitudes within a desired accuracy. The technique, initially proposed by Jayatilaka has been recently extended to the theory of the Extremely Localized Molecular Orbitals. In this context, we have first studied the effects of introducing a strict *a priori* localization on the electronic structure in X-ray constrained wave function calculations. Then, we have determined if the X-ray constrained wave function is intrinsically able to capture the electron correlation effects on the electron densities. Finally, also exploiting a novel X-ray constrained ELMO-based Valence Bond technique, we have reported theoretical studies on the *syn*-1,6:8,13-Biscarbonyl[14] annulene (BCA) to explain the partial rupture of the aromatic character of the molecule occurring at high-pressure.