



# Mise au point de nouveaux descripteurs théoriques pour la réactivité chimique

Guillaume Hoffmann

## ► To cite this version:

Guillaume Hoffmann. Mise au point de nouveaux descripteurs théoriques pour la réactivité chimique. Chimie théorique et/ou physique. Normandie Université, 2020. Français. NNT : 2020NORMR042 . tel-03123326

**HAL Id: tel-03123326**

**<https://theses.hal.science/tel-03123326>**

Submitted on 27 Jan 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Normandie Université

# THÈSE

**Pour obtenir le diplôme de doctorat**

**Spécialité Chimie**

**Préparée au sein de l'Université de Rouen Normandie**

## **Mise au point de nouveaux descripteurs théoriques pour la réactivité chimique**

**Présentée et soutenue par  
Guillaume HOFFMANN**

**Thèse soutenue publiquement le 29 Octobre 2020  
devant le jury composé de**

|                           |                                                                           |                      |
|---------------------------|---------------------------------------------------------------------------|----------------------|
| M / Christophe MORELL     | Professeur / Institut des Sciences Analytiques (Lyon)                     | Rapporteur           |
| M / Alejandro TORO-LABBÉ  | Professeur / Pontificia Universidad Católica (Santiago)                   | Rapporteur           |
| Mme / Isabelle CHATAIGNER | Professeur / Université de Rouen (Rouen)                                  | Examineur            |
| Mme / Tzonka MINEVA       | Directrice de Recherche au CNRS / Institut Charles Gerhardt (Montpellier) | Examineur            |
| M / Frank DE PROFT        | Professeur / Vrije Universiteit Brussel (Bruxelles)                       | Examineur            |
| M / Sami LAKHDAR          | Chargé de recherche CNRS / Université Paul Sabatier (Toulouse)            | Examineur            |
| M / Laurent JOUBERT       | Professeur / Université de Rouen (Rouen)                                  | Directeur de thèse   |
| M / Vincent TOGNETTI      | Maître de conférences / Université de Rouen (Rouen)                       | Codirecteur de thèse |

**Thèse dirigée par Laurent JOUBERT et Vincent TOGNETTI, laboratoire COBRA, UMR 6014**



## *Remerciements*

*Je souhaite premièrement remercier les membres du jury qui ont accepté de juger mon travail, les rapporteurs, le Pr. Christophe Morell et le Pr. Alejandro Toro-Labbé, les examinateurs, la Pr. Isabelle Chataigner, la Dr. Tzonka Mineva, le Pr. Frank De Proft et le Dr. Sami Lakhdar.*

*Je souhaite témoigner ma reconnaissance au Pr. Laurent Joubert, mon directeur de thèse, pour m'avoir accueilli au sein de son groupe et pour la confiance qu'il m'a accordée afin de mener à bien ce sujet de thèse, pour ses conseils, sa disponibilité permanente ainsi que son aide. Je remercie également le Dr. Vincent Tognetti, mon co-directeur de thèse, pour sa gentillesse et ses conversations scientifiques qui sont toujours passionnantes.*

*Je remercie la Pr. Annie-Claude Gaumont pour m'avoir accepté au sein du laboratoire LCMT, ainsi que le Dr. Sami Lakhdar pour sa participation dans cette thèse et dont les connaissances m'ont beaucoup éclairé sur le sujet.*

*Je remercie le LabEx SynOrg et la Région Normandie pour le financement de ces travaux de thèse, ainsi que le CRIANN et l'ensemble de son équipe pour les ressources de calcul et leur aide.*

*Je remercie le Pr. Hassan Oulyadi pour son accueil au sein de l'équipe 1 – Analyse et modélisation, ainsi que le Pr. Pierre-Yves Renard pour son accueil au sein du laboratoire COBRA et dans l'École Doctorale Normande de Chimie.*

*Je remercie les membres du laboratoire LITIS que j'ai pu connaître grâce au RIN AGAC porté par le Pr. Luc Brun, avec qui j'ai eu le plaisir de travailler dans le cadre d'une collaboration, le Pr. Sébastien Adam, le Dr. Pierre Héroux, le Dr. Benoît Gaüzère et particulièrement le Dr. Muhammet Balcilar.*

*Je tiens à remercier toutes les personnes que j'ai pu rencontrer durant mon temps passé au laboratoire LCMT, Fatima Rammal, William Lecroq, Alya Inial, Di Gao, Jean Francois Lohier et Victorien Martin.*

*Je remercie toutes les personnes que j'ai rencontrées au sein du laboratoire COBRA, Andrea Bonvicini, Asma Bourafai, Riham Najjar, Sébastien Hupin, Mattia Migliore, Johann*

*Lemaître, Mathieu Hédouin, Hend Besrour, Aurélien Viret, Estelle Deschamps, Clément Castilla, Lina Truong, Amélie Martin, Ghanem Hamdoun, Leïla Bechtella, Julien Maillard, Aël Cador pour son aide dans la relecture de ce manuscrit et Anaïs Corroy avec qui j'ai passé une année formidable.*

*Je remercie également les superbes personnes que j'ai eu la chance de rencontrer pendant ma thèse en Normandie et avec qui j'ai passé d'agréables moments, Claudia Sesma, Camille Sourdillat, Antoine Denain, Mathieu Morin et Jeanne Gosselin. Je tiens principalement à remercier Geoffrey Manrique, qui m'a accompagné durant cette aventure rouennaise, qui m'a aidé et qui a toujours été là pour me soutenir moralement.*

*Je voudrais exprimer ma reconnaissance envers tous les amis qui ont toujours été là, malgré les distances, qui m'ont encouragé et soutenu durant cette thèse, dont principalement Osian Fonquernie, Timothé Ferré, Emmanuel Queyla et Thibault Kaczmarek.*

*Un grand merci à Valentina Calabrese, pour son aide précieuse, son soutien, sa patience et ses encouragements dans tout ce que j'ai entrepris.*

*Mes derniers remerciements vont à ma famille à qui je dédie ces travaux de thèse.*

# Table des matières

|                                                                   |           |
|-------------------------------------------------------------------|-----------|
| Introduction générale.....                                        | 11        |
| <b>CHAPITRE 1 : Les méthodes théoriques calculatoires.....</b>    | <b>15</b> |
| 1. L'équation de Schrödinger dépendante du temps.....             | 17        |
| 2. L'équation de Schrödinger indépendante du temps.....           | 17        |
| 3. Du cas mono-électronique au système poly-électronique .....    | 18        |
| 4. La méthode d'Hartree-Fock.....                                 | 20        |
| 4.1. Les équations Hartree-Fock .....                             | 20        |
| 4.2. Champ auto-cohérent .....                                    | 22        |
| 5. Les méthodes post Hartree-Fock .....                           | 23        |
| 5.1. La théorie perturbative de Møller-Plesset .....              | 24        |
| 5.2. La théorie du cluster couplé .....                           | 26        |
| 6. Les fonctions de base .....                                    | 27        |
| 7. La théorie de la fonctionnelle de la densité .....             | 30        |
| 7.1. Le premier théorème de Hohenberg et Kohn .....               | 31        |
| 7.2. Le deuxième théorème de Hohenberg et Kohn .....              | 32        |
| 7.3. L'approche de Kohn-Sham .....                                | 33        |
| 7.4. Les fonctionnelles d'échange-corrélation .....               | 36        |
| <b>CHAPITRE 2 : Les méthodes théoriques interprétatives .....</b> | <b>43</b> |
| 1. La réactivité chimique .....                                   | 45        |
| 2. La surface d'énergie potentielle.....                          | 45        |
| 3. Modélisation du solvant.....                                   | 46        |
| 4. La DFT Conceptuelle .....                                      | 47        |
| 4.1. Fondements de la DFT conceptuelle .....                      | 47        |
| 4.2. Ensembles de représentations.....                            | 49        |
| 4.3. L'ensemble canonique.....                                    | 49        |
| 4.3.1. L'ensemble grand canonique.....                            | 51        |
| 4.3.2. L'ensemble isomorphique .....                              | 53        |
| 4.3.3. L'ensemble grand isomorphique .....                        | 54        |
| 4.3.4. Relations de Maxwell .....                                 | 55        |
| 4.4. Descripteurs globaux.....                                    | 56        |
| 4.4.1. Issus directement d'une dérivée de l'énergie .....         | 56        |
| 4.4.2. Descripteurs Ad hoc.....                                   | 59        |

|                                                                                |                                                                                                                                                     |            |
|--------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| 4.5.                                                                           | Descripteurs locaux et non locaux .....                                                                                                             | 62         |
| 4.5.1.                                                                         | Issus directement d'une dérivée de l'énergie .....                                                                                                  | 62         |
| 4.5.2.                                                                         | Descripteurs composites .....                                                                                                                       | 65         |
| 4.6.                                                                           | Principes de réactivité .....                                                                                                                       | 67         |
| 4.6.1.                                                                         | Principe d'égalisation des électronégativités .....                                                                                                 | 67         |
| 4.6.2.                                                                         | Principe des acides et des bases durs ou mous .....                                                                                                 | 71         |
| 4.6.3.                                                                         | Principe de dureté maximale.....                                                                                                                    | 70         |
| 4.6.4.                                                                         | Principe d'électrophilie minimale.....                                                                                                              | 71         |
| 4.6.5.                                                                         | Principe de polarisabilité minimale .....                                                                                                           | 73         |
| 5.                                                                             | Analyse de population .....                                                                                                                         | 74         |
| 5.1.                                                                           | Basées sur les orbitales moléculaires .....                                                                                                         | 74         |
| 5.1.1.                                                                         | Analyse de Population de Mulliken.....                                                                                                              | 74         |
| 5.1.2.                                                                         | Analyse de Population Naturelle .....                                                                                                               | 75         |
| 5.2.                                                                           | Basées sur le potentiel électrostatique.....                                                                                                        | 76         |
| 5.3.                                                                           | Basées sur la densité électronique.....                                                                                                             | 76         |
| 5.3.1.                                                                         | Analyse de population Hirshfeld .....                                                                                                               | 76         |
| 5.3.2.                                                                         | La théorie quantique AIM.....                                                                                                                       | 77         |
| <b>CHAPITRE 3 : Les échelles de réactivité expérimentale .....</b>             |                                                                                                                                                     | <b>91</b>  |
| 1.                                                                             | Introduction.....                                                                                                                                   | 93         |
| 2.                                                                             | L'échelle de réactivité de Mayr .....                                                                                                               | 95         |
| 2.1.                                                                           | Présentation générale .....                                                                                                                         | 95         |
| 2.2.                                                                           | Applications.....                                                                                                                                   | 100        |
| 2.3.                                                                           | Limites et extension.....                                                                                                                           | 101        |
| <b>CHAPITRE 4 : Prédiction théorique de l'électrophilie expérimentale.....</b> |                                                                                                                                                     | <b>105</b> |
| 1.                                                                             | <i>Les descripteurs moléculaires et atomiques peuvent-ils prédire l'électrophilie des accepteurs de Michaël ? .....</i>                             | <i>107</i> |
| 1.1.                                                                           | Préambule .....                                                                                                                                     | 107        |
| 1.2.                                                                           | Article .....                                                                                                                                       | 109        |
| 1.3.                                                                           | Conclusion et perspectives .....                                                                                                                    | 121        |
| 2.                                                                             | <i>Sur l'influence des effets dynamiques sur les descripteurs de réactivité .....</i>                                                               | <i>122</i> |
| 2.1.                                                                           | Préambule .....                                                                                                                                     | 122        |
| 2.2.                                                                           | Article .....                                                                                                                                       | 123        |
| 2.3.                                                                           | Conclusion et perspectives .....                                                                                                                    | 128        |
| 3.                                                                             | <i>Prédire l'électrophilie expérimentale à partir de descripteurs quantiques et topologiques : une approche par apprentissage automatique .....</i> | <i>130</i> |
| 3.1.                                                                           | Préambule .....                                                                                                                                     | 130        |

|                                                                                                                                             |            |
|---------------------------------------------------------------------------------------------------------------------------------------------|------------|
| 3.2. Article .....                                                                                                                          | 131        |
| 3.3. Conclusion .....                                                                                                                       | 144        |
| <b>CHAPITRE 5 : Indices d'électrophilie et liaisons halogènes : De nouvelles alternatives au potentiel électrostatique moléculaire.....</b> | <b>145</b> |
| 5.1. Préambule .....                                                                                                                        | 147        |
| 5.2. Article .....                                                                                                                          | 148        |
| 5.3 Conclusion et perspectives .....                                                                                                        | 160        |
| <b>CHAPITRE 6 : Synthèse de cations iminium vinylogues : Application de l'échelle de Mayr aux électrophiles ambidents.....</b>              | <b>161</b> |
| 1. Introduction .....                                                                                                                       | 163        |
| 2. Résultats et discussions .....                                                                                                           | 165        |
| 2.1 Synthèse.....                                                                                                                           | 165        |
| 2.2 Caractérisation .....                                                                                                                   | 168        |
| 2.3 Mesures cinétiques .....                                                                                                                | 170        |
| 3. Conclusion.....                                                                                                                          | 173        |
| Conclusion générale et perspectives.....                                                                                                    | 177        |
| Annexe 1 : Liste et formules des principaux descripteurs employées .....                                                                    | 179        |



# Liste des Figures

- Figure 1.1** : Schéma illustrant la méthode du champ auto-cohérent
- Figure 1.2** : Schéma illustrant la nomenclature générale des bases de Pople
- Figure 1.3** : Schéma illustrant la nomenclature générale des bases de Dunning
- Figure 1.4** : Schéma illustrant la procédure de résolution itérative des équations de Kohn-Sham
- Figure 1.5** : Echelle de Jacob de la théorie de la fonctionnelle de la densité
- Figure 2.1** : Dérivées successives au troisième ordre pour l'ensemble canonique
- Figure 2.2** : Dérivées successives au second ordre pour l'ensemble grand canonique
- Figure 2.3** : Dérivées successives au second ordre pour l'ensemble isomorphique
- Figure 2.4** : Dérivées successives au second ordre pour l'ensemble grand isomorphique
- Figure 2.5** : Diagramme de Parr illustrant les relations entre les différents ensembles
- Figure 2.6** : Représentation 2D de la topologie du champ de gradient de la densité électronique pour la molécule d'éthylène ( $C_2H_4$ )
- Figure 2.7** : Représentation 3D pour la molécule d'éthylène ( $C_2H_4$ )
- Figure 3.1** : Les deux références de l'échelle
- Figure 3.2** : Partie de l'échelle expérimentale de Mayr, avec à gauche les nucléophiles et à droite les électrophiles
- Figure 3.3** : Electrophilie expérimentale pour la 2,5-dichloroquinone (à droite) et la 2,3-dichloro-5,6-dicyano-p-benzoquinone (à gauche)
- Figure 3.4** : Réaction entre un nucléophile et l'ion bis (p-méthoxyphényl) carbénium
- Figure 3.5** : (1) Détermination de la longueur d'absorbance maximale (2) Mesure de la constante de vitesse observée expérimentalement (3) Courbe des constantes observées en fonction de la concentration du nucléophile
- Figure 3.6** : Représentation des différentes méthodes analytiques pour la détermination des constantes de vitesses expérimentales
- Figure 3.7** : Schéma illustrant la prédiction de réaction grâce à l'échelle d'électrophilie de Mayr
- Figure 3.8** : Réaction de Diels-Alder entre l'anhydride maléique et le cyclopentadiène
- Figure 4.1** : Réaction et mécanisme type d'une addition de Michael avec la buténone en exemple
- Figure 4.2** : Réaction entre les ylures de pyridinium et les accepteurs de Michael.
- Figure 4.3** : Quatre structures extraites de la dynamique moléculaire pour la molécule 7. Image générée avec GaussView 5.0
- Figure 4.4** : Quatre structures extraites de la dynamique moléculaire pour la molécule 6. Image générée avec GaussView 5.0
- Figure 6.1** : Formation des cations iminium vinylogues. Les points noirs correspondent aux centres électrophiles
- Figure 6.2** : Schéma illustrant le principe de la catalyse par LUMO Lowering (en bleu) et HOMO Raising (en rouge)
- Figure 6.3** : Exemple d'application des cations iminium vinylogues lors de synthèse en cascade [4].
- Figure 6.4** : Synthèse du sel de pyrrolidine
- Figure 6.5** : Première étape de la synthèse
- Figure 6.5** : Deuxième étape de la synthèse
- Figure 6.6** : Synthèse du sel de pyrrolidine
- Figure 6.8** : Dernière étape de la synthèse
- Figure 6.7** : Spectre RMN  $^1H$  à 500 MHz du produit 14. Le solvant utilisé est  $CDCl_3$ .
- Figure 6.8** : Spectre de masse, m/z 50 à 1000 en électrospray du composé 14, ions positifs
- Figure 6.11** : Structure RX du composé 14, générée avec SHELXL-2014
- Figure 6.12** : Représentation des trois sites électrophiles du cation iminium 14
- Figure 6.9** : Spectre UV/Visible après pulsation dans le dichlorométhane ( $CH_2Cl_2$ ) de la molécule 14.  $\lambda_{max} = 375\text{ nm}$
- Figure 6.10** : Exemple de détermination expérimentale de la constante cinétique  $k_2$  pour un nucléophile
- Figure 6.11** : Détermination expérimentale du paramètre d'électrophilie
- Figure 6.12** : Exemple d'ammoniums d'intérêt stériques pouvant être utilisés pour la synthèse de nouveaux cations iminium vinylogues
- Figure 6.13** : Valeurs des descripteurs calculées sur la molécule 14. Image généré avec GaussView 5.0

## Liste des acronymes

**HF** : Hartree-Fock

**MP2** : Møller-Plesset à l'ordre 2

**CC(S)(D)(T)(Q)** : Cluster couplé (simple) (double) (triple) (quadruple)

**OM** : Orbitale moléculaire

**STO** : Orbitale de type Slater

**GTO** : Orbitale de type gaussienne

**CGTO** : Orbitale de type gaussienne contractée

**BSSE** : Erreur de superposition de bases

**DFT** : Théorie de la fonctionnelle de la densité

**HK** : Hohenberg-Kohn

**KS** : Kohn-Sham

**LDA** : Approximation de densité locale

**GGA** : Approximation du gradient généralisée

**SCRf** : Champ de réaction auto-cohérent

**PCM** : Modèle de continuum polarisable

**CDFT** : Théorie de la fonctionnelle de la densité conceptuelle

**HOMO** : Orbitale moléculaire la plus haute occupée

**LUMO** : Orbitale la plus basse non occupée

**HSAB** : Principe des acides et des bases durs et mous

**QTAIM** : Théorie quantique des atomes dans les molécules

**BCP** : Point critique de liaison

**NNA** : Attracteur non nucléaire

**IQA** : Atomes quantique en interaction

**RMN** : Résonance magnétique nucléaire

**UV** : Ultraviolet

**RSP** : Principe de réactivité-sélectivité

**Me** : Méthyle

**Et** : Éthyle

**DABCO** : 1,4-Diazabicyclo[2,2,2]octane

*Mise au point de nouveaux descripteurs théoriques  
pour la réactivité chimique*

-

*Development of new theoretical descriptors  
for chemical reactivity*

# Introduction générale

C'est en se posant la question existentielle « *De quoi est faite la matière ?* », que Démocrite (460-370 avant J.C.), un philosophe grec de l'Antiquité, a émis l'hypothèse que la matière tangible est composée de minuscules unités qui peuvent être assemblées et démontées par diverses combinaisons, unités qu'il appela « *atomes* » [1].

Il aura alors fallu attendre le XIX<sup>ème</sup> siècle pour que naisse la première notion de *théorie atomique*, entre autres grâce à Dalton, afin d'interpréter et de rationaliser les phénomènes chimiques qui nous entourent, réunissant les notions d'atomes et de molécules. Puis, au XX<sup>ème</sup> siècle, avec l'essor de la physique atomique, l'augmentation considérable des connaissances permet d'établir un grand nombre de lois et de modèles permettant la compréhension des effets électroniques. Ces lois de la mécanique classique ne permettent pas en revanche de décrire le comportement des électrons ou des molécules. C'est grâce aux développements de la mécanique quantique, avec principalement l'équation de Schrödinger en 1926 [2], que la compréhension des systèmes à partir de la fonction d'onde va permettre une grande avancée, permettant le calcul et la prédiction des propriétés physiques et chimiques d'un système.

En substituant la fonction d'onde par la densité électronique, W. Kohn a développé la théorie de la fonctionnelle de la densité (la *DFT*) [3], qui représente à ce jour l'une des méthodes les plus utilisées dans les calculs quantiques. Ses calculs n'exigeant pas de connaître la fonction d'onde d'un système, ils ont permis l'expansion impressionnante des domaines de la chimie quantique et de la chimie computationnelle. Le succès de la DFT est également dû aux progrès réalisés dans le domaine de l'informatique, permettant de nos jours de réaliser des calculs de plus en plus conséquents en un temps plus rapide.

C'est en 1978 [4], initiée par R. G. Parr, que la DFT dite *conceptuelle* (CDFT) apparaît, afin de développer à partir des concepts de la DFT et notamment la densité électronique, une théorie de la réactivité chimique [5, 6]. Dans la CDFT, les concepts physiques et mathématiques de la DFT sont réutilisés et combinés pour former des *descripteurs* afin de quantifier les concepts chimiques, en particulier ceux liés à la réactivité. L'essence de la DFT conceptuelle est ainsi d'aborder la réactivité chimique et la sélectivité du point de vue de la densité électronique. Le développement de cette branche de la DFT depuis les dernières années a su démontrer ses vastes domaines d'applications, en passant de la chimie inorganique à celle des matériaux, à la

biochimie, à la chimie des polymères ainsi qu'à la chimie organique. Pour la réactivité chimique, la CDFT a rencontré un succès dans, entre autres, les réactions en phase gazeuse, les réactions en solution, les réactions et réarrangements dans les solides, les réactions de complexation acide-base, les réactions d'oxydoréduction et les réactions péricycliques.

On peut bien entendu également discuter de la réactivité chimique à partir de résultats expérimentaux. Savoir différencier ou classer les sites réactifs de différentes molécules est d'une importance capitale pour une meilleure compréhension de la réactivité chimique, que ce soit d'un point de vue théorique ou bien expérimental. C'est dans ce contexte que de nombreuses échelles expérimentales ont été établies au fil des années afin de classer les molécules par réactivité [7, 8]. Parmi ces échelles expérimentales, la plus décrite et utilisée est celle proposée par Mayr en 1994 [9, 10].

D'un point de vue théorique, une telle échelle expérimentale encourage la mise au point d'outils théoriques capable de prédire ces valeurs expérimentales, voire de prédire des valeurs sur des molécules jusqu'alors non synthétisées. C'est dans cette optique que s'inscrit ce travail de thèse, qui est la mise au point de nouvelles méthodes théoriques, basées notamment sur la CDFT, afin de prédire la réactivité et la sélectivité d'un grand nombre de composés organiques et d'étendre les domaines d'applications des descripteurs de réactivité.

La thèse est organisée de la manière suivante : les deux premiers chapitres vont se concentrer sur la présentation des méthodes quantiques utilisées, avec le premier chapitre qui présente les méthodes théoriques calculatoires employées et le second chapitre qui vise quant à lui à présenter les méthodes théoriques interprétatives, en mettant l'accent sur la théorie de la fonctionnelle de la densité conceptuelle et la présentation des descripteurs de réactivité.

Le troisième chapitre est consacré à la présentation des échelles expérimentales de réactivité, en présentant notamment l'échelle de Mayr qui est un sujet central de cette thèse. Le quatrième chapitre regroupe l'ensemble des travaux réalisés sur la prédiction théorique de l'électrophilie expérimentale, par le biais des descripteurs quantiques. Le cinquième chapitre présente les travaux obtenus dans l'utilisation des descripteurs issus de la CDFT sur la liaison halogène. Le dernier chapitre discute des résultats obtenus pour la synthèse de cations iminium vinylogues. Enfin, une conclusion générale et les perspectives proposées sont présentes en fin de manuscrit.

- [1] C.C.W. Taylor, *The atomists, Leucippus and Democritus: fragments, a text and translation with a commentary* University of Toronto, Press Incorporated (1999)
- [2] E. Schrödinger, *Annalen der Physik*, 384 (1926) 361-376
- [3] P. Hohenberg, W. Kohn, *Phys. Rev.* 136 (1964) B864
- [4] R. G. Parr, R. A. Donnelly, M. Levy, W. E. Palke. *J. Chem. Phys.* 68 (1978), 3801–3807
- [5] A. Toro-Labbé, *Theoretical aspects of chemical reactivity*, Elsevier, Amsterdam (2006)
- [6] P. K. Chattaraj, *Chemical reactivity theory, a density functional view* CRC Press, Boca Raton (2009)
- [7] C. Gardner Swain, C. B. Scott. *J. Am. Chem. Soc.* 75 (1953), 141–147
- [8] C. D. Ritchie. *Acc. Chem. Res.* 5 (1972), 348–354
- [9] H. Mayr, B. Kempf, A. R. Ofial. *Acc. Chem. Res.* 36 (2003), 66–77
- [10] H. Mayr, M. Patz. *Angew. Chemie Int. Ed. English* 33 (1994), 938–957



# CHAPITRE 1 : Les méthodes théoriques calculatoires

|                                                                |    |
|----------------------------------------------------------------|----|
| 1. L'équation de Schrödinger dépendante du temps.....          | 17 |
| 2. L'équation de Schrödinger indépendante du temps.....        | 17 |
| 3. Du cas mono-électronique au système poly-électronique ..... | 18 |
| 4. La méthode d'Hartree-Fock.....                              | 20 |
| 4.1. Les équations Hartree-Fock .....                          | 20 |
| 4.2. Champ auto-cohérent .....                                 | 22 |
| 5. Les méthodes post Hartree-Fock .....                        | 23 |
| 5.1 La théorie perturbative de Møller-Plesset .....            | 24 |
| 5.2. La théorie du cluster couplé .....                        | 26 |
| 6. Les fonctions de base .....                                 | 27 |
| 7. La théorie de la fonctionnelle de la densité .....          | 30 |
| 7.1. Le premier théorème de Hohenberg et Kohn .....            | 31 |
| 7.2. Le deuxième théorème de Hohenberg et Kohn .....           | 32 |
| 7.3. L'approche de Kohn-Sham .....                             | 33 |
| 7.4. Les fonctionnelles d'échange-corrélation .....            | 36 |





## 1. L'équation de Schrödinger dépendante du temps

Lorsque de Broglie émet en 1923 [1] l'hypothèse que toutes les particules se comportent aussi comme des ondes, Schrödinger postule une équation non relativiste [2], avec une solution analytique dans les cas simples, qui donne l'évolution dans l'espace ( $\vec{r}$ ) et le temps ( $t$ ) de la fonction d'onde  $\Psi$  d'une particule :

$$i\hbar \frac{\partial}{\partial t} \Psi(\vec{r}, t) = \hat{H} \Psi(\vec{r}, t), \quad (1.1)$$

avec  $\hbar$  la constante de Planck réduite,  $\hat{H}$  l'opérateur Hamiltonien et  $\Psi$  la fonction d'onde contenant toutes les propriétés quantiques du système. Le sens physique de cette fonction ne fut défini que plus tard en 1927 par Born et Jordan [3], par le biais du carré du module de la fonction d'onde :  $|\Psi|^2$ . Plus précisément, ils postulèrent que la densité de probabilité de présence  $dP$  d'une particule dans un volume infinitésimal ( $d^3r$ ) est donnée par :

$$dP(\vec{r}, t) = |\Psi(\vec{r}, t)|^2 d^3r. \quad (1.2)$$

Cette interprétation probabiliste implique la normalisation suivante :

$$\int_{\text{espace}} |\Psi(\vec{r}, t)|^2 d^3r = 1. \quad (1.3)$$

## 2. L'équation de Schrödinger indépendante du temps

Afin d'obtenir l'essentiel de l'information du système, il est nécessaire de passer par la résolution de l'équation de Schrödinger pour déterminer la fonction d'onde.

En se plaçant dans le cas où l'Hamiltonien n'est pas dépendant du temps, on peut écrire la fonction d'onde comme :

$$\Psi(\vec{r}, t) = A(t) \times \psi(\vec{r}), \quad (1.4)$$

où  $A$  décrit l'évolution temporelle et  $\psi$  la dépendance spatiale de la fonction d'onde. Les fonctions possibles (quantifiées)  $\psi_n$  représentent les états stationnaires. En utilisant (1.4), l'équation de Schrödinger

$$i\hbar \frac{\partial}{\partial t} [A(t)\psi(\vec{r})] = \hat{H}[A(t)\psi(\vec{r})] \quad (1.5)$$

se réécrit :

$$i\hbar \psi(\vec{r}) \frac{\partial A(t)}{\partial t} = A(t) \hat{H} \psi(\vec{r}). \quad (1.6)$$

En divisant les termes de l'équation (1.6) par  $A(t)\psi(\vec{r})$ , une fonction dépendant uniquement du temps doit être égale à une fonction spatiale. Elle doit donc être constante :

$$\frac{i\hbar}{A(t)} \frac{\partial A(t)}{\partial t} = \frac{\hat{H}\psi(\vec{r})}{\psi(\vec{r})} = \text{Constante} = \varepsilon, \quad (1.7)$$

$\varepsilon$  représente en fait l'énergie du système. Chaque état stationnaire  $\psi_n(\vec{r})$  est donc solution de l'équation aux valeurs propres suivante :

$$\hat{H}\psi_n(\vec{r}) = E_n\psi_n(\vec{r}). \quad (1.8)$$

Avec la notation « bra|ket » de Dirac on obtient l'équation de Schrödinger indépendante du temps aux valeurs propres :

$$E_n|\psi_n\rangle = \hat{H}|\psi_n\rangle \quad (1.9)$$

### 3. Du cas mono-électronique au système poly-électronique

La mécanique quantique présentée précédemment s'applique à des systèmes mono-électroniques, mais peut également s'appliquer à des systèmes moléculaires. Dans ce cas, une seule fonction d'onde régit l'ensemble du système. L'opérateur Hamiltonien  $\hat{H}$  (l'observable énergie) dans l'équation de Schrödinger s'exprime généralement comme la somme de l'opérateur d'énergie cinétique  $\hat{T}$  et de l'opérateur d'énergie potentielle  $\hat{V}$  :

$$\hat{H} = \hat{T} + \hat{V}. \quad (1.10)$$

Pour un système moléculaire à  $n$  électrons et  $M$  noyaux, l'opérateur Hamiltonien peut s'écrire en prenant en compte les opérateurs énergies des électrons (notés  $e$ ) et des noyaux (notés  $N$ ) :

$$\hat{H} = \hat{T}_N + \hat{T}_e + \hat{V}_{ee} + \hat{V}_{eN} + \hat{V}_{NN}. \quad (1.11)$$

L'opérateur d'énergie cinétique nucléaire  $\hat{T}_N$  s'écrit en unités atomiques (avec  $M_A$  désignant la masse du noyau A) :

$$\hat{T}_N = - \sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2, \quad (1.12)$$

tandis que l'opérateur d'énergie cinétique électronique  $\hat{T}_e$  s'écrit :

$$\hat{T}_e = - \sum_{i=1}^n \frac{1}{2} \nabla_i^2. \quad (1.13)$$

L'opérateur d'énergie potentielle de répulsion  $\hat{V}_{ee}$  entre deux électrons  $i$  et  $j$  a quant à lui l'expression suivante :

$$\hat{V}_{ee} = \sum_{i=1}^{n-1} \sum_{j>i}^n \frac{1}{\hat{r}_{ij}}, \quad (1.14)$$

où  $\hat{r}_{ij}$  est ici un opérateur de position relative ( $r_{ij} = |\vec{r}_i - \vec{r}_j|$ ).  $\hat{V}_{eN}$  est l'opérateur d'énergie potentielle d'attraction électrons-noyaux :

$$\hat{V}_{eN} = - \sum_{i=1}^n \sum_{A=1}^M \frac{Z_A}{\hat{r}_{iA}}, \quad (1.15)$$

où  $Z_A$  est le numéro atomique de l'atome  $A$  et  $\hat{r}_{iA}$  l'opérateur de position représentant la distance de l'électron  $i$  au noyau  $A$ . Enfin,  $\hat{V}_{NN}$  est l'opérateur d'énergie potentielle de répulsion entre deux noyaux qui s'écrit :

$$\hat{V}_{NN} = \sum_{A=1}^{M-1} \sum_{B>A}^M \frac{Z_A Z_B}{\hat{R}_{AB}}, \quad (1.16)$$

où  $\hat{R}_{AB}$  est l'opérateur de position représentant la distance entre les deux noyaux  $A$  et  $B$ . Tout ceci étant valable dans un cadre non relativiste.

En 1927, Born et Oppenheimer proposent une simplification de l'équation de Schrödinger appliquée aux molécules [4]. Cette approximation est basée sur l'importante différence de masse entre celles des électrons et celles des noyaux. On considère la position des noyaux dans l'espace comme fixe (le temps de résoudre le problème électronique), avec un mouvement des électrons dans l'espace considérablement plus rapide autour des noyaux.

Dans l'approximation de Born-Oppenheimer, la fonction d'onde totale du système qui dépend de la position des noyaux et des électrons,  $\psi(\{\vec{r}_i\}, \{\vec{R}_A\})$ , peut ainsi se découpler comme produit

d'une fonction d'onde nucléaire décrivant les noyaux  $\psi_{nuc}(\{\vec{R}_A\})$  et d'une fonction d'onde électronique décrivant les électrons  $\psi_{el}(\{\vec{r}_i\}; \{\vec{R}_A\})$  :

$$\psi(\{\vec{r}_i\}, \{\vec{R}_A\}) = \psi_{el}(\{\vec{r}_i\}; \{\vec{R}_A\}) \times \psi_{nuc}(\{\vec{R}_A\}). \quad (1.17)$$

A géométrie fixe la position des noyaux  $\{\vec{R}_A\}$  devient un ensemble de paramètres constants, introduisant l'équation de Schrödinger électronique. Dans ce cadre, on ne résout alors pas l'Hamiltonien complet du système  $\hat{H}$  mais on se concentre sur l'Hamiltonien électronique  $\hat{H}_{el}$  :

$$\hat{H}_n^{el} = \hat{T}_e + \hat{V}_{ee} + \hat{V}_{eN}. \quad (1.18)$$

Le terme  $\hat{V}_{ee}$  de répulsion électronique rend en pratique impossible la résolution au-delà de quelques électrons, c'est pour cela que l'on a besoin d'approximations supplémentaires permettant le calcul et la résolution de l'équation de Schrödinger électronique :

$$\hat{H}_n^{el} \psi_n^{el}(\vec{r}) = E_n^{el} \psi_n^{el}(\vec{r}). \quad (1.19)$$

$E^{el}$  représente l'énergie électronique du système moléculaire. L'état électronique de plus basse énergie sera appelé l'état électronique *fondamental* d'énergie  $E_0^{el}$ . Par la suite toutes les énergies et fonctions d'onde seront des énergies électroniques.

La résolution de l'équation (1.19) peut passer par l'utilisation du principe variationnel. Dans le cas de l'état fondamental, l'énergie sera calculée à partir d'une fonction d'onde dite d'essai paramétrée ( $\tilde{\psi}$ ). On peut montrer que cette énergie sera toujours supérieure (ou égale) à l'énergie exacte de l'état fondamental :

$$\tilde{E} = \frac{\langle \tilde{\psi} | \hat{H} | \tilde{\psi} \rangle}{\langle \tilde{\psi} | \tilde{\psi} \rangle} \geq E_0. \quad (1.20)$$

Les paramètres optimaux seront obtenus par une minimisation de l'énergie.

## 4. La méthode d'Hartree-Fock

### 4.1. Les équations Hartree-Fock

L'une des premières méthodes de résolution de l'équation de Schrödinger électronique dans le cadre de l'approximation de Born-Oppenheimer est la méthode de Hartree-Fock [5]. Cette méthode est basée sur l'approximation du champ moyen qui consiste à considérer seulement le

mouvement de chaque électron se déplaçant au sein d'un potentiel moyen causé par les autres électrons. On cherchera alors à exprimer la fonction d'onde polyélectronique à partir de fonctions d'onde monoélectroniques appelées spin-orbitales. En effet, il est important de préciser à ce stade qu'un électron (une particule qui possède un spin de  $\pm 1/2$ ) peut prendre exclusivement deux valeurs de spin,  $\alpha$  et  $\beta$  qui sont par définition de signes opposés.

En outre, la fonction d'onde  $\psi_{el}(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N)$  (où  $\vec{x}_i$  comprend la position des électrons  $\vec{r}$  ainsi que le spin des électrons  $\sigma_i$  ( $\alpha$  ou  $\beta$ ) :  $\vec{x}_i = (\vec{r}_i, \sigma_i)$ ) doit être antisymétrique par permutation des électrons. Pour assurer cela, on exprime la fonction d'onde  $\psi_{el}(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N)$  comme un déterminant de Slater :

$$\psi_{el}(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\vec{x}_1) & \phi_2(\vec{x}_1) & \dots & \phi_N(\vec{x}_1) \\ \phi_1(\vec{x}_2) & \phi_2(\vec{x}_2) & \dots & \phi_N(\vec{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\vec{x}_N) & \phi_2(\vec{x}_N) & \dots & \phi_N(\vec{x}_N) \end{vmatrix} \quad (1.21)$$

Dans ce formalisme, on respecte le principe d'exclusion de Pauli [6] puisque deux électrons ne peuvent pas occuper la même spin-orbitale. En effet si deux colonnes sont identiques, le déterminant est nul.

La méthode Hartree-Fock consiste alors à déterminer, par application du principe variationnel, les spin-orbitales définies précédemment qui sont solutions des équations Hartree-Fock mono électroniques :

$$\hat{F}_i(\vec{x}_i)\phi_i(\vec{x}_i) = \varepsilon_i\phi_i(\vec{x}_i). \quad (1.22)$$

L'opérateur de Fock  $\hat{F}_i$  est ici défini comme somme d'un Hamiltonien mono électronique  $\hat{h}_i$  et d'un terme qui approxime la moyenne de la répulsion électronique :

$$\hat{F}_i(\vec{x}_i) = \hat{h}_i(\vec{x}_i) + \sum_{j=1}^n \left[ \hat{J}_j(\vec{x}_i) - \hat{K}_j(\vec{x}_i) \right]. \quad (1.23)$$

On retrouve dans l'expression l'opérateur Coulombien  $\hat{J}_j(\vec{x}_i)$  entre l'électrons  $i$  de la spin-orbitale  $\phi_i(\vec{x}_i)$  et l'électron  $j$  de la spin-orbitale  $\phi_j(\vec{x}_j)$ . La somme des  $\hat{J}_j$  représente le potentiel électrostatique moyen créé par les autres électrons :

$$\hat{J}_i|\phi_j\rangle = \left\langle \phi_i \left| \frac{1}{r_{ij}} \right| \phi_i \right\rangle |\phi_j\rangle. \quad (1.24)$$

L'opérateur d'échange  $\hat{K}_j$  ( $\vec{x}_i$ ), lui, décrit l'effet quantique produit lors de l'échange de deux électrons et représente la correction du potentiel moyen à cause de l'antisymétrie de la fonction d'onde :

$$\hat{K}_i|\phi_j\rangle = \left\langle \phi_i \left| \frac{1}{\hat{r}_{ij}} \right| \phi_j \right\rangle |\phi_i\rangle. \quad (1.25)$$

L'énergie électronique du système que l'on appelle l'énergie Hartree-Fock  $E_{HF}$  peut alors s'exprimer comme :

$$E_{HF} = \sum_{i=1}^n \langle \phi_i | \hat{h}_i | \phi_i \rangle + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\langle \phi_j | \hat{J}_i | \phi_j \rangle - \langle \phi_j | \hat{K}_i | \phi_j \rangle). \quad (1.26)$$

## 4.2. Champ auto-cohérent

Lorsque l'on passe à la résolution des équations Hartree-Fock, on utilise en pratique la méthode du champ dit « auto-cohérent » [7].

En effet, chaque opérateur de Fock  $\hat{F}_i$  dépend de ses solutions (les orbitales). La méthode du champ auto-cohérent est une méthode itérative qui consiste à optimiser l'opérateur de Fock après chaque itération à partir des spin-orbitales  $\phi_i$  calculées à l'itération précédente, jusqu'à l'obtention d'orbitales qui remplissent un critère de convergence prédéfini.

Les spin-orbitales  $\phi_i$  sont exprimées en pratique comme une combinaison linéaire de fonctions de base  $\chi_\mu$  (qui sont décrites plus en détail dans les parties suivantes) :

$$\phi_i(\vec{x}) = \sum_{\mu=1}^{N_b} c_{\mu i} \chi_\mu(\vec{x}), \quad (1.27)$$

où  $N_b$  représente la taille de la base. On vise ici à déterminer variationnellement les coefficients  $c_{\mu i}$ . La solution des équations Hartree-Fock prend alors la forme :

$$\sum_v c_{vi} \langle \chi_\mu | \hat{F}_i | \chi_v \rangle = \varepsilon_i \sum_v c_{vi} \langle \chi_\mu | \chi_v \rangle. \quad (1.28)$$

Ce sont les équations dites de Roothaan et Hall [8] qui prennent la forme matricielle suivante :

$$\mathbf{FC} = \mathbf{SC}\boldsymbol{\varepsilon}, \quad (1.29)$$

où  $\mathbf{F}$  est la matrice de Fock dans la base choisie définie par :

$$F_{\mu\nu} = \langle \chi_\mu | \hat{F}_i | \chi_\nu \rangle, \quad (1.30)$$

$\mathbf{S}$  est la matrice du recouvrement des fonctions de base construite à partir de :

$$S_{\mu\nu} = \langle \chi_\mu | \chi_\nu \rangle, \quad (1.31)$$

$\mathbf{C}$  est une matrice carrée des coefficients  $c_{\mu i}$  et  $\mathbf{\varepsilon}$  est une matrice diagonale construite à partir des énergies orbitales  $\varepsilon_i$ . La méthode de résolution du champ auto-cohérent peut alors se résumer à un schéma qui se présente de la manière suivante :

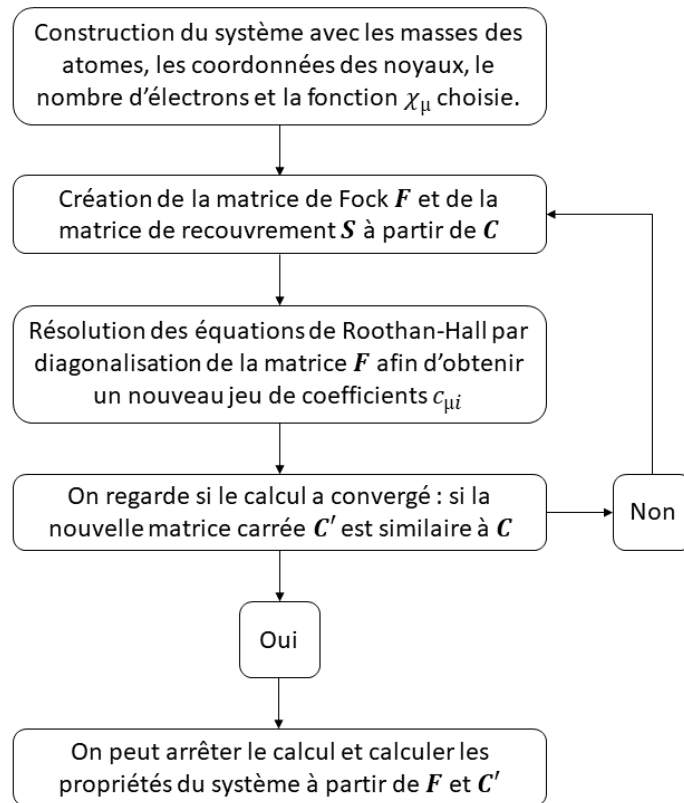


Figure 1.1: Schéma illustrant la méthode du champ auto-cohérent

## 5. Les méthodes post Hartree-Fock

Dans l'approche Hartree-Fock, qui approxime la fonction d'onde par un seul déterminant de Slater, l'antisymétrisation de la fonction d'onde ne permet pas à deux électrons de même spin de se trouver en un même point. La méthode Hartree Fock donnera toutefois une énergie  $E_{HF}$  plus élevée que l'énergie exacte  $E_{exacte}$ . La différence entre ces énergies est appelé l'énergie de corrélation électronique  $E_C$ , qui fut proposée en 1934 par Wigner et Seitz [9] puis définie par Löwdin en 1959 comme [10]:



$$E_C = E_{exacte} - E_{HF}. \quad (1.32)$$

Il existe ainsi plusieurs types de corrélations : dynamique, statique, longue-distance, courte-distance... Les méthodes Post Hartree-Fock prennent en compte une partie de la corrélation électronique. Seules deux approches seront présentées : la théorie perturbative de Møller-Plesset (notée MP par la suite) et la théorie du cluster couplé (CC).

### 5.1. La théorie perturbative de Møller-Plesset

La théorie de Møller-Plesset [11] (notée MP par la suite) est une théorie partant de l'approche perturbative de Rayleigh-Schrödinger, où l'Hamiltonien  $\hat{H}$  d'un système s'exprime comme la somme d'un Hamiltonien de référence non perturbé  $\hat{H}_0$  et d'une faible perturbation du système  $\lambda\hat{V}$  :

$$\hat{H} = \hat{H}_0 + \lambda\hat{V}, \quad (1.33)$$

$\lambda$  correspond à un paramètre réel sans dimension qui contrôle la force de cette perturbation (variant de 0 : pas de perturbation à 1 : perturbation totale). L'Hamiltonien non-perturbé peut être diagonalisée selon :

$$\hat{H}_0 |\psi_i^{(0)}\rangle = E_i^{(0)} |\psi_i^{(0)}\rangle. \quad (1.34)$$

L'équation de Schrödinger pour le système perturbé est :

$$\hat{H} |\psi_i\rangle = E_i |\psi_i\rangle, \quad (1.35)$$

$$(\hat{H}_0 + \lambda\hat{V}) |\psi_i\rangle = E_i |\psi_i\rangle. \quad (1.36)$$

On peut pour cela exprimer les termes  $E_i$  et  $\psi_i$  par un développement en série de Maclaurin :

$$E_i = E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \dots \quad (1.37)$$

$$|\psi_i\rangle = |\psi_i^{(0)}\rangle + \lambda |\psi_i^{(1)}\rangle + \lambda^2 |\psi_i^{(2)}\rangle + \dots \quad (1.38)$$

Ainsi :

$$E_i^{(k)} = \frac{1}{k!} \left. \frac{d^k E_i}{d\lambda^k} \right|_{\lambda=0}. \quad (1.39)$$

$$|\psi_i^{(k)}\rangle = \frac{1}{k!} \left. \frac{d^k |\psi_i\rangle}{d\lambda_k} \right|_{\lambda=0}. \quad (1.40)$$

On peut substituer le développement en série avec  $k \rightarrow \infty$  dans l'équation de Schrödinger générale pour obtenir :

$$(\hat{H}_0 + \lambda \hat{V}) \left( \sum_{i=0}^k \lambda_i |\psi_i\rangle \right) = \left( \sum_{i=0}^k \lambda_i E_i \right) \left( \sum_{i=0}^k \lambda_i |\psi_i\rangle \right). \quad (1.41)$$

Les facteurs  $\lambda_k$  dans l'équation de Schrödinger définiront l'ordre de perturbation  $k$  (l'ordre 1, 2, 3 ...  $k$ ).

Dans le cadre de cette théorie perturbative, l'approche MP consiste à définir l'Hamiltonien de référence non perturbé  $\hat{H}_0$  comme la somme des opérateurs de Fock  $\hat{F}_i$  :

$$\hat{H}_0 = \sum_{i=1}^n \hat{F}_i. \quad (1.42)$$

Si l'on reprend l'équation de Schrödinger (1.37), on peut alors exprimer les différents ordres de perturbation de l'approche MP :

#### A l'ordre zéro ( $k = 0$ ) :

L'expression de l'énergie de correction  $E_i^{(0)}$  est simplement la somme des énergies des orbitales occupées :

$$E_i^{(0)} = E_c^{MP0} = \sum_i^{occ.} \varepsilon_i. \quad (1.43)$$

#### Au premier ordre ( $k = 1$ ) :

Pour le terme correctionnel du premier ordre, l'énergie de correction correspond à la valeur moyenne prédite de  $\hat{H}$  telle que :

$$E_i^{(1)} = E_c^{MP1} = \langle \psi_i^{(0)} | \hat{V} | \psi_i^{(0)} \rangle. \quad (1.44)$$

Il faut noter que la somme de l'énergie du premier ordre et de l'ordre zéro permet de retrouver l'énergie Hartree-Fock :

$$E_i^{(0)} + E_i^{(1)} = E_{HF}. \quad (1.45)$$

**Au second ordre ( $k = 2$ ) :**

L'énergie de correction au second ordre se définit comme :

$$E_i^{(2)} = \sum_{j \neq i} \frac{\langle \psi_i^{(0)} | \hat{V} | \psi_j^{(0)} \rangle}{E_i^{(0)} - E_j^{(0)}}. \quad (1.46)$$

Ici, l'énergie totale a la forme suivante :

$$E_i = E_i^{(0)} + E_i^{(1)} + E_i^{(2)}, \quad (1.47)$$

$$E_i = E_{HF} + E_c^{MP2}. \quad (1.48)$$

Pour l'application numérique, l'approche MP2 consiste à calculer l'énergie Hartree-Fock et d'ajouter le terme correctionnel correspondant.

## 5.2. La théorie du cluster couplé

La théorie du cluster couplé (CC) [13] est souvent désignée comme la méthode Post Hartree-Fock la plus précise pour le calcul des effets de corrélation électronique dynamique. Contrairement à une extension perturbative (MP) ou à une extension linéaire de la fonction d'onde  $\psi$  (méthode Post Hartree-Fock de l'interaction de configuration (CI)), la méthode CC utilise, elle, une extension exponentielle :

$$|\psi_i\rangle = e^T |\psi_0\rangle, \quad (1.49)$$

$$|\psi_i\rangle = \left( 1 + \hat{T} + \frac{1}{2} \hat{T}^2 + \frac{1}{3!} \hat{T}^3 + \dots \right) |\psi_0\rangle. \quad (1.50)$$

L'opérateur  $\hat{T}$  est l'opérateur cluster qui se définit de la manière suivante :

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \dots + \hat{T}_N, \quad (1.51)$$

avec  $\hat{T}_1$  qui est l'opérateur des excitations singulières,  $\hat{T}_2$  l'opérateur des excitations doubles et ainsi de suite jusqu'à  $\hat{T}_N$ .

En reprenant l'équation de Schrödinger que l'on développe en série de Taylor, on peut retrouver :

$$|\psi_i\rangle = \left(1 + (\hat{T}_1 + \hat{T}_2 \dots) + \frac{1}{2}(\hat{T}_1 + \hat{T}_2 \dots)^2 + \dots\right) |\psi_0\rangle, \quad (1.52)$$

$$|\psi_i\rangle = \left(1 + \hat{T}_1 + \left(\hat{T}_2 + \frac{1}{2}\hat{T}_1^2\right) + \left(\hat{T}_3 + \hat{T}_1\hat{T}_2 + \frac{1}{3!}\hat{T}_1^3\right) + \dots\right) |\psi_0\rangle, \quad (1.53)$$

$$|\psi_i\rangle = |\psi_0\rangle + \sum_S c_S |S\rangle + \sum_D c_D |D\rangle + \sum_T c_T |T\rangle + \dots \quad (1.54)$$

Cette série est en pratique finie par le biais d'une troncation. Si l'on s'arrête au premier terme, en définissant  $\hat{T} = \hat{T}_1$ , on retrouve l'approximation CCS avec :

$$|\psi_i\rangle = |\psi_0\rangle + \sum_S c_S |S\rangle. \quad (1.55)$$

$$E_{cc} = E_0 + \sum_S c_S |S\rangle. \quad (1.56)$$

Si  $\hat{T} = \hat{T}_1 + \hat{T}_2$ , on appelle cette approximation CCSD (pour cluster couplé simple et double). On peut aller plus loin que cette approximation (CCSDT, CCSDTQ ...). Si  $\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3$ , les excitations triples  $\hat{T}_3$  peuvent être approximées, on notera alors la méthode CCSD(T).

Pour nos calculs, nous avons employé les méthodes CCSD et CCSD(T).

## 6. Les fonctions de base

Les fonctions de base représentent en chimie quantique un ensemble de fonctions permettant l'expression des orbitales. Il existe deux grands types de fonctions de bases, les fonctions de type Slater [14] (appelées STO, de l'anglais « *Slater Type Orbital* ») et les fonctions de type Gaussiennes [15] (appelées GTO, de l'anglais « *Gaussian Type Orbital* »).

Les fonctions de type Slater sont définies (en coordonnées cartésiennes) comme :

$$\chi_{a,b,c}^{STO}(x, y, z) = N x^a y^b z^c e^{-\zeta r}, \quad (1.57)$$

où  $N$  est une constante de normalisation,  $a, b$  et  $c$  contrôlent le moment angulaire  $L$  tel que  $L = a + b + c$  et  $\zeta$  est un exposant qui contrôle la taille de l'orbitale.

Pour des systèmes plus grands, le calcul numérique des intégrales de répulsion devient très coûteux en temps. Pour pallier cela on utilise souvent des fonctions GTO où l'on remplace la

partie exponentielle décroissante des fonctions STO par une fonction gaussienne  $e^{-\zeta r^2}$  ; avec la même notation, on a :

$$\chi_{a,b,c}^{GTO}(x, y, z) = N x^a y^b z^c e^{-\zeta r^2}. \quad (1.58)$$

Il faut noter que le produit de deux fonctions gaussiennes est également une gaussienne. Cependant, une fonction gaussienne ne peut pas représenter identiquement une fonction de Slater, les comportements étant différents lorsque  $r \rightarrow \infty$  et  $r \rightarrow 0$  (lorsque  $r$  tend vers 0 il y a présence d'un point de rebroussement, lorsque  $r$  tend vers l'infini la décroissance est trop rapide). Pour reproduire une fonction STO, on passe par une combinaison linéaire de plusieurs fonctions GTO afin d'obtenir une fonction dite « contractée » (CGTO) des fonctions gaussiennes définies précédemment, que l'on appelle « primitives » :

$$\chi_{a,b,c}^{CGTO}(x, y, z) = N \sum_{i=1}^n c_i x^a y^b z^c e^{-\zeta_i r^2}. \quad (1.59)$$

On obtient ainsi plusieurs types de fonctions de base :

- \_ Les bases dites « minimales », où l'on remplace chaque STO par une GTO ou CGTO.
- \_ Les bases « multiples » où l'on remplace chaque STO par plusieurs fonctions gaussiennes. Par exemple, dans une base double  $\zeta$ , on remplace chaque STO par deux GTO / CGTO, chacune étant caractérisée par son propre exposant.

Dans les bases « à valence séparée » seuls les électrons de la couche de valence sont traités par des bases multiples  $\zeta$ , les électrons de cœur sont traités à part avec une seule fonction GTO ou CGTO.

Il faut noter que l'on peut ajouter à chaque base des fonctions dites diffuses et des fonctions dites de polarisation. Une fonction diffuse est une fonction que l'on rajoute avec un faible exposant  $\zeta$ . Elles permettent de traiter la densité électronique loin du noyau dans le cas notamment des anions ou des interactions non covalentes.

Une fonction de polarisation est une fonction de moment angulaire plus élevé qui permet d'ajouter de la flexibilité aux fonctions de base, elles permettent une meilleure description de la distribution électronique (cela revient par exemple à rajouter une fonction de type d à une fonction de base avec des orbitales de valence p, une fonction de type f à une fonction de base avec des orbitales de valence d...).

En pratique, les bases les plus populaires sont les bases à valence séparées de Pople, qui ont typiquement comme forme  $N - MPP'G$  [16], dont la nomenclature générale est la suivante :

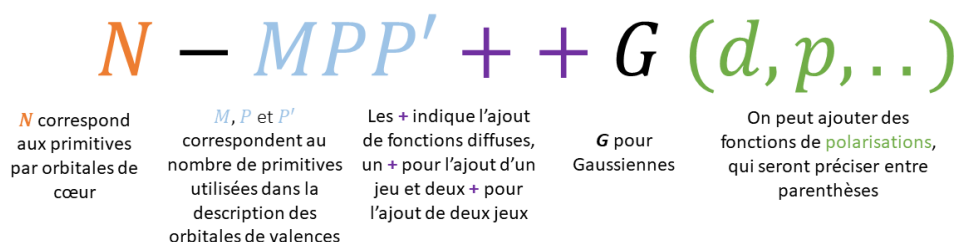


Figure 1.2: Schéma illustrant la nomenclature générale des bases de Pople

La base de Pople que nous avons principalement utilisée est  $6-311++G(2d,2p)$ .

Nous avons aussi utilisé des fonctions de bases plus récentes, les bases dites à corrélations consistantes polarisées (cc-p) développées par Dunning [17]. Ce sont des bases multiples pour les orbitales de valence seulement (dénotées  $V$ ) incluant des fonctions de polarisation (corrélées), permettant de converger vers la limite de base complète.

Une partie de l'ensemble des bases de Dunning peuvent se résumer dans le schéma suivant :

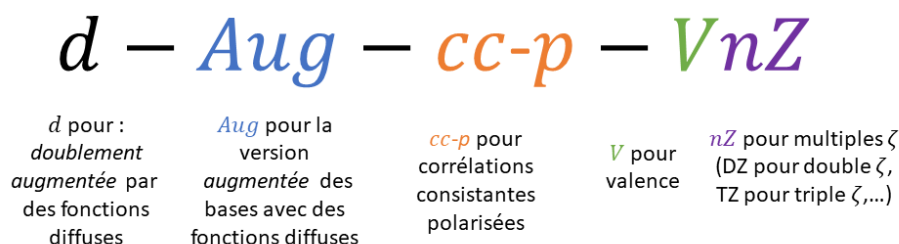


Figure 1.3: Schéma illustrant la nomenclature générale des bases de Dunning

La base de Dunning que nous avons principalement utilisée est  $Aug-cc-pVTZ$ .

Si l'on se place dans l'étude d'un système moléculaire  $AB$ , il est important de préciser qu'une utilisation de bases finies pour la détermination de l'énergie d'interaction entraîne une *erreur de superposition de bases*. Pour des calculs numériques, un complexe  $AB$  sera en général décrit par une base formée par la superposition des bases finies de  $A$  et de  $B$ . L'énergie d'interaction non corrigée se définit comme :

$$E_{int}^{AB} = E_{AB}^{AUB} - E_A^A - E_B^B, \quad (1.60)$$

où  $E_{int}^{AB}$  est l'énergie d'interaction du complexe  $AB$ ,  $E_{AB}^{AUB}$ ,  $E_A^A$  et  $E_B^B$  sont respectivement les énergies du complexe  $AB$  dans la base du complexe  $A \cup B$ , du fragment  $A$  dans la base associée à  $A$  et du fragment  $B$  dans la base associée à  $B$ .

La base  $A \cup B$  sera ainsi plus grande et entraînera un abaissement en valeur absolue de l'énergie d'interaction calculée, puisque l'énergie du fragment  $A$  dans la base  $A$  est supérieure à l'énergie du fragment  $A$  dans la base  $A \cup B$ , de même pour le fragment  $B$ . Cela crée l'erreur de superposition de base (appelée BSSE, de l'anglais « *Basis Set Superposition Error* »). Nous avons utilisé dans nos calculs d'énergie d'interaction l'approche proposée par Boys et Bernardi [18] du contrepoids, définissant l'énergie BSSE comme :

$$E_{AB}^{BSSE} = (E_A^A - E_A^{AUB}) + (E_B^B - E_B^{AUB}), \quad (1.61)$$

où les énergies de  $A$  et de  $B$  ( $E_A^A$  et  $E_B^B$ ) sont calculées dans la base du complexe  $AB$  ( $E_A^{AUB}$  et  $E_B^{AUB}$ ). L'énergie d'interaction *corrigée* devient ainsi :

$$E_{int}^{AB} = E_{AB}^{AUB} - E_A^A - E_B^B + E_{AB}^{BSSE}. \quad (1.62)$$

## 7. La théorie de la fonctionnelle de la densité

La théorie de la fonctionnelle de la densité est une alternative à la résolution de l'équation de Schrödinger dans le cadre de l'approximation de Born-Oppenheimer. Cette théorie, qui est fondée sur les deux principaux théorèmes d'Hohenberg et Kohn et par l'approche de Kohn et Sham qui seront détaillés par la suite, fait intervenir la densité électronique de l'état fondamental à la place de la fonction d'onde électronique. Pour une plus ample description, voir [19], [20] et [21].

La densité électronique de l'état fondamental  $\rho(\vec{r})$  d'un système à  $N$  électrons est définie comme l'intégrale sur toutes les coordonnées spatiales et de spin des électrons, à l'exception d'une coordonnée spatiale, du carré du module de la fonction d'onde du système :

$$\rho(\vec{r}_1) = N \int \dots \int |\psi_{el}(x_1, x_2, \dots x_N)|^2 d\sigma_1 d^4x_2 \dots d^4x_N, \quad (1.63)$$

$\rho(\vec{r})$  permet de déterminer le nombre moyen d'électrons dans un volume infinitésimal  $d^3r$  centré en  $\vec{r}$ . Ainsi l'intégration de la densité électronique sur l'ensemble des coordonnées spatiales permet de définir le nombre total d'électrons du système :

$$\int \rho(\vec{r}) d^3r = N. \quad (1.64)$$

Pour l'étude de systèmes chimiques, la densité électronique est une fonction définie positive à trois variables spatiales seulement, contrairement aux  $3N$  variables spatiales et  $N$  variables de spin de la fonction d'onde. Une autre caractéristique intéressante de la densité électronique est qu'elle est une observable qui peut être mesurée expérimentalement (par exemple par diffraction des rayons X) [22].

### 7.1. Le premier théorème de Hohenberg et Kohn

Le premier théorème qui constitue le fondement de la DFT a été établi par Hohenberg et Kohn en 1964 [23]. Si l'on considère un système chimique à  $N$  électrons soumis à un potentiel externe  $v_{ext}(\vec{r})$ , la résolution de l'équation de Schrödinger permet d'en obtenir la fonction d'onde électronique de l'état fondamental et de ce fait la densité électronique fondamentale correspondante  $\rho(\vec{r})$ . Il existe une correspondance univoque entre le potentiel externe et la densité électronique pour l'état fondamental stationnaire :

$$v_{ext}(\vec{r}) \rightarrow \rho(\vec{r}), \quad (1.65)$$

Hohenberg et Kohn ont alors montré que cette correspondance peut être inversée, la densité électronique peut ainsi permettre la détermination du potentiel externe (à une constante additive près) :

$$\rho(\vec{r}) \rightarrow v_{ext}(\vec{r}) + \text{constante}, \quad (1.66)$$

Par conséquent, cela permet de déduire les propriétés du système à partir de la densité électronique fondamentale. Dans ce formalisme, le potentiel externe est une fonctionnelle unique de la densité électronique, ainsi l'énergie totale du système est une fonctionnelle de la densité électronique.

Le premier théorème se démontre par un raisonnement par l'absurde. Considérons deux potentiels externes différents  $v_{ext}^{(1)}(\vec{r})$  et  $v_{ext}^{(2)}(\vec{r})$  qui ont des fonctions d'ondes différentes  $\psi_0^{(1)}$  et  $\psi_0^{(2)}$ , associés à une même densité de l'état fondamental  $\rho(\vec{r})$ , les deux potentiels auront ainsi deux hamiltoniens différents  $\hat{H}^{(1)}$  et  $\hat{H}^{(2)}$  avec leurs deux énergies associées  $E_1$  et  $E_2$ . Le principe variationnel permet d'écrire :



$$E_1 = \langle \psi_0^{(1)} | \hat{H}^{(1)} | \psi_0^{(1)} \rangle < \langle \psi_0^{(2)} | \hat{H}^{(1)} | \psi_0^{(2)} \rangle = \langle \psi_0^{(2)} | \hat{H}^{(2)} + \hat{V}^{(1)} - \hat{V}^{(2)} | \psi_0^{(2)} \rangle$$

$$E_1 < E_2 + \int [v_{ext}^{(1)}(\vec{r}) - v_{ext}^{(2)}(\vec{r})] \rho(\vec{r}) d^3r. \quad (1.67)$$

En échangeant les rôles des systèmes 1 et 2 par symétrie, on obtient similairement l'inégalité suivante :

$$E_2 < E_1 + \int [v_{ext}^{(2)}(\vec{r}) - v_{ext}^{(1)}(\vec{r})] \rho(\vec{r}) d^3r. \quad (1.68)$$

Une combinaison des deux équations précédentes revient à écrire le résultat suivant :

$$E_1 + E_2 < E_1 + E_2. \quad (1.69)$$

Il ne peut ainsi y avoir deux potentiels externes avec la même densité électronique. Cette inégalité stricte n'est valable que si l'état fondamental n'est pas dégénéré.

Hohenberg et Kohn ont ainsi défini la fonctionnelle universelle (indépendante du potentiel externe)  $F_{HK}[\rho]$  en exprimant l'énergie électronique  $E[\rho]$  comme une fonctionnelle de la densité électronique :

$$E[\rho] = F_{HK}[\rho] + \int \rho(\vec{r}) v_{ext}(\vec{r}) d^3r. \quad (1.70)$$

## 7.2. Le deuxième théorème de Hohenberg et Kohn

Le deuxième théorème de Hohenberg et Kohn fait appel au principe variationnel vu précédemment. Si l'on considère un système soumis à un potentiel externe  $v_{ext}$ , le théorème stipule que l'énergie électronique  $E[\rho]$  satisfait une propriété variationnelle : l'énergie de l'état fondamental  $E_0$  du système est obtenue en minimisant la fonctionnelle d'énergie, la densité associée correspondra à la densité exacte de l'état fondamental. Dans la reformulation (dite de « recherche contrainte ») par Levy des deux théorèmes d'Hohenberg-Kohn, on peut écrire exactement :

$$E_0 = \min_{\psi} \langle \psi | \hat{H} | \psi \rangle = \min_{\rho} \left\{ \min_{\psi \rightarrow \rho} \langle \psi | \hat{H} | \psi \rangle \right\} \quad (1.71)$$

$$\begin{aligned}
&= \min_{\rho} \left\{ \min_{\psi \rightarrow \rho} \left\{ \langle \psi | \hat{T}_e + \hat{V}_{ee} + \sum_{i=1}^N v_{ext}(\vec{r}) | \psi \rangle \right\} \right\} \\
&= \min_{\rho} \left\{ \min_{\psi \rightarrow \rho} \{ \langle \psi | \hat{T}_e + \hat{V}_{ee} | \psi \rangle \} + \int \rho(\vec{r}) v_{ext}(\vec{r}) d^3r \right\} \\
&= \min_{\rho} \left\{ F_{HK}[\rho] + \int \rho(\vec{r}) v_{ext}(\vec{r}) d^3r \right\}.
\end{aligned}$$

Ces deux théorèmes stipulent ainsi l'existence de la fonctionnelle exacte  $F_{HK}[\rho] = \min_{\psi \rightarrow \rho} \{ \langle \psi | \hat{T}_e + \hat{V}_{ee} | \psi \rangle \}$  qui permet d'exprimer l'énergie électronique par la seule densité électronique. En revanche, la forme analytique de la fonctionnelle est inconnue explicitement à l'heure actuelle.

### 7.3. L'approche de Kohn-Sham

Suite aux difficultés d'approximer directement la fonctionnelle d'énergie totale  $F_{HK}[\rho]$ , Kohn et Sham [24] ont proposé de décomposer la fonctionnelle d'énergie totale en passant par un système *fictif* sans interactions et de même densité que le système *réel* :

$$F_{HK}[\rho] = T_s[\rho] + E_{HXC}[\rho], \quad (1.72)$$

où  $T_s[\rho]$  est la fonctionnelle d'énergie cinétique des électrons sans interaction du système fictif, qui peut se définir par une formulation dans une approche variationnelle contrainte, comme :

$$T_s[\rho] = \min_{\phi \rightarrow \rho} \langle \phi | \hat{T} | \phi \rangle = \langle \phi[\rho] | \hat{T} | \phi[\rho] \rangle. \quad (1.73)$$

la minimisation est effectuée ici sur une fonction d'onde exacte du système sous la forme d'un déterminant de Slater  $\phi$  construit à partir des orbitales dites de Kohn-Sham  $\phi_i^{KS}$  (que l'on choisit ici par commodité réelles), l'énergie électronique s'exprime :

$$E[\{\phi_{KS}\}] = \sum_{i=1}^N \int \phi_i^{KS}(\vec{r}) \left( -\frac{1}{2} \nabla^2 + v_{ext}(\vec{r}) \right) \phi_i^{KS}(\vec{r}) d^3r + E_{HXC}[\rho], \quad (1.74)$$

où la densité électronique du système est exprimée en termes d'orbitales comme :

$$\rho(\vec{r}) = \sum_{i=1}^N |\phi_i^{KS}(\vec{r})|^2. \quad (1.75)$$

L'énergie et la densité exacte peuvent ainsi être obtenues par approche variationnelle :

$$\begin{aligned}
E_0 &= \min_{\rho} \left\{ F_{HK}[\rho] + \int \rho(\vec{r}) v_{ext}(\vec{r}) d^3r \right\} \\
&= \min_{\rho} \left\{ \min_{\phi \rightarrow \rho} \langle \phi | \hat{T} | \phi \rangle + E_{HXC}[\rho] + \int \rho(\vec{r}) v_{ext}(\vec{r}) d^3r \right\} \\
&= \min_{\rho} \left\{ \min_{\phi \rightarrow \rho} \{ \langle \phi | \hat{T} + \hat{V}_{ne} | \phi \rangle \} + E_{HXC}[\rho_{\phi}] \right\} \\
&= \min_{\phi} \{ \langle \phi | \hat{T} + \hat{V}_{ne} | \phi \rangle + E_{HXC}[\rho_{\phi}] \},
\end{aligned} \tag{1.76}$$

Seule la fonctionnelle d'Hartree-échange-corrélation  $E_{HXC}[\rho]$  doit ainsi être approximée par une fonctionnelle de la densité électronique, que l'on écrit :

$$E_{HXC}[\rho] = E_H[\rho] + E_{XC}[\rho]. \tag{1.77}$$

avec  $E_H[\rho]$  qui correspond à la fonctionnelle d'énergie de Hartree représentant la répulsion électrostatique :

$$E_H[\rho] = \frac{1}{2} \int \int \frac{\rho(\vec{r})\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d^3r d^3r'. \tag{1.78}$$

La fonctionnelle d'énergie d'échange-corrélation  $E_{XC}[\rho]$  se décompose elle aussi comme une somme de la fonctionnelle d'énergie d'échange  $E_X[\rho]$  et de corrélation  $E_C[\rho]$  :

$$E_{XC}[\rho] = E_X[\rho] + E_C[\rho]. \tag{1.79}$$

Ici, la fonctionnelle d'échange-corrélation inclut deux contributions : une correction d'énergie cinétique et une correction à la répulsion bi-électronique :

$$E_{XC}[\rho] = (T[\rho] - T_s[\rho]) + (V_{ee} - E_H[\rho]). \tag{1.80}$$

La minimisation de l'énergie s'effectue en introduisant le Lagrangien suivant :

$$\mathcal{L}[\{\phi_i^{KS}\}] = E[\{\phi_i^{KS}\}] - \sum_{i=1}^N \varepsilon_i \left( \int \phi_i^{KS}(\vec{r}) \phi_i^{KS}(\vec{r}) d^3r - 1 \right), \tag{1.81}$$

avec  $\varepsilon_i$  défini comme le multiplicateur de Lagrange associé à la condition de normalisation. La minimisation obéit à :

$$\frac{\delta \mathcal{L}}{\delta \phi_i^{KS}(\vec{r})} = 0. \tag{1.82}$$

Le calcul de la dérivée fonctionnelle  $\frac{\delta \mathcal{L}}{\delta \phi_i^{KS}(\vec{r})}$  amène à l'équation suivante :

$$\left(-\frac{1}{2}\nabla^2 + v_{ext}(\vec{r})\right)\phi_i^{KS}(\vec{r}) + \frac{\delta E_{HXC}[\rho]}{\delta \phi_i^{KS}(\vec{r})} = \varepsilon_i \phi_i^{KS}(\vec{r}), \quad (1.83)$$

où  $\frac{\delta E_{HXC}[\rho]}{\delta \phi_i^{KS}(\vec{r})}$  peut s'exprimer par une dérivation en chaîne comme :

$$\frac{\delta E_{HXC}[\rho]}{\delta \phi_i^{KS}(\vec{r})} = \int \frac{\delta E_{HXC}[\rho]}{\delta \rho(\vec{r}')} \frac{\delta \rho(\vec{r}')}{\delta \phi_i^{KS}(\vec{r})} d\vec{r}', \quad (1.84)$$

En définissant le potentiel d'Hartree-échange-corrélation  $v_{HXC}(\vec{r})$  :

$$v_{HXC}(\vec{r}) = \frac{\delta E_{HXC}[\rho]}{\delta \rho(\vec{r})}. \quad (1.85)$$

On aboutit aux équations de Kohn-Sham :

$$\left(-\frac{1}{2}\nabla^2 + v_{ext}(\vec{r}) + v_{HXC}(\vec{r})\right)\phi_i^{KS}(\vec{r}) = \varepsilon_i \phi_i^{KS}(\vec{r}), \quad (1.86)$$

où :

$$\hat{h}_{KS}(\vec{r}) = -\frac{1}{2}\nabla^2 + \hat{v}_{KS}(\vec{r}),$$

$$v_{KS}(\vec{r}) = v_{ext}(\vec{r}) + v_{HXC}(\vec{r}),$$

avec  $v_{KS}(\vec{r})$  défini comme le potentiel de Kohn-Sham et  $\hat{h}_{KS}$  l'Hamiltonien de Kohn-Sham. Elles définissent le système fictif à  $N$  électrons dans le potentiel externe  $v_{KS}(\vec{r})$  qui a la même densité électronique que l'état fondamental du système réel que l'on étudie.

Le potentiel  $v_{HXC}(\vec{r})$  peut se décomposer selon :

$$v_{HXC}(\vec{r}) = v_H(\vec{r}) + v_{XC}(\vec{r}), \quad (1.87)$$

ici  $v_H(\vec{r})$  est le potentiel de Hartree et  $v_{XC}(\vec{r})$  le potentiel d'échange-corrélation tel que :

$$v_H(\vec{r}) = \frac{\delta E_H[\rho]}{\delta \rho(\vec{r})}, \quad (1.88)$$

et

$$v_{XC}(\vec{r}) = \frac{\delta E_{XC}[\rho]}{\delta \rho(\vec{r})}. \quad (1.89)$$

Numériquement, la résolution itérative des équations de Kohn-Sham par un cycle auto-cohérent est similaire à la résolution des équations Hartree-Fock. Le schéma suivant détaille la procédure SCF pour une application DFT [25] :

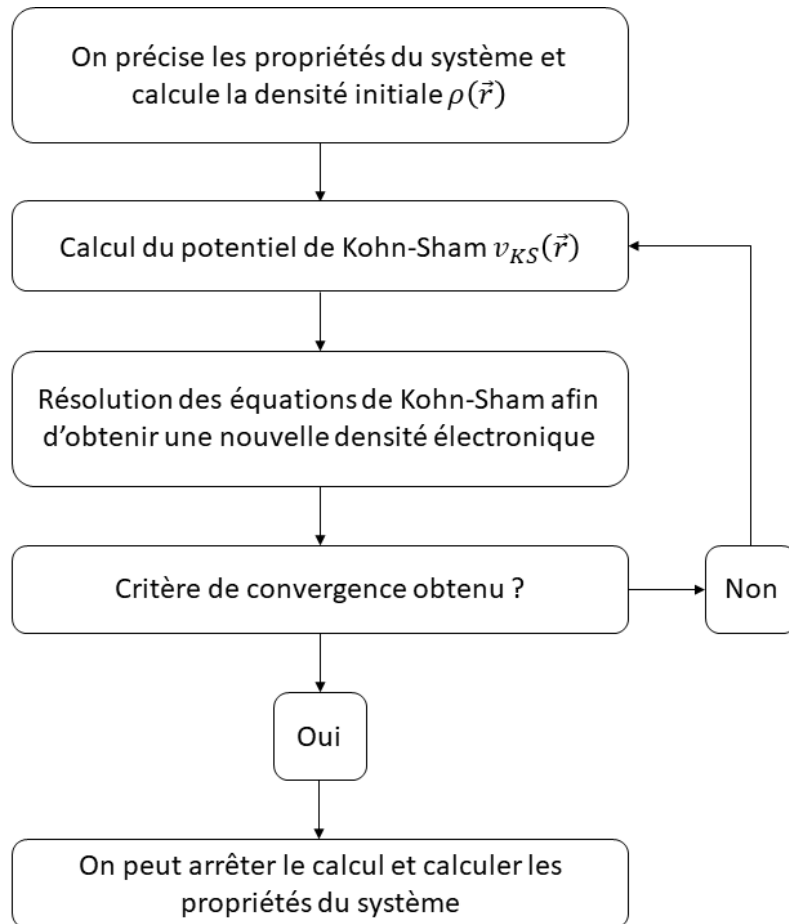


Figure 1.4: Schéma illustrant la procédure de résolution itérative des équations de Kohn-Sham

#### 7.4. Les fonctionnelles d'échange-corrélation

Pour une approche numérique, la fonctionnelle d'échange-corrélation  $E_{XC}[\rho]$  qui reste à déterminer dans la méthode de Kohn-Sham est approchée par plusieurs approximations. Les différentes  $E_{XC}$  qui ont été proposées peuvent être classées sur l'échelle suivante (dite échelle de Jacob de la DFT par Perdew [26]) :

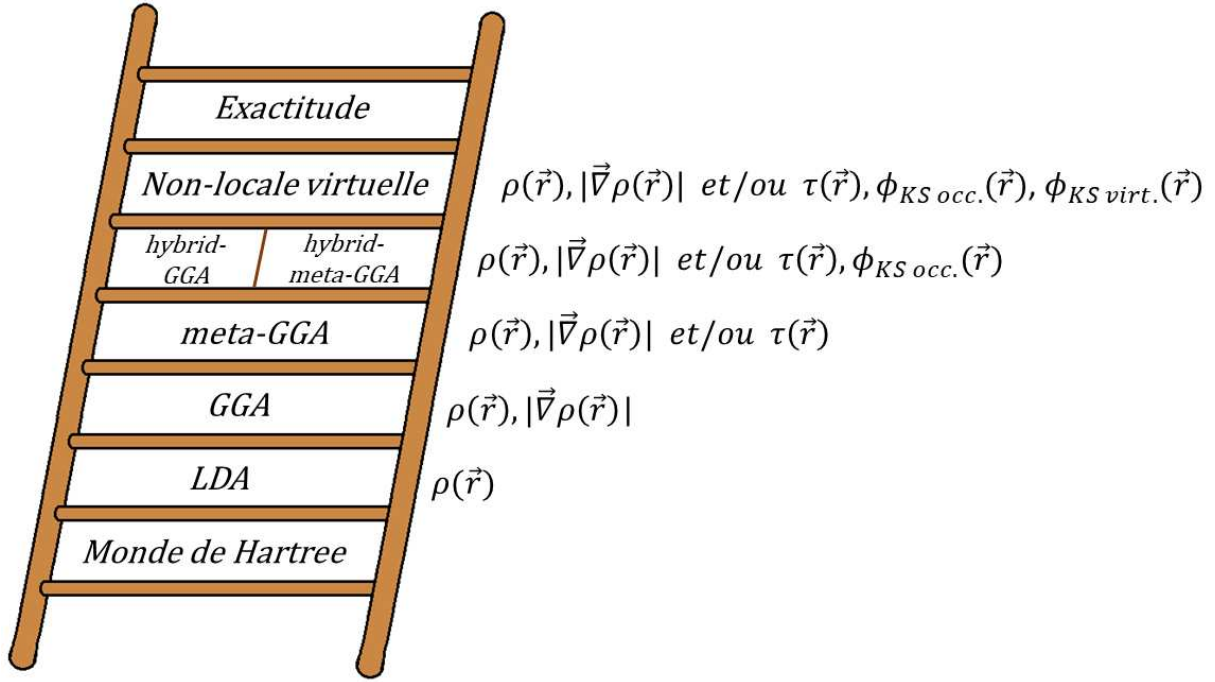


Figure 1.5: Echelle de Jacob de la théorie de la fonctionnelle de la densité

Les « échelons » de cette échelle correspondent aux différentes classes de fonctionnelles d'échange-corrélation et sont détaillés par la suite :

- Le premier échelon correspond aux premières générations de fonctionnelles, qui font appel à une approximation de la densité locale [27] (LDA, de l'anglais « *Local Density Approximations* »). Le système est considéré comme un gaz uniforme d'électrons, l'énergie d'échange-corrélation  $E_{XC}^{LDA}[\rho]$  est de la forme :

$$E_{XC}^{LDA}[\rho] = \int \rho(\vec{r}) \varepsilon_{XC}[\rho] d\vec{r}, \quad (1.90)$$

$\varepsilon_{XC}[\rho]$  désigne l'énergie d'échange-corrélation d'un gaz uniforme par électron, cette fonction peut être décomposée par une partie d'échange  $\varepsilon_X[\rho]$  et de corrélation  $\varepsilon_C[\rho]$ . La partie d'échange utilisée est celle définie par Dirac :

$$\varepsilon_X[\rho] = -\frac{3}{4} \left( \frac{3}{\pi} \right)^{1/3} \rho(\vec{r})^{1/3}, \quad (1.91)$$

deux exemples de fonctionnelles  $\varepsilon_C[\rho]$  sont les fonctionnelles *VWN* de Vosko, Wilk et Nusair [28] et *PW92L* de Perdew et Wang [29].

- Le second échelon est l'approximation du gradient généralisé (GGA, de l'anglais « *Generalized Gradient Approximations* »), ici cette approximation améliore la définition de la fonctionnelle d'échange corrélation LDA en incluant la dérivée spatiale de la densité électronique  $\vec{\nabla}\rho$  :

$$E_{XC}^{GGA}[\rho] = \int \rho(\vec{r}) \varepsilon_{XC}[\rho, \vec{\nabla}\rho] d\vec{r}. \quad (1.92)$$

En pratique,  $E_{XC}^{GGA}$  est séparée en une fonctionnelle d'échange et une fonctionnelle de corrélation. L'un des premiers exemples dans cette classe de fonctionnelle est BLYP [30], qui utilise B88 (Becke 1988) pour la fonctionnelle d'échange et qui est combinée avec LYP pour le terme de corrélation (1986). Parmi les plus utilisées, on peut citer PBE (Perdew et al.) [31] et PW91 (Ziesche, Eschrig et Perdew) [32].

- Le troisième échelon correspond aux fonctionnelles dites *meta-GGA* qui ajoute à la fonctionnelle une dépendance aux dérivées d'ordre supérieur, par le biais du Laplacien de la densité électronique  $\nabla^2\rho$  et/ou de la densité d'énergie cinétique  $\tau(\vec{r}) = \sum_i |\nabla\phi_i(\vec{r})|^2$ . Comme exemples de fonctionnelles méta-GGA, on peut citer *TPSS* [33], *VSXC* [34], ou *M06L* [35].
- Le quatrième échelon dénote les fonctionnelles hybrides introduites par Becke en 1993 [36]. Leurs noms viennent du fait qu'une partie d'échange exact  $E_{XX}$  est ajoutée à l'échange DFT  $E_{XC}^{DFT}$  dans leurs expressions. Cette approche est basée sur une connexion adiabatique entre le système fictif et réel. L'expression de ces fonctionnelles d'échange corrélation est de forme :

$$E_{XC}[\rho] = (1 - a)E_{XC}^{DFT} + aE_{XX}, \quad (1.93)$$

le paramètre  $a$  donne la part d'échange exact dans l'échange total. L'une des fonctionnelles les plus utilisées en chimie quantique est la fonctionnelle *B3LYP* [37] qui est de forme :

$$E_{XC}^{B3LYP} = E_X^{LDA} + a_0(E_{XX} - E_X^{LDA}) + a_X(E_X^{GGA} - E_X^{LDA}) + E_C^{LDA} + a_C(E_C^{GGA} - E_C^{LDA}). \quad (1.94)$$

Dans son expression, on retrouve les fonctionnelles d'échange et de corrélation des échelons précédents, ici  $a_0 = 0,20$ ,  $a_X = 0,72$  et  $a_C = 0,81$ . La fonctionnelle hybride que nous avons principalement utilisé est la fonctionnelle *M06-2X* [38], appartenant à l'ensemble des fonctionnelles *M06* dont les paramètres sont ajustés de manière empirique. *M06-2X* possède 54% d'échange exact.

- L'avant-dernier échelon représente les dernières classes de fonctionnelles, les fonctionnelles non locales virtuelles. Dans cette classe de fonctionnelles on trouve les fonctionnelles doubles hybrides, qui en plus d'une part d'échange exact contiennent une part de corrélation perturbative au second ordre (PT2), ajoutant une dépendance aux orbitales de Kohn-Sham non occupées. La fonctionnelle d'échange-corrélation s'exprime :

$$E_{XC}[\rho] = (1 - a_X)E_X^{DFT} + a_X E_{XX} + (1 - a_C)E_C^{DFT} + a_C E_C^{PT2}. \quad (1.95)$$

La quantité d'échange exact est en général supérieure à celle utilisée dans les simples hybrides. Quelques exemples de fonctionnelles doubles hybrides sont B2-PLYP (où  $a_X = 0.53$  et  $a_C = 0.27$ ) [39] et B2K-PLYP [40] (où  $a_X = 0.42$  et  $a_C = 0.72$ ).

L'équation (1.95) correspond à une fonctionnelle hybride globale. Il existe des hybrides locales, comme celle fondée sur la séparation de portée bi électronique [41]. Par exemple comme  $\omega$ B97X-D que nous avons utilisée dans nos calculs [42].

Une méthode similaire pour inclure des effets de corrélation électronique de longue portée, les effets de dispersion, est de compléter l'énergie calculée en ajoutant un terme énergétique de dispersion  $E_{disp}$  de type classique. Ce terme est ajouté indépendamment du cycle auto-cohérent, tel que  $E_{totale} = E_{DFT} + E_{disp}$ . Parmi les méthodes proposées, le terme le plus utilisé est la correction D3 développée par Grimme [43] où :

$$E_{disp} = -s_6 \sum_{i>j}^{N_{At}} \frac{C_6^{ij}}{(R_{ij})^6} f_{damp}(R_{ij}), \quad (1.96)$$

$$f_{damp}(R_{ij}) = \frac{1}{1 + e^{-\alpha_6(R_{ij}/R_{vdW}-1)}}, \quad (1.97)$$

ici  $C_6^{ij}$  sont les coefficients de dispersion,  $R_{ij}$  sont les distances inter atomiques. Une fonction d'« amortissement » (*damping* en anglais) permet alors de moduler le terme correctionnel (dont la variation est mesurée par  $\alpha_6$ ) en fonction des distances de Van der Waals  $R_{vdW}$ .  $s_6$  est un terme propre aux fonctionnelles utilisées.

Enfin, pour l'obtention de la fonctionnelle exacte, il est important de citer une partie des recherches en cours dans ce domaine [44].



- [1] L. De Broglie, *Note aux comptes rendus de l'Académie des sciences*. 177 (1923) 630-632.
- [2] E. Schrödinger, *Annalen der Physik*, 384 (1926) 361-376.
- [3] *Dans 2 lettres*, voir J. Lacki, *Archive for History of Exact Sciences*, 54 (2000) 279-318.
- [4] M. Born, J. R. Oppenheimer, *Annalen der Physik*, 84 (1927) 457-484.
- [5] a) D. R. Hartree, *Proc. Camb. Phil. Soc.* 24 (1928) 89–132.  
       b) V. Fock, *Zeitschrift für Physik*, 61 (1930) 126–148.
- [6] W. Pauli, *Zeitschrift für Physik*. 31 (1925) 765–783.
- [7] L. Brillouin, *Les champs 'self-consistents' de Hartree et de Fock*, Hermann & Cie, Paris, (1934).
- [8] a) C. C. J. Roothaan, *Reviews of Modern Physics*, 23 (1951) 69-89.  
       b) G. G. Hall, *Proceedings of the Royal Society A*, 205 (1951) 541-552.
- [9] E. Wigner, F. Seitz, *Physical Review* 46 (1934) 509-524.
- [10] P. O. Löwdin, *Advances in Chemical Physics*, 2 (1959) 207-322.
- [11] C. Møller, M. S. Plesset, *Physical Review*, 46 (1934) 618-622.
- [12] a) J. W. S. Rayleigh, *Theory of Sound. I (2nd ed.)*. London: Macmillan (1894).  
       b) E. Schrödinger, *Annalen der Physik*, 80 (1926) 437–490.
- [13] J. Čížek, *J. Chem. Phys.* 45 (1966) 4256–4266.
- [14] J. C. Slater, *Physical Review*, 36 (1930) 57-64.
- [15] S. F. Boys, *Proc. R. Soc. Lond. A*. 200 (1950) 542–554.
- [16] R. Ditchfield, W. J. Hehre, J. A. Pople, *J. Chem. Phys.* 54 (1971) 724–728.
- [17] T. H. Dunning, *J. Chem. Phys.* 90 (1989) 1007-1023.
- [18] S. F. Boys, F. Bernardi, *Molecular Physics* 19 (1970) 553–566.
- [19] R. G. Parr, W. Yang, *Density Functional Theory of Atoms and Molecules*, Oxford University Press, (1994)
- [20] A. Szabo, N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Dover Publications, (1996)
- [21] [https://www.lct.jussieu.fr/pagesperso/toulouse/enseignement/introduction\\_dft.pdf](https://www.lct.jussieu.fr/pagesperso/toulouse/enseignement/introduction_dft.pdf)

- [22] A. W. Hull, *J. Am. Chem. Soc.* 41 (1919) 1168–1195.
- [23] P. Hohenberg, W. Kohn, *Phys. Rev.* 136 (1964) B864
- [24] W. Kohn, L. J. Sham, *Phys. Rev.* 140 (1965) A1133
- [25] J. C. Slater, *The self-consistent field for Molecules and Solids : Quantum Thoery of Molecules and Solids*, McGraw Hill New York, (1974)
- [26] J. P. Perdew, K. Schmidt, In *Density Functional Theory and Its Applications to Materials* V. E. Van Doren, K. Van Alseoy, P. Geerlings, Eds New York, (2001).
- [27] J. P. Perdew, A. Zunger, *Phys. Rev.* 823 (1981) 5048-5079
- [28] S. H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* 58 (1980) 1200-1211.
- [29] J. P. Perdew, Y. Wang, *Phys. Rev. B* 45 (1992) 13244.
- [30] A. D. Becke, *Phys. Rev. A* 38 (1988) 3098.
- [31] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* 77 (1996) 3865.
- [32] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, C. Fiolhais *Phys. Rev. B* 46 (1992) 6671.
- [33] J. Tao, J. P. Perdew, V. N. Staroverov, G. E. Scuseria, *Phys. Rev. Lett.* 91 (2003) 146401.
- [34] T. V. Voorhis, G. E. Scuseria, *J. Chem. Phys.* 109 (1998) 400-410.
- [35] Y. Zhao, D. G. Truhlar, *J. Chem. Phys.* 125 (2006) 194101.
- [36] A. D. Becke, *J. Chem. Phys.* 98 (1993) 1372–1377.
- [37] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, *J. Phys. Chem.* 98 (1994) 11623-11627.
- [38] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* 120 (2006) 215–241.
- [39] S. Grimme, *J. Chem. Phys.* 124 (2006) 034108.
- [40] A. Tarnopolsky, A. Karton, R. Sertchook, D. Vuzman, J. M. L. Martin, *J. Phys. Chem. A.* 112 (2008) 3-8.
- [41] A. Savin, H. J. Flad, *Int. J. Quantum Chem.* 56 (1995) 327-332.
- [42] D. Chai, M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 10 (2008) 6615-6620
- [43] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* 132 (2010) 154104.
- [44] a) N. Mardirossian, M. Head-Gordon, *Mol. Phys.* 115 (2017) 2315-2372

- b) M. G. Medvedev, I. S. Bushmarinov, J. Sun, J. P. Perdew, K. A. Lyssenko, *Science* 355 (2017) 49-52
- c) S. A. Sato, A Rubio, *Phys. Rev. A*, 101 (2020) 012510

## CHAPITRE 2 : Les méthodes théoriques interprétatives

|                                                            |    |
|------------------------------------------------------------|----|
| 1. La réactivité chimique .....                            | 45 |
| 2. La surface d'énergie potentielle.....                   | 45 |
| 3. Modélisation du solvant.....                            | 46 |
| 4. La DFT Conceptuelle .....                               | 47 |
| 4.1. Fondements de la DFT conceptuelle .....               | 47 |
| 4.2. Ensembles de représentations.....                     | 49 |
| 4.2.1 L'ensemble canonique.....                            | 49 |
| 4.2.2. L'ensemble grand canonique.....                     | 51 |
| 4.2.3. L'ensemble isomorphique .....                       | 53 |
| 4.2.4. L'ensemble grand isomorphique .....                 | 54 |
| 4.2.5. Relations de Maxwell .....                          | 55 |
| 4.3. Descripteurs globaux.....                             | 56 |
| 4.3.1. Issus directement d'une dérivée de l'énergie.....   | 56 |
| 4.3.2. Descripteurs Ad hoc.....                            | 59 |
| 4.4. Descripteurs locaux et non locaux .....               | 62 |
| 4.4.1. Issus directement d'une dérivée de l'énergie.....   | 62 |
| 4.4.2. Descripteurs composites .....                       | 65 |
| 4.5. Principes de réactivité .....                         | 67 |
| 4.5.1. Principe d'égalisation des électronégativités ..... | 67 |
| 4.5.2. Principe des acides et des bases durs ou mous ..... | 69 |
| 4.5.3. Principe de dureté maximale.....                    | 70 |
| 4.5.4. Principe d'électrophilie minimale.....              | 71 |
| 4.5.5. Principe de polarisabilité minimale .....           | 73 |
| 5. Analyse de population .....                             | 74 |
| 5.1. Basées sur les orbitales moléculaires .....           | 74 |
| 5.1.1. Analyse de Population de Mulliken.....              | 74 |
| 5.1.2. Analyse de Population Naturelle .....               | 75 |
| 5.2. Basées sur le potentiel électrostatique.....          | 76 |
| 5.3. Basées sur la densité électronique.....               | 76 |
| 5.3.1. Analyse de population Hirshfeld .....               | 76 |
| 5.3.2. La théorie quantique AIM.....                       | 77 |



## 1. La réactivité chimique

Dans le chapitre précédent, nous avons abordé les notions théoriques principales permettant de calculer les propriétés d'une molécule. Nous allons dans cette partie nous concentrer sur l'interprétation des propriétés calculées d'une molécule et aborder la notion de réactivité chimique.

On définit une réaction chimique (ou « transformation chimique » si l'on souhaite paraphraser Lavoisier[1]) comme deux molécules qui interagissent entre elles afin d'obtenir de nouvelles molécules.

Grâce à l'interprétation des résultats des calculs par les chimistes théoriciens, on peut apporter une meilleure compréhension des réactions chimiques et une meilleure prédiction de ces dernières. Ainsi, un nombre important de concepts afin de rationaliser ces résultats ont été développés ; parmi eux, on peut citer la théorie des orbitales frontières, la théorie de la liaison de valence ou la théorie de l'état de transition.

L'une des principales théories récentes pour l'étude de la réactivité chimique au sein de la théorie de la fonctionnelle de la densité est la DFT *conceptuelle* qui est présentée dans ce chapitre.

## 2. La surface d'énergie potentielle

Pour modéliser une réaction chimique et les interactions entre des systèmes moléculaires, on emploie très souvent en chimie computationnelle la notion de surface d'énergie potentielle [2]. Ces surfaces d'énergie potentielle décrivent l'énergie du système en fonction d'une ou plusieurs coordonnées (c'est une fonction à  $3N-6$  variables lorsque tous les degrés de liberté nucléaires sont considérées). Si l'on utilise qu'une seule coordonnée, on parlera de courbe d'énergie potentielle ou de profil énergétique. Ces surfaces prennent alors en compte tous les changements de coordonnées et les différentes orientations possibles de la molécule. Ainsi, une molécule qui va changer de conformation (par exemple durant une réaction) va voir son énergie potentielle évoluer.

Sur la surface d'énergie potentielle, les points caractéristiques (points stationnaires) sont :

\_ Les minima qui vont correspondre à toutes les conformations dont un changement de coordonnées suivant une quelconque direction impliquera une augmentation de l'énergie. Ils

représentent les différents conformères de la molécule et les intermédiaires de réaction. Un minimum peut être *local* dans une région définie, ou *global* sur l'ensemble de la surface qui correspond à la géométrie du système de plus basse énergie

\_ Les points « selles » d'ordre 1 qui vont correspondre à un maximum local d'énergie dans une direction et à un minimum local dans toutes les autres directions. Un point selle est une structure (ou état) de transition de la molécule entre deux minima d'énergie.

Pour chacun des points stationnaires, les composantes du vecteur gradient (dérivées premières de l'énergie par rapport aux coordonnées nucléaires) sont nulles. Pour examiner une structure sur la surface d'énergie potentielle, on regardera la matrice Hessienne (ensemble des dérivées secondes de l'énergie par rapport aux coordonnées nucléaires). Si toutes les valeurs propres de la Hessienne sont positives on aura un minimum et un maximum si toutes les valeurs propres sont négatives. Si toutes les valeurs propres de la matrice sont positives sauf une, on aura un point selle d'ordre 1 qui s'interprète comme un état de transition.

Les valeurs propres de la Hessienne permettent de déterminer les fréquences de vibrations pour une structure donnée. Si une valeur propre est négative, on parlera de fréquence *imaginaire*. Le nombre de ces fréquences imaginaires définira un point stationnaire : une structure avec  $n$  fréquences imaginaires est un point selle d'ordre  $n$ . Un minimum n'aura pas de fréquence imaginaire et une structure de transition aura 1 fréquence imaginaire puisqu'il s'agit d'un point selle d'ordre 1.

### 3. Modélisation du solvant

La plupart des réactions chimiques s'effectuent en solution et le solvant utilisé a bien souvent des effets ou des interactions non négligeables pendant la réaction. En chimie numérique, il existe deux grandes méthodes pour prendre en compte le solvant lors des calculs : 1) de manière *directe* où l'on va placer des molécules de solvant autour des réactifs afin de représenter les effets du solvant, on parlera alors de méthode explicite, ou 2) de manière *indirecte* où les effets du solvant sont représentés par une perturbation, on parlera alors de méthode implicite. On peut également utiliser des méthodes hybrides, qui se situent entre les deux méthodes précédentes.

Dans nos études, nous avons choisi la méthode implicite car elle permet un gain de temps considérable lors des calculs. Le solvant n'est pas représenté directement par des molécules mais comme un continuum : la molécule étudiée est placée dans une *cavité* entourée d'un

continuum (on parle ainsi de champ réactionnel SCRF [3] de l'anglais « *Self Consistent Reaction Field* »). Le solvant qui est alors un milieu continu est caractérisé par sa constante diélectrique ( $\epsilon$ ) ou permittivité.

Différentes méthodes permettant de décrire la cavité ont été proposées. Celle que nous avons utilisée se place dans l'approche PCM [4, 5] (de l'anglais « *Polarizable Continuum Model*, »), dans laquelle la surface de la cavité est partitionnée en un ensemble d'éléments discrets  $T$  (tesselle ou *tesserae* en anglais) centrés sur les atomes. L'énergie libre de solvation est calculée de la manière suivante :

$$G_{\text{solvation}} = G_{\text{électrostatique}} + G_{\text{cavitation}} + G_{\text{dispersion-répulsion}}, \quad (2.1)$$

où l'interaction électrostatique  $G_{\text{électrostatique}}$  est calculée en plaçant une charge partielle polarisable au centre des éléments  $T$ . L'interaction  $G_{\text{cavitation}}$  est calculée le plus souvent en utilisant les rayons de van der Waals des atomes de la molécule avec un facteur constant. L'interaction de dispersion et de répulsion  $G_{\text{dispersion-répulsion}}$  se calcule avec la surface « accessible » au solvant. Pour une application numérique, les surfaces sont calculées par un formalisme de l'équation intégrale (appelées IEF-PCM [6]).

## 4. La DFT Conceptuelle

### 4.1. Fondements de la DFT conceptuelle

La Théorie de la fonctionnelle de la densité conceptuelle (notée *CDFT*<sup>1</sup>) est définie comme la théorie de la réactivité chimique, qui, à partir du point de vue de la densité électronique uniquement, vise à extraire des principes (et concepts) chimiques importants de la *DFT*.

Son fondement remonte à 1978 par Robert G. Parr [7], avec l'identification du multiplicateur de Lagrange  $\mu$  avec le potentiel chimique électronique dans l'équation suivante (équation d'Euler-Lagrange de la DFT) :

$$\mu = v(\vec{r}) + \frac{\delta F_{HK}}{\delta \rho(\vec{r})}. \quad (2.2)$$

---

<sup>1</sup> Ici CDFT correspond à *Conceptual Density Functional Theory*. À noter que l'on peut retrouver dans la littérature CDFT pour *Constrained Density Functional Theory*



Le deuxième théorème d'Hohenberg-Kohn impose que la densité électronique est obtenue, pour un potentiel externe  $v(\vec{r})$  fixé, par minimisation de  $E[\rho]$  sous la contrainte de normalisation de la densité électronique. Le Lagrangien correspondant est :

$$\mathcal{L} = E[\rho] - \mu \left( \int \rho(\vec{r}) d^3r - N \right), \quad (2.3)$$

où  $\mu$  est le multiplicateur de Lagrange associée à la contrainte de normalisation. La condition de stationnarité pour ce Lagrangien s'écrit donc :

$$\frac{\delta \mathcal{L}}{\delta \rho(\vec{r})} = 0 = \frac{\delta E[\rho]}{\delta \rho(\vec{r})} - \mu \frac{\delta}{\delta \rho(\vec{r})} \left[ \int \rho(\vec{r}) d^3r - N \right] = \frac{\delta E[\rho]}{\delta \rho(\vec{r})} - \mu, \quad (2.4)$$

d'où :

$$\mu = \left( \frac{\delta E}{\delta \rho(\vec{r})} \right)_{v(\vec{r})}. \quad (2.5)$$

Considérons maintenant une variation de la densité électronique toujours à  $v(\vec{r})$  fixé. L'énergie étant une fonctionnelle de la seule densité électronique, sa variation s'écrit :

$$\delta E[\rho] = \int \left( \frac{\delta E}{\delta \rho(\vec{r})} \right)_{v(\vec{r})} \delta \rho(\vec{r}) d^3r = \mu \int \delta \rho(\vec{r}) d^3r = \mu \delta N, \quad (2.6)$$

donc :

$$\mu = \left( \frac{\partial E}{\partial N} \right)_{v(\vec{r})}. \quad (2.7)$$

Ce multiplicateur de Lagrange est important, car il peut alors être relié à un concept expérimental : l'électronégativité. En 1961, Iczkowski et Margrave [8] ont montré par approche expérimentale que l'énergie d'un atome peut s'écrire simplement comme un polynôme en  $n$  (avec  $n$  la différence entre le nombre d'électrons et la charge nucléaire  $Z$  :  $n = N - Z$ ) comme suit :

$$E(N) = an^4 + bn^3 + cn^2 + dn, \quad (2.8)$$

en supposant la continuité et la différentiabilité de l'énergie, on peut remarquer que l'électronégativité d'un atome, qui est définie comme la pente à  $n = 0$ , correspond à l'inverse de l'équation (2.7) :

$$\chi = -\left(\frac{\partial E}{\partial N}\right). \quad (2.9)$$

Elle peut également être reliée à la définition de l'électronégativité de Mulliken [9], qui est une troncature à l'ordre 2 du polynôme précédent :

$$\chi = \frac{1}{2}(I + A), \quad (2.10)$$

avec  $A$  l'affinité électronique et  $I$  le potentiel de première ionisation verticaux. En effet, considérons une variation du nombre d'électrons du système à potentiel externe constant (processus vertical). Un développement de Taylor à l'ordre 1 en  $N$  pour l'énergie donne :

$$E(N \pm 1) = E(N) \pm \left(\frac{\partial E}{\partial N}\right)_{v(\vec{r})} = E(N) \pm \mu, \quad (2.11)$$

D'où :

$$\mu = \frac{1}{2}(E(N+1) - E(N-1)) = \frac{1}{2}(E(N+1) - E(N) + E(N) - E(N-1)). \quad (2.12)$$

Ainsi,  $\mu$  est un *descripteur de réactivité*.

De nombreux descripteurs de réactivité et dérivées successives supplémentaires, qui s'inscrivent dans différents ensembles de représentations, sont présentés par la suite.

## 4.2. Ensembles de représentations

La DFT conceptuelle se concentre initialement sur la fonctionnelle de l'énergie dans un ensemble canonique (comme nous l'avons fait pour le potentiel chimique) :  $E[\rho] = E[N, v(\vec{r})]$ . Il existe néanmoins plusieurs ensembles de représentations (équivalents) qui impliquent différentes variables telles que  $\rho$  ou  $\mu$ .

### 4.2.1 L'ensemble canonique

L'introduction du potentiel chimique comme un descripteur de réactivité émerge de la considération de la réponse d'un système lors d'une perturbation par un changement du nombre d'électrons (à  $v(\vec{r})$  fixe). On peut alors identifier la réponse complémentaire : la perturbation lors d'un changement du potentiel externe (à  $N$  fixe). Considérons une variation  $\delta v(\vec{r})$  de ce

potentiel à nombre d'électrons constant, qui correspond donc à une variation  $\delta\hat{H} = \sum_{i=1}^n \delta v(\hat{r}_i)$  de l'hamiltonien du système. La théorie des perturbations de Rayleigh-Schrödinger permet d'écrire que la variation d'énergie est égale au premier ordre à :

$$\delta E = \langle \psi_0 | \delta\hat{H} | \psi_0 \rangle = \int \rho(\vec{r}) \delta v(\vec{r}) d^3r, \quad (2.13)$$

où  $\psi_0$  est la fonction d'onde de l'état électronique non-perturbé et  $\rho(\vec{r})$  la densité électronique associée. Par définition de la dérivée fonctionnelle, on doit aussi avoir :

$$\delta E = \int \left( \frac{\delta E}{\delta v(\vec{r})} \right)_N \delta v(\vec{r}) d^3r. \quad (2.14)$$

On en déduit donc que :

$$\left( \frac{\delta E}{\delta v(\vec{r})} \right)_N = \rho(\vec{r}). \quad (2.15)$$

On a ainsi déterminé les deux premières dérivées de l'énergie en fonction du nombre d'électrons et du potentiel externe, ce qui reflète les transformations qui se produisent lors d'une réaction chimique.

On peut écrire l'expression différentielle de la fonctionnelle de l'énergie  $E[N, v(\vec{r})]$  de l'ensemble canonique comme [10] :

$$dE = \mu dN + \int \rho(\vec{r}) \delta v(\vec{r}) d^3r, \quad (2.16)$$

En exprimant la différentielle par une série de Taylor (ici nous ferons jusqu'à l'ordre 3), une grande quantité de fonctions de réponse en fonction de ces variables apparaît (les dérivées partielles successives, qui sont les descripteurs de réactivité), dont l'équation générale est la suivante :

$$\frac{\partial^n \delta^m E}{\partial N^n \delta v(\vec{r}_1) \delta v(\vec{r}_2) \dots \delta v(\vec{r}_m)}, \quad (2.17)$$

où  $n$  et  $m$  sont les exposants de la série de Taylor permettant l'expression d'un descripteur ( $n = 0$  et  $m = 1$  pour la densité électronique,  $n = 1$  et  $m = 0$  pour le potentiel chimique). D'une manière plus générale, on présente souvent l'ensemble des descripteurs de réactivité d'un ensemble sur le schéma suivant [11, 12] :

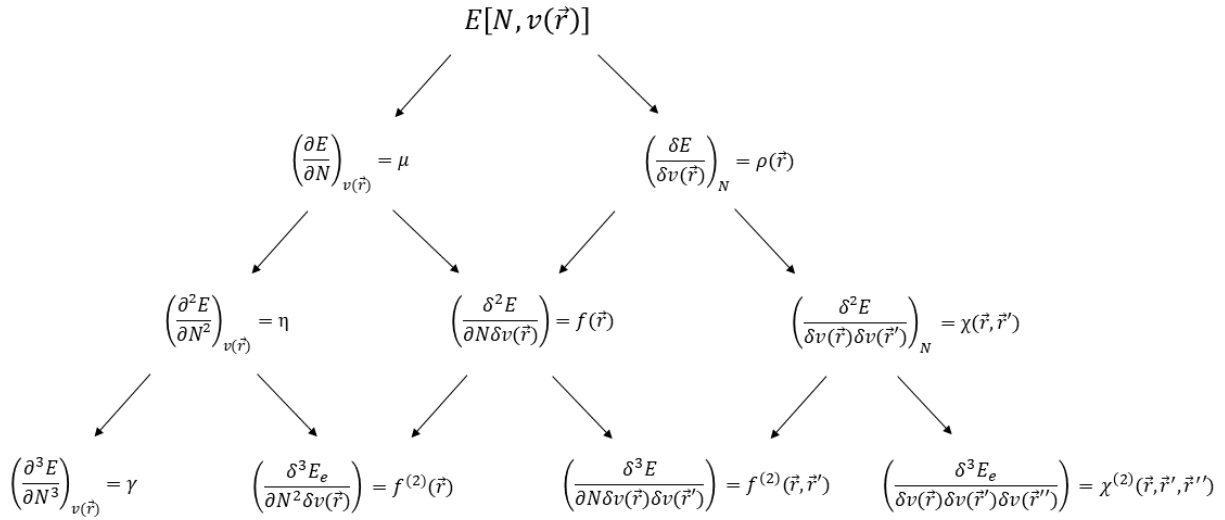


Figure 14.1: Dérivées successives au troisième ordre pour l'ensemble canonique

Il est important de discuter de la dérivabilité de l'énergie  $E$  par rapport au nombre total d'électrons  $N$  qui est une variable supposée discontinue [13]. Il faut être capable de définir proprement l'énergie électronique d'un système avec un nombre fractionnaire d'électrons [14]. Cela peut s'effectuer en considérant un système *ouvert* ou en interpolant l'énergie du système à des valeurs entières, ce qui demeure l'un des principaux problèmes de la théorie conceptuelle de la DFT [15].

De nombreuses formules d'interpolation avec un nombre entier d'électrons ont été considérées, les plus répandues sont des modèles quadratique, exponentiel, rationnel et linéaire [16].

Pour illustrer cela on peut regarder l'interpolation du potentiel chimique  $\mu$ . Il convient de différencier la réponse d'un système à un gain ou une perte d'électron ( $\mu^+$  et  $\mu^-$ ). Cela se traduit par les dérivées à droite et à gauche de l'énergie en fonction du nombre total d'électrons. Ainsi Parr, *et. al.* ont défini [14]:

$$\mu^+ = \left( \frac{\partial E}{\partial N} \right)_{v(\vec{r})}^+ = -A, \quad (2.18)$$

$$\mu^- = \left( \frac{\partial E}{\partial N} \right)_{v(\vec{r})}^- = -I. \quad (2.19)$$

#### 4.2.2. L'ensemble grand canonique

En partant de l'ensemble canonique, les variables de l'état fondamentales indépendantes  $N$  et  $v(\vec{r})$  peuvent être remplacées par leurs variables conjuguées  $\mu$  et  $v(\vec{r})$  par transformée de Legendre [10].

Par analogie avec la thermodynamique, on considère dans cet ensemble que le système est en contact avec un réservoir de  $\mu$  et  $\rho$ , ainsi la notation  $\Omega[\mu, v(\vec{r})]$  et le nom de « grand canonique » sont plus souvent employés. La fonctionnelle « grand potentiel »  $\Omega$  se définit par :

$$\Omega[\mu, v(\vec{r})] \equiv E - \mu N, \quad (2.20)$$

avec comme différentielle :

$$d\Omega[\mu, v(\vec{r})] = -Nd\mu + \int \rho(\vec{r})\delta v(\vec{r})d^3r. \quad (2.21)$$

En effet,

$$d\Omega = dE[N, v(\vec{r})] - Nd\mu - \mu dN = \cancel{\mu dN} + \int \rho(\vec{r})\delta v(\vec{r})d^3r - Nd\mu - \cancel{\mu dN} \quad (2.22)$$

Ainsi dans cet ensemble, on peut reproduire le schéma précédent et définir les descripteurs comme les dérivées successives telles que :

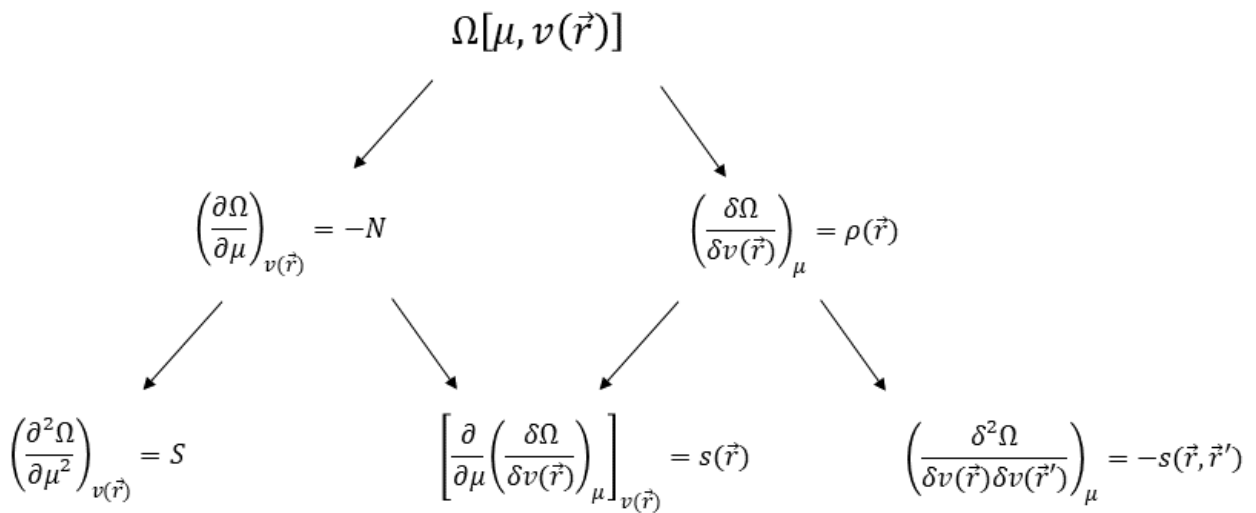


Figure 2.2 : Dérivées successives au second ordre pour l'ensemble grand canonique

La différence importante entre l'ensemble canonique et grand canonique vient du traitement de la variable  $N$ , le nombre d'électrons. Dans l'ensemble canonique,  $N$  est une variable discontinue et le nombre moyen d'électrons du système est constant (hormis lors de réactions de transfert d'électrons), le système étudié est isolé et l'influence du partenaire de la réaction ou de l'environnement n'est pas présente. L'ensemble grand canonique émerge du principe que le nombre d'électrons moyen d'une molécule peut varier en solution, puisque le solvant est perçu dans cet ensemble comme un réservoir d'électrons avec un potentiel chimique constant qui peut interagir avec la molécule.

Ainsi l'ensemble grand canonique est mieux adapté que le canonique lorsque les transferts d'électrons sont importants avec l'environnement, mais son utilisation reste moindre comparée à l'ensemble canonique, étant donné que l'utilisation d'un nombre  $N$  non constant d'une molécule en solution est contraire à l'image attendue par le chimiste.

#### 4.2.3. L'ensemble isomorphique

Les deux ensembles précédents se placent à potentiel externe constant et ne prennent pas en compte les variations possibles du potentiel externe  $v(\vec{r})$ . Ainsi, afin de « définir de nouveaux ensembles et de nouvelles fonctions à utiliser pour des problèmes spécifiques », les ensembles isomorphique et grand isomorphique apparaissent.

Historiquement, les deux ensembles ont été présentés avec les variables naturelles  $N$  et  $\rho(\vec{r})$ , mais en raison de la condition de normalisation, on utilise préférentiellement pour l'ensemble isomorphique la fonction de forme (*shape function*)  $\sigma(\vec{r})$  [17] telle que :

$$\sigma(\vec{r}) = \frac{\rho(\vec{r})}{N}. \quad (2.23)$$

Cette décomposition permet de définir la fonction de forme comme une densité électronique par électron ;  $\sigma(\vec{r})$  est alors toujours positive et suit la condition de normalisation :

$$\int \sigma(\vec{r}) d^3r = 1. \quad (2.24)$$

L'ensemble isomorphique se définit en utilisant la fonctionnelle universelle  $F$  comme fonction d'état avec la fonction de forme et le nombre d'électrons  $N$  comme variables [18] :

$$F[N, \sigma(\vec{r})] \equiv E - N \int v(\vec{r}) \sigma(\vec{r}) d^3r. \quad (2.25)$$

Sa forme différentielle est :

$$dF = \left[ \mu - \int v(\vec{r}) \sigma(\vec{r}) d^3r \right] dN - N \int v(\vec{r}) \delta\sigma(\vec{r}) d^3r. \quad (2.26)$$

Dans cet ensemble, les dérivées successives par rapport à  $N$  et  $\sigma(\vec{r})$  sont alors regroupées dans le schéma suivant :

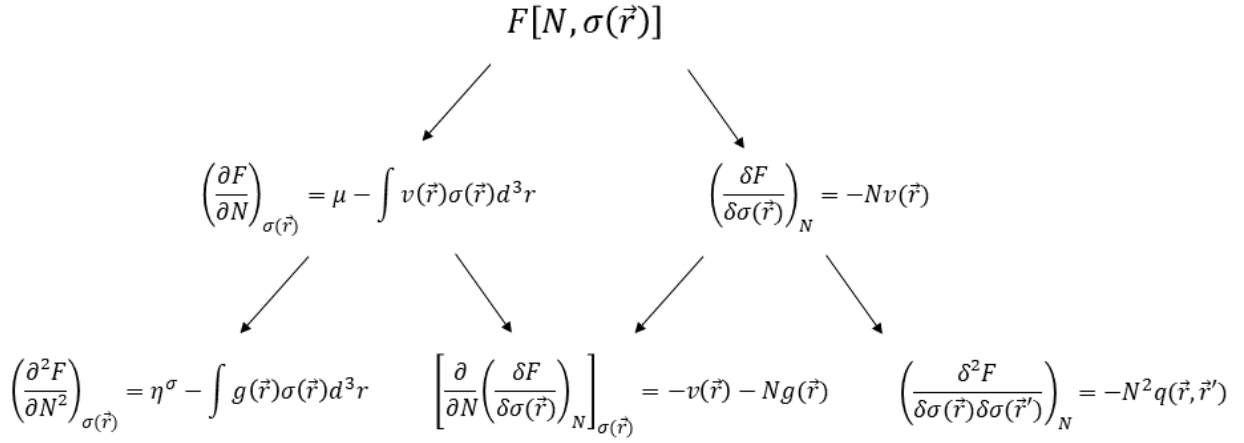


Figure 2.3: Dérivées successives au second ordre pour l'ensemble isomorphique

#### 4.2.4. L'ensemble grand isomorphique

Le dernier ensemble de cette partie est la représentation associée à des systèmes ouverts de l'ensemble précédent ( $N$  n'y est plus une variable naturelle) : l'ensemble grand isomorphique. Il demeure l'un des ensembles les moins étudiés en raison de son interprétation physique difficile [19, 20].

On retrouve, de la même manière que pour l'ensemble isomorphique, deux représentations de cet ensemble dans la littérature : soit avec la fonction de forme, soit avec la densité électronique, la dernière étant celle présentée ici. La fonctionnelle grand isomorphique  $R$  se définit :

$$R[\mu, \rho(\vec{r})] \equiv F - \mu N. \quad (2.27)$$

Avec comme différentielle :

$$dR = -Nd\mu - \int v(\vec{r}) \delta \rho(\vec{r}) d^3r. \quad (2.28)$$

Les descripteurs dans cet ensemble sont représentés dans le schéma suivant :

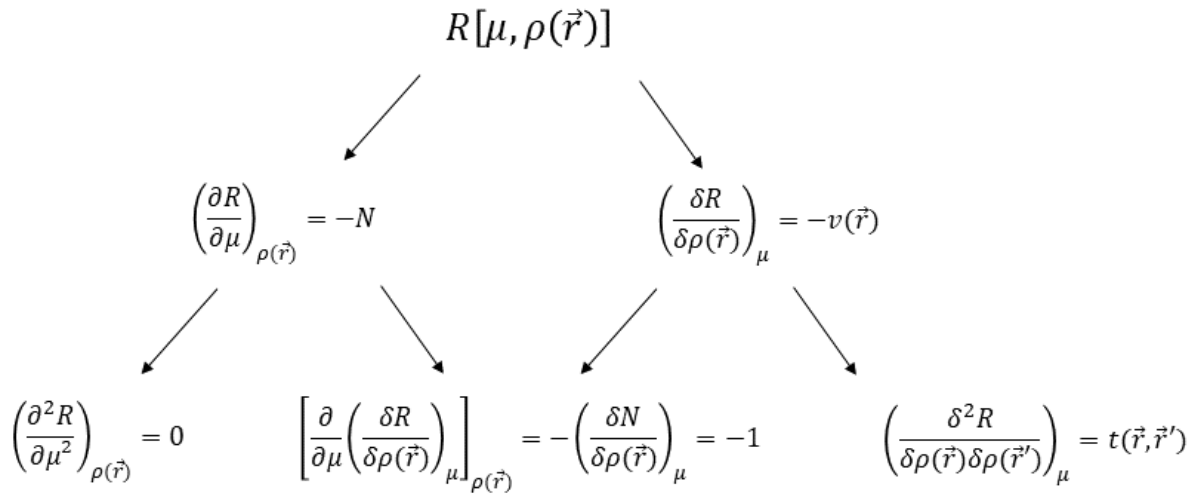


Figure 2.4: Dérivées successives au second ordre pour l'ensemble grand isomorphe

#### 4.2.5. Relations de Maxwell

Similairement à la thermodynamique, il est possible d'obtenir un ensemble de relations de Maxwell à partir des transformées de Legendre des différentes fonctionnelles  $E$ ,  $\Omega$ ,  $F$  et  $R$  présentées précédemment, permettant une connexion entre les différents ensembles.

Ces relations et les ensembles présentés précédemment, peuvent se résumer de manière simpliste dans le diagramme de Parr suivant [10] (analogue au carré de Born) :

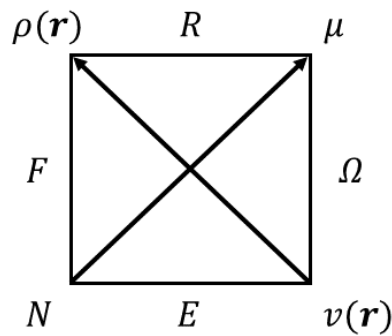


Figure 2.5: Diagramme de Parr illustrant les relations entre les différents ensembles

Dans ce diagramme, chaque segment du carré représente une fonctionnelle et chaque sommet représente une variable (à gauche les variables extensives  $\rho$  et  $N$  et à droite les variables intensives  $\mu$  et  $v$ ). Ainsi, chaque expression différentielle d'une fonctionnelle se retrouve avec les variables naturelles adjacentes, le signe du coefficient est indiqué par les flèches (négative



pour une flèche qui pointe vers une variable naturelle et positive lorsque la flèche part de la variable).

A partir du diagramme on retrouve les différentes relations de Maxwell :

$$\left(\frac{\partial \rho(\vec{r})}{\partial N}\right)_{v(\vec{r})} = \left(\frac{\delta \mu}{\delta v(\vec{r})}\right)_N, \quad (2.29)$$

$$\left(\frac{\partial \rho(\vec{r})}{\partial \mu}\right)_{v(\vec{r})} = -\left(\frac{\delta N}{\delta v(\vec{r})}\right)_\mu, \quad (2.30)$$

$$-\left(\frac{\partial v(\vec{r})}{\partial N}\right)_{\rho(\vec{r})} = \left(\frac{\delta \mu}{\delta \rho(\vec{r})}\right)_N, \quad (2.31)$$

$$\left(\frac{\partial v(\vec{r})}{\partial \mu}\right)_{\rho(\vec{r})} = \left(\frac{\delta N}{\delta \rho(\vec{r})}\right)_\mu. \quad (2.32)$$

Ces relations permettent de relier une réponse à une autre. Par exemple, l'équation (2.29), correspondant à la fonction de Fukui détaillée dans la suite, relie la réponse de la densité locale à un faible changement du nombre d'électrons (à  $v(\vec{r})$  constant) à la réponse du potentiel chimique à un changement localisé du potentiel externe (à  $N$  constant) en un point donné ( $\vec{r}$ ).

Enfin, en complément à ces descripteurs issus directement de dérivées successives (dépendant de l'ensemble considéré), un nombre important de descripteurs de réactivité dans le cadre de la CDFT sont apparus dans les 40 dernières années. Parmi eux, on retrouve les descripteurs *composites* (qui sont une combinaison de deux descripteurs ou plus), ainsi que des descripteurs *ad hoc* (destinés expressément à un usage particulier).

Les parties suivantes ne visent pas à lister exhaustivement l'ensemble des descripteurs mais à présenter une partie des descripteurs importants dans nos travaux. Une liste regroupant un plus grand ensemble de descripteurs est présente en Annexe 1. Numériquement, les descripteurs que nous avons calculés se font à l'aide du logiciel Gaussian 09, en utilisant la fonctionnelle M06-2X et la base Aug-cc-pVTZ (voir Chapitre 4).

### 4.3. Descripteurs globaux

#### 4.3.1. Issus directement d'une dérivée de l'énergie

#### Le potentiel chimique et l'électronégativité

Le potentiel chimique décrit précédemment est un descripteur global de la réactivité d'une molécule, dont l'évaluation permet de décrire la capacité d'une molécule à gagner ou perdre des électrons.

Si l'on considère deux molécules  $A$  et  $B$  en interaction qui échangent des électrons (à nombre total  $N = N_A + N_B$  constant), la variation d'énergie du système  $AB$  au premier ordre s'écrit :

$$\Delta E_{AB} = \Delta E_A + \Delta E_B = \mu_A \Delta N_A + \mu_B \Delta N_B = (\mu_A - \mu_B) \Delta N. \quad (2.33)$$

La différence  $(\mu_A - \mu_B)$  gouverne l'échange d'électrons entre les deux fragments. Ce transfert d'électrons entre l'espèce la moins électronégative (à haut potentiel chimique) vers l'espèce la plus électronégative (à faible potentiel chimique) se traduit par une diminution de l'énergie du système. Le transfert d'électrons s'arrête lorsque  $\mu_A$  et  $\mu_B$  sont « égalisées » ( $\mu_A = \mu_B$ ), ceci est directement lié au principe d'égalisation de Sanderson [21] (voir partie 4.5.1).

Dans le cas d'une réaction de transfert de charge énergétiquement favorable, la variation absolue du potentiel chimique, que l'on note  $|\Delta\mu|$ , aura une valeur élevée. Ceci correspond à l'un des premiers principes de réactivité dans le cadre de la CDFT, il s'agit de la règle «  $|\Delta\mu|$  big is good » (DMB) [22], dont l'énoncé par Parr et Yang en 1989 est le suivant [23] :

*Avec des dispositions généralement similaires pour réagir avec un réactif donné, le réactif préfère celui qui, à l'approche du réactif, est associé à la réponse maximale du potentiel chimique du système*

En raison de la discontinuité du nombre d'électrons discutée précédemment, il est également nécessaire de différencier la réponse d'un système à un gain ou une perte d'électrons ( $\mu^+$  et  $\mu^-$ ), on peut dans ce cadre définir le potentiel chimique comme :

$$\mu = \frac{(\mu^+ + \mu^-)}{2}, \quad (2.34)$$

par le biais d'un modèle à une parabole (avec  $\mu$  et  $\eta$ ). Récemment, Gázquez, Cedillo et Vela ont proposé de passer par une définition à deux paraboles [24] (avec  $\mu^+$  et  $\mu^-$ ,  $\eta^+$  et  $\eta^-$ ), avec une extension de la série de Taylor précédente pour définir deux nouvelles expressions :

$$\Delta E^+ = \mu^+ \Delta N + \frac{1}{2} \eta^+ (\Delta N)^2, \quad (2.35)$$

$$\Delta E^- = \mu^- \Delta N + \frac{1}{2} \eta^- (\Delta N)^2. \quad (2.36)$$

Ainsi, en supposant que  $\eta^+ + \eta^- = \eta$  et  $\eta = \mu^+ - \mu^-$ , on peut définir l'interpolation du potentiel chimique telle que :

$$\mu^+ = -\frac{1}{4}(I + 3A), \quad (2.37)$$

$$\mu^- = -\frac{1}{4}(3I + A). \quad (2.38)$$

Par cette approche, on peut alors différencier partiellement le gain de charge et la perte de charge d'une molécule en termes de potentiel d'ionisation et d'affinité électronique.

### La dureté chimique et la mollesse

Parr et Pearson identifient en 1983 [25] la dérivée seconde de l'énergie par rapport au nombre d'électrons comme la dureté chimique :

$$\eta = \left( \frac{\partial^2 E}{\partial N^2} \right)_{v(r)} = \left( \frac{\partial \mu}{\partial N} \right)_{v(r)}. \quad (2.39)$$

L'application de l'approximation des différences finies à l'équation précédente suit :

$$\eta = I - A. \quad (2.40)$$

Par application du théorème de Koopmans, la dureté correspond approximativement au gap HOMO-LUMO :

$$\eta = \varepsilon_{LUMO} - \varepsilon_{HOMO}. \quad (2.41)$$

La dureté chimique représente la résistance du potentiel chimique au changement du nombre total d'électrons et ainsi la résistance d'un système au transfert de charge. Pour une molécule S, la dureté correspond à l'énergie de la réaction de dismutation suivante :



avec toujours  $I_S > A_S$ . La valeur minimale de la dureté chimique est alors de 0, ce qui correspond à une absence de réaction. La propriété inverse est alors maximisée. Il s'agit de la mollesse chimique S :

$$S = \frac{1}{\eta}, \quad (2.43)$$

qui correspond à la capacité d'une molécule à conserver des électrons. À partir des deux descripteurs  $S$  et  $\eta$ , un grand nombre de descripteurs composites (qui font appel à ces descripteurs) sont apparus, en plus des principes chimiques liés à la réactivité du système basés sur ces descripteurs (HSAB, MHP,...).

### L'hyperdureté

La dérivée d'ordre trois de l'énergie a été définie comme *l'hyperdureté* en 1991 par Parr et Fuentealba [26] telle que :

$$\gamma = \left( \frac{\partial^3 E}{\partial N^3} \right)_{v(\vec{r})} = \left( \frac{\partial \eta}{\partial N} \right)_{v(\vec{r})}. \quad (2.44)$$

L'application du théorème de Koopmans permet d'écrire :

$$\gamma_{FMO} = \varepsilon_{LUMO} - 2\varepsilon_{HOMO} + \varepsilon_{HOMO-1}. \quad (2.45)$$

L'hyperdureté ( $\gamma$ ) mesure la variation de la dureté en fonction de celle du nombre total d'électrons. Elle peut être positive ou négative. En revanche, étant un descripteur du troisième ordre, la valeur de  $\gamma$  est généralement très faible (caractéristique d'une série convergente). Ainsi, le sens chimique de l'hyperdureté reste difficile à évaluer.

Dans un article récent [27], il a été montré qu'une valeur positive de  $\gamma$  semble indiquer une stabilité de la molécule tandis qu'une valeur plutôt négative indique une réactivité importante ; ainsi, pour une série de molécules, l'évaluation de l'hyperdureté permet de les décrire et de les classer par leurs réactivité relatives, mais ne permet pas de déterminer si une molécule est électrophile ou nucléophile.

#### 4.3.2. Descripteurs Ad hoc

### L'indice d'électrophilie et nucléophilie

Dans une étude de docking moléculaire, Maynard *et al.* postulent, en 1998 [28], que la plus forte corrélation entre des descripteurs et la mesure du pouvoir électrophile s'obtient avec la relation :

$$\Delta E_{covalent}^A \approx -\frac{\chi^2}{\eta}. \quad (2.46)$$

En partant de cette définition, Parr *et al.* [29] ont proposé de définir l'indice d'électrophilie  $\omega$  comme :

$$\omega = \frac{\mu^2}{2\eta}. \quad (2.47)$$

Pour le démontrer, Parr *et al.* ont considéré une molécule immergée dans une mer d'électrons avec un potentiel chimique et une température nulle. La molécule électrophile aurait alors une variation de son énergie électronique  $\Delta E$  à potentiel constant lors d'un transfert d'électrons  $\Delta N$  tel que :

$$\Delta E = \mu\Delta N + \frac{1}{2}\eta\Delta N^2. \quad (2.48)$$

Pour le transfert optimal correspondant à  $d\Delta E/d\Delta N = 0$ , le système accepte une quantité optimale d'électrons  $\Delta N_{opt}$  :

$$\Delta N_{opt} = -\frac{\mu}{\eta}, \quad (2.49)$$

associée à une stabilisation énergétique :

$$\Delta E = -\frac{\mu^2}{2\eta}. \quad (2.50)$$

Puisque  $\Delta E < 0$  et  $\eta > 0$ , le processus de transfert de charge est énergétiquement favorable, l'indice d'électrophilie est alors défini comme la stabilisation énergétique lors d'un transfert de charge.

Ce descripteur est ainsi majoritairement utilisé lors d'études comparatives de composés électrophiles et constitue un descripteur important pour nos travaux (voir chapitres 5, 6 et 7).

Il est important de noter que de manière similaire au potentiel chimique, à cause de la discontinuité de l'énergie, différentes définitions pour l'indice d'électrophilie correspondant à un gain d'électrons  $\omega^+$  et à une perte d'électrons  $\omega^-$  sont disponibles dans la littérature [24, 30].

Enfin, pour une définition de l'indice de nucléophilie  $N$  [31], Domingo *et. al.* ont proposé la définition en partant du principe qu'un faible électrophile est un fort nucléophile :

$$N = \frac{1}{\omega}. \quad (2.51)$$

Néanmoins, si une molécule possède plusieurs groupes fonctionnels, la relation inverse n'est pas vérifiée. Ainsi un nombre important de définitions pour les indices de nucléophilie sont présentes dans la littérature, dont une partie est présente en annexes.

### Les indices électrofuge et nucléofuge

Afin de différencier les différents groupes partants lors d'une réaction de substitution ou d'élimination, Ayers *et. al.* ont proposé en 2005 [32] les indices électrofuge et nucléofuge comme descripteurs de réactivité permettant de classer les différents groupes au sein même d'une molécule.

Si l'on considère la réaction de dismutation suivante :



de la molécule  $EN$  avec une charge  $q$  associée, formant un électrofuge  $E$  et un nucléofuge  $N$ , la variation de l'énergie à l'ordre 2 s'écrit :

$$E(q + \Delta q) - E(q) = \Delta q \frac{dE}{dq} + \frac{(\Delta q)^2}{2} \frac{d^2E}{dq^2} = -\mu \Delta q + \frac{\eta}{2} (\Delta q)^2. \quad (2.53)$$

La charge que le nucléofuge peut accepter avec un donneur d'électrons parfait (un nucléophile avec une électronégativité et une dureté nulle) correspond à :

$$\Delta q_{id\acute{e}ale} = \frac{\mu}{\eta} = -\frac{I + A}{2(I - A)}. \quad (2.54)$$

Ainsi, puisque l'énergie d'un donneur d'électrons parfait ne varie pas, on peut définir l'énergie de dissociation nucléofuge  $\Delta E_{\text{nucléofuge}}$  :

$$\Delta E_{\text{nucléofuge}} = \frac{(I - 3A)^2}{8(I - A)}. \quad (2.55)$$

Cette énergie est définie comme la « pénalité énergétique » d'un fragment moléculaire à accepter des électrons d'un nucléophile parfait. De manière analogue, on peut ainsi définir l'énergie électrofuge  $\Delta E_{\text{électrofuge}}$  :

$$\Delta E_{\text{électrofuge}} = \frac{(3I - A)^2}{8(I - A)}. \quad (2.56)$$

Ces énergies permettent ainsi de décrire pour des fragments moléculaires les variations avec des donneurs/accepteurs parfaits d'électrons, mais ne sont pas des mesures directes du comportement électrofuge/nucléofuge d'un fragment moléculaire. En effet, une valeur faible de  $\Delta E_{\text{nucléofuge}}$  et  $\Delta E_{\text{électrofuge}}$  correspond à un fort comportement électrofuge/nucléofuge.

C'est pour cela que l'on utilise plus communément les indices nucléofuge et électrofuge  $\lambda_N$  et  $\lambda_E$  :

$$\lambda_N \equiv e^{-\beta N \Delta E_{\text{nucléofuge}}}, \quad (2.57)$$

$$\lambda_E \equiv e^{-\beta E \Delta E_{\text{électrofuge}}}, \quad (2.58)$$

avec

$$\beta_N = 1,841 \text{ eV et } \beta_E = 0,148 \text{ eV},$$

qui correspondent aux valeurs requises pour que l'indice électrofuge et nucléofuge d'un proton soit de 1. Cette transformation *ad hoc* permet de définir des valeurs positives dans une gamme de données permettant de différencier les groupes partants d'une molécule en comparant avec un hydrogène en utilisant  $\lambda_N$  et  $\lambda_E$ , mais ces dernières ne rentrent dans aucune représentation contrairement à  $\Delta E_{\text{nucléofuge}}$  et  $\Delta E_{\text{électrofuge}}$ .

#### 4.4. Descripteurs locaux et non locaux

##### 4.4.1. Issus directement d'une dérivée de l'énergie

##### Fonction de Fukui

La fonction de Fukui [33] correspond à la dérivée mixte de l'énergie dans l'ensemble canonique par rapport au potentiel externe et au nombre d'électrons, c'est un descripteur local de réactivité puisqu'elle est fonction d'une position de l'espace ( $\vec{r}$ ) :

$$f(\vec{r}) = \left( \frac{\delta^2 E}{\partial N \delta v(\vec{r})} \right). \quad (2.59)$$

Elle se définit également comme la dérivée de la densité électronique par rapport au nombre d'électrons, ou, en utilisant les relations de Maxwell, comme la variation du potentiel chimique par rapport au potentiel externe :

$$f(\vec{r}) = \left( \frac{\partial \rho(\vec{r})}{\partial N} \right)_{v(\vec{r})} = \left( \frac{\delta \mu}{\delta v(\vec{r})} \right)_N. \quad (2.60)$$

L'évaluation de la fonction de Fukui en un point donné ( $\vec{r}$ ) permet ainsi de connaître la réponse d'un système à un changement de nombre d'électrons.

Pour une molécule, en raison de la discontinuité des dérivées par rapport au nombre d'électrons, il faut distinguer la réponse lors d'un gain d'électrons ou d'une perte d'électrons. On définit ainsi deux fonctions de Fukui :

- La fonction de Fukui électrophile  $f^+(\vec{r})$  qui correspond à un gain d'électrons (la variation  $\partial N$  est positive) et qui représente une attaque nucléophile en un point  $\vec{r}$  :

$$f^+(\vec{r}) = \left( \frac{\partial \rho(\vec{r})}{\partial N} \right)_{v(\vec{r})}^+. \quad (2.61)$$

- La fonction de Fukui nucléophile  $f^-(\vec{r})$  qui correspond à une perte d'électrons (la variation  $\partial N$  est négative) et qui représente une attaque électrophile en un point  $\vec{r}$  :

$$f^-(\vec{r}) = \left( \frac{\partial \rho(\vec{r})}{\partial N} \right)_{v(\vec{r})}^-. \quad (2.62)$$

Ces deux fonctions permettent de déterminer les sites les plus aptes à gagner des électrons (sites électrophiles) et ceux qui sont les plus aptes à perdre des électrons (sites nucléophiles).

Pour une évaluation numérique, plusieurs approches sont disponibles dans la littérature [34–37]. Dans l'approche de Kohn-Sham, on a exactement :

$$f^+(\vec{r}) = |\phi_{LUMO}^{KS}(\vec{r})|^2 + \sum_{i=1}^{n_{occ}} \left( \frac{\phi_i^{KS}(\vec{r})}{\partial N} \right)_{v(\vec{r})}^+, \quad (2.63)$$

$$f^-(\vec{r}) = |\phi_{HOMO}^{KS}(\vec{r})|^2 + \sum_{i=1}^{n_{occ}} \left( \frac{\phi_i^{KS}(\vec{r})}{\partial N} \right)_{v(\vec{r})}^-. \quad (2.64)$$

Dans le cadre de l'approximation de Koopmans, on néglige la relaxation des orbitales non frontières :



$$f^+(\vec{r}) \cong |\phi_{LUMO}^{KS}(\vec{r})|^2 = \rho_{LUMO}(\vec{r}), \quad (2.65)$$

$$f^-(\vec{r}) \cong |\phi_{HOMO}^{KS}(\vec{r})|^2 = \rho_{HOMO}(\vec{r}), \quad (2.66)$$

correspondant aux densités électroniques des orbitales LUMO (de plus basse énergie non occupée) et HOMO (de plus haute énergie occupée).

Dans l'approche par différences finies, les fonctions de Fukui sont évaluées selon :

$$f^+(\vec{r}) = \rho_{N+1}(\vec{r}) - \rho_N(\vec{r}), \quad (2.67)$$

$$f^-(\vec{r}) = \rho_N(\vec{r}) - \rho_{N-1}(\vec{r}). \quad (2.68)$$

Ces dernières font appel aux densités électroniques du système à  $N$ ,  $N + 1$  et  $N - 1$  électrons à géométrie fixe.

### Descripteur Dual

Le descripteur dual [38] (ou fonction de Fukui du second ordre) est un descripteur local correspondant à la dérivée mixte du troisième ordre suivante :

$$f^{(2)}(\vec{r}) = \left( \frac{\delta^3 E_e}{\partial N^2 \delta v(\vec{r})} \right). \quad (2.69)$$

Il peut être défini comme la dérivée de la fonction de Fukui en fonction du nombre d'électrons, comme la variation de la dureté par rapport à la variation du potentiel externe ou en tant que la dérivée seconde de la densité électronique :

$$f^{(2)}(\vec{r}) = \left( \frac{\partial f(\vec{r})}{\partial N} \right)_{v(\vec{r})} = \left( \frac{\delta \eta}{\delta v(\vec{r})} \right)_N = \left( \frac{\partial^2 \rho(\vec{r})}{\partial N^2} \right)_{v(\vec{r})}. \quad (2.70)$$

Cette dernière permet par application de l'approximation des différences finies, de calculer le descripteur dual comme la différence en un point ( $\vec{r}$ ) des fonctions de Fukui électrophile et nucléophile :

$$f^{(2)}(\vec{r}) = f^+(\vec{r}) - f^-(\vec{r}). \quad (2.71)$$

Dans le cadre de l'approximation de Koopmans selon :

$$f^{(2)}(\vec{r}) = \rho_{LUMO}(\vec{r}) - \rho_{HOMO}(\vec{r}). \quad (2.72)$$

Si la valeur obtenue en un point ( $\vec{r}$ ) est négative, cela équivaut à  $f^+(\vec{r}) < f^-(\vec{r})$ , indiquant que le site est plus nucléophile qu'électrophile. Inversement, si la valeur est positive, cela équivaut à  $f^+(\vec{r}) > f^-(\vec{r})$ , indiquant que le site sera plus électrophile que nucléophile. Ainsi, dans une molécule, l'évaluation du descripteur dual permet de définir les régions les plus électrophiles et les plus nucléophiles.

L'une des principales applications du descripteur dual est la rationalisation des interactions électrophiles/nucléophiles. On peut citer par exemple les règles de Woodward et Hoffmann [39].

### Réponse linéaire

Le dernier descripteur présenté est un descripteur non-local : il dépend de deux positions de l'espace ( $\vec{r}$ ) et ( $\vec{r}'$ ). La dérivée à l'ordre 2 non locale de l'ensemble canonique correspond à la réponse linéaire [40]:

$$\chi(\vec{r}, \vec{r}') = \left( \frac{\delta^2 E}{\delta v(\vec{r}) \delta v(\vec{r}')} \right)_N, \quad (2.73)$$

que l'on peut définir comme la réponse de la densité électronique en un point ( $\vec{r}$ ) induite par une variation du potentiel externe en un point ( $\vec{r}'$ ) :

$$\chi(\vec{r}, \vec{r}') = \left( \frac{\delta \rho(\vec{r})}{\delta v(\vec{r}')} \right)_N. \quad (2.74)$$

Cette dernière définition permet de relier la réponse linéaire à la polarisabilité. L'évaluation mathématique de ce noyau (*kernel*) reste difficile puisque c'est une fonction dépendant de six variables.

Concernant son interprétation chimique, la fonction de réponse linéaire est reliée à la structure des couches électroniques atomiques, aux effets inductifs et mésomères d'une molécule ainsi qu'à l'aromaticité dans les molécules [41].

#### 4.4.2. Descripteurs composites

### Dureté et mollesse locale

Dans l'ensemble grand canonique, Parr et Yang identifient en 1985 [42] la mollesse locale comme la dérivée de la densité électronique par rapport au potentiel chimique :

$$s(\vec{r}) = \left( \frac{\partial \rho(\vec{r})}{\partial \mu} \right)_{v(\vec{r})}, \quad (2.75)$$

avec comme relation avec la mollesse :

$$S = \int s(\vec{r}) d^3r. \quad (2.76)$$

Ainsi par application de la règle de dérivée des fonctions composées, on peut directement relier la mollesse locale à la fonction de Fukui :

$$s(\vec{r}) = \left( \frac{\partial N}{\partial \mu} \right)_{v(\vec{r})} \left( \frac{\partial \rho(\vec{r})}{\partial N} \right)_{v(\vec{r})} = Sf(\vec{r}). \quad (2.77)$$

La mollesse locale contient la même information que les fonctions de Fukui (c'est-à-dire la sensibilité du potentiel chimique d'un système à une perturbation externe locale) plus une information supplémentaire sur la mollesse moléculaire totale. Par conséquent, la fonction de Fukui ou la mollesse locale peuvent être utilisées pour comparer les différents sites de réactivité au sein d'une même molécule, mais seule  $s(\vec{r})$  permet de comparer différents sites de réactivité entre différentes molécules.

Pour la définition de la dureté locale, plusieurs définitions sont disponibles dans la littérature. On peut citer la définition analytique de Ghosh et Berkowitz [43] qui définissent de manière analogue :

$$\tilde{\eta}(\vec{r}) = \left( \frac{\delta \mu}{\delta \rho(\vec{r})} \right)_{v(\vec{r})}. \quad (2.78)$$

### L'électrophilie locale

Afin de définir de manière identique à la mollesse locale un descripteur capable de comparer les sites électrophiles entre plusieurs molécules et en se basant sur la définition de l'indice d'électrophilie, Chattaraj, Maiki et Sakar proposent en 2003 l'électrophilie locale [44]:

$$\omega(\vec{r}) = \omega f(\vec{r}). \quad (2.79)$$

Ainsi, au sein d'une molécule :

$$\omega = \int \omega(\vec{r}) d^3r. \quad (2.80)$$

Ce descripteur contient ainsi de l'information sur la fonction de Fukui et sur l'indice d'électrophilie globale. Pour une réaction donnée avec un nucléophile quelconque, on peut ainsi comparer les meilleurs sites électrophiles entre deux molécules puisque leurs indices d'électrophilie seront différents.

Une application similaire de cette stratégie avec le descripteur dual, permet d'obtenir le descripteur multiphilique [45] :

$$\omega^{(2)}(\vec{r}) = \omega f^{(2)}(\vec{r}). \quad (2.81)$$

#### 4.5. Principes de réactivité

En complément des descripteurs de réactivité, on peut retrouver dans la CDFT des principes et des concepts de réactivité, qui permettent une interprétation et une compréhension des réactions chimiques.

Le cadre théorique, les définitions mathématiques et une discussion sur la validité de cinq de ces principes sont présentés dans cette partie.

##### 4.5.1. Principe d'égalisation des électronégativités

À partir d'une définition de l'électronégativité dépendante de la longueur d'une liaison chimique, Sanderson énonce à partir de 1951 les bases du principe d'égalisation des électronégativités [21]:

*Lorsque deux ou plusieurs atomes initialement différents en électronégativité se combinent chimiquement, leurs électronégativités deviennent égalisées dans la molécule.*

Ce principe peut se démontrer dans le cadre de la DFT conceptuelle, avec les définitions de l'électronégativité et du potentiel chimique vues précédemment (équations 2.7 et 2.9).

Si l'on considère deux atomes  $A$  et  $B$  avec  $\mu_A^0 < \mu_B^0$ , qui interagissent afin de former une molécule  $AB$ , il y aura un flux d'électrons de  $B$  vers  $A$ , associé à des variations d'énergies :

$$\Delta E_A = E_A^0 + \mu_A^0(N_A - N_A^0), \quad (2.82)$$

$$\Delta E_B = E_B^0 + \mu_B^0(N_B - N_B^0). \quad (2.83)$$

Ainsi, l'énergie totale du système sera :

$$\Delta E_{AB} = \Delta E_A + \Delta E_B = E_A^0 + E_B^0 + (\mu_A^0 - \mu_B^0)\Delta N, \quad (2.84)$$

avec :

$$\Delta N = N_B^0 - N_B = N_A - N_A^0. \quad (2.85)$$

Une valeur positive de  $\Delta N$  correspondra à une stabilisation du système. L'optimisation de l'énergie lors de l'échange d'électrons minimisera alors  $\Delta E_A + \Delta E_B$  et donnera :

$$\mu_A = \mu_B = \mu_{AB}, \quad (2.86)$$

où au second ordre :

$$\mu_A = \left( \frac{\partial E}{\partial N_A} \right)_{v(\vec{r})} = \mu_A^0 + 2\eta_A \Delta N = \mu_A(N_A^0 + \Delta N), \quad (2.87)$$

$$\mu_B = \left( \frac{\partial E}{\partial N_B} \right)_{v(\vec{r})} = \mu_B^0 - 2\eta_B \Delta N = \mu_B(N_B^0 - \Delta N). \quad (2.88)$$

Ainsi :

$$\mu_{AB} = -(\mu_A^0 \mu_B^0)^{1/2}. \quad (2.89)$$

Si l'on exprime en termes d'électronégativités, on retrouve le principe d'égalisation :

$$\chi_{AB} = (\chi_A^0 \chi_B^0)^{1/2}. \quad (2.90)$$

Il est important de noter ici qu'à partir de l'équation (2.85), on peut déterminer directement le transfert de charge entre les deux fragments dans ce principe :

$$\Delta N = \frac{\mu_B^0 - \mu_A^0}{2(\eta_A + \eta_B)}. \quad (2.91)$$

#### 4.5.2. Principe des acides et des bases durs ou mous

Le principe des acides et des bases durs ou mous (HSAB de l'anglais "*Hard and Soft Acids and Bases*") fut introduit en 1963 par Pearson [46] et demeure l'un des principes fondamentaux pour la rationalisation de la réactivité chimique. Le principe s'énonce de la manière suivante :

*Lorsque les « autres effets » tels que la force des acides et des bases sont similaires, les acides durs préfèrent se lier aux bases dures et les acides mous préfèrent se lier aux bases molles.*

Dans l'article originel de Pearson, la dureté chimique n'était pas clairement définie, mais en affirmant la validité du principe HSAB, Pearson a pu établir les principales caractéristiques des espèces *dures* et *molles*. En particulier, il a observé que le site réactif des espèces *dures* a tendance à être petit, fortement chargé et relativement non polarisable, tandis que le site réactif d'un réactif *mou* a tendance à être grand, moins chargé et hautement polarisable. En utilisant ces caractéristiques et les tendances connues de la réactivité chimique, les acides et les bases pourraient être classés comme réactifs durs, mous ou limites.

Ce n'est alors qu'avec l'identification de la dureté en 1983 par Parr et Pearson [25] qu'une échelle quantitative de la dureté pour le principe HSAB a pu être proposée, permettant de donner par le biais de la CDFT le cadre théorique de ce principe. Celui-ci peut se démontrer de la même manière que le principe précédent : à partir de l'équation du transfert de charge (2.91) entre les deux fragments, on peut écrire l'énergie de stabilisation au second ordre :

$$\Delta E_{AB} = -\frac{1}{2}(\eta_A + \eta_B)\Delta N^2 = -\frac{(\mu_B - \mu_A)^2}{4(\eta_A + \eta_B)}, \quad (2.92)$$

que l'on peut réécrire en termes de grand potentiel :

$$\Delta E_{AB} = \Delta\Omega_A + \Delta\Omega_B, \quad (2.93)$$

avec :

$$\Delta\Omega_A = -\frac{(\mu_B - \mu_A)^2}{4} \frac{\eta_A}{(\eta_A + \eta_B)^2}, \quad (2.94)$$

$$\Delta\Omega_B = -\frac{(\mu_B - \mu_A)^2}{4} \frac{\eta_B}{(\eta_A + \eta_B)^2}, \quad (2.95)$$

Ainsi quand  $\mu_B \neq \mu_A$  et que l'on fixe  $\eta_B$ , on peut voir que  $\Delta\Omega_A$  est minimisée par rapport à  $\eta_A$  et la stabilisation est obtenue lorsque  $\eta_A = \eta_B$ , confirmant le principe HSAB. Ainsi, à l'équilibre :

$$\Delta\Omega_A = \Delta\Omega_B. \quad (2.96)$$

Dans le cadre de la CDFT,  $A$  et  $B$  sont des systèmes ouverts. Ces équations démontrent que la quantité thermodynamique naturelle pour un atome dans une molécule n'est pas l'énergie  $E$  mais le grand potentiel  $\Omega$ . À l'équilibre, après égalisation du potentiel chimique, les grands potentiels de  $A$  et  $B$  préfèrent être aussi négatifs que possible.

#### 4.5.3. Principe de dureté maximale

À partir de la théorie des orbitales frontières et du principe HSAB, Pearson propose en 1987 [47] le principe qu'une molécule vise à augmenter sa stabilité lors d'une réaction chimique, en augmentant son gap HOMO-LUMO (c'est-à-dire sa dureté chimique), énonçant le principe de dureté maximale (MHP) :

*Il semble y avoir une règle de la nature selon laquelle les molécules s'organisent de manière à être aussi dures que possible*

On peut démontrer simplement ce principe en étudiant la variation de la dureté  $\Delta\eta$  [48], lors d'une réaction de transfert de charge entre deux molécules  $A$  et  $B$  :



La variation d'énergie  $\Delta E$  lors de cette réaction est définie par :

$$\Delta E = -\frac{1}{2} \frac{(\mu_A - \mu_B)^2}{(\eta_A + \eta_B)}, \quad (2.98)$$

et la dureté de la molécule  $AB$  peut se calculer avec :

$$\eta_{AB} = \frac{\eta_A \eta_B}{\eta_A + \eta_B}. \quad (2.99)$$

Si l'on relie la variation de la dureté à la variation d'énergie, on obtient la relation suivante :

$$\Delta\eta = \eta_{AB} - (\eta_A + \eta_B), \quad (2.100)$$

$$\Delta\eta = -\left(\frac{\eta_A^2 + \eta_B^2 + \eta_A\eta_B}{\eta_A + \eta_B}\right). \quad (2.101)$$

Afin de montrer les variations de la dureté au cours d'une réaction chimique (variations de  $\Delta E$ ), on étudie pour la suite de la démonstration, les variations de  $\eta_A$  avec  $B$  constant (les descripteurs de  $B$  ne varient pas) :

$$\left(\frac{\partial\Delta\eta}{\partial\Delta\eta_A}\right)_B = -\eta_A \left(\frac{\eta_A + 2\eta_B}{\eta_A + \eta_B}\right) < 0, \quad (2.102)$$

$$\left(\frac{\partial\Delta E}{\partial\Delta\eta_A}\right)_{\mu_A,B} = \frac{1}{2} \left(\frac{\mu_A - \mu_B}{\eta_A + \eta_B}\right)^2 > 0. \quad (2.103)$$

Ainsi :

$$\left(\frac{\partial\Delta E}{\partial\Delta\eta}\right)_{\mu_A,B} < 0. \quad (2.104)$$

Une réaction de transfert de charge est ainsi favorisée si l'on modifie un réactif d'une manière à augmenter la dureté totale. Plus  $\Delta\eta$  est élevée, plus la réaction est exothermique. Au cours d'une réaction, une molécule vise ainsi à augmenter ainsi sa dureté chimique.

Ce principe demeure une règle générale, une étude récente sur 101 réactions a permis de montrer que le principe de dureté maximale n'est valide que dans 61% des cas [49]. Cela peut s'expliquer par les différentes approximations effectuées par le principe : on néglige les effets électrostatiques et entropiques et la démonstration s'appuie sur les modèles de perturbations de  $E$  et  $N$  simplifiés.

#### 4.5.4. Principe d'électrophilie minimale

Le principe d'électrophilie minimale ( $MEP^2$ ) est une extension du principe précédent. En 2003, Chamorro, Chattaraj et Fuentealba ont montré que si l'on suit l'évolution de la dureté  $\eta$  lors d'une réaction chimique et qu'elle évolue vers son maximum, l'indice d'électrophilie  $\omega$  (équation 2.47) sera minimisé durant la réaction [50, 51].

---

<sup>2</sup> Ici MEP correspond à *Minimum Electrophilicity Principle*. À noter que l'on peut retrouver dans la littérature MEP pour *Molecular Electrostatic Potential*, MEP pour *Maximum Entropy Principle* ainsi que MEP pour *Minimum Energy Path*.



Cette relation peut se démontrer par l'expression de l'indice d'électrophilie  $\omega$  comme une fonctionnelle dans l'ensemble canonique :

$$\omega \equiv \omega[N, v(\vec{r})], \quad (2.105)$$

avec comme différentielle :

$$d\omega = \left( \frac{\partial \omega}{\partial N} \right)_{v(\vec{r})} dN + \int \left( \frac{\delta \omega}{\delta v(\vec{r})} \right)_N dv(\vec{r}) d^3r, \quad (2.106)$$

où les dérivées peuvent être réécrites avec les expressions des descripteurs usuels ( $\mu, \eta, \gamma, f(\vec{r})$  et  $f^{(2)}(\vec{r})$ ) :

$$\left( \frac{\partial \omega}{\partial N} \right)_{v(\vec{r})} = \frac{\mu}{\eta} - \frac{\mu^2}{2\eta^2} \gamma, \quad (2.107)$$

$$\left( \frac{\delta \omega}{\delta v(\vec{r})} \right)_N = \frac{\mu}{\eta} f(\vec{r}) - \frac{\mu^2}{2\eta^2} f^{(2)}(\vec{r}), \quad (2.108)$$

permettant l'expression :

$$d\omega = \frac{\mu}{\eta} \left[ \eta dN + \int f(\vec{r}) \Delta v(\vec{r}) d^3r \right] - \frac{1}{2} \left( \frac{\mu}{\eta} \right)^2 \left[ \gamma dN + \int f^{(2)}(\vec{r}) \Delta v(\vec{r}) d^3r \right], \quad (2.109)$$

$$d\omega = \frac{\mu}{\eta} d\mu - \frac{1}{2} \left( \frac{\mu}{\eta} \right)^2 d\eta. \quad (2.110)$$

On peut alors voir dans l'équation 2.110 que l'étude des variations de l'indice d'électrophilie passe par l'étude des variations du potentiel chimique et de la dureté. Lors d'une réaction à potentiel chimique constant, les variations de  $\omega$  et de  $\eta$  sont directement reliées, connectant les deux principes. Le principe s'énonce ainsi de la manière suivante :

*La direction naturelle de l'évolution d'un système est vers un état d'électrophilie minimum*

La validité de ce principe semble suivre la validité du MHP, avec une supériorité pour le MEP : sur une étude de 101 réactions, le MEP est valide dans 69% des cas contre 61% pour le MHP, son évaluation apporte ainsi une information importante lors d'une réaction chimique.

Il est important également de noter que l'on peut démontrer que l'indice d'électrophilie suit un principe d'égalisation lors d'une réaction chimique [52]. Si l'on regarde une réaction entre un électrophile et un nucléophile, l'indice d'électrophilie diminuera pour l'électrophile et augmentera pour le nucléophile. L'indice d'électrophilie du produit sera alors égalisé jusqu'à une moyenne des deux indices d'électrophilie.

#### 4.5.5. Principe de polarisabilité minimale

En suivant le principe de dureté maximale et la définition de la dureté, Chattaraj et Sengupta ont proposé en 1996 [53] le principe de polarisabilité minimale (notée *MPP*) .

En partant de la dureté chimique, qui mesure la capacité de résistance de la densité électronique à subir différentes déformations, Politzer [54] (1987) a montré que l'on peut la relier à l'inverse de la polarisabilité, une propriété physique qui est une mesure de l'inverse de la résistance d'un système à subir différentes déformations.

Le principe MPP se démontre à partir de la définition de l'électronégativité  $\chi_Z$  . L'électronégativité d'un atome à  $N$  électrons, avec une charge nucléaire  $Z$ , s'exprime comme  $\chi_Z = -(\partial E / \partial N)_Z$ . De par la dépendance de l'électronégativité à la charge atomique, on peut reformuler cette dérivée en utilisant la charge de l'atome  $Q$  (avec  $Q = Z - N$ ) comme  $\chi_Z = -(\partial E / \partial Q)_Z$ . L'expression de la fonction  $\chi_Z(Q)$  par une série de Taylor autour du point  $Q = 0$  à l'ordre 2 suit :

$$\chi_Z(Q) = \chi_{Q=0} + \left( \frac{\partial \chi}{\partial Q} \right)_{Q=0} . \quad (2.111)$$

Le premier terme  $\chi_{Q=0}$  correspond à l'électronégativité de l'atome neutre, tandis que le second terme correspond au changement de charge sur l'atome, qui est l'inverse de la capacité de charge  $\kappa$  définie par Huheey :

$$\kappa = \frac{1}{\left( \frac{\partial \chi}{\partial Q} \right)} . \quad (2.112)$$

À partir des relations  $\chi = -(\partial E / \partial N)_Z$  et  $Q = Z - N$ , on peut ainsi relier  $\chi$  à la définition de la dureté :

$$\left( \frac{\partial \chi}{\partial Q} \right)_{Q=0} = \left( \frac{\partial^2 E}{\partial N^2} \right)_{Z, Q=0} . \quad (2.113)$$

Si l'on exprime la dureté chimique en termes du potentiel d'ionisation  $I$  et de l'affinité électronique  $A$ , on retrouve alors la définition de la capacité de charge de Huheey :

$$\kappa = \frac{1}{(I - A)} . \quad (2.114)$$

Ainsi, en suivant le principe MHP, on peut définir le principe de polarisabilité minimum (MPP) :

*La direction naturelle de l'évolution d'un système est vers un état de polarisabilité minimum*

La polarisabilité est ainsi minimisée lorsque qu'un système électronique évolue vers un état stable. L'importance de la polarisabilité par rapport à la dureté chimique vient du fait que la polarisabilité peut, elle, être déterminée expérimentalement.

Ce principe défini comme une règle générale est généralement accepté [55, 56] et vise à apporter une information complémentaire lors d'études de réactions chimiques. Une évaluation récente de Blair et Thakkar (2013) de ce principe à partir de 2400 polarisabilités expérimentales et théoriques ont montré une violation du principe pour seulement 3,6% des cas [57].

## **5. Analyse de population**

Les analyses de population sont une méthode de partition de la fonction d'onde ou de la densité électronique qui permettent notamment de définir les charges atomiques. Celles-ci demeurent parmi les premiers descripteurs de la structure électronique qui ne soient pas physiquement observables, si bien qu'il n'existe pas de définition univoque d'un atome dans une molécule ou d'une charge atomique.

Dans la littérature, il existe un nombre important de modèles théoriques permettant d'évaluer des charges atomiques à partir de partitions de la population électronique. Les principales méthodes théoriques d'analyse de populations utilisées seront présentées dans cette partie.

### **5.1. Basées sur les orbitales moléculaires**

#### **5.1.1. Analyse de Population de Mulliken**

C'est en 1955 que Mulliken [58] propose la première analyse de population pour calculer les charges atomiques au sein d'une molécule. Cette analyse est basée sur l'utilisation de la matrice

densité et du recouvrement orbitalaire pour répartir les électrons d'une entité moléculaire de manière fractionnée entre ses différentes parties.

Soit la matrice densité dans la base de fonctions  $\mu$  (avec ses coefficients  $c_{\mu i}$  pour l'orbitale moléculaire  $i$ ) :

$$D_{\mu\nu} = 2 \sum_i c_{\mu i} c_{\nu i}. \quad (2.115)$$

La charge atomique de Mulliken s'exprime :

$$q_{Mulliken}(A) = Z_A - \left[ \sum_{\mu \in A} D_{\mu\mu} S_{\mu\mu} + \frac{1}{2} \sum_{v \neq \mu} D_{\mu\nu} S_{\mu\nu} \right], \quad (2.116)$$

avec  $S$  la matrice de recouvrement des fonctions de base  $\mu$  et  $\nu$ .

En pratique cette analyse dépend fortement des fonctions de base employées. Un autre défaut supplémentaire, souligné par Mulliken, provient de la répartition arbitraire des électrons situés dans les orbitales moléculaires entre deux atomes : une moitié à un atome et une moitié pour le second.

### 5.1.2. Analyse de Population Naturelle

Afin d'améliorer l'analyse de population précédente, Weinhold *et. al.* proposent en 1985 [59] une analyse de population qui ne dépend pas directement des fonctions de base utilisées. La densité électronique y est projetée sur un jeu d'orbitales atomiques *naturelles* afin d'en construire les orbitales atomiques. La construction de ces orbitales naturelles se base sur une approche variationnelle à partir du partitionnement de l'espace de Hilbert de la molécule.

Dans cette analyse de population, c'est la méthode *NBO* (NPA) pour le calcul des charges atomiques et des propriétés des liaisons de Weinhold *et. al.* [60] qui demeure la plus utilisée, elle peut se résumer par le schéma suivant :

$$\chi_\mu \rightarrow NAO \rightarrow NHO \rightarrow NBO \rightarrow NLMO, \quad (2.117)$$

où les orbitales atomiques naturelles (*NAO*) sont construites à partir des fonctions de base initiales  $\chi_\mu$ . Ensuite, les combinaisons linéaires des *NAO* conduisent à la formation des orbitales hybrides naturelles (*NHO*) sur chacun des atomes, dont le recouvrement orbitalaire permet de définir les orbitales naturelles de liaisons (*NBO*). Enfin, par approche

perturbationnelle sur une structure de Lewis idéale, on peut construire les orbitales naturelles localisées (*NLMO*).

La somme des éléments diagonaux de la matrice de densité formée sur les *NAO* (orthonormées  $\theta_i$ ) correspondent aux populations atomiques *NPA* (voir [59]) :

$$N_{NPA} = \langle \theta_i | \hat{F} | \theta_i \rangle. \quad (2.118)$$

## 5.2. Basées sur le potentiel électrostatique

Certaines analyses de population sont ajustées sur le potentiel électrostatique moléculaire (*ESP*) [61]:

$$ESP(\vec{r}) = \sum_A \frac{Z_A}{|\vec{R}_A - \vec{r}|} - \int \frac{\rho(\vec{r}') d^3 r'}{|\vec{r}' - \vec{r}|}, \quad (2.119)$$

où  $\vec{R}_A$  correspond à la position de la charge  $Z_A$ .

Dans nos travaux de thèse, nous avons utilisé la méthode CHELPG proposé par Breneman et Wiberg en 1990 [62] (de l'anglais « *Charges from electrostatic potentials using a grid based method* ») dans laquelle les charges atomiques d'une molécule sont ajustées pour reproduire l'*ESP*.

Dans cette méthode, l'*ESP* est calculé sur chaque point d'une grille. Chaque point est espacé de 3,0 pm à l'intérieur d'un cube. Les dimensions du cube sont choisies de telle sorte que la molécule soit située au centre du cube, avec 28,0 pm d'espace libre entre la molécule et l'extrémité du cube dans les trois dimensions.

L'un des défauts importants de la méthode CHELPG est néanmoins le traitement des systèmes de grande taille, dont l'évaluation de l'*ESP* en un point  $\vec{r}$  peut négliger une partie des atomes qui se trouvent loin du point  $\vec{r}$ . [63]

## 5.3. Basées sur la densité électronique

### 5.3.1. Analyse de population Hirshfeld

Dans l'analyse de population Hirshfeld [64], la densité électronique moléculaire est distribuée entre tous les atomes de la molécule. Les charges calculées utilisent des densités atomiques qui partitionnent la densité électronique moléculaire.

La densité de la *promolécule* est initialement définie comme une somme des densités atomiques à partir de la géométrie de la molécule que l'on étudie :

$$\rho_{promolécule}(\vec{r}) = \sum_A^{N_{atomes}} \rho_A^{atomique}(\vec{r}). \quad (2.120)$$

La densité électronique moléculaire *réelle* en tout point de l'espace est ensuite partitionnée par des poids  $w_A(\vec{r})$  :

$$w_A(\vec{r}) = \frac{\rho_A^{atomique}(\vec{r})}{\rho_{promolécule}(\vec{r})}. \quad (2.121)$$

Les charges atomiques se calculent ensuite selon :

$$q_{Hirshfeld} = Z_A - \int w_A(\vec{r}) \rho(\vec{r}) d^3r. \quad (2.122)$$

Les charges Hirshfeld sont néanmoins très dépendantes du choix de la densité de la promolécule initiale [64]. Usuellement, cette densité est obtenue de manière computationnelle par superposition des densités atomiques des atomes neutres.

Une alternative plus récente proposée par Bultinck *et al.* [65] pour résoudre le choix arbitraire de la densité de la promolécule est de calculer les charges Hirshfeld de manière itérative. La procédure, notée Hirshfeld-I, permet de calculer les populations atomiques à partir de la théorie de l'information et les populations atomiques sont déterminées seulement à partir de la densité électronique moléculaire (pour une description de l'algorithme de calcul, voir [65]).

### 5.3.2. La théorie quantique AIM

La théorie quantique AIM [66, 67] (QTAIM de l'anglais « *Quantum theory of Atoms in Molecules* ») est une analyse topologique du champ de gradients de la densité électronique. Nous allons tout d'abord définir quelques éléments clés (comme la définition d'un point critique, chemins de gradient) nécessaires à la compréhension de cette théorie.

## Caractéristiques topologiques

Dans la théorie quantique AIM, un point critique est un point de l'espace réel (repéré par le vecteur position  $\vec{r}_c$ ) où le gradient de la densité électronique s'annule :

$$\vec{\nabla}\rho(\vec{r}_c) = \vec{0}. \quad (2.123)$$

Ses caractéristiques topologiques peuvent ensuite être obtenues à partir de la matrice Hessienne  $\mathcal{H}_\rho$ , ici en coordonnées cartésiennes, de la densité électronique :

$$\mathcal{H}_\rho = \begin{pmatrix} \frac{\partial^2 \rho(\vec{r}_c)}{\partial x^2} & \frac{\partial^2 \rho(\vec{r}_c)}{\partial x \partial y} & \frac{\partial^2 \rho(\vec{r}_c)}{\partial x \partial z} \\ \frac{\partial^2 \rho(\vec{r}_c)}{\partial y \partial x} & \frac{\partial^2 \rho(\vec{r}_c)}{\partial y^2} & \frac{\partial^2 \rho(\vec{r}_c)}{\partial y \partial z} \\ \frac{\partial^2 \rho(\vec{r}_c)}{\partial z \partial x} & \frac{\partial^2 \rho(\vec{r}_c)}{\partial z \partial y} & \frac{\partial^2 \rho(\vec{r}_c)}{\partial z^2} \end{pmatrix}. \quad (2.124)$$

Pour définir le type de point critique  $\mathcal{H}_\rho$  doit être diagonalisée dans une nouvelle base orthonormée  $(x', y', z')$  selon :

$$\mathcal{H}_\rho = \begin{pmatrix} \frac{\partial^2 \rho(\vec{r}_c)}{\partial x'^2} & 0 & 0 \\ 0 & \frac{\partial^2 \rho(\vec{r}_c)}{\partial y'^2} & 0 \\ 0 & 0 & \frac{\partial^2 \rho(\vec{r}_c)}{\partial z'^2} \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}. \quad (2.125)$$

Cela correspond à une rotation des axes du système permettant d'obtenir les valeurs propres  $\lambda_1, \lambda_2$  et  $\lambda_3$  qui correspondent aux principales courbures.

Les différents types de points critiques sont ainsi définis par la matrice précédente avec leurs *rangs* et *signatures*, où le *rang* désigne le nombre de valeurs propres non nulles de  $\mathcal{H}_\rho$  au point  $\vec{r}$  et la *signature* la somme algébrique des signes des valeurs propres. On peut les classer selon :

Table 2.1 : Classement des différents points critiques

| (Rang, Signature) | Point critique                                  |
|-------------------|-------------------------------------------------|
| (3, -3)           | Attracteur (non) nucléaire (Maximum local)      |
| (3, -1)           | Point critique de liaison (2 Maxima, 1 Minimum) |
| (3, +1)           | Point critique d'anneau (2 Minima, 1 Minimum)   |
| (3, +3)           | Point critique de cage (Minimum local)          |

Par exemple, un point critique de liaison (*BCP*), point critique de type (3,-1) pour une structure à l'équilibre, correspond à un minimum de densité électronique entre deux atomes liés (le long du chemin de liaison) et à un maximum dans le plan perpendiculaire à ce chemin de liaison. D'un point de vue QTAIM, les *BCP* sont des indicateurs de liaisons chimiques (ioniques ou covalentes).

Les caractéristiques de ce BCP renseignent sur la nature d'une liaison. Pour caractériser cette liaison, on calcule généralement la densité électronique et son Laplacien (égal à la somme des valeurs propres de matrice Hessienne) en ce point. Si la densité électronique est élevée et si le Laplacien de la densité est négatif, cela désigne une liaison plutôt covalente. A l'inverse, si la densité électronique en ce point est faible et si son Laplacien est positif, la liaison peut être décrite comme ionique.

Les points critiques d'anneau (*Ring Critical Point* ou *RCP*) sont caractéristiques des structures cycliques, par exemple un cycle aromatique. Dans ce cas, le point critique correspond à un minimum de densité électronique au niveau de la surface du cycle et à un maximum dans la direction perpendiculaire. L'ensemble de ces points critiques constitue le « graphe moléculaire » en QTAIM qui représente la connectivité topologique entre les atomes.

### Chemin de gradient et champ de vecteur gradient

Le gradient de la densité électronique en un point  $\vec{r}$  de l'espace réel en trois dimensions ( $\mathbb{R}^3$ ), est défini par un vecteur  $\vec{\nabla}\rho(\vec{r})$ , qui pointe dans le sens de l'augmentation maximale de  $\rho(\vec{r})$ .

Pour tracer un chemin de gradient à partir d'un point  $\vec{r}$ , on effectue un pas infinitésimal dans la direction  $\vec{\nabla}\rho(\vec{r})$ , puis on recalcule le gradient en ce nouveau point afin d'obtenir une nouvelle



direction. Ainsi, en répétant continuellement ce processus, on construit un chemin de gradient complet.

Comme la densité électronique présente des maxima aux positions des noyaux, chaque chemin de gradient se termine au niveau d'un noyau atomique appelé attracteur nucléaire (dénotés *NA* pour *Nuclear Attractor*). Le champ de vecteur gradient de la densité électronique est ainsi constitué de l'ensemble des chemins de gradient. A noter que l'on peut parfois avoir des attracteurs en dehors des positions des noyaux. On parlera alors de *NNA* pour *Non Nuclear Attractors*.

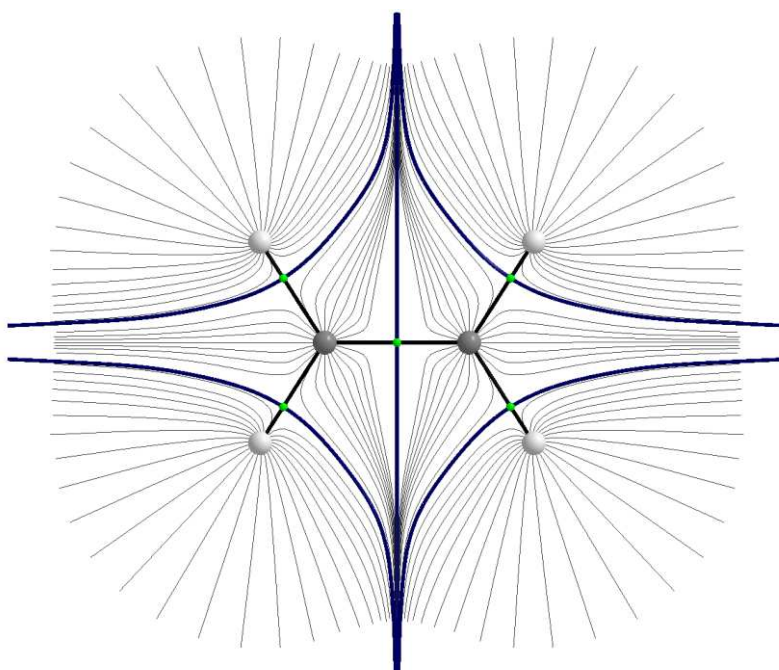


Figure 2.6 : Représentation 2D de la topologie du champ de gradient de la densité électronique pour la molécule d'éthylène (C<sub>2</sub>H<sub>4</sub>). (image AIMAll) (les lignes grises correspondent aux chemins de gradient, celles en bleu aux surfaces interatomiques et celles en noir pour les chemins de liaisons entre atomes)

### Définition d'un atome

La région de l'espace formée par l'ensemble des chemins de gradient qui se terminent à un attracteur donné définit un bassin atomique  $\Omega$ . Un atome se définit simplement comme l'union d'un attracteur et de son bassin associé.

Certains chemins de vecteurs gradient n'aboutissent pas à un attracteur mais à un *BCP*. Ils définissent les surfaces interatomiques (*IAS*) qui séparent les bassins d'atomes voisins. Chaque point de ces surfaces remplit la condition de flux nul suivante :

$$\vec{\nabla}\rho(\vec{r}) \cdot \vec{n}(\vec{r}) = 0, \quad (2.126)$$

où  $\vec{n}(\vec{r})$  est un vecteur normal à la surface interatomique.

L'espace réel est ainsi partitionné en bassins atomiques (qui ne se recouvrent pas). Ils représentent des systèmes ouverts, en pratique, il peut être usuel de définir une limite à ces volumes (en principe d'extension infinie) par le choix d'une isosurface de densité électronique adéquate. Généralement une valeur de 0,001 u.a. est considérée comme suffisante pour recouvrir l'essentiel de la densité électronique. Pour illustrer cela, un exemple de la théorie quantique AIM partitionnant la molécule d'éthylène ( $C_2H_4$ ) avec une valeur d'isodensité égale à 0,001 u.a. peut être vu sur la figure suivante :

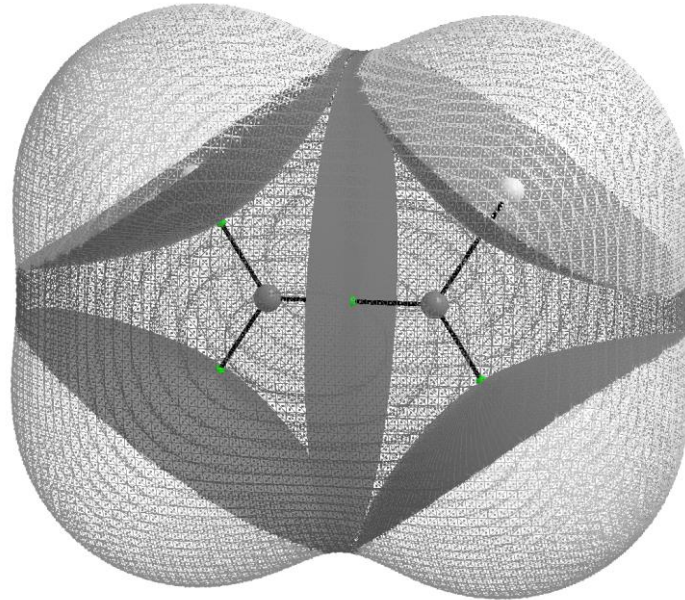


Figure 2.7: Représentation 3D pour la molécule d'éthylène ( $C_2H_4$ ). (image AIMAll) (En gris les bassins atomiques des atomes de carbone et en blanc les bassins atomiques des atomes d'hydrogène)

### Les propriétés atomiques

A partir de la définition de l'atome, il est possible d'intégrer n'importe quelle fonction dépendant de  $\vec{r}$  dans le volume d'un bassin atomique. Ceci permet d'obtenir ainsi une grandeur atomique  $G(A)$  :

$$G(A) = \int_{\Omega_A} g(\vec{r}) d^3r. \quad (2.127)$$

Par exemple, la population électronique d'un atome  $N_{\Omega_A}$  correspond à l'intégration de la densité électronique sur le bassin :

$$N_{\Omega_A} = \int_{\Omega_A} \rho(\vec{r}) d^3r. \quad (2.128)$$

Ainsi, la charge atomique QTAIM pour un atome  $A$  s'obtient selon :

$$q_{QTAIM}(\Omega_A) = Z_A - N_{\Omega_A}. \quad (2.129)$$

Pour un descripteur local de la CDFT, des valeurs atomiques (dites « condensées ») pourront ainsi être obtenues par intégration sur les volumes des bassins atomiques à partir de l'équation (2.126). Par exemple, pour le descripteur dual atomique :

$$f^{(2)}(A) = \int_{\Omega_A} f^{(2)}(\vec{r}) d^3r. \quad (2.130)$$

Ainsi le descripteur dual condensé peut permettre de caractériser les sites atomiques nucléophiles/électrophiles au sein d'une molécule, ce qui est conceptuellement plus parlant pour les chimistes organiciens. On peut noter que toutes les propriétés atomiques sont additives. Cela permet de définir des propriétés pour certains groupes fonctionnels.

### L'indice de localisation et l'indice de délocalisation électronique

Les propriétés électroniques peuvent être étudiées grâce au formalisme des matrices de densité réduite (*RDM*) [68]. Conformément à la convention de McWeeny [69], les RDM du premier et du deuxième ordre ( $\rho_1$  et  $\rho_2$ ) sont respectivement définies comme :

$$\rho_1(\vec{r}_1; \vec{r}_1') = N \int \Psi_{el}(\vec{x}_1, \dots, \vec{x}_N) \Psi_{el}^*(\vec{x}_1', \dots, \vec{x}_N') ds_1 ds_1' d^4\vec{x}_2 \dots d^4\vec{x}_N, \quad (2.131)$$

$$\rho_2(\vec{r}_1, \vec{r}_2; \vec{r}_1', \vec{r}_2') = N(N-1) \int \Psi_{el}(\vec{x}_1, \dots, \vec{x}_N) \Psi_{el}^*(\vec{x}_1', \dots, \vec{x}_N') ds_1 ds_2 ds_1' ds_2' d^4\vec{x}_3 \dots d^4\vec{x}_N, \quad (2.132)$$

dans lesquelles les vecteurs positions  $\vec{x}_i$  comprennent la position des électrons  $\vec{r}$  ainsi que le spin des électrons  $\sigma_i$  ( $\alpha$  ou  $\beta$ ) :  $\vec{x}_i = (\vec{r}_i, \sigma_i)$ .

A partir de la matrice de densité du second ordre, il est possible d'obtenir sa partie diagonale afin de définir la fonction de pair du second ordre  $\rho_2(\vec{r}_1, \vec{r}_2)$ , qui correspond à la densité de probabilité de trouver une paire d'électrons aux positions  $d^3\vec{r}_1$  et  $d^3\vec{r}_2$ . Cette grandeur a une partition naturelle en trois termes : Coulombienne  $\rho_1(\vec{r}_1)\rho_1(\vec{r}_2)$ , d'échange et de corrélation. Les deux derniers termes sont généralement regroupés dans la densité d'échange-corrélation

$\rho_{xc}(\vec{r}_1, \vec{r}_2)$ , qui est égale à  $\rho(\vec{r}_1)\rho(\vec{r}_2) - \rho_2(\vec{r}_1, \vec{r}_2)$ . Si la densité d'échange-corrélation est moyennée sur une paire de bassins, on obtient une mesure du nombre direct de paires d'électrons qui sont partagées entre les atomes quantiques A et B.

A partir de cette définition, on peut définir l'indice de localisation électronique  $LI$  [70] qui correspond à un nombre moyen de paires d'électrons situées sur l'atome A :

$$LI(A, A) = \int_{\Omega_A} d^3\vec{r}_1 \int_{\Omega_A} d^3\vec{r}_2 \rho_{xc}(\vec{r}_1, \vec{r}_2), \quad (2.133)$$

l'indice de délocalisation électronique  $DI$  correspond quant à lui au nombre de paires d'électrons délocalisées entre deux atomes A et B :

$$DI(A, B) = 2 \int_{\Omega_A} d^3\vec{r}_1 \int_{\Omega_B} d^3\vec{r}_2 \rho_{xc}(\vec{r}_1, \vec{r}_2). \quad (2.134)$$

La somme des  $LI$  et  $DI$  sur l'ensemble d'une molécule permet de retrouver le nombre d'électrons :

$$N = \sum_A LI(A, A) + \frac{1}{2} \sum_{A \neq B} DI(A, B). \quad (2.135)$$

En QTAIM, le  $DI$  correspond à un indice de liaison topologique.

### La décomposition énergétique IQA

L'interaction des atomes quantiques (IQA de l'anglais « *Interacting Quantum Atoms* ») est une méthode de décomposition énergétique basée sur une partition AIM de l'espace. Introduit initialement par Blanco *et. al.* en 2005 [71], IQA décompose, en principe exactement, l'énergie moléculaire totale en une somme de termes énergétiques pertinents chimiquement sans avoir recours à une quelconque référence arbitraire. On peut les diviser en deux catégories : une somme de termes intra-atomiques et une somme de termes inter-atomiques. L'énergie totale d'un système moléculaire s'écrit comme :

$$E_{totale} = \sum_A E_{intra}(A) + \frac{1}{2} \sum_A \sum_{A \neq B} E_{inter}(A, B), \quad (2.136)$$

où A et B correspondent aux atomes A et B.

Les parties suivantes visent à présenter les termes énergétiques principaux (pour une description complète, voir [71, 72]).

➤ L'énergie intra atomique IQA

Chaque terme énergétique  $E_{intra}(A)$  peut se décomposer selon :

$$E_{intra}(A) = T(A) + V_{en}(A) + V_{ee}(A, A), \quad (2.137)$$

avec  $T(A)$  l'énergie cinétique pour le bassin atomique :

$$T(A) = -\frac{1}{2} \int_{\Omega_A} \nabla_{\vec{r}_1}^2 \rho_1(\vec{r}_1; \vec{r}'_1) |_{\vec{r}'_1 = \vec{r}_1} d^3 \vec{r}_1, \quad (2.138)$$

$V_{en}(A)$  l'interaction électrostatique électron-noyau dans le bassin :

$$V_{en}(A) = - \int_{\Omega_A} \frac{Z_A \rho(\vec{r})}{\|\vec{r} - \vec{r}_A\|} d^3 \vec{r}, \quad (2.139)$$

$V_{ee}(A, A)$  l'interaction électrostatique électron-électron fut introduite par Pendás *et al.* en 2004 [73] se décompose selon :

$$V_{ee}(A, A) = V_{ee,cl}(A, A) + V_{ee,xC}(A, A), \quad (2.140)$$

où  $V_{ee,cl}(A, A)$  représente la partie coulombienne de la répulsion inter-électronique :

$$V_{ee,cl}(A, A) = \frac{1}{2} \int_{\Omega_A} d^3 \vec{r}_1 \int_{\Omega_A} r_{12}^{-1} \rho(\vec{r}_1) \rho(\vec{r}_2) d^3 \vec{r}_2, \quad (2.141)$$

et  $V_{ee,xC}(A, A)$  représente la contribution d'échange-corrélation à cette même répulsion bi-électronique. Dans le cas d'une fonction d'onde mono-déterminante, elle s'exprime par :

$$V_{ee,xC}(A, A) = - \sum_{ij} \int_{\Omega_A} d^3 \vec{r}_1 \int_{\Omega_A} r_{12}^{-1} f_{ij}(\vec{r}_1) f_{ij}(\vec{r}_2) d^3 \vec{r}_2, \quad (2.142)$$

avec  $r_{12}$  qui correspond à la distance inter-électronique et avec  $f_{ij}(\vec{r}) = \varphi_i(\vec{r})\varphi_j(\vec{r})$  où  $\varphi$  correspond à une orbitale moléculaire.

### ➤ L'énergie inter-atomique IQA

L'énergie d'interaction inter-atomique  $E_{inter}(A, B)$  se définit comme :

$$E_{inter}(A, B) = V_{nn}(A, B) + V_{en}(A, B) + V_{ne}(A, B) + V_{ee}(A, B), \quad (2.143)$$

avec  $V_{nn}(A, B)$  qui correspond à l'énergie de répulsion noyau-noyau selon :

$$V_{nn}(A, B) = \frac{Z_A Z_B}{\|\vec{R}_A - \vec{R}_B\|}, \quad (2.144)$$

$V_{en}(A, B)$  qui correspond à l'énergie d'interaction des électrons de A avec le noyau de B et  $V_{ne}(A, B)$  l'énergie d'interaction du noyau de A avec les électrons de B. Par exemple, l'expression de  $V_{en}(A, B)$  est :

$$V_{en}(A, B) = - \int_{\Omega_A} \frac{Z_B \rho(\vec{r})}{\|\vec{r} - \vec{r}_B\|} d^3\vec{r}, \quad (2.145)$$

$V_{ee}(A, B)$  correspond à l'interaction électrostatique électron-électron selon :

$$V_{ee}(A, B) = V_{ee,Cl}(A, B) + V_{ee,XC}(A, B). \quad (2.146)$$

L'énergie électrostatique  $V_{ee,Cl}(A, B)$  entre A et B s'écrit comme :

$$V_{ee,Cl}(A, B) = \int_{\Omega_A} d^3\vec{r}_1 \int_{\Omega_B} r_{12}^{-1} \rho(\vec{r}_1) \rho(\vec{r}_2) d^3\vec{r}_2, \quad (2.147)$$

et l'énergie d'échange corrélation  $V_{ee,XC}(A, B)$  s'écrit dans le cas d'un fonction d'onde mono déterminantale:

$$V_{ee,XC}(A, B) = -2 \sum_{ij} \int_{\Omega_A} d^3\vec{r}_1 \int_{\Omega_B} r_{12}^{-1} f_{ij}(\vec{r}_1) f_{ij}(\vec{r}_2) d^3\vec{r}_2. \quad (2.148)$$

On peut donc écrire :

$$E_{inter}(A, B) = V_{Cl}(A, B) + V_{ee,XC}(A, B), \quad (2.149)$$

où  $V_{cl}(A, B)$  regroupe l'ensemble des termes électrostatiques :

$$V_{cl}(A, B) = V_{en}(A, B) + V_{ne}(A, B) + V_{ee,cl}(A, B), \quad (2.150)$$

et  $V_{ee,xc}(A, B)$  représente donc la « covalence » entre les atomes  $A$  et  $B$ .

Parmi les applications de la décomposition énergétique, on peut citer notamment les réactions impliquant une formation de liaisons intra et/ou intermoléculaires [74, 75], dont le calcul des énergies d'interactions permet une interprétation chimique pertinente. Elle a également été utilisée avec succès comme un lien direct avec la théorie de la liaison chimique [76–78].

Pour plus de détails sur l'utilisation d'IQA dans un cadre DFT, voir [79].

- [1] A. Lavoisier. *Traité élémentaire de chimie éd. Cuchet 1*, (1793), 107
- [2] E. G. Lewars. *Computational Chemistry. Introduction to the Theory and Applications of Molecular and Quantum Mechanics*. Springer International Publishing (2011)
- [3] M. L. Connolly. *Science*. 221 (1983), 709–713
- [4] S. Miertuš, E. Scrocco, J. Tomasi. *Chem. Phys.* 55 (1981), 117–129
- [5] J. Tomasi, R. Bonaccorsi, R. Cammi, F. J. O. del Valle. *J. Mol. Struct. THEOCHEM* 234 (1991), 401–424
- [6] J. Tomasi, B. Mennucci, R. Cammi. *Chem. Rev.* 105 (2005), 2999–3093
- [7] R. G. Parr, R. A. Donnelly, M. Levy, W. E. Palke. *J. Chem. Phys.* 68 (1978), 3801–3807
- [8] R. P. Iczkowski, J. L. Margrave. *J. Am. Chem. Soc.* 83 (1961), 3547–3551
- [9] R. S. Mulliken. *J. Chem. Phys.* 2 (1934), 782–793
- [10] R. F. Nalewajski, R. G. Parr. *J. Chem. Phys.* 77 (1982), 399–407
- [11] H. Chermette. *J. Comput. Chem.* 20 (1999), 129–154
- [12] P. Geerlings, F. De Proft, W. Langenaeker. *Chem. Rev.* 103 (2003), 1793–1873
- [13] R. A. Miranda-Quintana, P. W. Ayers. *Phys. Chem. Chem. Phys.* 18 (2016), 15070–15080
- [14] J. P. Perdew, R. G. Parr, M. Levy, J. L. J. Balduz. *Phys. Rev. Lett.* 49 (1982), 1691
- [15] P. Geerlings, E. Chamorro, P. K. Chattaraj, et al. *Theor. Chem. Acc.* 139 (2020), 1–18
- [16] F. Heidar-Zadeh, R. A. Miranda-Quintana, T. Verstraelen, et al. *J. Chem. Theory Comput.* 12 (2016), 5777–5787
- [17] A. Cedillo. *Int. J. Quantum Chem.* 52 (1994), 231–240
- [18] F. De Proft, S. Liu, R. G. Parr. *J. Chem. Phys.* 107 (1997), 3000–3006
- [19] S. Liu, R. G. Parr. *J. Chem. Phys.* 106 (1997), 5578–5586
- [20] P. Geerlings, F. De Proft, S. Fias. *Acta Phys. - Chim. Sin.* 34 (2018), 699–707
- [21] R. T. Sanderson. *Science*. 114 (1951), 670–672
- [22] R. A. Miranda-Quintana, P. W. Ayers. *Theor. Chem. Acc.* 138 (2019), 1–6
- [23] R. G. Parr, W. Yang, *Density Functional Theory of Atoms and Molecules* Oxford University Press, New York, (1989).
- [24] J. L. Gázquez, A. Cedillo, A. Vela. *J. Phys. Chem. A* 111 (2007), 1966–1970
- [25] R. G. Parr, R. G. Pearson. *J. Am. Chem. Soc.* 105 (1983), 7512–7516
- [26] P. Fuentealba, R. G. Parr. *J. Chem. Phys.* 94 (1991), 5559–5564
- [27] C. Morell, A. Grand, A. Toro-Labbé, H. Chermette. *J. Mol. Model.* 19 (2013), 2893–2900
- [28] A. T. Maynard, M. Huang, W. G. Rice, D. G. Covell. *Proc. Natl. Acad. Sci. U. S. A.* 95 (1998), 11578–11583



- [29] R. G. Parr, L. V. Szentpály, S. Liu. *J. Am. Chem. Soc.* 121 (**1999**), 1922–1924
- [30] E. Chamorro, M. Duque-Noreña, P. Pérez. *J. Mol. Struct. THEOCHEM* 901 (**2009**), 145–152
- [31] L. R. Domingo, P. Pérez. *Org. Biomol. Chem.* 9 (**2011**), 7168–7175
- [32] P. W. Ayers, J. S. M. Anderson, J. I. Rodriguez, Z. Jawed. *Phys. Chem. Chem. Phys.* 7 (**2005**), 1918–1925
- [33] R. G. Parr, W. Yang. *J. Am. Chem. Soc.* 106 (**1984**), 4049–4050
- [34] R. R. Contreras, P. Fuentealba, M. Galván, P. Pérez. *Chem. Phys. Lett.* 304 (**1999**), 405–413
- [35] W. Yang, W. J. Mortier. *J. Am. Chem. Soc.* 108 (**1986**), 5708–5711
- [36] P. Bultinck, D. Clarisse, P. W. Ayers, R. Carbo-Dorca. *Phys. Chem. Chem. Phys.* 13 (**2011**), 6110–6115
- [37] J. Melin, P. W. Ayers, J. V. Ortiz. *J. Phys. Chem. A* 111 (**2007**), 10017–10019
- [38] C. Morell, A. Grand, A. Toro-Labbé. *J. Phys. Chem. A* 109 (**2005**), 205–212
- [39] P. Geerlings, P. W. Ayers, A. Toro-Labbé, et al. *Acc. Chem. Res.* 45 (**2012**), 683–695
- [40] P. Geerlings, S. Fias, Z. Boisdenghien, F. De Proft. *Chem. Soc. Rev.* 43 (**2014**), 4989–5008
- [41] P. Geerlings, S. Fias, T. Stuyver, et al. *Density Funct. Theory* (**2019**), 3–30
- [42] W. Yang, R. G. Parr. *Proc. Natl. Acad. Sci. U. S. A.* 82 (**1985**), 6723–6726
- [43] M. Berkowitz, R. G. Parr, S. K. Ghosh. *J. Am. Chem. Soc.* 107 (**1985**), 6811–6814
- [44] P. K. Chattaraj, B. Maiti, U. Sarkar. *J. Phys. Chem. A* 107 (**2003**), 4973–4975
- [45] J. Padmanabhan, R. Parthasarathi, M. Elango, et al. *J. Phys. Chem. A* 111 (**2007**), 9130–9138
- [46] R. G. Pearson. *J. Am. Chem. Soc.* 85 (**1963**), 3533–3539
- [47] R. G. Pearson. *J. Chem. Educ.* 64 (**1987**), 561–567
- [48] R. G. Parr, P. K. Chattaraj. *J. Am. Chem. Soc.* 113 (**1991**), 1854–1855
- [49] R. Saha, S. Pan, P. K. Chattaraj. *Molecules* 21 (**2016**), 1–16
- [50] E. Chamorro, P. K. Chattaraj, P. Fuentealba. *J. Phys. Chem. A* 107 (**2003**), 7068–7072
- [51] R. A. Miranda-Quintana, P. W. Ayers. *J. Chem. Phys.* 148 (**2018**), 2017–2019
- [52] P. K. Chattaraj, S. Giri, S. Duley. *J. Phys. Chem. Lett.* 1 (**2010**), 1064–1067
- [53] P. K. Chattaraj, S. Sengupta. *J. Phys. Chem.* 100 (**1996**), 16126–16130
- [54] P. Politzer. *J. Chem. Phys.* 86 (**1986**), 1072–1073
- [55] P. W. Ayers. *Faraday Discuss.* 135 (**2007**), 161–190
- [56] Y. Simón-Manso, P. Fuentealba. *J. Phys. Chem. A* 102 (**1998**), 2029–2032

- [57] S. A. Blair, A. J. Thakkar. *Chem. Phys. Lett.* 556 (2013), 346–349
- [58] R. S. Mulliken. *J. Chem. Phys.* 23 (1955), 1833–1840
- [59] A. E. Reed, R. B. Weinstock, F. Weinhold. *J. Chem. Phys.* 83 (1985), 735–746
- [60] F. Weinhold, C. R. Landis. *Chem. Educ. Res. Pr.* 2 (2001), 91–104
- [61] R. H. Petrucci, W. S. Harwood, F. G. Herring, M. J. D. *General Chemistry: Principles & Modern Applications*, 9th Ed. Pearson Education, New Jersey (2011)
- [62] C. M. Breneman, K. B. Wiberg. *J. Comput. Chem.* 11 (1990), 361–373
- [63] E. Sigfridsson, U. Ryde. *J. Comput. Chem.* 19 (1998), 377–395
- [64] F. L. Hirshfeld. *Theor. Chim. Acta* 44 (1977), 129–138
- [65] P. Bultinck, C. Van Alsenoy, P. W. Ayers, R. Carbó-Dorca. *J. Chem. Phys.* 126 (2007)
- [66] R. F. W. Bader. *Chem. Rev.* 91 (1991), 893–928
- [67] P. Popelier. *Atoms in Molecules: An Introduction* Prentice Hall, London, (2000)
- [68] C. Outeiral, M. A. Vincent, Á. Martín Pendás, P. L. A. Popelier. *Chem. Sci.* 9 (2018), 5517–5529
- [69] R. McWeeny, B.T. Sutcliffe, *Methods of molecular quantum mechanics*, Academic press, London (1969).
- [70] X. Fradera, M. A. Austen, R. F. W. Bader. *J. Phys. Chem. A* 103 (1999), 304–314
- [71] M. A. Blanco, A. M. Pendás, E. Francisco. *J. Chem. Theory Comput.* 1 (2005), 1096–1109
- [72] E. Francisco, A. M. Pendás, M. A. Blanco. *J. Chem. Theory Comput.* 2 (2006), 90–102
- [73] A. Martín Pendás, M. A. Blanco, E. Francisco. *J. Chem. Phys.* 120 (2004), 4581–4592
- [74] M. Yahia-Ouahmed, V. Tognetti, L. Joubert. *Theor. Chem. Acc.* 135 (2016), 1–15
- [75] M. Yahia-Ouahmed, V. Tognetti, L. Joubert. *Comput. Theor. Chem.* 1053 (2015), 254–262
- [76] A. Martín Pendás, E. Francisco, M. A. Blanco, C. Gatti. *Chem. - A Eur. J.* 13 (2007), 9362–9371
- [77] A. Martín Pendás, E. Francisco, M. A. Blanco. *Faraday Discuss.* 135 (2007), 423–438
- [78] A. M. Pendás, E. Francisco, M. A. Blanco. *Phys. Chem. Chem. Phys.* 9 (2007), 1087–1092
- [79] V. Tognetti, L. Joubert. *Phys. Chem. Chem. Phys.* 16 (2014), 14539–14550



## CHAPITRE 3 : Les échelles de réactivité expérimentale

|                                          |     |
|------------------------------------------|-----|
| 1. Introduction.....                     | 93  |
| 2. L'échelle de réactivité de Mayr ..... | 95  |
| 2.1. Présentation générale .....         | 95  |
| 2.2. Applications.....                   | 100 |
| 2.3. Limites et extension.....           | 101 |



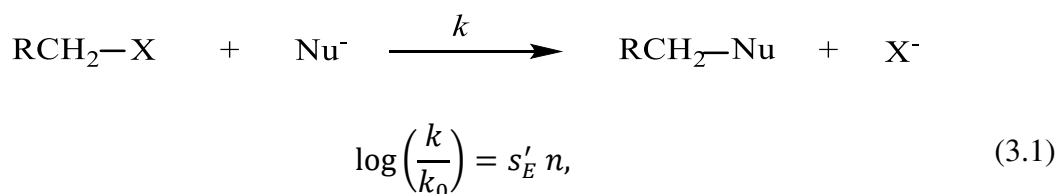
## 1. Introduction

Lorsque l'on mentionne la notion de réactivité chimique, on utilise le plus souvent les termes d'électrophile et de nucléophile introduits par Ingold en 1934 [1], afin de caractériser les molécules qui ont tendance à accepter ou à donner des électrons.

L'électrophilie se définit comme « la réactivité relative d'un réactif électrophile ». On parle également de « pouvoir électrophile » [2]. On peut alors établir une échelle de réactivité entre différentes molécules en fonction de cette propriété intrinsèque. Expérimentalement, on mesure l'électrophilie d'une molécule de différentes manières.

Legon et Millen ont obtenu des paramètres d'électrophilie à partir de constantes de force de liaisons hydrogène, mesurées par spectroscopie rotationnelle [3]. Pasanen *et. al.* ont proposé une relation à partir des fréquences de vibration de la liaison C=O et des déplacements chimiques en RMN du  $^{13}\text{C}$ , afin d'obtenir l'électrophilie expérimentale de composés carbonylés [4]. Topol *et. al.* [5] ont quant à eux utilisé les potentiels d'oxydo-réduction afin de déterminer la réactivité de composés électrophiles, obtenus expérimentalement par voie électrochimique. L'électrophilie expérimentale peut également être obtenue par le biais de mesures cinétiques [6–8], qui est actuellement l'approche recommandée par l'IUPAC (Union internationale de chimie pure et appliquée) puisque même si l'électrophilie est étroitement liée à l'acidité de Lewis (LA) qui « est mesurée par des constantes d'équilibre relatives » (c'est un point de vue thermodynamique), l'électrophilie « est mesurée par des constantes de vitesse relatives » (donc une propriété cinétique).

Pour les méthodes cinétiques, c'est en 1953 que Swain et Scott [9] proposent la première méthode expérimentale, afin d'estimer la nucléophilie  $n$  d'une réaction entre un nucléophile et un électrophile halogéné :



où  $k$  est la constante de vitesse d'une réaction entre un électrophile donné (ici  $\text{RCH}_2\text{X}$ ) et d'un nucléophile (à  $25^\circ\text{C}$ ),  $k_0$  la constante de vitesse de la réaction entre le même nucléophile et le bromure de méthyle ( $\text{MeBr}$ ) et  $s'_E$  une caractéristique propre à l'électrophile (correspondant à une mesure de sa sensibilité à la nucléophilie du réactif).

Dans le cas où la loi d'Arrhenius est vérifiée, c'est-à-dire que les réactions avec les énergies d'activation les plus faibles sont les plus rapides et inversement celles qui ont les énergies d'activation les plus élevées sont les plus lentes, la relation est équivalente à une relation d'énergie libre permettant de construire une échelle de réactivité d'une série de réactifs nucléophiles, à partir des constantes de vitesse  $k$  pour la réaction du bromure de méthyle avec les nucléophiles dans l'eau à 25°C,  $s'_E$  étant défini comme égal à 1 pour ces réactions et  $n$  étant défini comme égal à 0 pour l'hydrolyse du bromure de méthyle.

$$n = \log k_0 - \log k_{MeBr}^{H_2O}. \quad (3.2)$$

Parmi les nombreuses applications de cette échelle [10], on peut citer notamment l'étude mécanistique du réarrangement de Winstein à partir de l'équation (3.1) [11, 12]. Il est néanmoins important de préciser que cette échelle suppose que tous les nucléophiles ont la même réactivité vis-à-vis d'un même site réactif donné. C'est pour cela qu'en 1954, Edwards proposa d'étendre l'équation (3.1) en ajoutant les termes de basicité et de polarisabilité du nucléophile [13, 14] :

$$\log \left( \frac{k}{k_0} \right) = \alpha E_N + \beta H_N, \quad (3.3)$$

où  $\alpha$  et  $\beta$  sont des paramètres propres à un nucléophile qui relient la sensibilité du substrat aux facteurs de basicité et de polarisabilité.  $H_N$  est une mesure de la basicité du nucléophile par rapport aux protons, définie par  $H_N = pK_{a_{NuH}} + 1,74$  où le pKa est celui de l'acide conjugué du nucléophile et la constante 1,74 est la correction du pKa avec  $H_3O^+$ .  $E_N$  est le terme introduit par Edwards pour rendre compte de la polarisabilité du nucléophile. Il est lié au potentiel d'oxydation  $E_0$  de la réaction de dimérisation du nucléophile par  $E_N = E_0 + 2,60$ .

La deuxième échelle importante de nucléophilie est celle proposée par Ritchie en 1972 [15], basée sur l'utilisation d'un paramètre de nucléophilie  $N_+$  indépendant de l'électrophile. Cette échelle est construite à partir de réactions entre des nucléophiles et des cations organiques stables (comme des arènediazoniums, des triaryleméthyles et les cations aryle tropylium) et sur la relation d'énergie libre suivante :

$$\log \left( \frac{k_N}{k_0} \right) = N_+, \quad (3.4)$$

où  $k_N$  est la constante de vitesse d'une réaction entre un cation et un nucléophile dans un solvant donné et  $k_0$  la constante de vitesse de la réaction du même cation avec l'eau dans l'eau.

$k_0$  est alors une propriété propre à l'électrophile (notée  $E_+$ ) et indépendante du nucléophile, ce qui peut permettre d'écrire l'équation (3.4) comme une équation dépendant de deux paramètres spécifiques au nucléophile et à l'électrophile :

$$\log k_N - \log k_0 = N_+, \quad (3.5)$$

$$\log k_N = \log k_0 + N_+, \quad (3.6)$$

$$\log k_N = E_+ + N_+. \quad (3.7)$$

Néanmoins, l'échelle de Ritchie omet l'influence de l'électrophile sur la réactivité et la sélectivité relative du nucléophile [16, 17]: si l'on étudie particulièrement les réactions avec des nucléophiles possédant plusieurs sites de réactivité, la *force* relative de l'électrophile a une influence sur la sélectivité [18]. L'application de l'échelle de Ritchie à un nombre important d'électrophiles organométalliques par Kane-MaGuire et Sweigart [19, 20] a permis de démontrer la sélectivité constante et la capacité de prédiction de l'échelle expérimentale [21, 22]. Elle va ainsi à l'encontre du principe de réactivité-sélectivité (le *RSP* que l'on peut énoncer : « *Au plus la réactivité d'une molécule augmente, au plus la sélectivité diminue* ») démontrant l'inefficacité de ce dernier [23].

## 2. L'échelle de réactivité de Mayr

### 2.1 Présentation générale

En conséquence des limites de l'équation de Ritchie, une expression qui s'applique aux différentes séries de réactions doit inclure un paramètre de sensibilité spécifique aux nucléophiles. C'est ainsi qu'en 1994 Mayr et Patz ont postulé l'équation d'énergie libre suivante [18] :

$$\log k_{20^\circ\text{C}} = s_N(E + N), \quad (3.8)$$

reliant  $k_{20^\circ\text{C}}$  la constante cinétique d'une réaction à  $20^\circ\text{C}$  entre un électrophile et un nucléophile à  $E$  qui correspond au paramètre d'électrophilie (qui ne dépend pas du solvant) et à  $N$  et  $s_N$  qui sont respectivement le paramètre de nucléophilie et la sensibilité spécifique au nucléophile.



Initialement, un électrophile de référence, l'ion bis(p-méthoxyphényl)carbénium, a été sélectionné avec  $E = 0$  pour la construction de l'échelle, à la température de 20°C et en utilisant le dichlorométhane ( $\text{CH}_2\text{Cl}_2$ ) en solvant de référence, afin de construire l'échelle complète applicable à ces conditions.

Cependant, lorsque on utilise seulement l'ion bis (p-méthoxyphényl) carbénium à cette fin, cet électrophile de référence n'est pas adapté pour générer une échelle de nucléophilie complète. Alors qu'il est possible de différencier la réactivité nucléophile des arènes et des alcènes riches en électrons ainsi que des allylsilanes et des éthers d'énol grâce à leurs réactions avec cet ion carbénium, les énamines et les carbanions les plus courants réagissent en revanche, eux, jusqu'à atteindre la limite de diffusion. Pour les arènes et les alcènes non activés, ils réagissent trop lentement ou pas du tout [24]. L'utilisation de cet électrophile ne couvre alors ainsi qu'une portion de l'échelle complète.

Quel que soit l'électrophile sélectionné comme référence, une partie des nucléophiles sera toujours en dehors de la portion couverte par cet électrophile. Pour cette raison, l'échelle n'est pas basée sur un, mais sur plus de 29 électrophiles de structures proches (des cations benzhydrylium et des quinones méthides), dont chacun permet de comparer un groupe différent de nucléophiles, permettant ainsi de construire une échelle complète.

Pour les nucléophiles, la seule référence utilisé est l'allyltriméthylsilane avec  $s_N = 1$ .

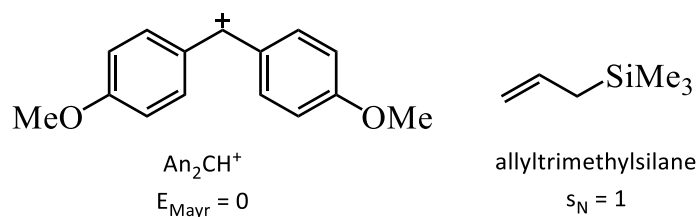


Figure 3.1: Les deux références de l'échelle

A partir de ces deux paramètres, 285 constantes cinétiques du second ordre issues de réactions entre les 29 électrophiles de références et plusieurs nucléophiles permettent de construire l'équation (3.8) [18]. La figure suivante regroupe une portion de l'échelle de Mayr et démontre la capacité de l'échelle à classer des molécules de familles différentes :

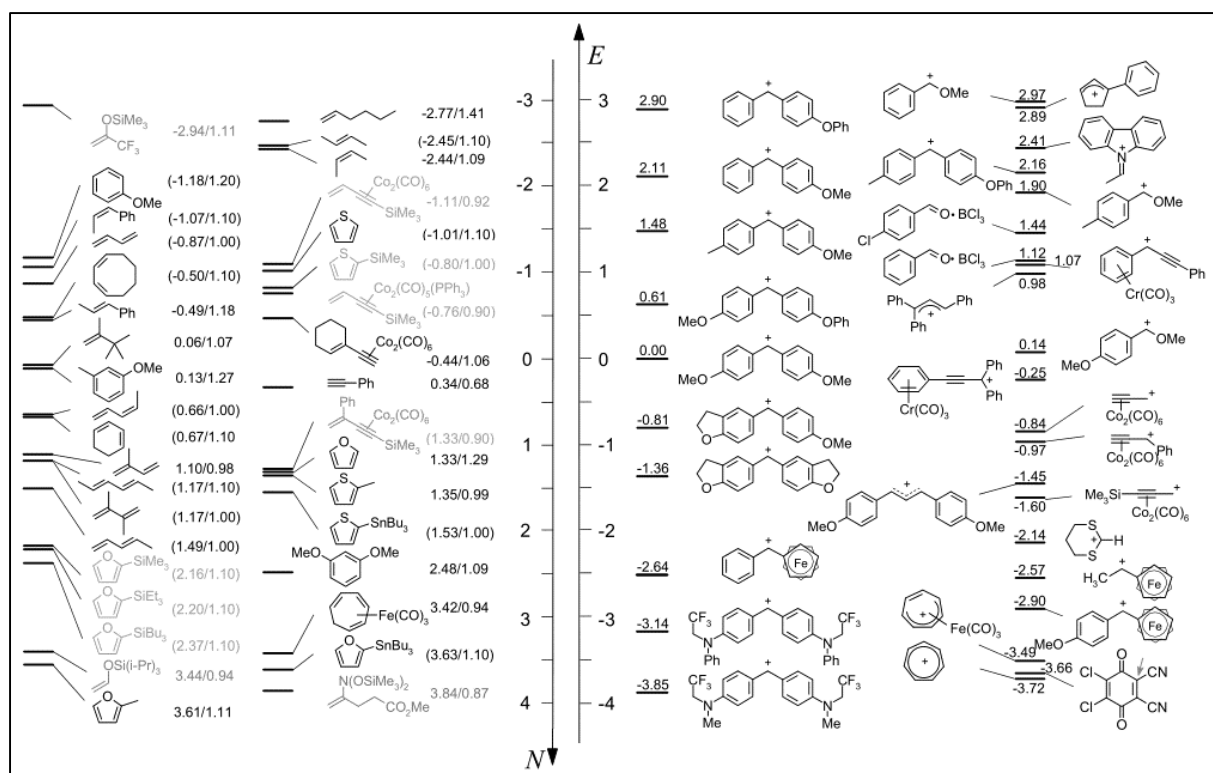


Figure 3.2: Partie de l'échelle expérimentale de Mayr, avec à gauche les nucléophiles et à droite les électrophiles. Les molécules grisées ont une valeur expérimentale avec seulement 1 électrophile de référence [25].

Actuellement, près de 1552 paramètres d'électrophilie et de nucléophilie publiés par Mayr *et al.* sont disponibles et regroupés dans une base de données en ligne [25].

Il est important de noter ici qu'aucune molécule n'est uniquement électrophile ou nucléophile (à certaines exceptions près), on parle de centres (ou sites) électrophiles/nucléophiles sur une molécule. Ainsi, l'application de l'équation de Mayr sur une molécule ambidente (avec deux sites de réactivité) peut permettre une classification des sites sur une même molécule :

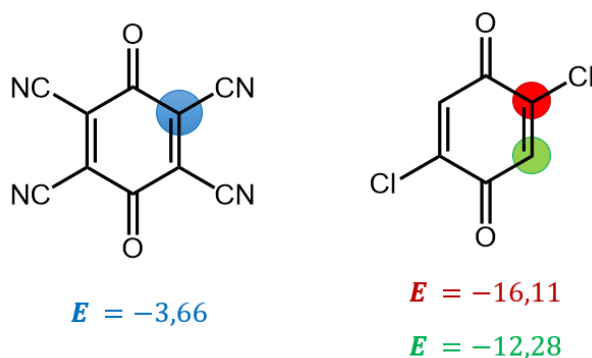


Figure 3.3 : Electrophilie expérimentale pour la 2,3-dichloro-5,6-dicyano-p-benzoquinone (à gauche) et la 2,5-dichloroquinone (à droite)

avec ici un exemple de comparaison de la réactivité électrophile entre la 2,5-dichloroquinone [26] et la 2,3-dichloro-5,6-dicyano-p-benzoquinone [27] où il existe deux types de sites électrophiles.

Expérimentalement, les mesures cinétiques se réalisent par trois méthodes différentes, dont chacune d'entre elles fait appel à une mesure de l'intensité lumineuse absorbée par l'échantillon analysé en fonction du temps. L'application de la loi de Beer-Lambert :

$$A = \varepsilon l C, \quad (3.9)$$

où  $A$  désigne l'absorbance,  $\varepsilon$  la constante d'absorptivité propre à l'entité chimique,  $l$  la longueur du trajet parcouru par la lumière dans le milieu considéré et  $C$  la concentration molaire, permet de suivre la concentration chimique en fonction du temps et ainsi d'obtenir les constantes cinétiques expérimentales.

L'exemple suivant (Figure 3.5) illustre la détermination expérimentale de la constante de vitesse d'une réaction pour la détermination des paramètres  $N$  et  $s_N$ , ici une addition nucléophile est réalisée avec l'ion bis (p-méthoxyphényl) carbénium [24]:

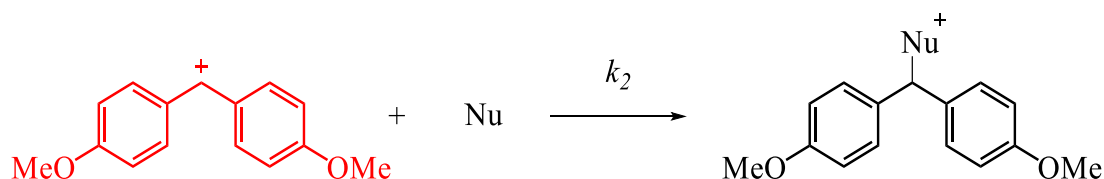


Figure 3.4: Réaction entre un nucléophile et l'ion bis (p-méthoxyphényl) carbénium

Dans un premier temps (1), la longueur d'onde d'absorbance maximale de l'électrophile seul est mesurée expérimentalement (ici  $\lambda = 520$  nm). Puis (2) en se plaçant à la longueur d'onde d'absorption maximale, le nucléophile est ajouté à la solution de l'électrophile et l'absorbance est suivie en fonction du temps, permettant l'obtention de la constante de vitesse observée  $k_{obs}$ . La réaction est ensuite répétée à plusieurs concentrations du nucléophile connues  $[Nu]$  (3) et la pente de la droite des concentrations observées  $k_{obs}$  en fonction des différentes concentrations de nucléophiles utilisées permet de retrouver la constante de vitesse expérimentale de la réaction  $k_2$ .

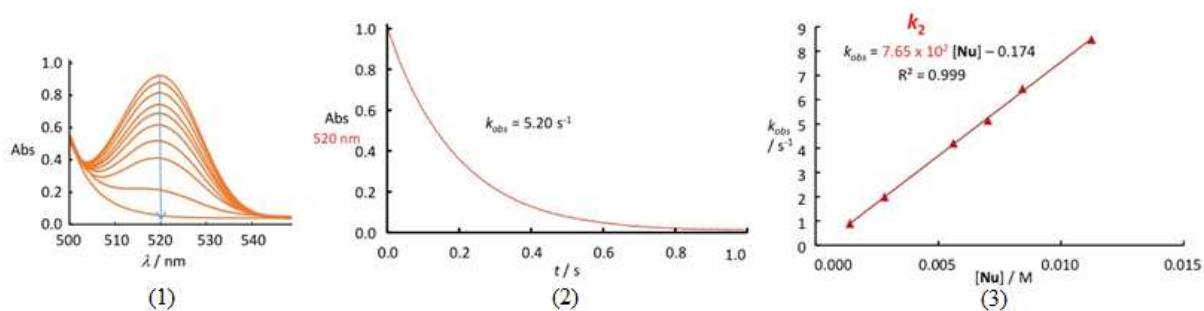


Figure 3.5 : (1) détermination de la longueur d'absorbance maximale (2) mesure de constante de vitesse observée expérimentalement (3) Courbe des constantes observées en fonction de la concentration du nucléophile

Lorsque les réactions sont relativement *lentes*, on peut utiliser un spectromètre UV-visible classique qui mesure l'absorbance directement dans le milieu réactionnel. Lorsque les réactions sont plus rapides, on utilise premièrement un spectromètre à flux stoppé (*stopped flow* en anglais), qui sous flux d'argon, utilise de très faibles quantités et apporte une meilleure précision. Enfin, en dernier, on utilise la méthode de photolyse laser-flash dans lequel un échantillon est d'abord excité par une forte impulsion (appelée impulsion de pompe) de lumière provenant d'un laser pulsé d'une largeur d'impulsion d'ordre de la nanoseconde, picoseconde ou femtoseconde ou par une autre source lumineuse à impulsion courte telle qu'une lampe flash. Le choix de l'instrument dépendra ainsi de l'ordre de grandeur de la constante de vitesse de la réaction étudiée.

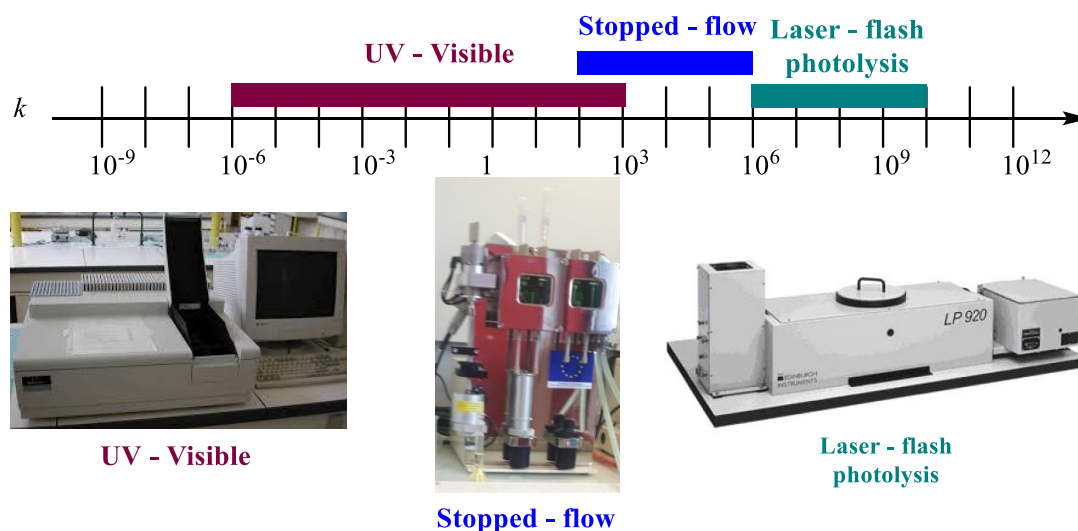


Figure 3.6 : Représentation des différentes méthodes analytiques pour la détermination des constantes de vitesses expérimentales

## 2.2 Applications

En pratique, ces données expérimentales peuvent être facilement utilisées pour prédire la « faisabilité » d'une réaction, ainsi que la sélectivité en chimie organique. En effet, en négligeant l'aspect du solvant, on peut estimer rapidement la faisabilité de la réaction en additionnant les paramètres  $E$  et  $N$ . Puisqu'une réaction entre deux molécules ne peut être plus rapide que la limite de diffusion, si la condition «  $E + N > -5$  » est respectée, la réaction est dite faisable [28, 29].

On peut illustrer cette règle en prenant l'exemple de la réaction avec un électrophile disponible dans la base de données de Mayr (ici  $C_6H_7^+-Fe(CO)_3$ ) et deux nucléophiles, (Nu1 et Nu2) dont on connaît au préalable les valeurs expérimentales :

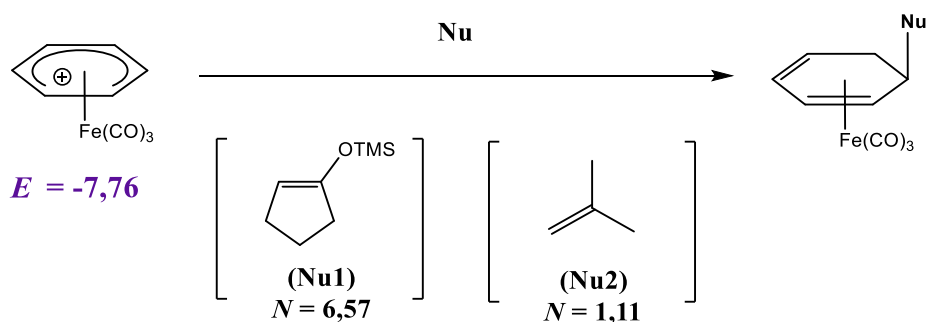


Figure 3.7 : Schéma illustrant la prédiction de réaction grâce à l'échelle d'électrophilie de Mayr

Si l'on regarde la réaction avec le premier nucléophile (Nu1), le 1-triméthylsiloxy)cyclopentène dont  $N = 6,57$ , la condition est respectée ( $E + N = -1,19$ ) et expérimentalement la réaction est observée avec une constante cinétique de  $k = 0,0856 \text{ s}^{-1}$  [18, 29–32]. En revanche si l'on prend le second nucléophile (Nu2), l'isobutène dont  $N = 1,11$ , la condition n'est pas respectée et la réaction n'est pas observée expérimentalement.

En plus de la prédiction rapide de réactions chimiques, l'échelle de Mayr peut être appliquée dans le design de nouveaux composés lorsque des paramètres expérimentaux de molécules de même famille sont disponibles [33–36].

Pour une liste détaillée des exemples d'applications, il convient de regarder les différentes revues publiées par le groupe de Mayr *et al.* [18, 37, 38].

## 2.3 Limites et extension

L'équation (3.8) et l'ensemble des échelles de réactivité basées sur des mesures cinétiques ne garantissent pas qu'avec une valeur prédite de constantes cinétiques élevée, le produit obtenu en fin de réaction ait une stabilité élevée. Une prédiction cinétiquement *favorable* peut être thermodynamiquement *instable*, c'est le cas par exemple de la réaction du DABCO avec les cations benzhydrylium [39]. Pour ce genre de cas, il conviendra d'utiliser l'échelle thermodynamique d'acidité et de basicité de Lewis [40] proposée par Mayr *et al.*.

Un autre point limitant l'utilisation de l'échelle de Mayr est qu'elle est valide uniquement dans le cas de la formation d'une liaison  $\sigma$  : une réaction de chimie organique qui implique un mécanisme concerté (sans intermédiaire réactionnel) ne suivra pas l'équation (3.7). Si l'on prend par exemple une réaction de Diels-Alder entre l'anhydride maléique ( $E = -11.31$ ) [41] et le cyclopentadiène ( $N = 2,30$  et  $s_N = 1,06$ ) [42], l'application de l'échelle de Mayr et de la règle précédente prédit la formation lente de l'intermédiaire zwitterionique :

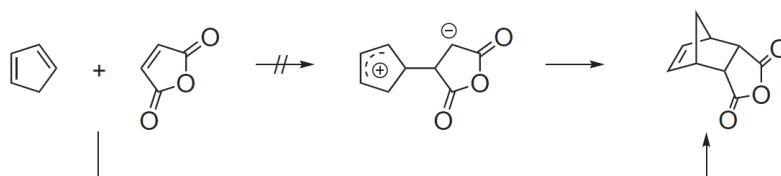


Figure 3.8: Réaction de Diels-Alder entre l'anhydride maléique et le cyclopentadiène

Néanmoins, la réaction suivant un mécanisme concerté s'effectue beaucoup plus rapidement en évitant la formation de l'espèce zwitterionique [43], l'équation (3.8) n'est pas ainsi valable pour une réaction péricyclique classique.

Enfin, la dernière limitation de l'échelle de réactivité dérivée à partir des cations benzhydrylium et des quinone méthides est sa viabilité pour l'ensemble des autres électrophiles sans aucun critère de sélectivité [44]. Si l'on mesure expérimentalement les constantes cinétiques  $\log k_{20^\circ\text{C}}$  pour des électrophiles et que l'on trace  $\log k_{20^\circ\text{C}} / s_N$  en fonction de  $N$ , l'obtention d'une pente de 1 démontre la validité de l'équation (3.8) sur ces mêmes électrophiles. C'est ainsi le cas pour des accepteurs de Michael de structures variées [45] et des ions iminium [46], qui sont alors inclus dans l'échelle de réactivité.

En revanche, si l'on étudie en détail la corrélation entre  $\log k_{20^\circ\text{C}} / s_N$  en fonction de  $N$  pour un ensemble de réactions  $S_N2$  entre des nucléophiles et des  $C_{sp^3}$ -électrophiles (comme par exemple des halogénoalcane), la pente observée diffère de 1 (par exemple, la pente observée

est de 0,61 avec le cation S-methyldibenzothiophenium [47]). Leur réactivité relative suit alors la réactivité des cations benzhydrylium, mais un critère de sélectivité spécifique à l'électrophile est nécessaire pour appliquer l'équation (3.8). Mayr, Breugst et Phan ont ainsi proposé en 2006 une extension de l'échelle de réactivité à appliquer dans ces cas, l'équation générale d'énergie libre de réactivité électrophile-nucléophile est alors :

$$\log k_{20^\circ\text{C}} = s_N s_E (E + N), \quad (3.10)$$

où  $s_E$  est le paramètre de susceptibilité spécifique à l'électrophile. Lorsque  $s_E$  est égal à 1, on retrouve l'équation (3.8). L'échelle de Swain et Scott, qui est alors basée sur des réactions d'halogénoalcane et de nucléophiles pour lesquels  $s_N \approx 0,6$ , permet de relier l'équation (3.1) à l'équation (3.10), dont elle est un cas particulier. L'échelle de Ritchie est quant à elle un cas particulier de l'équation (3.8) lorsque la sélectivité des nucléophiles n'est pas prise en compte. Le schéma suivant regroupe ainsi les équations des différentes échelles de réactivité [47] :

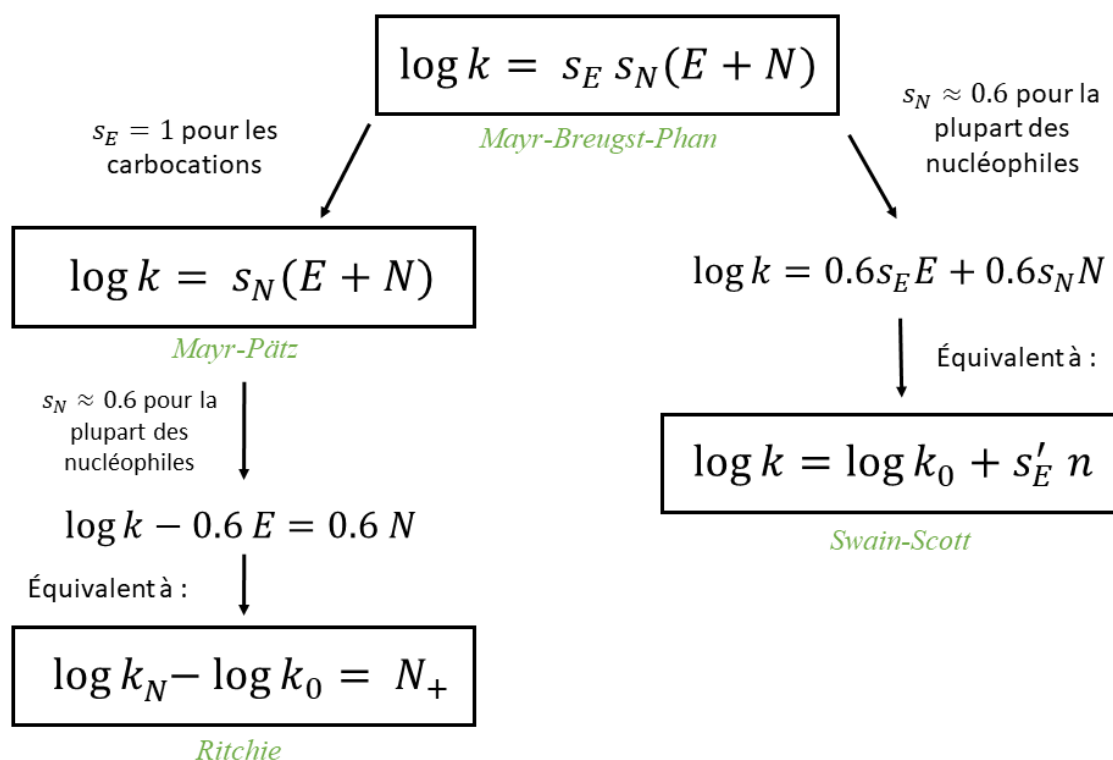


Figure 3.9: Schéma illustrant les relations entre les différentes échelles expérimentales de réactivité

- [1] C. K. Ingold. *Chem. Rev.* 15 (1934), 225–274
- [2] P. Muller. *Pure Appl. Chem* 66 (1994), 1077–1184
- [3] A. C. Legon, D. J. Millen. *J. Am. Chem. Soc.* 109 (1987), 356–358
- [4] H. Neuvonen, K. Neuvonen, A. Koch, et al. *J. Org. Chem.* 67 (2002), 6995–7003
- [5] I. A. Topol, C. Mcgrath, E. Chertova, et al. (2001), 1434–1445
- [6] Y. Zhou, L. Li, H. Ye, et al. *J. Am. Chem. Soc.* 138 (2016), 381–389
- [7] N. Santschi, T. Nauser. *ChemPhysChem* 18 (2017), 2973–2976
- [8] N. Rozatian, I. W. Ashworth, G. Sandford, D. R. W. Hodgson. *Chem. Sci.* 9 (2018), 8692–8702
- [9] C. Gardner Swain, C. B. Scott. *J. Am. Chem. Soc.* 75 (1953), 141–147
- [10] R. G. Pearson, H. Sobel, J. Songstad. *J. Am. Chem. Soc.* 90 (1968), 319–326
- [11] A. Gagneux, S. Winstein, W. Young. *J. Am. Chem. Soc.* 82 (1960), 5956–5957
- [12] A. S. Carlson, J. J. Topczewski. *Org. Biomol. Chem.* 17 (2019), 4406–4429
- [13] J. O. Edwards. *J. Am. Chem. Soc.* 76 (1954), 1540–1547
- [14] J. O. Edwards. *J. Am. Chem. Soc.* 78 (1956), 1819–1820
- [15] C. D. Ritchie. *Acc. Chem. Res.* 5 (1972), 348–354
- [16] P. Denton, C. D. Johnson. *J. Chem. Soc. Perkin Trans. 2* (1995), 477–481
- [17] C. D. Ritchie, *Revue canadienne de chimie*, 64, (1986), 2239–2250
- [18] H. Mayr, M. Patz. *Angew. Chemie Int. Ed. English* 33 (1994), 938–957
- [19] L. A. P. Kane-Maguire, E. D. Honig, D. A. Sweigart. *J. Chem. Soc. Chem. Commun.* (1984), 345–347
- [20] L. A. P. Kane-Maguire, E. D. Honig, D. A. Sweigart. *Chem. Rev.* 84 (1984), 525–543
- [21] D. J. Evans, L. A. P. Kane-Maguire. *J. Organomet. Chem.* 312 (1986), 24–26
- [22] L. A. P. Kane-Maguire, P. A. Mouncher, A. Salzer. *J. Organomet. Chem.* 347 (1988), 383–390
- [23] H. Mayr, A. R. Ofial. *Angew. Chemie - Int. Ed.* 45 (2006), 1844–1854
- [24] H. Mayr. *Tetrahedron* 71 (2015), 5095–5111
- [25] <https://www.cup.lmu.de/oc/mayr/reaktionsdatenbank2/>
- [26] X. Guo, H. Mayr. *J. Am. Chem. Soc.* 136 (2014), 11499–11512
- [27] X. Guo, H. Mayr. *J. Am. Chem. Soc.* 135 (2013), 12377–12387
- [28] R. Lucius, R. Loos, H. Mayr. *Angew. Chemie - Int. Ed.* 41 (2002), 91–95
- [29] H. Mayr, T. Bug, M. F. Gotta, et al. *J. Am. Chem. Soc.* 123 (2001), 9500–9512
- [30] R. Lucius, R. Loos, H. Mayr. *Angew. Chemie - Int. Ed.* 41 (2002), 91–95



- [31] H. Mayr, B. Kempf, A. R. Ofial. *Acc. Chem. Res.* 36 (2003), 66–77
- [32] A. J. Birch, L. F. Kelly, A. S. Narula. *Tetrahedron* 38 (1982), 1813–1823
- [33] K. Umezawa, M. Yoshida, M. Kamiya, et al. *Nat. Chem.* 9 (2017), 279–286
- [34] R. J. Mayer, A. R. Ofial. *Angew. Chemie - Int. Ed.* 58 (2019), 17704–17708
- [35] M. C. Holland, S. Paul, W. B. Schweizer, et al. *Angew. Chemie - Int. Ed.* 52 (2013), 7967–7971
- [36] S. Lakhdar, R. Goumont, F. Terrier, et al. *Org. Biomol. Chem.* 5 (2007), 1744–1751
- [37] H. Mayr, A. R. Ofial. *SAR QSAR Environ. Res.* 26 (2015), 619–646
- [38] H. Mayr, B. Kempf, A. R. Ofial. *Acc. Chem. Res.* 36 (2003), 66–77
- [39] M. Baidya, S. Kobayashi, F. Brotzel, et al. *Angew. Chemie - Int. Ed.* 46 (2007), 6176–6179
- [40] H. Mayr, J. Ammer, M. Baidya, et al. *J. Am. Chem. Soc.* 137 (2015), 2580–2599
- [41] D. S. Allgäuer, H. Mayr. *European J. Org. Chem.* 2014 (2014), 2956–2963
- [42] H. Mayr, B. Kempf, A. R. Ofial. *Acc. Chem. Res.* 36 (2003), 66–77
- [43] J. Sauer, H. Wiest, a Mielert. *Chem. Ber.* 97 (1964), 3183–3207
- [44] H. Mayr, A. R. Ofial. *SAR QSAR Environ. Res.* 26 (2015), 619–646
- [45] H. Asahara, H. Mayr. *Chem. - An Asian J.* 7 (2012), 1401–1407
- [46] R. Appel, S. Chelli, T. Tokuyasu, et al. *J. Am. Chem. Soc.* 135 (2013), 6579–6587
- [47] T. B. Phan, M. Breugst, H. Mayr. *Angew. Chem.* 118 (2006), 3869–3874

## CHAPITRE 4 : Prédiction théorique de l'électrophilie expérimentale

|                                                                                                                                                        |     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 1. <i>Les descripteurs moléculaires et atomiques peuvent-ils prédire l'électrophilie des accepteurs de Michaël ?</i> .....                             | 107 |
| 1.1. Préambule.....                                                                                                                                    | 107 |
| 1.2. Article.....                                                                                                                                      | 109 |
| 1.3. Conclusion et perspectives .....                                                                                                                  | 121 |
| 2. <i>Sur l'influence des effets dynamiques sur les descripteurs de réactivité</i> .....                                                               | 122 |
| 2.1. Préambule.....                                                                                                                                    | 122 |
| 2.2. Article.....                                                                                                                                      | 123 |
| 2.3. Conclusion et perspectives .....                                                                                                                  | 128 |
| 3. <i>Prédire l'électrophilie expérimentale à partir de descripteurs quantiques et topologiques : une approche par apprentissage automatique</i> ..... | 130 |
| 3.1. Préambule.....                                                                                                                                    | 130 |
| 3.2. Article.....                                                                                                                                      | 131 |
| 3.3. Conclusion.....                                                                                                                                   | 144 |



# 1. Les descripteurs moléculaires et atomiques peuvent-ils prédire l'électrophilie des accepteurs de Michael ?

## 1.1 Préambule

Prédire l'électrophilie d'une molécule est d'une importance capitale pour une meilleure compréhension de la réactivité chimique, que ce soit d'un point de vue théorique ou bien expérimental. Dans ce sens, les descripteurs quantiques issus de la DFT conceptuelle ont déjà montré leurs capacités de prédiction pour les cations benzhydrylium qui sont les références de l'échelle d'électrophilie de Mayr. Bien que celle-ci couvre une multitude d'électrophiles, très peu d'études théoriques discutent de la prédiction de l'électrophilie expérimentale pour les autres familles.

Pour cette première étude, notre choix s'est porté sur la famille des accepteurs de Michael (des composés  $\alpha$ - $\beta$ -insaturés), grâce, entre autres, à l'ajout récent par Mayr et son équipe de 15 paramètres expérimentaux d'électrophilie en 2017, rendant le nombre de valeurs disponibles suffisantes pour une étude théorique. L'addition de Michael est une réaction importante en chimie organique puisqu'elle résulte en la formation d'une liaison carbone-carbone (ou carbone-hétéroatome), sans assistance par des complexes organométalliques ni catalyseur(s) à base de métaux.

Le schéma type d'une addition de Michael avec la buténone est le suivant :

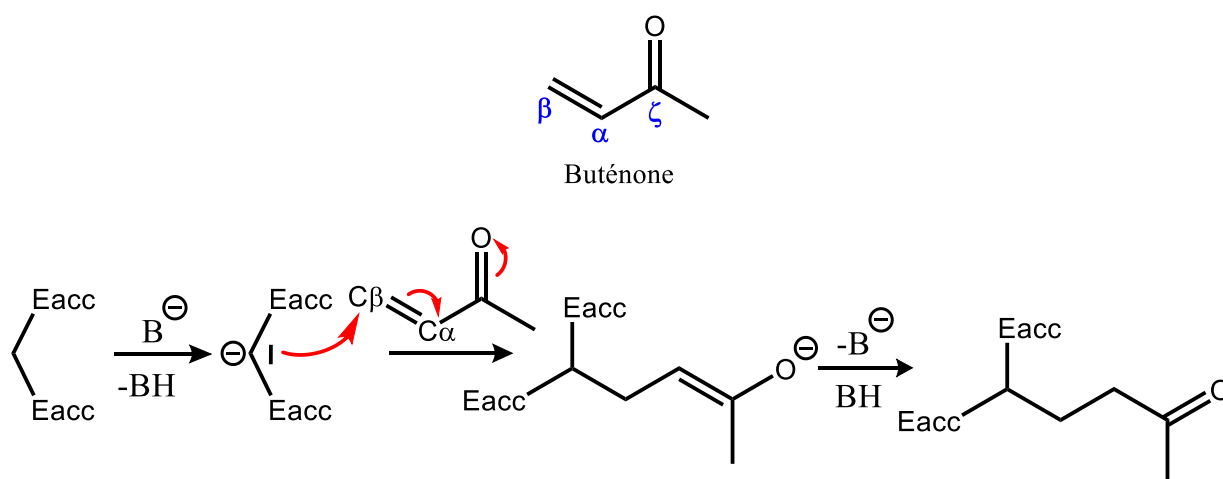


Figure 4.1 : Réaction et mécanisme type d'une addition de Michael avec la buténone en exemple

Sur cette molécule (ainsi que sur l'ensemble des accepteurs de Michael), la valeur expérimentale d'électrophilie (-16,76), ainsi que toutes les valeurs discutées par la suite, correspondent à l'électrophilie du carbone  $\beta$ . Elle a été mesurée dans le DMSO avec 3 nucléophiles de référence (pour lesquels les valeurs des paramètres nucléophiles sont connues). Expérimentalement, la réaction avec les nucléophiles est réalisée dans des conditions telles que l'attaque se fasse uniquement sur le carbone  $\beta$ , carbone le plus réactif de la molécule. Cela se produit en utilisant des ylures de pyridinium, permettant la formation d'adduits qui subissent des réactions intramoléculaires ultérieures rapides, supprimant ainsi la réversibilité de l'étape de formation de la liaison C-C :

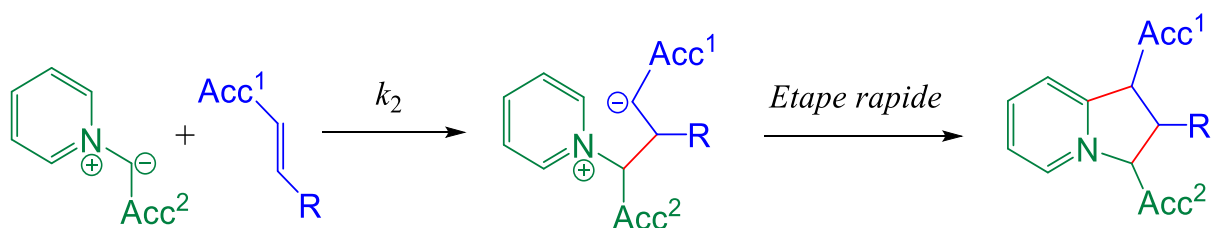


Figure 4.2: Réaction entre les ylures de pyridinium et les accepteurs de Michael. (Acc signifie groupement accepteur)

Nous nous sommes alors intéressés à la capacité de prédiction des descripteurs de réactivité chimique disponibles dans la littérature. La liste complète des descripteurs utilisés dans cette étude est présentée dans le tableau 2 de l'article ci-après.



# Can molecular and atomic descriptors predict the electrophilicity of Michael acceptors?

Guillaume Hoffmann<sup>1</sup> · Vincent Tognetti<sup>1</sup> · Laurent Joubert<sup>1</sup>

Received: 22 June 2018 / Accepted: 16 August 2018  
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

## Abstract

In this paper, we assess the ability of various intrinsic molecular and atomic descriptors, grounded in the conceptual density functional theory and the quantum theory of atoms-in-molecules frameworks, to predict the electrophilicity of Michael acceptors, which are fundamental organic reagents involved in the formation of carbon–carbon bonds. To this aim, linear and multilinear regressions between these theoretical properties and the experimental values gathered in Mayr–Patz’ scale were performed. The relevance of quantum chemical descriptors are then discussed.

**Keywords** Mayr’s electrophilicity scale · Michael additions · Molecular descriptors · Atomic descriptors · Reactivity indices · Conceptual DFT · Quantum theory of atoms-in-molecules (QTAIM)

## Introduction

Being able to differentiate and classify molecules by their propensity to accept or donate electrons, from either a theoretical or experimental point of view, is of particular importance in organic chemistry for a better understanding of reactivity. For this purpose, several quantitative scales for electrophilicity and nucleophilicity have been established throughout the years [1–3]. The scale proposed by Mayr et al. [4] is certainly among the most comprehensive and useful available. It basically relies on the following equation, linking the rate constant measured at 20 °C ( $k_{20\text{ °C}}$ ) for the reaction between a given nucleo-

phile and electrophile with the intrinsic reactivity properties of the reactants:

$$\log k_{20\text{ °C}} = s_N (N_{\text{Mayr}} + E_{\text{Mayr}}) \quad (1)$$

In Eq. 1,  $E_{\text{Mayr}}$  denotes the electrophilicity parameter;  $N_{\text{Mayr}}$  the nucleophilicity parameter; and  $s_N$ , the so-called nucleophile-specific susceptibility parameter. It is worth mentioning that  $E_{\text{Mayr}}$  is related almost exclusively to electronic effects and is independent of solvent. In practice, these molecular descriptors can then easily be used to predict rates and selectivity in organic chemistry [5, 6], and to guide synthetic chemists to selecting the most relevant reagents. From a theoretical aspect, such an extensive scale with an online freely available database of over 1400 molecules [7] fosters the design of new theoretical tools capable of recovering and extending (for instance on not yet synthesized molecules) these experimental values.

With this aim in mind, a large number of theoretical reactivity indices, mainly within the framework of conceptual density functional theory (CDFT) [8, 9], have been proposed in recent years and tested with respect to Mayr’s scale. One can, for instance, cite the recent papers by Pereira et al. [10] and González et al. [11]. One should notice that, in the literature, efforts have been concentrated mainly on retrieving the experimental electrophilic reactivities of benzhydrylium ions [12–14], which constitute a subclass of particular importance because they were established

This paper belongs to Topical Collection International Conference on Systems and Processes in Physics, Chemistry and Biology (ICSPCB-2018) in honor of Professor Pratim K. Chattaraj on his sixtieth birthday

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00894-018-3802-9>) contains supplementary material, which is available to authorized users.

✉ Vincent Tognetti  
vincent.tognetti@univ-rouen.fr

✉ Laurent Joubert  
laurent.joubert@univ-rouen.fr

<sup>1</sup> COBRA UMR 6014 & FR 3038, INSA Rouen, CNRS, Université de Rouen Normandie, 1 rue Tesnière, 76821 Mont St Aignan, France

as reference electrophiles for building Mayr-Patz' scale [6, 15]. The use of frontier molecular orbitals [16, 17], or of the intrinsic reactivity index [18], or of descriptors derived from information theory, have been recently discussed [19]. From our side, we showed that the constant sign domains of the dual descriptor [20], as well as the atomic electronegativities [13] and atomic energies [21] were also successful.

This prominence of benzhylium ions in electrophilicity studies is notable due to their ability to correctly mimic the behavior of the carbocations that are generated as typical intermediates in organic synthesis [22]. However, despite their predating omnipresence in theoretical studies devoted to Mayr-Patz' scale, many other electrophiles, such as aldehydes, ketones and quinones are of considerable interest and should not be overlooked.

For such reasons, we will focus in this work on the electrophilic reactivity of  $\alpha$ - $\beta$ -unsaturated compounds known as Michael acceptors, already investigated by Domingo and co-workers [23]. The generic Michael addition [24] is epitomized by the reaction depicted in Scheme 1: reagent **1** features an acidic proton in the  $\alpha$  position with respect to two electron-withdrawing groups (EWGs) that can be abstracted by a base **B**. The generated carbanion can attack an enone molecule in the  $\beta$  position of the C=C double bond, forming an enolate intermediate that can be subsequently protonated to afford the final product.

Such 1,4-nucleophilic addition reactions (which can be extended to compounds involving heteroatoms) are often key processes in total synthesis since they result in the formation of a C–C bond between the two initial reactants. From a physicochemical point of view, notably thanks to the very recent addition of 15 new empirical electrophilicity ( $E_{\text{Mayr}}$ ) parameters by Mayr's group [25], the amount of available experimental data has become large enough to assess the performances of quantum theoretical descriptors for this particular reaction.

In this paper, we will focus mainly on chemical descriptors stemming from CDFT and from the quantum theory of atoms-in-molecules (QTAIM) [26, 27]. Both frameworks are indeed based on the same ingredient—the electron density—which is the primary observable in quantum chemistry. They are thus complementary from our point of view. In the last decade, we have notably exploited the synergy of CDFT and QTAIM in organic chemistry and organometallics [28–37].

All CDFT and QTAIM descriptors will be described in detail in the next section. They will be then applied (using the protocol explained in the computational section) to the experimental dataset (taken from Mayr's Database of Reactivity Parameters [7]) in order to design a predicting tool for the potential reactivity of Michael acceptors.

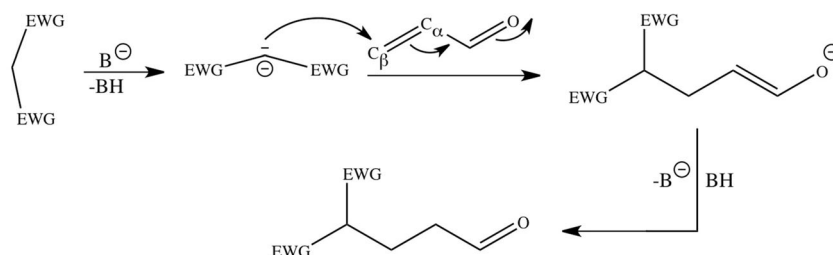
It should be stressed that this approach is centered on the intrinsic properties of the reactants, independently of the way they interact along the reaction, and relies on a strong assumption that we explain now. First, we presume that the reaction is under kinetic (rather than thermodynamical) control. Then, according to Arrhenius' law, the rate constant depends on the activation barrier and thus on the property of the transition state, which is a point on the potential energy surface where the reagents may strongly interact and might be significantly distorted. However, in case the reaction is very exothermic, one can then conjecture that the transition state (TS) is early. The application of the Hammond postulate implies that the physicochemical TS properties are expected to be enough close to the reactants, or at least can be safely extrapolated from them.

## The studied reactivity descriptors

The first category of descriptors gathers those from CDFT. They can be classified into two categories: basic descriptors (BD) and composite descriptors (CD). We define BDs as those corresponding to direct derivatives of the electronic energy  $E_e$  with respect to the natural variables within the chosen ensemble. For instance, in the canonical representation,  $E_e$  is expressed as a function of the number of electrons  $N$  and as a functional of the external potential generated by the nuclei,  $v(\vec{r})$ . The primary aim of CDFT is to provide a chemical gist to these derivatives [38, 39]. They can be either global (one value for the whole molecule), local (depending on each real space position  $\vec{r}$ ), or non-local [depending on two ( $\vec{r}$ ,  $\vec{r}'$ ) or more positions].

The three global BDs (GBDs) we will consider are the electronic chemical potential  $\mu$  [40], the molecular hardness  $\eta$  [41], and the hyperhardness  $\gamma$  [42, 43] according to:

**Scheme 1** A prototypical Michael reaction



$$\mu = \left( \frac{\partial E_e[N, v(\vec{r})]}{\partial N} \right)_{v(\vec{r})}; \eta = \left( \frac{\partial^2 E_e[N, v(\vec{r})]}{\partial N^2} \right)_{v(\vec{r})}; \quad (2)$$

$$\gamma = \left( \frac{\partial^3 E_e[N, v(\vec{r})]}{\partial N^3} \right)_{v(\vec{r})}$$

The three local BSs (LBDs) under study are the electron density  $\rho$ , the Fukui function  $f$  [44], and the standard dual descriptor  $f^{(2)}$  [45]:

$$\rho(\vec{r}) = \left( \frac{\delta E_e[N, v(\vec{r})]}{\delta v(\vec{r})} \right)_N; \quad (3)$$

$$f(\vec{r}) = \frac{\delta^2 E_e[N, v(\vec{r})]}{\partial N \delta v(\vec{r})};$$

$$f^{(2)}(\vec{r}) = \frac{\delta^3 E_e[N, v(\vec{r})]}{\partial^2 N \delta v(\vec{r})}$$

Note that global descriptors may characterize the whole reactivity of the molecule, but they do not give insight into regioselectivity, unlike local functions.

In practice, there exist two main models to evaluate these derivatives: the so-called parabolic (P) and straight-lines (S) models. In the P model, the value for a non-integer number of electrons between  $N-1$  and  $N+1$  is obtained through the unique continuously differentiable parabola that gives the correct energy for the  $N-1$ ,  $N$ , and  $N+1$  points. In the S model, there is a derivative discontinuity [46] for the energy when crossing an integer value for the electron number: the line connecting  $E_e(N-1)$  to  $E_e(N)$  has a different slope than that

linking  $E_e(N)$  to  $E_e(N+1)$ . In this model, one thus distinguishes between the right and left derivatives, giving rise, for instance, to the  $\mu^+$ ,  $\mu^-$  global descriptors, and to the  $f^+$  and  $f^-$  local functions.

Once the energy model is chosen, one can use two different approaches for the numerical evaluation: the finite difference linearization (FDL) and the frontier molecular orbital (FMO) one (in the following,  $\varepsilon_{\text{HOMO}}$  and  $\varepsilon_{\text{LUMO}}$  will denote the Kohn-Sham orbital energy of the highest occupied MO and of the lowest unoccupied one, respectively). Table 1 collects the working equations for both approaches.

As previously mentioned, these basic descriptors can be mixed to design CDs, which can be evaluated within either the FDL or FMO approaches, and which are presented in Table 2. One of the most popular CD is probably the global electrophilicity index, whose scope has been extensively reviewed by Chattaraj [47, 48]:

$$\omega = \frac{\mu^2}{2\eta} \quad (4)$$

Similarly, Fukui functions can be multiplied by selected global descriptors to afford new local descriptors, as notably proposed by Chattaraj and co-workers [49–51]. They can be theoretically justified either by shifting towards other ensembles (for instance, the global softness  $S$  [52], and the local softness  $s(\vec{r}) = f(\vec{r})/\eta$  [53] are basic descriptors in the so-called grand-canonical ensemble, recommended to compare species with different electron numbers) or by scrutinizing relevant Taylor expansions (this is the case for  $\omega$ ,  $\Delta\rho^+$  and  $\Delta\rho^-$ ) [54], in either the P or S models.

However, experimentalists may be not well accustomed with local functions that convey infinite information (at any

**Table 1** Basic descriptors from conceptual density functional theory (DFT) considered in this study. *P* Parabolic model, *S* straight-lines model, *GBD* global basic descriptors, *LBD* local basic descriptors, *FMO* frontier

| Name                          | Category | Formula                                                                                                                                                                                                                                                                                                               |
|-------------------------------|----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Electronic chemical potential | P-GBD    | $\begin{cases} \mu_{\text{FDL}} = \frac{E_e(N+1) - E_e(N-1)}{2} \\ \mu_{\text{FMO}} = \frac{\varepsilon_{\text{HOMO}} + \varepsilon_{\text{LUMO}}}{2} \end{cases}$                                                                                                                                                    |
|                               | S-GBD    | $\begin{cases} \mu_{\text{FDL}}^+ = E_e(N+1) - E_e(N) \\ \mu_{\text{FMO}}^+ = \varepsilon_{\text{LUMO}} \end{cases} \quad \begin{cases} \mu_{\text{FDL}}^- = E_e(N) - E_e(N-1) \\ \mu_{\text{FMO}}^- = \varepsilon_{\text{HOMO}} \end{cases}$                                                                         |
| Molecular hardness            | P-GBD    | $\begin{cases} \eta_{\text{FDL}} = E_e(N+1) - 2E_e(N) + E_e(N-1) \\ \eta_{\text{FMO}} = \varepsilon_{\text{LUMO}} - \varepsilon_{\text{HOMO}} \end{cases}$                                                                                                                                                            |
| Hyperhardness                 | P-GBD    | $\gamma_{\text{FMO}} = \varepsilon_{\text{LUMO}} - 2\varepsilon_{\text{HOMO}} + \varepsilon_{\text{HOMO}-1}$                                                                                                                                                                                                          |
| Fukui functions               | S-LBD    | $\begin{cases} f_{\text{FDL}}^+(\vec{r}) = \rho_{N+1}(\vec{r}) - \rho_N(\vec{r}) \\ f_{\text{FMO}}^+(\vec{r}) = \rho_{\text{LUMO}}(\vec{r}) \end{cases} \quad \begin{cases} f_{\text{FDL}}^-(\vec{r}) = \rho_N(\vec{r}) - \rho_{N-1}(\vec{r}) \\ f_{\text{FMO}}^-(\vec{r}) = \rho_{\text{HOMO}}(\vec{r}) \end{cases}$ |
|                               | S-LBD    | $\begin{cases} f_{\text{FDL}}^{(2)}(\vec{r}) = \rho_{N+1}(\vec{r}) - 2\rho_N(\vec{r}) + \rho_{N-1}(\vec{r}) \\ f_{\text{FMO}}^{(2)}(\vec{r}) = \rho_{\text{LUMO}}(\vec{r}) - \rho_{\text{HOMO}}(\vec{r}) \end{cases}$                                                                                                 |
| Dual descriptor               |          |                                                                                                                                                                                                                                                                                                                       |

molecular orbital, *FDL* finite difference linearization, *HOMO* highest occupied molecular orbital, *LUMO* lowest unoccupied molecular orbital



**Table 2** Composite descriptors from conceptual DFT considered in this study. *GCD* General composite descriptors, *LCD* local composite descriptors

| Name                       | Category | Formula                                                                                                                                                                                                                                                   |
|----------------------------|----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Molecular softness         | P-GCD    | $S_{\text{FDL, FMO}} = 1/\eta_{\text{FDL, FMO}}$                                                                                                                                                                                                          |
| Nucleofugality             | P-GCD    | $\Delta E_{\text{FDL, FMO}}^{\text{nucleo}} = \frac{(\mu_{\text{FDL, FMO}} + \eta_{\text{FDL, FMO}})^2}{2\eta_{\text{FDL, FMO}}}$ ; $\lambda_{\text{FDL, FMO}}^{\text{nucleo}} = e^{-\beta_N \Delta E_{\text{FDL, FMO}}^{\text{nucleo}}}$                 |
| Electrofugality            | P-GCD    | $\Delta E_{\text{FDL, FMO}}^{\text{electro}} = \frac{(\mu_{\text{FDL, FMO}} - \eta_{\text{FDL, FMO}})^2}{2\eta_{\text{FDL, FMO}}}$ ; $\lambda_{\text{FDL, FMO}}^{\text{electro}} = e^{-\beta_E \Delta E_{\text{FDL, FMO}}^{\text{electro}}}$              |
| Electrophilicity index     | P-GCD    | $\omega_{\text{FDL, FMO}} = \frac{\mu_{\text{FDL, FMO}}^2}{2\eta_{\text{FDL, FMO}}}$                                                                                                                                                                      |
| Electrodonating power      | S-GCD    | $\omega_{\text{FDL, FMO}}^- = \frac{(3\mu_{\text{FDL, FMO}}^- + \mu_{\text{FDL, FMO}}^+)^2}{16(\mu_{\text{FDL, FMO}}^+ - \mu_{\text{FDL, FMO}}^-)}$                                                                                                       |
| Electroaccepting power     | S-GCD    | $\omega_{\text{FDL, FMO}}^+ = \frac{(\mu_{\text{FDL, FMO}}^- + 3\mu_{\text{FDL, FMO}}^+)^2}{16(\mu_{\text{FDL, FMO}}^+ - \mu_{\text{FDL, FMO}}^-)}$                                                                                                       |
| Local softness             | S-LCD    | $\begin{cases} s_{\text{FDL, FMO}}^+(\vec{r}) = S f_{\text{FDL, FMO}}^+(\vec{r}) \\ s_{\text{FDL, FMO}}^-(\vec{r}) = S f_{\text{FDL, FMO}}^-(\vec{r}) \end{cases}$                                                                                        |
| Composite Fukui functions  | S-LCD    | $\mu_{\text{FDL, FMO}}^+ f_{\text{FDL, FMO}}^+(\vec{r}); \mu_{\text{FDL, FMO}}^- f_{\text{FDL, FMO}}^-(\vec{r}); \omega_{\text{FDL, FMO}} f_{\text{FDL, FMO}}^+(\vec{r})$                                                                                 |
| Local chemical potential   | S-LCD    | $\mu_{\text{FDL, FMO}}(\vec{r}) = \frac{1}{2} (\mu_{\text{FDL, FMO}}^- f_{\text{FDL, FMO}}^-(\vec{r}) + \mu_{\text{FDL, FMO}}^+ f_{\text{FDL, FMO}}^+(\vec{r}))$                                                                                          |
| Local hardness             | S-LCD    | $\eta_{\text{FDL, FMO}}(\vec{r}) = \mu_{\text{FDL, FMO}}^+ f_{\text{FDL, FMO}}^+(\vec{r}) - \mu_{\text{FDL, FMO}}^- f_{\text{FDL, FMO}}^-(\vec{r})$                                                                                                       |
| Multiphilic descriptor     | S-LCD    | $f_{\omega}^{(2)}(\vec{r}) = \omega_{\text{FDL, FMO}} f_{\text{FDL, FMO}}^{(2)}(\vec{r})$                                                                                                                                                                 |
| Composite dual descriptors | S-LCD    | $S_{\text{FDL, FMO}} f_{\text{FDL, FMO}}^{(2)}(\vec{r}); s_{\text{FDL, FMO}}^{(2)}(\vec{r}) = S_{\text{FDL, FMO}} f_{\text{FDL, FMO}}^{(2)}(\vec{r})$                                                                                                     |
| Local electrophilicity     | S-LCD    | $\Delta \rho_{\text{FDL, FMO}}^+(\vec{r}) = \frac{\mu_{\text{FDL, FMO}}^- f_{\text{FDL, FMO}}^+(\vec{r})}{\eta_{\text{FDL, FMO}}} + \frac{1}{2} \left( \frac{\mu_{\text{FDL, FMO}}}{\eta_{\text{FDL, FMO}}} \right)^2 f_{\text{FDL, FMO}}^{(2)}(\vec{r})$ |
| Local nucleophilicity      | S-LCD    | $\Delta \rho_{\text{FDL, FMO}}^-(\vec{r}) = \frac{\mu_{\text{FDL, FMO}}^+ f_{\text{FDL, FMO}}^-(\vec{r})}{\eta_{\text{FDL, FMO}}} + \frac{1}{2} \left( \frac{\mu_{\text{FDL, FMO}}}{\eta_{\text{FDL, FMO}}} \right)^2 f_{\text{FDL, FMO}}^{(2)}(\vec{r})$ |

real space point). Indeed, organic chemistry is often explained in terms of reactive sites, which can be viewed as a coarse-grain representation of the molecule. One thus needs a “translation” of local functions into this usual language of reactive centers. This can be conveniently achieved by a procedure coined “condensation” [36]. In this paper, we will mainly consider atomic condensation, but the same approach can be used for whole substituents (several atoms grouped together) of interest. Several non-equivalent ways for partitioning a molecule into its atoms have been proposed over the last decades: Hirshfeld partition [55], fuzzy atoms [56], Parr’s atoms [57], etc.

Here, as already stated in [Introduction](#), we will rely mainly on QTAIM that divides real space into non-overlapping basins  $\Omega_A$ . For any local function  $a(\vec{r})$ , one can derive the corresponding atomic value by summing all contributions of points pertaining to the chosen atom:

$$a(A) = \int_{\Omega_A} a(\vec{r}) d^3r \quad (5)$$

For instance, in the FMO approximation, the atomic electrophilic Fukui function will be evaluated simply by:

$$f_{\text{FMO}}^+(A) = \int_{\Omega_A} \rho_{\text{LUMO}}(\vec{r}) d^3r \quad (6)$$

Within the FDL approach, the situation is a little bit trickier. Two main schemes actually exist [58]: the response of molecular fragment (RMF) and the fragment of molecular response (FMR) schemes. In the latter, the partition of the species with  $N$  electrons is also used for the condensation of the functions involving  $N+1$  or  $N-1$  electrons. In the first approach, the atomic partition is specifically recalculated from the electron densities of the oxidized or reduced molecules. One then gets:

$$\begin{aligned} f_{\text{FDL}}^+(A) &= \int_{\Omega_A(N+1)} \rho_{N+1}(\vec{r}) d^3r - \int_{\Omega_A(N)} \rho_N(\vec{r}) d^3r \\ &= q_N(A) - q_{N+1}(A) \end{aligned} \quad (7)$$

where  $q_N(A)$  denotes the atomic charge in the molecule with  $N$  electrons. This is the famous Yang-Mortier scheme [59]. It can be applied on any population analysis. Thus, in order to enrich the number of descriptors, we decided to use it (besides QTAIM) with the natural population analysis

(NPA) [60], Mulliken [61], Hirshfeld, and electrostatic-fitted ChelpG charges [62] for all local functions derived from Fukui functions and the dual descriptor. Note that these atomic values can also be combined, as for instance in the relative electrophilicity defined by the  $s^k(A) = s^+(A)/s^-(A)$  ratio [53] that we also computed.

Furthermore, atomic descriptors specific to QTAIM have been included, for instance the atomic localization index,  $LI(A)$ . Some others are energies, such as the atomic Kohn-Sham kinetic energy  $T_S(A)$ , the atomic Thomas-Fermi [ $T_{TF}(A)$ ] and von Weizsäcker kinetic energies  $T_W(A)$  (part of Liu's steric energy [63]), the atomic virial energy [ $T_W(A)$ ], and the interacting quantum atoms additive atomic energy,  $E_{IQA}(A)$ . We refer the interested reader to some of our recent publications for the corresponding formulas [21, 64, 65].

To be more comprehensive about atomic descriptors, it should be stressed that another method for condensing Fukui functions was proposed by Chamorro, Pérez and coworkers (CP) [66, 67], and was used in particular by Mayr and coworkers [25, 68]. The CP method is no longer based atomic charges, but on discriminating localized Gaussian primitives. This approach will also be considered in our paper.

Lastly, three descriptors recently proposed by Demir [69] were added: the electronic part of the molecular electrostatic potential evaluated at a given nucleus,  $h(\vec{R}_A)$ , and the two finite difference combinations ( $\Delta^+ h(\vec{R}_A) = h_{N+1}(\vec{R}_A) - h_N(\vec{R}_A)$  and  $\Delta^- h(\vec{R}_A) = h_N(\vec{R}_A) - h_{N-1}(\vec{R}_A)$ ).

## Computational details

The molecular structures and the values for the different chemical reactivity descriptors presented in the previous section were obtained using the following steps. First, the structures of the most stable conformers were generated with the Balloon [70] software (for molecules 9 and 11, using RMSD filtering) or by relaxed energy scans along relevant coordinates (for instance, dihedral angles for rotations around bonds), which were subsequently optimized with the Gaussian 09 [71] software at the M06-2X/aug-cc-pVTZ level of theory using the recommended 'ultrafine' integration grid. Note that, for some species with  $N+1$  electrons, the quadratic QC algorithm has to be used to properly converge the self-consistent field (SCF) energy (leading to correct positive electron affinities). For analogous reasons, the FDL calculation of the hyperhardness was not considered (since it would involve the  $N+2$  electron species).

Solvent effects (the selected experimental values were actually measured in dimethylsulfoxide (DMSO), a non-coordinating solvent) were taken into account using a polarizable continuum model [72], in the latest implementation of the integral equation formalism (IEFPCM). QTAIM calculations were performed using the AIMALL software [73] (IQA energies are obtained using the M06-2X tag model), while CDFT descriptors were obtained from homemade scripts. Condensation according to the Chamorro-Pérez scheme was performed with a procedure described by these authors and with a program available online (<https://github.com/dmsteglenko/Fukui-function-calculation>).

Atomic units are used for all descriptors, except for  $\mu$ ,  $\eta$ ,  $\omega$ ,  $\Delta E^{\text{nucleo/electro}}$  for which electron-volts are more commonly used (see also Table S2 in the supplementary file for the units of all descriptors).

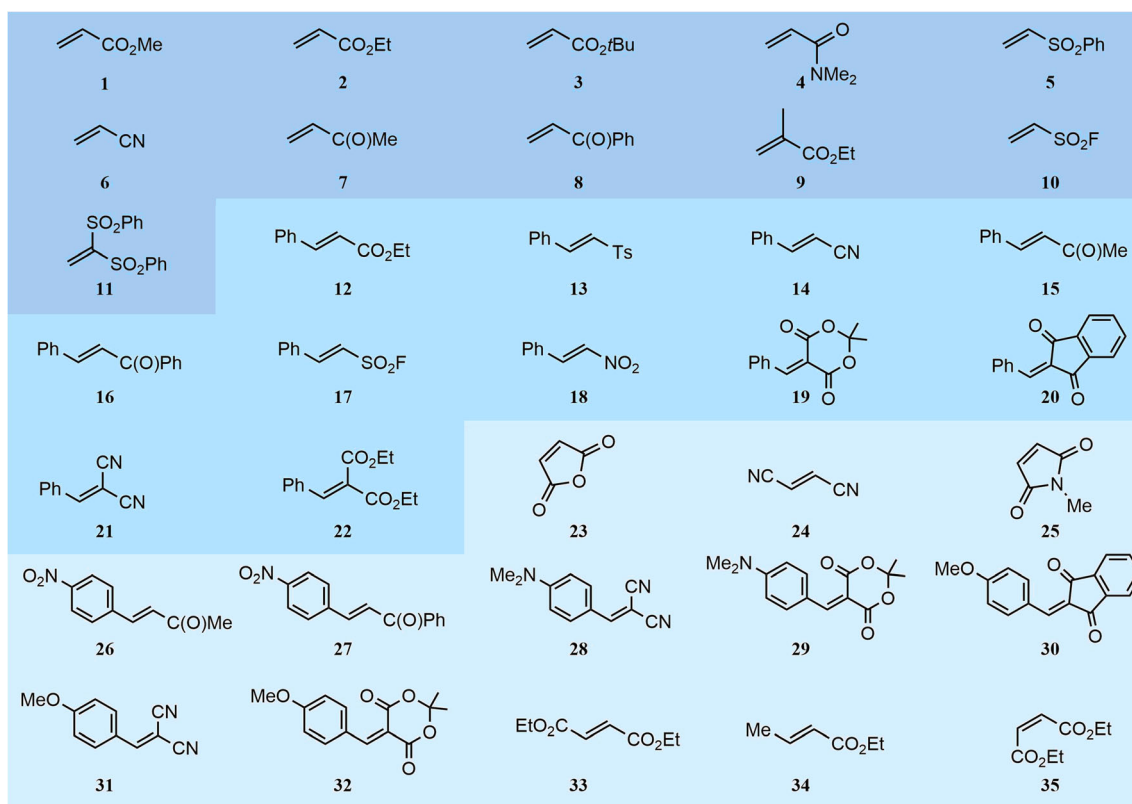
## Results and discussion

Our dataset consists of the 35 molecules represented in Scheme 2. They can be classified into three groups: mono-acceptor-substituted ethylenes (11 molecules, in dark blue, numbered 1–11), phenyl-substituted Michael acceptors (11 molecules, in bright blue, numbered from 12–22), and the so-called "other Michael acceptors" (13 molecules, in violet, numbered from 23–25).

The first step of our study consisted of a conformational analysis. Table S1 in the supplementary information file shows selected conformers and their relative energy (for instance, four minima were identified for compounds 1, 2, and 3). In all cases, the most stable conformer was found to be the same in terms of SCF and standard Gibbs energies, and also corresponded to that reported by the research group of Mayr, who used a slightly different computational protocol (with only one exception, compound 8, but the energy difference between the two minima is much too low to be significant).

Then, for each molecule, 163 descriptors were evaluated as described in the preceding sections, and whose values are gathered in Table S2 in the supplementary information file. One may wonder whether some are redundant. In particular, it is of methodological interest to investigate if FDL and FMO descriptors are correlated. Figure 1 illustrates the discrepancies between both approaches for four selected descriptors (two atomic descriptors relative to the  $C_\beta$  carbon atom, condensed within QTAIM, and two global descriptors):  $\omega$ ,  $\Delta E^{\text{nucleo}}$ ,  $f^+$ , and  $f^{(2)}$ .

Interestingly, a high correlation between FDL and FMO was observed for the electrophilicity index [with the coefficient of determination ( $R^2$ ) equal to 0.94], which fully vanished for  $\Delta E^{\text{nucleo}}$ , while both descriptors were based on the same ingredients ( $\mu$  and  $\eta$ ). This well illustrates to what



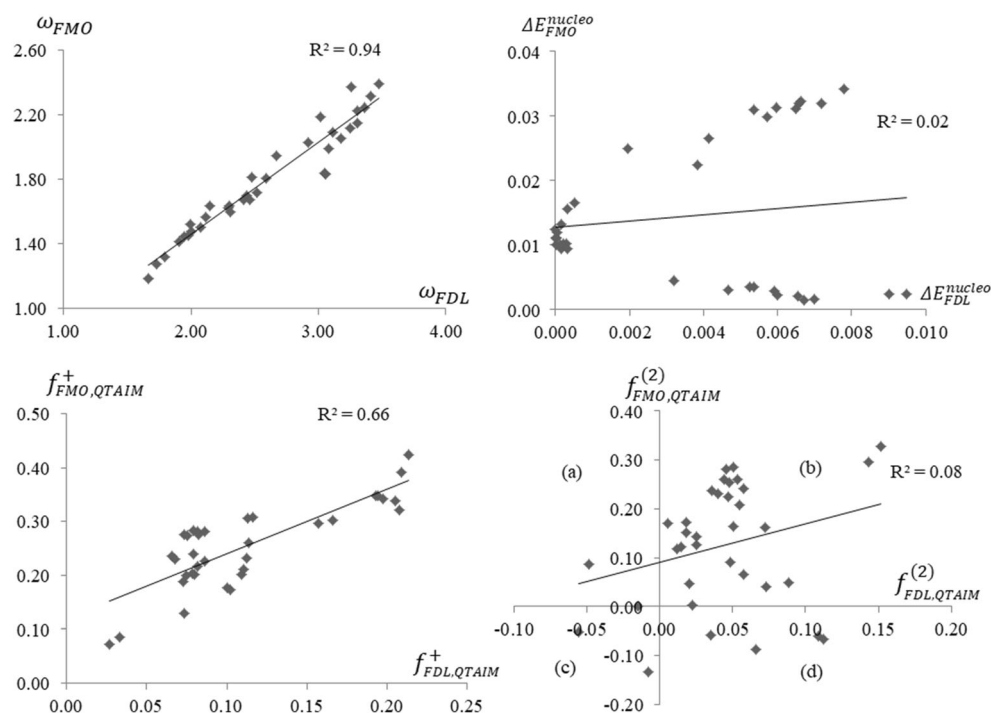
**Scheme 2** The studied dataset

extent error compensations may be or not be at work for CDs, with no clear rule to account for them.

For the atomic electrophilic Fukui function  $f^+$  (which measures the amount of electrons an atom can gain when one

excess electron is spread over the whole molecule), the  $R^2$  value remains quite low (0.66) even if both approaches afford a similar qualitative trend within the whole molecular dataset. The case of the dual descriptor is also instructive. Indeed, let

**Fig. 1** Comparison between the finite difference linearization and the frontier molecular orbital (FMO) schemes for the whole set on selected descriptors at the M06-2X/aug-cc-pVTZ PCM level of theory.  $\omega_{FMO}$  and  $\Delta E_{FMO}^{nucleo}$  in eV, other descriptors in atomic units



us recall that it allows for quantifying the competition between electrophilicity and nucleophilicity: when  $f^{(2)}(A)$  is positive, electrophilicity dominates at the atomic scale, while nucleophilicity is predominant for negative values.

The quadrants (b) and (c) in the bottom right panel in Fig. 1 collect points for which the condensed dual descriptor features the same sign for the FDL and FMO treatments. On the other hand, points in quadrants (a) and (d) correspond to different signs depending on the computational protocol. The fact that there is a non-negligible number of compounds in these areas indicates that the two approaches predict opposite reactivity type.

Such an analysis could also be led to unravel the influence of the population scheme chosen in the Yang-Mortier procedure. The results are pictured in Fig. 2 for the condensed dual descriptor. It immediately appears that the five population schemes give results that are not correlated. The highest  $R^2$  value between the QTAIM quantities and the others was obtained with the Hirshfeld partition, but remained low (0.50).

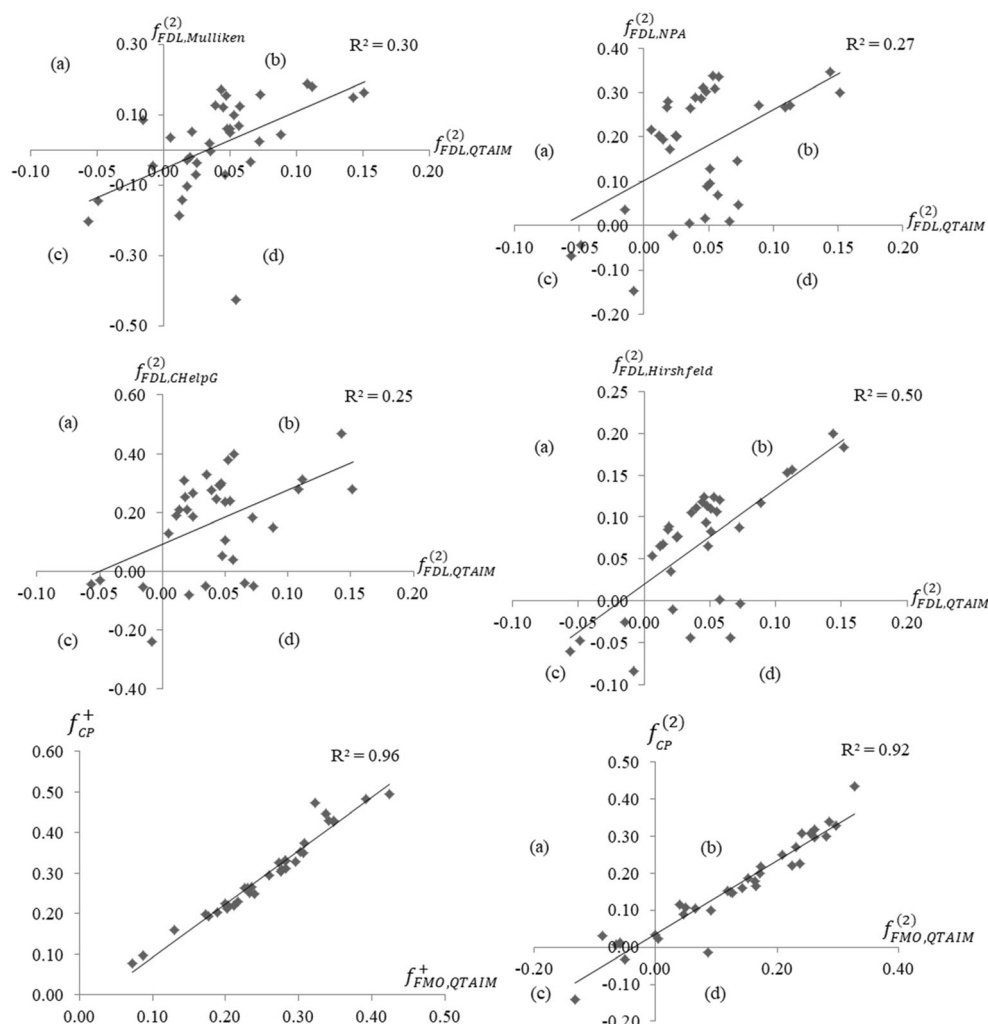
However, one should underline that (with the exception of the Mulliken case, which displays many negative values) most

of the points are located in quadrant (b), meaning that the  $C_\beta$  carbon is predicted to be electrophilic. Henceforth, even if several Yang-Mortier schemes consistently agree from a qualitative point of view, they lead to significantly different quantitative estimates. Conversely, the good agreement (see bottom of Fig. 2) between the QTAIM values (in the FMO approach) and those obtained from the Chamorro-Pérez scheme for  $f^+$  and  $f^{(2)}$  should be mentioned.

Such a comparison can also be led for atomic charges (see Fig. S1 in the supplementary information file): it appears that, apparently counterintuitively, ChelpG and Mulliken predict negative atomic charge for the electrophilic carbon atom  $C_\beta$  in 94% of the cases, while the situation is much more balanced for Hirshfeld and QTAIM. The quadrant analysis shows that these both last approaches are consistent with one another about the charge sign, while considerable discrepancies are observed with the other schemes, so that extrapolating reactivity behavior from these atomic charges may remain elusive.

One final comment about methodology is related to the influence of the level of theory. Figure S2 in the supplementary information compares our FMO electrophilicity

**Fig. 2** Four top panels Influence of the population analysis in the Yang-Mortier scheme for the condensation of the dual descriptor on carbon  $\beta$ . Two bottom panels Comparison between QTAIM Yang-Mortier and Chamorro-Pérez (CP) schemes. M06-2X/aug-cc-pVTZ PCM level of theory. All values in atomic units



indices with those reported by Mayr and coworkers at the B3LYP/6–31 + G (d) level of theory [25] with SMD solvation model [74]. High correlation was found ( $R^2 = 0.95$ ), suggesting good transferability of these descriptors with respect to the computational protocol that is used to generate the wavefunction or electron density.

Now, it follows from our previous findings that all selected descriptors will be actually considered as generally independent the one from the other, so that none of them can be excluded a priori. We then investigated monolinear regressions between each of them and Mayr's experimental values for electrophilicity. Unfortunately, no high correlations were obtained, since the higher  $R^2$  value we obtained was with  $\mu^+ f^+$ , amounting to 0.64, as depicted in Fig. 3, according to:

$$E_{\text{Mayr}} = -768.6 \mu_{\text{FMO}}^+ f_{\text{FMO,QTAIM}}^+(C_\beta) - 26.00 \quad (8)$$

Disappointingly, the electrophilicity index, which was previously recommended in the literature [23], only gave  $R^2 = 0.40$  (see Fig. S3 in the supplementary information file).

These results imply that a single descriptor may not be sufficient to describe the electrophilicity of the whole set of these Michael acceptors. One may ascribe this failure to the fact that one needs to take into account not only the reactivity of the carbon  $\beta$ , but the reactivity of the whole carbon–carbon double bond since carbon  $\alpha$  has also a role to play. This can be achieved by considering a combination of the  $(f_\alpha^+ + f_\beta^+)$

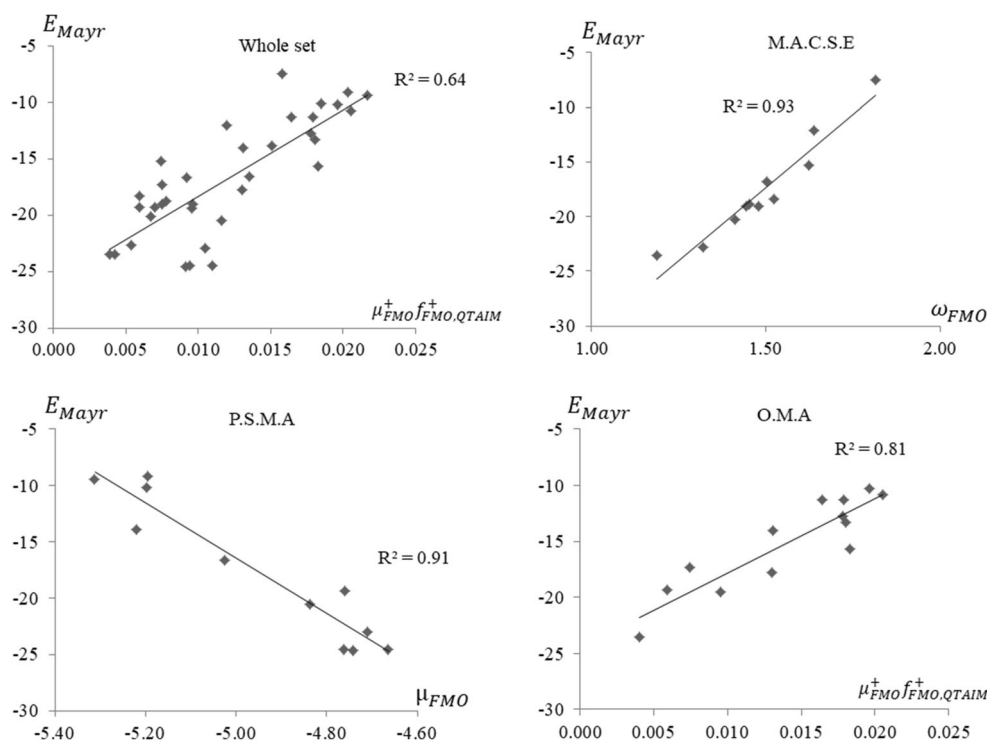
type. This leads to a descriptor that will be denoted  $f^+(C_\alpha C_\beta)$ , with its natural grand-canonical (and composite) extensions  $Sf^+(C_\alpha C_\beta)$ ,  $\mu^+ f^+(C_\alpha C_\beta)$ ,  $\omega f^+(C_\alpha C_\beta)$ ...

From a QTAIM point of view, the C=C bond can be basically characterized by the delocalization index,  $DI(C_\alpha C_\beta)$ , between carbons  $\alpha$  and  $\beta$  (which can be considered as the QTAIM  $C_\alpha$ – $C_\beta$  bond order), the electron density  $\rho_C$  and the electron density laplacian values  $\nabla^2 \rho_C$  at the corresponding bond critical point (BCP) that roughly represent bond strength. Unfortunately, none of these  $C_\alpha$ – $C_\beta$  QTAIM bond descriptors was more successful [higher  $R^2 = 0.21$ , with  $DI(C_\alpha C_\beta)$ ].

A more general strategy is then to consider bilinear regressions. They could in fact provide a more thorough description of reactivity. For instance, one descriptor will be linked to orbital control, while another will be related to charge control. In such cases, one will obtain a combination reminiscent of Ayers' general purpose reactivity descriptor [75, 76]. As we considered 163 descriptors, 13,203 pairs should be considered. Nevertheless, we decided to reduce this number by grouping only descriptors calculated using the same approach (FDL or FMO) or the same population analysis for the sake of consistency. Indeed, we believe that mixing (for instance) a NPA descriptor with a QTAIM one would generate models that could not be interpreted easily in a chemical sense. For such reasons, we considered only 7732 models over the 13,203 possible ones.

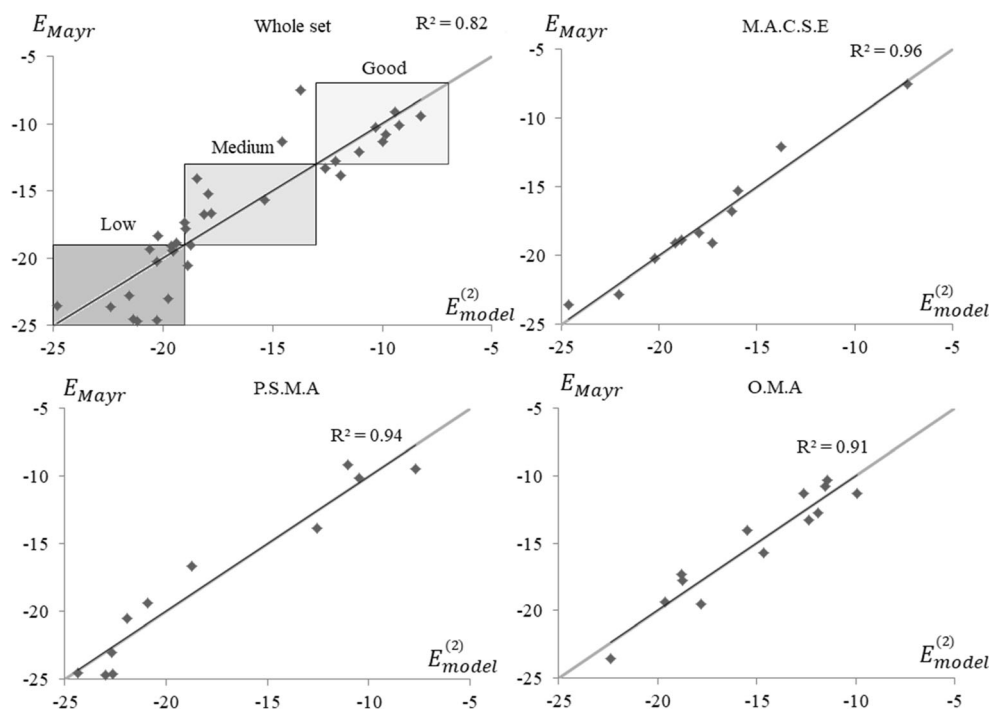
The best suitable bilinear model (see Fig. 4) that emerged consisted in the combination of  $\omega f_\beta^+$  and  $f^+(C_\alpha C_\beta)$  with  $R^2 =$

**Fig. 3** Best monolinear regressions for Mayr's electrophilicity  $E_{\text{Mayr}}$  (corresponding to highest  $R^2$  values) and regression lines for the whole set (top left), mono acceptor substituted ethylenes (M.A.C.S.E.; top right), phenyl substituted acceptors (P.S.M.A.; bottom left), and others (O.M.A.; bottom right) subgroups





**Fig. 4** Best bilinear models for Mayr's electrophilicity  $E_{\text{Mayr}}$  for the whole set (top left), the M.A.C.S.E. (top right), the P.S.M.A. (bottom left), and others O.M.A. (bottom right) subgroups

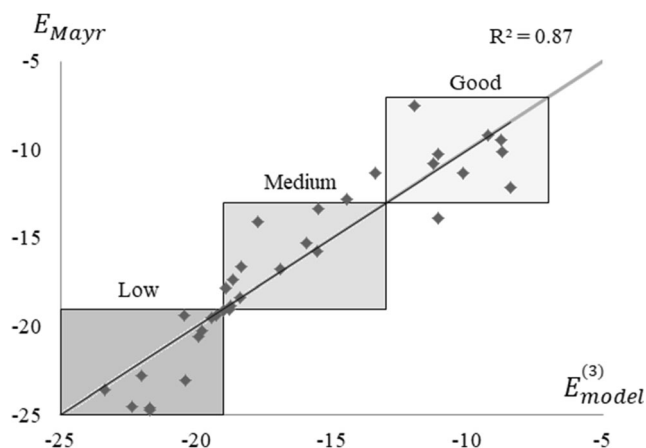


0.82, according to the following equation:

$$E_{\text{Mayr}} = 52.12 \omega_{\text{FMO}} f_{\text{FMO,QTAIM}}^+(C_\beta) - 44.35 f_{\text{FMO,QTAIM}}^+(C_\alpha C_\beta) - 21.96 \quad (9)$$

$$E_{\text{Mayr}} = 65.37 \omega_{\text{FMO}} f_{\text{FMO,QTAIM}}^+(C_\beta) - 59.94 f_{\text{FMO,QTAIM}}^+(C_\alpha C_\beta) + 16.54 DI(C_\alpha C_\beta) - 49.74 \quad (10)$$

Even if substantial improvement was obtained from this bilinear model [the corresponding mean absolute error (MAE) equals 1.6], it fails in being very accurate. Keeping the two descriptors entering Eq. 9, we then performed a trilinear analysis (see Fig. 5). The double bond delocalization index appears to be the more efficient partner, according to Eq. 10 (with  $R^2 = 0.87$ , see Fig. 3):



**Fig. 5** Best trilinear model for Mayr's electrophilicity  $E_{\text{Mayr}}$  for the whole set

The MAE value was found to be equal to 1.3. This might seem still too much high. An alternative approach is to consider regressions from subsets, as already proposed by Domingo [23]. With this aim, we performed independent monolinear regressions for the mono acceptor substituted ethylenes (11 molecules, M.A.C.S.E. family), the phenyl substituted Michael acceptors (11 molecules, P.S.M.A.), and the “other Michael acceptors” (13 molecules, O.M.A.). The best monolinear models (see Fig. 3) were the following:

$$E_{\text{Mayr}} = 26.80 \omega_{\text{FMO}} - 57.53 \text{ (compounds 1–11)} \quad (11a)$$

$$E_{\text{Mayr}} = -24.52 \mu_{\text{FMO}} - 139.23 \text{ (compounds 12–22)} \quad (11b)$$

$$E_{\text{Mayr}} = -662.7 \mu_{\text{FMO}}^+ f_{\text{FMO,QTAIM}}^+(C_\beta) - 24.42 \text{ (compounds 23–35)} \quad (11c)$$

The high  $R^2$  values for the two first groups should be noted: 0.93 for the first group and 0.91 for the second, while it is significantly lower for the third group (0.81). This is in fact quite expected since the “other” family is a heterocycle set, made of very different molecules. Bilinear regressions were also considered (see Fig. 4), with  $R^2 = 0.96$  for the first family,

0.94 for the second, and  $R^2 = 0.91$  for the third:

$$E_{\text{Mayr}} = -(1447\mu_{\text{FMO}}^+ + 212.1 S_{\text{FMO}})f_{\text{FMO,QTAIM}}^+(C_\beta) \quad (12a)$$

−21.91 (compounds 1–11)

$$E_{\text{Mayr}} = -(2358\mu_{\text{FMO}}^+ - 904.2 S_{\text{FMO}})f_{\text{FMO,QTAIM}}^+(C_\beta) \quad (12b)$$

−19.55 (compounds 12–22)

$$E_{\text{Mayr}} = -5.91\mu_{\text{FMO}} + 1264 S_{\text{FMO}}^2 f_{\text{FMO,CP}}^{(2)}(C_\beta) \quad (12c)$$

−49.98 (compounds 13–35)

On the other hand, one can wonder whether for practical applications one needs to predict an accurate value or, instead, to be able to classify the reactants. We thus decided to create a more qualitative typology for Michael acceptors. More specifically, as experimental electrophilicity values range from −7.50 to −25, we divided it into three subclasses of reactivity: C1 = good electrophiles (−7 to −13), C2 = medium (from −13 to −19), and C3 = low (from −19 to −25).

In Figs. 4 and 5, rectangles were drawn that correspond to compounds for which the statistical models and the experimental values predict the same reactivity category. Conversely, points outside these rectangles are for molecules for which the subclass predicted by the model differs from that from the experimental value. One can see that almost all points with the trilinear model belong to these three rectangles, enlightening its potential interest for applied chemists.

## Conclusions

In this paper, we investigated the ability of descriptors issued from CDFT and the QTAIM to predict Mayr's experimental electrophilicity values for a series of 35 Michael acceptors. Various calculation schemes (finite difference linearization, FMOs) and condensation techniques (Yang-Mortier applied on different population analyses, CP approach) were first compared and shown to be in general uncorrelated.

Monolinear regressions were then performed, which proved satisfying for the mono acceptor substituted ethylenes and the phenyl substituted subsets, but inadequate for the whole dataset. Bilinear and trilinear regressions afforded consequent improvement, involving mainly the global electrophilicity index  $\omega$  and the atomic  $\mu^+ f^+$  generalized electrophilic Fukui function (which are obviously chemically meaningful), but they fail in being truly predictive for the whole set.

In summary, this study thus proves that these CDFT-QTAIM descriptors can be efficient to quantify the reactivity difference for similar compounds, and to retrieve qualitative ranking for different families. They also allow for establishing a new typology of “Michael acceptance power” that could be finally useful for experimentalists.

**Acknowledgments** The authors would like to gratefully acknowledge the LABEX SynOrg and the Normandy Region for funding and support, and the Centre Régional Informatique et d'Applications Numériques de Normandie (CRIANN) high-performance computing facility. It is our pleasure to contribute to this special issue honoring the fundamental contributions made in chemistry by Prof. Chattaraj, a pioneer in conceptual density functional theory.

## References

- Gardner Swain C, Scott CB (1953) Quantitative correlation of relative rates. Comparison of hydroxide ion with other nucleophilic reagents toward alkyl halides, esters, epoxides and acyl halides. *J Am Chem Soc* 75:141–147. <https://doi.org/10.1021/ja01097a041>
- Ritchie D (1986) Cation-anion combination reactions. 26. A review. *Can J Chem* 64:2239–2250. <https://doi.org/10.1139/v86-370>
- Heo CKM, Bunting JW (1994) Nucleophilicity towards a vinylic carbon atom: Rate constants for the addition of amines to the 1-methyl-4-vinylpyridinium cation in aqueous solution. *J Chem Soc Perkin Trans 2*:2279–2290
- Mayr H, Patz M (1994) Scales of nucleophilicity and electrophilicity: A system for ordering polar organic and organometallic reactions. *Angew Chem Int Ed Eng* 33:938–957. <https://doi.org/10.1002/anie.199409381>
- Lucius R, Loos R, Mayr H (2002) Kinetic studies of carbocation–carbanion combinations: Key to a general concept of polar organic reactivity. *Angew Chem Int Ed* 41:91–95. [https://doi.org/10.1002/1521-3773\(20020104\)41:1<91::AID-ANIE91>3.0.CO;2-P](https://doi.org/10.1002/1521-3773(20020104)41:1<91::AID-ANIE91>3.0.CO;2-P)
- Mayr H, Bug T, Gotta MF et al (2001) Reference scales for the characterization of cationic electrophiles and neutral nucleophiles. *J Am Chem Soc* 123:9500–9512. <https://doi.org/10.1021/ja010890y>
- <http://www.cup.lmu.de/oc/mayr/reaktionsdatenbank/>. Mayr's Database of Reactivity Parameters
- Chermette H (1999) Chemical reactivity indexes in density functional theory. *J Comput Chem* 20:129–154. [https://doi.org/10.1002/\(SICI\)1096-987X\(19990115\)20:1<129::AID-JCC13>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1096-987X(19990115)20:1<129::AID-JCC13>3.0.CO;2-A)
- Geerlings P, De Proft F, Langenaeker W (2003) Conceptual density functional theory. *Chem Rev* 103:1793–1874. <https://doi.org/10.1021/cr990029p>
- Pereira F, Latino DARS, Aires-De-Sousa J (2011) Estimation of mayr electrophilicity with a quantitative structure-property relationship approach using empirical and DFT descriptors. *J Organomet Chem* 76:9312–9319. <https://doi.org/10.1021/jo201562f>
- González MM, Cárdenas C, Rodríguez JI et al (2018) Quantitative electrophilicity measures. *Wuli Huaxue Xuebao/Acta Phys - Chim Sin* 34:662–674. <https://doi.org/10.3866/PKU.WHXB201711021>
- Pérez P, Toro-Labbé A, Aizman A, Contreras R (2002) Comparison between experimental and theoretical scales of electrophilicity in benzhydryl cations. *J Organomet Chem* 67:4747–4752. <https://doi.org/10.1021/jo020255q>
- Tognetti V, Morell C, Joubert L (2015) Atomic electronegativities in molecules. *Chem Phys Lett* 635:111–115. <https://doi.org/10.1016/j.cplett.2015.05.057>

14. Liu S, Rong C, Lu T (2014) Information conservation principle determines electrophilicity, nucleophilicity, and regioselectivity. *J Phys Chem A* 118:3698–3704. <https://doi.org/10.1021/jp5032702>
15. Mayr H, Ofial AR (2015) A quantitative approach to polar organic reactivity. *SAR QSAR Environ Res* 26:619–646. <https://doi.org/10.1080/1062936X.2015.1078409>
16. Zhuo LG, Liao W, Yu ZX (2012) A frontier molecular orbital theory approach to understanding the Mayr equation and to quantifying nucleophilicity and electrophilicity by using HOMO and LUMO energies. *Asian J Org Chem* 1:336–334. <https://doi.org/10.1002/ajoc.201200103>
17. Wang C, Fu Y, Guo QX, Liu L (2010) First-principles prediction of nucleophilicity parameters for  $\pi$  nucleophiles: Implications for mechanistic origin of Mayr's equation. *Chem A Eur J* 16:2586–2598. <https://doi.org/10.1002/chem.200902484>
18. Chamorro E, Melin J (2015) On the intrinsic reactivity index for electrophilicity/nucleophilicity responses. *J Mol Model* 21:53. <https://doi.org/10.1007/s00894-015-2608-2>
19. Kiyooka SI, Kaneno D, Fujiyama R (2013) Intrinsic reactivity index as a single scale directed toward both electrophilicity and nucleophilicity using frontier molecular orbitals. *Tetrahedron* 69:4247–4258. <https://doi.org/10.1016/j.tet.2013.03.083>
20. Tognetti V, Morell C, Joubert L (2015) Quantifying electro/nucleophilicity by partitioning the dual descriptor. *J Comput Chem* 36:649–659. <https://doi.org/10.1002/jcc.23840>
21. Tognetti V, Joubert L (2017) On atoms-in-molecules energies from Kohn–Sham calculations. *ChemPhysChem* 18:2675–2687. <https://doi.org/10.1002/cphc.201700637>
22. Dau-Schmidt J-P, Mayr H (1994) Relative reactivities of alkyl chlorides under Friedel–Crafts conditions. *Chem Ber* 127:205–212. <https://doi.org/10.1002/cber.19941270129>
23. Domingo LR, Pérez P, Contreras R (2004) Reactivity of the carbon-carbon double bond towards nucleophilic additions. A DFT analysis. *Tetrahedron* 60:6585–6591. <https://doi.org/10.1016/j.tet.2004.06.003>
24. Carey FA, Sundberg RJ (2007) Advanced organic chemistry Part A: structure and mechanisms. Springer, New York
25. Allgäuer DS, Jangra H, Asahara H et al (2017) Quantification and theoretical analysis of the electrophilicities of Michael acceptors. *J Am Chem Soc* 139:13318–13329. <https://doi.org/10.1021/jacs.7b05106>
26. Bader RFW (1990) Atoms in molecules: A quantum theory. Oxford University Press, Oxford
27. Popelier PLA (2000) Atoms in molecules: An Introduction. Pearson Education, Harlow
28. Aliyenne A, Pin F, Nimbarte VD et al (2016) Bi (OTf)<sub>3</sub>-catalysed access to 2,3-substituted isindolinones and tricyclic N,O-acetals by trapping of bis-N-acyliminium species in a tandem process. *Eur J Org Chem* 2016:3592–3602. <https://doi.org/10.1002/ejoc.201600530>
29. Tognetti V, Boulangé A, Peixoto PA et al (2014) A theoretical study on diastereoselective oxidative dearomatization by iodoxybenzoic acid. *J Mol Model* 20:1–10. <https://doi.org/10.1007/s00894-014-2342-1>
30. Falkowska E, Laurent MY, Tognetti V et al (2015) Synthesis of SF<sub>5</sub>-substituted isoxazolidines using 1,3-dipolar cycloaddition reactions of nitrones with pentafluorosulfanyl acrylic esters and amides. *Tetrahedron* 71:8067–8076. <https://doi.org/10.1016/j.tet.2015.08.050>
31. Tognetti V, Guégan F, Luneau D et al (2017) Structural effects in octahedral carbonyl complexes: an atoms-in-molecules study. *Theor Chem Accounts* 136:1–12. <https://doi.org/10.1007/s00214-017-2116-9>
32. Inostroza-Rivera R, Yahia-Ouahmed M, Tognetti V et al (2015) Atomic decomposition of conceptual DFT descriptors: application to proton transfer reactions. *Phys Chem Chem Phys* 17:17797–17808. <https://doi.org/10.1039/C5CP01515D>
33. Guégan F, Tognetti V, Joubert L et al (2016) Towards the first theoretical scale of the trans effect in octahedral complexes. *Phys Chem Chem Phys* 18:982–990. <https://doi.org/10.1039/C5CP04982B>
34. Tognetti V, Bouzbou S, Joubert L (2017) A theoretical study of the diastereoselective allylation of aldehydes with new chiral allylsilanes. *J Mol Model* 23:5. <https://doi.org/10.1007/s00894-016-3173-z>
35. Jouanno LA, Di Mascio V, Tognetti V et al (2014) Metal-free decarboxylative Hetero-Diels-Alder synthesis of 3-hydroxypyridines: A rapid access to N-fused bicyclic hydroxypiperidine scaffolds. *J Organomet Chem* 79:1303–1319. <https://doi.org/10.1021/jo402729a>
36. Zielinski F, Tognetti V, Joubert L (2013) A theoretical study on the gas-phase protonation of pyridine and phosphinine derivatives. *J Mol Model* 19:4049–4058. <https://doi.org/10.1007/s00894-013-1925-6>
37. Falkowska E, Tognetti V, Joubert L et al (2015) First efficient synthesis of SF<sub>5</sub>-substituted pyrrolidines using 1,3-dipolar cycloaddition of azomethine ylides with pentafluorosulfanyl-substituted acrylic esters and amides. *RSC Adv* 5:6864–6868. <https://doi.org/10.1039/C4RA14075C>
38. Chattaraj PK, Lee H, Parr RG (1991) HSAB principle. *J Am Chem Soc* 113:1855–1856. <https://doi.org/10.1021/ja00005a073>
39. Chattaraj PK (2009) Chemical reactivity theory. CRC, Boca Raton
40. Parr RG, Donnelly RA, Levy M, Palke WE (1978) Electronegativity: The density functional viewpoint. *J Chem Phys* 68:3801–3807. <https://doi.org/10.1063/1.436185>
41. Parr RG, Pearson RG (1983) Absolute hardness: Companion parameter to absolute electronegativity. *J Am Chem Soc* 105:7512–7516. <https://doi.org/10.1021/ja00364a005>
42. Geerlings P, De Proft F (2008) Conceptual DFT: the chemical relevance of higher response functions. *Phys Chem Chem Phys* 10:3028. <https://doi.org/10.1039/b717671f>
43. Morell C, Grand A, Toro-Labbé A, Chermette H (2013) Is hyperhardness more chemically relevant than expected? *J Mol Model* 19:2893–2900. <https://doi.org/10.1007/s00894-013-1778-z>
44. Parr RG, Yang W (1984) Density functional approach to the frontier-electron theory of chemical reactivity. *J Am Chem Soc* 106:4049–4050. <https://doi.org/10.1021/ja00326a036>
45. Morell C, Grand A, Toro-Labbé A (2005) New dual descriptor for chemical reactivity. *J Phys Chem A* 109:205–212. <https://doi.org/10.1021/jp046577a>
46. Perdew JP, Parr RG, Levy M, Balduz JL (1982) Density-functional theory for fractional particle number: derivative discontinuities of the energy. *Phys Rev Lett* 49:1691–1694. <https://doi.org/10.1103/PhysRevLett.49.1691>
47. Parr RG, Szentpály LV, Liu S (1999) Electrophilicity index. *J Am Chem Soc* 121:1922–1924. <https://doi.org/10.1021/ja983494x>
48. Chattaraj PK, Sarkar U, Roy DR (2006) Electrophilicity Index. *Chem Rev* 106:2065–2091. <https://doi.org/10.1021/cr040109f>
49. Padmanabhan J, Parthasarathi R, Elango M et al (2007) Multiphilic descriptor for chemical reactivity and selectivity. *J Phys Chem A* 111:9130–9138. <https://doi.org/10.1021/jp0718909>
50. Chattaraj PK, Maiti B, Sarkar U (2003) Philicity: A unified treatment of chemical reactivity and selectivity. *J Phys Chem A* 107:4973–4975. <https://doi.org/10.1021/jp034707u>
51. Parthasarathi R, Padmanabhan J, Elango M et al (2004) Intermolecular reactivity through the generalized philicity concept. *Chem Phys Lett* 394:225–230. <https://doi.org/10.1016/j.cplett.2004.07.002>
52. Yang W, Parr RG (1985) Hardness, softness, and the fukui function in the electronic theory of metals and catalysis. *Proc Natl Acad Sci USA* 82:6723–6726. <https://doi.org/10.1073/pnas.82.20.6723>



53. Roy RK, Krishnamurti S, Geerlings P, Pal S (1998) Local softness and hardness based reactivity descriptors for predicting intra- and intermolecular reactivity sequences: Carbonyl compounds. *J Phys Chem A* 102:3746–3755. <https://doi.org/10.1021/jp973450v>
54. Morell C, Gázquez JL, Vela A et al (2014) Revisiting electroaccepting and electrodonating powers: proposals for local electrophilicity and local nucleophilicity descriptors. *Phys Chem Chem Phys* 16:26832–26842. <https://doi.org/10.1039/C4CP03167A>
55. Hirshfeld FL (1977) Bonded-atom fragments for describing molecular charge densities. *Theor Chim Acta* 44:129–138. <https://doi.org/10.1007/BF00549096>
56. Becke AD (1988) A multicenter numerical integration scheme for polyatomic molecules. *J Chem Phys* 88:2547–2553. <https://doi.org/10.1063/1.454033>
57. Li L, Parr RG (1986) The atom in a molecule: A density matrix approach. *J Chem Phys* 84:1704–1711. <https://doi.org/10.1063/1.450468>
58. Bultinck P, Van Alsenoy C, Ayers PW, Carbó-Dorca R (2007) Critical analysis and extension of the Hirshfeld atoms in molecules. *J Chem Phys* 126:144101. <https://doi.org/10.1063/1.2715563>
59. Yang W, Mortier WJ (1986) The use of global and local molecular parameters for the analysis of the gas-phase basicity of amines. *J Am Chem Soc* 108:5708–5711. <https://doi.org/10.1021/ja00279a008>
60. Reed AE, Weinstock RB, Weinhold F (1985) Natural population analysis. *J Chem Phys* 83:735–746. <https://doi.org/10.1063/1.449486>
61. Mulliken RS (1955) Electronic population analysis on LCAO-MO molecular wave functions. I *J Chem Phys* 23:1833–1840. <https://doi.org/10.1063/1.1740588>
62. Breneman CM, Wiberg KB (1990) Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J Comput Chem* 11:361–373. <https://doi.org/10.1002/jcc.540110311>
63. Liu S (2007) Steric effect: A quantitative description from density functional theory. *J Chem Phys* 126:244103. <https://doi.org/10.1063/1.2747247>
64. Tognetti V, Joubert L (2014) Density functional theory and Bader's atoms-in-molecules theory: Towards a vivid dialogue. *Phys Chem Chem Phys* 16:14539–14550. <https://doi.org/10.1039/c3cp55526g>
65. Patrikeev L, Joubert L, Tognetti V (2016) Atomic decomposition of Kohn-Sham molecular energies: The kinetic energy component. *Mol Phys* 114:1285–1296. <https://doi.org/10.1080/00268976.2015.1113314>
66. Chamorro E, Perez P (2005) Condensed-to-atoms electronic Fukui functions within the framework of spin polarized density-functional theory. *J Chem Phys* 123:1–10. <https://doi.org/10.1063/1.2033689>
67. Contreras RR, Fuentealba P, Galván M, Pérez P (1999) A direct evaluation of regional Fukui functions in molecules. *Chem Phys Lett* 304:405–413. [https://doi.org/10.1016/S0009-2614\(99\)00325-5](https://doi.org/10.1016/S0009-2614(99)00325-5)
68. Li Z, Jangra H, Chen Q et al (2018) Kinetics and mechanism of oxirane formation by Darzens condensation of ketones: Quantification of the electrophilicities of ketones. *J Am Chem Soc* 140:8b01657. <https://doi.org/10.1021/jacs.8b01657>
69. Demir S, Engin MS, Çay S et al (2017) DFT and conceptual-DFT assessment on selective tertiary amine functionalized calix[4]arene-anion interaction. *Comput Theor Chem* 1117:292–298. <https://doi.org/10.1016/j.comptc.2017.08.008>
70. Vainio MJ, Johnson MS (2007) Generating conformer ensembles using a multiobjective genetic algorithm. *J Chem Inf Model* 47:2462–2474. <https://doi.org/10.1021/Ci6005646>
71. Frisch MJ, Trucks GW, Schlegel HB et al Gaussian 09 Revision D.01. Gaussian 09 Revis. D.01. Gaussian, Inc., Wallingford
72. Tomasi J, Mennucci B, Cammi R (2005) Quantum mechanical continuum solvation models. *Chem Rev* 105:2999–3093. <https://doi.org/10.1021/cr9904009>
73. Keith TA (2017) AIMAll (Version 17.11.14). TK Gristmill Software, Overland Park. [aim.tkgristmill.com](http://aim.tkgristmill.com)
74. Marenich AV, Cramer CJ, Truhlar DG (2009) Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J Phys Chem B* 113:6378–6396. <https://doi.org/10.1021/jp810292n>
75. Anderson JSM, Melin J, Ayers PW (2007) Conceptual density-functional theory for general chemical reactions, including those that are neither charge- nor frontier-orbital-controlled. 1. Theory and derivation of a general-purpose reactivity indicator. *J Chem Theory Comput* 3:358–374. <https://doi.org/10.1021/ct600164j>
76. Anderson JSM, Ayers PW (2007) Predicting the reactivity of ambidentate nucleophiles and electrophiles using a single, general-purpose, reactivity indicator. *Phys Chem Chem Phys* 9:2371–2378. <https://doi.org/10.1039/b700960g>

### 1.3 Conclusion et perspectives

L'un des principaux résultats obtenus durant cette étude est la démonstration de la faible capacité de prédiction par le biais d'un seul descripteur quantique de réactivité chimique, sur un ensemble d'accepteurs de Michael de structures variées ( $R^2 = 0,64$  avec un descripteur sur les 35 molécules, voir la **Figure 3** de l'article).

En revanche, si l'on se concentre sur une *sous-famille* d'accepteurs de Michael (de structures relativement proches), nous avons confirmé que les descripteurs d'électrophilie issus de la CDFT tels que l'indice d'électrophilie et la fonction de Fukui atomique étaient suffisants pour prédire l'électrophilie de Mayr (pour la sous-famille des accepteurs monosubstituées éthyléniques (M.A.C.S.E.)  $R^2 = 0,93$  avec un descripteur, voir la **Figure 3**).

Si l'on considère un ensemble de molécules de structures variées, il est nécessaire de passer par un modèle de régression faisant appel à plusieurs descripteurs (par exemple,  $R^2 = 0,82$  avec un modèle bilinéaire, voir la **Figure 4**). Avec un modèle obtenu par une régression trilineaire dans l'article ( $R^2 = 0,87$ , voir équation 10 et **Figure 5**), il est ainsi possible de déterminer la force relative (forte, moyenne ou faible) d'un accepteur de Michael de manière qualitative.

La méthodologie employée dans cet article peut être adaptée à d'autres familles importantes de molécules organiques dont les paramètres expérimentaux sont disponibles. Durant les deux années qui ont suivi la publication de l'article, près de 70 nouveaux paramètres expérimentaux sont ainsi apparus dans la littérature.

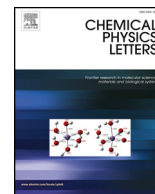
Afin d'améliorer le modèle obtenu dans cet article, nous pouvons envisager trois axes principaux. Le premier reviendrait à utiliser d'autres descripteurs quantiques de réactivité, en supposant que les descripteurs que nous avons utilisés ne soient pas les plus adaptés au problème. C'est une voie que nous n'avons pas explorée. Dans un second temps nous pouvons supposer que la structure calculée des accepteurs de Michael utilisée ici n'est pas assez représentative et envisager de réaliser une respiration de la structure optimisée ; cette voie correspond à la seconde partie de ce chapitre. Enfin, nous pouvons envisager des méthodes plus sophistiquées, tels que les méthodes non linéaires. C'est la voie que nous avons explorée dans la troisième partie, en utilisant des méthodes d'apprentissages automatisés.

## ***2. Sur l'influence des effets dynamiques sur les descripteurs de réactivité***

### **2.1 Préambule**

La majeure partie des applications de la DFT conceptuelle repose sur une approche statique. Les descripteurs sont ainsi calculés sur les structures optimisées des espèces réactives. Or, au cours d'une réaction chimique, la molécule ne reste pas dans sa structure d'équilibre sans mouvement : elle est en constante évolution.

Afin d'améliorer le modèle précédent quant à la prédiction qualitative de l'électrophilie des accepteurs de Michael, nous avons ainsi décidé lors d'une nouvelle étude d'examiner l'influence d'une respiration de la structure d'équilibre sur les descripteurs de réactivité chimique, par dynamique moléculaire en solution. Les différents descripteurs de réactivité sont calculés sur des « *snapshots* » extraits le long de la dynamique moléculaire et traités statistiquement en termes de moyenne, médiane, maximum et minimum. Les régressions vis-à-vis des valeurs expérimentales d'électrophilie de Mayr sont ensuite effectuées à partir de descripteurs *statistiques* et comparées à celles obtenues avec les descripteurs purement *statiques*.



## Research paper

## On the influence of dynamical effects on reactivity descriptors

Guillaume Hoffmann, Vincent Tognetti\*, Laurent Joubert\*

Normandy Univ., COBRA UMR 6014 &amp; FR 3038, Université de Rouen, INSA Rouen, CNRS, 1 rue Tesnière, 76821 Mont St Aignan. Cedex, France



## HIGHLIGHTS

- The influence of dynamical structural effects on conceptual DFT reactivity descriptors is investigated.
- They were shown to feature significant differences with respect to standard static approaches.
- They provide new reactivity descriptors that are much more correlated to Mayr's experimental electrophilicity values.

## ARTICLE INFO

## Keywords:

Michael acceptors  
Electrophilicity  
Conceptual DFT  
Reactivity indices  
Molecular dynamics  
Mayr's scale

## ABSTRACT

In this paper, we investigate the role of structural dynamical effects on reactivity properties of selected Michael acceptors. To this aim, quantum molecular dynamics simulations were performed. Conceptual density functional theory descriptors were then evaluated on selected snapshots, providing statistical indicators suitable to assess the discrepancies between static and dynamical approaches. The implications of these results for building a predictive model correlated to Mayr's electrophilicity index are then discussed, paving the way towards a more realistic account of experimental results.

## 1. Introduction

Conceptual density functional theory (CDFT) [1–4] and the respective chemical reactivity indices that it generates were shown successful in the prediction and characterization of empirical and theoretical concepts, such as electrophilicity [5–9], nucleophilicity [10–12], gas-phase basicities [13], Woodward-Hoffmann rules [14],  $\sigma$ -holes [15,16], *trans*-effect [17,18], and covering all research fields ranging from organic to inorganic chemistries. In particular, they provide an efficient prediction of the feasibility of a chemical reaction based only on the reactants physicochemical properties. In the case of electrophilicity and nucleophilicity, these theoretical descriptors can be confronted to the corresponding experimental parameters defining the popular Mayr's scales [19–21].

Noteworthy, these scales are widely used by experts in organic synthesis for a quick assessment of the outcome of a reaction between known compounds. Indeed, a reliable rule of thumb is that the reaction of a nucleophile and an electrophile will be quantitative if the sum of their Mayr's parameters is higher than  $-5$ . Currently, there are 1456 available experimental parameters for reagents such as diverse as carbocations, aldehydes, ketones, imines, alkenes, ylides, aliphatic amines..., thus covering most of the usual arsenal used in organic

synthesis.

It currently remains a theoretical challenge to predict these experimental parameters (based on kinetic measurements) for any type of molecule. Certainly, such a predicting tool would be highly valuable for organic chemists dealing with compounds for which experimental parameters have not been yet determined, and even for biologists, for example in the assessment of toxicity of molecules [22]. Besides this predictive power, such a model could also convey chemical rationalization, which could help improving the current reagents by a well-targeted *in silico* strategy.

In a preceding study [23], we developed a model able of categorizing the electrophilicity reactivity of  $\alpha$ - $\beta$ -unsaturated compounds, which epitomize Michael acceptors, a family of chemical reagents instrumental in the formation of carbon-carbon bonds and ubiquitous in organic synthesis. This model, as almost all of related works (see also Ref. [24]), was based on a static approach of chemical reactivity, and failed at providing a quantitative model with respect to experimental data, a fact also underlined by Mayr in recent studies devoted to Michael acceptors [25,26] and later extended to other reagents like ketones [27].

One may thus wonder whether this failure might be ascribable to the neglect of dynamical features. Indeed, due to thermal agitation, the

\* Corresponding authors.

E-mail addresses: [vincent.tognetti@univ-rouen.fr](mailto:vincent.tognetti@univ-rouen.fr) (V. Tognetti), [laurent.joubert@univ-rouen.fr](mailto:laurent.joubert@univ-rouen.fr) (L. Joubert).

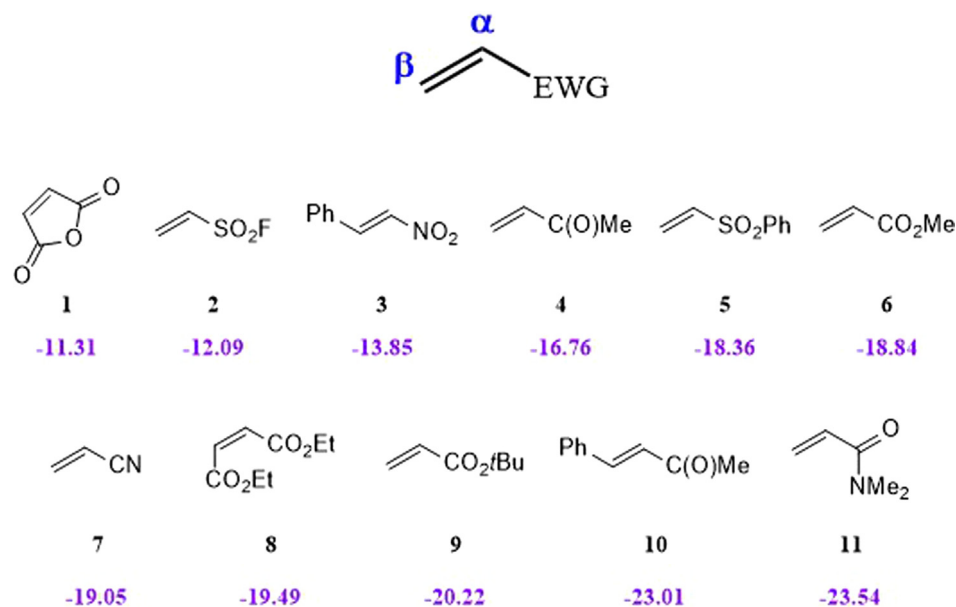


Fig. 1. View of the studied Michael acceptors. EWG stands for electron-withdrawing group. Numbers refer to experimental electrophilicities as measured by the Mayr's group (relative units).

reactants inside a beaker are generally not frozen in their most stable conformation. This structural breathing may in some cases lead to explore geometries that can significantly differ from the optimized ones, and which would accordingly exhibit significantly different reactivity behaviors.

From the best of our knowledge, the calculation of chemical reactivity descriptors extracted from dynamical trajectories is extremely scarce, probably because static approaches often revealed sufficient to correctly predict experimental results, as exemplified by the benzhylium carbocations family for which many high linear correlations were reported and that has served as a workhorse for electrophilicity studies for a long time.

As a matter of fact, the predominant focus on rigid or hindered species of such type could account for the prevalence of the static theoretical approach that is, furthermore, much less time-consuming from a computational perspective. Nevertheless, we want to point out that the dynamical behavior of chemical reactivity indices was seminally investigated by Liu [28] in gas phase for few hydrated clustered cations or anions (like  $\text{Ca}^{2+}(\text{H}_2\text{O})_{15}$  or  $\text{Cl}^-(\text{H}_2\text{O})_{30}$ ), but with the main focus not being organic chemical reactivity. Another important study is that presented as a poster by Oller and Vöhringer-Martinez during the QUITEL 2018 conference held in Santiago (Chile), which reports a methodology based on molecular dynamics to define Boltzmann weighted reactivity descriptors [29].

In the current work, molecular dynamics simulations in solution on selected Michael acceptors were performed at a semi-empirical quantum chemical level of theory in order to follow the evolution over time of some of the most relevant reactivity descriptors considered in our previous paper [23], and to assess to what extent they differ from their static properties and may afford hints on how to improve reactivity prediction by means of conceptual DFT.

## 2. Computational details

All density functional theory calculations were performed using the Gaussian 09 package [30]. Consistently with our previous study [23], molecular geometries were first optimized using the *meta*-hybrid M06-2X exchange-correlation functional [31] in conjunction with the aug-cc-pVTZ basis set, with recommended 'ultrafine' integration grid. Solvent effects (dimethylsulfoxide, DMSO) were taken into account by a

polarizable continuum model (in the last implementation of the integral equation formalism (IEFPCM) [32]) for all computations (static and dynamical approaches).

These optimized structures were then used to start unbiased molecular dynamics simulations at the quantum chemical semi-empirical PM6 level of theory [33] (within IEFPCM), which allows long simulations at a low computational cost, in the NVT canonical ensemble. The time step for propagation was set to 1 fs. The PM6 wavefunction was converged at any step (Born-Oppenheimer-type dynamics). The total simulation time was equal to 100 ps, ensuring representative sampling. Temperature ( $T = 298.15$  K) was controlled by velocity scaling.

Structures were then extracted every 1 ps. The 20 first ones were considered to be part of the equilibration process, while the 80 following ones were considered for data production. Indeed, as shown in Graph S1 in the supplementary information file, temperature was found to be properly converged after 20 ps for every simulation. Subsequent single point calculations at the M06-2X/aug-cc-pVTZ level of theory were carried out to evaluate the chosen conceptual DFT reactivity descriptors.

These descriptors can be classified into two main classes: molecular (also known as "global") and atomic ones. Those belonging to the first category are here: three avatars of the electronic chemical potential ( $\mu$ ,  $\mu^+$ ,  $\mu^-$ ), the molecular hardness  $\eta$ , softness  $S$ , and hyperhardness  $\gamma$  [34], the electrophilicity index  $\omega$  [35,36], and the electroaccepting power  $\omega^+$ . Atomic descriptors considered here are: condensed electrophilic and nucleophilic Fukui functions  $f^+$ ,  $f^-$ , and the dual descriptor  $f^{(2)}$  [37,38], supplemented by grand-canonical extensions ( $\mu^+ f^+$ ,  $\mu^- f^-$ ,  $\omega f^+$ ,  $S f^{(2)}$ ,  $s^{(2)}$ ,  $s^+ = S f^+$ ,  $s^- = S f^-$ ). We refer the interested reader to our previous paper for the corresponding mathematical definitions [23].

Besides, two main computational approaches can be implemented to evaluate these descriptors: either finite difference linearization (FDL) or the Koopmans-like frozen molecular orbital (FMO) scheme, which are not equivalent [39]. For each descriptor, both were here considered. Atomic condensation was performed within the framework of Bader's atoms-in-molecules theory (QTAIM) [40,41] using Todd Keith's AIMAll software [42].

It is important to notice that two carbon atoms denoted  $C_\alpha$  and  $C_\beta$  (see Fig. 1) may be involved in the carbon-carbon bond formation process during a Michael addition, so that condensation of the two

atomic corresponding basins was performed. In Ref. [23], we also defined the total electrophilicity of the  $C_\alpha=C_\beta$  double bond,  $f^+(C_{\alpha\beta})$ , by simply summing the electrophilicity of each atom,  $f^+(C_\alpha)$  and  $f^+(C_\beta)$ .

In total, for each molecule, 9 global descriptors and 22 condensed ones were evaluated at both the FDL and FMO theory levels at each selected point along the dynamical trajectory.

Atomic units (a.u.) are used throughout the paper unless explicitly otherwise stated. Numerical values corresponding to the statistical analysis of all descriptors are gathered in the [supplementary information file](#) (Tables S1–S80). Experimental values were extracted from Mayr's online database [43].

### 3. Results and discussion

In order to investigate dynamical effects on chemical reactivity indices, we selected 11 molecules (represented in Fig. 1) out of our previous dataset [23] that consisted of 35 Michael acceptors. Let us recall that we divided it into three subgroups: “good” Michael acceptors, “medium”, and “low” ones. In order to be as representative as possible, we thus selected in the current paper molecules stemming from each of these three representative families. Note that two molecules of these sets are symmetric (compounds 1 and 8) at their energy minimum, a property that will not be kept during the molecular dynamics at a given time. However, the final time-averaged properties for the two carbons that are equivalent in the optimized structures were found, as expected, almost equal.

In summary, this new working subset covers a range of experimental electrophilicity values from  $-11.31$  to  $-23.54$ , which is that of the known experimental one for Michael acceptors, with values well distributed along it. To be more comprehensive, let us recall that the nucleophilic addition on a Michael acceptor that features conjugated multiple bonds (for instance  $C=C$  and  $C=O$  as in 11, or  $C=C$  and  $C\equiv N$  as in 7) affords the formation of a bond between the nucleophile (often a carbanion) and carbone  $C_\beta$  of the acceptor (see Fig. 1 for atom labeling).

As a first example for the analysis of variation of descriptors along a trajectory, we now scrutinize the electrophilicity index (calculated using FDL) for N,N-dimethylprop-2-enamide (molecule 11). Its values during the simulation are displayed in the left part of Fig. 2. The static value was found equal to 0.061 a.u. It can be compared to the average value (calculated on the last 80 ps): 0.057 a.u., and to the median: 0.054 a.u. The obtained values spanned between 0.046 a.u. (min) and 0.077 a.u. (max), and the standard deviation is 0.009 a.u., so that the coefficient of variation is equal to 16%. This last value is not at all

negligible, showing that even for a rigid molecule as 11, dynamical effects can be of importance. Interestingly, the average and median values suggest that the molecule is actually less reactive than expected from the static approach.

One may wonder if this is a general result for the FDL electrophilicity index. It is actually the case for compound 10 (static value: 0.091, average: 0.075) and 3 (static: 0.122, average: 0.115). Note that these two molecules show different behaviors: important discrepancies between the two approaches for 10, small ones for 3. Conversely, the trend was found opposite for 2 (static value: 0.079, average: 0.094) and 1 (static: 0.120, average: 0.124). It is also instructive to compare 10 and 2: the electrophilicity index is higher from the static point view for 10, while it is higher in 2 from the average side.

It is also instructive to compare values for FMO electrophilicity index values. At the static level, the values are almost identical for compounds 2 and 10 (0.060 and 0.062, respectively), while the average values considerably differ: 0.069 (molecule 2) with respect to 0.052 (10). This is in line with the fact that Mayr's electrophilicity is significantly higher for 2 than for 10 ( $-12.1$  vs.  $-23.0$ ). All these preliminary results prove that considering a dynamical approach can invert reactivity orders, and possibly significantly alter the reactivity hierarchy.

As a second example, we now have a look at an atomic descriptor, and more specifically at the dual descriptor condensed on carbon  $C_\beta$  for compound 10, namely (E)-4-phenylbut-3-en-2-one. Let us recall that the dual descriptor allows for predicting if an atom is mainly electrophilic (positive condensed value) or nucleophilic (negative condensed value). The time evolution of the dual descriptor  $f_{FDL}^{(2)}(C_\beta)$  is represented in the right part of Fig. 2. It appears that the static, median and average values are all positive, but that the two first quantities differ a lot, since these three values were found equal to 0.012, 0.057, 0.055, respectively.

Another salient feature is that there are several sign changes along the trajectory, the most important one occurring around  $t = 55$  ps. This means that the carbon atom then changes its reactivity from electrophilic to nucleophilic. One then would have expected the average value to be lower than the static one due to these negative contributions. This is actually not the case because they are largely compensated by important positive values (see for instance the positive peak before  $t = 80$  ps). As a consequence, this compound exhibits a very high coefficient of variation value since it amounts to 84%.

All of these statistical quantities were evaluated for all molecules. By doing so, the average, minimal, maximal, and median values become new reactivity descriptors that can be used to predict Mayr's

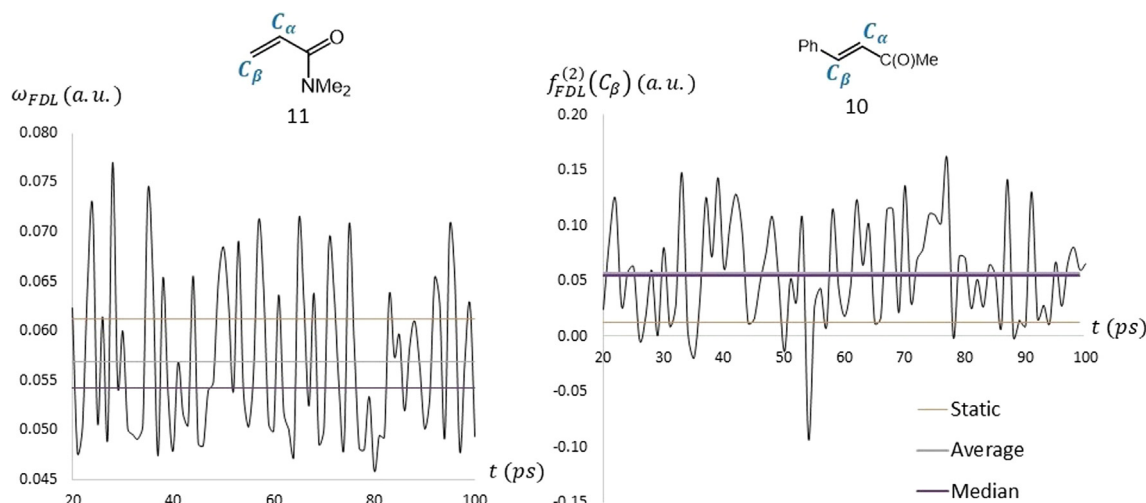
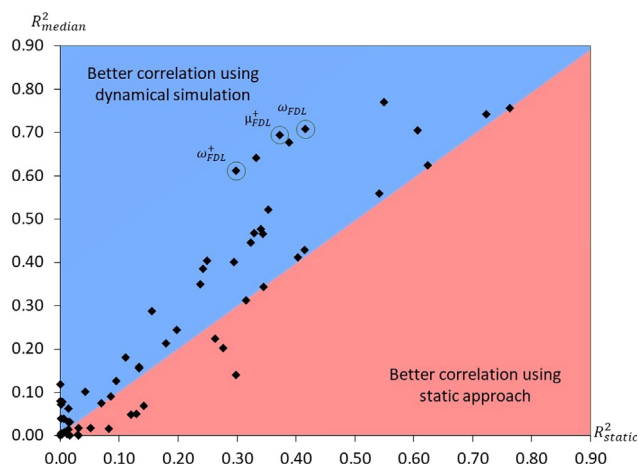


Fig. 2. Time (in ps) evolution of the electrophilicity index for molecule 11 (left part, atomic units) and of  $C_\beta$  condensed dual descriptor for molecule 10 (right part, atomic units), obtained by single-point calculations at the M06-2X/aug-cc-pVTZ level of theory on snapshots from PM6 NVT molecular dynamics simulation.





**Fig. 3.** Coefficients of determination ( $R^2_{\text{median}}$ ) for monolinear regressions using median values of each reactivity descriptor with respect to coefficients of determination ( $R^2_{\text{static}}$ ) using static values of the same descriptors.

experimental electrophilicity, denoted  $E_{\text{Mayr}}$ , for the subset of 11 Michael acceptors. In fact, for a given descriptor, the dynamical values are generally not highly correlated to the static ones, as shown in Tables S81–S88 in the supplementary information file that gather selected correlation matrices. For instance, the coefficient of determination between the average and the static  $f_{\text{FDL}}^{(2)}(C_\beta)$  values was found equal to 0.66. This implies that these dynamical descriptors are true new descriptors.

A simple way for comparing their prediction ability with static ones is to compare their coefficient of determination (for monolinear regressions) towards Mayr's experimental electrophilicity. To this aim, Fig. 3 shows the  $R^2_{\text{median}}$  values from dynamical simulations with respect to the  $R^2_{\text{static}}$  ones for all tested descriptors.

This analysis showed that for some descriptors, dynamics effects do not significantly modify the correlation to the experimental value. This is for instance the case for  $\mu_{\text{FDL}}$ ,  $S_{\text{FDL}}$  and  $\gamma_{\text{FDL}}$ . Indeed, for the electronic chemical potential, we found  $R^2_{\text{median}} = 0.74$  and  $R^2_{\text{static}} = 0.72$ . Likewise, all descriptors with very low  $R^2$  values (lower than 0.2) in the static case also exhibit low  $R^2$  dynamical values.

On the contrary, we obtained important differences for some other reactivity descriptors. This is notably the case for  $\omega_{\text{FDL}}$ ,  $\mu_{\text{FDL}}^+$  and  $\omega_{\text{FDL}}^{+}$ , the corresponding points being circled in Fig. 3. For the electrophilicity index, we got  $R^2_{\text{median}} = 0.71$  and  $R^2_{\text{static}} = 0.42$ . These results indicate that some descriptors that were not highly correlated to experimental electrophilicity in the static approach can be now considered as promising tools for the global prediction. More generally, it was found that  $R^2_{\text{median}}$  values are generally higher than static ones, corresponding to points located in the blue area in Fig. 3. Similar conclusions hold for the other statistical descriptors, in particular the time-averaged values.

More precisely, the highest  $R^2$  values obtained using all static and

dynamical values in monolinear regression procedure was found with the median value of  $\omega_{\text{FMO}}$ , for which  $R^2$  is equal to 0.77. If only static values were retained, this  $R^2$  value dropped down to 0.55. Henceforth, there is a little improvement, but it can be considered not sufficient to achieve quantitative correlation. In fact, as discussed in our previous paper [23], at least two descriptors should be used for correlation when the dataset is composed of very different molecules in order to grasp the various possible reactivity behaviors (charge or orbital control, role of the two carbons of the double bond in the chemical process...).

We then performed bilinear regression analysis. In order to only deal with consistent combinations, only FDL (resp. FMO) descriptors were paired with another FDL (resp. FMO) one. For the very same reasons, we only coupled statistical descriptors of the same kind: for instance, an average value with another average value, excluding mixed combinations (an average value with a maximal one, for example). Ten descriptor subsets were thus subjected to bilinear regressions, defined by FMO static, FDL static, FMO average, FDL average, FMO median, FDL median, FMO maximal, FDL maximal, FMO minimal, and finally FMO maximal values. Table 1 collects the obtained models with the highest  $R^2$  values.

When using only static descriptors, the best models showed  $R^2$  of about 0.8. Noteworthy, such values indicate a prediction capacity that is similar to the one we reported for the full dataset of 35 molecules of our previous paper [23] (this likeness also confirms that the current subset can be considered as statistically representative of the whole one). The model with the highest  $R^2$  value was the following one:

$$E_{\text{Mayr}} = -54.39 + 344.92\mu_{\text{FMO,static}}^+ + 821.34\omega_{\text{FMO,static}} \quad (1)$$

It is worth noticing that it involves only global descriptors, which is an advantage since it is independent of any atomic partition and since its evaluation is more straightforward than condensed values. Nevertheless, the  $R^2$  value remains rather low. One can then wonder if it can be improved using statistical global descriptors. This is actually the case, involving other descriptors:

$$E_{\text{Mayr}} = -42.99 - 169.60\mu_{\text{FDL,average}} - 55.35\gamma_{\text{FDL,average}} \quad (2)$$

with  $R^2 = 0.86$ . Substantial predictivity enhancement can thus be achieved only incorporating condensed descriptors from the dynamical approach. The most promising models were actually obtained using the medians ( $R^2 = 0.93$ ) according to:

$$E_{\text{Mayr}} = -14.04 - 42.46s_{\text{FDL,median}}^+(C_\alpha) + 1836.84\omega_{\text{FDL,median}}^+(C_\alpha) \quad (3)$$

and with the average values ( $R^2 = 0.97$ ) following:

$$E_{\text{Mayr}} = -14.58 - 44.84s_{\text{FDL,average}}^+(C_\alpha) + 1996.99\omega_{\text{FDL,average}}^+(C_\alpha) \quad (4)$$

Note that here  $\omega_{\text{median}}^+$  refers to the median of the whole  $\omega(t_i)f^+(t_i)$  value distribution along the trajectory (at times  $t_i$ ). It is **not** equal to the average value of  $\omega$  multiplied by the average of  $f^+$  values.

Our last model (represented in Fig. 4) is thus particularly appealing since its mean absolute deviation with respect to Mayr's experimental

**Table 1**

Bilinear regressions toward Mayr's electrophilicity scale with highest coefficient of determination values, depending on the nature of the used descriptors.

| Static                                                       |       | Average                                                       |       | Median                                                        |       | Max.                                                            |       | Min.                                               |       |
|--------------------------------------------------------------|-------|---------------------------------------------------------------|-------|---------------------------------------------------------------|-------|-----------------------------------------------------------------|-------|----------------------------------------------------|-------|
| Pair                                                         | $R^2$ | Pair                                                          | $R^2$ | Pair                                                          | $R^2$ | Pair                                                            | $R^2$ | Pair                                               | $R^2$ |
| $\eta_{\text{FMO}}, \omega_{\text{FMO}}$                     | 0.79  | $\mu_{\text{FMO}}, f_{\text{FMO}}^+(C_\alpha)$                | 0.92  | $\mu_{\text{FMO}}, f_{\text{FMO}}^+(C_\alpha)$                | 0.90  | $\mu_{\text{FMO}}, \omega_{\text{FMO}}^+$                       | 0.83  | $\mu_{\text{FDL}}, f_{\text{FDL}}^{(2)}(C_\alpha)$ | 0.83  |
| $\omega_{\text{FMO}}, \omega_{\text{FMO}}^+$                 | 0.79  | $\mu_{\text{FDL}}, \omega_{\text{FDL}}^+(C_\alpha)$           | 0.92  | $\mu_{\text{FDL}}, f_{\text{FDL}}^+(C_\alpha)$                | 0.91  | $\mu_{\text{FDL}}, \omega_{\text{FDL}}^+$                       | 0.83  | $\mu_{\text{FMO}}, S_{\text{FMO}}$                 | 0.84  |
| $f_{\text{FMO}}^+(C_\alpha), \omega_{\text{FMO}}^+(C_\beta)$ | 0.80  | $\mu_{\text{FMO}}, S_{\text{FMO}}^+(C_\alpha)$                | 0.93  | $\mu_{\text{FMO}}, S_{\text{FMO}}^+(C_\alpha)$                | 0.91  | $\mu_{\text{FDL}}, \omega_{\text{FDL}}^+(C_\beta)$              | 0.83  | $\mu_{\text{FDL}}, \mu_{\text{FDL}}$               | 0.84  |
| $\eta_{\text{FMO}}, \omega_{\text{FMO}}^+$                   | 0.80  | $\mu_{\text{FDL}}, f_{\text{FDL}}^+(C_\alpha)$                | 0.93  | $\mu_{\text{FMO}}, S_{\text{FMO}}^+(C_\alpha)$                | 0.91  | $S_{\text{FDL}}, \omega_{\text{FDL}}^+$                         | 0.83  | $\mu_{\text{FDL}}, f_{\text{FDL}}^+(C_\beta)$      | 0.84  |
| $\mu_{\text{FMO}}, f_{\text{FMO}}^{(2)}(C_\beta)$            | 0.80  | $\mu_{\text{FMO}}, \omega_{\text{FMO}}^+(C_\alpha)$           | 0.94  | $\mu_{\text{FDL}}, S_{\text{FDL}}^+(C_\alpha)$                | 0.92  | $S_{\text{FMO}}, \omega_{\text{FMO}}^+$                         | 0.83  | $\mu_{\text{FDL}}, S_{\text{FDL}}$                 | 0.85  |
| $\mu_{\text{FMO}}, S_{\text{FMO}}$                           | 0.81  | $\mu_{\text{FDL}}, S_{\text{FDL}}^+(C_\alpha)$                | 0.94  | $S_{\text{FDL}}^+(C_\alpha), \omega_{\text{FDL}}^+(C_\alpha)$ | 0.93  | $\mu_{\text{FDL}}, \omega_{\text{FDL}}^+$                       | 0.83  | $\mu_{\text{FMO}}, S_{\text{FMO}}$                 | 0.85  |
| $\mu_{\text{FMO}}, \omega_{\text{FMO}}$                      | 0.81  | $S_{\text{FDL}}^+(C_\alpha), \omega_{\text{FDL}}^+(C_\alpha)$ | 0.97  | $\mu_{\text{FMO}}, \omega_{\text{FMO}}^+(C_\alpha)$           | 0.94  | $f_{\text{FDL}}^{(2)}(C_\beta), \omega_{\text{FDL}}^+(C_\beta)$ | 0.84  | $\mu_{\text{FDL}}, f_{\text{FDL}}^+(C_\alpha)$     | 0.88  |

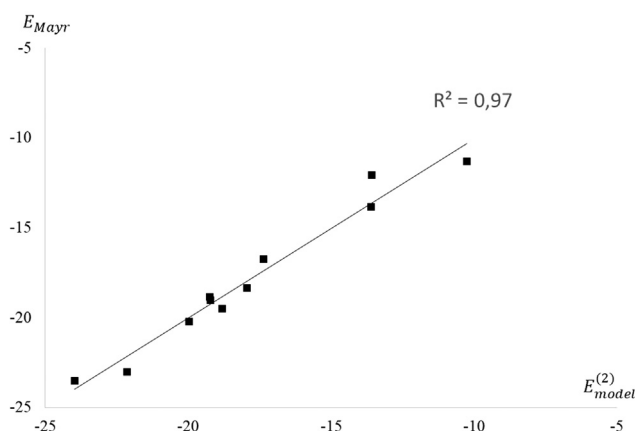


Fig. 4. Best bilinear model for the prediction of Mayr's electrophilicity scale (relative units) according to Eq. (4). The straight line represents perfect agreement with experimental data.

values is equal to 0.6 relative units, to be compared to 1.6 that was obtained with Eq. (1) from the static approach. A striking feature of Eqs. (3) and (4) is that they involve properties related to carbon  $\alpha$  and not to carbon  $\beta$ , while the bond between nucleophile and electrophile occurs on carbon  $\beta$ . This is an important point we already discussed in our previous paper [23]. Actually, if we restrained the bilinear regression to  $C_\beta$  descriptors, the best model corresponds to:

$$E_{\text{Mayr}} = -32.37 - 14.59f_{\text{FDL,maximal}}^{(2)}(C_\beta) + 1053.59\omega_{\text{FDL,maximal}}^+(C_\beta) \quad (5)$$

with  $R^2 = 0.84$ , featuring an obvious worsening of the model.

Obviously, bilinear models are relevant if they combine descriptors that are not highly correlated. We thus computed the intercorrelation matrix gathering the  $R^2$  values between each pair of descriptors entering Eqs. (1)–(5). It corresponds to Table S88 in the supplementary information file. It immediately appears that such descriptors cannot be considered correlated, so that those involved in the bilinear models are really complementary.

Lastly, in order to roughly assess the reliability of such models, we considered a molecule that was not included in the fitting set, namely fumaronitrile. Its Mayr's electrophilicity is equal to  $-15.71$ , filling the missing gap between compounds 4 ( $E_{\text{Mayr}} = -16.76$ ) and 3 ( $E_{\text{Mayr}} = -13.85$ ). The static values for  $s_{\text{FDL}}^+(C_\alpha)$  and  $\omega_{\text{FDL}}^+(C_\beta)$  were found to be equal to 0.527 and 0.013, respectively, to be compared to the corresponding time-averaged ones: 0.564 and 0.012. The use of Eq. (4) then predicted an electrophilicity equal to  $-15.31$ . The difference between this last value and the experimental one is equal to 0.4, and it is thus lower than the mean absolute deviation calculated on the 11 molecules used to build the model. This new result supports once more the promising predictive power of dynamical reactivity descriptors.

Finally, it should also be noticed that, unfortunately, such models are far from being easily interpreted since it involves a competition between  $s_{\text{FDL,average}}^+(C_\alpha)$  and  $\omega_{\text{FDL,average}}^+(C_\beta)$  that both enter Eq. (4) with opposite signs. There are in fact cases where the sum of these two contributions is positive and other ones for which it reveals negative. We believe that such facts could not simply be accounted for without resorting to a detailed mechanistic study of these reactions, which is outside the scope of this paper and that we will report in due course.

#### 4. Conclusions

In this letter, we have shown that dynamical effects can trigger dramatic changes of reactivity descriptor values with respect to those obtained from a static approach. We also gave evidence that average and median statistical data gathered along a trajectory allows for a

much better prediction of Mayr's electrophilicity parameters. It also paves the way toward the definition of temperature-dependent reactivity indices (by choosing the appropriate thermostat temperature for the molecular dynamics simulations), and therefore toward more realistic explanation of experimental data. Besides, such a protocol is general and could be applied to other families than Michael acceptors.

#### Conflict of interest

Author declares that there is no conflict of interest.

#### Acknowledgements

The authors would like to gratefully acknowledge the LABEX SynOrg and the Normandy Region for funding and support, and the CRIANN high-performance computing facility. Part of this project was supported by the RIN AGAC project.

#### Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cplett.2019.03.040>.

#### References

- [1] R.G. Parr, W. Yang, *Annu. Rev. Phys. Chem.* 46 (1995) 701.
- [2] H. Chermette, *J. Comput. Chem.* 20 (1999) 129.
- [3] P. Geerlings, F. De Proft, W. Langenaeker, *Chem. Rev.* 103 (2003) 1793.
- [4] F. De Proft, P. Geerlings, *Chem. Rev.* 101 (2001) 1451.
- [5] L.R. Domingo, P. Pérez, R. Contreras, *Tetrahedron* 60 (2004) 6585.
- [6] A. Aizman, R. Contreras, P. Pérez, *Tetrahedron* 61 (2005) 889.
- [7] V. Tognetti, C. Morell, L. Joubert, *Chem. Phys. Lett.* 635 (2015) 111.
- [8] L.G. Zhuo, W. Liao, Z.X. Yu, *Asian J. Org. Chem.* 1 (2012) 336.
- [9] X.Y. Zhou, C.Y. Rong, T. Lu, S. Liu, *Acta Phys. Chim. Sin.* 30 (2014) 2055.
- [10] S. Liu, C. Rong, T. Lu, *J. Phys. Chem. A* 118 (2014) 3698.
- [11] C. Wang, Y. Fu, Q.X. Guo, S. Liu, *Chem. Eur. J.* 16 (2010) 2586.
- [12] R. Contreras, J. Andres, V.S. Safont, P. Campodonico, J.G. Santos, *J. Phys. Chem.* 107 (2003) 5588.
- [13] S.K. Rajak, N. Islam, D.C. Ghosh, *J. Quant. Inf. Sci.* 1 (2011) 87.
- [14] P. Geerlings, P.W. Ayers, A. Toro-Labbé, P.K. Chattaraj, F. De Proft, *Acc. Chem. Res.* 45 (2012) 683.
- [15] V. Tognetti, C. Morell, L. Joubert, *J. Comput. Chem.* 36 (2015) 649.
- [16] N. Galland, G. Montavon, J.-Y. Le Questel, J. Graton, *New J. Chem.* 42 (2018) 10510.
- [17] F. Guégan, V. Tognetti, L. Joubert, H. Chermette, D. Luneau, C. Morell, *Phys. Chem. Chem. Phys.* 18 (2016) 982.
- [18] V. Tognetti, F. Guégan, D. Luneau, H. Chermette, C. Morell, L. Joubert, *Theor. Chem. Acc.* 136 (2017) 85.
- [19] H. Mayr, M. Patz, *Angew. Chem. Int. Ed.* 33 (1994) 938.
- [20] R. Lucius, R. Loos, H. Mayr, *Angew. Chem. Int. Ed.* 41 (2002) 91.
- [21] H. Mayr, *Angew. Chem. Int. Ed.* 50 (2011) 3612.
- [22] J.A.H. Schwob, Y.K. Koleva, S.J. Enoch, F. Bajot, M. Hewitt, J.C. Madden, D.W. Roberts, T.W. Schultz, *Chem. Rev.* 111 (2011) 2562.
- [23] G. Hoffmann, V. Tognetti, L. Joubert, *J. Mol. Model.* 24 (2018) 281.
- [24] J. Oller, P. Pérez, P.W. Ayers, E. Vöhringer-Martinez, *Int. J. Quant. Chem.* 118 (2018) e25706.
- [25] D.S. Allgäuer, H. Jangra, H. Asahara, Z. Li, Q. Chen, H. Zipse, A.R. Ofial, H. Mayr, *J. Am. Chem. Soc.* 139 (2017) 13318.
- [26] H. Jangra, Q. Chen, E. Fuks, I. Zenz, P. Mayer, A.R. Ofial, H. Zipse, H. Mayr, *J. Am. Chem. Soc.* 140 (2018) 16758.
- [27] Z. Li, H. Jangra, Q. Chen, P. Mayer, A.R. Ofial, H. Zipse, H. Mayr, *J. Am. Chem. Soc.* 140 (2018) 5500.
- [28] S. Liu, *J. Chem. Sci.* 117 (2005) 477.
- [29] J. Oller, E. Vöhringer-Martinez, Boltzmann weighted local reactivity descriptors of  $\alpha,\beta$ -unsaturated systems by molecular dynamics simulations, Poster presentation in the XLIV International Congress of Theoretical Chemists of Latin Expression, Santiago, (2018).
- [30] M. J. Frisch, et al. Gaussian 09 Revision D.01, Gaussian Inc., Wallingford, CT; 2013.
- [31] Y. Zhao, D.G. Truhlar, *Theor. Chem. Acc.* 120 (2008) 215.
- [32] E. Cancès, B. Mennucci, J. Tomasi, *J. Chem. Phys.* 107 (1997) 3032.
- [33] J.J.P. Stewart, *J. Mol. Model.* 13 (2007) 1173.
- [34] C. Morell, A. Grand, A. Toro-Labbé, H. Chermette, *J. Mol. Model.* 19 (2013) 2893.
- [35] R.G. Parr, L. von Szentpály, S. Liu, *J. Am. Chem. Soc.* 121 (1999) 1922.
- [36] P.K. Chattaraj, U. Sarkar, D.R. Roy, *Chem. Rev.* 106 (2006) 2065.
- [37] C. Morell, A. Grand, A. Toro-Labbé, *J. Phys. Chem. A* 109 (2005) 205.
- [38] C. Morell, A. Grand, A. Toro-Labbé, *Chem. Phys. Lett.* 425 (2006) 342.
- [39] F. Zielinski, V. Tognetti, L. Joubert, *Chem. Phys. Lett.* 527 (2012) 67.
- [40] R.F.W. Bader, *Atoms in Molecules: A Quantum Theory*, Oxford University Press, Oxford, U.K., 1990.
- [41] P.L.A. Popelier, *Atoms in Molecules An Introduction*, Pearson Education, Harlow, 2000.
- [42] T.A. Keith, AIMAll (Version 17.11.14), [aim.tkgristmill.com](http://aim.tkgristmill.com); 2017.
- [43] <https://www.cup.lmu.de/oc/mayr/reaktionsdatenbank2/>.



## 2.3 Conclusion et perspectives

Certaines molécules dont les structures varient faiblement le long de la dynamique moléculaire (comme la molécule 7, page 25 de l'article) présentent une faible variation de leurs valeurs de descripteurs de réactivité chimique.

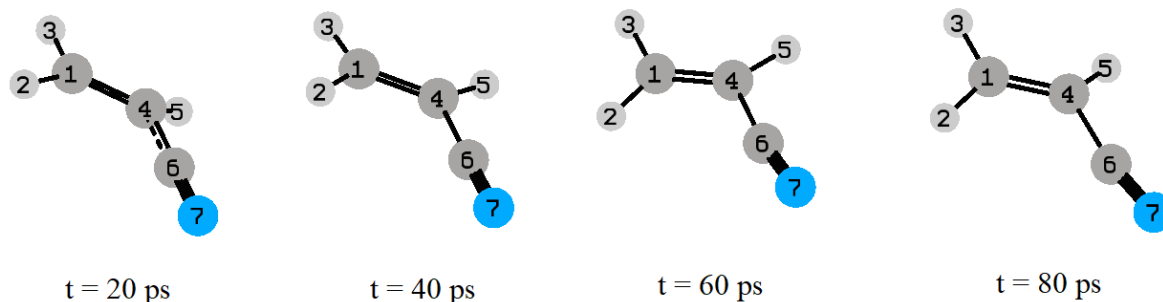


Figure 4.3: Quatre structures extraites de la dynamique moléculaire pour la molécule 7. Image générée avec GaussView 5.0

Néanmoins, pour d'autres molécules on observe des changements conformationnels lors d'une dynamique moléculaire (comme la molécule 6), qui subissent une influence importante de la dynamique moléculaire sur les valeurs calculées des descripteurs de réactivité chimique. La molécule 6 passant de la géométrie *cis* à *trans* par rotation autour de la liaison simple C-C durant la dynamique ( $\Delta E = 5,1 \text{ kcal. mol}^{-1}$  pour la barrière d'activation).

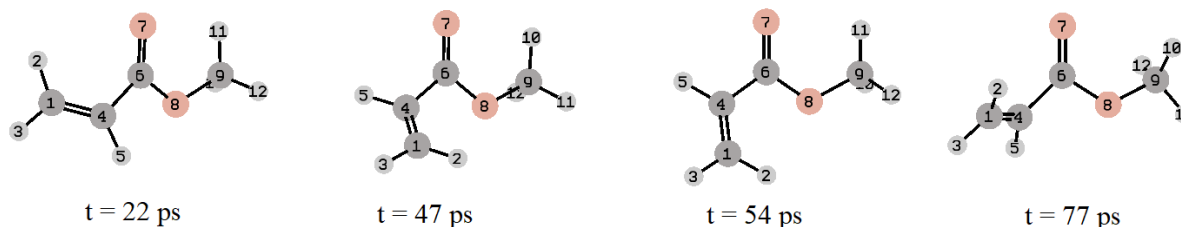


Figure 4.4 : Quatre structures extraites de la dynamique moléculaire pour la molécule 6. Image générée avec GaussView 5.0

Les descripteurs dynamiques ainsi calculés démontrent une capacité supérieure de prédiction de l'électrophilie (voir la **Figure 3** de l'article) par rapport aux descripteurs issus de l'approche statique, permettant une prédiction suffisante de l'électrophilie expérimentale.

Pour aller plus loin dans cette étude et pour arriver à obtenir des descripteurs de réactivité qui correspondraient à une représentation plus *réaliste* de molécules en solution vis-à-vis de certains paramètres expérimentaux, on pourrait envisager d'étudier plus finement l'influence

du solvant et de la température, en ajoutant par exemple à la dynamique moléculaire une boîte de solvant explicite et de comparer l'ensemble des descripteurs de réactivité obtenus. On pourrait également envisager de réaliser une dynamique par approche hybride QM / MM (*quantum mechanics and molecular mechanics*) [1-4].

- [1] M. T. C. Martins-Costa, F. F. García-Prieto, M. F. Ruiz-López. *Org. Biomol. Chem.* 13 (2015), 1673–1679
- [2] J. L. Rivail, M. Ruiz-Lopez, X. Assfeld. *Quantum Model. Complex Mol. Syst.* (2015), 1–523
- [3] M. T. C. Martins-Costa, M. F. Ruiz-López. *Chem. Phys.* 332 (2007), 341–347
- [4] J. Oller, D. A. Saez, E. Vöhringer-Martinez. *J. Phys. Chem. A* 124 (2020), 849–857

### ***3. Prédire l'électrophilie expérimentale à partir de descripteurs quantiques et topologiques : une approche par apprentissage automatique***

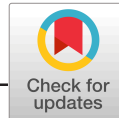
#### 3.1 Préambule

*Le travail réalisé ici a été mené en collaboration avec le groupe du LITIS (Laboratoire d'Informatique, de Traitement de l'Information et des Systèmes) (Sébastien Adam, Pierre Héroux, Benoit Gaüzere, Muhammet Balcilar) et le laboratoire COBRA (Guillaume Hoffmann, Vincent Tognetti, Laurent Joubert).*

Dans cette étude, nous avons choisi d'aller au-delà des méthodes de régressions linéaires classiques. En effet, on trouve dans la littérature seulement deux études qui utilisent des méthodes d'apprentissage automatique pour la prédiction de l'électrophilie expérimentale. Ces études se limitent en outre à l'utilisation de seulement trois descripteurs issus de la CDFT, ou à l'utilisation d'une base de données expérimentales réduite (voir réfs 39,40 de l'article).

Nous avons donc décidé dans cette nouvelle étude, de sélectionner un grand nombre de molécules disponibles dans la base de données de Mayr (près de 111 électrophiles soit 35% des données disponibles dans la littérature) afin d'utiliser ces méthodes non-linéaires. En effet, un modèle d'apprentissage automatique (qui est un sous-ensemble de l'intelligence artificielle) améliore sa performance en analysant une très grande quantité de données. Le modèle affine ses paramètres afin de s'ajuster aux nouvelles données qu'il reçoit, améliorant ainsi progressivement sa performance. Nous avons alors choisi de sélectionner des accepteurs de Michael (42 molécules dans cette étude) et des cations benzhydrylium (64 molécules) en supplément.

Nous avons ensuite calculé un nombre important de descripteurs pour chacune de ces molécules (en moyenne 700 descripteurs atomiques, moléculaires et topologiques par molécule) et mis au point un modèle de prédiction basé sur des méthodes dites d'apprentissage supervisé (un modèle d'apprentissage qui apprend une fonction de prédiction à partir d'exemples).



# Predicting experimental electrophilicities from quantum and topological descriptors: A machine learning approach

Guillaume Hoffmann<sup>1</sup> | Muhammet Balcilar<sup>2</sup> | Vincent Tognetti<sup>1</sup> |  
Pierre Héroux<sup>2</sup> | Benoît Gaüzère<sup>2</sup> | Sébastien Adam<sup>2</sup> | Laurent Joubert<sup>1</sup>

<sup>1</sup>Normandy Univ., COBRA UMR 6014 & FR 3038, Université de Rouen, INSA Rouen, CNRS, Mont St Aignan, France

<sup>2</sup>Normandie Univ., UNIROUEN, UNIHAVRE, INSA Rouen, LITIS, Rouen, France

## Correspondence

Vincent Tognetti, Normandy Univ., COBRA UMR 6014 & FR 3038, Université de Rouen, INSA Rouen, CNRS, 1 rue Tesnière 76821 Mont St Aignan Cedex, France.

Email: [vincent.tognetti@univ-rouen.fr](mailto:vincent.tognetti@univ-rouen.fr)

Benoît Gaüzère, Normandie Univ., UNIROUEN, UNIHAVRE, INSA Rouen, LITIS, Rouen 76000, France.

Email: [benoit.gauzere@insa-rouen.fr](mailto:benoit.gauzere@insa-rouen.fr)

## Abstract

In this paper, we assess the ability of various machine learning methods, either linear or non-linear, to efficiently predict Mayr's experimental scale for electrophilicity. To this aim, molecular and atomic descriptors rooted in conceptual density functional theory and in the quantum theory of atoms-in-molecules as well as topological features defined within graph theory were evaluated for a large set of molecules widely used in organic chemistry. State-of-the-art regression tools belonging to the support vector machines family and decision tree models were in particular considered and implemented. They afforded a promising predictive model, validating the use of such methodologies for the study of chemical reactivity.

## KEYWORDS

conceptual density functional theory, decision trees, electrophilicity, machine learning, nonlinear approaches, reactivity descriptors

## 1 | INTRODUCTION

Introduced in 1934 by Ingold,<sup>[1]</sup> the concepts of “electrophiles” and “nucleophiles”, which relate to molecules that have tendency to respectively accept or donate electrons, remain to this day cornerstones in experimental chemistry. They notably allow for efficiently predicting the outcome of a reaction, for accounting for reactivity trends between reagents, or for tackling regioselectivity issues within a unified rationalizing framework. Correlatively, the associated propensity of a molecule to accept/donate electrons was later defined as electrophilicity and nucleophilicity.

Thereafter, several methods were developed to quantify these concepts in an experimental way, and several scales were subsequently proposed, based, among others, on redox potentials,<sup>[2]</sup> hydrogen-bond stretching force constants,<sup>[3]</sup> or on rate constants. This last approach is actually the one recommended by the International Union of Pure and Applied Chemistry (IUPAC)<sup>[4]</sup> that stresses that even if electrophilicity is closely related to Lewis acidity (LA), LA “is measured by relative equilibrium constants” (this is a thermodynamical point of view), while electrophilicity “is measured by relative rate constants” (thus a kinetic property).

In particular, the scales stemming from kinetic studies have been the subject of many studies since the seminal work by Swain and Scott,<sup>[5]</sup> followed by those by Edwards<sup>[6]</sup> and Ritchie.<sup>[7]</sup> In this respect, the experimental scale proposed by Mayr's and coworkers in early 1990s<sup>[8]</sup> certainly stands as the most comprehensive one. Freely available online, it includes reagents used for a wide variety of polar organic reactions (such as Friedel-Crafts,<sup>[9,10]</sup> Michael additions,<sup>[11,12]</sup> Tsuji-Trost, Hosomi-Sakurai...), and is constantly updated and expanded. More precisely, this double scale is based on the following linear relationship:

$$\log k_{20^\circ\text{C}} = s_N(N + E). \quad (1)$$

where  $E$  and  $N$  denote, respectively, the electrophilicity and nucleophilicity parameters,  $k_{20^\circ\text{C}}$  the absolute rate constant of the reaction at  $20^\circ\text{C}$ , and  $s_N$  the nucleophile-specific sensitivity parameter.

Besides providing insight on the feasibility of reactions using the quick “ $N + E > -5$ ”<sup>[13]</sup> assessment, Mayr's scale also affords reliable ranking between reactants and is valuable to inspect selectivity in organic reactions.<sup>[14]</sup> Moreover, it can be applied to ambident compounds (this refers to molecules featuring two or more reactive sites),

since specific parameters can be experimentally derived for the different sites. This is in particular the case for the quinone family. For instance, in 2,5-dichloroquinone, there is one electrophilicity parameter associated to the  $C_{Cl}$  carbon (Mayr's experimental value:  $-16.11$ ) and one for the  $C_H$  atom ( $-12.28$ ).<sup>[15]</sup> Similar investigations were led by the Boubaker's group on 5-nitro-3-X-thiophenes,<sup>[16]</sup> demonstrating that kinetic measurements are instrumental in determining which site is the most reactive.

From a theoretical perspective, one of the main remaining challenges is to accurately estimate or predict these parameters on a given site. Such a model would give helpful insight to experimentalists, and will have strong impact on the development of new chemical reagents.<sup>[17,18]</sup> Up to now, the most predictive models have been those based on the calculation of methyl anion affinities (MAAs).<sup>[19,20]</sup> This promising approach however differs from that retained in this paper. In fact, MAA corresponds to the reaction energy of the various electrophiles with a unique selected nucleophile ( $CH_3^-$ ). It is thus not an intrinsic property of the electrophilic species (with reactive atom X), and its evaluation requires determining the addition product featuring the new X- $CH_3$  bond that is formed.

Interestingly, this is a thermodynamical descriptor and not a kinetic one. Note that if one assumes that all reactions follow Arrhenius' law,  $\log k_{20^\circ C}$  is proportional to the activation energy. So, the linear relationship observed in References 19 and 20 between MAA and Mayr's electrophilicity implies that reaction energy and activation energy are also linearly correlated. This would suggest that the Brønsted-Evans-Polanyi relations<sup>[21,22]</sup> are valid for these systems.

Here, we do not consider such assumptions, and we aim at theoretically predicting experimental values from the physico-chemical properties of the sole reactive species. Among the various possible approaches of reactivity in quantum chemistry, conceptual density functional theory (CDFT) and the several indices that come along<sup>[23–25]</sup> have attracted much interest over the last decades due to its deep physical roots. The CDFT proponents have regularly emphasized that, within this framework, many formal definitions of known empirical chemical concepts (such as electronegativity, chemical potential,<sup>[26]</sup> hardness and softness<sup>[27,28]</sup>) naturally emerged.

In the case of electrophilicity, the main CDFT reactivity index is certainly the global (i.e., molecular) index  $\omega$  defined by Parr et al.,<sup>[29]</sup> which was shown to be successful in retrieving Mayr's electrophilicity on series of molecules inside a given family or sharing close structures, such as diarylcarbenium<sup>[30]</sup> and diazonium<sup>[31]</sup> ions, benzyldenemalononitriles,<sup>[32]</sup> or Michael acceptors of similar formulas,<sup>[33]</sup> even if  $\omega$  was shown to be quite ineffective for this last family as a whole.<sup>[11]</sup> Nevertheless, when simultaneously dealing with different molecular families, the use of one single descriptor generally reveals insufficient to accurately predict Mayr's electrophilicity. In such cases, combinations of descriptors were shown to be quite effective,<sup>[33,34]</sup> even if only semi-quantitative.

In this work, we decided to focus on electrophiles. In this large group, up to 319 experimental parameters are currently determined and available in Mayr's online database. As already stated, most of the previous methods to predict experimental electrophilicities rely on

simple mono- or multi-linear regression analyses, which clearly exhibit prohibitive deficiencies when structures with too much variable chemical patterns are at stake.

For such reason, we looked into more advanced technics based on machine learning (ML). The use of this artificial intelligence subfield in theoretical chemistry is indeed currently booming, gaining more and more popularity in general physical chemistry.<sup>[35–38]</sup> Nevertheless, its application to the prediction of reactivity parameters still remains highly scarce, or far from being exhaustive. In particular, the use of ML in conjunction with state-of-the-art CDFT is even rarer, while this last theory has definitely proven its chemical significance to rationalize chemical reactions.

Indeed, previous works in this field were either centered on global reactivity descriptors omitting atomic descriptors (so that no discussion on ambident molecules is possible)<sup>[39]</sup> or using empirical molecular descriptors.<sup>[40]</sup> Conversely, we decided in this work to focus on global and atomic descriptors based on quantum chemistry, and in particular on CDFT, which could treat in principle all types of molecules (displaying one or several competing reactive sites) and that involve physico-chemical properties that convey clear chemical information.

From a methodological point of view, this article will also be grounded in state-of-the-art ML tools, in contrast with previous works based on "old-fashioned" approach (such as genetic algorithms). Our contribution also differs from studies in References 39,40. Indeed, in the last one, the authors, while also aiming at estimating electrophilicity, combined this objective with the quest for the smallest number of descriptors in order to keep an interpretative viewpoint. A complex processing pipeline was then proposed, involving feature selection in conjunction with regression methods, but each step was optimized separately, leading to a non-optimal model since the regressions do not take all available information nor their combination into account.

On the other hand, in Reference 39 the authors focused on obtaining a functional form of the electrophilicity through Genetic Programming and Grammatical Evolution only using ionization energy and electron affinity as descriptors. Such an approach offers a full understanding of the computation that was done, but at the expense of the estimation accuracy.

Our contribution here targets the opposite point of view by focusing on the quality of the electrophilicity prediction through the use of complex ML models without caring at first at their interpretability. However, as for any work involving ML, the choice of the dataset is crucial. In order to design a representative one that contains a large but reliable number of molecules, we hence picked all electrophiles available in Mayr's Database (319 electrophiles) and subjected them to successive selections to finally obtain a database made of 111 molecules.

To this aim, we first decided to take into account the number of reference nucleophiles used for the determination of electrophilicity as a fundamental criterion. Indeed, the nucleophile-specific sensitivity parameters  $s_N$  may be not accurately experimentally estimated when only one or two reference electrophiles are used. As a consequence, we decided to take only electrophilicity parameters determined with

at least three or more reference nucleophiles. By doing this, we retained only the most trustable experimental values. This actually corresponds to the "star classification system" in Mayr's database.

More precisely, the "one star" compounds have an experimental value determined roughly using only one nucleophile. This method (used for instance for ketones<sup>[20]</sup>) only provides broadly estimated experimental values.<sup>[21]</sup> As Mayr stated, they should be used only to "give a good idea about relative reactivities of strong nucleophiles towards weak electrophiles and of weak nucleophiles toward strong electrophiles."<sup>[13]</sup> Another important point is the fact that experimental values for one or two stars compounds are also subject to re-evaluation. For example, benzaldehyde has a previously reported electrophilicity value of  $-19.50$ <sup>[41]</sup> but was recently re-adjusted to  $-12.90$ .<sup>[20]</sup> For such reasons, we excluded all one and two stars compounds in order to use only reliable and accurate experimental values (even if the excluded molecules could have been used as an additional test set).

Since our method is based on the properties of the reactants that may depend on their environment, the solvent was also selected as a criterion: only parameters determined in dichloromethane were considered (they currently represent 73% of all the available experimental values). At last, the molecules containing metallic atoms were removed, giving us these 111 pure organic molecules (see list in Supporting Information). By doing so, this set contains respectively three different classes of electrophiles: 64 carbocations (with diarylcarbenium ions, indolyl-stabilized carbocations, *N*-substituted carbocations and other conjugated carbocations), 42 Michael acceptors (with benzylidenemalonates, bissulfonyl ethylenes, quinone methides, trans- $\beta$ -nitrostyrenes and other Michael acceptors), and 5 arenes (electron-deficient compounds). It can be noticed that either carbocations, Michael acceptors and arenes belong to the more general  $\pi$ -electrophiles category, which is of primary importance in organic synthesis.

In the next section, we have purposely chosen to present the quantum chemistry and the machine learning approaches in a pedagogical (even though not so concise) way to facilitate the dialogue between both research communities (ML and quantum chemistry ones) that may be not fully aware of the techniques used by each of them.

## 2 | COMPUTATIONAL DETAILS

### 2.1 | Molecular and atomic descriptors from quantum chemistry

The quantum chemical descriptors used in this study can be classified into three categories. The first one gathers *global CDFT* molecular descriptors: they give only one value for the whole molecule (they, thus, are not appropriate for the study of regioselectivity). In turn, they can be split into two subclasses: basic and composite ones.

Basic molecular descriptors are expressed as direct energy derivatives with respect to fundamental variables (the number of electrons

$N$  and the external potential  $v(\vec{r})$  in the case of the canonical ensemble), generally used up to the third order (other orders were in fact proposed but rarely considered).<sup>[42]</sup> They are epitomized by the electronic chemical potential  $\mu$  and the chemical hardness  $\eta$  (where  $E_e$  denotes the total electronic energy):

$$\mu = \left( \frac{\partial E_e[N, v(\vec{r})]}{\partial N} \right)_{v(\vec{r})}, \quad (2)$$

$$\eta = \left( \frac{\partial^2 E_e[N, v(\vec{r})]}{\partial N^2} \right)_{v(\vec{r})}, \quad (3)$$

The second subclass of global CDFT descriptors are coined "composite", because they are generated by combination of basic molecular descriptors. A notable example is the already mentioned Parr's electrophilicity index  $\omega$ .<sup>[29]</sup>

$$\omega = \frac{\mu^2}{2\eta}. \quad (4)$$

The second family of quantum descriptors considered in this study collects *atomic CDFT* descriptors. In CDFT, the functional derivatives of the molecular energy with respect to a local function such as the external potential engender functions of one real space variable (of the  $g(\vec{r})$  3D-type) or (when such differentiation is iterated) of several real space variables (such as  $h(\vec{r}, \vec{r}')$ ). These  $g(\vec{r})$  are called "local" functions, while  $h(\vec{r}, \vec{r}')$  are said to be "non-local" and are called "kernels". Obviously, local quantities are well suited to discriminate between various reactive sites on the same molecule since they encode spatial information. Representative local functions are the popular Fukui function  $f(\vec{r})$ <sup>[28]</sup> and dual descriptor  $f^{(2)}(\vec{r})$ <sup>[43]</sup>:

$$f(\vec{r}) = \frac{\delta^2 E_e[N, v(\vec{r})]}{\partial N \delta v(\vec{r})}, \quad (5)$$

$$f^{(2)}(\vec{r}) = \frac{(\delta^3 E_e[N, v(\vec{r})])}{(\partial N^2 \delta v(\vec{r}))}. \quad (6)$$

Following the same dichotomy, local functions can be divided into basic (this is the case for the Fukui function and dual descriptor mentioned above) or composite functions such as the multiphilic descriptor defined by<sup>[44]</sup>:

$$f_{\omega}^{(2)}(\vec{r}) = \omega f^{(2)}(\vec{r}). \quad (7)$$

Actually, while the multiphilic and dual descriptors "contain the same intramolecular reactivity information the former is expected to be a better intermolecular descriptor because of its global information content",<sup>[44]</sup> and it was successfully applied to describe (among others) the reactivity differences within ketones or amines reagents.

Finally, all these global or local CDFT descriptors can be evaluated within two different main frameworks: either using finite difference linearization (FDL) or within the frozen or frontier molecular orbital (FMO) approximation.

However, local functions actually embody infinite information (since defined at each real space point), while chemists are accustomed to think at the atomic level. An atomic *condensation* is thus required to translate the information gathered in local functions into the usual chemistry language and to afford a *coarse-grained* view of chemical reactivity. This implies that one defines what an atom is inside a molecule, an intricate and not univocal task.<sup>[45]</sup> In this study we will exclusively rely on the quantum theory of atoms in molecules (QTAIM) pioneered by Bader<sup>[46,47]</sup> that divides real space into non-overlapping atomic basins thanks to the zero-flux condition. Each chemical atom A is then associated to an atomic volume  $\Omega_A$  in real space. Given any local CDFT descriptor, an atomic value is simply obtained by integration over this basin:

$$g(A) = \int_{\Omega_A} g(\vec{r}) d^3r. \quad (8)$$

The third and last descriptor category we will consider corresponds to *pure* QTAIM atomic descriptors, which are not defined within CDFT, but only within QTAIM, such as the atomic localization index (LI), atomic Thomas-Fermi kinetic energy, virial or interacting quantum atoms (IQA) atomic energies. We refer the interested reader to our recent paper<sup>[48]</sup> for an extensive description of these QTAIM energy quantities and their use in understanding chemistry.

Please note that the comprehensive list of all considered descriptors can be found in Supporting Information (see Tables S1–S3).

## 2.2 | Quantum chemistry protocol

Initially, a search of the most stable conformer was performed for all molecules using the Balloon software<sup>[49]</sup> with the MMFF94 force field. Full geometrical optimizations were then carried out at the M06-2X/6-311++G(2d,2p) level of theory with the Gaussian 09 software,<sup>[50]</sup> using the “ultrafine” integration grid. Solvent effects (dichloromethane) were described by means of a polarizable continuum model (namely, IEFPCM<sup>[51]</sup>). QTAIM calculations were achieved with the AIMAll software<sup>[52]</sup> using the “M06-2X model” tag to ensure correct energy decompositions.

Homemade scripts were then used to extract and calculate all atomic and molecular descriptors. More technical details about the quantum chemistry protocol could be found in our previous study.<sup>[33]</sup>

## 2.3 | Topological descriptors from graph theory

In order to enlarge the scope of chemical descriptors, we also considered topological features computed on the molecular graph associated to each molecule. Note that the term “topological” is here to be understood in its meaning in graph theory and not within the

framework of Quantum Chemical Topology (QCT),<sup>[47]</sup> a full branch of theoretical chemistry that subsumes QTAIM.

The topological features here present the advantage of being very quickly estimated since they do not require any quantum chemistry computations, but only the Simplified Molecular Input Line Entry Specification (SMILES) code of the molecule. More specifically, we used the set of extracted treelets,<sup>[53]</sup> which describes a molecular graph as a bag of patterns, these patterns corresponding to all labeled subtrees having at most six nodes. Such a topological approach was already successful in chemistry to predict the boiling point, the optical rotation angle or to evaluate toxicity.<sup>[54–56]</sup>

Considering our dataset, 864 different treelets have been extracted. To this aim, we first created the ChemDraw Table (CT) format of each molecule that shows atom name, atom-to-atom connections and connection types. Then we fed the “moleculesTreeletData.cpp” code in given repository<sup>[57]</sup> by this CT formatted dataset and took 864 number of bins histogram for each molecule.

## 2.4 | Molecular classes

Finally, to these quantum and topological descriptors, we had an additional feature that refers to the electrophile type. Indeed, according to Mayr, the whole set can be divided into three families: carbocations, Michael acceptors, and arenes. We thus defined three descriptors, coined CV, MV and AV that will indicate to which category a given molecule belongs to. For instance, CV (resp. MV) will be equal to 1 if the molecule is a carbocation (resp. a Michael acceptor) and to 0 if it is not.

## 2.5 | General machine learning methodology

ML gathers different frameworks having a same core principle: learning from examples. More specifically, according to Mitchell,<sup>[58]</sup> “a computer program is said to learn from experience  $E$  with respect to some class of tasks  $T$  and performance measure  $P$ , if its performance at tasks in  $T$ , as measured by  $P$ , improves with experience  $E$ ”. Generally,  $T$  is defined as a prediction task, and  $P$  is used to know how accurate the prediction is. Finally, experience  $E$  can be understood as a bunch of paired data where each data is associated to its particular property.

Most of ML methods operate on Euclidean spaces. The data is then generally represented by a matrix  $X \in \mathbb{R}^{N \times d}$  where each line noted  $x_i$  corresponds to a particular data encoded by  $d$  values, also referred as characteristics, descriptors or features. The number of lines  $N$  corresponds to the number of data available in the dataset. The property to predict, encoded by a vector  $Y \in \mathbb{R}^N$ , characterizes the kind of prediction task.

First, classification problems correspond to a prediction problem where each available data is assigned to a particular class, for example,  $Y \in \mathbb{N}^N$ . They can be classification of images (do they contain a cat or a dog?), drugs (are they toxic or not?). Second, regression problems aim to assign a continuous value to each data:  $Y \in \mathbb{R}^N$ . They include

the prediction of a boiling point of a molecule,<sup>[59,60]</sup> the price of a given house.

Once the prediction problem has been established, that is, we identified a matrix  $X$  and a corresponding vector of properties  $Y$ , the learning process may find a function  $f: \mathbb{R}^d \rightarrow y$  that approximates the property associated to the data. Such learning process is generally performed by minimizing the following function:

$$f^* = \arg\min_f \left\{ \sum_{i=1}^N \mathcal{L}(f(x_i), y_i) + \lambda \Omega(f) \right\}, \quad (9)$$

where  $\mathcal{L}: y \times y \rightarrow \mathbb{R}$  is a loss function encoding the ability of  $f$  to fit the pairs of data-property. Considering regression problems, this function may correspond to Square Loss, Mean Absolute Error (MAE). The second term, weighted by a hyperparameter  $\lambda$ , penalizes some  $f$  functions having undesired properties (for instance non-smoothness) through<sup>[61]</sup>  $\Omega: (X^d \rightarrow y) \rightarrow \mathbb{R}$ . For example, a classical choice is the Euclidean norm, which promotes smooth vectors, that is, homogeneous vectors rather than vectors composed of very different numbers. Note that any penalty can be associated to this approach, depending on the a priori information about the task to predict. The  $\lambda$  value allows controlling the influence of this penalty relatively to the loss with respect to the data term.

Particular well-known problems in ML are overlearning and lack of generalization. For instance, the model may predict very accurately on  $X$  but fails when exposed to unknown data during learning process, while its main purpose is to predict unseen data. To overcome these issues, the dataset is generally divided into three disjoint subsets: training, validation and test subsets. The first one is employed to learn an optimal model  $f^*$ , and the second to estimate its performance on data not taken into account during the learning process. This performance is also used to pick the best  $\lambda$  value across a predefined range. Finally, to avoid an optimistic performance measure for a given  $\lambda$  value, we measure the final performance on the test subset that has not been used during the whole process. Repeating these divisions allows averaging the random fluctuations due to particular data subsets. Different strategies actually exist to split the initial data, such as, among others,  $k$ -Fold Cross Validation,<sup>[62–64]</sup> Random Sampling, and Leave one out.

In this article, we choose several state-of-the-art ML methods to work on these sets. They can be roughly divided into three categories: linear-based regression, non-linear regression and decision tree methods. One may notice that deep learning methods are here missing although they represent methods of choice in other contexts. This is due to the very limited number of data (111) available to learn a neural network.

First, a linear regression (LR) consists in finding a vector of coefficients  $w \in \mathbb{R}^d$  so that the loss function is defined by the following Euclidean norm:

$$\mathcal{L}(Xw, Y) = \|Xw - Y + b\|_2^2. \quad (10)$$

As previously discussed, we can add a regularization operator. In the case of Lasso Regression (LaR),<sup>[65]</sup> we penalize vectors  $w$  having a

high usual  $L_1$  ("taxicab") norm, and thus we favor vectors having a lot of null values:

$$w^* = \arg\min_w \left\{ \|Xw - Y + b\|_2^2 + \lambda \|w\|_1 \right\} \quad (11)$$

Another way to increase the generalization ability is to reduce the degrees of freedom, that is, the number of features encoding each data. One method of choice to achieve it is Principal Component Analysis (PCA),<sup>[66]</sup> which represents the matrix  $X \in \mathbb{R}^{N \times d}$  into a new matrix  $X_p \in \mathbb{R}^{N \times d'}$  with  $d' < d$ , encoding as much information as possible. This is achieved by computing a projection matrix  $P \in \mathbb{R}^{d \times d'}$  such as  $XP$  maximizes the explained variance. The PCA regression is then defined by:

$$w^* = \arg\min_w \left\{ \|XPw + b - y\|_2^2 \right\}. \quad (12)$$

Such linear approaches are quite easy to implement but are intrinsically limited. For instance, if the relationship between the data and the property to predict is quadratic, no linear model can be adequate. To overcome this limitation, one method of choice is the family of Support Vector Machines (SVM),<sup>[67]</sup> initially defined as linear prediction models that maximize the margin separating two classes. In the regression version of this approach, noted SVR, we aim to find a hyperplane including all data points within an epsilon-insensitive ( $\epsilon$ ) tube

$$w^* = \arg\min_w \left\{ \frac{1}{2} \|w\|_2^2 \right\} \quad \text{subject to:} \quad (13)$$

$$|Y_i - w^T x_i + b| < \epsilon, \quad \forall i \in \{1, \dots, N\}$$

Thanks to kernel theory (not to be confused with CDFT kernels) and the representer theorem,<sup>[62]</sup> the hyperplane can be expressed by a linear combination of similarities functions on input data, so that the prediction of a new data  $x$  is equal to:

$$f(x) = \sum_{i=1}^N \alpha_i Y_i k(x, x_i), \quad (14)$$

where  $\alpha_i$  represents the optimal dual variables and  $k: X \times X \rightarrow \mathbb{R}$  is a particular function called kernel corresponding to a scalar product between two data embeddings into an Hilbert space. If the embedding underlying the kernel function is not linear, then the decision function is not linear in the input space. Many kernels have been proposed in the literature. Here, we will focus on two of them, namely the Tanimoto<sup>[68]</sup> (that proved its efficiency in bioinformatics and chemoinformatics as such for the prediction of toxicity in the rat and mutagenicity in Salmonella<sup>[69–72]</sup>) and the well-known Radial Basis Function (RBF) Gaussian kernels<sup>[66]</sup>:

$$k_{\text{Tanimoto}}(x_i, x_j) = \frac{x_i^T x_j}{x_i^T x_i + x_j^T x_j - x_i^T x_j}, \quad (15)$$



$$k_{\text{RBF}}(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|_2^2}{2\sigma^2}\right), \quad (16)$$

where  $\sigma$  controlling the distance of influence of each data.

Another method called Gaussian Process Regression (GPR)<sup>[73]</sup> similarly aims to extend standard linear regression scheme to higher dimensional spaces through the use of a kernel function. This method assumes that the observed properties include some Gaussian noise in addition to the contributions of descriptors and associated coefficients. GPR has this useful ability to provide uncertainty measurements on the predictions and works empirically well on small datasets. This method includes two hyperparameters: the kernel used to compute the covariance and the amplitude  $\varepsilon$  of the noise characterized by the variance  $\sigma$ .

Finally, decision trees<sup>[74]</sup> use a tree-like model to design a prediction function. The strategy of Decision Trees is to split the input data into  $M$  homogeneous regions and to fit a particular linear prediction model  $f_m$  for each region  $R_m$  in  $\{1 \dots M\}$ . Then the prediction function  $f$  is defined as a combination of these local  $f_m$  prediction functions:

$$f(x_i) = \sum_{m=1}^M f_m(x_i) I(x_i \in R_m). \quad (17)$$

where  $I(\cdot)$  equals 1 if the input statement is true or 0 otherwise, and  $M$  is an hyperparameter to tune according to dataset. The tree-like structure is modeled by successive splits of data, leading to  $M$  leaves coding the final regions where local models are fitted. One particular interest of these methods resides in their human interpretability, close to human thinking. To add more learning capacity, an extension consists in considering ensembles of independent decision trees whose decisions are merged together. These methods are then denoted as Additive Trees or Random Forests. However, these extensions affect the human interpretability of the prediction model.

In this framework, one method of choice is gradient boosting, denoted XGBoost,<sup>[74,75]</sup> which has recently been dominating applied ML challenges for structured or tabular data<sup>[76,77]</sup> and whose implementation is very effective,<sup>[78]</sup> and thus was successfully applied in various fields.<sup>[79,80]</sup> It can also be noticed that XGBoost is able to deal with missing values (even if it is not the case here), which expands the application domain of this method. This method will be denoted by Gradient Boosting Decision Tree (GBDT) within the results section.

## 2.6 | Machine learning protocol

All input features were initially normalized (0 for mean and 1 for standard deviation) according to the training set. Lasso regressions were performed using Matlab's built-in implementation,<sup>[81]</sup> while Kernel Methods were based on the libsvm Matlab library with our own implementation from scratch of the kernels. For the GBDT algorithm, we used the Python implementation of XGboost library<sup>[82]</sup> with

default parameters, except for the number of decision trees chosen to 10,000. Shapley additive explanation values are calculated in Python.<sup>[83]</sup> GPR was performed using the GaussianProcessRegressor in the scikit-learn library.<sup>[84]</sup>

Given results for the test set have been computed using  $k$ -fold cross validation model validation technique as depicted in Figure 1. This protocol has been repeated 100 times for each algorithm. The hyperparameters (variance for PCA regression,  $\lambda$  for Lasso regression,  $\sigma$  for RBF kernels and  $C$  for SVR) were optimized through a grid search and the performance of each hyperparameter set was estimated on a validation set.

To apply this protocol, we first randomly divided our data into 10 folds and we assigned one fold as test set and the remaining folds as training sets. As recalled by Kuhn and Johnson,<sup>[85]</sup> there is actually no formal rule for the choice of  $k$ , but "as  $k$  gets larger, the difference in size between the training set and the resampling subsets gets smaller. As this difference decreases, the bias of the technique becomes smaller", so that "there is a bias-variance trade-off associated with the choice of  $k$  in  $k$ -fold cross-validation. Typically, given these considerations, one performs  $k$ -fold cross-validation using  $k = 5$  or  $k = 10$ , as these values have been shown empirically to yield test error rate estimates that suffer neither from excessively high bias nor from very high variance."<sup>[86]</sup>

Then, the performance of each hyperparameter configuration was estimated on one fold (the validation set) among the nine remaining, the eight other folds (learning sets) being used as learning data. To have a better estimation of the performance, each fold is successively assigned to the validation set, leading to nine different pairs of learning/validation sets and thus nine different estimations of the performance for one hyperparameter configuration. Best hyperparameter configuration was determined by the one having the best average performance over the nine validation sets.

Finally, performance on test set is obtained after training our model on the nine training sets using the best hyperparameters and predicting the test set. It is worth noticing that our test set has not been used through the whole learning stage, including the learning of optimal hyperparameters. This may prevent some overfitting bias that may lead to optimistic performances.

## 3 | RESULTS AND DISCUSSION

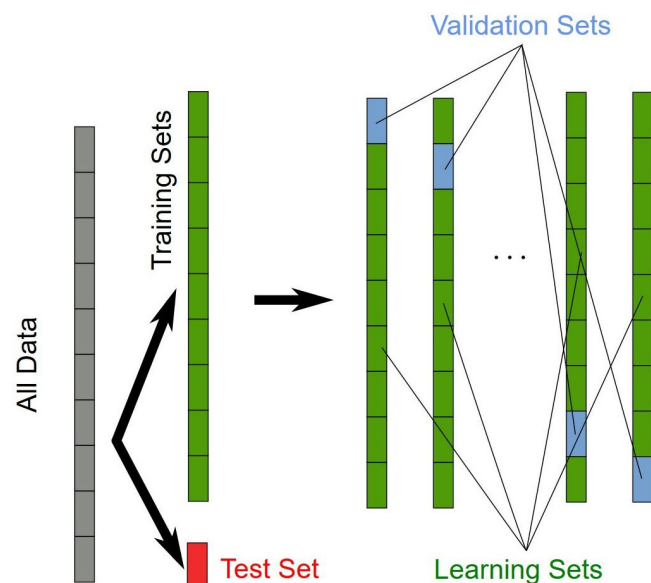
In our approach, each molecule can be represented by different levels of description. As previously stated, there are 19 global features consisting of 16 CDFT molecular descriptors (see Section 2.1) and the 3 family codes (see Section 2.4). As explained in Section 2.3, we evaluated 864 topological features. The case of *atomic* descriptors (defined within QTAIM, see Section 2.1) deserves more caution when included in a ML process. One choice is to use only information related to the reactive atom, that-is-to-say the atom (highlighted by a circle in Table S4) that will be bonded to the nucleophile at the end of the reaction. This atom is formally identified in Mayr's online database, and it is described by a 50 length feature vector.

Nevertheless, it is expected that the properties of other atoms can considerably impact the nucleophilic attack process. For instance, in case of Michael additions, we recently provided compelling evidence<sup>[33,34]</sup> that the properties of the carbon atom that is adjacent to the reactive one are of fundamental importance to retrieve experimental data. Indeed, this is the full C=C double bond that reacts

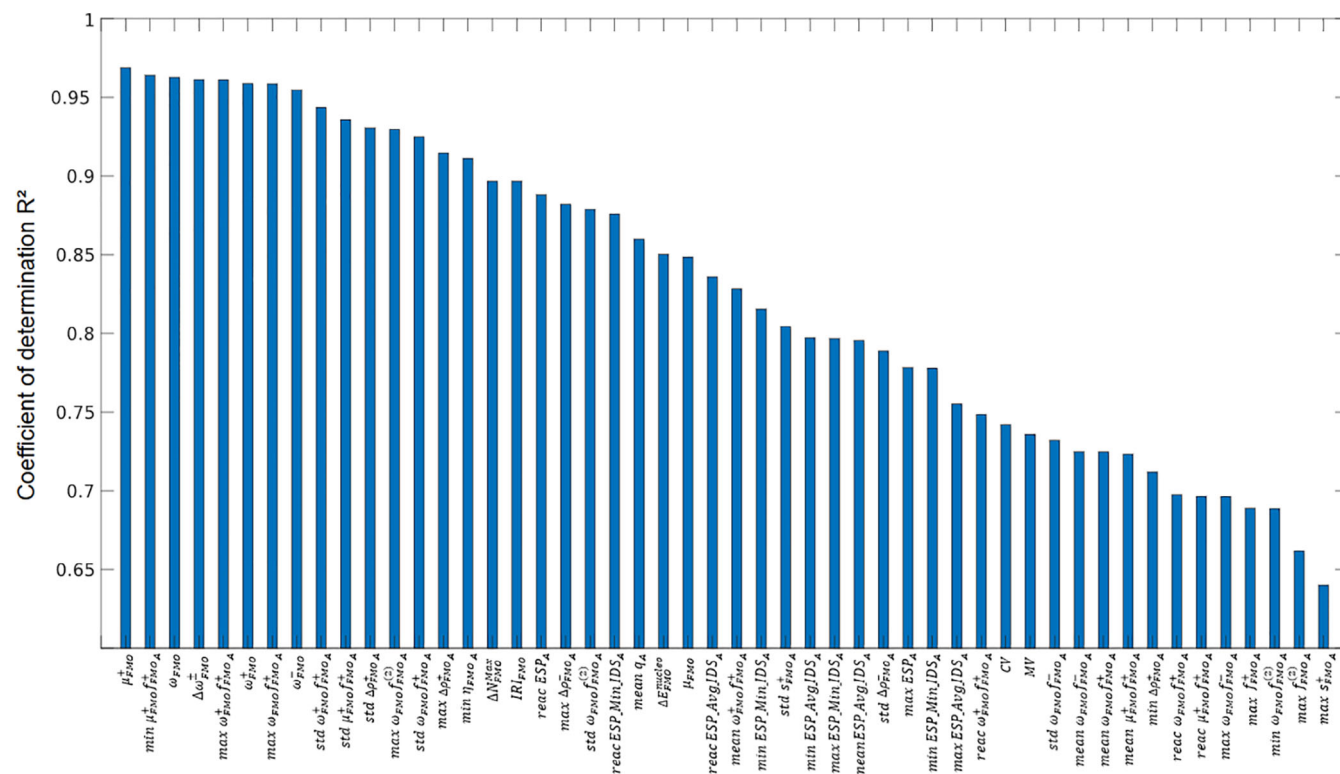
during such chemical reactions. More generally, nucleophilic additions are characterized by substantial electron transfer (ET) from the nucleophilic species to the electrophilic one. This ET does not involve only the reactive atom, but it is spread over the whole molecule since electrons are delocalized in the entire space. Besides, it is known that many chemical reactions are affected by secondary interactions that may increase (in case they are stabilizing) or decrease (if destabilizing) significantly rate constants, as illustrated by the *endo* preference in Diels-Alder cycloadditions. For such reasons, atomic information for all atoms may be meaningful to account for the electrophilicity measured at a given atomic site.

However, molecules are of various sizes, while traditional ML methods require a fixed dimension to represent data. Thus, to overcome the issue rose by this variable number of atoms, we chose to aggregate any of the 50 atomic features into four simple statistical descriptors, namely minimum, maximum, mean, and standard deviation values. Such operation led to a fixed length (200) atomic feature vector for each molecule. Even if such treatment induced loss of information, it is regularly used in other fields, since, for instance, wavelet features are computed this way in image processing,<sup>[87]</sup> and a similar aggregation approach in graph theory has also received large attention.<sup>[88]</sup> These examples actually exhibited a good tradeoff between the induced loss and the need of being independent of the number of dimensions.

With all the descriptors in hand, we then calculated the 50 most correlated features (in the sense of monolinear regressions) to the actual Mayr's electrophilicity value, as shown in Figure 2. According to



**FIGURE 1** Data handling for machine learning [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



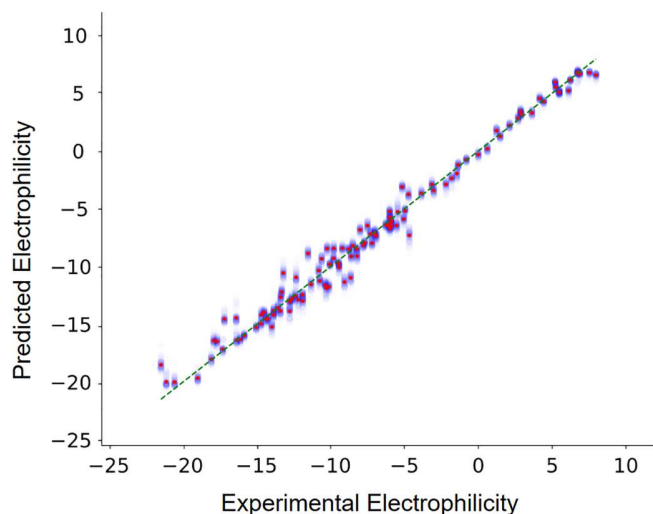
**FIGURE 2** Top 50 for the most correlated features in terms of  $R^2$  values [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

**TABLE 1** Best models for the prediction of Mayr's electrophilicity in terms of mean absolute error (MAE) and  $R^2$  values (for the test sets) according to the prediction model and the selected descriptors

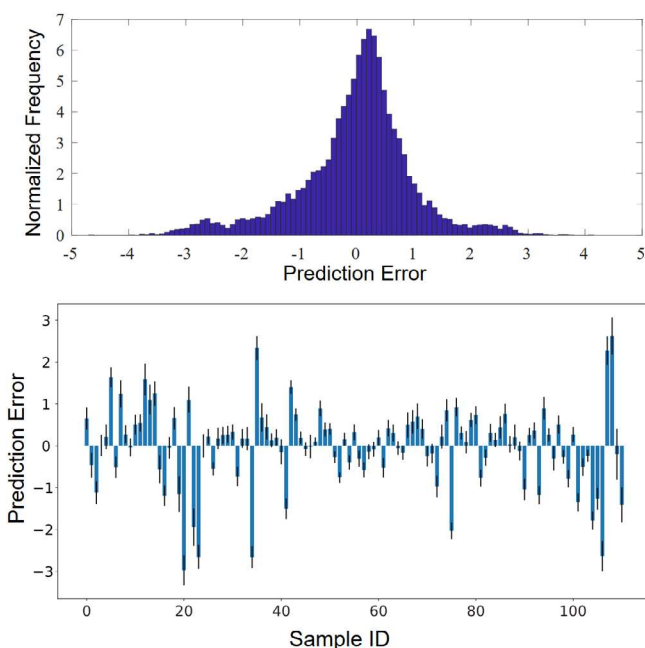
| Entry | Descriptors                | Perf. | LR            | PCA           | LaR           | GPR           | RBF           | Tanimoto      | GBDT                 |
|-------|----------------------------|-------|---------------|---------------|---------------|---------------|---------------|---------------|----------------------|
| 1     | G                          | MAE   | 1.27 ± 0.02   | 1.60 ± 0.09   | 1.45 ± 0.01   | 1.27 ± 0.02   | 1.20 ± 0.07   | 1.31 ± 0.11   | 1.37 ± 0.04          |
|       |                            | $R^2$ | 0.949 ± 0.015 | 0.924 ± 0.007 | 0.936 ± 0.001 | 0.948 ± 0.002 | 0.947 ± 0.007 | 0.928 ± 0.023 | 0.946 ± 0.002        |
| 2     | $A_{\text{react}}$         | MAE   | 1.43 ± 1.07   | 1.34 ± 0.08   | 1.39 ± 0.12   | 0.96 ± 0.01   | 0.82 ± 0.07   | 0.96 ± 0.08   | 1.13 ± 0.03          |
|       |                            | $R^2$ | 0.840 ± 0.274 | 0.943 ± 0.020 | 0.930 ± 0.041 | 0.948 ± 0.041 | 0.965 ± 0.012 | 0.950 ± 0.011 | 0.950 ± 0.030        |
| 3     | $A_{\text{stat}}$          | MAE   | —             | 1.37 ± 0.12   | 1.29 ± 0.03   | 1.19 ± 0.05   | 1.13 ± 0.15   | 1.03 ± 0.06   | 0.96 ± 0.03          |
|       |                            | $R^2$ | —             | 0.913 ± 0.051 | 0.951 ± 0.004 | 0.953 ± 0.009 | 0.915 ± 0.060 | 0.937 ± 0.012 | 0.976 ± 0.002        |
| 4     | T                          | MAE   | —             | 2.91 ± 0.23   | 2.88 ± 0.13   | 2.49 ± 0.09   | 1.71 ± 0.07   | 1.71 ± 0.07   | 1.78 ± 0.06          |
|       |                            | $R^2$ | —             | 0.709 ± 0.039 | 0.715 ± 0.023 | 0.756 ± 0.019 | 0.893 ± 0.009 | 0.885 ± 0.006 | 0.934 ± 0.005        |
| 5     | G + $A_{\text{react}}$     | MAE   | —             | 1.05 ± 0.13   | 1.04 ± 0.06   | 0.76 ± 0.04   | 0.74 ± 0.09   | 0.78 ± 0.05   | 0.85 ± 0.02          |
|       |                            | $R^2$ | —             | 0.956 ± 0.020 | 0.964 ± 0.015 | 0.972 ± 0.003 | 0.968 ± 0.022 | 0.972 ± 0.004 | 0.973 ± 0.002        |
| 6     | G + $A_{\text{stat}}$      | MAE   | —             | 1.18 ± 0.13   | 0.87 ± 0.10   | 0.98 ± 0.04   | 0.96 ± 0.14   | 0.90 ± 0.06   | 0.76 ± 0.03          |
|       |                            | $R^2$ | —             | 0.927 ± 0.068 | 0.967 ± 0.051 | 0.962 ± 0.006 | 0.943 ± 0.055 | 0.948 ± 0.012 | 0.980 ± 0.001        |
| 7     | G + $A_{\text{stat}}$ + T  | MAE   | —             | 0.92 ± 0.09   | 0.76 ± 0.09   | 0.82 ± 0.08   | 0.92 ± 0.11   | 0.87 ± 0.06   | 0.73 ± 0.03          |
|       |                            | $R^2$ | —             | 0.956 ± 0.009 | 0.974 ± 0.043 | 0.964 ± 0.011 | 0.947 ± 0.024 | 0.950 ± 0.012 | 0.981 ± 0.001        |
| 8     | G + $A_{\text{react}}$ + T | MAE   | —             | 0.85 ± 0.06   | 0.80 ± 0.03   | 0.78 ± 0.04   | 0.85 ± 0.08   | 0.75 ± 0.04   | 0.75 ± 0.03          |
|       |                            | $R^2$ | —             | 0.969 ± 0.005 | 0.979 ± 0.002 | 0.971 ± 0.004 | 0.955 ± 0.013 | 0.975 ± 0.003 | 0.978 ± 0.002        |
| 9     | G + $A_{\text{max}}$ + T   | MAE   | —             | 0.95 ± 0.13   | 0.73 ± 0.03   | 0.82 ± 0.04   | 0.95 ± 0.10   | 0.73 ± 0.05   | <b>0.72 ± 0.03</b>   |
|       |                            | $R^2$ | —             | 0.951 ± 0.069 | 0.982 ± 0.002 | 0.967 ± 0.003 | 0.942 ± 0.025 | 0.974 ± 0.006 | <b>0.982 ± 0.001</b> |

Note: Methods: LR, linear regression; PCA, principal component analysis; LaR, lasso regression; GPR, Gaussian process regression; RBF, radial basis function Gaussian kernel; GBDT, gradient boosting decision tree. MAE, mean absolute error. Descriptors: G, global;  $A_{\text{react}}$ , atomic on only reactive atoms;  $A_{\text{stat}}$ , aggregated statistics on all atoms; T, topological.

Bold values correspond to the best model.



**FIGURE 3** Predicted versus actual values for Mayr's electrophilicity [Color figure can be viewed at [wileyonlinelibrary.com](#)]



**FIGURE 4** Error distribution of the best prediction model for Mayr's electrophilicity [Color figure can be viewed at [wileyonlinelibrary.com](#)]

this analysis, the vertical electron affinity  $\mu_{\text{FMO}}^+$  global descriptor is the most correlated feature ( $R^2 = 0.969$ ), which is not surprising since it is equal to the energy of the lowest unoccupied Kohn-Sham molecular orbital (LUMO). More generally, top 50 consists of 11 global and 39 atomic features, but no topological features since the most correlated one lies in the 59th position. This observation notably underlines the importance of atomic features (even if used within simple statistics).

Table 1 now shows the general performances of the various models in terms of  $R^2$  and mean absolute error (MAE) values with respect to Mayr's experimental values. Let us recall that these

performances were measured only on the test sets, and thus they are related to molecules that were explicitly excluded from the training process. We actually tested various complexity levels: models with only global (labeled "G") descriptors (entry 1), models with only atomic ones using either the reactive atom ( $A_{\text{react}}$  approach, entry 2) or the aggregated statistics ( $A_{\text{stat}}$ , entry 3), models with only topological ones (entry 4), models with both global and atomic features (entries 5 and 6), and finally models with all (global + atomic + topological) descriptors (entries 7, 8 and 9).

It is interesting to notice that the two ways for using atomic information revealed strongly unequivalent. For instance, using RBF, the  $A_{\text{react}}$  descriptors (entry 2) afford a better model than  $A_{\text{stat}}$  (entry 3), with MAE equal to 0.82 and 1.13, respectively. The opposite trend was instead found using GBDT (the MAEs being, respectively, 1.13 and 0.96).

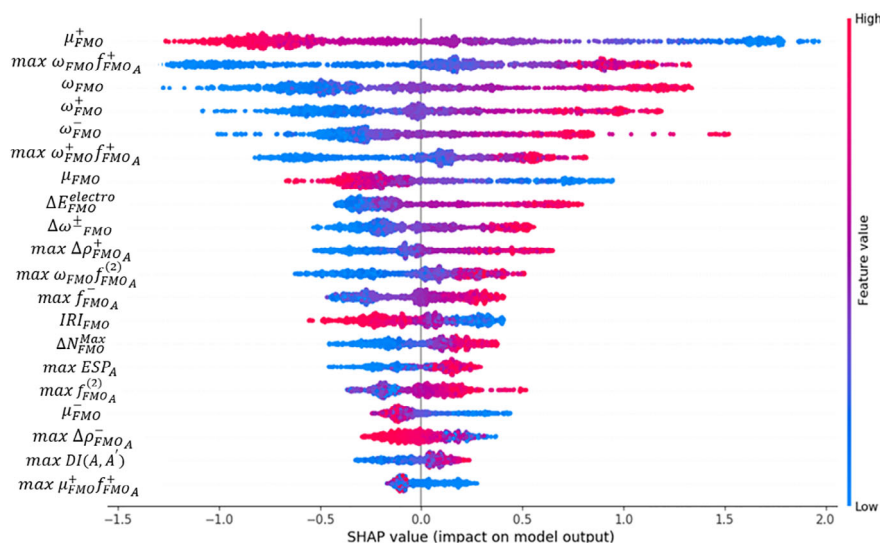
Another observation deserving to be discussed is that, while some models include more descriptive information, their final performance may be counter-intuitively overperformed by simpler models, that is, taking into account less descriptive features. This is for instance the case when comparing the RBF performances with  $A_{\text{react}}$  (MAE = 0.82) and with the considerably larger  $\{G + A_{\text{react}} + T\}$  feature set (MAE = 0.92). This phenomenon is actually related to overfitting. Indeed, considering a lot of descriptors of some data may allow to perfectly predict properties of this particular set of data, all known during the learning stage. However, this learning may be so perfect that it learns specificities of this particular set of data that are not encountered in other sets. Thus, the prediction of unseen data may suffer from these particularities, leading to a worse performance.

All in all, according to our tests, gradient boosting decision tree with all extracted features clearly outperformed all other trials. The corresponding code is freely available online.<sup>[89]</sup> Besides, even if performance of only topological features is the worst (and even worse than baseline linear regression using global features), adding them significantly increased the performance of GBDT (the MAE going from 0.76 to 0.71). Although Kernel based methods did not reach the top of Table 1, this does not mean that they do not work well on that kind of problem. In fact, they are especially useful when dealing with topological information.

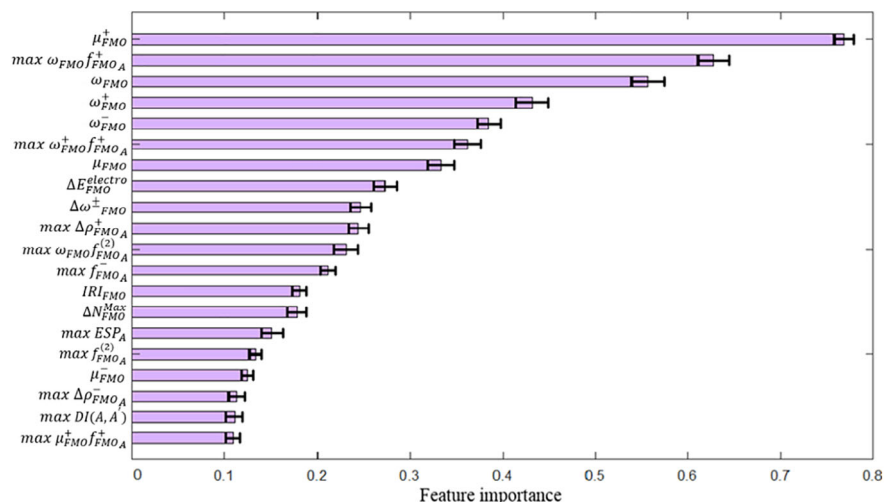
It can also be noticed that the GPR approach (which can be suitable in principle for small datasets as it is the case here) performs quite well on many descriptor combinations, but it never reaches top accuracy in any descriptor set. Here, the difficulty of a prior modeling of the noise may alter the prediction results.

Figure 3 now represents the scatter plot for the best method prediction (see Table S5 for the corresponding mean and standard deviation values over 100 runs). In this plot, red points are the mean prediction over 100 random runs, while blue points show the prediction result in each run. Finally, the blue color distribution indicates how much frequently this prediction is made by GBDT. It can be observed that, in most cases, the prediction was satisfying. It is also instructive to decompose the performances of this best model according to chemical families. Actually, we found that the MAE value

**FIGURE 5** SHAP values of the 20 most important features for the best prediction model of Mayr's electrophilicity [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 6** Feature importance for the best prediction model of Mayr's electrophilicity according to the SHAP analysis [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



was equal to 0.50 for the 64 carbocations, 1.90 for the 5 arenes (which is certainly not statistically relevant), and 0.83 for Michael acceptors. This last value can be compared to our best model previously reported: 1.30,<sup>[33]</sup> pointing out the strong improvement obtained by using modern ML methods.

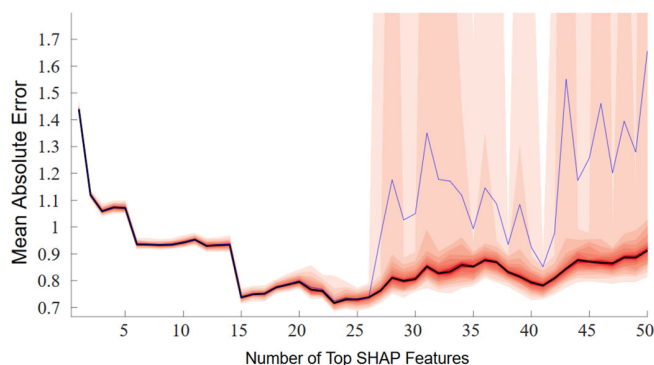
Additionally, the error distribution over 100 random runs is then depicted in Figure 4: the left plot shows distribution of the error regardless of the sample number over all runs, while the right plot shows the mean error and standard deviation for each sample. The mean value of the left plot is  $-0.02$  and its standard deviation is 1.00. We also found that there is no significant relation between the errors and any collected feature since the maximum absolute correlation between error and feature is not higher than 0.1.

Nevertheless, despite these promising results, interpreting input features still remains important if the model is expected to become more understandable to human understanding, making it gray box instead of black box. To the best of our knowledge, the most suitable way for interpretation of features from tree ensemble is SHapley Additive exPlanation (SHAP).<sup>[90,91]</sup> Figure 5 shows the SHAP

interaction values distribution over the most effective top 20 features: each point in each line shows one sample in the dataset, and the color of the points refers to the normalized value of the corresponding features.

Since the bigger the mean absolute value of the feature's SHAP, the more important that feature, we calculated the mean absolute value of each feature over 100 random runs, and we show the importance of the most important 20 features as well as their standard deviation of importance in Figure 6.

Noteworthy, in the top 20 list, there are 10 global features and 10 atomic features without topological features (the first one appearing at the 37th position). This last observation is perfectly in line with the fact that chemical reactivity is essentially of quantum nature. According to this analysis, the vertical electron affinity  $\mu_{FMO}^+$  global feature, which was already found to be the most correlated one, was again identified as the most important for the best gradient boosting decision tree model. Noticeably, the second most relevant descriptor is the maximal value over all atoms of the atomic  $\omega_{FMO} f_{FMO}^+$  (composite electrophilic Fukui function) descriptor.



**FIGURE 7** Performances of multilinear regressions after feature selection by GBDT [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

In summary, CDFT indices (either global or local) involving Parr's electrophilicity index  $\omega$  demonstrated their predictive capacity towards Mayr's experimental electrophilicities parameters as 7 of the top 10 most relevant descriptors of our GBDT model are based on a simple representation of the energy by a parabola. It is interesting to note that most of the other present descriptors are as expected related to electrophilicity according to CDFT. For instance, the maximal values for the dual descriptor  $f_{\text{FMO}}^{(2)}$  which is here positive as expected and thus describing electrophilicity, lies in 16th position,<sup>[43]</sup> the maximal local atomic electrophilicity<sup>[92]</sup>  $\Delta\rho$  in 10th and 18th positions. Noteworthy, in 12th position we have the presence of  $f^-$  in our model. This Fukui function incorporates aspects related to electron donation process (hence, nucleophilicity), but Fuentealba et al.<sup>[93]</sup> recently suggested that this descriptor could be included in a new formula proposal for a local electrophilicity index.

From a more chemical point of view, one could classically expect Pearson's hard and soft acids and bases (HSAB) theory<sup>[94]</sup> or the Klopman-Salem scheme<sup>[95,96]</sup> to be the results of our model. This would mean that the model would involve a combination of a so-called "charge control" descriptor and an "orbital controlled" descriptor (resulting in a single expression similarly to the general purpose of reactivity indicator<sup>[97]</sup>). We note that our results do not provide enough evidence to confirm these theories.

This statement actually confirms the trend that descriptors following HSAB and Klopman-Salem principles were shown<sup>[98]</sup> to be inappropriate towards a structurally diverse set of compounds and especially for molecules with two or more reactive sites (ambident reagents),<sup>[99,100]</sup> leading to Mayr's conclusion that these theories should be abandoned.<sup>[101]</sup>

The SHAP analysis can also be used as a feature selector for multilinear regressions (MRs). Let us consider that we want to perform a ML with  $k$  descriptors. These ones will be those featuring the  $k$  highest SHAP values. Figure 7 presents the 10-fold test MAE values with respect to  $k$ . According to this approach, MAE reaches  $0.72 \pm 0.03$ , with  $R^2 = 0.976 \pm 0.004$  when the top 23 SHAP features were selected as inputs. Interestingly, if more than 26 features are used, the model becomes more unstable, since the differences

between mean and median values of MAE increases. This 23-linear model can be seen as a kind of light version of the best model of GBDT.

## 4 | CONCLUSION

The prediction of the electrophilicity value of chemical reagents is of utmost importance in organic chemistry and is still a challenge in computational chemistry. Being able to get rid of long or high budget laboratory experiments will certainly lead to tremendous improvements of synthetic strategies. To this aim, we proposed here a machine learning approach. Our tests showed that conceptual density functional theory global and atomic descriptors are relevant features to reach this aim. In particular, the gradient boosting decision tree reached accuracy that is much better than previous studies on that topic. More specifically, we designed a model achieving a mean absolute error equal to 0.72 (with a coefficient of determination of 0.98) with respect to Mayr's experimental values (based on rate constant measurements) for a set of 111 representative and common electrophilic reagents in organic chemistry. Such results constitute, from our point of view, incentive reasons for a still wider use of machine learning methods for the prediction of reactivity properties.

## ACKNOWLEDGMENTS

The authors would like to gratefully acknowledge the LABEX SynOrg (ANR-11-2ABX-0029) and the Normandy Region for funding and support (project AGAC), as well as the Centre Régional Informatique et d'Applications Numériques de Normandie (CRIANN) for computing resources.

## ORCID

Guillaume Hoffmann  <https://orcid.org/0000-0003-1851-5451>

Vincent Tognetti  <https://orcid.org/0000-0002-0292-3509>

## REFERENCES

- [1] C. K. Ingold, *Chem. Rev.* **1934**, 15, 225.
- [2] I. A. Topol, C. Mcgrath, E. Chertova, C. Dasenbrock, W. R. Lacourse, M. A. Eissenstat, S. K. Burt, L. E. Henderson, J. R. Casas-finet, *Protein Sci.* **2001**, 10, 1434.
- [3] A. C. Legon, D. J. Millen, *J. Am. Chem. Soc.* **1987**, 109, 356.
- [4] A. D. McNaught, A. Wilkinson, *IUPAC. Compendium of chemical terminology*, 2nd ed. (the "Gold Book"), Blackwell Scientific Publications, Oxford **1997**.
- [5] C. Gardner Swain, C. B. Scott, *J. Am. Chem. Soc.* **1953**, 75, 141.
- [6] J. O. Edwards, *J. Am. Chem. Soc.* **1954**, 76, 1540.
- [7] C. D. Ritchie, *Acc. Chem. Res.* **1972**, 5, 348.
- [8] H. Mayr, M. Patz, *Angew. Chemie Int. Ed. Engl.* **1994**, 33, 938.
- [9] J. -P. Dau-Schmidt, H. Mayr, *Chem. Ber.* **1994**, 127, 205.
- [10] M. F. Gotta, H. Mayr, *J. Org. Chem.* **1998**, 63, 9769.
- [11] D. S. Allgäuer, H. Jangra, H. Asahara, Z. Li, Q. Chen, H. Zipse, A. R. Ofial, H. Mayr, *J. Am. Chem. Soc.* **2017**, 139, 13318.
- [12] A. Levens, F. An, J. E. M. Fernando, A. R. Ofial, D. W. Lupton, H. Mayr, *Top. Catal.* **2018**, 61, 585.
- [13] H. Mayr, *Tetrahedron* **2015**, 71, 5095.
- [14] H. Mayr, B. Kempf, A. R. Ofial, *Acc. Chem. Res.* **2003**, 36, 66.



- [15] X. Guo, H. Mayr, *J. Am. Chem. Soc.* **2014**, 136, 11499.
- [16] W. Gabsi, K. Essalah, R. Goumont, B. Tangour, T. Boubaker, *Int. J. Chem. Kinet.* **2018**, 50, 659.
- [17] S. Lakhdar, R. Goumont, F. Terrier, T. Boubaker, J. M. Dust, E. Buncel, *Org. Biomol. Chem.* **2007**, 5, 1744.
- [18] M. C. Holland, S. Paul, W. B. Schweizer, K. Bergander, C. Mück-Lichtenfeld, S. Lakhdar, H. Mayr, R. Gilmour, *Angew. Chem. Int. Ed.* **2013**, 52, 7967.
- [19] A. Mood, M. Tavakoli, E. Gutman, D. Kadish, P. Baldi, D. L. Van Vranken, *J. Org. Chem.* **2020**, 85, 4096.
- [20] Z. Li, H. Jangra, Q. Chen, P. Mayer, A. R. Ofial, H. Zipse, H. Mayr, *J. Am. Chem. Soc.* **2018**, 140, 5500.
- [21] J. N. Brønsted, *Chem. Rev.* **1928**, 5, 231.
- [22] M. G. Evans, M. Polanyi, *Trans. Faraday Soc.* **1938**, 34, 11.
- [23] P. Geerlings, F. De Proft, W. Langenaeker, *Chem. Rev.* **2003**, 103, 1793.
- [24] F. De Proft, P. Geerlings, *Chem. Rev.* **2001**, 101, 1451.
- [25] H. Chermette, *J. Comput. Chem.* **1999**, 20, 129.
- [26] R. G. Parr, R. A. Donnelly, M. Levy, W. E. Palke, *J. Chem. Phys.* **1978**, 68, 3801.
- [27] R. G. Parr, R. G. Pearson, *J. Am. Chem. Soc.* **1983**, 105, 7512.
- [28] R. G. Parr, W. Yang, *J. Am. Chem. Soc.* **1984**, 106, 4049.
- [29] R. G. Parr, L. V. Szentpály, S. Liu, *J. Am. Chem. Soc.* **1999**, 121, 1922.
- [30] P. Pérez, A. Toro-Labbé, A. Aizman, R. Contreras, *J. Org. Chem.* **2002**, 67, 4747.
- [31] P. Pérez, *J. Org. Chem.* **2003**, 68, 5886.
- [32] L. R. Domingo, P. Pérez, R. Contreras, *Tetrahedron* **2004**, 60, 6585.
- [33] G. Hoffmann, V. Tognetti, L. Joubert, *J. Mol. Model.* **2018**, 24, 281.
- [34] G. Hoffmann, V. Tognetti, L. Joubert, *Chem. Phys. Lett.* **2019**, 724, 24.
- [35] F. Peccati, E. Desmedt, J. Contreras-García, *Comput. Theor. Chem.* **2019**, 1159, 23.
- [36] J. N. Wei, D. Duvenaud, A. Aspuru-Guzik, *ACS Cent. Sci.* **2016**, 2, 725.
- [37] B. Maryasin, P. Marquetand, N. Maulide, *Angew. Chem. Int. Ed.* **2018**, 57, 6978.
- [38] P. O. Dral, *J. Phys. Chem. Lett.* **2020**, 11, 2336.
- [39] M. M. González, C. Cárdenas, J. I. Rodríguez, S. Liu, F. Heidar-Zadeh, R. A. Miranda-Quintana, P. W. Ayers, *Wuli Huaxue Xuebao / Acta Phys. Chim. Sin.* **2018**, 34, 662.
- [40] D. A. R. S. Latino, F. Pereira, *Helv. Chim. Acta* **2015**, 98, 863.
- [41] R. Appel, H. Mayr, *J. Am. Chem. Soc.* **2011**, 133, 8240.
- [42] A. F. Jalbout, M. Solimannejad, *J. Mol. Struct. THEOCHEM* **2003**, 623, 83.
- [43] C. Morell, A. Grand, A. Toro-Labbé, *J. Phys. Chem. A* **2005**, 109, 205.
- [44] J. Padmanabhan, R. Parthasarathi, M. Elango, V. Subramanian, B. S. Krishnamoorthy, S. Gutierrez-Oliva, A. Toro-Labbé, D. R. Roy, P. K. Chattaraj, *J. Phys. Chem. A* **2007**, 111, 9130.
- [45] R. G. Parr, P. W. Ayers, R. F. Nalewajski, *J. Phys. Chem. A* **2005**, 109, 3957.
- [46] R. F. W. Bader, *Atoms in Molecules: A Quantum Theory*, Oxford University Press, Oxford, UK **1990**.
- [47] P. L. A. Popelier, *Atoms in Molecules An Introduction*, Pearson Education, Harlow **2000**.
- [48] V. Tognetti, L. Joubert, *ChemPhysChem* **2017**, 18, 2675.
- [49] M. J. Vainio, M. S. Johnson, *J. Chem. Inf. Model.* **2007**, 47, 2462.
- [50] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, *Gaussian 09, Revision D.01*, Gaussian Inc., Wallingford, CT **2013**.
- [51] J. Tomasi, B. Mennucci, R. Cammi, *Chem. Rev.* **2005**, 105, 2999.
- [52] T. A. Keith, AIMAll (Version 19.02.13), TK Gristmill Software, Overland Park KS, **2019** (aim.tkgristmill.com).
- [53] B. Gaüzère, L. Brun, D. Villemin, *Pattern Recognit. Lett.* **2012**, 33, 2038.
- [54] P. Foggia, G. Percannella, M. Vento, *Int. J. Pattern Recognit. Artif. Intell.* **2014**, 28, 1450001.
- [55] B. Gaüzère, P. A. Grenier, L. Brun, D. Villemin, *Pattern Recognit.* **2015**, 48, 356.
- [56] A. Feragen, N. Kasenburg, J. Petersen, M. De Bruijne, K. Borgwardt, *Adv. Neural Inf. Process. Syst.* **2013**, 2, 1.
- [57] <https://github.com/bgauzere/ChemoKernel>.
- [58] T. M. Mitchell, *Machine Learning*, McGraw-Hill, New York, USA **1997**.
- [59] D. Cherqaoui, D. Villemin, A. Mesbah, J. M. Cense, V. Kvasnicka, *J. Chem. Soc. Faraday Trans.* **1994**, 90, 2015.
- [60] D. Cherqaoui, D. Villemin, V. Kvasnicka, *Chemom. Intell. Lab. Syst.* **1994**, 24, 117.
- [61] A. N. Tikhonov, A. V. Goncharyk, V. V. Stepanov, A. G. Yagola, *Numerical Methods for the Solution of Ill-Posed Problems*, Dordrecht **1995**.
- [62] P. A. Devijver, J. Kittler, *Pattern Recognition: A Statistical Approach*, Prentice Hall, London **1982**.
- [63] R. Kohavi, *Proc. Fourteenth Int. Jt. Conf. Artif. Intell.* **1995**, 2, 1137.
- [64] S. Geisser, *Predictive Inference*, Chapman and Hall, New York, NY **1993**.
- [65] R. Tibshirani, *J. R. Stat. Soc. Ser. B* **1996**, 58, 267.
- [66] S. Wold, K. Esbensen, P. Geladi, *Chemom. Intell. Lab. Syst.* **1987**, 2, 37.
- [67] B. Schölkopf, A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, Part II. MIT Press, Cambridge, MA **2002**.
- [68] B. Schölkopf, R. Herbrich, A. J. Smola, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **2001**, 2111, 416.
- [69] L. Ralaivola, S. J. Swamidass, H. Saigo, P. Baldi, *Neural Networks* **2005**, 18, 1093.
- [70] A. Smalter, J. Huan, G. Lushington, 8th IEEE Int. Conf. Bioinforma. Bioeng. BIBE 2008, **2008**.
- [71] G. Hinselmann, A. Jahn, N. Fechner, A. Zell, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, **2009**, 5483 LNCS, pp. 25–36.
- [72] S. J. Swamidass, J. Chen, J. Bruand, P. Phung, L. Ralaivola, P. Baldi, *Bioinformatics* **2005**, 21, 359.
- [73] C. E. Rasmussen, Gaussian processes in machine learning. in *Summer School on Machine Learning*, Springer, Berlin, Heidelberg **2003**, p. 63.
- [74] A. Maunz, C. Helma, *SAR QSAR Environ. Res.* **2008**, 19, 413.
- [75] J. Friedman, T. Hastie, R. Tibshirani, *The Elements of Statistical Learning*, Springer, New York **2001**.
- [76] T. Chen, C. Guestrin, *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* **2016**, 13, 785.
- [77] T. Chen, T. He, J. M. L. R. Work, *Conf. Proc.* **2015**, 42, 69.
- [78] L. Torlay, M. Perrone-Bertolotti, E. Thomas, M. Baciú, *Brain Informatics* **2017**, 4, 159.
- [79] J. H. Friedman, *Ann. Stat.* **2001**, 29, 1189.
- [80] D. Tamayo, A. Silburt, D. Valencia, K. Menou, M. Ali-Dib, C. Petrovich, C. X. Huang, H. Rein, C. van Laerhoven, A. Paradise, A. Obertas, N. Murray, *Astrophys. J.* **2016**, 832, L22.

- [81] A. Möller, V. Ruhlmann-Kleider, C. Leloup, J. Neveu, N. Palanque-Delabrouille, J. Rich, R. Carlberg, C. Lidman, C. Pritchett, *J. Cosmol. Astropart. Phys.* **2016**, 2016(12), 8.
- [82] <https://www.mathworks.com/help/stats/lasso.html>.
- [83] <https://xgboost.readthedocs.io/en/latest/python/index.html>.
- [84] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, *J. Mach. Learn. Res.* **2011**, 12, 2825.
- [85] M. Kuhn, K. Johnson, *Applied Predictive Modeling*, 1st ed., Springer, New York **2013**.
- [86] G. James, D. Witten, T. Hastie, R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R (Springer Texts in Statistics)*, 1st ed., Springer, New York **2013**.
- [87] B. S. Manjunath, W. Y. Ma, *IEEE Trans. Pattern Anal. Mach. Intell.* **1996**, 18, 837.
- [88] M. Defferrard, X. Bresson, P. Vandergheynst, Convolutional neural networks on graphs with fast localized spectral filtering, in *Advances in neural information processing systems* **2016**, 3844.
- [89] <https://github.com/balcilar/Predicting-electrophilicity>.
- [90] S. M. Lundberg, S. I. Lee, *Adv. Neural Inf. Process. Syst.* **2017**, 2017, 4766.
- [91] S. M. Lundberg, G. G. Erion, S.-I. Lee, Consistent individualized feature attribution for tree ensembles, **2018**.
- [92] C. Morell, J. L. Gázquez, A. Vela, F. Guégan, H. Chermette, *Phys. Chem. Chem. Phys.* **2014**, 16, 26832.
- [93] A. Robles, M. Franco-Pérez, J. L. Gázquez, C. Cárdenas, P. Fuentealba, *J. Mol. Model.* **2018**, 24, 245.
- [94] R. G. Pearson, *J. Am. Chem. Soc.* **1963**, 85, 3533.
- [95] G. Klopman, *J. Am. Chem. Soc.* **1968**, 90, 223.
- [96] L. Salem, *J. Am. Chem. Soc.* **1968**, 90, 543.
- [97] J. S. M. Anderson, P. W. Ayers, *Phys. Chem. Chem. Phys.* **2007**, 9, 2371.
- [98] M. Torrent-Sucarrat, F. De Proft, P. W. Ayers, P. Geerlings, *Phys. Chem. Chem. Phys.* **2010**, 12, 1072.
- [99] M. Torrent-Sucarrat, F. De Proft, P. Geerlings, P. W. Ayers, *Chem. A Eur. J.* **2008**, 14, 8652.
- [100] T. Bettens, M. Alonso, F. De Proft, T. A. Hamlin, F. M. Bickelhaupt, *Chem. A Eur. J.* **2020**, 26, 1.
- [101] H. Mayr, M. Breugst, A. R. Ofial, *Angew. Chem. Int. Ed.* **2011**, 50, 6470.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Hoffmann G, Balcilar M, Tognetti V, et al. Predicting experimental electrophilicities from quantum and topological descriptors: A machine learning approach. *J Comput Chem.* 2020;41:2124–2136. <https://doi.org/10.1002/jcc.26376>



### 3.3 Conclusion

Bien que l'on qualifie souvent de *boîte noire* les méthodes d'apprentissage automatique, l'utilisation de la méthode SHAP dans l'article a permis à notre modèle de fournir des explications a posteriori sur l'apport de chaque descripteur au modèle. On a ainsi pu montrer que les descripteurs usuels d'électrophilie (tels que le descripteur dual atomique, l'indice d'électrophilie et la fonction de Fukui atomique) participent majoritairement à la construction d'un modèle capable de prédire l'électrophilie de Mayr. L'un des principaux résultats attendus est la capacité supérieure des descripteurs quantiques sur les descripteurs classiques issus du graphe moléculaire.

L'utilisation de la méthode *gradient boosting* (GBDT) a ainsi permis de mettre au point un modèle surpassant les méthodes actuelles de prédiction de l'électrophilie et de mettre en avant l'utilisation de méthodes d'apprentissage automatique en combinaison avec les descripteurs quantiques de réactivité.

Pour la suite de ces travaux, qui ont montré la grande capacité de prédiction des méthodes d'apprentissage supervisé sur les paramètres de Mayr, on pourrait envisager de considérer la mise au point d'un modèle de prédiction des paramètres de nucléophilie ( $N$  et  $s_N$ ) puisqu'un plus grand nombre de données expérimentales sont disponibles (près de 1227 paramètres disponibles).

Enfin, on pourrait également envisager d'effectuer une comparaison entre la capacité de prédiction du modèle présenté dans l'article et les modèles de la littérature, sur des molécules dont la valeurs d'électrophilie ne sont à ce jour qu'estimées [1–4].

- [1] Z. Li, H. Jangra, Q. Chen, et al. *J. Am. Chem. Soc.* 140 (2018), 5500–5515
- [2] A. Mood, M. Tavakoli, E. Gutman, et al. *J. Org. Chem.* 85 (2020), 4096–4102
- [3] R. J. Mayer, A. R. Ofial. *Angew. Chemie - Int. Ed.* 58 (2019), 17704–17708
- [4] K. Umezawa, M. Yoshida, M. Kamiya, et al. *Nat. Chem.* 9 (2017), 279–286

CHAPITRE 5 : *Indices d'électrophilie et liaisons halogènes :  
De nouvelles alternatives au potentiel électrostatique  
moléculaire*

|    |                                  |     |
|----|----------------------------------|-----|
| 1. | Préambule .....                  | 147 |
| 2. | Article .....                    | 148 |
| 3  | Conclusion et perspectives ..... | 160 |



## 1. Préambule

Une question que l'on peut se poser est de savoir si les descripteurs quantiques issus principalement de la DFT conceptuelle peuvent être appliqués à un second problème d'importance en chimie : la nature de la liaison chimique.

Nous nous sommes intéressés dans ce chapitre à la liaison halogène dont la nature exacte demeure à ce jour un sujet controversé dans la littérature. Elle se définit dans un complexe moléculaire comme l'interaction entre une région électrophile (*trou sigma*) associée à l'atome d'halogène (RX) et une région nucléophile d'un accepteur (B), le plus souvent une base de Lewis :

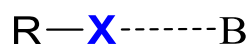


Figure 5.1 : Représentation d'une liaison halogène

Dans la littérature, la liaison halogène est définie principalement comme une interaction majoritairement électrostatique. Le concept de trou sigma fait référence à une région de l'espace située dans la partie externe de l'atome d'halogène le long de cette liaison, que l'on caractérise en termes de potentiel électrostatique moléculaire (MEP). Elle reste néanmoins une liaison dépendant de plusieurs paramètres importants, telles que les effets d'induction, de dispersion, de transfert de charge, d'interaction d'échange [1-2] ...

L'objectif de cette nouvelle étude est de se concentrer principalement sur le caractère électrophile de l'halogène. Plus précisément nous avons réalisé une étude comparative entre les descripteurs d'électrophilie issus de la CDFT et le potentiel électrostatique moléculaire, sur un grand nombre de complexes halogénés (près de 70 molécules étudiées).

Enfin, suite à l'important nombre de définitions de l'indice d'électrophilie présentes dans la littérature, que ce soit au niveau moléculaire ou atomique, nous avons réalisé une étude comparative entre ces différentes définitions et proposé une extension de l'indice d'électrophilie usuel (voir Chapitre 2, partie 4.3.1).

[1] M. H. Kolář, P. Hobza, *Chem. Rev.* 116 (2016), 5155–5187

[2] E. Parisini, P. Metrangolo, T. Pilati, G. Resnati, G. Terraneo, *Chem. Soc. Rev.* 40 (2011), 2267–2278.

# Electrophilicity Indices and Halogen Bonds: Some New Alternatives to the Molecular Electrostatic Potential

Published as part of The Journal of Physical Chemistry virtual special issue "Paul Geerlings Festschrift".

Guillaume Hoffmann, Vincent Tognetti,\* and Laurent Joubert\*

Cite This: *J. Phys. Chem. A* 2020, 124, 2090–2101

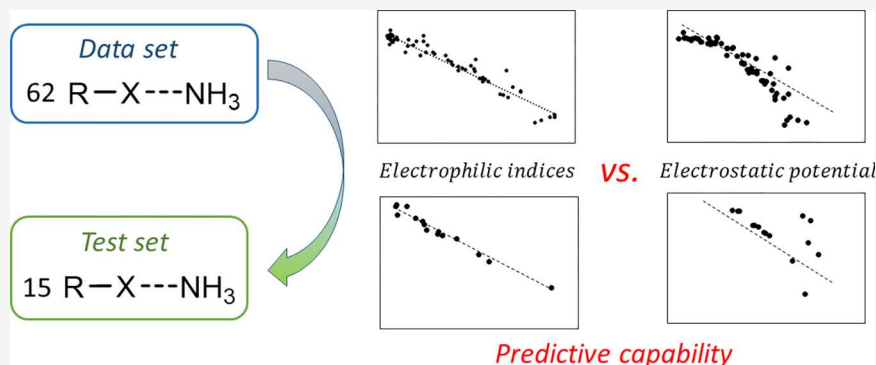
Read Online

ACCESS |

Metrics & More

Article Recommendations

Supporting Information



**ABSTRACT:** We assess the ability of various atomic and molecular electrophilicity descriptors to predict the strength of halogen bonds. To this aim, several physicochemical quantities rooted within the framework of conceptual density functional theory were derived using second and third order Taylor expansions of the electronic energy, and their correlation to binding energies were compared with those obtained for more usual electronic descriptors. This benchmark was performed for a large and representative database of noncovalent complexes involving fluorine, chlorine, and bromine atoms, and showed that some of these new quantities, in particular the atomic cubic electrophilicity index, exhibited more transferability and broadness of application than did more common descriptors such as the molecular electrostatic potential.

## 1. INTRODUCTION

In 1814, Colin and Gaultier de Claubry,<sup>1,2</sup> while working on the combination of iodine with vegetal and animal substances, succeeded in the formation of the  $I_2 \cdots NH_3$  compound, which was later identified as the first complex involving a halogen bond (XB). This specific bond was later defined in 1863<sup>3</sup> with a proposed structure. However, chemists had to wait for many more years until the International Union of Pure and Applied Chemistry (IUPAC) proposed a general and recommended definition of the halogen bond,<sup>4</sup> describing it as an interaction that involves a halogen donor, R-X, featuring an electrophilic region associated with the halogen atom, and a nucleophilic region located on an acceptor, B, which is in most cases a Lewis base.

In the two last decades, intense experimental and theoretical analyses of this noncovalent bond have led to unveil its important role not only in chemistry and biology, but also in various other fields, such as materials science,<sup>5–8</sup> crystal engineering,<sup>9–11</sup> supramolecular chemistry,<sup>12,13</sup> drug development,<sup>14,15</sup> and organocatalysis.<sup>16–18</sup> More recently, and as a particular interest in the field of medicinal chemistry,<sup>19</sup> astatine was also shown to form halogen bonds,<sup>20,21</sup> thus completing

the XB family and its potential applications. From a theoretical perspective, despite all these advances and the considerably high number of computational studies devoted to these bonds, both the accurate prediction of the properties of R-X $\cdots$ B complexes and the determining of the exact nature of such bonds still remain, from our viewpoint, a challenge.

In the literature, the most popular rationalizing approach is that proposed by Politzer, Clark, and co-workers, which they named "the  $\sigma$ -hole concept".<sup>22,23</sup> It commonly refers to a real space region in the outer part of the halogen atom along the R-X bond characterized in terms of the molecular electrostatic potential (MEP). It is in general associated with a local deficiency in electron density, and thus to an electrophilic character. This simple but highly powerful paradigm was actually found to be quite effective in the prediction of

Received: October 31, 2019

Revised: February 6, 2020

Published: March 2, 2020

interaction energies in small data sets (typically a dozen of studied compounds) of halogen-bond complexes.<sup>24–27</sup>

However, several issues may arise from the single use of the electrostatic potential and of the  $\sigma$ -hole model for the description of halogen-bonded complexes, as it has been pointed out in recent years. One problem resides in the fact that  $\sigma$ -holes are not always positive. It was initially shown by Brinck et al.<sup>28</sup> in the case of chlorine in  $\text{CH}_3\text{Cl}$ , a result that was subsequently confirmed by other studies that included larger sets of molecules.<sup>29–31</sup> Later on, it was also proven that electrostatic attraction may not be the only important interaction that contributes to the general XB complexes, in particular when dispersion and induction contribute substantially.<sup>32,33</sup>

In addition, Wolters and Bickelhaupt demonstrated, using energy decomposition analysis (EDA), that the charge-transfer interaction also plays an important role.<sup>34</sup> Besides, exchange repulsion has also proven to play a part in the general description of halogen bonding geometries.<sup>35,36</sup> More recently, Popelier et al. showed, using the Interacting Quantum Atoms (IQA) decomposition and the Relative Energy Gradient (REG) tool,<sup>37</sup> that the secondary interactions and the so-called “self” atomic energies are mainly responsible for the bond directionality,<sup>38</sup> in line with some of our previous findings both on intermolecular or in intramolecular halogen bonds.<sup>39–44</sup>

These few examples (among a plethora literature that we do not plan here to review extensively) support our previous claim that the exact nature of these bonds still remains a subject of controversy,<sup>29,45,46</sup> as the  $\sigma$ -hole model, while widespread, has appeared over the last years to be sometimes too crude to grasp the complexity of these intriguing bonds.

Our aim in this paper is to focus on the cornerstone property that is the electrophilic character of the halogen atom in the R–X molecule before the formation of the bond itself, and to discuss how it can be measured by various descriptors. For this purpose, we will extensively rely on conceptual density functional theory (CDFT), also known as the theory of chemical reactivity,<sup>47–50</sup> which provides a mathematical and physical framework for chemical reactivity concepts such as (to cite only few of them) electronegativity,<sup>51</sup> hardness<sup>52</sup> (directly linked to Pearson’s Hard and Soft Acids and Bases (HSAB) theory<sup>53</sup>), and the maximum hardness principle.<sup>54,55</sup>

It should be noticed that the use of general CDFT indices for halogen bonds were already reported by us<sup>56</sup> and by other groups.<sup>34,57–59</sup> In this contribution, we rather intend to compare some specific CDFT electrophilic indices, either already known or newly derived, with existing electrostatic potential models. All these descriptors, and their associated mathematical derivation, will be described in detail in the next section. They will then be applied (see fourth section) for the characterization of a large set of halogen-bonded complexes (more than 70 in total), at variance with many studies that only test descriptors on reduced data sets (typically less than 20). Some conclusions and guidelines will finally underline the general relevance of such CDFT descriptors.

## 2. ELECTROPHILICITY DESCRIPTORS IN CONCEPTUAL DENSITY FUNCTIONAL THEORY

Conceptual DFT is based, within the canonical ensemble, on the successive derivatives  $\frac{\delta^{m+n}E_e}{\delta v(\mathbf{r})^m \delta N^n}$  of the electronic energy  $E_e$  with respect to the number of electrons  $N$ , to the external

potential generated by the nuclei  $v(\mathbf{r})$ , and on the use of perturbative expansions of  $E_e$  through these ingredients. The primary aim of conceptual DFT is actually to extract chemical information from such quantities. As a consequence, they can be called “chemical descriptors”.

From a general point of view, any chemical descriptor in conceptual DFT can be classified into three categories: global descriptors that only involve derivatives with respect to  $N$ , local descriptors that are real space functions depending on only one spatial coordinate of the  $a(\mathbf{r})$  type, and nonlocal descriptors (also called “kernels”), which are functions of multiple space coordinates  $b(\mathbf{r}, \mathbf{r}', \dots)$ .

In a recent paper,<sup>60</sup> we coined “basic descriptors” as those that are directly expressed as energy derivatives in the canonical ensemble. On the other hand, “composite descriptors” are obtained by a combination of basic descriptors. Among the global basic descriptors, one finds the electronic chemical potential  $\mu$ ,<sup>51</sup> the molecular hardness  $\eta$ ,<sup>52</sup> and the molecular hyperhardness  $\gamma$ ,<sup>61,62</sup> respectively, defined by

$$\mu = \left( \frac{\partial E_e}{\partial N} \right)_{v(\mathbf{r})} \quad (1a)$$

$$\eta = \left( \frac{\partial^2 E_e}{\partial N^2} \right)_{v(\mathbf{r})} \quad (1b)$$

$$\gamma = \left( \frac{\partial^3 E_e}{\partial N^3} \right)_{v(\mathbf{r})} \quad (1c)$$

Global indices related to electrophilicity are in general composite descriptors based on these last quantities. In particular, the global electrophilicity index  $\omega$ , initially qualitatively proposed by Maynard et al.<sup>63</sup> and later defined by Parr et al.,<sup>64</sup> remains one of the most used reactivity descriptors for describing global electrophilicity in a molecule.<sup>65</sup> Defined as the stabilization energy of a chemical system (molecule) in interaction with a perfect electron donor (a bath of electrons with zero chemical potential), this intrinsic property was derived in terms of the chemical potential  $\mu$  and the chemical hardness  $\eta$  through a parabolic model:

$$\omega_{\text{parabolic}} = \frac{\mu^2}{2\eta} \quad (2)$$

We will here consider an extension of this second-order model by explicitly considering a cubic Taylor expansion of the total energy  $E = E_e + E_{\text{NN}}$ , where  $E_{\text{NN}}$  stands for the repulsion energy between nuclei, which are here considered fixed:

$$E(N + \Delta N) = E(N) + \left( \frac{\partial E}{\partial N} \right)_{v(\mathbf{r})} \Delta N + \frac{1}{2} \left( \frac{\partial^2 E}{\partial N^2} \right)_{v(\mathbf{r})} \Delta N^2 + \frac{1}{6} \left( \frac{\partial^3 E}{\partial N^3} \right)_{v(\mathbf{r})} \Delta N^3 \quad (3)$$

At constant external potential, at any order  $p$ ,  $\frac{\partial^p E}{\partial N^p} = \frac{\partial^p E_e}{\partial N^p}$ , so that, identifying the first, second, and third derivatives to the chemical potential, molecular hardness, and molecular hyperhardness (eqs 1a–1c):

$$\begin{aligned}\Delta E_{\text{cubic}} &= E(N + \Delta N) - E(N) \\ &= \mu \Delta N + \frac{\eta}{2} \Delta N^2 + \frac{\gamma}{6} \Delta N^3\end{aligned}\quad (4)$$

The energy variation is optimal for  $\frac{d\Delta E_{\text{cubic}}}{d\Delta N} = 0$ , thus for

$$\mu + \eta \Delta N + \frac{\gamma}{2} \Delta N^2 = 0 \quad (5)$$

the corresponding roots of this binomial equation are

$$\Delta N = \frac{-\eta \pm \sqrt{\eta^2 - 2\mu\gamma}}{\gamma} \quad (6)$$

Taking into account that we are modeling electrophilicity, the  $\Delta N > 0$  condition (using the fact that  $\mu < 0$  and  $\eta > 0$  for all known systems due to the conjectured energy convexity), only one solution can be retained:

$$\Delta N = \frac{-\eta + \sqrt{\eta^2 - 2\mu\gamma}}{\gamma} \quad (7)$$

In cases for which hyperhardness is small, one can perform a first-order Taylor expansion in  $\gamma$  of eq 7, giving

$$\Delta N = -\frac{\mu}{\eta} - \frac{\mu^2}{2\eta^3}\gamma + O(\gamma^2) \quad (8)$$

It should be noticed that in the case for which  $\gamma = 0$ , this equation reduces to  $\Delta N = -\mu/\eta$ , which is precisely that obtained by Parr and co-workers through the parabolic model.

Setting now  $\Delta N = -\frac{\mu}{\eta} - \frac{\mu^2}{2\eta^3}\gamma$  in eq 4 gives

$$\Delta E_{\text{cubic}} = -\frac{\mu}{2\eta} - \frac{\mu^3}{6\eta^3}\gamma - \frac{\mu^4}{8\eta^5}\gamma^2 + A\gamma^3 + B\gamma^4 \quad (9)$$

For internal consistency, this expression must be also truncated at first-order in  $\gamma$ . Note also that the zeroth-order term is nothing else than the opposite of Parr's electrophilicity index, so that

$$\Delta E_{\text{cubic}} = -\omega_{\text{parabolic}} - \frac{\mu^3}{6\eta^3}\gamma \quad (10)$$

Following a similar approach as Parr, we will henceforth define the cubic electrophilicity index as being equal to the opposite of  $\Delta E_{\text{cubic}}$ . It can be rearranged according to

$$\omega_{\text{cubic}} = \omega_{\text{parabolic}} \left( 1 + \frac{\mu}{3\eta^2}\gamma \right) \quad (11)$$

explicitly showing how  $\gamma$  perturbs the original index.

We now discuss the practical evaluation of this equation. There are two usual working schemes to compute chemical descriptors, the first working scheme is based on the frontier molecular orbital (FMO) approach and the second one is the finite difference linearization (FDL), which considers the addition or removal of one (or more) electron(s) at fixed geometry (vertical processes in order to work at fixed external potential). It should be stressed that other alternative approaches, such as those based on the extended Koopmans' theorem (EKT), may be valuably used, providing a straightforward way in direct comparison with experimental values.<sup>66</sup>

If coming back to the FMO scheme, one has the following approximations:

$$\mu^{\text{FMO}} = \frac{\epsilon_{\text{HOMO}} + \epsilon_{\text{LUMO}}}{2} \quad (12a)$$

$$\eta^{\text{FMO}} = \epsilon_{\text{LUMO}} - \epsilon_{\text{HOMO}} \quad (12b)$$

$$\gamma^{\text{FMO}} = \epsilon_{\text{LUMO}} - 2\epsilon_{\text{HOMO}} + \epsilon_{\text{HOMO}-1} \quad (12c)$$

where, as usual, HOMO and LUMO acronyms denote the highest occupied and lowest unoccupied Kohn–Sham (KS) molecular orbitals, respectively. Equations 12a–12c allow for defining the corresponding  $\omega_{\text{cubic}}^{\text{FMO}}$  global descriptor.

On another side with the FDL approach, once more eq 4 is considered, but now with  $\Delta N = 1$ . The energy variation is, by definition, the opposite of the electron affinity EA:

$$\Delta E_{\text{cubic}}^{\Delta N=1} = -\text{EA}(N) = \mu + \frac{\eta}{2} + \frac{\gamma}{6} \quad (13a)$$

We can also consider the removal of the electron, directly linked to the first ionization potential  $\text{IP}_1$ :

$$\Delta E_{\text{cubic}}^{\Delta N=-1} = \text{IP}_1(N) = -\mu + \frac{\eta}{2} - \frac{\gamma}{6} \quad (13b)$$

We can also consider a decrement of two electrons, which now involves both  $\text{IP}_1$  and the second ionization potential  $\text{IP}_2$ :

$$\Delta E_{\text{cubic}}^{\Delta N=-2} = \text{IP}_1(N) + \text{IP}_2(N) = -2\mu + 2\eta - \frac{4}{3}\gamma \quad (13c)$$

Equations 13a–13c define a system of linear equations in  $\mu$ ,  $\eta$ , and  $\gamma$ , which can be exactly solved, giving (where the  $N$  dependence for energies has been dropped out for sake of clarity)

$$\mu_{\text{cubic}}^{\text{FDL}} = \frac{1}{6}(-2\text{EA} - 5\text{IP}_1 + \text{IP}_2) \quad (14a)$$

$$\eta_{\text{cubic}}^{\text{FDL}} = -\text{EA} + \text{IP}_1 \quad (14b)$$

$$\gamma_{\text{cubic}}^{\text{FDL}} = -\text{EA} + 2\text{IP}_1 - \text{IP}_2 \quad (14c)$$

These equations could be compared to those obtained for the parabolic model based on:

$$-\text{EA}(N) = \mu + \frac{\eta}{2} \quad (15a)$$

$$\text{IP}_1(N) = -\mu + \frac{\eta}{2} \quad (15b)$$

whose solution is

$$\mu_{\text{parabolic}}^{\text{FDL}} = -\frac{1}{2}(\text{EA} + \text{IP}_1) \quad (16a)$$

$$\eta_{\text{parabolic}}^{\text{FDL}} = -\text{EA} + \text{IP}_1 \quad (16b)$$

Equation 16a is no more than the opposite of Mulliken's electronegativity.<sup>66,67</sup> Interestingly, confronting eqs 14 and 16 shows that the third order (hyperhardness) effect is ineffective for the molecular hardness, while it affects the chemical potential. All these equations then allow us to define two approximate descriptors within the FDL approach:

$$\omega_{\text{cubic}}^{\text{FDL}} = \frac{(\mu_{\text{cubic}}^{\text{FDL}})^2}{2\eta_{\text{cubic}}^{\text{FDL}}} \left( 1 + \frac{\mu_{\text{cubic}}^{\text{FDL}}}{3(\eta_{\text{cubic}}^{\text{FDL}})^2} \gamma_{\text{cubic}}^{\text{FDL}} \right) \quad (17)$$



$$\omega_{\text{parabolic}}^{\text{FDL}} = \frac{(\mu_{\text{parabolic}}^{\text{FDL}})^2}{2\eta_{\text{parabolic}}} \quad (18)$$

In total, we have thus defined up to now four global electrophilicity indices:  $\omega_{\text{parabolic}}^{\text{FMO}}$ ,  $\omega_{\text{cubic}}^{\text{FMO}}$ ,  $\omega_{\text{parabolic}}^{\text{FDL}}$ ,  $\omega_{\text{cubic}}^{\text{FDL}}$ .

It should be underlined that we used here the  $N - 2$  electron species, and not the  $N + 2$  one. This choice was notably governed by pragmatic reasons. Indeed, evaluating electron affinities is sometimes not an easy task within current DFT approximations. Moreover, there are some species to which no extra electrons can be added (it will be expelled to infinity), so that EA is then not properly defined. As a consequence, the second EA (which would be needed using  $\Delta N = 2$ ) is not either defined. This is in particular the case for neutral species. To show this, we recall that the asymptotic KS potential (far from the nuclei) for an atom of nuclear charge  $Z$  is  $(N - Z - 1)/r$ , which is repulsive for  $N = Z + 2$ . As all of the considered molecules in our set will be neutral, considering the  $N - 2$  species instead of the  $N + 2$  one is justified. See also some recent discussion about the related topic of negative electron affinities by Geerlings and co-workers.<sup>68</sup>

Besides, another implicit hypothesis in the present work lies in the fact that both parabolic and cubic energy interpolations assume that  $E(N)$  is continuously differentiable. It is not our purpose here to deal with the well-known derivative discontinuity issue.<sup>69</sup>

It should also be emphasized that other schemes were already proposed in the literature. For instance, Gázquez et al.<sup>70</sup> suggested an alternative scheme to the unique parabola model using two models for charge accepting and charge donating processes with different parabolas, resulting to

$$\omega_G^+ = \frac{(\text{IP}_1 + 3\text{EA})^2}{16(\text{IP}_1 - \text{EA})} \quad (19)$$

where  $\omega_G^+$  is defined as the electroaccepting power of a given molecule.

On the other hand, considering an alternative linear interpolation for the chemical potential, Chamorro et al.<sup>71,72</sup> proposed two other expressions for the electrophilicity power:

$$\omega_{\text{C1}}^+ = \frac{(\text{EA})^2}{(\text{IP}_1 - \text{EA})} \quad (20)$$

and:

$$\omega_{\text{C2}}^+ = \frac{(\text{EA})^2}{2\text{IP}_1} \quad (21)$$

However, all of the presented avatars of the electrophilicity index are actually global descriptors that describe the reactivity of the whole molecule. They do not afford any insight into regioselectivity. For this purpose, one has to rely on local functions. The first basic local descriptor is the derivative of  $E_e$  with respect to the external potential. First-order perturbation theory shows that it identifies to the electron density:

$$\left( \frac{\delta E_e}{\delta v(\underline{r})} \right)_N = \rho(\underline{r}) \quad (22)$$

From this equation, one can define the Fukui functions<sup>73</sup> defined by

$$f^+(\underline{r}) = \left( \frac{\delta^2 E_e}{\delta v(\underline{r}) \delta N} \right)_{\partial N > 0} \quad (23a)$$

$$f^-(\underline{r}) = \left( \frac{\delta^2 E_e}{\delta v(\underline{r}) \delta N} \right)_{\partial N < 0} \quad (23b)$$

where  $f^+$  is the electrophilic Fukui function since it informs on the real space regions that are the most prone to gain electrons, while  $f^-$  is the nucleophilic counterpart. As a given zone can be at the same time electrophilic and nucleophilic, it is very often enlightening to consider the dual descriptor:<sup>74</sup>

$$f^2(\underline{r}) = \left( \frac{\delta^3 E_e}{\delta v(\underline{r}) \delta N^2} \right) \approx f^+(\underline{r}) - f^-(\underline{r}) \quad (24)$$

so that positive values correspond to points in which the electrophilicity predominates over the nucleophilicity.

However, it has been often advocated, when comparing molecules with different electron numbers, to shift (by the appropriate Legendre transform) to the grand-canonical ensemble  $\Omega[\mu, v(\underline{r})]$ . In this ensemble, the basic local descriptors are the so-called local softnesses:

$$s^\pm(\underline{r}) = \frac{\delta^2 \Omega}{\delta v(\underline{r}) \delta \mu} = \frac{f^\pm(\underline{r})}{\eta} \quad (25)$$

which can be considered as composite descriptors within the canonical ensemble. Other composite local descriptors were later defined, such as the local electrophilicity  $\omega(\underline{r})$  according to

$$\omega(\underline{r}) = \omega_{\text{parabolic}} f^+(\underline{r}) \quad (26)$$

Chattaraj et al. proposed instead the local philicity index:<sup>75</sup>

$$\omega_{G,\text{C1},\text{C2}}^+(\underline{r}) = \omega_{G,\text{C1},\text{C2}}^+ f^+(\underline{r}) \quad (27)$$

On the other hand, Fuentealba et al.<sup>76</sup> introduced a new description for local electrophilicity based on the changes in the electron density:

$$\omega_{\text{elec}-2}(\underline{r}) = \omega_{\text{parabolic}} \left[ f^-(\underline{r}) + f^+(\underline{r}) + \frac{(\text{IP} + \text{EA})}{2(\text{IP} - \text{EA})} f^2(\underline{r}) \right] \quad (28)$$

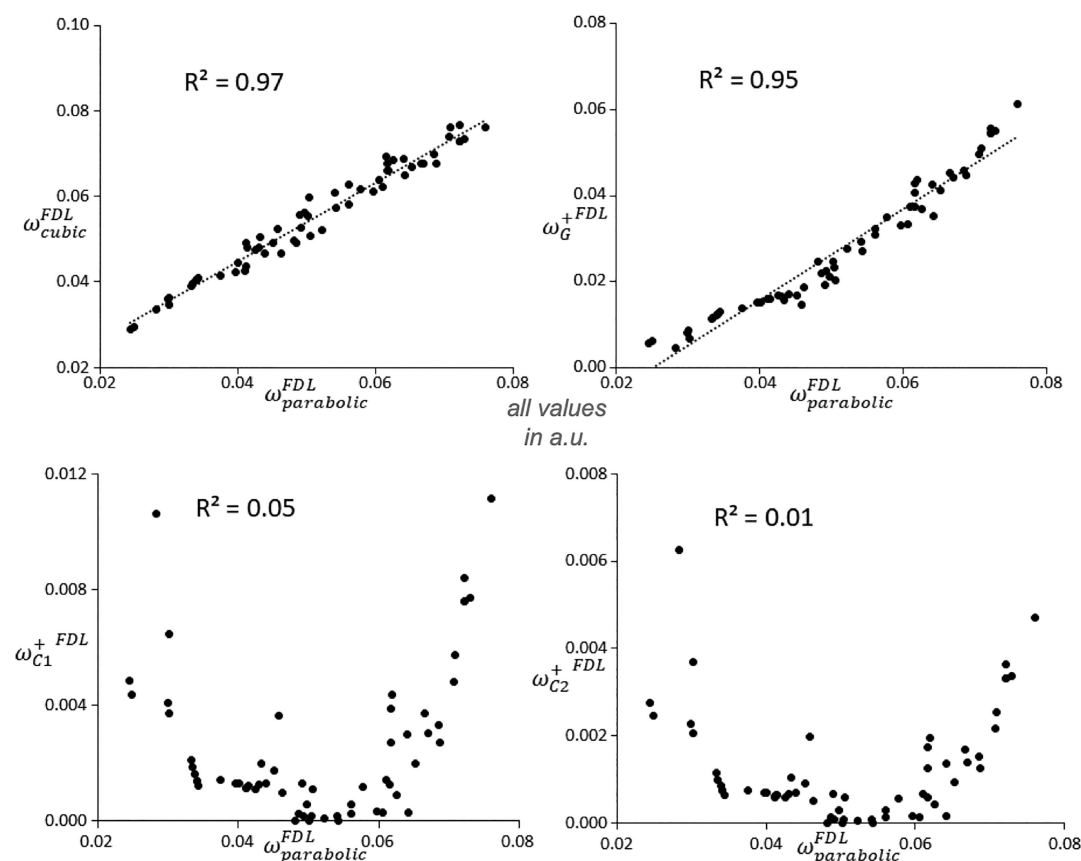
Once more, all these local functions can be evaluated either by the FMO approach, in which  $f^+(\underline{r})$  (respectively  $f^-(\underline{r})$ ) corresponds to the LUMO (respectively HOMO) electron density, or by the FDL one,  $f^+(\underline{r})$  being, in that case, the (vertical) electron density difference between the  $N + 1$  and  $N - 1$  electron species.

Nonlocal kernels could also be considered. For instance, the static linear response kernel is defined by<sup>77</sup>

$$\chi(\underline{r}, \underline{r}') = \left( \frac{\delta^2 E_e}{\delta v(\underline{r}) \delta v(\underline{r}')} \right)_N \quad (29)$$

Consider now an infinitesimal variation of the external potential in the form of an infinitesimal point charge located at  $\underline{R}_{pc}$  ( $\delta v(\underline{r}) = -\delta q/|\underline{r} - \underline{R}_{pc}|$ ). The associated energy variation reads:<sup>78</sup>





**Figure 1.** Comparison of various global electrophilicity indices at the M06-2X/aug-cc-pVTZ level of theory. All values in atomic units.

$$\delta E_e = -\delta q \int \frac{\rho(\underline{r})}{|\underline{r} - \underline{R}_{pc}|} d^3r \quad (30)$$

which is proportional to the electrostatic component of the molecular electrostatic potential (MEP):

$$\text{MEP}(\underline{R}_{pc}) = v(\underline{R}_{pc}) - \int \frac{\rho(\underline{r})}{|\underline{r} - \underline{R}_{pc}|} d^3r \quad (31)$$

In other words, the MEP function is a conceptual DFT descriptor.

Local descriptors actually convey an infinite amount of information since they are defined at each real space point. On the other hand, a chemist very often adopts a coarse-grained view of reactivity based on atomic or substituent properties, which are in finite number. It is thus appealing to build a bridge between both approaches. This can be achieved by noticing that the information embodied in any local descriptor  $a(\underline{r})$  (either basic or composite) can be condensed at the atomic level by summing up all contributions inside a given atom:

$$a(A) = \int_{\Omega_A} a(\underline{r}) d^3r \quad (32)$$

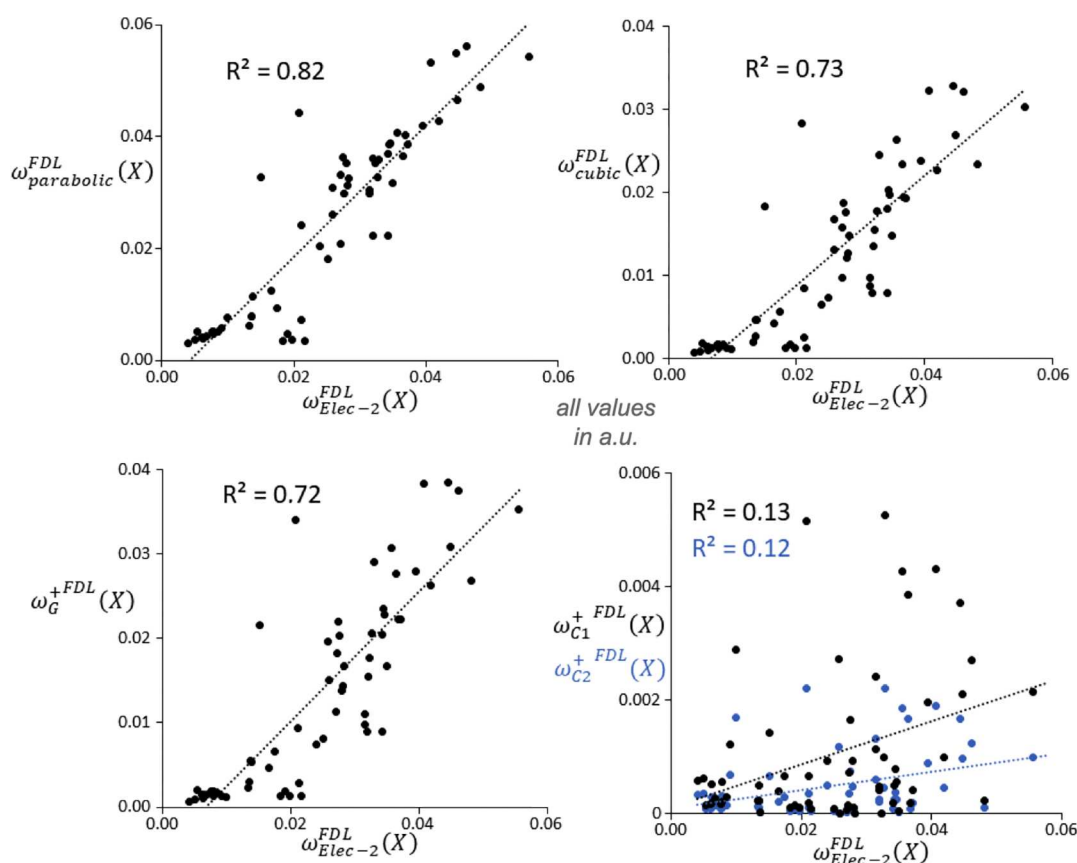
where real space is divided into nonoverlapping volumes (often called “atomic basins”)  $\Omega_A$ . When considering the electron density as  $a(\underline{r})$ , one directly obtains the atomic electronic population. It follows from this that atomic charges are true conceptual DFT condensed descriptors. Other atomic values can be straightforwardly obtained by considering, among others, the Fukui functions or the dual descriptor.<sup>79</sup>

In this paper, we choose to make exclusive use of the quantum theory of atoms in molecules (QTAIM)<sup>80,81</sup> that allows the partitioning of a molecule into proper open subsystems, which is commonly applied when looking at the halogen bonds interaction,<sup>82–85</sup> and especially at the computed interaction energies.<sup>39,86,87</sup> Obviously such a study could have been performed using alternative partitions such as Hirshfeld,<sup>88</sup> fuzzy atoms, etc.<sup>89</sup>

### 3. COMPUTATIONAL DETAILS

All calculations were performed using the Gaussian 09 package<sup>93</sup> and with the M06-2X functional,<sup>90</sup> which has been shown to be efficient for noncovalent interaction and in the description of halogen bonds.<sup>29,91</sup> The standard aug-cc-pVTZ triple- $\zeta$  basis set was used for all atoms. The nature of all stationary points obtained by the optimization procedure was checked by vibrational analysis (no imaginary frequencies were found). Basis set superposition errors (BSSEs) were taken into account using Boys and Bernardi’s counterpoise scheme.<sup>92</sup> The recommended “ultrafine” integration grid was used throughout the study.

QTAIM calculations were carried out using the AIMAll package.<sup>94</sup> Chemical reactivity descriptors were then calculated using the frontier molecular orbital scheme (denoted FMO) and with the finite difference linearization (FDL). Within the FDL approach, which involves single points calculations for the  $N + 1$  or  $N - 1$ -electron species, the Yang-Mortier scheme was applied<sup>95</sup> using the QTAIM population analysis. Further details on these specific calculations are provided in the [Supporting Information](#) and in our previous articles on electrophilicity prediction.<sup>60,96</sup>



**Figure 2.** Comparison of various QTAIM condensed electrophilicities of halogens at the M06-2X/aug-cc-pVTZ level of theory. All values in atomic units.

All descriptors values will be given in atoms units (au).

#### 4. RESULTS AND DISCUSSION

Our main data set consists of 62 common halogen-bonded complexes (inspired by our literature survey) formed between R–X and NH<sub>3</sub> as the base partner. The involved R moieties are H, F, Br, HO, FO, ClO, BrO, NC, H<sub>3</sub>C, F<sub>3</sub>C, Cl<sub>3</sub>C, H<sub>3</sub>Si, HCC, H<sub>2</sub>CCH, CH<sub>3</sub>CH<sub>2</sub>, Br–CH<sub>2</sub>CH<sub>2</sub>, H<sub>3</sub>CO, FH<sub>2</sub>CO, ClH<sub>2</sub>CO, F<sub>2</sub>HCO, Cl<sub>2</sub>HCO, Cl<sub>3</sub>CO, F<sub>2</sub>ClCO, F<sub>3</sub>CO, Br<sub>3</sub>CO, Ph, FPh, ClPh, BrPh, H<sub>2</sub>NPh, and O<sub>2</sub>NPh. It covers both aliphatic and aromatic compounds, and features some of the most common substituents in organic chemistry, making this set representative of most of the halogen bonds encountered experimentally.

For each of these R substituents, we considered two different halogen atoms X: Cl and Br, excluding the astatine complexes that would require spin–orbit calculations, but also iodine as we found, during preliminary calculations, several discrepancies due to the choice of the basis sets (an issue outside of the scope of this paper and that we intend to report in due course). However, despite considering only two halogens, this large collection stands among the biggest available ones dedicated to theoretical predictions of halogen bonds.

Let us recall that our strategy here is based on the intrinsic properties of the reactants, that-is-to-say the R–X donor molecule (note that the descriptors are evaluated on the optimized geometry). For such reasons, geometrical and chemical properties of the complexes, which are available in the [Supporting Information](#), will not be discussed here. In total, for each R–X molecule, we calculated over 19 global

reactivity descriptors and 16 atomic descriptors, which were either numerically computed from the FDL or FMO approaches (see the [Supporting Information](#) for all numerical values).

At first, we will compare the various expressions obtained from the global and local electrophilicities. For the sake of simplicity, we are restraining here to the FDL scheme. [Figure 1](#) shows the variations of the four less used global electrophilicity indices, namely  $\omega_{\text{cubic}}$ ,  $\omega_G^+$ ,  $\omega_{C1}^+$ ,  $\omega_{C2}^+$  with respect to the most popular index, which is Parr's venerable one ( $\omega_{\text{parabolic}}$ ). From the left top panel, it appears that the cubic and parabolic models are well correlated ( $R^2 = 0.97$ ). The mean absolute deviation (MAD) between them was found equal to 0.006, which represents more than 10% of the average value for  $\omega_{\text{parabolic}}$  (0.051). This shows that the hyperhardness correction is not negligible at all (a conclusion that is also valid for the chemical potential since the  $R^2$  value between  $\mu_{\text{cubic}}^{\text{FDL}}$  and  $\mu_{\text{parabolic}}^{\text{FDL}}$  is no more than 0.92). It is worthy noticing that  $\omega_G^+$  also exhibited a strong linear relationship with Parr's electrophilicity index ( $R^2 = 0.95$ ), even if they are numerically quite different since the average for the  $\omega_G^+/\omega_{\text{parabolic}}$  ratio equals 0.49.

Interestingly such a value can be easily justified. Indeed, in general the first ionization potential is much higher in magnitude than the electron affinity. The descriptors values can then be safely approximated by the first terms of a Taylor expansion in the EA variable according to

$$\omega_{\text{parabolic}} = \frac{1}{8}IP_1 + \frac{3}{8}EA + O(EA^2) \quad (33a)$$

$$\omega_G^+ = \frac{1}{16}IP_1 + \frac{7}{16}EA + O(EA^2) \quad (33b)$$

so that

$$\frac{\omega_G^+}{\omega_{\text{parabolic}}} = \frac{1}{2} + 2\frac{EA}{IP_1} + O(EA^2) \quad (34)$$

From these results, one may also wonder whether  $\omega_{\text{cubic}}$  and  $\omega_G^+$  are also actually correlated. The coefficient of determination was found equal to 0.92, corroborating our previous analysis.

Conversely, it is plain from the inspection of the two bottom panels of Figure 1 that neither  $\omega_{C1}^+$  nor  $\omega_{C2}^+$  are correlated to  $\omega_{\text{parabolic}}$  ( $R^2 < 0.1$  in both cases). Once more, this result can be understood by inspecting Taylor expansions in the EA variable:

$$\omega_{C1}^+ = \frac{EA^2}{IP_1} + O(EA^3) \quad (35a)$$

$$\omega_{C2}^+ = \frac{EA^2}{IP_1} + O(EA^5) \quad (35b)$$

which considerably differs from those presented in eqs 33a and 33b. Equation 35a and eq 35b also suggest that  $\omega_{C1}^+$  and  $\omega_{C2}^+$  are correlated. This is the case as illustrated by the high  $R^2$  value (0.97). Finally, as a consequence of all previous remarks,  $\omega_G^+$  and  $\omega_{C1}^+$  or  $\omega_{C2}^+$  are not correlated.

We now compare the condensed values on every halogen atom X. Let us recall that the four local functions  $\omega(r)$  and  $\omega_{G,C1,C2}^+(r)$  share the same local ingredient (the electrophilicity Fukui function) so that the relationship between corresponding condensed values is only driven by the global component we have just discussed. On the other hand, Fuentealba's descriptor  $\omega_{\text{elec-2}}(r)$  may considerably differ from the other  $\omega(r)$  functions since it also involves the nucleophilic Fukui function. Figure 2 then shows the halogen QTAIM condensed  $\omega(X)$  and  $\omega_{G,C1,C2}^+(X)$  values as a function of  $\omega_{\text{elec-2}}(X)$  in the FDL approach.

Once more,  $\omega_{C1}^+(X)$  and  $\omega_{C2}^+(X)$  were revealed apart (with  $R^2 < 0.20$ ). This can be accounted for by the fact that  $\omega_{\text{elec-2}}(r)$  includes  $\omega_{\text{parabolic}}$ , just like  $\omega_G^+(r)$  and  $\omega_{\text{parabolic}}(r)$ . However, the correlations between the condensed values involving the other descriptors are not so high ( $R^2$  value below than 0.83). It can be concluded that even if the  $f^-(r)$  does not change, the basic trends (low values (respectively high) for  $\omega_{\text{elec-2}}(X)$  generally correspond to low (respectively high) values for  $\omega_{\text{parabolic}}(X)$  or  $\omega_G^+(X)$ ); it has a non-negligible impact on the measure of the halogen electrophilicity. This is certainly due to the fact that halogens are also well-known nucleophilic sites through their lone pairs, which implies that  $f^-(X)$  may take important values.

After this methodological benchmark, we now assess to what extent these electrophilicity indices could be used to predict halogen-bond strength. This last quantity may actually present several faces, and may thus be evaluated from nonequivalent approaches. Let us consider the generic formation of a noncovalent complex,  $M_1 \cdots M_2$ , between two molecules  $M_1$  and  $M_2$ . We can define (among others) the complexation self-consistent field (SCF) energy by

$$\Delta E_{\text{SCF}}^{\text{compl}} = E_{\text{SCF}}(M_1 \cdots M_2) - E_{\text{SCF}}(M_1^{\text{opt}}) - E_{\text{SCF}}(M_2^{\text{opt}}) \quad (36)$$

where  $E_{\text{SCF}}(M_1^{\text{opt}})$  denotes the SCF energy of molecule  $M_1$  in its optimized geometry. Vibrational and thermodynamical corrections could then be added on top of it. For instance, it is common practice to include vibrational zero-point energies (ZPEs) to  $\Delta E_{\text{SCF}}^{\text{compl}}$ . When considering entropy effects, the most natural quantity is the standard complexation Gibbs energy according to

$$\Delta G^{0,\text{compl}} = G^\circ(M_1 \cdots M_2) - G^\circ(M_1^{\text{opt}}) - G^\circ(M_2^{\text{opt}}) \quad (37)$$

On the contrary, one could prefer to look at the isolated molecules in the conformation they adopt once in the complex. In such cases, we will consider, instead, binding SCF energies:

$$\Delta E_{\text{SCF}}^{\text{bind}} = E_{\text{SCF}}(M_1 \cdots M_2) - E_{\text{SCF}}(M_1^{M_1 \cdots M_2}) - E_{\text{SCF}}(M_2^{M_1 \cdots M_2}) \quad (38)$$

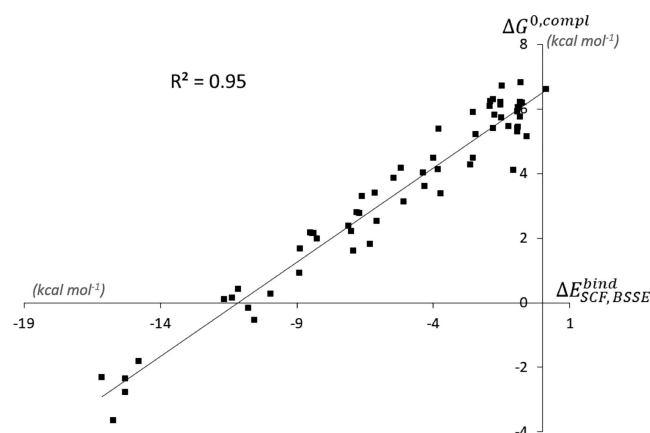
The difference between  $\Delta E_{\text{SCF}}^{\text{bind}}$  and  $\Delta E_{\text{SCF}}^{\text{compl}}$  could be coined the deformation energy, since it measures the energy required for each molecule  $M_1$  and  $M_2$  to deform from their most stable geometry to the final one in the complex. It should be noticed that, as the common way for evaluating thermodynamic corrections is based on the harmonic approximation around equilibrium positions, one cannot straightforward evaluate Gibbs binding energies.

From a practical point of view, finite basis sets are used to solve the SCF Kohn–Sham equations. This approximation generates the so-called basis set superposition errors. The corresponding correction can be added to  $\Delta E_{\text{SCF}}^{\text{bind}}$ , leading to the  $\Delta E_{\text{SCF,BSSE}}^{\text{bind}}$  quantity. Note however that it is far from trivial to define the BSSE-corrected version of complexation energies (for instance, where to locate ghost functions?). In this paper, we will only consider the two quantities that are the most often scrutinized:  $\Delta E_{\text{SCF,BSSE}}^{\text{bind}}$  and  $\Delta G^{0,\text{compl}}$ .

We found that the calculated values for the BSSE-corrected interaction energies for the 62 complexes vary from  $-0.56 \text{ kcal mol}^{-1}$  ( $\text{H}_3\text{SiBr} \cdots \text{NH}_3$ ) to  $-16.13 \text{ kcal mol}^{-1}$  ( $\text{Br}_3\text{COBr} \cdots \text{NH}_3$ ). These values cover a significant range, which makes our chosen set statistically relevant. The average values were found equal to  $-4.04 \text{ kcal mol}^{-1}$  for  $X = \text{Cl}$  and equal to  $-6.59 \text{ kcal mol}^{-1}$  for  $X = \text{Br}$ , consistently with the fact that halogen bonds are in general stronger when the halogen mass increases. More precisely, for a given substituent R, we always found, as expected, that  $|\Delta E_{\text{SCF,BSSE}}^{\text{bind}}(\text{R}-\text{Cl} \cdots \text{NH}_3)| < |\Delta E_{\text{SCF,BSSE}}^{\text{bind}}(\text{R}-\text{Br} \cdots \text{NH}_3)|$ .

Figure 3 shows that  $\Delta G^{0,\text{compl}}$  is rather well correlated to  $\Delta E_{\text{SCF,BSSE}}^{\text{bind}}$  since  $R^2 = 0.95$ . It should however be noticed that the points for very low (in absolute value)  $\Delta E_{\text{SCF,BSSE}}^{\text{bind}}$  (typically lower than  $2.0 \text{ kcal mol}^{-1}$ ), describe a kind of anarchic cloud. This could be ascribable to the fact that this is the entropy term that then clearly dominates in the corresponding Gibbs energies. This entropy term also explains why most of the Gibbs complexation energies were found positive, since only 22% of the values are negative for bromine and there are none in the chlorine case.

This observation is fully in line with Politzer and Murray's valuable and fundamental statement: "for sufficiently strong interactions, therefore,  $\Delta H$  dominates and  $\Delta G < 0$ , i.e. thermodynamic stability".<sup>97</sup> Let us also recall that all values here are gas phase ones. Quoting once more Politzer and Murray,<sup>97</sup> "in solution and solid phases, the situation is complicated by additional factors, the overall effect of which can produce  $\Delta G < 0$  even in cases that are thermodynamically unstable in the gas phase". However, the evaluation of



**Figure 3.** Standard Gibbs complexation energies with respect to BSSE-corrected binding energies at the M06-2X/aug-cc-pVTZ level of theory. All values in kcal mol<sup>-1</sup>.

entropies in solvent is a hard task, clearly outside the scope of our paper (see however references<sup>98,99</sup> for some recent attempts). For such reasons, only  $\Delta E_{\text{SCF,BSSE}}^{\text{bind}}$  will be further considered.

We are now seeking correlation between these interaction energies and the computed CDFD descriptors. It is instructive first to look at  $\omega_{\text{parabolic}}$ . It was found to perform quite disappointingly since  $R^2 = 0.69$  and  $0.65$  for the whole set of 62 molecules, in the FDL and FMO approaches, respectively. The use of the genuine condensed electrophilic Fukui function  $f^+(X)$  was not more successful since  $R^2 = 0.66$  (FDL) and  $0.53$  (FMO). This last point could be rationalized by remarking that, due to symmetry,  $f^+(\text{Cl})$  and  $f^+(\text{Br})$  are both exactly equal to  $0.50$  (whatever the atomic partition) in  $\text{Cl}_2$  and  $\text{Br}_2$ , while the interactions energies are considerably different for the  $\text{Cl}_2 \cdots \text{NH}_3$  ( $-5.08$  kcal mol<sup>-1</sup>) and the  $\text{Br}_2 \cdots \text{NH}_3$  ( $-6.91$  kcal mol<sup>-1</sup>) complexes. It is the reason why when one compares the reactivity of sites on different molecules, it is strongly recommended to shift to the grand-canonical ensemble or to use composite descriptors (for instance by multiplying  $f^+(X)$  by a well chosen modulating global property that would help to discriminate between the  $\text{Cl}_2$  and  $\text{Br}_2$  molecules).

This can be achieved by considering the  $\omega_{\text{parabolic}}(X)$ ,  $\omega_{\text{cubic}}(X)$ ,  $\omega_G^+(X)$ ,  $\omega_{\text{C1}}^+(X)$ ,  $\omega_{\text{C2}}^+(X)$ , and  $\omega_{\text{elec-2}}(X)$  condensed

descriptors. In the FDL approach, the  $R^2$  values were found equal to  $0.85$ ,  $0.94$ ,  $0.87$ ,  $0.39$ ,  $0.35$ , and  $0.60$ , respectively, the “best” model (represented in the left panel of Figure 4) corresponding to the following equation (energies in kcal mol<sup>-1</sup>):

$$\Delta E_{\text{SCF,BSSE}}^{\text{bind}} = -373.46\omega_{\text{cubic}}(X) - 0.16 \quad (39)$$

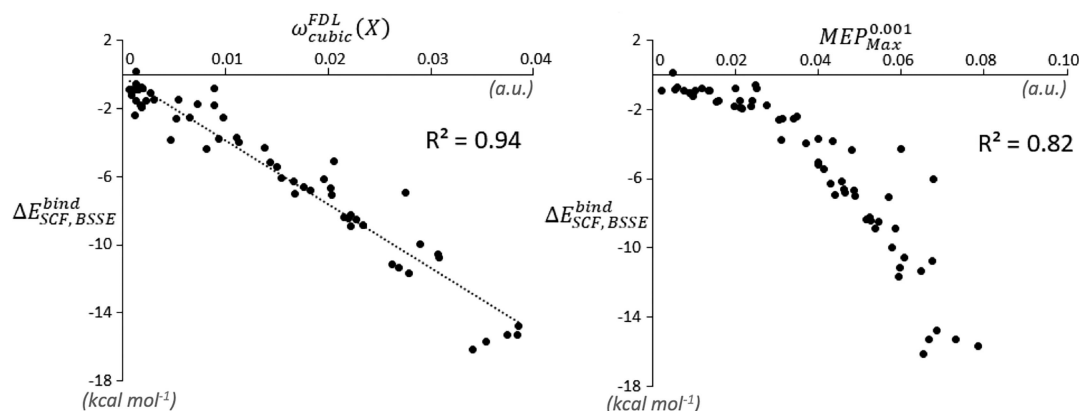
Within the FMO framework, the corresponding  $R^2$  values were  $0.80$ ,  $0.90$ ,  $0.76$ ,  $0.56$ ,  $0.81$ , and  $0.56$ , respectively. While the  $R^2$  coefficient between FDL and FMO approaches remains of the same order of magnitude for five descriptors, a considerable difference was observed for  $\omega_{\text{C2}}^+(X)$ . A closer inspection of the FDL values for these descriptors show that they are sometimes very small due to the  $\text{EA}^2$  term (see eq 35) and then close to the default SCF convergence threshold (of the  $10^{-5}$  or  $10^{-6}$  order of magnitude), making them irrelevant from the pure numerical point of view. The best model was obtained by (energies in kcal mol<sup>-1</sup>):

$$\Delta E_{\text{SCF,BSSE}}^{\text{bind}} = -231.01\omega_{\text{cubic}}(X) - 0.28 \quad (40)$$

We can also assess the usefulness of the condensed local softness  $s^+(X)$ , which is, as already discussed, the natural extension of the electrophilic Fukui function within the grand-canonical ensemble. The  $R^2$  values were found equal to  $0.79$  (FDL) and  $0.68$  (FMO), the correlation being thus lower than with the composite electrophilicity indices.

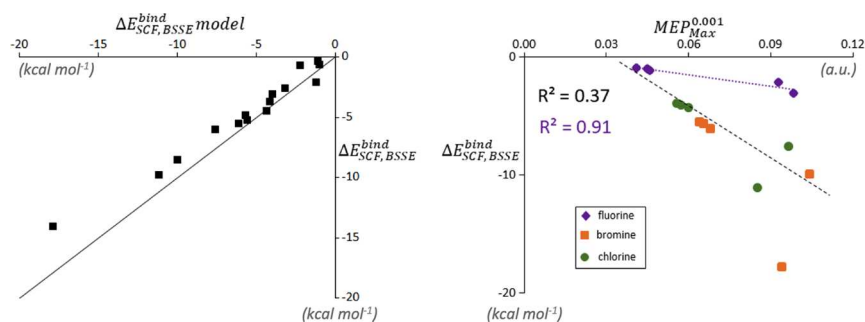
It is now instructive to compare such performances with that obtained for the MEP, which is by far the most used descriptor in the halogen bond community. More specifically, many studies employed the maximal MEP value located on the  $0.001$  au isodensity surface, which we will denote  $\text{MEP}_{\text{max}}^{0.001}$ . As shown in the right panel of Figure 4, even if there is a clear trend ( $\Delta E_{\text{SCF,BSSE}}^{\text{bind}}$  is all the more negative as  $\text{MEP}_{\text{max}}^{0.001}$  is more positive), the coefficient of determination is not so high since it equals  $0.82$ . More precisely, it clearly appears in the graph that weaker complexes ( $|\text{MEP}_{\text{max}}^{0.001}| < 2$  kcal mol<sup>-1</sup>) tend to correlate differently than stronger complexes. This might be due to the charge transfer character of stronger complexes that cannot be described with only electrostatics, as also conjectured by Hobza and co-workers.<sup>29</sup>

This MEP failure may thus seem in contradiction with the existing literature. We believe that it is not the case for the



**Figure 4.** (Left panel) BSSE-corrected binding energies (in kcal mol<sup>-1</sup>) of 62 halogen-bonded complexes with respect to QTAIM condensed electrophilicities of halogens (in atomic units) and corresponding linear regression. (Right panel) BSSE-corrected binding energies (in kcal mol<sup>-1</sup>) of 62 halogen-bonded complexes with respect to maximal molecular electrostatic potential values (in atomic units) on the  $0.001$  au isodensity surface and corresponding linear regression. M06-2X/aug-cc-pVTZ level of theory.





**Figure 5.** (Left panel) Comparison of BSSE-corrected binding energies (in  $\text{kcal mol}^{-1}$ ) predicted by the best linear model (eq 39) with the DFT computed ones for the test set of 15 halogen-bonded complexes. (Right panel) BSSE-corrected binding energies (in  $\text{kcal mol}^{-1}$ ) for the test set with respect to maximal molecular electrostatic potential values (in atomic units) on the 0.001 au isodensity surface.

following reason. Indeed, let us consider a subfamily of the R–X molecules, for instance the one composed by the nine complexes of the  $R_1R_2R_3C-O-Cl$  type. The correlation with  $MEP_{Max}^{0.001}$  is then particularly high ( $R^2 = 0.97$ ). When enlarging the data set by including the  $R_1R_2R_3C-O-Br$  compounds, the correlation begins to decrease, but it is still acceptable ( $R^2 = 0.91$ ). One might then conclude that the MEP descriptor is actually not the most versatile one, but that it certainly remains reliable for restricted (sometimes too much narrow) sets. Interestingly, one has also to observe that  $\omega_{cubic}^{FDL}(X)$  still outperforms the MEP for these subgroups since it affords a model with  $R^2 = 0.98$  for the nine chlorine complexes, and  $R^2 = 0.99$  for the 18 chlorine and bromine ones.

The model given by eq 39 (FDL approach) features a mean absolute error (MAE) equal to  $0.81 \text{ kcal mol}^{-1}$  for the data set of 62 complexes, while eq 40 (FMO approximation) gives a MAE equal to  $1.03 \text{ kcal mol}^{-1}$ . For the sake of comparison, the corresponding MAE for  $MEP_{Max}^{0.001}$  is significantly higher, amounting to  $1.35 \text{ kcal mol}^{-1}$ , while that for the FDL grand-canonical  $s^+(X)$  descriptor was  $1.53 \text{ kcal mol}^{-1}$ , hinting once more that composite electrophilicities may be preferable to standard grand-canonical derivatives.

We will now assess the robustness of the two selected best models by evaluating its performance on a validation test set. To this aim, and in order to enlarge the scope of application of the electrophilicity indices, we consider R–X molecules for which halogen bonds can also be formed with amonia when the halogen atom in R–X is fluorine. Five of such R–F compounds were then selected: NC–F, NCCC–F, NCCCC–F and two  $BeCl_2$  complexes following Yáñez and co-workers' strategy of using beryllium bonds to generate or enhance  $\sigma$ -holes:  $BeCl_2-NC-F$  et  $BeCl_2-NO_2-F$ .<sup>100</sup> It is indeed known that fluorine bonds albeit rare are not impossible.<sup>101,102</sup>

The corresponding analogues for Cl and Br were also added, creating a set of 15 new complexes that were fully optimized (note that geometric linearity was obtained for all these XBs, as revealed by the R–X–N angle values close to  $180^\circ$ ). As expected, the fluorine complexes were found to be weaker than the chlorine and bromine counterparts in each case, the  $|\Delta E_{SCF, BSSE}^{bind}(R-F\cdots NH_3)| < |\Delta E_{SCF, BSSE}^{bind}(R-Cl\cdots NH_3)| < |\Delta E_{SCF, BSSE}^{bind}(R-Br\cdots NH_3)|$  hierarchy being always valid. Besides, the maximal (in absolute value) interaction energy ( $17.80 \text{ kcal mol}^{-1}$ ) was reached for the  $BeCl_2NO_2Br\cdots NH_3$  complex, in perfect agreement with Yáñez' approach.

The left part of Figure 5 illustrates the performances of eq 39, involving  $\omega_{cubic}(X)$ , for the validation subgroup. The corresponding MAE was found equal to  $1.51 \text{ kcal mol}^{-1}$ .

Equation 40, instead, provides a MAE equal to  $3.21 \text{ kcal mol}^{-1}$ . It is then instructive to compare the MEP performance (see right panel in Figure 5) for these 15 complexes: no significant correlation is visible. Finally, we can now assess the accuracy of our monolinear models on all molecules (that-is-to-say by gathering the training and the test sets together). For the 75 studied complexes, the MAE is  $0.98 \text{ kcal mol}^{-1}$  (FDL) and  $1.71 \text{ kcal mol}^{-1}$  (FMO). These results stress that FMO and FDL frameworks can significantly differ (the first one not taking into account orbital relaxation upon electron number variation), in particular for the strongest halogen bonds. They also enlighten the robustness of the reported prediction by means of eq 39, which is largely independent of the halogen and substituent types and which can cover a large range of XB strengths (from the weakest to the strongest ones).

Finally, we used this validation set in order to assess the accuracy of the chosen DFT protocol. To this aim, we considered two (frozen core) post-Hartree–Fock strategies (both using the aug-cc-pVTZ basis set): (i) single-point (SP) calculations on the DFT geometries at second-order Møller–Plesset perturbation theory (MP2)<sup>103</sup> and at the coupled cluster level using single and double substitutions with noniterative triple excitations<sup>104</sup> (CCSD(T)), (ii) MP2 and CCSD(T) SP calculations on the MP2 fully optimized geometries.

The corresponding BSSE-corrected interaction energies are gathered in Table S11. We found that the average absolute deviation between the CCSD(T) (respectively MP2) bonding energies between DFT and MP2 geometries is equal to  $0.11 \text{ kcal mol}^{-1}$  (respectively,  $0.12 \text{ kcal mol}^{-1}$ ) for these 15 systems, indicating that the DFT geometries can be considered as reliable. It is then interesting to estimate the MAE for the full M06-2X results with respect to the CCSD(T) ones (here at MP2 geometries) for the three different bonding regimes. The first MAE, corresponding to  $|\Delta E_{SCF, BSSE}^{bind}| < 5 \text{ kcal mol}^{-1}$ , equals  $0.21 \text{ kcal mol}^{-1}$ . For complexes with  $5 < |\Delta E_{SCF, BSSE}^{bind}| < 10 \text{ kcal mol}^{-1}$ , the MAE was found to be equal to  $0.40 \text{ kcal mol}^{-1}$ . Only the two strongest bonds are a little bit less satisfactorily described by M06-2X because the beryllium bond is also itself strong, but the correct trend is absolutely conserved. The MAE for the full set is equal to  $0.48 \text{ kcal mol}^{-1}$  with a very high coefficient of determination value ( $R^2 = 0.996$ ), corroborating previous findings by Varadwaj and co-workers<sup>105</sup> about the accuracy of the M06-2X exchange–correlation functional for this kind of complexes.

## 5. CONCLUSION

In summary, this work, which involved an unusual large number of halogen-bonded complexes, reports compelling evidence for the ability of conceptual DFT descriptors to predict chemical bond properties. In particular, several expressions for global and local electrophilicities were here discussed (in particular the importance of the third order correction), and their correlation with interaction energies of halogen bonds was assessed. It was found that the QTAIM condensed cubic electrophilicity index was in particular successful in predicting bond strength for a wide variety of complexes, independently of the considered halogen atom or of its substituents. This versatility strongly contrasts with the electrostatic descriptors that should be used only within restricted (and sometimes too small) subgroups. All these results foster the use of conceptual DFT not only for reactivity purposes, but for many other phenomena that are actually driven by electronic properties.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpca.9b10233>.

Numerical values for each calculated descriptor, Cartesian coordinates for all complexes, and selected geometric parameters. Binding energies at the MP2 and CCSD(T) levels (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

**Vincent Tognetti** – Normandy University, COBRA UMR 6014 & FR 3038, Université de Rouen INSA Rouen, CNRS 76821 Mont St Aignan. Cedex, France; [orcid.org/0000-0002-0292-3509](https://orcid.org/0000-0002-0292-3509); Email: [vincent.tognetti@univ-rouen.fr](mailto:vincent.tognetti@univ-rouen.fr)

**Laurent Joubert** – Normandy University, COBRA UMR 6014 & FR 3038, Université de Rouen INSA Rouen, CNRS 76821 Mont St Aignan. Cedex, France; Email: [laurent.joubert@univ-rouen.fr](mailto:laurent.joubert@univ-rouen.fr)

### Author

**Guillaume Hoffmann** – Normandy University, COBRA UMR 6014 & FR 3038, Université de Rouen INSA Rouen, CNRS 76821 Mont St Aignan. Cedex, France

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jpca.9b10233>

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The authors would like to gratefully acknowledge the LABEX SynOrg and the Normandy Region for funding and support, and the Centre Régional Informatique et d'Applications Numériques de Normandie (CRIANN) for computing resources.

## ■ REFERENCES

- (1) Colin, J.; De Claubry, G. Sur les combinaisons de l'iode avec les substances végétales et animales. *Anal. Chim.* **1814**, 90, 87–100.
- (2) Colin, M. Note sur quelques combinaisons de l'iode. *Annal. Chim.* **1814**, 91, 252–272.

- (3) Guthrie, F. On the iodide of iodammonium. *J. Chem. Soc.* **1863**, 16, 239–244.
- (4) Desiraju, G. R.; Ho, P. S.; Kloo, L.; Legon, A. C.; Marquardt, R.; Metrangola, P.; Politzer, P.; Resnati, G.; Rissanen, K. Definition of the halogen bond (IUPAC recommendations 2013). *Pure Appl. Chem.* **2013**, 85, 1711–1713.
- (5) Meyer, F.; Dubois, P. Halogen bonding at work: Recent applications in synthetic chemistry and materials science. *CrystEngComm* **2013**, 15, 3058–3071.
- (6) Saccone, M.; Catalano, L. Halogen bonding beyond crystals in materials science. *J. Phys. Chem. B* **2019**, 123, 9281–9290.
- (7) Berger, G.; Soubhye, J.; Meyer, F. Halogen bonding in polymer science: From crystal engineering to functional supramolecular polymers and materials. *Polym. Chem.* **2015**, 6, 3559–3580.
- (8) Zhang, C.; Wu, Y.; Xu, X.; Li, Y.; Feng, Li.; Wu, G. Synthesis of Polyisobutylene with  $\alpha$ -amino terminal group by combination of cationic polymerization with alkylation. *J. Polym. Sci., Part A: Polym. Chem.* **2008**, 45, 936–946.
- (9) Rissanen, K. Halogen bonded supramolecular complexes and networks. *CrystEngComm* **2008**, 10, 1107–1113.
- (10) Mukherjee, A.; Tothadi, S.; Desiraju, G. R. Halogen bonds in crystal engineering: Like hydrogen bonds yet different. *Acc. Chem. Res.* **2014**, 47, 2514–2524.
- (11) Metrangola, P.; Meyer, F.; Pilati, T.; Resnati, G.; Terraneo, G. Halogen bonding in supramolecular chemistry. *Angew. Chem., Int. Ed.* **2008**, 47, 6114–6127.
- (12) Langton, M. J.; Robinson, S. W.; Marques, I.; Félix, V.; Beer, P. D. Halogen bonding in water results in enhanced anion recognition in acyclic and rotaxane hosts. *Nat. Chem.* **2014**, 6, 1039–1043.
- (13) Robinson, S. W.; Mustoe, C. L.; White, N. G.; Brown, A.; Thompson, A. L.; Kennepohl, P.; Beer, P. D. Evidence for halogen bond covalency in acyclic and interlocked halogen-bonding receptor anion recognition. *J. Am. Chem. Soc.* **2015**, 137 (1), 499–507.
- (14) Wilcken, R.; Zimmermann, M. O.; Lange, A.; Joerger, A. C.; Boeckler, F. M. Principles and applications of halogen bonding in medicinal chemistry and chemical biology. *J. Med. Chem.* **2013**, 56, 1363–1388.
- (15) Lu, Y.; Shi, T.; Wang, Y.; Yang, H.; Yan, X.; Luo, X.; Jiang, H.; Zhu, W. Halogen bonding - A novel interaction for rational drug design? *J. Med. Chem.* **2009**, 52, 2854–2862.
- (16) Kniep, F.; Jungbauer, S. H.; Zhang, Q.; Walter, S. M.; Schindler, S.; Schnapperelle, I.; Herdtweck, E.; Huber, S. M. Organocatalysis by neutral multidentate halogen-bond donors. *Angew. Chem., Int. Ed.* **2013**, 52, 7028–7032.
- (17) Bruckmann, A.; Pena, M. A.; Bolm, C. Organocatalysis through halogen-bond activation. *Synlett* **2008**, 2008 (6), 900–902.
- (18) Bulfield, D.; Huber, S. M. Halogen bonding in organic synthesis and organocatalysis. *Chem. - Eur. J.* **2016**, 22, 14434–14450.
- (19) Vaidyanathan, G.; Zalutsky, M. Astatine radiopharmaceuticals: prospects and problems. *Curr. Radiopharm.* **2010**, 1, 177–196.
- (20) Guo, N.; Maurice, R.; Teze, D.; Graton, J.; Champion, J.; Montavon, G.; Galland, N. Experimental and computational evidence of halogen bonds involving astatine. *Nat. Chem.* **2018**, 10, 1–7.
- (21) Galland, N.; Montavon, G.; Le Questel, J. Y.; Graton, J. Quantum calculations of At-mediated halogen bonds: On the influence of relativistic effects. *New J. Chem.* **2018**, 42, 10510–10517.
- (22) Clark, T.  $\sigma$ -Holes. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, 3, 13–20.
- (23) Clark, T.; Hennemann, M.; Murray, J. S.; Politzer, P. Halogen bonding: The  $\sigma$ -hole: Proceedings of 'Modeling interactions in biomolecules II', Prague, September 5th–9th. *J. Mol. Model.* **2007**, 13, 291–296.
- (24) Alkorta, I.; Sánchez-Sanz, G.; Elguero, J. Linear free energy relationships in halogen bonds. *CrystEngComm* **2013**, 15, 3178–3186.
- (25) Riley, K. E.; Murray, J. S.; Fanfrlík, J.; Rezáč, J.; Solá, R. J.; Concha, M. C.; Ramos, F. M.; Politzer, P. Halogen bond tunability I: The effects of aromatic fluorine substitution on the strengths of halogen-bonding interactions involving chlorine, bromine, and iodine. *J. Mol. Model.* **2011**, 17, 3309–3318.

- (26) Riley, K. E.; Murray, J. S.; Politzer, P.; Concha, M. C.; Hobza, P. Br...O complexes as probes of factors affecting halogen bonding: Interactions of bromobenzenes and bromopyrimidines with acetone. *J. Chem. Theory Comput.* **2009**, *5*, 155–163.
- (27) Jalbout, A. F.; Li, X.; Abou-Rachid, H. Analytical potential energy functions and theoretical spectroscopic constants for MX/MX- (M = Ge, Sn, Pb; X = O, S, Se, Te, Po) and LuA (A = H, F) systems: density functional theory calculations. *Int. J. Quantum Chem.* **2007**, *107*, 522–539.
- (28) Brinck, T.; Murray, J. S.; Politzer, P. Surface electrostatic potentials of halogenated methanes as indicators of directional intermolecular interactions. *Int. J. Quantum Chem.* **1992**, *44*, 57–64.
- (29) Kolář, M. H.; Hobza, P. Computer modeling of halogen bonds and other  $\sigma$ -hole interactions. *Chem. Rev.* **2016**, *116*, 5155–5187.
- (30) Rezáč, J.; Riley, K. E.; Hobza, P. Benchmark calculations of noncovalent interactions of halogenated molecules. *J. Chem. Theory Comput.* **2012**, *8*, 4285–4292.
- (31) Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S. Halogen bonds in biological molecules. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 16789–16794.
- (32) Stone, A. J. Are halogen bonded structures electrostatically driven? *J. Am. Chem. Soc.* **2013**, *135*, 7005–7009.
- (33) Riley, K. E.; Hobza, P. The relative roles of electrostatics and dispersion in the stabilization of halogen bonds. *Phys. Chem. Chem. Phys.* **2013**, *15*, 17742–17751.
- (34) Wolters, L. P.; Bickelhaupt, F. M. Halogen bonding versus hydrogen bonding: a molecular orbital perspective. *ChemistryOpen* **2012**, *1*, 96–105.
- (35) Adhikari, U.; Scheiner, S. Sensitivity of pnictogen, chalcogen, halogen and H-bonds to angular distortions. *Chem. Phys. Lett.* **2012**, *532*, 31–35.
- (36) Kolář, M.; Hostaš, J.; Hobza, P. The strength and directionality of a halogen bond are co-determined by the magnitude and size of the  $\sigma$ -hole. *Phys. Chem. Chem. Phys.* **2014**, *16*, 9987–9996.
- (37) Thacker, J. C. R.; Popelier, P. L. A. The ANANKE relative energy gradient (REG) method to automate IQA analysis over configurational change. *Theor. Chem. Acc.* **2017**, *136*, 1–13.
- (38) Orangi, N.; Eskandari, K.; Thacker, J. C. R.; Popelier, P. L. A. Directionality of halogen bonds: an interacting quantum atoms (IQA) and relative energy gradient (REG) study. *ChemPhysChem* **2019**, *20*, 1922–1930.
- (39) Syzgantseva, O. A.; Tognetti, V.; Joubert, L. On the physical nature of halogen bonds: A QTAIM study. *J. Phys. Chem. A* **2013**, *117*, 8969–8980.
- (40) Tognetti, V.; Joubert, L. On atoms-in-molecules energies from Kohn–Sham calculations. *ChemPhysChem* **2017**, *18*, 2675–2687.
- (41) Tognetti, V.; Joubert, L. Unraveling charge transfer processes with the quantum theory of atoms-in-molecules. *Theor. Chem. Acc.* **2016**, *135*, 1–13.
- (42) Yahia-Ouahmed, M.; Tognetti, V.; Joubert, L. Intramolecular halogen bonding: an interacting quantum atoms study. *Theor. Chem. Acc.* **2016**, *135*, 1–15.
- (43) Tognetti, V.; Joubert, L. Following halogen bond formation with Bader's atoms-in-molecules theory. In *Challenges and Advances in Computational Chemistry and Physics*; Chauvin, R., Lepetit, C., Silvi, B., Alikhani, E., Eds.; Springer International Publishing: Switzerland, 2016; Vol. 22, p 435.
- (44) Yahia-Ouahmed, M.; Tognetti, V.; Joubert, L. Halogen-halogen interactions in perhalogenated ethanes: An interacting quantum atoms study. *Comput. Theor. Chem.* **2015**, *1053*, 254–262.
- (45) Varadwaj, P.; Varadwaj, A.; Marques, H. Halogen bonding: a halogen-centered noncovalent interaction yet to be understood. *Inorganics* **2019**, *7*, 40.
- (46) Robertson, C. C.; Wright, J. S.; Carrington, E. J.; Perutz, R. N.; Hunter, C. A.; Brammer, L. Hydrogen bonding: Vs. halogen bonding: The solvent decides. *Chem. Sci.* **2017**, *8*, 5392–5398.
- (47) Geerlings, P.; De Proft, F.; Langenaeker, W. Conceptual density functional theory. *Chem. Rev.* **2003**, *103*, 1793–1873.
- (48) Chermette, H. Chemical reactivity indexes in density functional theory. *J. Comput. Chem.* **1999**, *20*, 129–154.
- (49) Parr, R. G.; Yang, W. *Density-functional theory*; Oxford University Press: New York, Oxford, 1989.
- (50) Fuentealba, P.; Cárdenas, C. Density functional theory of chemical reactivity. *Chem. Model.* **2014**, *11*, 151–174.
- (51) Parr, R. G.; Donnelly, R. A.; Levy, M.; Palke, W. E. Electronegativity: The density functional viewpoint. *J. Chem. Phys.* **1978**, *68*, 3801–3807.
- (52) Parr, R. G.; Pearson, R. G. Absolute hardness: companion parameter to absolute electronegativity. *J. Am. Chem. Soc.* **1983**, *105*, 7512–7516.
- (53) Pearson, R. G. Hard and soft acids and bases. *J. Am. Chem. Soc.* **1963**, *85*, 3533–3539.
- (54) Pearson, R. G.; Sen, K. D. Eds. *Chemical hardness. structure and bonding*; Springer: Berlin, Heidelberg, 1993; Vol. 80.
- (55) Miranda-Quintana, R. A.; Ayers, P. W. Note: Maximum hardness and minimum electrophilicity principles. *J. Chem. Phys.* **2018**, *148*, 2017–2019.
- (56) Tognetti, V.; Morell, C.; Joubert, L. Quantifying electro/nucleophilicity by partitioning the dual descriptor. *J. Comput. Chem.* **2015**, *36*, 649–659.
- (57) Kozuch, S.; Martin, J. M. L. Halogen bonds: Benchmarks and theoretical analysis. *J. Chem. Theory Comput.* **2013**, *9*, 1918–1931.
- (58) Pinter, B.; Nagels, N.; Herrebout, W. A.; De Proft, F. Halogen bonding from a hard and soft acids and bases perspective: Investigation by using density functional theory reactivity indices. *Chem. - Eur. J.* **2013**, *19*, 519–530.
- (59) Nagels, N.; Geboes, Y.; Pinter, B.; De Proft, F.; Herrebout, W. A. Tuning the halogen/hydrogen bond competition: A spectroscopic and conceptual dft study of some model complexes involving CHF<sub>2</sub>I. *Chem. - Eur. J.* **2014**, *20*, 8433–8443.
- (60) Hoffmann, G.; Tognetti, V.; Joubert, L. Can molecular and atomic descriptors predict the electrophilicity of Michael acceptors? *J. Mol. Model.* **2018**, *24*, 281.
- (61) Fuentealba, P.; Parr, R. G. Higher-order derivatives in density-functional theory, especially the hardness derivative  $\partial\eta/\partial N$ . *J. Chem. Phys.* **1991**, *94*, 5559–5564.
- (62) Morell, C.; Grand, A.; Toro-Labbé, A.; Chermette, H. Is hyper-hardness more chemically relevant than expected? *J. Mol. Model.* **2013**, *19*, 2893–2900.
- (63) Maynard, A. T.; Huang, M.; Rice, W. G.; Covell, D. G. Reactivity of the HIV-1 nucleocapsid protein p7 zinc finger domains from the perspective of density-functional theory. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 11578–11583.
- (64) Parr, R. G.; Szentpály, L. V.; Liu, S. Electrophilicity index. *J. Am. Chem. Soc.* **1999**, *121*, 1922–1924.
- (65) Chattaraj, P. K.; Sarkar, U.; Elango, M.; Parthasarathi, R.; Subramanian, V. Electrophilicity as a possible descriptor of the kinetic behaviour. <http://arXiv:physics/0509089v1> 2005.
- (66) Yildiz, D.; Bozkaya, U. Assessment of the extended Koopmans' theorem for the chemical reactivity: Accurate computations of chemical potentials, chemical hardnesses, and electrophilicity indices. *J. Comput. Chem.* **2016**, *37*, 345–353.
- (67) Mulliken, R. S. A new electroaffinity scale; Together with data on valence states and on valence ionization potentials and electron affinities. *J. Chem. Phys.* **1934**, *2*, 782–793.
- (68) Cárdenas, C.; Ayers, P.; De Proft, F.; Tozer, D. J.; Geerlings, P. Should negative electron affinities be used for evaluating the chemical hardness? *Phys. Chem. Chem. Phys.* **2011**, *13*, 2285–2293.
- (69) Perdew, J. P.; Parr, R. G.; Levy, M.; Balduz, J. L. J. DFT for fractional particle number: Derivative Discontinuity. *Phys. Rev. Lett.* **1982**, *49*, 1691.
- (70) Gázquez, J. L.; Cedillo, A.; Vela, A. Electrodonating and electroaccepting powers. *J. Phys. Chem. A* **2007**, *111*, 1966–1970.
- (71) Chamorro, E.; Duque-Noreña, M.; Pérez, P. A comparison between theoretical and experimental models of electrophilicity and nucleophilicity. *J. Mol. Struct.: THEOCHEM* **2009**, *896*, 73–79.



- (72) Chamorro, E.; Duque-Noreña, M.; Pérez, P. Further relationships between theoretical and experimental models of electrophilicity and nucleophilicity. *J. Mol. Struct.: THEOCHEM* **2009**, *901*, 145–152.
- (73) Parr, R. G.; Yang, W. Density functional approach to the frontier-electron theory of chemical reactivity. *J. Am. Chem. Soc.* **1984**, *106*, 4049–4050.
- (74) Morell, C.; Grand, A.; Toro-Labbé, A. New dual descriptor for chemical reactivity. *J. Phys. Chem. A* **2005**, *109*, 205–212.
- (75) Chattaraj, P. K.; Maiti, B.; Sarkar, U. Philicity: A unified treatment of chemical reactivity and selectivity. *J. Phys. Chem. A* **2003**, *107*, 4973–4975.
- (76) Robles, A.; Franco-Pérez, M.; Gázquez, J. L.; Cárdenas, C.; Fuentealba, P. Local electrophilicity. *J. Mol. Model.* **2018**, *24*, 245.
- (77) Geerlings, P.; Fias, S.; Boisdenghien, Z.; De Proft, F. Conceptual DFT: chemistry from the linear response function. *Chem. Soc. Rev.* **2014**, *43*, 4989–5008.
- (78) Guégan, F.; Mignon, P.; Tognetti, V.; Joubert, L.; Morell, C. Dual descriptor and molecular electrostatic potential: Complementary tools for the study of the coordination chemistry of ambiphilic ligands. *Phys. Chem. Chem. Phys.* **2014**, *16*, 15558–15569.
- (79) Zielinski, F.; Tognetti, V.; Joubert, L. Condensed descriptors for reactivity: A methodological study. *Chem. Phys. Lett.* **2012**, *527*, 67–72.
- (80) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; The International Series of Monographs on Chemistry: No. 22; Oxford University Press: New York, 1990.
- (81) Popelier, P. L. *Atoms in Molecules: An Introduction*; Pearson Education: Harlow, U.K., 2000.
- (82) Eskandari, K.; Lesani, M. Does fluorine participate in halogen bonding? *Chem. - Eur. J.* **2015**, *21*, 4739–4746.
- (83) Matta, C. F.; Hernández-Trujillo, J.; Tang, T. H.; Bader, R. F. W. Hydrogen - Hydrogen bonding: A stabilizing interaction in molecules and crystals. *Chem. - Eur. J.* **2003**, *9*, 1940–1951.
- (84) Poater, J.; Solà, M.; Bickelhaupt, F. M. Hydrogen-hydrogen bonding in planar biphenyl, predicted by atoms-in-molecules theory, does not exist. *Chem. - Eur. J.* **2006**, *12*, 2889–2895.
- (85) Grabowski, S. J. QTAIM characteristics of halogen bond and related interactions. *J. Phys. Chem. A* **2012**, *116*, 1838–1845.
- (86) Wang, W. Halogen bond involving hypervalent halogen: CSD search and theoretical study. *J. Phys. Chem. A* **2011**, *115*, 9294–9299.
- (87) Forni, A. Experimental and theoretical study of the Br...N halogen bond in complexes of 1,4-dibromotetrafluorobenzene with dipyrilidyl derivatives. *J. Phys. Chem. A* **2009**, *113*, 3403–3412.
- (88) Hirshfeld, F. L. Bonded-atom fragments for describing molecular charge densities. *Theor. Chim. Acta* **1977**, *44*, 129–138.
- (89) Becke, A. D. A multicenter numerical integration scheme for polyatomic molecules. *J. Chem. Phys.* **1988**, *88*, 2547–2553.
- (90) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, non-covalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other function. *Theor. Chem. Acc.* **2008**, *120*, 215–241.
- (91) Shaw, R. A.; Hill, J. G. A simple model for halogen bond interaction energies. *Inorganics* **2019**, *7*, 19.
- (92) Boys, S. F.; Bernardi, F. T. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553–566.
- (93) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision D.01; Gaussian, Inc.: Wallingford CT, 2013.
- (94) AIMAll version 19.02.13; Todd A. Keith, TK Gristmill Software: Overland Park KS, USA, 2019; ([aim.tkgristmill.com](http://aim.tkgristmill.com)).
- (95) Yang, W.; Mortier, W. J. The use of global and local molecular parameters for the analysis of the gas-phase basicity of amines. *J. Am. Chem. Soc.* **1986**, *108*, 5708–5711.
- (96) Hoffmann, G.; Tognetti, V.; Joubert, L. On the influence of dynamical effects on reactivity descriptors. *Chem. Phys. Lett.* **2019**, *724*, 24–28.
- (97) Politzer, P.; Murray, J. S. Enthalpy and entropy factors in gas phase halogen bonding: Compensation and competition. *CrystEngComm* **2013**, *15*, 3145–3150.
- (98) Izato, Y.; Matsugi, A.; Koshi, M.; Miyake, A. A simple heuristic approach to estimate the thermochemistry of condensed-phase molecules based on the polarizable continuum model. *Phys. Chem. Chem. Phys.* **2019**, *21*, 18920–18929.
- (99) Besora, M.; Vidossich, P.; Lledós, A.; Ujaque, G.; Maseras, F. Calculation of reaction Free energies in solution: a comparison of current approaches. *J. Phys. Chem. A* **2018**, *122*, 1392–1399.
- (100) Brea, O.; Mö, O.; Yáñez, M.; Alkorta, I.; Elguero, J. Creating  $\sigma$ -holes through the formation of beryllium bonds. *Chem. - Eur. J.* **2015**, *21*, 12676–12682.
- (101) Pavan, M. S.; Durga Prasad, K.; Guru Row, T. N. Halogen bonding in fluorine: Experimental charge density study on intermolecular F...F and F...S donor-acceptor contacts. *Chem. Commun.* **2013**, *49*, 7558–7560.
- (102) Metrangola, P.; Murray, J. S.; Pilati, T.; Politzer, P.; Resnati, G.; Terraneo, G. Fluorine-centered halogen bonding: A factor in recognition phenomena and reactivity. *Cryst. Growth Des.* **2011**, *11* (9), 4238–4246.
- (103) Möller, C.; Plesset, M. S. Note on an approximation treatment for many-electron systems. *Phys. Rev.* **1934**, *46*, 618–622.
- (104) Purvis, G. D., III; Bartlett, R. J. A full coupled-cluster singles and doubles model – the inclusion of disconnected triples. *J. Chem. Phys.* **1982**, *76*, 1910–1918.
- (105) Varadwaj, A.; Marques, H. M.; Varadwaj, P. R. Is the fluorine in molecules dispersive? Is molecular electrostatic potential a valid property to explore fluorine-centered non-covalent interactions? *Molecules* **2019**, *24*, 379.



### 3. Conclusion et perspectives

Dans cet article, les descripteurs d'électrophilie issus de la CDFT, dont l'indice d'électrophilie cubique condensé  $\omega_{cubic}(\vec{r})$ , ont présenté une plus grande étendue d'application que les descripteurs couramment utilisés (le potentiel électrostatique moléculaire). De plus, contrairement aux travaux présents dans la littérature qui se concentrent usuellement sur l'étude de certaines familles de molécules ou sur un nombre réduit de systèmes étudiés, nous nous sommes intéressés à plus de 77 complexes non covalents impliquant des liaisons halogène avec des atomes de fluor, de chlore et de brome.

L'indice d'électrophilie cubique  $\omega_{cubic}$ , qui ajoute la perturbation au troisième ordre à la définition parabolique de l'électrophilie de Parr, semble apporter une information supplémentaire vis-à-vis de la prédiction de propriétés électroniques, confirmant ainsi l'utilisation de ces descripteurs et l'importance des descripteurs d'ordre supérieur.

En perspective de ces travaux, grâce à la démonstration récente de la capacité des complexes d'astate à former des liaisons halogènes, on pourrait envisager d'appliquer la méthodologie de cet article sur des complexes d'astate. On pourrait ainsi, avec les descripteurs quantiques issus de la DFT conceptuelle, effectuer une comparaison des propriétés d'un ensemble de complexes halogénés variés (comportant des atomes de *F*, *Cl*, *Br*, *I* et *At*).

CHAPITRE 6 : *Synthèse de cations iminium vinylogues :  
Application de l'échelle de Mayr aux électrophiles ambivalents*

|                                   |     |
|-----------------------------------|-----|
| 1. Introduction .....             | 163 |
| 2. Résultats et discussions ..... | 165 |
| 2.1 Synthèse .....                | 165 |
| 2.2 Caractérisation.....          | 168 |
| 2.3 Mesures cinétiques.....       | 170 |
| 3. Conclusion.....                | 173 |



## 1. Introduction

*Le travail réalisé ici a été mené en collaboration avec le groupe du LCMT (Laboratoire de Chimie Moléculaire et Thio-organique) (Sami Lakhdar, Annie Claude-Gaumont, William Lecroq) et le laboratoire COBRA (Guillaume Hoffmann, Vincent Tognetti, Laurent Joubert).*

Dans ce dernier chapitre, nous nous sommes intéressés à la synthèse et à la caractérisation de cations iminium vinylogues.

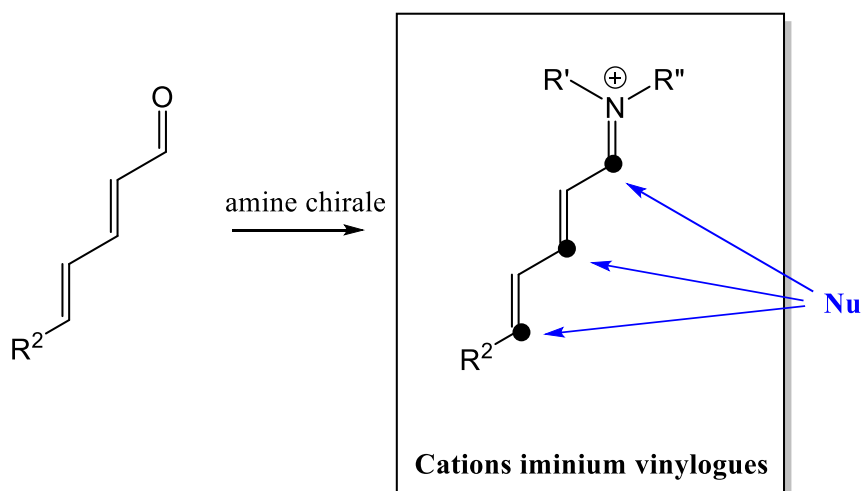


Figure 6.1: Formation des cations iminium vinylogues. Les points noirs correspondant aux centres électrophiles.

En effet, les cations iminium vinylogues sont des composés servant d'intermédiaires réactifs en organocatalyse [1, 2]. Ils utilisent le principe de catalyse par activation de l'ion iminium, entraînant au niveau orbitalaire un abaissement de l'énergie de la LUMO (que l'on appelle en anglais *LUMO-lowering*) du composé initial. Cette combinaison de composés vinylogues et d'ions iminium donne accès à un outil important en synthèse énantiosélective permettant de favoriser un site électrophile du composé initial. Bien que le phénomène inverse (*HOMO-raising*) soit très bien connu et décrit en organocatalyse, l'utilisation de cations iminium vinylogues pour le *LUMO-lowering* en est encore au stade initial.

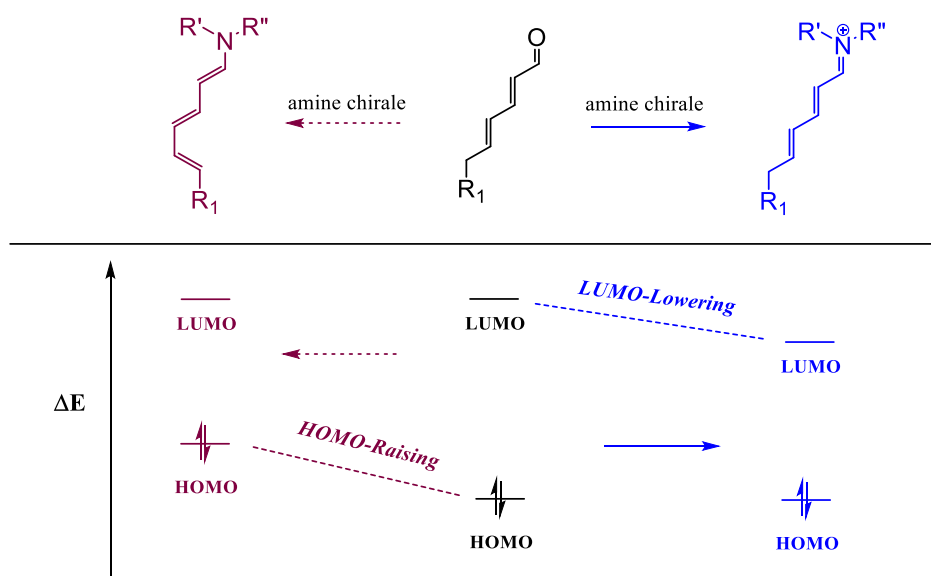


Figure 6.2: Schéma illustrant le principe de la catalyse par LUMO Lowering (en bleu) et HOMO Raising (en rouge)

Dans leurs principales utilisations en synthèse en « cascade » [3, 4], le résultat correspond en général à un mélange des différents produits. Le schéma suivant représente l'exemple de la synthèse en cascade de cœurs chromans [4], avec les 4 voies possibles lors du mécanisme :

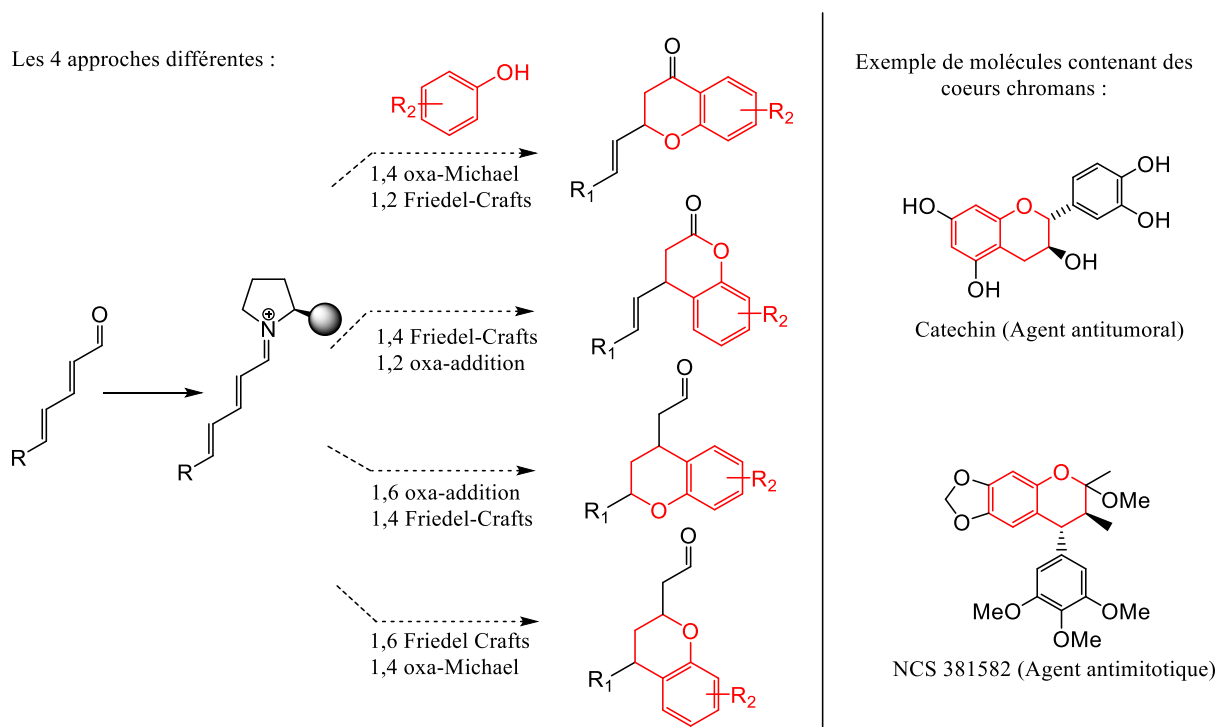


Figure 6.3: Exemple d'application des cations iminium vinylogues lors de synthèse en cascade [4].

Pour pallier ces problèmes mécanistiques lors de ces synthèses d'importance, le but initial de ce projet est d'appliquer l'échelle de Mayr aux différents sites électrophiles de ces catalyseurs, ce qui permettrait *in fine* d'effectuer un classement par ordre de réactivité de ces sites, pour différents ions iminium. Cela pourrait ainsi apporter une meilleure compréhension des facteurs contrôlant la régiosélectivité et permettre de guider la synthèse.

## 2. Résultats et discussion

### 2.1 Synthèse

La première partie concerne la synthèse des cations iminium vinylogues, obtenus ici à partir d'aldéhydes vinylogues *trans*, basée sur l'étude de Woodward *et al.* [4, 5]. Le schéma général de la synthèse totale réalisée est le suivant :

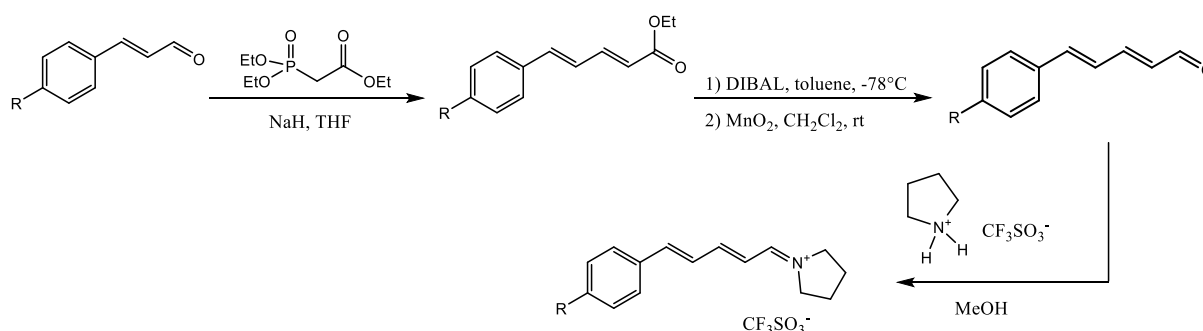


Figure 6.4: Synthèse totale des cations iminiums vinylogues

Quatre réactifs disponibles dans le commerce ont ainsi été considérés pour cette voie de synthèse **1-4**. La procédure générale pour l'obtention des quatre aldéhydes vinylogues **9-12** est ainsi présentée :

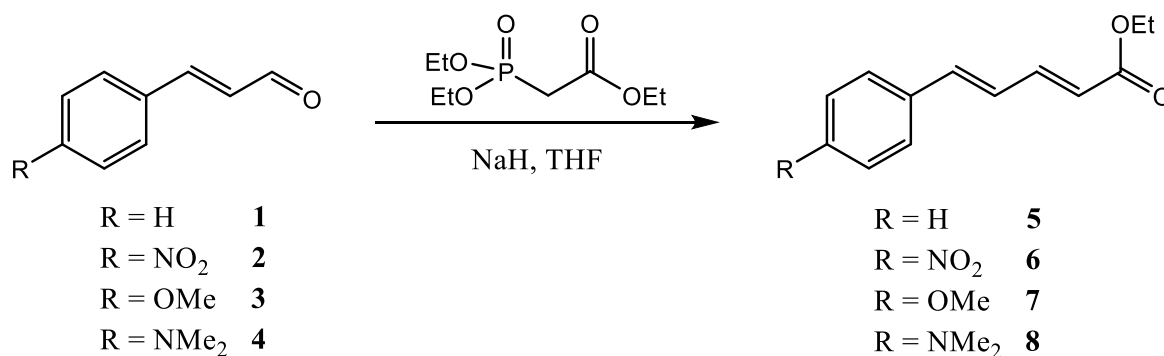


Figure 6.5 : Première étape de la synthèse

Dans un premier temps, les aldéhydes de départ sont transformés en esters par une réaction d'Horner–Wadsworth–Emmons [6]. Le triéthylphosphonoacétate (1,1 équivalents) est premièrement ajouté goutte à goutte à une solution de NaH (1,1 équivalents) dans le THF, à 0°C, sous agitation et sous argon. Après 1 heure, l'aldéhyde initial **1-4** (1 équivalent) est à son tour ajouté au mélange réactionnel au goutte à goutte. Le mélange réactionnel est ensuite ramené à température ambiante, puis neutralisé avec une solution de NH<sub>4</sub>Cl saturée. La réaction est ensuite extraite 3 fois avec l'éther diéthylique, puis lavée consécutivement avec de l'eau et de la saumure. Le produit est ensuite séché sur sulfate de magnésium anhydre (MgSO<sub>4</sub>), filtré, puis évaporé sous pression réduite. Les produits bruts obtenus sont ensuite utilisés pour la prochaine étape sans purification.

Dans un second temps, les esters obtenus précédemment **5-8** sont réduits en aldéhydes en deux étapes, grâce au DIBAL (l'hydruure de diisobutylaluminium).

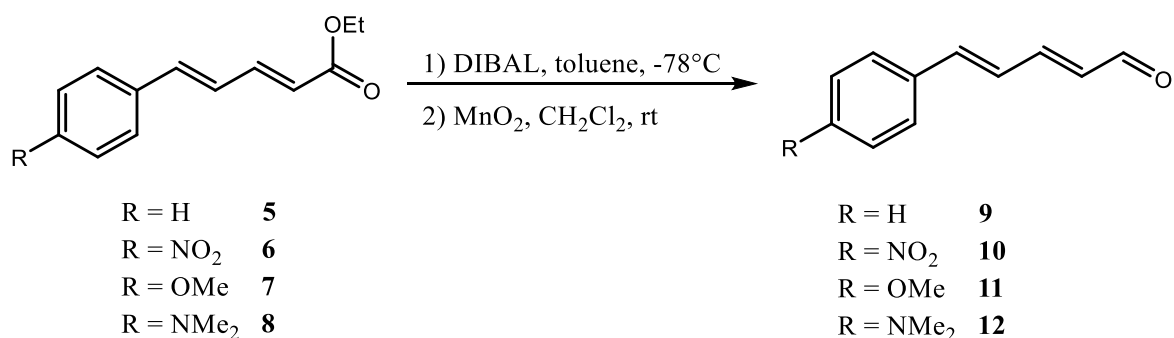


Figure 6.6 : Deuxième étape de la synthèse

Les esters précédemment obtenus **5-8** sont dissous dans une grande quantité de toluène sous argon et refroidis à -78°C (mélange carboglace-acétone). Puis, une faible quantité de DIBAL (3 équivalents) est ajoutée au goutte à goutte au mélange réactionnel. Après 30 minutes d'agitation, le milieu réactionnel est neutralisé avec un faible ajout d'eau au goutte à goutte, puis ramené à température ambiante. Le mélange réactionnel est ensuite ajouté à une solution de NaHCO<sub>3</sub>, MgSO<sub>4</sub> et d'acétate d'éthyle (EtOAc). Après filtration, le produit est alors évaporé puis séché sous vide pendant 24 heures.

Le produit de la première étape est alors dissous dans une grande quantité de dichlorométhane, auquel on ajoute un excès de MnO<sub>2</sub> (11 équivalents). Après deux heures d'agitation, la solution est filtrée, lavée avec de l'acétate d'éthyle puis évaporée sous pression réduite. Le produit brut est alors purifié par chromatographie sur colonne (avec un mélange de EtOAc/pentane 5/95).

Les rendements des réactions sont respectivement de 13% pour **9**, 41% pour **10**, 24% pour **11** et 38% pour **12**.

Enfin, la dernière étape consiste en la formation des cations iminium vinylogues à partir des aldéhydes obtenus. L'ammonium choisi ici pour la synthèse est un sel de pyrrolidine, que l'on obtient par traitement de la pyrrolidine avec de l'acide trifluorométhanesulfonique :

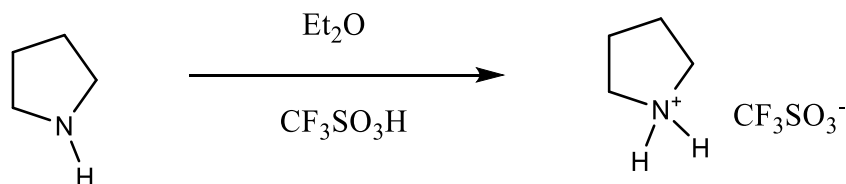


Figure 6.7: Synthèse du sel de pyrrolidine

La procédure employée ici est basée sur des travaux antérieurs réalisés par Lakhdar *et al.* sur des cations iminium [7].

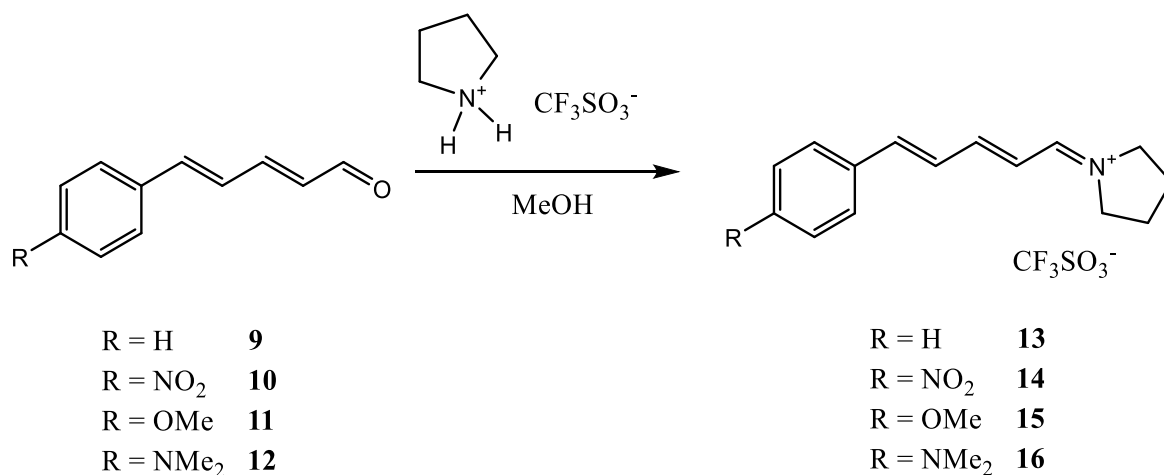


Figure 6.8: Dernière étape de la synthèse

Les aldéhydes sont initialement dissous dans une faible quantité de méthanol, à laquelle est ajoutée au goutte à goutte la solution du sel d'ammonium (1,5 équivalent). La réaction est ainsi réalisée à température ambiante et sous agitation pendant 24 heures. Dès lors, sur les quatre synthèses effectuées, seule la réaction avec l'aldéhyde **10** présente un changement de couleur (du jaune au marron foncé) et seule cette réaction présente, après une analyse RMN rapide (spectre proton réalisé sur un spectromètre RMN 500 MHz), une transformation du réactif initial. Le solide obtenu est lavé avec de l'éther diéthylique et séché sur rampe à vide. Le



rendement de cette étape est de 62%. La partie suivante concerne alors la caractérisation de la molécule **14** ainsi obtenue.

## 2.2 Caractérisation

La première caractérisation est effectuée sur le solide pur obtenu en fin de réaction. L'analyse RMN  $^1\text{H}$  (Résonance Magnétique Nucléaire) est ainsi réalisée avec un spectromètre BRUKER Avance III de 500 MHz, disponible au LCMT à l'ENSI de Caen avec Rémi Legay (Ingénieur de Recherche au LCMT) :

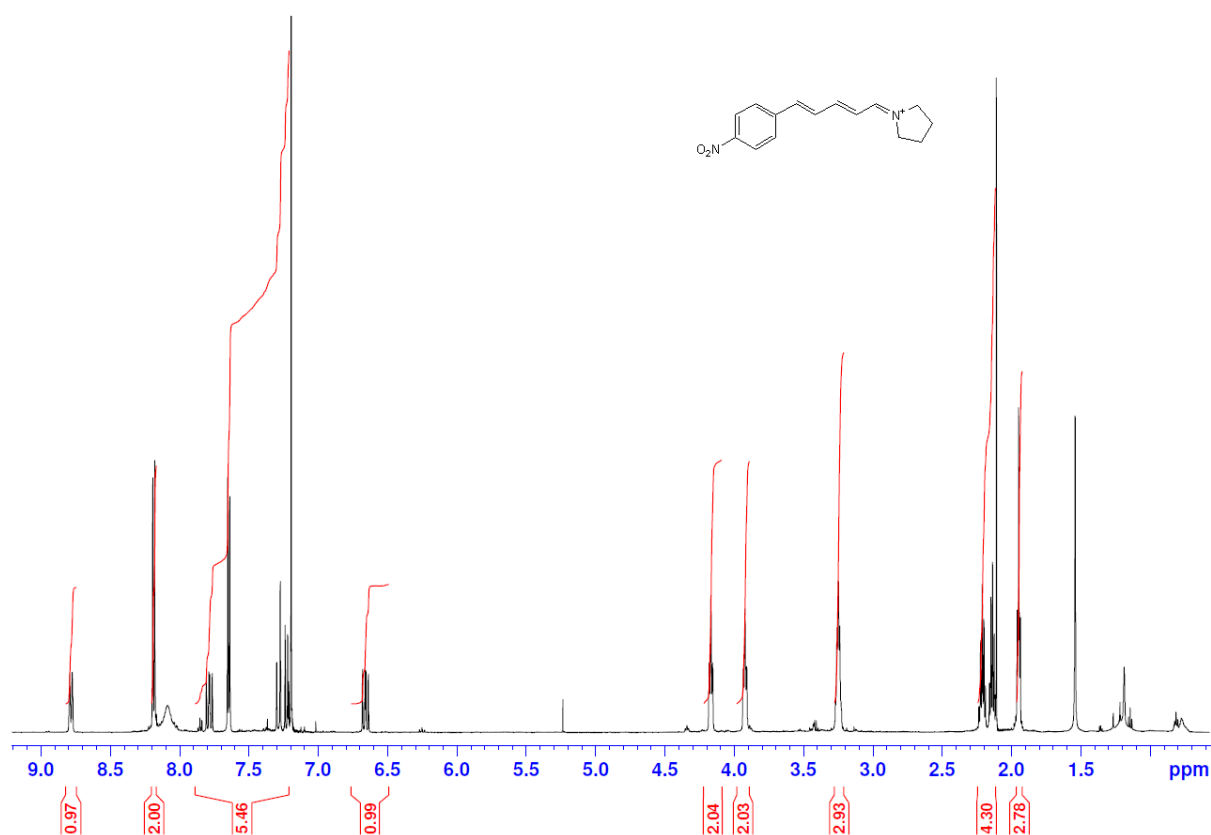


Figure 6.9: Spectre RMN  $^1\text{H}$  à 500 MHz du produit **14**. Le solvant utilisé est  $\text{CDCl}_3$ . Image générée par le Dr. Ghanem Hamdoun (Groupe du Pr. Oulyadi au laboratoire COBRA) avec TopSpin.

L'analyse du spectre permet de confirmer la présence de notre molécule dans notre mélange : il y a présence de 9 protons entre 6,5 et 8,7 ppm qui correspondent aux 4 protons du cycle aromatique ainsi que des 5 protons aliphatiques, le plus déblindé correspondant au proton proche de l'iminium. Le solvant utilisé ici est le  $\text{CDCl}_3$  dont le pic caractéristique est visible à 7,2 ppm. On peut également noter une faible présence d'impuretés dans notre mélange : acétone, eau, graisse ainsi que l'ammonium pyrrolidine.

Une analyse par spectrométrie de masse en complément d'une analyse élémentaire a également permis de confirmer la structure de notre molécule. L'analyse a été réalisée sur un instrument Acquity UPLC H-Class Xevo G2-XS QToF (WATERS) disponible au LCMT à l'ENSI de Caen par Karine Jarsale (Ingénieur de Recherche au LCMT) sur le solide obtenu après cristallisation. Le spectre suivant correspond au produit **14** :

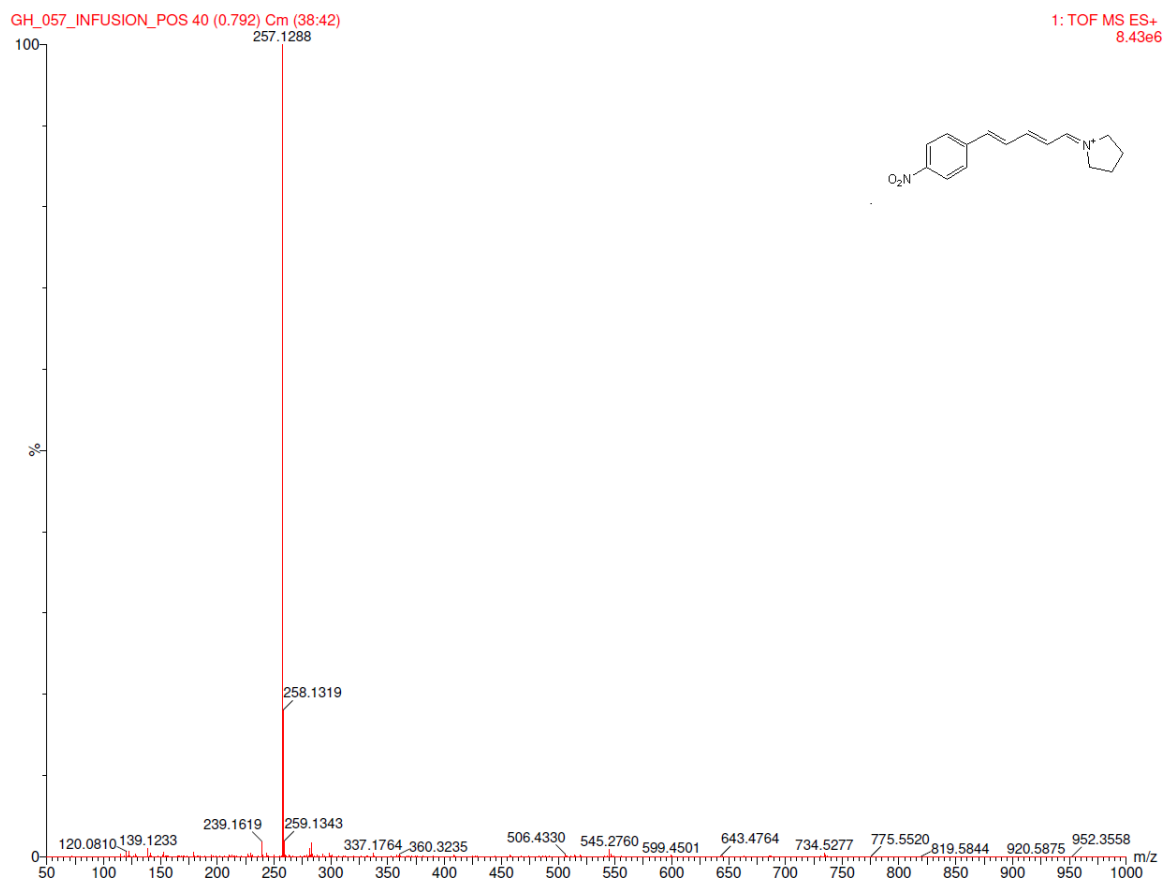


Figure 6.10: Spectre de masse, m/z 50 à 1000 en électrospray du composé **14**, ions positifs

La masse calculée ( $m/z$ ) initiale est de 257,1290 pour  $C_{15}H_{17}N_2O_2$  et la masse trouvée ( $m/z$ ) est de 257,1288. Cette analyse permet de confirmer la présence de notre produit attendu et permet de confirmer que la réaction s'est bien produite.

Enfin, pour la dernière caractérisation de la molécule **14**, des cristaux purs adaptés à l'analyse cristallographique par diffraction des rayons X ont été obtenus par diffusion lente (dans un cristalliseur), à partir d'une faible quantité de **14** solide dans une solution de dichlorométhane avec un faible ajout d'éther diéthylique :

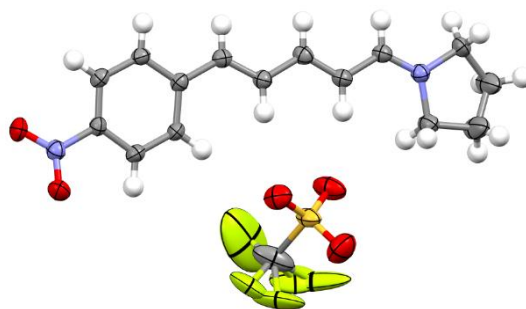


Figure 6.11 : Structure RX du composé **14**, générée avec SHELXL-2014

Les analyses de diffraction aux rayons X ont ainsi été réalisées sur un instrument Bruker APEXII Kappa CCD diffractomètre ( $\text{MoK}\alpha$   $\lambda = 0,71073 \text{ \AA}$  ; monochromateur en graphite) par Jean-François Lohier (Ingénieur de Recherche au LCMT). Pour résoudre la structure, les programmes utilisés sont SHELXS-97 et SHELXL-2014. Cette analyse permet de donner la structure exacte de notre molécule et de confirmer la prédominance de la structure *trans* de notre produit.

### 2.3 Mesures cinétiques

Après avoir obtenu le cation iminium vinylogue **14**, nous avons réalisé des mesures cinétiques afin de déterminer des paramètres expérimentaux d'électrophilie de Mayr sur les trois sites caractéristiques ( $C_A$ ,  $C_B$  et  $C_C$ ) :

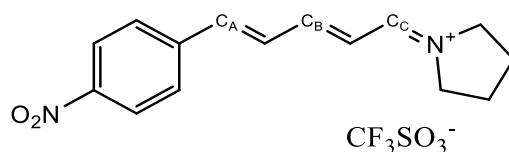


Figure 6.12: Représentation des trois sites électrophiles du cation iminium **14**

Afin de calibrer l'appareil de flux stoppé pour réaliser un suivi de l'absorbance en fonction du temps, il est nécessaire de mesurer la longueur d'onde d'absorbance maximale. La longueur d'onde maximale d'absorbance mesurée pour notre molécule est de  $\lambda_{\text{max}} = 375 \text{ nm}$  :

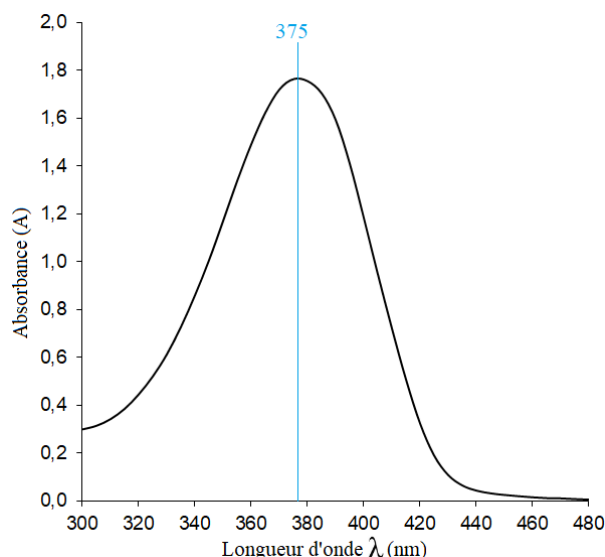


Figure 6.13: Spectre UV/Visible après pulsation dans le dichlorométhane ( $\text{CH}_2\text{Cl}_2$ ) de la molécule 14.  $\lambda_{\text{max}} = 375 \text{ nm}$

Nous avons ensuite sélectionné 4 nucléophiles disponibles dans la base de données de Mayr [8] afin de réaliser les mesures cinétiques pour la détermination du paramètre électrophile d'un premier site. En utilisant plusieurs concentrations, on peut alors tracer la courbe des constantes cinétiques observées expérimentalement  $k_{\text{obs}}$  en fonction des concentrations utilisées pour chaque mesure, afin d'obtenir la pseudo-constante du second ordre expérimental  $k_2$  (qui correspond à la pente de la courbe). La figure 7 représente ainsi la courbe pour la réaction entre le nucléophile 1 (le triphénylphosphane) et la molécule **14** :

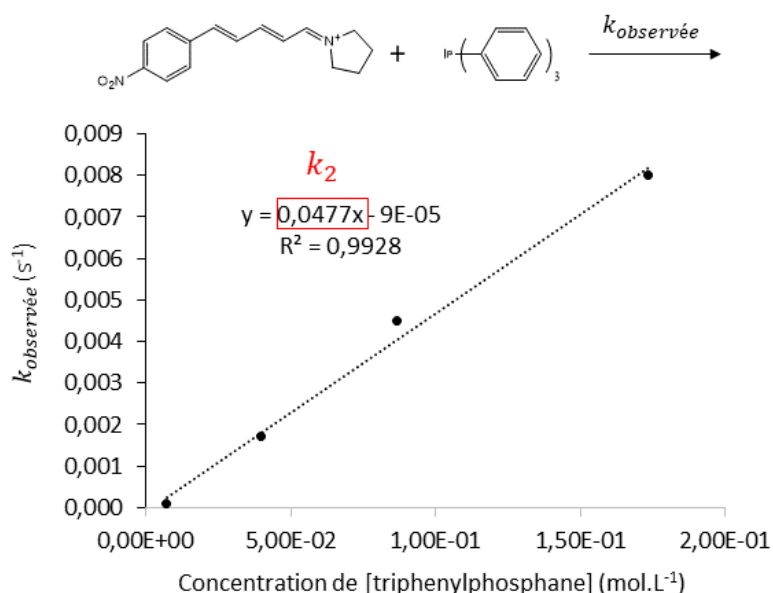
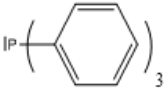
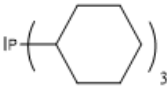
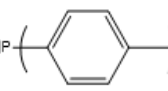
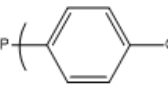


Figure 6.14 : Exemple de détermination expérimentale de la constante cinétique  $k_2$  pour un nucléophile

Les résultats pour les mesures cinétiques sont regroupés dans la table suivante :

Tableau 1 : Ensemble des résultats expérimentaux pour les 4 nucléophiles

| Nucléophiles                                                                       | $N / s_N$    | $k_2[14] \text{ (M}^{-1} \text{s}^{-1})$ |
|------------------------------------------------------------------------------------|--------------|------------------------------------------|
|   | 14,33 / 0,65 | 0,0477                                   |
|   | 14,64 / 0,68 | 0,1221                                   |
|   | 15,44 / 0,64 | 0,9201                                   |
|  | 16,17 / 0,62 | 4,7872                                   |

On peut alors utiliser l'équation de Mayr (équation 3.8) afin de déterminer expérimentalement l'électrophilie  $E$  du site réactionnel. Pour cela, on trace simplement la courbe de  $\frac{\log k_2}{s_N}$  en fonction de  $N$  :

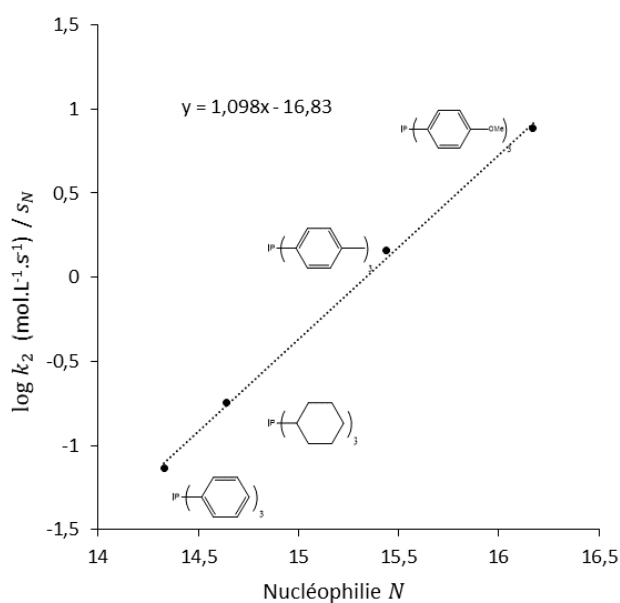


Figure 6.15 : Détermination expérimentale du paramètre d'électrophilie

La pente de la droite, qui est ici égale à 1,1, respecte la condition de l'équation de Mayr (la pente doit être de 1) [9, 10]. L'ordonnée à l'origine permet de retrouver expérimentalement  $E$  pour un site réactif qui est ici de  $-16,83$ .

### 3. Conclusion

Dans ces travaux, nous avons réussi à réaliser la première synthèse d'un nouveau cation iminium vinylogue stable que l'on a pu isoler et caractériser par différentes méthodes analytiques.

D'un point de vue de la synthèse, nous pourrions envisager d'employer un plus grand nombre d'aldéhydes à la première étape, mais également d'utiliser un plus grand nombre d'ammoniums à la dernière étape afin d'obtenir de nouveaux cations iminium vinylogues. La dernière étape de la synthèse (Figure 6.8), a été réalisée dans les mêmes conditions en essayant le catalyseur de MacMillan 1<sup>ère</sup> génération, mais n'a pas abouti sur la synthèse avec les quatre aldéhydes obtenus (pas de produit observé).

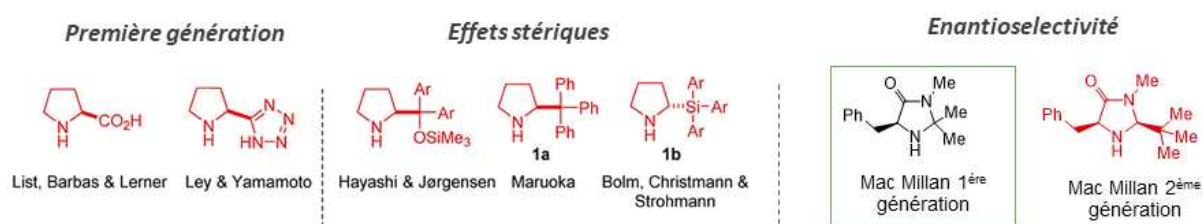


Figure 6.16: Exemple d'ammoniums d'intérêt stériques pouvant être utilisées pour la synthèse de nouveaux cations iminium vinylogues

Une fois la molécule **14** obtenue, nous avons réussi à déterminer un paramètre d'électrophilie par le biais de mesures cinétiques sur un des trois sites électrophiles, nous n'avons en revanche pas pu expérimentalement déterminer le site correspondant au paramètre obtenu. Il est nécessaire à cette étape de poursuivre le travail réalisé en effectuant une analyse des produits formés afin de déterminer le site électrophile correspondant.

Afin d'aller plus loin dans la détermination des paramètres électrophiles de la molécule **14**, nous avons effectué la caractérisation des sites électrophiles par approche théorique. La fonction de Fukui électrophile et le descripteur dual (voir Chapitre 2, partie 4.5.1) ont alors été calculés avec la fonctionnelle M06-2X et la base 6-311++G(2d,2p).

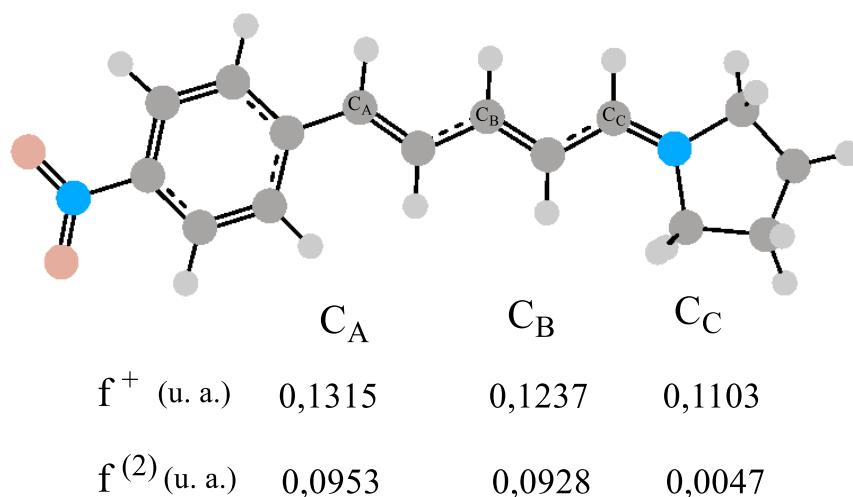


Figure 6.17: Valeurs des descripteurs calculées sur la molécule **14**. Image généré avec GaussView 5.0. Calcul réalisée avec Gaussian 09 au niveau M06-2X 6-311++G(2d,2p) et données en approximation FMO. Condensation QTAIM réalisée avec AIMALL v19.10.12.

Les résultats obtenus à partir des descripteurs suggèrent un classement d'électrophilie de  $C_A > C_B > C_C$ . On pourrait ainsi supposer que l'électrophilie mesurée expérimentalement correspondrait au site  $C_A$ .

Il reste néanmoins à continuer les travaux réalisés dans ce chapitre et d'effectuer la détermination expérimentale des autres paramètres d'électrophilie.

- [1] K. A. Ahrendt, C. J. Borths, D. W. C. MacMillan. *J. Am. Chem. Soc.* 122 (**2000**), 4243–4244
- [2] P. Melchiorre, M. Marigo, A. Carlone, G. Bartoli. *Angew. Chemie - Int. Ed.* 47 (**2008**), 6138–6171
- [3] X. Li, L. Yang, C. Peng, et al. *Chem. Commun.* 49 (**2013**), 8692–8694
- [4] P. H. Poulsen, K. S. Feu, B. M. Paz, et al. *Angew. Chemie - Int. Ed.* 54 (**2015**), 8203–8207
- [5] S. V. Ley, S. C. Smith, P. R. Woodward. *Tetrahedron* 48 (**1992**), 1145–1174
- [6] W. S. Wadsworth. *Org. React.* 73–253
- [7] S. Lakhdar, T. Tokuyasu, H. Mayr. *Angew. Chemie - Int. Ed.* 47 (**2008**), 8723–8726
- [8] B. Kempf, H. Mayr. *Chem. - A Eur. J.* 11 (**2005**), 917–927
- [9] H. Mayr. *Tetrahedron* 71 (**2015**), 5095–5111
- [10] H. Mayr, M. Patz. *Angew. Chemie Int. Ed. English* 33 (**1994**), 938–957





## Conclusion générale et perspectives

Ces travaux de thèse ont été consacrés premièrement à la prédiction théorique de l'électrophilie expérimentale. Dans un premier temps [1] nous avons montré, pour une série d'accepteurs de Michael, l'intérêt de prendre en compte plusieurs atomes et plusieurs descripteurs pour un seul site afin de prédire l'électrophilie. Les résultats obtenus nous ont amenés à chercher de nouvelles voies afin d'améliorer le modèle de prédiction. Dans une seconde étude [2], nous avons obtenu une meilleure précision du modèle, en montrant la forte influence d'une dynamique moléculaire sur les descripteurs de réactivité. Bien que ces calculs soient plus coûteux en temps, ils permettent, grâce à une respiration de la structure moléculaire, d'obtenir une représentation plus réaliste lors d'une réaction chimique. En dernier lieu, nous avons réussi à mettre au point un modèle performant grâce à des méthodes d'apprentissage supervisé (*machine learning*) [3]. Ces méthodes, qui commencent à émerger dans de nombreux travaux scientifiques, demeuraient faiblement explorées en combinaison avec la CDFT. Elles ont montré, dans nos travaux, leurs forts potentiels une fois combinées avec les descripteurs de réactivité.

Dans un second temps, les descripteurs issus de la CDFT se sont avérés prometteurs lors d'une étude de la liaison chimique. Plus particulièrement nous nous sommes intéressés à la liaison halogène. Les descripteurs d'électrophilie de la réactivité ont en effet ainsi été présentés en alternative aux descripteurs usuels employés [4].

Enfin, dans la dernière partie, une nouvelle voie de synthèse de cations iminium vinylogues est présentée. La molécule ainsi obtenue a pu être caractérisée par un ensemble de méthodes analytiques et son électrophilie de Mayr a été mesurée sur un site réactif que nous n'avons pas pu identifier. Enfin, nous avons initié théoriquement les prédictions d'électrophilie des sites réactifs de la molécule. En perspective de ces travaux, il convient d'obtenir un nombre plus important de ces cations iminium vinylogues et de réaliser la caractérisation complète de leurs sites électrophiles par mesures cinétiques, ainsi que d'appliquer par une approche théorique les méthodes de prédictions mises au point dans nos précédents travaux. Une fois leur réactivité connue, une synthèse en organocatalyse les impliquant pourrait être envisagée afin de démontrer leurs domaines d'application.

Bien que la CDFT connaisse un succès grandissant depuis son apparition, l'une des perspectives que l'on pourrait envisager serait de réaliser, sous forme d'une revue scientifique ou via une interface web, une base de données qui regrouperait l'ensemble des descripteurs et des indices de réactivité disponibles dans la littérature. Un tel travail préciserait alors, pour chacun des descripteurs, leurs principaux domaines d'application. Ceci permettrait, comme souhaité par R. G. Parr, de présenter à un novice dans le domaine « à *chaque problème chimique sa solution* » [7].

La perspective précédente va de pair avec l'ensemble des principes de réactivité disponibles au sein de la CDFT. On pourrait par la suite envisager la modélisation d'un ensemble de réactions nucléophile-électrophile et de suivre le long du chemin réactionnel les indices de réactivité. La validité d'un ensemble de principes de réactivité pour des réactions électrophile-nucléophile serait ainsi évalué et une comparaison de ces derniers pourrait être discutée. Enfin, nous pourrions envisager d'aller plus loin et de suivre le long du chemin réactionnel les différentes contributions énergétiques des réactifs grâce à la décomposition IQA. En combinaison avec les méthodes récentes d'identification des contributions les plus pertinentes (l'approche du gradient d'énergie relative REG [8]), on pourrait ainsi évaluer les termes les plus importants lors de ces réactions.

- [1] G. Hoffmann, V. Tognetti, L. Joubert. *J. Mol. Model.* 24 (2018), 281
- [2] G. Hoffmann, V. Tognetti, L. Joubert. *Chem. Phys. Lett.* 724 (2019), 24–28
- [3] G. Hoffmann, M. Balcilar, V. Tognetti, et al. *J. Comput. Chem.* 41 (2020), 2124–2136
- [4] G. Hoffmann, V. Tognetti, L. Joubert. *J. Phys. Chem. A* 124 (2020), 2090–2101
- [5] P. Geerlings, E. Chamorro, P. K. Chattaraj, et al. *Theor. Chem. Acc.* 139 (2020), 1–18
- [6] X. Guo, H. Mayr. *J. Am. Chem. Soc.* 136 (2014), 11499–11512
- [7] R. G. Parr. *Annu. Rev. Phys. Chem.* 46 (1995), 701–728
- [8] J. C. R. Thacker, P. L. A. Popelier, *Theor. Chem. Acc.* 136 (2017), 136, 86–98.

# Annexe 1 : Liste et formules des principaux descripteurs de réactivité employés

L'ensemble des formules pour les descripteurs issus de la DFT conceptuelle est donné ici dans l'approximation des orbitales frontières (FMO). Pour l'approche par différence finies et les formules appropriées, il convient de se reporter au premier article page 105.

Tableau 1 : Descripteurs globaux

| Nom                             | Formule                                                                                                             | Référence                                                                                                              |
|---------------------------------|---------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------|
| Potentiel d'ionisation          | $\mu_{FMO}^- = \varepsilon_{HOMO}$                                                                                  | J. P. Perdew, R. G. Parr, M. Levy, J. L. Balduz, <i>Phys. Rev. Lett.</i> , <b>1982</b> , 49, 1691-1694                 |
| Affinité électronique           | $\mu_{FMO}^+ = \varepsilon_{LUMO}$                                                                                  |                                                                                                                        |
| Potentiel chimique électronique | $\mu_{FMO} = \frac{\varepsilon_{HOMO} + \varepsilon_{LUMO}}{2}$                                                     | R. G. Parr, R. A. Donnelly, M. Levy, W. E. Palke, <i>J. Chem. Phys.</i> , <b>1978</b> , 68, 3801-3807                  |
| Dureté moléculaire              | $\eta_{FMO} = \varepsilon_{LUMO} - \varepsilon_{HOMO}$                                                              | R. G. Parr, R. G. Pearson, <i>J. Am. Chem. Soc.</i> , <b>1983</b> , 105, 7512-7516                                     |
| Mollesse moléculaire            | $S_{FMO} = \frac{1}{\eta_{FMO}}$                                                                                    | W. Yang, R. G. Parr, <i>Proc. Natl. Acad. Sci. USA.</i> , <b>1985</b> , 82, 6723-6726                                  |
| Hyper dureté                    | $\gamma_{FMO} = \varepsilon_{LUMO} - 2\varepsilon_{HOMO} + \varepsilon_{HOMO-1}$                                    | P. Fuentealba, R. G. Parr, <i>J. Chem. Phys.</i> , <b>1991</b> , 94, 5559-5564                                         |
| Indice d'électrophilie          | $\omega_{FMO} = \frac{\mu_{FMO}^2}{2\eta_{FMO}}$                                                                    | R. G. Parr, L. von Szentpaly, S. Liu, <i>J. Am. Chem. Soc.</i> , <b>1999</b> , 121, 1922-1924                          |
| Pouvoir électro-accepteur       | $\omega_{FMO}^+ = \frac{(\varepsilon_{HOMO} + 3\varepsilon_{LUMO})^2}{16(\varepsilon_{LUMO} - \varepsilon_{HOMO})}$ | J. L. Gázquez, A. Cedillo, A. Vela, <i>J. Phys. Chem. A.</i> , <b>2007</b> , 111, 1966-1970                            |
| Pouvoir électro-donneur         | $\omega_{FMO}^- = \frac{(3\varepsilon_{HOMO} + \varepsilon_{LUMO})^2}{16(\varepsilon_{LUMO} - \varepsilon_{HOMO})}$ |                                                                                                                        |
| Nucléofugalité                  | $\Delta E_{FMO}^{nucleo} = \frac{(\mu_{FMO} + \eta_{FMO})^2}{2\eta_{FMO}}$                                          | P. W. Ayers, J. S. M. Anderson, J. I. Rodriguez, Z. Jawed, <i>Phys. Chem. Chem. Phys.</i> , <b>2005</b> , 7, 1918-1925 |

|                                          |                                                                                                       |                                                                                                                        |
|------------------------------------------|-------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------|
| Électrofugalité                          | $\Delta E_{FMO}^{electro} = \frac{(\mu_{FMO} - \eta_{FMO})^2}{2 \eta_{FMO}}$                          | P. W. Ayers, J. S. M. Anderson, J. I. Rodriguez, Z. Jawed, <i>Phys. Chem. Chem. Phys.</i> , <b>2005</b> , 7, 1918-1925 |
| Hyper mollesse                           | $S_{FMO}^{(2)} = -\frac{\gamma_{FMO}}{(\eta_{FMO})^3}$                                                | P. W. Ayers, R. G. Parr, <i>J. Chem. Phys.</i> , <b>2008</b> , 129, 054111                                             |
| Indice d'électrophilie relative          | $\Delta\omega_{FMO}^{\pm} = \omega_{FMO}^{+} + \omega_{FMO}^{-}$                                      | P. K. Chattaraj, A. Chakraborty, S. Giri, <i>J. Phys. Chem. A.</i> , <b>2009</b> , 113, 10068-10074                    |
| Capacité maximale d'acceptance de charge | $\Delta N_{FMO}^{Max} = -\frac{\mu_{FMO}}{\eta_{FMO}}$                                                | R. G. Parr, L. von Szentpaly, S. Liu, <i>J. Am. Chem. Soc.</i> , <b>1999</b> , 121, 1922-1924                          |
| Inverse de l'hyper dureté                | $P_{FMO} = \frac{1}{\gamma_{FMO}}$                                                                    | P. Fuentealba, R. G. Parr, <i>J. Chem. Phys.</i> , <b>1991</b> , 94, 5559-5564                                         |
| Indice de réactivité intrinsèque         | $IRI_{FMO} = \frac{\varepsilon_{HOMO} + \varepsilon_{LUMO}}{\varepsilon_{LUMO} - \varepsilon_{HOMO}}$ | S. I. Kiyooka, D. Kaneno, R. Fujiyama, <i>Tetrahedron</i> , <b>2013</b> , 69, 4247-4258                                |

Tableau 2 : Descripteurs locaux.

| Nom                                               | Formule                                                                | Référence                                                                                                                  |
|---------------------------------------------------|------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------|
| Fonction de Fukui (pour une attaque nucléophile)  | $f_{FMO}^{+}(\vec{r}) = \rho_{LUMO}(\vec{r})$                          | R. G. Parr, W. Yang, <i>J. Am. Chem. Soc.</i> , <b>1984</b> , 106, 4049-4050                                               |
| Fonction de Fukui (pour une attaque électrophile) | $f_{FMO}^{-}(\vec{r}) = \rho_{HOMO}(\vec{r})$                          |                                                                                                                            |
| Descripteur dual                                  | $f_{FMO}^{(2)}(\vec{r}) = \rho_{LUMO}(\vec{r}) - \rho_{HOMO}(\vec{r})$ | C. Morell, A. Grand, A. Toro-Labbé, <i>J. Phys. Chem. A.</i> , <b>2005</b> , 109, 205-212                                  |
| Descripteur multiphilique                         | $\omega_{FMO} f_{FMO}^{(2)}(\vec{r})$                                  | J. Padmanabhan, R. Parthasarathi, V. Subramanian, P. K. Chattaraj, <i>J. Phys. Chem. A.</i> , <b>2007</b> , 111, 1358-1361 |
| Fonctions de Fukui composites                     | $\mu_{FMO}^{+} f_{FMO}^{+}(\vec{r})$                                   | A. Toro-Labbe, P. Jaque, J. S. Murray, P. Politzer, <i>Chem. Phys. Lett.</i> , <b>2005</b> , 407, 143-146                  |
|                                                   | $\mu_{FMO}^{-} f_{FMO}^{-}(\vec{r})$                                   |                                                                                                                            |

|                                                                  |                                                                                                                                                                    |                                                                                                                               |
|------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|
| Indice de philicité locale<br>(pour une attaque<br>nucléophile)  | $\omega_{FMO} f_{FMO}^+(\vec{r})$                                                                                                                                  | P. K. Chattaraj, B. Maiti, U. Sarkar,<br><i>J. Phys. Chem. A.</i> , <b>2003</b> , 107, 4973–<br>4975                          |
| Indice de philicité locale<br>(pour une attaque<br>électrophile) | $\omega_{FMO} f_{FMO}^-(\vec{r})$                                                                                                                                  |                                                                                                                               |
| Fonctions de Fukui<br>composites                                 | $\omega_{FMO}^+ f_{FMO}^+(\vec{r})$                                                                                                                                | A. Toro-Labbe, P. Jaque, J. S.<br>Murray, P. Politzer, <i>Chem. Phys.<br/>Lett.</i> , <b>2005</b> , 407, 143–146              |
|                                                                  | $\omega_{FMO}^- f_{FMO}^-(\vec{r})$                                                                                                                                |                                                                                                                               |
| Mollesse locale                                                  | $s_{FMO}^+(\vec{r}) = S f_{FMO}^+(\vec{r})$                                                                                                                        | R. K. Roy, S. Krishnamurti, P.<br>Geerlings, S. Pal, <i>J. Phys. Chem. A.</i> ,<br><b>1998</b> , 102, 3746–3755               |
|                                                                  | $s_{FMO}^-(\vec{r}) = S f_{FMO}^-(\vec{r})$                                                                                                                        | R. K. Roy, S. Krishnamurti, P.<br>Geerlings, S. Pal, <i>J. Phys. Chem. A.</i> ,<br><b>1998</b> , 102, 3746–3755               |
| Descripteur dual dans<br>l'ensemble grand<br>canonique           | $S_{FMO} f_{FMO}^{(2)}(\vec{r})$                                                                                                                                   | C. Morell, A. Grand, A. Toro-Labbé,<br><i>J. Phys. Chem. A.</i> , <b>2005</b> , 109, 205–<br>212                              |
| Electrophilie relative                                           | $s_{K_{FMO}}^+(\vec{r}) = \frac{s_{FMO}^+(\vec{r})}{s_{FMO}^-(\vec{r})}$                                                                                           | R. K. Roy, S. Krishnamurti, P.<br>Geerlings, S. Pal, <i>J. Phys. Chem. A.</i> ,<br><b>1998</b> , 102, 3746–3755               |
| Nucléophilie relative                                            | $s_{K_{FMO}}^-(\vec{r}) = \frac{s_{FMO}^-(\vec{r})}{s_{FMO}^+(\vec{r})}$                                                                                           | R. K. Roy, S. Krishnamurti, P.<br>Geerlings, S. Pal, <i>J. Phys. Chem. A.</i> ,<br><b>1998</b> , 102, 3746–3755               |
| Hyper mollesse locale                                            | $S_{FMO}^{(2)} f_{FMO}^{(2)}(\vec{r}) = (S_{FMO})^2 \times f_{FMO}^{(2)}(\vec{r})$                                                                                 | P. W. Ayers, R. G. Parr, <i>J. Chem.<br/>Phys.</i> , <b>2008</b> , 129, 054111                                                |
| Electrophilie locale                                             | $\Delta\rho_{FMO}^+(\vec{r}) = \frac{\mu_{FMO}}{\eta_{FMO}} f_{FMO}^+(\vec{r}) + \frac{1}{2} \left( \frac{\mu_{FMO}}{\eta_{FMO}} \right)^2 f_{FMO}^{(2)}(\vec{r})$ | C. Morell, J. L. Gázquez, A. Vela, F.<br>Guégan, H. Chermette, <i>Phys. Chem.<br/>Chem. Phys.</i> , <b>2014</b> , 26832-26842 |
| Nucléophilie locale                                              | $\Delta\rho_{FMO}^-(\vec{r}) = \frac{\mu_{FMO}}{\eta_{FMO}} f_{FMO}^-(\vec{r}) + \frac{1}{2} \left( \frac{\mu_{FMO}}{\eta_{FMO}} \right)^2 f_{FMO}^{(2)}(\vec{r})$ | C. Morell, J. L. Gázquez, A. Vela, F.<br>Guégan, H. Chermette, <i>Phys. Chem.<br/>Chem. Phys.</i> , <b>2014</b> , 26832-26842 |
| Potentiel chimique locale                                        | $\mu_{FMO}(\vec{r}) = \frac{1}{2} (\mu_{FMO}^+ f_{FMO}^+(\vec{r}) + \mu_{FMO}^- f_{FMO}^-(\vec{r}))$                                                               | J. Soto-Delgado, A. Aizman, L. R.<br>Domingo, R. Contreras, <i>Chem. Phys.<br/>Lett.</i> , <b>2010</b> , 499, 272–277         |
| Dureté locale                                                    | $\eta_{FMO}(\vec{r}) = \mu_{FMO}^+ f_{FMO}^+(\vec{r}) - \mu_{FMO}^- f_{FMO}^-(\vec{r})$                                                                            | L. Meneses, W. Tiznado, R.<br>Contreras, P. Fuentealba, <i>Chem.<br/>Phys. Lett.</i> , <b>2004</b> , 383, 181–187             |

|                                  |                                 |                                                                                |
|----------------------------------|---------------------------------|--------------------------------------------------------------------------------|
| Fonctions de Fukui<br>composites | $P_{FMO}f_{FMO}^+(\vec{r})$     | P. Fuentealba, R. G. Parr, <i>J. Chem. Phys.</i> , <b>1991</b> , 94, 5559-5564 |
|                                  | $P_{FMO}f_{FMO}^-(\vec{r})$     |                                                                                |
| Descripteur dual<br>composite    | $P_{FMO}f_{FMO}^{(2)}(\vec{r})$ |                                                                                |





## **Résumé :**

L'étude des descripteurs de réactivité globaux, locaux et non locaux d'un système réactif est d'une importance capitale afin de comprendre la réactivité de la totalité des processus chimiques lors d'une réaction. Le but de cette thèse a ainsi été de mettre au point de nouveaux descripteurs de réactivités, ainsi que des modèles de prédiction basés sur ces derniers, afin d'étudier la réactivité chimique. Les principales méthodes théoriques employées ont été la Théorie de la fonctionnelle de la densité conceptuelle (CDFT) et la théorie quantique « *Atoms in Molecules* » (QTAIM) qui sont toutes deux basées sur la densité électronique. Notre domaine d'étude se place principalement dans le cadre de l'échelle expérimentale de Mayr, qui permet par le biais de mesures cinétiques d'effectuer un classement des molécules par ordre de réactivité. Dans un premier temps, de grandes avancées ont été réalisées durant cette thèse vis-à-vis de la prédiction théorique de l'électrophilie expérimentale des accepteurs de Michael. Puis dans un second temps, nous nous sommes intéressées à l'application des descripteurs de réactivité sur la liaison chimique, et particulièrement la liaison halogène. Enfin, une partie de synthèse réalisée au cours de cette thèse est présentée, en proposant une nouvelle voie de synthèse des cations iminium vinylogues.

Mots clés : CDFT – QTAIM – Electrophilie – Addition de Michael – Descripteurs - Réactivité

## **Abstract :**

The study of global, local and non-local reactivity descriptors of a reactive system is of paramount importance in order to understand the reactivity of all chemical processes during a reaction. The goal of this thesis was then to develop new reactivity descriptors, as well as prediction models based on them, in order to study chemical reactivity.

The main theoretical methods used were the Conceptual Density Functional Theory (CDFT) and Quantum Theory of "Atoms in Molecules" (QTAIM), which are both based on electron density. Our field of study is mainly within the framework of the Mayr experimental scale, which allows, through kinetic measurements a classification of molecules in order of reactivity. In the first part, great advances were made during this thesis with respect to the theoretical prediction of experimental electrophilicity of Michael acceptors. Then in a second step, we looked at the application of reactivity descriptors on the chemical bond, and in particular the halogen bond. Finally, a part of synthesis carried out during the course of this thesis is presented, by proposing a new way of synthesis of vinylogous iminium cations.

Keywords : CDFT – QTAIM – Electrophilicity – Michael addition– Descriptors - Reactivity