HAL
open science

Vers une sélection guidée des descripteurs de réactivité

Aël Cador

To cite this version:

Aël Cador. Vers une sélection guidée des descripteurs de réactivité. Chimie organique. Normandie Université, 2022. Français. NNT : 2022NORMR027 . tel-03899835

HAL Id: tel-03899835

https://theses.hal.science/tel-03899835

Submitted on 15 Dec 2022

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

Vers une sélection guidée des descripteurs de réactivité

Présentée et soutenue par
AEL CADOR

Thèse soutenue le 19/09/2022 devant le jury composé de		
M. CHRISTOPHE MORELL	PROFESSEUR DES UNIVERSITES, UNIVERSITE LYON 1 CLAUDE BERNARD	Rapporteur du jury
MME EMILIE-LAURE ZINS	MAITRE DE CONFERENCES HDR, SORBONNE UNIVERSITE	Rapporteur du jury
M. PAUL POPELIER	PROFESSEUR DES UNIVERSITES, Univ. de Sciences de Manchester (RU)	Membre du jury
MME ISABELLE MILAZZO	PROFESSEUR DES UNIVERSITES, Université de Rouen Normandie	Président du jury
M. LAURENT JOUBERT	PROFESSEUR DES UNIVERSITES, Université de Rouen Normandie	Directeur de thèse
M. VINCENT TOGNETTI	MAITRE DE CONFERENCES, Université de Rouen Normandie	Co-directeur de thèse

Thèse dirigée par LAURENT JOUBERT (CHIMIE ORGANIQUE, BIOORGANIQUE, REACTIVITE, ANALYSE) et VINCENT TOGNETTI (CHIMIE ORGANIQUE, BIOORGANIQUE, REACTIVITE, ANALYSE)

UNIVERSITÉ DE ROUEN NORMANDIE

ed 508 NC Normande de Chimie

Laboratoire Cobra Chimie organique, Bioorganique, Réactivité et Analyse.

Remerciements

Je tiens tout d'abord à remercier la Région Normandie et le Labex SynOrg pour le financement de cette thèse, ainsi que les membres du jury, en particulier les rapporteurs, le Pr Christophe Morell et la Dr Émilie-Laure Zins pour avoir accepté d'évaluer ces travaux.

Je remercie mes directeurs de thèse, le Pr Laurent Joubert et le Dr Vincent Tognetti, pour leur présence et leur accompagnement continuel au cours de ces trois années de thèse, ainsi que pour tous leurs conseils. Merci également au Dr Guillaume Hoffmann et à Sébastien Menant pour leur aide inestimable lors du début et de la fin de ma thèse et pour m'avoir donné, même pour un temps, le sentiment d'appartenir à un groupe de chercheurs.

Je tiens également à remercier le Pr Paul Popelier ainsi que Leonardo Duarte et Fabio Falcioni pour leur aide, particulièrement lors de mon séjour à Manchester en février 2020, et pendant le reste de ma thèse.

Merci à mes parents et à ma famille pour leur soutien pendant ces trois années dont la majeure partie a été plutôt mouvementée en raison des conditions sanitaires exceptionnelles.

Merci à mes amis de longue date, Adrien, Nathan, Joséphine, Nicolas, Émeline, Karolina, Salomé, Louisiane, pour leur soutien au cours de cette période difficile.

Merci à mes amis de la toile, que ce soit chez Cronos : Esme, Mirta, Gash, Draca, Aza, Guth, Bog, Grant, Grim, Halrek, Krikk, Maghen, Ytrakx, ou sur Sirannon : Joly, Nago, Khao, Lomë, Edward, Kaels ; sans oublier la Fondation SCP : Hero, Slow, Uplyx, Gabi, Henry, Stan, Maël, Ilu, Andro, D, Macro, Archi, Cyrielle, et tous les autres auteurs, traducteurs et critiqueurs que j'ai pu connaître là-bas mais aussi Natan, Fariose, Voratox, Fropp, Fana, Blood et le reste de la bande. Sans eux, je n'aurais sans doute pas eu le courage d'aller jusqu'au bout des choses lors des périodes d'isolement.

Merci également à tous mes collègues du laboratoire COBRA, doctorants, post-doctorants et permanents : Aurélien, Valentina, Leïla, Oscar, Olivier, Johann, Julien, Charlotte, Clément, Estelle, Nathaniel, Linlin, Solveig, Katia, Marian, Youssou, Titouan, Valmir, Alan, Marie, Corinne, Muriel… Même si je n'ai pas pu les connaître très longtemps, c'était un plaisir d'être dans la même équipe.

Table des matières

Introduction générale **13**

Chapitre 1 – Méthodes de calcul en chimie quantique **21**

1.1 L'équation de Schrödinger 22

1.1.1 L'équation de Schrödinger dépendante du temps 22

1.1.2 L'équation de Schrödinger indépendante du temps 22

1.1.3 Du cas monoélectronique au cas polyélectronique 23

1.2 Méthode Hartree-Fock 26

1.2.1 Les équations de Hartree-Fock et opérateurs associés 26

1.2.2 La méthode du champ auto-cohérent 28

1.3 Méthodes post-Hartree-Fock 29

1.4 Fonctions de base 31

1.5 La théorie de la fonctionnelle de la densité 34

1.5.1 Le premier théorème de Hohenberg et Kohn 35

1.5.2 Le second théorème de Hohenberg et Kohn 36

1.5.3 L'approche de Kohn-Sham 37

1.5.4 Les fonctionnelles d'échange-corrélation 39

1.5.5 Les équations de Kohn-Sham 43

Chapitre 2 – Méthodes d'analyse en chimie quantique **49**

2.1 Modélisation des surfaces d'énergie potentielle 50

2.2 Modélisation du solvant (implicite et explicite) 53

2.3 DFT conceptuelle 55

2.3.1 Fondements de la DFT conceptuelle 55

2.3.2 Descripteurs globaux 57

2.3.3 Descripteurs locaux 59

2.3.4 Principes de réactivité 62

2.4 Analyses de population 62

2.4.1 Analyses de population basées sur les orbitales moléculaires 63

2.4.2 Analyses de population basées sur le potentiel électrostatique 66

2.4.3 Analyses de population basées sur la densité électronique 67

2.5 Méthode REG 77

2.5.1 Objectif 77

2.5.2 Principe 78

2.5.3 Applications 83

Chapitre 3 – Étude de la torquosélectivité de l'ouverture thermique de cyclobutènes substitués 89

3.1 État de l'art 90

3.2 Ajouts de nouvelles procédures pour l'analyse REG-IQA et REGdiff 92

3.3 Implémentations pour l'analyse REG-NBO et REG-NBOdiff 103

3.4 Résultats 107

3.4.1 Systèmes et détails de calcul 107

3.4.2 Étude des profils d'énergie 108

3.4.3 Étude REG-NBOdiff de la différence des profils 110

3.4.4 Étude REG-IQAdiff de la différence des profils 120

Chapitre 4 – Étude théorique des additions d'aza-Michael 131

4.1 Introduction 132

4.2 Article 137

4.3 Conclusion 165

Chapitre 5 – Étude théorique des additions d'aza-Michael en solvant explicite 167

5.1 Introduction 167

5.2 Article 168

5.3 Conclusion 212

Conclusion générale 217

Liste des figures

Figure 0.1 - Exemple de réaction d'ouverture thermique de cycle (borylcyclobutène)..... 15

Figure 0.2 - Exemple typique de réaction d'aza-Michael. 16

Figure 2.3 - Représentation schématique d'une surface d'énergie potentielle 50

Figure 2.4 – Graphique des zones de la réaction définies par Toro-Labbé et al. [1]..... 52

Figure 2.5 - Schéma des trois surfaces utilisées dans la méthode PCM..... 54

Figure 2.6 – Représentation d'orbitales moléculaires et localisées du borycyclobutène 65

Figure 2.7 – Représentation du graphe moléculaire et d'une partie des bassins atomiques et des chemins de gradient de l'état de transition du borylcyclobutène lors de l'ouverture thermique du cycle..... 71

Figure 2.8 - Exemple de découpage du profil réactionnel de l'ouverture thermique du borylcyclobutène. 78

Figure 2.9 – Diagrammes en violon représentant la densité des coefficients de détermination des régressions des termes d'une analyse REG sur chaque segment 79

Figure 2.10 - Évolution de l'énergie totale et de deux termes IQA (haut) et régressions linéaires de ces termes par rapport à l'énergie totale (bas)..... 80

Figure 2.11 – Graphique à tiges permettant la visualisation d'une analyse REG-IQA restreinte aux R^2 supérieurs à 0,90, pour la réaction de la méthylamine avec l'acrylate de méthyle en solvant explicite (DMSO)..... 81

Figure 2.12 - Schéma du fonctionnement de la librairie REG.py 83

Figure 3.13 - Mécanisme de l'ouverture thermique de cyclobutènes substitués..... 90

Figure 3.14 - Explication orbitalaire de l'ouverture de cyclobutènes substitués. 91

Figure 3.15 - Normalisation de la coordonnée de réaction des profils de réaction de l'ouverture thermique du borylcyclobutène. 94

Figure 3.16 - Régressions polynômiales d'un terme IQA pour le difluorosilylcyclobutène sans et avec modification de correction..... 95

Figure 3.17 - Régressions polynômiales et différences d'un terme IQA pour le difluorosilylcyclobutène. 96

Figure 3.18 - Régressions rééchantillonnées (rouge et bleu) et différence des régressions (vert) des profils d'énergie de l'ouverture thermique du chlorocyclobutène..... 97

Figure 3.19 - Schéma d'illustration de la procédure de changements de segmentation 99

Figure 3.20 - Tracé de l'analyse REG-IQAdiff de la différence des termes IQA principaux sur le 1er segment de la réaction d'ouverture du borylcyclobutène. 101

Figure 3.21 – Schéma du fonctionnement de la librairie REG.py modifiée (analyse REGdiff) 102

Figure 3.22 – Schéma de Lewis des étapes de l'ouverture du borylcyclobutène du point de vue NBO 103

Figure 3.23 - Schéma du fonctionnement de la librairie REG.py modifiée (analyse NBO et NBOdiff) 106

Figure 3.24 - Représentation "boules-bâtons" des systèmes étudiés. 107

Figure 3.25 - Profils d'énergie de l'ouverture thermique du borylcyclobutène 109

Figure 3.26 - Corrélation entre la barrière d'activation ΔE_{act} et la différence d'énergie entre les réactifs et le premier point d'inflexion du profil d'énergie pour toutes les réactions de tous les systèmes 109

Figure 3.27 - Profils normalisés selon l'équation 3.1 et différence des profils « in » et « out » pour l'ouverture thermique du borylcyclobutène 111

Figure 3.28 - Termes significatifs (R^2>0,90) issus de l'analyse REG-NBOdiff pour le premier segment de la réaction d'ouverture thermique du borylcyclobutène..... 112

Figure 3.29 - Termes significatifs (R^2>0,90) issus de l'analyse REG-NBOdiff pour le second segment de la réaction d'ouverture thermique du borylcyclobutène..... 112

Figure 3.30 - Profils normalisés selon l'équation 3.7 et différence des profils « in » et « out » pour l'ouverture thermique de l'aminocyclobutène 113

Figure 3.31 - Termes significatifs (R^2>0,90) issus de l'analyse REG-NBOdiff pour le second segment de la réaction d'ouverture thermique de l'aminocyclobutène..... 113

Figure 3.32 - Profils normalisés selon l'équation 3.1 et différences des profils « in » et « out » pour l'ouverture thermique du carbonylcyclobutène 115

Figure 3.33 - Termes significatifs (R^2>0,90) issus de l'analyse REG-NBOdiff sur le premier segment de l'ouverture thermique du carbonylcyclobutène 115

Figure 3.34 - Profils normalisés selon l'équation 3.1 et différences des profils « in » et « out » pour l'ouverture thermique du nitrosocyclobutène 116

Figure 3.35 - Termes significatifs (R^2>0,90) issus de l'analyse REG-NBOdiff sur le premier segment de l'ouverture thermique du nitrosocyclobutène 116

Figure 3.36 - Profils normalisés selon l'équation 3.7 et différences des profils « in » et « out » pour l'ouverture thermique du méthylcyclobutène 117

Figure 3.37 - Termes significatifs (R^2>0,90) issus de l'analyse REG-NBOdiff sur le second segment de l'ouverture thermique du méthylcyclobutène 118

Figure 3.38 - Termes significatifs (R^2>0,90) issus de l'analyse REG-NBOdiff pour le second segment de la réaction d'ouverture thermique du nitrocyclobutène (normalisation via l'équation 3.7) 119

Figure 3.39 – Termes significatifs (R^2>0,90) issus de l'analyse REG-IQAdiff pour le premier segment de la réaction d'ouverture thermique du borylcyclobutène 121

Figure 3.40 - Termes significatifs (R^2>0,90) issus de l'analyse REG-IQAdiff pour le second segment de la réaction d'ouverture thermique du borylcyclobutène 121

Figure 3.41 - Termes significatifs (R^2>0,90) issus de l'analyse REG-IQAdiff pour le premier segment de la réaction d'ouverture thermique du borylcyclobutène (G1 = BH_2). Normalisation des segments suivant l'équation 3.1. 122

Figure 3.42 - Termes significatifs (R^2>0,90) issus de l'analyse REG-IQAdiff pour le second segment de la réaction d'ouverture thermique de l'aminocyclobutène (G1 = NH_2). Normalisation des segments suivant l'équation 3.8. 123

Figure 3.43 - Termes significatifs (R^2>0,90) issus de l'analyse REG-IQAdiff pour le premier segment de la réaction d'ouverture thermique du carbonylcyclobutène (G1 = CHO). Normalisation des segments suivant l'équation 3.1. 124

Figure 3.44 - Termes significatifs (R^2>0,90) issus de l'analyse REG-IQAdiff pour le premier segment de la réaction d'ouverture thermique du nitrosocyclobutène (G1 = NO). Normalisation des segments suivant l'équation 3.1. 125

Figure 3.45 - Termes significatifs (R^2>0,90) issus de l'analyse REG-IQAdiff pour le second segment de la réaction d'ouverture thermique du chlorocyclobutène (G1 = Cl). Normalisation des segments suivant l'équation 3.8. 125

Figure 3.46 - Termes significatifs (R^2>0,90) issus de l'analyse REG-IQAdiff pour le second segment de la réaction d'ouverture thermique du nitrocyclobutène (G1 = NO_2). Normalisation des segments suivant l'équation 3.8. 126

Figure 3.47 - Termes significatifs (R^2>0,90) issus de l'analyse REG-IQAdiff pour le second segment de la réaction d'ouverture thermique du méthylcyclobutène (G1 = CH_3). Normalisation des segments suivant l'équation 3.8. 127

Liste des acronymes

ASC – *Apparent Surface Charges* (charges apparentes de surface)

BCP – *Bond Critical Point* (point critique de liaison)

CBS – *Complete Basis Set* (ensemble complet de fonctions de base)

CDFT – *Conceptual Density Functional Theory* (théorie conceptuelle de la fonctionnelle de la densité)

CGTO – *Contracted Gaussian-Type Orbital* (orbitale contractée de type gaussienne)

ChElPG – *Charges from Electrostatic Potentials using a Grid-based method* (charges obtenues à partir du potentiel électrostatique en utilisant une méthode basée sur une grille)

DFT – *Density Functional Theory* (théorie de la fonctionnelle de la densité)

DMB – « $\Delta\mu$ *big is good* »

DMSO – Diméthylsulfoxyde

ESP – *ElectroStatic Potential* (potentiel électrostatique)

FDL – *Finite Difference Linearization* (différences finies)

FMO – *Frozen Molecular Orbitals* (orbitales moléculaires gelées)

GGA – *Generalized Gradient Approximation* (approximation du gradient généralisé)

GTO – *Gaussian-Type Orbital* (orbitale de type gaussienne)

HF – Hartree-Fock

HK – Hohenberg-Kohn

HOMO – *Highest Occupied Molecular Orbital* (orbitale moléculaire la plus haute occupée)

HSAB – *Hard and Soft Acids and Bases* (acides et bases durs et mous)

IEFPCM – *Integral Equation Formalism Polarizable Continuum Model* (modèle du continuum polarisable utilisant le formalisme des équations intégrales)

IAS – *InterAtomic Surface* (surface interatomique)

IQA – *Interacting Quantum Atoms* (atomes quantiques en interaction)

IRC – *Intrinsic Reaction Coordinate* (coordonnée de réaction intrinsèque)

KS – Kohn-Sham

LDA – *Local Density Approximation* (approximation de la densité locale)

LUMO – *Lowest Unoccupied Molecular Orbital* (orbitale moléculaire la plus basse vacante)

MEP – *Molecular Electrostatic Potential* (potentiel électrostatique moléculaire)

MEP – *Minimum Electrophilicity Principle* (principe d'électrophilie minimale)

MHP – *Maximum Hardness Principle* (principe de dureté maximale)

MP – Møller-Plesset

NAO – *Natural Atomic Orbitals* (orbitales naturelles atomiques)

NBO – *Natural Bonding Orbitals* (orbitales naturelles de liaison)

NHO – *Natural Hybrid Orbitals* (orbitales naturelles hybrides)

NLMO – *Natural Localised Molecular Orbitals* (orbitales naturelles moléculaires localisées)

NPA – *Natural Population Analysis* (analyse de population naturelle)

OA – Orbitale Atomique

OM – Orbitale Moléculaire

ONIOM – *Our own n-layered Integrated molecular Orbital and Molecular mechanics* (notre propre méthode à n niveaux intégrant les mécaniques moléculaires et les orbitales moléculaires)

PES – *Potential Energy Surface* (surface d'énergie potentielle)

QM/MM – *Quantum Mechanics / Molecular Mechanics* (mécanique quantique / mécanique moléculaire)

QTAIM – *Quantum Theory of Atoms In Molecules* (théorie quantique des atomes dans les molécules)

REG – *Relative Energy Gradient* (gradient d'énergie relative)

SCF – *Self-Consistent Field* (champ auto-cohérent)

SCRF – *Self-Consistent Reaction Field* (champ de réaction auto-cohérent)

STO – *Slater-Type Orbital* (orbitale de type slatérienne)

Introduction générale

Depuis la nuit des temps, les hommes ont tenté de mieux comprendre le monde qui les entourait. Les réactions chimiques les ont accompagnés depuis la découverte du feu, mais ce n'est qu'au siècle des Lumières que la chimie entre dans son ère moderne, avec notamment les découvertes de Lavoisier. Les progrès accomplis depuis lors ont été faramineux, et la chimie est de nos jours omniprésente au quotidien. Au-delà des avancées des synthèses et des analyses, les théories les plus récentes permettent désormais de calculer et d'étudier les propriétés des composés chimiques avec une précision remarquable. L'équation de Schrödinger proposée en 1926 [1], la méthode de calcul d'Hartree-Fock en 1930 [2,3] et la théorie de la fonctionnelle de la densité en 1964 [4,5] ont été autant de bonds en avant dans la compréhension des systèmes chimiques. L'avènement de la théorie conceptuelle de la fonctionnelle de la densité dans les années 1980 [6,7] a également permis d'établir des ponts entre la chimie quantique et des notions empiriques chères aux chimistes telles que l'électronégativité et la dureté.

Cependant, comme l'a dit le chimiste théoricien Parr [8], « *accurate calculation is not synonymous with useful interpretation. To calculate a molecule is not to understand it.* ». Sans outils d'interprétation, les données obtenues par calcul – quand bien même elles seraient extrêmement précises – ne peuvent pas être pleinement exploitées.

Qui plus est, si l'on souhaite étudier des systèmes de plus en plus complexes et étendus, et donc plus intéressants d'un point de vue chimique, la quantité d'informations fournies par les calculs va également croissant. Pour s'en convaincre, il suffit de voir que le nombre d'interactions possibles entre atomes augmente avec le carré du nombre d'atomes, et même davantage si ces interactions sont décomposées en plusieurs termes.

Il est alors bien sûr possible de supposer quels descripteurs seront pertinents grâce à l'intuition chimique. Cependant, cela reste une sélection réduite et il est facile de passer à côté d'effets *a priori* négligeables ou moins importants mais qui s'avèrent en fait avoir un rôle de premier plan. Le développement d'une

méthode permettant donc de trier et sélectionner automatiquement les descripteurs les plus pertinents sans sélection arbitraire ni recours à l'intuition est donc particulièrement crucial.

Dans ce but, l'analyse du gradient d'énergie relative, ou « REG » (*Relative Energy Gradient*) [9], développée à partir de 2017 dans le groupe du Pr Popelier à Manchester, est idéale pour contourner ces obstacles. Appliquée à un profil d'énergie pour une transformation quelconque, cette méthode s'appuie sur des régressions linéaires entre l'énergie totale du système et les descripteurs étudiés afin de les classer selon l'importance et la similarité de leur variation par rapport à l'énergie totale.

La méthode a été appliquée dans plus d'une dizaine de travaux au cours des dernières années sur des sujets aussi variés que le polymorphisme de l'acide succinique [10], les réactions S_N2 [11], la caractérisation de liaisons halogène [12] et l'énergie stérique [13].

Toutefois, les applications en réactivité sont restées restreintes à l'étude de réactions simples. En effet, la méthode ne peut à l'origine traiter qu'une seule étape élémentaire, avec un seul type d'analyse, la méthode de décomposition énergétique des atomes quantiques en interaction dite « IQA » (*Interacting Quantum Atoms*). De plus, elle ne peut pas tenir compte des effets d'environnement comme les effets de solvant.

L'objectif de cette thèse est donc de développer la méthode REG pour pouvoir l'appliquer à des réactions complexes d'intérêt en chimie de synthèse, notamment en étudiant des réactions concurrentes du même système et en s'intéressant aux effets de solvant. Ce sont des problématiques importantes pour la synthèse, où il faut savoir s'il y a possibilité de compétition, pourquoi une réaction est favorisée par rapport à une autre, et comment le milieu réactionnel peut favoriser ou au contraire interférer avec la réaction. Il suffit de voir l'effet du solvant sur des réactions archétypiques comme les S_N1, S_N2, E1 et E2 où le solvant seul peut déjà orienter la réaction vers un mécanisme plutôt qu'un autre [14].

Attardons-nous désormais sur deux exemples de réactions qui permettront d'aborder ces deux problématiques : la réaction électrocyclique d'ouverture thermique de cycles insaturés et la réaction d'aza-Michael.

L'ouverture thermique de cycles insaturés est un exemple typique de compétition de réactions. En effet, si l'on considère le cas le plus simple, un cyclobutène, l'ouverture du cycle par réaction électrocyclique provoque la rupture de la liaison simple à l'opposé de la double liaison, la transformation de cette double liaison en liaison simple et des deux liaisons simples restantes en doubles liaisons. Cela induit un changement d'hybridation des deux atomes de carbone entre lesquels la liaison se rompt. Si ces atomes portent des substituants, le changement d'hybridation provoque une rotation de ces substituants. Pour l'ouverture thermique d'un cyclobutène, soit une réaction à 2 électrons π, cette rotation des substituants se fait de façon conrotatoire. Il y a deux possibilités de rotation en compétition, soit vers l'intérieur du cycle (menant à une configuration Z) soit vers l'extérieur du cycle (menant à une configuration E), comme illustré sur la Figure 0.1 ci-dessous.

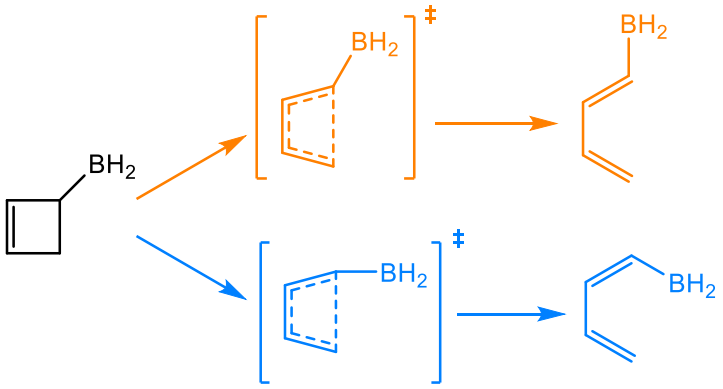


Figure 0.1 - Exemple de réaction d'ouverture thermique de cycle (borylcyclobutène)

Les deux états de transition et produits possibles sont représentés : la rotation vers l'extérieur du cycle en orange et la rotation vers l'intérieur du cycle en bleu.

Expérimentalement, une nette préférence pour l'un ou l'autre mode de rotation a pu être observée pour certains substituants, parfois pour des raisons stériques (le substituant encombrant se retrouve alors plutôt à l'extérieur du cycle), parfois pour des raisons orbitalaires. Ainsi, les substituants présentant une lacune électronique (BH_2…) tourneront préférentiellement vers l'intérieur du cycle grâce à l'interaction stabilisante de la lacune avec l'orbitale liante de la liaison qui se rompt, tandis que l'inverse est constaté pour les substituants présentant une paire libre (NH_2…) à cause de la répulsion entre orbitales occupées. Cette préférence pour l'un ou l'autre mode de rotation a été appelée torquosélectivité par Houk [15].

Cette étude est donc idéale pour étudier des compétitions entre deux réactions : il n'y a que deux chemins possibles, la réaction n'implique qu'une molécule avec un seul réactif et un produit par chemin, et les systèmes étudiés peuvent être de très petite taille.

L'objectif est donc de trouver quels facteurs physico-chimiques permettent d'expliquer les différences entre les deux chemins et qui font que l'un ou l'autre est favorisé selon le substituant.

Quant à la réaction d'aza-Michael, il s'agit d'une variante de l'addition de Michael initialement découverte par le chercheur du même nom en 1887 [16] ; cette réaction implique de façon générale une attaque nucléophile en 1,4 sur une double liaison carbone-carbone conjuguée à un groupement électroattracteur (EA). La réaction d'origine implique l'attaque d'un atome de carbone du nucléophile, mais la réaction peut également impliquer un hétéroatome, menant aux variantes telles que les réactions d'oxa-Michael [17] (avec un atome d'oxygène), phospha-Michael [18] (avec un atome de phosphore), thia-Michael [19] (avec un atome de soufre), et en particulier la réaction d'aza-Michael [20] lorsqu'un atome d'azote du nucléophile, souvent une amine, attaque l'électrophile (voir la Figure 0.2 ci-dessous).

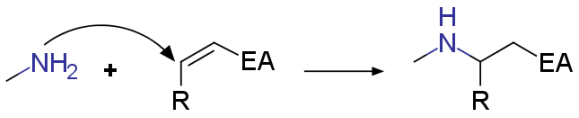


Figure 0.2 - Exemple typique de réaction d'aza-Michael.
EA signifie ici « groupement électroattracteur ».

Il s'agit d'une famille de réactions particulièrement usitées en chimie organique du fait de leur simplicité (la réaction ne nécessite pas de conditions extrêmes ni de catalyseurs onéreux) et de leur efficacité (économie d'atomes totale). La réaction peut être réalisée de façon asymétrique [21] grâce à des catalyseurs de transfert de phase ou des catalyseurs organiques, ainsi que sans solvant [22] (le nucléophile remplissant alors ce rôle). La réaction est favorisée dans les solvants polaires protiques du fait de la formation de liaisons hydrogène entre le solvant, le nucléophile et le groupement électroattracteur de l'électrophile, augmentant respectivement la nucléophilie et l'électrophilie des

réactifs. En revanche, dans des solvants aprotiques ou apolaires, la réaction est plus lente et difficile. Ainsi, il s'agit d'une réaction de choix pour l'étude des effets du solvant sur un mécanisme réactionnel.

C'est également une réaction intéressante du point de vue de l'étude de l'électrophilie, une problématique chère à notre groupe : la prédiction de l'électrophilie de Mayr [23,24], une grandeur expérimentale issue d'études cinétiques et permettant de déterminer *a priori* si une réaction électrophile/nucléophile est possible. La prédiction de cette électrophilie grâce à des descripteurs de chimie quantique reste à l'heure actuelle un vrai défi, et permettrait de jeter un pont supplémentaire entre expérience et théorie. Une méthode robuste et efficace de prédiction *in silico* de cette grandeur aiderait grandement les chimistes de synthèse.

L'organisation de la thèse est donc comme suit :

- Les deux premiers chapitres de cette thèse présenteront les principes sous-jacents des méthodes de calcul et d'analyse de chimie quantique utilisées dans ces travaux.
    - Le premier chapitre se concentre sur les méthodes calculatoires de chimie quantique, en allant de l'équation de Schrödinger à la DFT en passant par la méthode de Hartree-Fock.
    - Le second chapitre aborde quant à lui les méthodes interprétatives, dont la théorie des orbitales naturelles de liaison, ou « NBO » (*Natural Bonding Orbitals*), la théorie quantique des atomes dans les molécules, ou « AIM » (*Atoms In Molecules*), la décomposition énergétique IQA et l'analyse REG toutes deux évoquées précédemment.
- Le troisième chapitre présente l'étude de la torquosélectivité de l'ouverture de cyclobutènes substitués, ainsi que les implémentations apportées à l'analyse REG dans le cadre de cette étude. En effet, puisque celle-ci n'est à l'origine applicable que pour l'analyse IQA d'une seule réaction, de nouvelles routines pour l'analyse NBO ainsi que pour comparer deux réactions concurrentes du même système ont été implémentées.
- Le quatrième et le cinquième chapitre consistent en deux articles sur la réaction d'aza-Michael et l'électrophilie de Mayr.

- Le premier de ces articles est une étude en solvant implicite du mécanisme complet, de la cinétique, des transferts de charge et des variations de divers descripteurs de DFT conceptuelle. Il s'attache en particulier à relier les constantes de vitesse, les charges et les descripteurs à l'électrophilie de Mayr ainsi qu'à vérifier quels principes de réactivité sont applicables pour les réactions des systèmes étudiés.
- Le second article est une étude d'une partie des systèmes du premier article en solvant explicite. L'article se focalise sur l'étape d'addition et vise à expliquer l'influence du solvant sur la réaction. Un lien entre un terme énergétique et l'électrophilie de Mayr est présenté, ce qui en fait le premier lien entre analyse REG et résultats expérimentaux.

Enfin, la thèse se termine par une conclusion générale sur l'ensemble des travaux effectués ainsi que des perspectives pour poursuivre les axes de recherche développés dans le manuscrit.

Bibliographie de l'introduction

1. E. Schrödinger. *Ann. Phys.* **384**, 361–376 (1926).
2. D. R. Hartree. *Math. Proc. Cambridge Philos. Soc.* **24**, 111–132 (1928).
3. V. Fock. *Zeitschrift für Phys.* **61**, 126–148 (1930).
4. P. Hohenberg & W. Kohn. *Phys. Rev.* **136**, B864–B871 (1964).
5. W. Kohn & L. J. Sham. *Phys. Rev.* **140**, A1133–A1138 (1965).
6. R. G. Parr, R. A. Donnelly, M. Levy & W. E. Palke. *J. Chem. Phys.* **68**, 3801–3807 (1978).
7. R. G. Parr & R. G. Pearson. *J. Am. Chem. Soc.* **105**, 7512–7516 (1983).
8. W. Kohn, M. Levy, J. P. Perdew, E. H. Lieb, E. K. U. Gross, R. M. Dreizler, R. G. Parr, A. K. Ragajopal, H. Stoll, A. Savin, C. O. Almbladh, U. von Barth, N. D. Lang, R. K. Bhaduri, M. Brack, G. Holzwarth, J. A. Maruhn, P. Schuck, J. Treiner, *et al.* in *Proceedings of a NATO ASI on density functional methods in physics, held September 5-16, 1983, in Alcabideche, Portugal* (eds. Dreizler, R. M. & Da Providência, J.) vol. 123 (Plenum Press, 1983).
9. J. C. R. Thacker & P. L. A. Popelier. *Theor. Chem. Acc.* **136**, 86 (2017).
10. A. S. De la Vega, L. J. Duarte, A. F. Silva, J. M. Skelton, T. Rocha-Rinza & P. L. A. Popelier. *Phys. Chem. Chem. Phys.* **24**, 11278–11294 (2022).
11. I. Alkorta, J. C. R. Thacker & P. L. A. Popelier. *J. Comput. Chem.* **39**, 546–556 (2018).
12. N. Orangi, K. Eskandari, J. C. R. Thacker & P. L. A. Popelier. *ChemPhysChem* **20**, 1922–1930 (2019).
13. B. C. B. Symons, D. J. Williamson, C. M. Brooks, A. L. Wilson & P. L. A. Popelier. *ChemistryOpen* **8**, 560-570 (2019).
14. N. E. Schore & P. C. Vollhardt. *Traité de chimie organique* (De Boeck, 1981).
15. W. Kirmse, N. G. Rondan & K. N. Houk. *J. Am. Chem. Soc.* **106**, 7989–7991 (1984).
16. A. Michael. *J. für Prakt. Chemie* **35**, 349–356 (1887).
17. C. F. Nising & S. Brase. *Chem. Soc. Rev.* **41**, 988–999 (2012).
18. D. Enders, A. Saint-Dizier, M. Lannou & A. Lenzen. *European J. Org. Chem.* **2006**, 29–49 (2006).
19. P. Wadhwa, A. Kharbanda & A. Sharma. *Asian J. Org. Chem.* **7**, 634–661 (2018).
20. A. Y. Rulev. *Russ. Chem. Rev.* **80**, 197–218 (2011).
21. P. R. Krishna, A. Sreeshailam & R. Srinivas. *Tetrahedron* **65**, 9657–6972 (2009).
22. M. Bláha, O. Trhlíková, J. Podešva, S. Abbrent, M. Steinhart, J. Dybal & M. Dušková-Smrčková. *Tetrahedron* **74**, 58–67 (2018).
23. H. Mayr & M. Patz. *Angew. Chemie Int. Ed. English* **33**, 938–957 (1994).
24. Mayr's Database Of Reactivity Parameters: https://www.cup.lmu.de/oc/mayr/reaktionsdatenbank/fe/.

Chapitre 1 – Méthodes de calcul en chimie quantique_Toc105148538

1.1 L'équation de Schrödinger 22

1.1.1 L'équation de Schrödinger dépendante du temps 22

1.1.2 L'équation de Schrödinger indépendante du temps 22

1.1.3 Du cas monoélectronique au cas polyélectronique 23

1.2 Méthode Hartree-Fock 26

1.2.1 Les équations Hartree-Fock et opérateurs associés 26

1.2.2 La méthode du champ auto-cohérent 28

1.3 Méthodes post-Hartree-Fock 29

1.4 Fonctions de base 31

1.5 La théorie de la fonctionnelle de la densité 34

1.5.1 Le premier théorème de Hohenberg et Kohn 35

1.5.2 Le second théorème de Hohenberg et Kohn 36

1.5.3 L'approche de Kohn-Sham 37

1.5.4 Les fonctionnelles d'échange-corrélation 39

1.5.5 Les équations de Kohn-Sham 43

1.1 L'équation de Schrödinger

1.1.1 L'équation de Schrödinger dépendante du temps

En 1923, De Broglie énonce l'hypothèse de la dualité onde-corpuscule [1] : toutes les particules se comportent également comme des ondes. À sa suite, Schrödinger propose en 1926 une équation non relativiste permettant de trouver l'évolution dans l'espace ($\vec{r}$) et le temps (t) de la fonction d'onde Ψ d'un système [2] :

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

où $\hbar$ est la constante de Planck réduite, $\Psi(\vec{r},t)$ la fonction d'onde comprenant toutes les propriétés quantiques du système, et $\hat{H}$ l'opérateur Hamiltonien. Il est possible de trouver une solution analytique à cette équation mais seulement dans des cas très simples. Cependant, le sens physique de cette fonction d'onde n'est pas alors encore connu.

Born et Jordan postulent en 1927 [3] que le carré du module de la fonction d'onde $|\Psi|^2$ est relié à la densité de probabilité de présence dP d'une particule dans un volume infinitésimal (d^3r) :

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

Ce postulat implique la normalisation suivante de $\Psi(\vec{r},t)$:

$$\iiint |\Psi(\vec{r},t)|^2 d^3r = 1 \tag{1.3}$$

1.1.2 L'équation de Schrödinger indépendante du temps

La résolution de l'équation de Schrödinger permet d'obtenir la fonction d'onde du système ; cependant, celle-ci est trop complexe dans la grande majorité des cas et nécessite de passer par plusieurs approximations.

On peut d'abord se placer dans le cas où l'Hamiltonien ne dépend pas du temps. On peut alors écrire la fonction d'onde en séparant les variables :

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

où A décrit l'évolution de la fonction d'onde dans le temps et ψ représente sa composante spatiale. Il existe plusieurs fonctions ψ_n correspondant aux différents états stationnaires du système.

On peut ainsi réécrire l'équation de Schrödinger en utilisant l'expression (1.4) :

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

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

Si l'on divise l'équation (1.6) par $A(t)\psi(\vec{r})$, on obtient une fonction dépendant uniquement du temps égale à une fonction dépendant seulement de l'espace, ce qui n'est possible que si ces fonctions sont toutes deux constantes :

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

Cette constante (ε) est l'énergie totale du système. Avec les deux derniers membres de l'équation (1.7), on peut donc écrire l'équation aux valeurs propres suivante, que doit satisfaire chaque état stationnaire ψ_n d'énergie E_n :

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

Ce qui peut également s'écrire avec la notation bra-ket de Dirac :

$$\hat{H}|\psi_n\rangle = E_n|\psi_n\rangle \tag{1.9}$$

Il s'agit de l'équation de Schrödinger indépendante du temps.

1.1.3 Du cas monoélectronique au cas polyélectronique

L'équation (1.9) obtenue précédemment permet d'étudier des systèmes monoélectroniques mais également des systèmes poly-électroniques comme les molécules. La fonction d'onde moléculaire contient alors l'intégralité des propriétés quantiques du système.

L'opérateur Hamiltonien $\hat{H}$ permettant de trouver les énergies E_n dans l'équation de Schrödinger s'écrit habituellement comme une somme de deux opérateurs, l'opérateur d'énergie cinétique $\hat{T}$ et l'opérateur d'énergie potentielle $\hat{V}$:

$$\hat{H} = \hat{T} + \hat{V} \tag{1.10}$$

Pour un système à N électrons et M noyaux, on peut réécrire l'Hamiltonien en prenant en compte les différentes interactions entre particules, avec des opérateurs relatifs aux électrons (notés e) et aux noyaux (notés n) :

$$\hat{H} = \hat{T}_n + \hat{T}_e + \hat{V}_{nn} + \hat{V}_{en} + \hat{V}_{ee} \tag{1.11}$$

où $\hat{T}_n$ et $\hat{T}_e$ sont respectivement les opérateurs d'énergie cinétique nucléaire et électronique, et $\hat{V}_{nn}$, $\hat{V}_{en}$ et $\hat{V}_{ee}$ sont respectivement les opérateurs d'énergie potentielle des interactions noyau-noyau, électron-noyau et électron-électron.

En unités atomiques (où $\hbar = m_e = \frac{e^2}{4\pi\varepsilon_0} = 1$), l'opérateur d'énergie cinétique nucléaire $\hat{T}_n$ peut s'écrire :

$$\hat{T}_n = -\sum_{a=1}^{M} \frac{1}{2m_a} \nabla_a^2 \tag{1.12}$$

avec une somme courant sur chaque noyau a de masse m_a.

Similairement, l'opérateur d'énergie cinétique électronique $\hat{T}_e$ s'écrit en unités atomiques :

$$\hat{T}_e = -\sum_{i=1}^{N} \frac{1}{2} \nabla_i^2 \tag{1.13}$$

Quant à lui, l'opérateur d'énergie potentielle de répulsion nucléaire $\hat{V}_{nn}$ entre des noyaux a et b s'écrit :

$$\hat{V}_{nn} = \sum_{a=1}^{M} \sum_{b>a}^{M} \frac{Z_a Z_b}{r_{ab}} \tag{1.14}$$

avec Z_a et Z_b les numéros atomiques des noyaux a et b ; r_{ab} correspond à la distance relative entre les noyaux a et b ($r_{ab} = |\vec{r_a} - \vec{r_b}|$).

Enfin, l'opérateur d'énergie potentielle d'attraction $\hat{V}_{ne}$ entre un noyau a et un électron i s'écrit :

$$\hat{V}_{ne} = -\sum_{a=1}^{M}\sum_{i=1}^{N}\frac{Z_a}{r_{ai}} \tag{1.15}$$

Comme précédemment, r_{ai} est la distance entre le noyau a et l'électron i. Enfin, l'opérateur d'énergie potentielle de répulsion électronique $\hat{V}_{ee}$ entre les électrons i et j s'écrit :

$$\hat{V}_{ee} = \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} \tag{1.16}$$

Encore une fois, r_{ij} est la distance entre les électrons i et j. Il faut noter qu'aucune de ces expressions d'opérateurs ne tient compte des effets relativistes.

En 1927, Born et Oppenheimer proposent une simplification de l'équation de Schrödinger appliquée aux molécules [4] en se basant sur l'importante différence de masse entre noyaux et électrons (en effet, le rapport des masses est déjà de l'ordre de 10^3 pour les noyaux les plus légers). On peut donc considérer que la vitesse de déplacement des noyaux est négligeable par rapport à celle des électrons et que les noyaux sont donc fixes le temps du réarrangement électronique.

Dans le cadre de l'approximation de Born-Oppenheimer, on peut découpler la position des noyaux et celle des électrons. La fonction d'onde totale du système dépendant de ces deux ensembles de paramètres $\psi(\{\vec{r_a}\};\{\vec{r_i}\})$ peut donc s'écrire comme un produit d'une fonction d'onde nucléaire décrivant uniquement les noyaux $\psi_{nuc}(\{\vec{r_a}\})$ et d'une fonction d'onde électronique décrivant uniquement les électrons $\psi_{el}(\{\vec{r_a}\};\{\vec{r_i}\})$:

$$\psi(\{\vec{r_a}\};\{\vec{r_i}\}) = \psi_{nuc}(\{\vec{r_a}\}) \times \psi_{el}(\{\vec{r_a}\};\{\vec{r_i}\}) \tag{1.17}$$

Si l'on considère une géométrie fixe, la position des noyaux devient un ensemble de paramètres constants. L'équation de Schrödinger (1.8) se scinde alors en deux équations, l'une électronique et l'autre décrivant les rotations et les vibrations du système (incluant donc les termes nucléaires). On ne résout plus l'équation électronique avec l'Hamiltonien total du système ($\hat{H}$) mais avec un Hamiltonien électronique ($\hat{H}_{el}$). Il s'ensuit que l'équation (1.12) se simplifie en :

$$\widehat{H}_{el} = \widehat{T}_e + \widehat{V}_{en} + \widehat{V}_{ee} \tag{1.18}$$

Cependant, le terme $\widehat{V}_{ee}$ rend en pratique la résolution impossible dès que le nombre d'électrons augmente, nécessitant des approximations supplémentaires.

L'équation de Schrödinger électronique définie par (1.8) se réécrit également :

$$\widehat{H}_n^{el}\psi_n^{el}(\vec{r}) = E_n^{el}\psi_n^{el}(\vec{r}) \tag{1.19}$$

où E_n^{el} est l'énergie électronique du système. L'énergie la plus basse de l'ensemble des E_n^{el} est notée E_0^{el} et correspond à l'état électronique fondamental du système. Par la suite, toutes les énergies et fonctions d'onde seront électroniques (sauf mention contraire).

L'équation (1.19) peut être résolue à l'aide du principe variationnel. Ainsi, pour calculer l'énergie de l'état fondamental, une fonction d'onde d'essai paramétrée ($\widetilde{\psi}$) est utilisée. On peut montrer que l'énergie $\tilde{E}$ obtenue via ces fonctions d'onde sera toujours supérieure ou égale à l'énergie exacte de l'état fondamental :

$$\tilde{E} = \frac{\langle\widetilde{\psi}|\widehat{H}|\widetilde{\psi}\rangle}{\langle\widetilde{\psi}|\widetilde{\psi}\rangle} \geq E_0 \tag{1.20}$$

En optimisant les paramètres de $\widetilde{\psi}$ de sorte à minimiser l'énergie, il est possible de tendre vers la fonction d'onde exacte.

1.2 Méthode Hartree-Fock

1.2.1 Les équations de Hartree-Fock et opérateurs associés

La méthode de Hartree-Fock (HF) [5,6] est une méthode de résolution approchée de l'équation de Schrödinger. Elle se place dans le cadre de l'approximation de Born-Oppenheimer. Plutôt que de considérer explicitement l'influence de tous les électrons les uns sur les autres, les électrons sont vus comme s'ils étaient en train de se déplacer dans un potentiel moyen engendré par les autres électrons. C'est l'approximation dite du champ moyen.

À cette fin, on cherche à exprimer la fonction d'onde polyélectronique du système à partir de fonctions d'onde monoélectroniques dites spinorbitales, qui sont chacune le produit d'une fonction spatiale (orbitale) et d'une fonction de spin. Les électrons peuvent en effet avoir uniquement deux valeurs de spin, +½ et –½. Par convention, les électrons de spin +½ sont notés α et les électrons de spin –½ sont notés β.

Pour respecter le principe d'exclusion de Pauli [7], la fonction d'onde polyélectronique $\psi_{el}(\vec{x}_1,\vec{x}_2,\dots\vec{x}_n)$ dépendant des $\vec{x}_i$ (prenant en compte la position $\vec{r}_i$ et le spin σ_i (α ou β) de chaque électron : $\vec{x}_i = (\vec{r}_i, \sigma_i)$) doit être antisymétrique par permutation de deux électrons quelconques :

$$\psi_{el}(\vec{x}_1,\dots\vec{x}_i,\dots\vec{x}_j,\dots\vec{x}_N) = -\psi_{el}(\vec{x}_1,\dots\vec{x}_j,\dots\vec{x}_i,\dots\vec{x}_N) \tag{1.21}$$

Une expression de la fonction d'onde respectant cette condition est celle du déterminant de Slater :

$$\psi_{el}(\vec{x}_1,\vec{x}_2,\dots\vec{x}_N) = \frac{1}{\sqrt{N!}}\begin{vmatrix} \Phi_1(\vec{x}_1) & \Phi_2(\vec{x}_1) & \cdots & \Phi_N(\vec{x}_1) \\ \Phi_1(\vec{x}_2) & \Phi_2(\vec{x}_2) & \cdots & \Phi_N(\vec{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_1(\vec{x}_N) & \Phi_2(\vec{x}_N) & \cdots & \Phi_N(\vec{x}_N) \end{vmatrix} \tag{1.22}$$

En utilisant ce formalisme déterminantal, on remarque qu'il est impossible à deux électrons d'occuper la même spinorbitale, car dans ce cas, deux des colonnes du déterminant seraient identiques et le déterminant serait nul. C'est le principe d'exclusion de Pauli.

L'application du principe variationnel à ce déterminant de Slater permet de montrer que l'ensemble des spinorbitales est solution des équations de Hartree-Fock monoélectroniques suivantes :

$$\hat{F}_i(\vec{x}_i)\Phi_i(\vec{x}_i) = \varepsilon_i\Phi_i(\vec{x}_i) \tag{1.23}$$

où $\hat{F}_i$ est l'opérateur de Fock. Il s'agit de la somme d'un opérateur Hamiltonien monoélectronique $\hat{h}_i(\vec{x}_i)$ et d'un terme prenant partiellement en compte la répulsion électronique :

$$\hat{F}_i(\vec{x}_i) = \hat{h}_i(\vec{x}_i) + \sum_{j=1}^{N}\left(\hat{J}_j(\vec{x}_i) - \hat{K}_j(\vec{x}_i)\right) \tag{1.24}$$

Dans cette équation, $\hat{J}_j(\vec{x}_i)$ est l'opérateur Coulombien prenant en compte l'interaction entre l'électron i de la spinorbitale $\Phi_i(\vec{x}_i)$ et l'électron j de la spinorbitale $\Phi_j(\vec{x}_j)$:

$$\hat{J}_j|\Phi_i\rangle = \left\langle \Phi_j \middle| \frac{1}{\hat{r}_{ij}} \middle| \Phi_j \right\rangle |\Phi_i\rangle \tag{1.25}$$

Quant à lui, $\widehat{K}_j(\vec{x}_i)$ est l'opérateur d'échange qui décrit l'effet quantique qui se produit lors de l'échange de deux électrons, provenant du caractère antisymétrique de la fonction d'onde polyélectronique :

$$\widehat{K}_j|\Phi_i\rangle = \left\langle \Phi_j \middle| \frac{1}{\hat{r}_{ij}} \middle| \Phi_i \right\rangle |\Phi_j\rangle \tag{1.26}$$

On peut donc exprimer l'énergie électronique (ou Hartree-Fock) E_{HF} du système comme :

$$E_{HF} = \sum_{i=1}^{N} \langle \Phi_i | \hat{h}_i | \Phi_i \rangle + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\langle \Phi_j | \hat{J}_i | \Phi_j \rangle - \langle \Phi_j | \widehat{K}_i | \Phi_j \rangle \right) \tag{1.27}$$

1.2.2 La méthode du champ auto-cohérent

La résolution des équations de Hartree-Fock repose sur la méthode dite du « champ auto-cohérent » [8] (ou SCF pour « *self-consistent field* ») : les opérateurs de Fock et les orbitales (qui sont les solutions des équations de Hartree-Fock) sont optimisés simultanément et se corrigent d'eux-mêmes, d'où le nom d'auto-cohérent. Il faut donc proposer des spinorbitales et des opérateurs d'essai au début de chaque calcul, puis optimiser itérativement les orbitales et les opérateurs jusqu'à ce que les orbitales correspondent au critère de convergence fixé au début du calcul.

Les spinorbitales $\Phi_i(\vec{x}_i)$ sont souvent exprimées sous forme d'une combinaison linéaire de N_b fonctions de base (abordées plus en détail dans la section 1.4), comme proposé par Roothaan en 1951 [9,10] :

$$\Phi_i(\vec{x}_i) = \sum_{\mu=1}^{N_b} c_{\mu i} \chi_\mu(\vec{x}) \tag{1.28}$$

L'optimisation des spinorbitales porte sur l'optimisation des coefficients $c_{\mu i}$ en utilisant le principe variationnel. Les équations de Hartree-Fock prennent alors la forme suivante :

$$\sum_{\nu} c_{\nu i} \langle \chi_\mu | \hat{F}_i | \chi_\nu \rangle = \varepsilon_i \sum_{\nu} c_{\nu i} \langle \chi_\mu | \chi_\nu \rangle \tag{1.29}$$

Ce sont les équations de Roothaan et Hall, que l'on peut également écrire sous forme matricielle :

$$FC = SC\varepsilon \tag{1.30}$$

où F est la matrice de Fock (le premier terme dans (1.29)), dont les éléments $F_{\mu\nu}$ sont définis par :

$$F_{\mu\nu} = \langle \chi_\mu | \hat{F}_i | \chi_\nu \rangle \tag{1.31}$$

et S est la matrice de recouvrement entre les orbitales (le second terme de l'équation 1.29), dont les éléments $S_{\mu\nu}$ sont définis par :

$$S_{\mu\nu} = \langle \chi_\mu | \chi_\nu \rangle \tag{1.32}$$

Enfin, C est une matrice carrée de coefficients $c_{\mu\nu}$ et ε une matrice diagonale contenant les énergies orbitalaires ε_i.

La méthode SCF peut ainsi se résumer en quatre étapes :

1. Obtenir un premier jeu de coefficients, C à partir du système (avec les masses des atomes, les coordonnées des noyaux, le nombre d'électrons, les fonctions de base) ;
2. Générer les matrices de Fock F et de recouvrement S à partir de C ;
3. Résoudre les équations de Roothaan-Hall en diagonalisant F pour obtenir un nouveau jeu de coefficients C' ;
4. Vérifier la convergence du calcul (a-t-on $C' \approx C$?) ;
    a. Si oui, on peut arrêter le calcul SCF et déterminer les propriétés du système à partir de F et C' ;
    b. Sinon, il faut retourner à l'étape 2 en utilisant C'.

1.3 Méthodes post-Hartree-Fock

La méthode Hartree-Fock permet d'obtenir une valeur approximative de l'énergie exacte. Elle fournit une base sur laquelle s'appuient diverses méthodes, dites post-Hartree-Fock. Celles-ci visent à améliorer la précision des résultats obtenus, notamment en prenant en compte la corrélation électronique, l'effet quantique absent de la méthode Hartree-Fock.

En effet, la méthode Hartree-Fock ne permet pas de trouver l'énergie exacte E_{exacte} mais seulement une valeur plus élevée E_{HF}. La différence a été définie par Löwdin [11] comme étant l'énergie de corrélation électronique E_c :

$$E_c = E_{exacte} - E_{HF} \tag{1.33}$$

Il existe plusieurs types de corrélation, comme les corrélations statique, dynamique, à longue portée, à courte portée, etc. Les méthodes post-HF peuvent se focaliser sur un type de corrélation plutôt qu'un autre. On peut notamment citer la théorie perturbationnelle de Møller-Plesset [12] (notée MP) et la théorie du cluster couplé [13] (CC).

La première approche repose sur une perturbation de l'Hamiltonien du système $\hat{H}_0$ du type $\hat{H} = \hat{H}_0 + \lambda\hat{V}$, avec λ un paramètre contrôlant la force de la perturbation. L'injection de cet Hamiltonien perturbé $\hat{H}$ dans l'équation de Schrödinger aboutit à une expression de l'énergie sous forme d'un polynôme infini en λ :

$$E = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \cdots \tag{1.34}$$

En pratique, on tronque le polynôme à un certain ordre, d'où l'appellation MPx où x est l'ordre maximal des termes correctifs considérés.

Au second ordre, l'application de la théorie des perturbations de Rayleigh-Schrödinger donne l'énergie de correction au second ordre :

$$E_0^{(2)} = \sum_{j>0} \frac{\left\langle \psi_0^{(0)} \middle| \hat{V} \middle| \psi_j^{(0)} \right\rangle}{E_0^{(0)} - E_j^{(0)}} \tag{1.35}$$

L'énergie totale a alors la forme suivante :

$$E_0 = E_0^{(0)} + E_0^{(1)} + E_0^{(2)} \tag{1.36}$$

$$E_0 = E_{HF} + E_c^{MP2} \tag{1.37}$$

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

Pour la méthode du cluster couplé, on utilise une extension exponentielle de la fonction d'onde de la forme $|\psi_i\rangle = e^T|\psi_n\rangle$. Cette exponentielle peut s'exprimer par un développement en série qui contient entre autres l'opérateur cluster $\hat{T}$, lui-même une somme d'opérateurs $\hat{T}_i$ correspondant à des excitations à l'ordre i. La fonction d'onde peut alors s'exprimer sous la forme d'une somme de la fonction d'onde non excitée et de termes d'excitations simples, doubles, triples, etc.

En pratique, on tronque cette somme à un certain ordre, et la méthode est notée CCS (cluster couplé simple) si l'on s'arrête à l'ordre un, CCSD (cluster couplé simple et double) si l'on s'arrête à l'ordre deux, etc.

1.4 Fonctions de base

Les fonctions de base évoquées précédemment dans la section 1.2 sont des ensembles de fonctions permettant de représenter les orbitales. Il existe notamment deux grands types de fonctions de base, les fonctions de type Slater [14] (« *Slater-type orbitals* », STO) et les fonctions de type gaussiennes [15] (« *Gaussian-type orbitals* », GTO).

Historiquement, les premières fonctions de base ayant été développées sont les fonctions de type Slater, qui s'expriment en coordonnées cartésiennes sous la forme suivante :

$$\chi_{a,b,c}^{STO} = Nx^a y^b z^c e^{-\zeta r} \tag{1.38}$$

Dans cette équation, N est une constante permettant de normaliser la fonction, les exposants a, b et c doivent respecter la règle $L = a + b + c$ où L est le moment angulaire, et ζ est l'exposant orbital qui contrôle la taille de l'orbitale. Les STO s'inspirent des solutions de l'équation de Schrödinger appliquée aux atomes hydrogénoïdes (possédant un seul électron).

Cependant, si l'on étudie des systèmes de plus en plus grands, l'utilisation de STO devient rapidement un problème en raison du temps nécessaire à l'évaluation des éléments de matrice $F_{\mu\nu}$. C'est pour cela que Boys [15] a proposé les fonctions de type gaussiennes, qui s'expriment ainsi en coordonnées cartésiennes :

$$\chi_{a,b,c}^{GTO} = N x^a y^b z^c e^{-\zeta r^2} \tag{1.39}$$

L'avantage principal d'utiliser des GTO au lieu de STO est qu'un produit de gaussiennes peut s'écrire sous la forme d'une seule gaussienne (ce qui n'est pas le cas pour les STO), facilitant les intégrations et réduisant le temps de calcul.

Cependant, elles ne reproduisent pas les variations radiales de la densité électronique aussi bien que les orbitales de type Slater : à courte portée (si r tend vers 0), les STO présentent un point de rebroussement tandis que les GTO ont une forme de courbe en cloche, et à longue portée (si r tend vers +∞), les GTO décroissent plus rapidement que les STO. Il est possible de pallier en partie ce problème en utilisant des fonctions de type gaussiennes dites contractées, CGTO, qui sont des sommes de GTO (les « primitives ») avec différentes valeurs du paramètre ζ :

$$\chi_{a,b,c}^{GTO} = N \sum_{i=1}^{n} c_i x^a y^b z^c e^{-\zeta_i r^2} \tag{1.40}$$

On peut ainsi retrouver un comportement qui s'approche plus de celui des STO en jouant sur les coefficients c_i et ζ_i.

Il est possible d'utiliser une base dite « minimale » où une fonction de base (STO, GTO ou CGTO) est utilisée pour décrire chaque orbitale, ou bien une base dite « étendue » où plusieurs fonctions de base sont utilisées pour décrire chaque orbitale. Ainsi, on parle de base « double zêta » si deux fonctions sont utilisées pour chaque orbitale, « triple zêta » avec trois fonctions, etc. On peut également séparer les électrons de cœur et de valence et utiliser des bases dites « à valence séparée » où chaque orbitale de cœur est décrite par une base minimale et chaque orbitale de valence est décrite par une base étendue.

Pour améliorer la précision des résultats, des fonctions supplémentaires peuvent être ajoutées ; on distingue notamment :

- Les fonctions diffuses, qui sont des fonctions à grande portée avec un faible ζ permettant d'ajuster le comportement des orbitales à grande distance du noyau ; cela est important pour décrire les anions et les interactions intermoléculaires.

- Les fonctions de polarisation, qui sont des fonctions de nombre quantique azimutal supérieur : par exemple, pour des atomes d'hydrogène que l'on décrit avec des fonctions de base approximant l'orbitale atomique 1s, on pourrait ajouter des fonctions de base correspondant à des orbitales p ; pour des atomes de carbone, on pourrait ajouter des fonctions de base correspondant à des orbitales d, etc. Cela permet de décrire des effets de polarisation, notamment dans les liaisons chimiques.

Bien évidemment, plus on utilise de fonctions de base, plus les résultats seront (en principe) précis, mais plus les calculs prendront de temps. Il faudrait une infinité de fonctions de base pour décrire parfaitement les systèmes : c'est ce que l'on appelle la limite de l'ensemble complet de fonctions de base (« *Complete Basis Set* » ou CBS).

Les bases les plus couramment utilisées incluent les bases à valence séparée de Pople [16], qui sont notées *X-YZ+G(T)* :

- *X* est le nombre de primitives utilisées pour la CGTO décrivant chaque orbitale de cœur ;
- *Y* et *Z* correspondent au nombre de primitives utilisées pour les différentes CGTO décrivant chaque orbitale de valence. Le tiret indique la séparation de valence. Il peut y avoir plus de deux chiffres si la base est triple-zêta ou plus (exemple : 6-311G) ;
- Le + est optionnel et indique si la base est suppléée de fonctions diffuses : il peut y avoir un ou deux + correspondant respectivement à l'ajout de jeux de fonctions diffuses soit uniquement sur les atomes « lourds » (seconde période et suivants) ou sur tous les atomes ;
- La lettre *G* indique que les fonctions de base sont des gaussiennes (GTO ou CGTO) ;
- *T* est optionnel et indique si la base comporte des fonctions de polarisation, en ajoutant le nombre et le type de fonctions. Par exemple, 6-31G(2d,p) indique que deux fonctions de polarisation d sont ajoutées sur les atomes ayant des orbitales de valence p et qu'une fonction de polarisation p est ajoutée sur les atomes ayant des orbitales de valence s.

Les bases de Pople que nous avons le plus souvent utilisées sont 6-31+G(d) et 6-311++G(2d,2p).

Les bases de Dunning [17] sont également un choix très populaire : elles comportent des jeux de fonctions polarisées calibrées pour obtenir des résultats proches de la limite CBS pour des calculs post-Hartree-Fock. Elles ont la syntaxe générale *d-aug-cc-pVXZ* :

- La lettre *d* (optionnelle) indique si l'on utilise un ou deux jeux de fonctions de polarisation ;
- Le préfixe *aug* (optionnel) indique si la base est suppléée de fonctions diffuses ou non ;
- Les lettres *cc-p* indiquent qu'il s'agit d'une base polarisée à corrélation cohérente (« *correlation-consistent polarised* ») ;
- La lettre *V* indique la séparation de valence ;
- *XZ* indique combien de fonctions sont utilisées pour chaque orbitale de valence : DZ pour double zêta, TZ pour triple, etc.

De façon similaire, les bases d'Ahlrichs [18] ont pour syntaxe *defn-XZVPD* où *n* est un nombre optionnel qui correspond à la première ou à la seconde définition de ces bases et *X* indique si la base est double zêta, triple zêta, etc. Les lettres *P* et *D* indiquent si la base inclut des fonctions de polarisation et/ou diffuses.

Nous avons utilisé la base def2-TZVP dans certains de nos calculs.

1.5 La théorie de la fonctionnelle de la densité

La théorie de la fonctionnelle de la densité (« *Density Functional Theory* », DFT) [19,20] est une alternative aux méthodes post-Hartree-Fock qui repose non pas sur la fonction d'onde mais sur la densité électronique de l'état fondamental du système.

Cette densité électronique $\rho(\vec{r})$ est définie par l'intégrale sur toutes les coordonnées spatiales et de spin de tous les électrons, sauf une coordonnée spatiale, du carré du module de la fonction d'onde :

$$\rho(\vec{r}_1) = N \int \dots \int |\psi_{el}(x_1, x_2, \dots x_N)|^2 d\sigma_1 d^4x_2 \dots d^4x_N \tag{1.41}$$

$\rho(\vec{r})$ correspond au nombre moyen d'électrons dans un volume infinitésimal d^3r centré en $\vec{r}$. Il s'ensuit que l'intégrer sur tout l'espace revient à retrouver le nombre total d'électrons N :

$$\int \rho(\vec{r}_1) d^3 r_1 = N \tag{1.42}$$

La densité électronique est ainsi une grandeur réelle positive dépendant de seulement trois coordonnées spatiales, là où la fonction d'onde du système dépend de $4N$ variables ($3N$ coordonnées spatiales et N coordonnées de spin), ce qui permet de réduire considérablement la complexité des calculs pour de grands systèmes. Il faut également noter que la densité électronique est une observable (on peut notamment la déterminer par diffraction des rayons X [21]), tandis que la fonction d'onde ne l'est pas.

1.5.1 Le premier théorème de Hohenberg et Kohn

Le théorème fondateur de la DFT a été démontré en 1964 par Hohenberg et Kohn [22]. Pour un système donné à N électrons soumis à un potentiel externe v_{ext}, il est possible d'obtenir l'énergie de l'état fondamental et la densité électronique correspondante $\rho(\vec{r})$ via la résolution de l'équation de Schrödinger. À un potentiel externe donné correspond une unique densité électronique :

$$v_{ext}(\vec{r}) \Rightarrow \rho(\vec{r}) \tag{1.43}$$

Le premier théorème de Hohenberg-Kohn montre alors que la relation peut être inversée et qu'à une densité électronique donnée correspond un unique potentiel externe (à une constante près) :

$$\rho(\vec{r}) \Rightarrow v_{ext}(\vec{r}) + C \tag{1.44}$$

Ainsi, la densité électronique permet d'obtenir toutes les propriétés du système. Dans ce cadre, on peut écrire le potentiel externe (en pratique le plus souvent généré par les noyaux) comme une fonctionnelle de la densité électronique, et l'énergie totale du système également.

On peut démontrer ce premier théorème par l'absurde. Si l'on considère deux potentiels externes différents $v_{ext}^{(1)}(\vec{r})$ et $v_{ext}^{(2)}(\vec{r})$ auxquels correspondent deux fonctions d'onde différentes $\psi_0^{(1)}$ et $\psi_0^{(2)}$ et une même densité électronique de l'état fondamental $\rho(\vec{r})$, il y a alors un Hamiltonien associé à chaque potentiel externe : $\hat{H}^{(1)}$ et $\hat{H}^{(2)}$ et il en résulte deux énergies totales différentes E_1 et E_2. En appliquant le principe variationnel (dans l'hypothèse où l'état fondamental n'est pas dégénéré), on a :

$$E_1 = \left\langle \psi_0^{(1)} \middle| \hat{H}^{(1)} \middle| \psi_0^{(1)} \right\rangle < \left\langle \psi_0^{(2)} \middle| \hat{H}^{(1)} \middle| \psi_0^{(2)} \right\rangle = \left\langle \psi_0^{(2)} \middle| \hat{H}^{(2)} + \hat{V}^{(1)} - \hat{V}^{(2)} \middle| \psi_0^{(2)} \right\rangle \tag{1.45}$$

$$E_1 < E_2 + \int \left[v_{ext}^{(1)}(\vec{r}) - v_{ext}^{(2)}(\vec{r}) \right] \rho(\vec{r}) d^3r \tag{1.46}$$

Par symétrie, on obtient :

$$E_2 < E_1 + \int \left[v_{ext}^{(2)}(\vec{r}) - v_{ext}^{(1)}(\vec{r}) \right] \rho(\vec{r}) d^3r \tag{1.47}$$

La combinaison de (1.42) et (1.43) permet d'écrire :

$$E_2 + E_1 < E_1 + E_2 \tag{1.48}$$

ce qui est impossible. Deux potentiels externes différents ne peuvent donc pas avoir la même densité électronique.

Ce résultat a permis à Hohenberg et Kohn d'établir l'existence d'une fonctionnelle universelle $F_{HK}[\rho]$ (valable pour tous les systèmes quel que soit le potentiel externe) en exprimant l'énergie totale de l'état fondamental électronique comme une fonctionnelle de la densité électronique :

$$E[\rho] = F_{HK}[\rho] + \int \rho(\vec{r}) v_{ext}(\vec{r}) d^3r \tag{1.49}$$

1.5.2 Le second théorème de Hohenberg et Kohn

Le second théorème de Hohenberg-Kohn fait à nouveau appel au théorème variationnel. Pour un système soumis à un potentiel externe $v_{ext}(\vec{r})$, le théorème indique que l'énergie totale $E[\rho]$ obéit au principe variationnel : l'énergie de l'état fondamental (E_0) est obtenue en minimisant la fonctionnelle de l'énergie, et la densité électronique associée à cette fonctionnelle sera celle de l'état fondamental. Levy [23] a reformulé les deux théorèmes de la façon suivante par l'intermédiaire d'une recherche contrainte :

$$E_0 = \min_{\psi}\langle\psi|\hat{H}|\psi\rangle = \min_{\rho}\left\{\min_{\psi\to\rho}\langle\psi|\hat{H}|\psi\rangle\right\} \tag{1.50}$$

$$= \min_{\rho}\left\{\min_{\psi\Rightarrow\rho}\{\langle\psi|\hat{T}_e + \hat{V}_{ee} + \hat{v}_{ext}|\psi\rangle\}\right\} \tag{1.51}$$

$$= \min_{\rho}\left\{\min_{\psi\Rightarrow\rho}\{\langle\psi|\hat{T}_e + \hat{V}_{ee}|\psi\rangle\} + \int \rho(\vec{r})v_{ext}(\vec{r})d^3r\right\} \tag{1.52}$$

$$= \min_{\rho}\left\{F_{HK}[\rho] + \int \rho(\vec{r})v_{ext}(\vec{r})d^3r\right\} \tag{1.53}$$

Ce théorème démontre donc qu'il existe une fonctionnelle exacte $F_{HK}[\rho] = \min_{\psi\to\rho}\{\langle\psi|\hat{T}_e + \hat{V}_{ee}|\psi\rangle\}$ grâce à laquelle il est possible d'exprimer l'énergie totale en se basant uniquement sur la densité électronique. L'expression analytique exacte de cette fonctionnelle reste toutefois inconnue à ce jour.

1.5.3 L'approche de Kohn-Sham

Après des tentatives infructueuses de trouver la forme de la fonctionnelle d'énergie exacte $F_{HK}[\rho]$, Kohn et Sham [24] ont proposé une décomposition de la fonctionnelle en deux composantes en se basant sur un système électronique fictif sans interactions entre particules mais possédant la même densité électronique que le système réel :

$$F_{HK}[\rho] = T_S[\rho] + E_{HXC}[\rho] \tag{1.54}$$

$T_S[\rho]$ est la fonctionnelle d'énergie cinétique des électrons n'interagissant pas entre eux dans le système fictif et $E_{HXC}[\rho]$ est la fonctionnelle d'Hartree-échange-corrélation.

Comme pour l'énergie, on peut définir $T_S[\rho]$ en utilisant une recherche contrainte :

$$T_S[\rho] = \min_{\phi\to\rho}\langle\phi|\hat{T}|\phi\rangle = \langle\phi[\rho]|\hat{T}|\phi[\rho]\rangle \tag{1.55}$$

Dans ce cadre, ϕ est la fonction d'onde exacte du système exprimée par un déterminant de Slater ϕ construit à partir des orbitales de Kohn-Sham (ϕ_i^{KS}). L'énergie électronique s'écrit alors :

$$E[\{\phi_i^{KS}\}] = \sum_{i=1}^{N}\int \phi_i^{KS}(\vec{r})\left(-\frac{1}{2}\nabla^2 + v_{ext}(\vec{r})\right)\phi_i^{KS}(\vec{r})d^3r + E_{HXC}[\rho] \tag{1.56}$$

et la densité électronique s'écrit comme une somme du carré des modules des orbitales de Kohn-Sham :

$$\rho(\vec{r}) = \sum_{i=1}^{N} \left|\phi_i^{KS}(\vec{r})\right|^2 \tag{1.57}$$

On peut ainsi réécrire les équations (1.50) à (1.53) pour obtenir l'énergie et la densité électronique exactes grâce au principe variationnel :

$$E_0 \quad = \min_{\rho}\left\{F_{HK}[\rho] + \int \rho(\vec{r})v_{ext}(\vec{r})d^3r\right\} \tag{1.58}$$

$$= \min_{\rho}\left\{\min_{\phi\to\rho}\{\langle\phi|\hat{T}|\phi\rangle\} + E_{HXC}[\rho] + \int \rho(\vec{r})v_{ext}(\vec{r})d^3r\right\} \tag{1.59}$$

$$= \min_{\rho}\left\{\min_{\phi\to\rho}\{\langle\phi|\hat{T} + \hat{V}_{ne}|\phi\rangle\} + E_{HXC}[\rho_\phi]\right\} \tag{1.60}$$

$$= \min_{\phi}\{\langle\phi|\hat{T} + \hat{V}_{ne}|\phi\rangle + E_{HXC}[\rho_\phi]\} \tag{1.61}$$

En revanche, la fonctionnelle d'Hartree-échange-corrélation $E_{HXC}[\rho]$ doit être approximée ; on l'écrit sous la forme d'une somme de deux fonctionnelles :

$$E_{HXC}[\rho] = E_H[\rho] + E_{XC}[\rho] \tag{1.62}$$

$E_H[\rho]$ est la fonctionnelle d'énergie de Hartree et $E_{XC}[\rho]$ la fonctionnelle d'échange-corrélation. $E_H[\rho]$ correspond à la répulsion électrostatique classique engendrée par la densité électronique :

$$E_H[\rho] = \frac{1}{2}\iint \frac{\rho(\vec{r})\rho(\vec{r}')}{|\vec{r} - \vec{r}\,'|} d^3r d^3r' \tag{1.63}$$

Quant à elle, la fonctionnelle d'échange-corrélation se décompose à nouveau en une somme de deux fonctionnelles, la fonctionnelle d'énergie d'échange $E_X[\rho]$ et la fonctionnelle d'énergie de corrélation $E_C[\rho]$:

$$E_{XC}[\rho] = E_X[\rho] + E_C[\rho] \tag{1.64}$$

$E_{XC}[\rho]$ peut également s'écrire comme la somme de deux corrections : une correction d'énergie cinétique de corrélation et une correction de répulsion bi-électronique, puisqu'il faut tenir compte des répulsions entre électrons dans le système réel :

$$E_{XC}[\rho] = (T[\rho] - T_S[\rho]) + (V_{ee} - E_H[\rho]) \tag{1.65}$$

1.5.4 Les fonctionnelles d'échange-corrélation

Dans la méthode de Kohn-Sham, le seul élément ne pouvant être calculé de façon exacte est la fonctionnelle d'échange-corrélation $E_{XC}[\rho]$. Il est possible de s'en approcher par plusieurs approximations que l'on peut classer sur une échelle (dite « échelle de Jacob de la DFT » [25]) :

Échelon	Type de fonctionnelle	Termes pris en compte
« ∞ »	Exactitude	?
5	Non-locales virtuelles	$\rho, \vec{\nabla}\rho, \nabla^2\rho, \tau, E_X^{HF}, E_C^{PT2}$
4	Hybrides	$\rho, \vec{\nabla}\rho, \nabla^2\rho, \tau, E_X^{HF}$
3	Meta-GGA	$\rho, \vec{\nabla}\rho, \nabla^2\rho, \tau$
2	GGA	$\rho, \vec{\nabla}\rho$
1	LDA	ρ
0	Hartree-Fock	-

Voici à quoi correspondent ces différents niveaux d'approximation :

- L'échelon « zéro » correspond à la méthode Hartree-Fock, sans aucune corrélation. On peut en effet voir cette méthode comme un cas particulier de la DFT.
- Le premier échelon regroupe les fonctionnelles utilisant l'approximation LDA [26] (« *Local Density Approximation* », approximation de la densité locale). Dans ce cadre, le système est considéré comme un gaz uniforme d'électrons dans lequel baignent les noyaux. L'énergie d'échange-corrélation est de la forme :

$$E_{XC}^{LDA}[\rho] = \int \rho(\vec{r})\varepsilon_{XC}[\rho]d^3r \tag{1.66}$$

$\rho(\vec{r})\varepsilon_{XC}[\rho]$ correspond à la densité d'énergie d'échange-corrélation, où $\varepsilon_{XC}[\rho]$ est l'énergie d'échange-corrélation de chaque électron d'un gaz uniforme d'électrons. On peut décomposer cette énergie en un terme d'échange $\varepsilon_X[\rho]$ et un terme de corrélation $\varepsilon_C[\rho]$. La solution de Dirac pour le gaz uniforme d'électrons est utilisée pour le terme d'échange :

$$\varepsilon_X[\rho] = -\frac{3}{4}\sqrt[3]{\frac{3}{\pi}\rho(\vec{r})} \tag{1.67}$$

Pour l'énergie de corrélation, il existe plusieurs propositions dans la littérature, dont notamment la fonctionnelle VWN introduite par Vosko, Wilk et Nusai en 1980 [27] et la fonctionnelle PW92L proposée par Perdew et Wang en 1992 [28].

- Le second échelon correspond aux fonctionnelles reposant sur l'approximation GGA [29] (« *Generalized Gradient Approximation* », approximation du gradient généralisé), qui prend en compte le gradient de la densité électronique $\vec{\nabla}\rho$ en plus de la densité, améliorant ainsi la précision des méthodes GGA par rapport aux méthodes LDA :

$$E_{XC}^{GGA}[\rho] = \int \rho(\vec{r})\varepsilon_{XC}\left[\rho, \vec{\nabla}\rho\right]d^3r \tag{1.68}$$

Il existe de nombreuses fonctionnelles GGA d'échange et de corrélation. Les exemples les plus populaires incluent d'une part la fonctionnelle d'échange-corrélation BLYP utilisant la fonctionnelle d'échange B88 (proposée par Becke en 1988 [29]) et la fonctionnelle de corrélation LYP (proposée par Lee, Yang et Parr la même année [30]), et d'autre part la fonctionnelle PBE développée par Perdew, Burke et Ernzerhof [31] et qui comprend à la fois échange et corrélation.

- Au troisième échelon, on trouve les fonctionnelles dites méta-GGA [32], qui prennent en compte des dérivées de la densité électronique d'ordre supérieur (notamment le Laplacien $\nabla^2\rho$) et/ou la densité d'énergie cinétique définie par $\tau(\vec{r}) = \sum_i \left|\nabla\phi_i^{KS}(\vec{r})\right|^2$. Les fonctionnelles de ce type comprennent les fonctionnelles TPSS [32], VSXC [33] et certaines fonctionnelles du Minnesota (comme M06L [34]).
- Le quatrième échelon comprend les fonctionnelles hybrides [35] qui sont divisées en deux catégories : les fonctionnelles hybrides globales et les fonctionnelles à séparation de portée. Elles incluent une contribution d'échange exact E_{XX} en plus de la contribution DFT E_{XC}^{DFT} qui peut provenir de fonctionnelles des échelons précédents. Elles ont la forme générale suivante :

$$E_{XC}^{hyb} = aE_{XX}[\{\phi_i^{KS}\}] + (1-a)E_{XC}^{DFT}[\rho] \quad (1.69)$$

où le paramètre a indique la proportion d'échange exact utilisé. L'une des fonctionnelles les plus utilisées en chimie quantique est une fonctionnelle hybride globale, B3LYP [36], qui a la forme :

$$E_{XC}^{B3LYP} = aE_{XX} + bE_X^{LDA} + (1-a-b)E_X^{B88} + cE_C^{LDA} + (1-c)E_C^{LYP} \quad (1.70)$$

Pour B3LYP, les valeurs de ces paramètres sont $a = 0{,}20$, $b = 0{,}08$ et $c = 0{,}19$. Nous avons utilisé cette fonctionnelle pour certains de nos calculs. Une autre fonctionnelle couramment utilisée est M06-2X [37], qui utilise des paramètres optimisés empiriquement et inclut 54 % d'échange exact. Quant à elles, les fonctionnelles hybrides à séparation de portée divisent les contributions d'échange en une partie à courte portée et une partie à longue portée, et utilisent des contributions énergétiques différentes à courte et longue portée. Cela permet de mieux tenir compte des interactions à longue portée. L'expression générale est donnée ci-dessous :

$$E_{XC}^{RSH} = c_{X,SR}E_{XX,SR} + c_{X,LR}E_{XX,LR} + (1-c_{X,SR})E_{X,SR}^{DFT} + (1-c_{X,LR})E_{X,LR}^{DFT} + E_C^{DFT} \quad (1.71)$$

Un exemple de telle fonctionnelle est ωB97X-D [38], qui utilise 22 % d'échange exact à courte distance, 78 % d'échange DFT à courte distance et un facteur ω = 0,3 bohr^{-1} pour contrôler l'étendue spatiale de la contribution à courte portée.

Nous avons également utilisé cette fonctionnelle.

- Le cinquième échelon inclut les dernières générations de fonctionnelles, dites « non-locales virtuelles ». Celles-ci comprennent les fonctionnelles doubles hybrides [39], qui comprennent une part d'échange exact comme les hybrides mais également une part de corrélation issue de méthodes perturbatives au second ordre (PT2) de type Görling-Levy qui font intervenir les orbitales de Kohn-Sham non occupées (d'où le nom de « virtuelles »). Ce type de fonctionnelle se base sur le suivi d'un chemin adiabatique entre le système fictif sans interactions et le système réel. Dans ce cas, la fonctionnelle d'échange-corrélation s'exprime par :

$$E_{XC}^{DH} = aE_{XX} + (1-a)E_X^{DFT} + cE_C^{PT2} + (1-c)E_C^{DFT} \quad (1.72)$$

Quelques exemples incluent les fonctionnelles B2-PLYP [39] (avec $a = 0{,}53$ et $b = 0{,}27$) et B2K-PLYP [40] (avec $a = 0{,}42$ et $b = 0{,}72$).

Enfin, il faut mentionner les corrections dites de dispersion [41,42], qui permettent de mieux décrire les interactions de van der Waals (non-covalentes). L'une des méthodes les plus notables est celle de Grimme, qui consiste à ajouter un terme de correction classique à l'énergie. Le modèle DFT-D2 [43] est un des premiers modèles de cette famille et utilise des termes en $-\frac{C}{R^6}$:

$$E_{disp}^{D2} = -s_6 \sum_{A}^{M} \sum_{B<A}^{M} \frac{C_{6,AB}}{R_{AB}^6} f_{damp}^{D2}(R_{AB}) \tag{1.73}$$

$$f_{damp}^{D2}(R_{AB}) = \frac{1}{1 + e^{-D_6(\frac{R_{AB}}{R_{0,AB}} - 1)}} \tag{1.74}$$

où s_6 et D_6 sont des facteurs dépendant de la fonctionnelle, les coefficients $C_{6,AB}$ sont paramétrés empiriquement pour les différents atomes, et f_{damp}^{D2} [44] est une fonction de la distance R_{AB} permettant d'atténuer la correction à courte portée (puisque $\frac{1}{R_{AB}^6}$ diverge à courte distance) en utilisant une distance de coupure $R_{0,AB}$ (la somme des rayons de van der Waals des atomes considérés).

La méthode DFT-D3 [45] améliore la précision de la méthode DFT-D2 en ajoutant des termes $-\frac{C}{R^8}$:

$$E_{disp}^{D3} = -\sum_{A}^{M} \sum_{B<A}^{M} \left[s_6 \frac{C_{6,AB}}{R_{AB}^6} f_{damp,6}^{D3}(R_{AB}) + s_8 \frac{C_{8,AB}}{R_{AB}^8} f_{damp,8}^{D3}(R_{AB}) \right] \tag{1.75}$$

$$f_{damp,n}^{D3}(R_{AB}) = \frac{1}{1 + \frac{D_n}{\left(\frac{R_{AB}}{s_{R_{AB},n} R_{0,AB}}\right)^n}} \tag{1.76}$$

Ici, $s_{R_{AB},n}$ est un facteur d'échelonnement de la distance de coupure dépendant de l'ordre n des termes considérés. Il est également possible de continuer à ajouter des termes d'ordre pair supérieur (10, 12…).

La méthode DFT-D4 [46] est la méthode la plus récente de cette famille ; les coefficients sont calculés à partir d'intégrations de la polarisabilité atomique en utilisant les charges partielles atomiques. Cette méthode prend également en compte des interactions à trois centres.

Nous avons utilisé les méthodes DFT-D2 et DFT-D3 en conjonction avec la fonctionnelle B3LYP (qu'on note alors B3LYP-D2 ou B3LYP-D3) dans nos calculs.

Il est également possible de calculer les termes de dispersion par le biais de fonctionnelles explicites de la densité électronique [47], d'ajouter un potentiel effectif non-local au potentiel de Kohn-Sham [48] ou d'optimiser les paramètres des fonctionnelles de sorte à prendre en compte cet effet sans forcément avoir de terme correctif.

1.5.5 Les équations de Kohn-Sham

L'énergie est minimisée avec le Lagrangien suivant :

$$\mathcal{L}\left[\{\phi_i^{KS}\}\right] = E\left[\{\phi_i^{KS}\}\right] - \sum_{i=1}^{N} \varepsilon_i \left(\int \phi_i^{KS}(\vec{r})\phi_i^{KS}(\vec{r})d^3r - 1 \right) \tag{1.77}$$

où ε_i est le multiplicateur de Lagrange associé à la condition de normalisation suivante :

$$\int \phi_i^{KS}(\vec{r})\phi_i^{KS}(\vec{r})d^3r = 1 \tag{1.78}$$

La minimisation du Lagrangien amène à :

$$\frac{\delta\mathcal{L}}{\delta\phi_i^{KS}(\vec{r})} = 0 \tag{1.79}$$

Le calcul de cette dérivée fonctionnelle amène aux N équations suivantes (1 par orbitale) :

$$\left(-\frac{1}{2}\nabla^2 + v_{ext}(\vec{r})\right)\phi_i^{KS}(\vec{r}) + \frac{\delta E_{HXC}[\rho]}{\delta\phi_i^{KS}(\vec{r})} = \varepsilon_i\phi_i^{KS}(\vec{r}) \tag{1.80}$$

On peut utiliser une dérivation en chaîne pour exprimer le terme $\frac{\delta E_{HXC}[\rho]}{\delta\phi_i^{KS}(\vec{r})}$:

$$\frac{\delta E_{HXC}[\rho]}{\delta\phi_i^{KS}(\vec{r})} = \int \frac{\delta E_{HXC}[\rho]}{\delta\rho(\vec{r}')}\frac{\delta\rho(\vec{r}')}{\delta\phi_i^{KS}(\vec{r})}d^3r' \tag{1.81}$$

Le terme $\frac{\delta E_{HXC}[\rho]}{\delta\rho(\vec{r}')}$ est appelé le potentiel d'Hartree-échange-corrélation $v_{HXC}(\vec{r})$:

$$v_{HXC}(\vec{r}) = \frac{\delta E_{HXC}[\rho]}{\delta \rho(\vec{r})} \tag{1.82}$$

On obtient ainsi les équations dites de Kohn-Sham :

$$\left(-\frac{1}{2}\nabla^2 + v_{HXC}(\vec{r}) + v_{ext}(\vec{r})\right)\phi_i^{KS}(\vec{r}) = \varepsilon_i \phi_i^{KS}(\vec{r}) \tag{1.83}$$

$$\hat{h}_{KS}(\vec{r})\phi_i^{KS}(\vec{r}) = \varepsilon_i \phi_i^{KS}(\vec{r}) \tag{1.84}$$

où $\hat{h}_{KS}$, l'Hamiltonien de Kohn-Sham, correspond à :

$$\hat{h}_{KS}(\vec{r}) = -\frac{1}{2}\nabla^2 + v_{KS}(\vec{r}) \tag{1.85}$$

$$v_{KS}(\vec{r}) = v_{HXC}(\vec{r}) + v_{ext}(\vec{r}) \tag{1.86}$$

$v_{KS}(\vec{r})$ est appelé potentiel de Kohn-Sham.

Les équations de Kohn-Sham définissent le système fictif à N électrons non-interagissants soumis au potentiel de Kohn-Sham et possédant la même densité électronique que le système réel à son état fondamental.

Comme la fonctionnelle $E_{HXC}[\rho]$, le potentiel $v_{HXC}(\vec{r})$ peut se décomposer :

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

$v_H(\vec{r})$ est le potentiel de Hartree et $v_{XC}(\vec{r})$ le potentiel d'échange-corrélation, définis comme :

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

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

En pratique, la résolution des équations de Kohn-Sham suit le même principe que celle des équations de Hartree-Fock, avec un cycle auto-cohérent [49] :

1. Calculer les orbitales de Kohn-Sham initiales ;
2. Calculer la densité électronique initiale $\rho(\vec{r})$ à partir des orbitales de Kohn-Sham ;
3. Calculer le potentiel de Kohn-Sham $v_{KS}(\vec{r})$;
4. Résoudre les équations de Kohn-Sham pour obtenir une nouvelle densité électronique $\rho'(\vec{r})$;

5. Vérifier la convergence du calcul (a-t-on $\rho' \approx \rho$?) ;

    a. Si oui, on peut arrêter le calcul SCF et déterminer les propriétés du système à partir de $\rho(\vec{r})$;

    b. Sinon, il faut retourner à l'étape 3 en utilisant $\rho'(\vec{r})$.

Bibliographie du chapitre 1

1. L. de Broglie. *Comptes Rendus* **177**, 507–510 (1923).
2. E. Schrödinger. *Ann. Phys.* **384**, 361–376 (1926).
3. M. Born & P. Jordan. *Zeitschrift für Phys.* **34**, 858–888 (1925).
4. M. Born & R. Oppenheimer. *Ann. Phys.* **389**, 457–484 (1927).
5. D. R. Hartree. *Math. Proc. Cambridge Philos. Soc.* **24**, 111–132 (1928).
6. V. Fock. *Zeitschrift für Phys.* **61**, 126–148 (1930).
7. W. Pauli. *Zeitschrift für Phys.* **31**, 765–783 (1925).
8. S. M. Blinder. *Am. J. Phys.* **33**, 431–443 (1965).
9. C. C. J. Roothaan. *Rev. Mod. Phys.* **23**, 69–89 (1951).
10. G. G. Hall. *Proc. R. Soc. London. Ser. A. Math. Phys. Sci.* **205**, 541–552 (1951).
11. P.-O. Löwdin. in *Advances in Chemical Physics, vol. 2* 207–322 (John Wiley & Sons, Ltd, 1958).
12. C. Møller & M. S. Plesset. *Phys. Rev.* **46**, 618–622 (1934).
13. J. Ĉíźek. *J. Chem. Phys.* **45**, 4256–4266 (1966).
14. J. C. Slater. *Phys. Rev.* **36**, 57–64 (1930).
15. S. F. Boys. *Proc. R. Soc. London. Ser. A. Math. Phys. Sci.* **200**, 542–554 (1950).
16. R. Ditchfield, W. J. Hehre & J. A. Pople. *J. Chem. Phys.* **54**, 720–723 (1971).
17. T. H. Dunning. *J. Chem. Phys.* **90**, 1007–1023 (1989).
18. A. Schäfer, H. Horn & R. Ahlrichs. *J. Chem. Phys.* **97**, 2571–2577 (1992).
19. R. G. Parr & Y. Weitao. *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, 1989).
20. A. Szabo & N. S. Ostlund. *Modern quantum chemistry : introduction to advanced electronic structure theory* (Dover Publications, 1996).
21. P. Coppens. *Annu. Rev. Phys. Chem.* **43**, 663–692 (1992).
22. P. Hohenberg & W. Kohn. *Phys. Rev.* **136**, B864–B871 (1964).
23. M. Levy. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 6062–6065 (1979).
24. W. Kohn & L. J. Sham. *Phys. Rev.* **140**, A1133–A1138 (1965).
25. J. P. Perdew. in *AIP Conference Proceedings* vol. 577 (American Institute of Physics, 2001).
26. J. P. Perdew & A. Zunger. *Phys. Rev. B* **23**, 5048 (1981).
27. S. H. Vosko, L. Wilk & M. Nusair. *Can. J. Phys.* **58**, 1200–1211 (1980).
28. J. P. Perdew & Y. Wang. *Phys. Rev. B* **45**, 13244 (1992).
29. A. D. Becke. *Phys. Rev. A* **38**, 3098–3100 (1988).
30. C. Lee, W. Yang & R. G. Parr. *Phys. Rev. B* **37**, 785–789 (1988).

31. J. P. Perdew, K. Burke & M. Ernzerhof. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).

32. J. Tao, J. P. Perdew, V. N. Staroverov & G. E. Scuseria. *Phys. Rev. Lett.* **91**, 146401 (2003).

33. T. Van Voorhis & G. E. Scuseria. *J. Chem. Phys.* **109**, 400–410 (1998).

34. Y. Zhao & D. G. Truhlar. *J. Chem. Phys.* **125**, 194101 (2006).

35. A. D. Becke. *J. Chem. Phys.* **98**, 5648–5652 (1993).

36. P. J. Stephens, F. J. Devlin, C. F. Chabalowski & M. J. Frisch. *J. Phys. Chem.* **98**, 11623–11627 (1994).

37. Y. Zhao & D. G. Truhlar. *Theor. Chem. Acc.* **120**, 215–241 (2008).

38. J. Da Chai & M. Head-Gordon. *Phys. Chem. Chem. Phys.* **10**, 6615–6620 (2008).

39. S. Grimme. *J. Chem. Phys.* **124**, 034108 (2006).

40. A. Tarnopolsky, A. Karton, R. Sertchook, D. Vuzman & J. M. L. Martin. *J. Phys. Chem. A* **112**, 3–8 (2008).

41. S. Grimme. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**, 211–228 (2011).

42. S. Grimme, A. Hansen, J. G. Brandenburg & C. Bannwarth. *Chem. Rev.* **116**, 5105–5154 (2016).

43. S. Grimme. *J. Comput. Chem.* **27**, 1787–1799 (2006).

44. S. Grimme, S. Ehrlich & L. Goerigk. *J. Comput. Chem.* **32**, 1456–1465 (2011).

45. S. Grimme, J. Antony, S. Ehrlich & H. Krieg. *J. Chem. Phys.* **132**, 154104 (2010).

46. E. Caldeweyher, C. Bannwarth & S. Grimme. *J. Chem. Phys.* **147**, 034112 (2017).

47. M. Dion, H. Rydberg, E. Schröder, D. C. Langreth & B. I. Lundqvist. *Phys. Rev. Lett.* **92**, 246401 (2004).

48. O. A. Von Lilienfeld, I. Tavernelli, U. Rothlisberger & D. Sebastiani. *Phys. Rev. Lett.* **93**, 153004 (2004).

49. J. C. Slater. *Quantum theory of molecules and solids.* (McGraw-Hill, 1974).

Chapitre 2 – Méthodes d'analyse en chimie quantique

2.1 Modélisation des surfaces d'énergie potentielle 50

2.2 Modélisation du solvant (implicite et explicite) 53

2.3 DFT conceptuelle 55

2.3.1 Fondements de la DFT conceptuelle 55

2.3.2 Descripteurs globaux 57

2.3.3 Descripteurs locaux 59

2.3.4 Principes de réactivité 62

2.4 Analyses de population 62

2.4.1 Analyses de population basées sur les orbitales moléculaires 63

2.4.2 Analyses de population basées sur le potentiel électrostatique 66

2.4.3 Analyses de population basées sur la densité électronique 67

2.5 Méthode REG 77

2.5.1 Objectif 77

2.5.2 Principe 78

2.5.3 Applications 83

2.1 Modélisation des surfaces d'énergie potentielle

Lors de l'étude d'une molécule en chimie théorique, il n'y a pas une structure unique à considérer mais le plus souvent un large éventail : il y a bien sûr les différents conformères (stables ou métastables) mais également des structures instables telles que les états de transition. Il ne faut pas non plus oublier les divers réarrangements intramoléculaires possibles (résultant entre autres de tautomérisations ou de transpositions). Ces structures ont évidemment des énergies très variées, et leur ensemble forme ce que l'on appelle la surface d'énergie potentielle (PES, « *potential energy surface* » en anglais). Cette hypersurface a autant de dimensions qu'il y a de degrés de liberté dans le système, ce qui rend très rapidement impossible sa visualisation complète. Cependant, en suivant la variation d'un ou deux degrés de liberté, il est possible de dresser des profils énergétiques (un degré de liberté) ou des surfaces d'énergie potentielle (deux degrés de liberté) comme celle représentée sur la Figure 2.3.

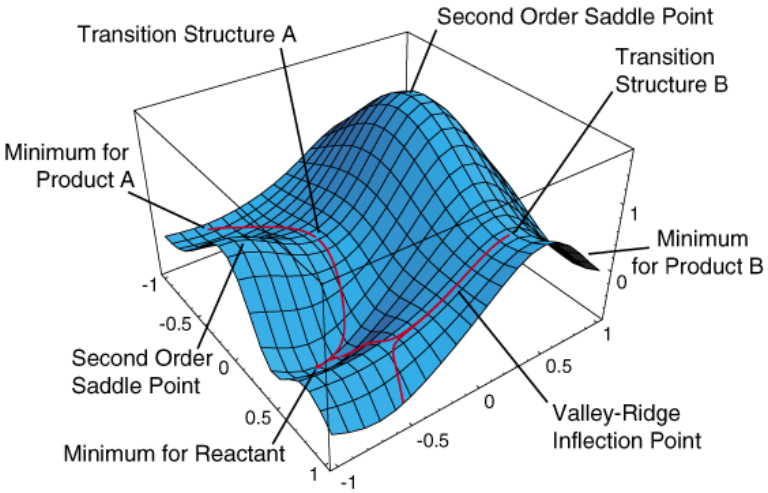


Figure 2.3 - Représentation schématique d'une surface d'énergie potentielle

Les axes des abscisses correspondent chacun à un degré de liberté du système tandis que l'énergie est tracée en ordonnée. Les chemins d'énergie minimale reliant les points stationnaires sont indiqués en rouge.

Les points remarquables sur cette surface d'énergie potentielle sont :

- Les minima énergétiques, qui peuvent être un minimum global (stable) ou des minima locaux (métastables), qui sont des conformères ou des isomères que l'on peut souvent caractériser expérimentalement ;

- Les points selle, ou cols, qui sont des minima énergétiques selon tous les degrés de liberté du système sauf un, suivant lequel ce sont des maxima énergétiques ; ce sont des états de transition qui ont expérimentalement une durée de vie très brève et ne peuvent être caractérisés qu'indirectement (en les piégeant, par exemple). Il faut également noter que les états de transition sont les points les plus hauts en énergie sur le chemin de plus basse énergie reliant deux minima énergétiques.

Lors des calculs, la nature des minima et états de transition est caractérisée par le calcul de la matrice Hessienne de l'énergie, dont les valeurs propres correspondent aux courbures de la surface d'énergie potentielle selon les modes normaux de vibration du système. Ces valeurs propres sont également reliées aux constantes de force et *in fine* aux fréquences infrarouge et Raman. Dans chaque cas :

- Les fréquences des minima sont toutes réelles ;
- Les fréquences des états de transition sont toutes réelles, sauf une qui est imaginaire. Cette unique fréquence imaginaire correspond au mode de vibration menant aux deux minima situés de part et d'autre de l'état de transition.

Enfin, il est possible de suivre la progression d'une réaction des réactifs aux produits en passant par l'état de transition en utilisant une coordonnée intrinsèque de réaction (IRC, *Intrinsic Reaction Coordinate*) notée ξ. Il s'agit d'un nombre réel prenant en compte les variations des différents paramètres structuraux variant au cours de la réaction ; il correspond à l'abscisse curviligne du chemin de plus basse énergie reliant les réactifs aux produits.

Pour des cas très simples, la coordonnée intrinsèque de réaction est simplement une distance ou un angle : par exemple, dans la pyramidalisation de l'ammoniac, l'angle H-N-H est la coordonnée intrinsèque de réaction.

Toro-Labbé et al. [1] ont proposé de définir une « force » de réaction comme l'opposé de la dérivée de l'énergie par rapport à la coordonnée intrinsèque de réaction :

$$F(\xi) = -\frac{\partial E}{\partial \xi} \tag{2.1}$$

Pour une étape élémentaire (un seul état de transition), cette grandeur, qui a la dimension d'une force, est donc négative avant l'état de transition et positive après. Ses extrema, qui correspondent aux points d'inflexion du profil énergétique, permettent de définir trois zones représentées sur la Figure 2.4 :

- La zone des réactifs, qui va jusqu'au premier point d'inflexion ;
- La zone de l'état de transition, qui va du premier point d'inflexion au second, et englobe ainsi l'état de transition ;
- La zone des produits, après le second point d'inflexion.

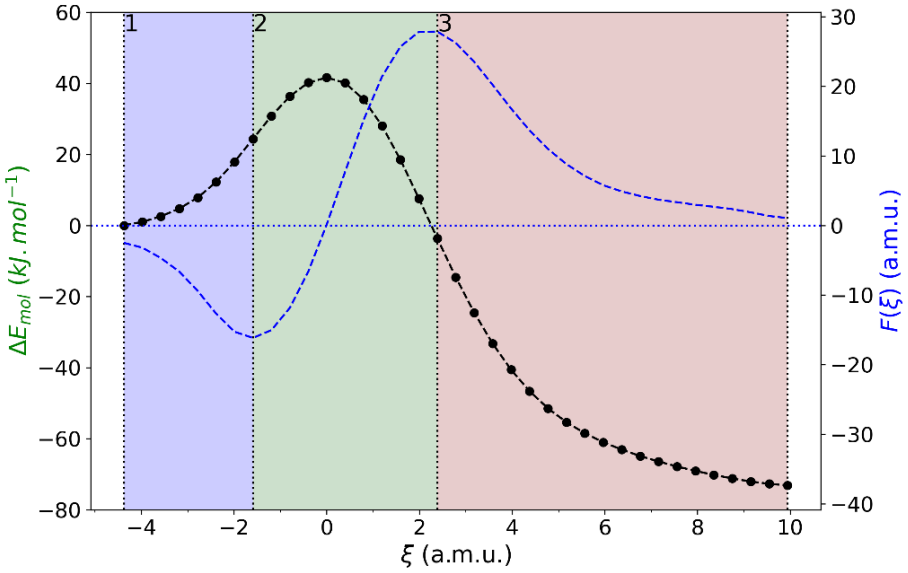


Figure 2.4 – Graphique des zones de la réaction définies par Toro-Labbé et al. [1]

Le profil énergétique est représenté en pointillés noirs, la force de réaction en pointillés bleus, la zone des réactifs est en bleu, la zone de l'état de transition en vert et celle des produits en rouge.

Les études de Toro-Labbé et al. [2,3] ont ainsi montré que ce découpage permet de bien distinguer les évènements principaux de la réaction : le plus souvent, les zones des réactifs et des produits correspondent majoritairement à des réorganisations structurales, sans transfert d'électrons, tandis que la zone de l'état de transition est celle où la réaction proprement dite a lieu, avec les ruptures et formations de liaisons.

2.2 Modélisation du solvant (implicite et explicite)

La modélisation de réactions chimiques peut se faire en phase gaz (en considérant que les molécules sont dans le vide), mais il est également possible de prendre en compte des effets de solvant.

Cela peut se faire de deux façons : en prenant en compte l'effet du solvant de façon implicite, ou en ajoutant de façon explicite des molécules de solvant autour du soluté.

Parmi les méthodes de solvatation implicites, la plus simple est sans doute la méthode du champ de réaction d'Onsager [4]. Dans ce cadre, l'influence du solvant sur le soluté peut se voir comme un effet de polarisation qui va déformer la densité électronique du soluté, et qui interagit à son tour avec le solvant. À l'Hamiltonien des équations de Kohn-Sham s'ajoute alors un Hamiltonien correspondant à la perturbation du système par le solvant.

Dans ce modèle, le solvant est décrit comme un continuum possédant une permittivité diélectrique constante et le soluté est placé dans une cavité au sein de ce continuum où la constante diélectrique relative (ε_r) est égale à 1. La surface de la cavité est définie comme étant une simple sphère centrée sur le soluté.

Les calculs se déroulent alors en deux temps : le cycle SCF du soluté seul en phase gaz est mené jusqu'à convergence, puis l'influence des charges est reportée sur le continuum, qui cause à son tour une déformation de la densité électronique du soluté ; ce second cycle auto-cohérent continue jusqu'à convergence. On parle dans ce cas de SCRF (« *Self-Consistent Reaction Field* », champ de réaction auto-cohérent).

La méthode PCM [5,6] (« *Polarizable Continuum Model* », modèle du continuum polarisable), l'une des méthodes de solvatation implicite les plus utilisées, raffine ce modèle en utilisant une cavité non pas sphérique mais dépendant de la structure du soluté. En plus du continuum diélectrique du modèle d'Onsager, le solvant est représenté comme un ensemble de charges ponctuelles parsemées sur la surface de la cavité (suivant le modèle ASC, « *apparent surface charge* », charges apparentes de surface). On y distingue trois surfaces différentes, de la plus proche du soluté à la plus éloignée :

- La surface de van der Waals est la surface externe d'un ensemble de sphères se recouvrant entre elles, centrées sur les noyaux et dont le rayon est égal au rayon de van der Waals des différents atomes lourds (excluant donc l'hydrogène). C'est sur cette surface que sont positionnées les charges ponctuelles représentant le solvant.
- La surface d'exclusion du solvant est construite à partir de la surface de van der Waals : tous les rayons des sphères sont multipliés par un facteur dépendant du solvant, puis la surface obtenue est lissée en ajoutant des sphères non centrées sur des noyaux.
- Enfin, la surface accessible au solvant incrémente la surface d'exclusion du solvant en tenant compte du rayon idéal du solvant (la taille des molécules de solvant si ces dernières étaient sphériques).

Ces trois surfaces sont représentées sur le schéma de la Figure 2.5 ci-dessous.

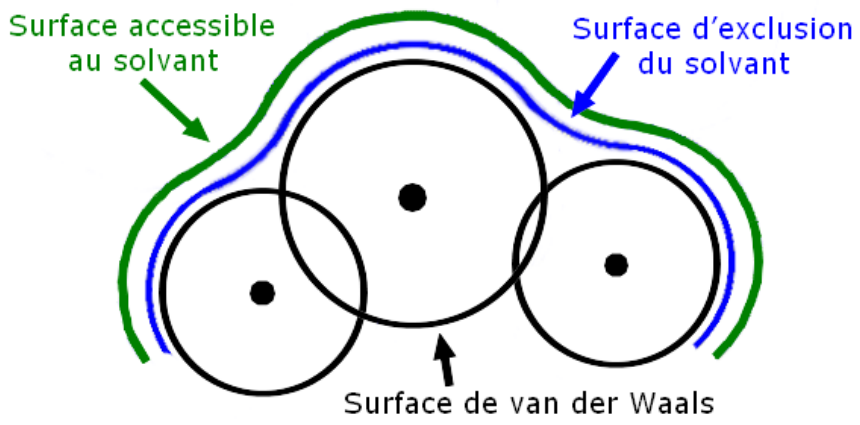


Figure 2.5 - Schéma des trois surfaces utilisées dans la méthode PCM

Dans la méthode PCM, l'énergie libre totale est calculée comme une somme de trois termes :

$$G = G_{cav} + G_{es} + G_{dr} \tag{2.2}$$

G_{cav} est l'énergie de cavitation, qui se calcule sur la base de la surface de van der Waals, G_{es} est l'énergie électrostatique, qui se base sur la surface d'exclusion du solvant, et G_{dr} est l'énergie de dispersion-répulsion qui prend en compte la surface accessible au solvant. L'énergie est calculée via un cycle auto-

cohérent comme pour la méthode précédente. Nous avons utilisé cette méthode sous sa variante IEF-PCM [7], qui utilise un formalisme d'équations intégrales (« *Integral Equation Formalism* »).

Ces méthodes ont un coût computationnel faible, mais elles ne permettent pas d'étudier les interactions spécifiques avec le solvant (par exemple les liaisons hydrogène dans la première couche de solvatation). Pour cela, il reste nécessaire de modéliser le solvant de façon explicite en plaçant des molécules de solvant autour du soluté. Cependant, le nombre de configurations possibles, le temps de calcul et les interactions possibles augmentent très rapidement avec le nombre de molécules. De plus, l'utilisation d'un nombre fini de molécules entraîne des effets de bord qui n'existent pas dans le cas réel.

Il est possible de réduire le coût computationnel de cette méthode en utilisant des stratégies de calcul hybrides qui utilisent deux (ou plus) niveaux de théorie différents pour différentes parties du système. Par exemple, dans la méthode QM/MM [8] (« *Quantum Mechanics / Molecular Mechanics* », mécanique quantique / mécanique moléculaire), le système est divisé en deux parties : le système d'intérêt, qui est traité avec un niveau de théorie élevé (calcul coûteux mais précis) et son environnement, qui est traité par un niveau de théorie plus faible (calcul plus rapide mais moins précis). Les interactions entre les deux systèmes sont ensuite calculées de façon classique. La méthode ONIOM [9] (« *Our own n-layered Integrated molecular Orbital and Molecular mechanics* ») généralise cette méthode à N niveaux et à toute une panoplie de niveaux de théorie.

2.3 DFT conceptuelle

2.3.1 Fondements de la DFT conceptuelle

Pendant longtemps, les chimistes ont rationalisé des réactions et des phénomènes chimiques en utilisant différentes grandeurs comme l'électronégativité ou la dureté. Cependant, ces grandeurs ne s'identifiaient à aucun des résultats fournis par les calculs de chimie quantique.

C'est en 1978 que Parr et al. publient un article fondateur [10] dans lequel ils identifient le potentiel chimique μ à la dérivée de l'énergie du système E par rapport au nombre d'électrons N à potentiel externe $v(\mathbf{r})$ constant, et à l'opposé de l'électronégativité χ :

$$\mu = \left(\frac{\partial E}{\partial N}\right)_{v(\vec{r})} = -\chi \tag{2.3}$$

Par la suite, de nombreuses autres grandeurs, que l'on appellera descripteurs, ont été identifiées, comme la dureté moléculaire η [11], qui est la dérivée seconde de l'énergie par rapport au nombre d'électrons. C'est également l'inverse de la mollesse S :

$$\eta = \left(\frac{\partial^2 E}{\partial N^2}\right)_{v(\vec{r})} = \frac{1}{S} \tag{2.4}$$

La densité électronique elle-même s'avère également être un descripteur relié à l'énergie par sa dérivée par rapport au potentiel externe :

$$\rho(\vec{r}) = \left(\frac{\delta E}{\delta v(\vec{r})}\right)_N \tag{2.5}$$

Ensemble, ces descripteurs forment les bases de la CDFT (*Conceptual Density Functional Theory*, théorie conceptuelle de la fonctionnelle de la densité en français). Aujourd'hui, plusieurs centaines de descripteurs ont été proposés pour différents usages [12].

On distingue trois types de descripteurs :

- Les descripteurs globaux, qui ont une valeur unique pour une molécule de structure donnée,
- Les descripteurs locaux, dont la valeur dépend de la position dans l'espace,
- Les descripteurs non-locaux, dont la valeur dépend d'au moins deux positions dans l'espace.

Parmi tous ces descripteurs, on trouve des descripteurs issus de dérivées successives de l'énergie comme l'électronégativité ou la dureté mentionnées ci-dessus, mais également des descripteurs composites formés à partir des descripteurs précédents afin de reproduire certains phénomènes, comme l'indice d'électrophilie.

Il faut toutefois mentionner le problème de la dérivation de l'énergie par rapport au nombre d'électrons N : celui-ci est une variable discontinue, puisqu'il est physiquement impossible d'avoir un nombre fractionnaire d'électrons dans un système isolé. On peut toutefois définir l'énergie du système avec un nombre fractionnaire d'électrons en considérant un système ouvert ou en interpolant l'énergie du système à $N - 1$ et $N + 1$ électrons.

En pratique, pour calculer ces descripteurs, deux grandes approches sont utilisées :

- L'approche des orbitales moléculaires gelées (FMO, « *Frozen Molecular Orbitals* ») qui s'appuie sur les propriétés (énergie, densité…) de la HOMO et de la LUMO du système à N électrons.
- L'approche des différences finies (FDL, « *Finite Difference Linearization* ») qui s'appuie sur des calculs du système à géométrie gelée à $N-1$, N et $N+1$ électrons.

Ces deux approches permettent d'évaluer les mêmes descripteurs mais ne donnent pas nécessairement les mêmes valeurs.

2.3.2 Descripteurs globaux

2.3.2.1 Potentiel chimique (μ)

Le potentiel chimique μ évoqué précédemment est défini comme la première dérivée de l'énergie par rapport au nombre d'électrons à potentiel externe constant [10]. C'est également l'opposé de l'électronégativité χ.

C'est un descripteur global indiquant la capacité d'une molécule à gagner ou perdre des électrons : en effet, quand deux espèces A et B réagissent et échangent des électrons, à nombre constant d'électrons $N = N_A + N_B$, on peut écrire la relation suivante pour la variation de l'énergie du système (au premier ordre) :

$$\Delta E_{tot} = \Delta E_A + \Delta E_B = \mu_A \Delta N_A + \mu_B \Delta N_B = (\mu_A - \mu_B)\Delta N \qquad (2.6)$$

Supposons que $\mu_A > \mu_B$. Si les électrons sont transférés de l'espèce ayant le potentiel chimique le plus élevé (donc l'électronégativité la plus faible) vers l'espèce ayant le potentiel chimique le plus faible (donc l'électronégativité la plus élevée), on a $\Delta N = \Delta N_A < 0$ et $\Delta N_B > 0$, soit $\Delta E_{tot} < 0$, le système est stabilisé. Une fois que la différence des potentiels chimiques s'annule, le transfert d'électrons cesse. $\mu_A - \mu_B$ permet donc de décrire et de prédire l'écoulement des électrons dans les systèmes en réaction.

En pratique, μ est calculé soit à partir des énergies des orbitales frontières, soit à partir de l'énergie du système avec un électron manquant ou supplémentaire :

$$\mu_{FMO} = \frac{E_{HOMO} + E_{LUMO}}{2} \tag{2.7}$$

$$\mu_{FDL} = \frac{E(N+1) + E(N-1)}{2} \tag{2.8}$$

2.3.2.2 Dureté chimique (η) et mollesse (S)

En 1983, Parr et Pearson [11] ont défini la dureté chimique η comme la dérivée seconde de l'énergie par rapport au nombre d'électrons à potentiel externe constant ; c'est donc également la dérivée première du potentiel chimique par rapport au nombre d'électrons à potentiel externe constant. C'est aussi l'inverse de la mollesse S :

$$\eta = \left(\frac{\partial^2 E}{\partial N^2}\right)_{v(\vec{r})} = \left(\frac{\partial \mu}{\partial N}\right)_{v(\vec{r})} = \frac{1}{S} \tag{2.9}$$

η est un descripteur global qui représente la résistance du système au transfert d'électrons, tandis que la mollesse caractérise l'inverse, la capacité du système à garder des électrons.

η est directement reliée à l'énergie de réaction de la dismutation suivante pour une molécule M :

$$2\,M \rightarrow M^+ + M^- \tag{2.10}$$

η et S sont eux aussi calculés soit à partir des énergies des orbitales frontières, soit à partir de l'énergie du système avec un électron manquant ou supplémentaire :

$$\eta_{FMO} = I - A \tag{2.11}$$

$$\eta_{FDL} = (E(N+1) - E(N)) - (E(N) - E(N-1)) \tag{2.12}$$

Dans l'équation (2.11), I est le potentiel d'ionisation vertical du système et A son affinité électronique. Le théorème de Koopmans permet alors d'approcher η_{FMO} par le *gap* HOMO-LUMO :

$$\eta_{FMO} = E_{LUMO} - E_{HOMO} \tag{2.13}$$

2.3.2.3 Indices d'électrophilie (ω) et de nucléophilie (N)

Après avoir identifié des concepts chers aux chimistes tels que l'électronégativité, on peut se demander s'il est possible de retrouver un descripteur décrivant l'électrophilie et la nucléophilie de composés réactifs. Suite à une observation de corrélation de descripteurs avec l'énergie d'interaction en docking par Maynard et al [13], Parr et al [14] ont proposé un descripteur composite, c'est-à-dire qu'il n'est pas directement issu d'une dérivée de l'énergie mais plutôt calculé à partir d'autres descripteurs :

$$\omega = \frac{\mu^2}{2\eta} \tag{2.14}$$

ω est l'indice d'électrophilie et correspond à la stabilisation énergétique du système lors d'un transfert de charge, plus précisément au transfert du nombre optimal d'électrons vers le composé électrophile plongé dans une mer d'électrons à température et potentiel chimique nuls et à potentiel externe constant.

De façon similaire au couple dureté/mollesse, Domingo et al. [15] ont proposé de définir un indice de nucléophilie N comme l'inverse de l'indice d'électrophilie :

$$N = \frac{1}{\omega} \tag{2.15}$$

2.3.3 Descripteurs locaux

2.3.3.1 Fonctions de Fukui (f(r))

Les fonctions de Fukui [16] sont définies comme la dérivée mixte de l'énergie électronique par rapport au nombre d'électrons et au potentiel externe :

$$f(\vec{r}) = \frac{\delta^2 E}{\delta N \delta v(\vec{r})} = \left(\frac{\partial \rho(\vec{r})}{\partial N}\right)_{v(\vec{r})} = \left(\frac{\partial \mu}{\partial v(\vec{r})}\right)_N \tag{2.16}$$

On parle bien de fonction**s** de Fukui au pluriel car la dérivée $\frac{\partial \rho}{\partial N}$ n'est pas continue pour chaque valeur entière de N. Ainsi, il existe trois fonctions de Fukui selon que l'on considère la dérivée à gauche, à droite ou en zéro :

- la fonction de Fukui nucléophile $f^{-}(\vec{r})$, qui décrit une attaque électrophile (donc une perte d'électrons, ∂N est négatif) ;
- la fonction de Fukui électrophile $f^{+}(\vec{r})$, qui décrit une attaque nucléophile (donc un gain d'électrons, ∂N est positif) ;
- et la fonction de Fukui radicalaire $f^{0}(\vec{r})$, qui décrit une attaque radicalaire lorsque ∂N est nul.

Ces fonctions permettent de déterminer quel site est le plus susceptible d'attaquer ou d'être attaqué lors d'une réaction nucléophile/électrophile ou radicalaire. Elles sont nommées en hommage à K. Fukui, qui a développé la théorie des orbitales frontières, que l'on peut rationaliser entre autres grâce aux fonctions de Fukui (même si elles ont été développées après).

Pour le calcul des fonctions électrophile et nucléophile, on utilise les définitions suivantes :

$$f_{FMO}^{-}(\vec{r}) = \rho_{HOMO}(\vec{r}) \tag{2.17}$$

$$f_{FMO}^{+}(\vec{r}) = \rho_{LUMO}(\vec{r}) \tag{2.18}$$

$$f_{FDL}^{-}(\vec{r}) = \rho_{N}(\vec{r}) - \rho_{N-1}(\vec{r}) \tag{2.19}$$

$$f_{FDL}^{+}(\vec{r}) = \rho_{N+1}(\vec{r}) - \rho_{N}(\vec{r}) \tag{2.20}$$

2.3.3.2 Descripteur dual (f^2)

Le descripteur dual proposé par Morell et al. [17] est défini comme la dérivée mixte de l'énergie suivante, ainsi que la dérivée seconde de la densité électronique par rapport au nombre d'électrons :

$$f^{(2)}(\vec{r}) = \frac{\delta^3 E}{\delta N^2 \delta v(\vec{r})} = \left(\frac{\partial^2 \rho(\vec{r})}{\partial N^2}\right)_{v(\vec{r})} = \left(\frac{\partial f(\vec{r})}{\partial N}\right)_{v(\vec{r})} = \left(\frac{\delta \eta}{\delta v(\vec{r})}\right)_{N} \tag{2.21}$$

En pratique, le descripteur dual est calculé comme la différence des fonctions de Fukui électrophile et nucléophile en un point $\vec{r}$:

$$f^{(2)}(\vec{r}) = f^{+}(\vec{r}) - f^{-}(\vec{r}) \tag{2.22}$$

Le signe du descripteur dual permet de déterminer si un site est plus électrophile ou nucléophile : si $f^{(2)}(\vec{r}) > 0$, $f^{+}(\vec{r})$ l'emporte sur $f^{-}(\vec{r})$, le site est plus électrophile que nucléophile. Si $f^{(2)}(\vec{r}) < 0$, $f^{-}(\vec{r})$ est supérieur à $f^{+}(\vec{r})$, le site est plus nucléophile qu'électrophile.

2.3.3.3 Dureté locale (η(A)) et mollesse locale (s(A))

De la même façon que la dureté permet de savoir quel composé est le plus susceptible de réagir, la dureté locale [18] permet de savoir en quel site la réaction va se dérouler. On peut la définir par la dérivée du potentiel chimique par rapport à la densité électronique :

$$\eta(\vec{r}) = \left(\frac{\delta\mu}{\delta\rho(\vec{r})}\right)_{v(\vec{r})} \tag{2.23}$$

Il existe plusieurs façons de la calculer ; on peut notamment citer les définitions de Meneses et al. [19] ($\eta_M(\vec{r})$) et Gál et al. [20] ($\eta_G(\vec{r})$) :

$$\eta_M(\vec{r}) = \left(\frac{f^{+}(\vec{r}) + f^{-}(\vec{r})}{2}\right) + \mu f^{(2)}(\vec{r}) \tag{2.24}$$

$$\eta_G(\vec{r}) = If^{-}(\vec{r}) - Af^{+}(\vec{r}) \tag{2.25}$$

De façon similaire à la mollesse, on peut définir une mollesse locale, qui n'est cependant pas l'inverse de la dureté locale, mais le produit de la mollesse globale avec les fonctions de Fukui :

$$s^{\pm}(\vec{r}) = Sf^{\pm}(\vec{r}) \tag{2.26}$$

En intégrant la mollesse locale sur tout l'espace, on retrouve la mollesse globale (puisque les fonctions de Fukui intègrent à 1). À l'échelle d'une molécule, la mollesse locale apporte les mêmes informations que les fonctions de Fukui, mais son utilité principale est de pouvoir comparer des sites réactifs dans différentes molécules, ce qu'il n'est pas possible de faire avec les fonctions de Fukui seules.

2.3.4 Principes de réactivité

Au fil des années, des observations sur la variation des descripteurs entre les réactifs et les produits ou le long du profil réactionnel ont mené à l'établissement de divers principes de réactivité, dont quelques-uns parmi les principaux sont listés ci-dessous :

- Le principe des acides et des bases durs et mous (HSAB, « *Hard and Soft Acids and Bases* ») [21] : les composés durs réagissent préférentiellement avec les composés durs ; de même pour les composés mous.
- Le principe de dureté maximale (MHP, « *Maximum Hardness Principle* ») [22] : les molécules réagissent de sorte à maximiser leur dureté.
- Le principe d'électrophilie minimale (MEP, « *Minimum Electrophilicity Principle* ») [23] : les molécules réagissent de sorte à minimiser leur indice d'électrophilie.
- Le principe « $\Delta\mu$ *big is good* » (DMB) [24,25] : les atomes qui se lient lors d'une réaction sont ceux menant à la plus grande variation du potentiel chimique.
- Le principe de la somme de Fukui maximale [26] : les atomes qui se lient lors d'une réaction sont ceux dont la somme des fonctions de Fukui électrophile d'une part et nucléophile d'autre part est maximale.

Ces principes sont détaillés plus avant dans le chapitre 4.

2.4 Analyses de population

Les analyses de population désignent l'ensemble des méthodes visant à déterminer le nombre d'électrons de chaque atome dans une molécule (et donc leur charge) et les propriétés atomiques liées. La charge atomique n'étant pas une observable, il n'existe pas de définition unique de celle-ci, et il n'existe pas non plus de manière unique de définir un atome dans une molécule.

Il existe un grand nombre de méthodes d'analyse de population basées soit sur les orbitales moléculaires, soit sur le potentiel électrostatique, soit sur la densité électronique ; quelques-unes de ces méthodes sont présentées dans cette partie.

2.4.1 Analyses de population basées sur les orbitales moléculaires

2.4.1.1 Analyse de population de Mulliken

Historiquement, la première approche permettant de calculer théoriquement les charges atomiques est celle proposée par Mulliken en 1955 [27]. Elle repose sur les matrices densité et de recouvrement orbitalaire et permet de fractionner les électrons d'une molécule sur ses différentes parties.

On peut exprimer les coefficients $D_{\mu\nu}$ de la matrice densité $\boldsymbol{D}$ comme une somme de produits de coefficients dans une base de fonctions (μ, ν) pour les différentes orbitales i :

$$D_{\mu\nu} = \sum_i n_i C_{\mu i} C_{\nu i} \tag{2.27}$$

n_i est ici l'occupation de chaque orbitale (0 pour une orbitale virtuelle, 1 pour une orbitale à demi occupée et 2 pour une orbitale occupée).

La charge atomique de Mulliken s'exprime alors :

$$q_{Mulliken,A} = Z_A - \left[\sum_{\mu \in A} D_{\mu\mu} S_{\mu\mu} + \frac{1}{2} \sum_{\nu \neq \mu} D_{\mu\nu} S_{\mu\nu}\right] \tag{2.28}$$

où $S_{\mu\mu}$ et $S_{\mu\nu}$ sont des éléments de $\boldsymbol{S}$, la matrice de recouvrement des fonctions de base μ et ν.

Cependant, cette approche souffre de plusieurs problèmes : les charges obtenues sont très dépendantes des fonctions de base employées [28], les populations atomiques de Mulliken peuvent être négatives, ce qui n'a pas de sens physique, et les électrons sont répartis arbitrairement lorsqu'ils se trouvent dans des orbitales moléculaires placées entre deux atomes, la moitié pour l'un des atomes et l'autre moitié pour le second, indépendamment de la différence d'électronégativité entre les deux éléments.

2.4.1.2 Analyse de population naturelle

Les méthodes d'analyse de population naturelle (NPA, « *Natural Population Analysis* ») sont une alternative à la méthode de Mulliken ; elles permettent de s'affranchir d'une partie des inconvénients et d'améliorer la précision des résultats.

Parmi ces méthodes, la plus utilisée est l'approche NBO (« *Natural Bonding Orbitals* ») de Weinhold et al. [29]. Elle repose sur une suite de transformations d'orbitales à partir des fonctions de base, afin d'obtenir des orbitales localisées :

1. Le jeu des fonctions de base est transformé en un ensemble d'orbitales atomiques naturelles (NAO, « *Natural Atomic Orbitals* »), qui sont l'ensemble des orbitales atomiques (localisées sur un seul centre) orthogonales dont l'occupation des orbitales de cœur et de valence est maximale pour la fonction d'onde considérée ;
2. L'ensemble des orbitales hybrides naturelles (NHO, « *Natural Hybrid Orbitals* ») est formé par combinaison linéaire des NAO sur chaque centre : les NHO sont également un ensemble d'orbitales orthogonales ;
3. Enfin, les orbitales naturelles de liaison (NBO) sont obtenues à partir des NHO seules ou combinées : ce sont l'ensemble des orbitales à 1, 2 ou (plus rarement) 3 centres construites de sorte à ce que les $\frac{N}{2}$ premières NBO d'un système de N électrons à couches fermées décrivent le mieux possible la structure de Lewis de la molécule, avec une occupation très proche ou égale à 2 pour toutes.

On distingue plusieurs types de NBO (avec leur notation en NBO entre parenthèses) :

- Les NBO de type « cœur » (CR), constituées d'une seule NHO de cœur,
- Les NBO de type « paire libre » (LP), constituées d'une seule NHO de valence occupée,
- Les NBO de type « liante » (BD) et « antiliante » (BD*), constituées d'une combinaison linéaire de 2 ou plus NHO de valence respectivement occupées et inoccupées, de sorte que la somme des carrés des coefficients soit égale à 1 ;
- Les NBO de type « lacune » (LV), constituées d'une seule NHO de valence inoccupée,
- Les NBO de type « Rydberg » (RY), constituées d'une seule NHO de Rydberg inoccupée : ces orbitales sont une conséquence du passage des fonctions de base aux NAO et sont les orbitales supplémentaires nécessaires à l'obtention d'un ensemble orthogonal. Elles permettent entre autres de décrire les effets de délocalisation.

Les coefficients des NHO composant chaque NBO liante permettent de savoir si cette dernière a un caractère plus ou moins ionique : si les deux coefficients sont de valeurs proches, la NBO a un caractère liant ; s'ils sont très différents, la NBO a un caractère ionique.

La Figure 2.6 ci-dessous permet de comparer une orbitale moléculaire (délocalisée) et deux NBO (localisées).

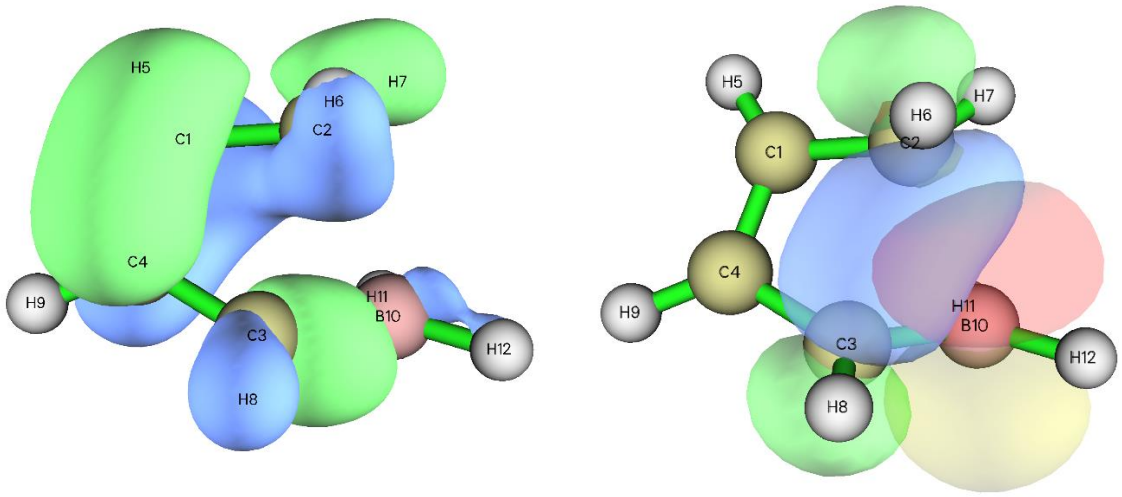


Figure 2.6 – Représentation d'orbitales moléculaires et localisées du borycyclobutène

Gauche : HOMO / Droite : deux NBO superposées de l'état de transition de l'ouverture thermique du borylcyclobutène. Les deux NBO représentées correspondent à la NBO liante des atomes de carbone 2 et 3 (en bleu et vert) et à la NBO vacante de l'atome de bore (en jaune et rouge). Les orbitales sont délimitées par une isosurface de 0,05 u.a. Image obtenue avec le logiciel Multiwfn.

On peut voir que cette approche permet de retrouver des concepts familiers aux chimistes tels qu'une liaison, une lacune électronique, une paire libre… sans la difficulté liée à l'interprétation d'orbitales délocalisées sur la molécule toute entière.

Dans cette approche, les charges atomiques sont obtenues grâce aux éléments diagonaux de la matrice densité $\boldsymbol{D}$ formée en utilisant les NAO θ_i :

$$q_{NPA} = Z_A - \langle \theta_i | \widehat{D} | \theta_i \rangle \tag{2.29}$$

Il est possible de calculer par la théorie des perturbations des énergies d'interaction entre les différentes NBO, notamment l'énergie de stabilisation donneur-accepteur $E(2)$:

$$E(2) = -n_{\varphi_i} \frac{\langle \varphi_i | \hat{F} | \varphi_j^* \rangle^2}{\varepsilon_{\varphi_i} - \varepsilon_{\varphi_j^*}} \tag{2.30}$$

où φ_i et φ_j^* sont deux NBO respectivement occupée et inoccupée, n_{φ_i} est l'occupation de la NBO occupée φ_i et $\langle \varphi_i | \hat{F} | \varphi_j^* \rangle^2$ est le carré de l'élément de la matrice de Fock relatif aux NBO φ_i et φ_j^*. On obtient ainsi $\left(\frac{N}{2} - 1\right)^2$ termes $E(2)$, qui sont toujours stabilisants.

On peut également aller plus loin en calculant l'énergie d'interaction $E(4)$ entre deux NBO occupées :

$$E(4) = \left(F_{I,I}^{NLMO} - F_{I,I}^{PNLMO/2}\right) + \left(F_{J,J}^{NLMO} - F_{J,J}^{PNLMO/2}\right) \tag{2.31}$$

Celle-ci est approximée par une somme d'éléments diagonaux de la matrice de Fock relatifs à des NLMO (« *Natural Localised Molecular Orbitals* »), des orbitales apparentées aux NBO mais étant plus délocalisées de sorte à être totalement occupées ou inoccupées, et à des PNLMO (des NLMO « partielles » non-orthogonalisées). On obtient ainsi $\left(\frac{N}{2} - 1\right)^2$ termes $E(4)$ (déstabilisants), tout comme pour les termes $E(2)$. Nous avons utilisé ces termes $E(2)$ et $E(4)$ comme descripteurs de réactivité.

2.4.2 Analyses de population basées sur le potentiel électrostatique

Pour définir une charge atomique, il peut sembler naturel de s'intéresser au potentiel électrostatique moléculaire [30] appelé ESP, « *ElectroStatic Potential* » ou MEP, « *Molecular Electrostatic Potential* ». Ce potentiel est défini comme :

$$ESP(\vec{r}) = \sum_A \frac{Z_A}{\|\overrightarrow{R_A} - \vec{r}\|} - \int \frac{\rho(\vec{r}')}{\|\vec{r}' - \vec{r}\|} d^3r' \tag{2.32}$$

avec $\overrightarrow{R_A}$ la position de la charge nucléaire Z_A.

Nous avons notamment utilisé la méthode ChElPG (« *Charges from Electrostatic Potentials using a Grid-based method* ») [31] proposée par Breneman et Wiberg en 1990. Dans ce cadre, les charges sont celles permettant de retrouver le potentiel électrostatique, et sont obtenues par ajustement à l'ESP : celui-ci est calculé en chaque point d'une grille cubique couvrant l'ensemble du système.

Cette méthode est néanmoins coûteuse pour des systèmes de grande taille, pour lesquels il est plus intéressant d'utiliser une grille dite de Lebedev (grille sphérique centrée sur les atomes).

2.4.3 Analyses de population basées sur la densité électronique

2.4.3.1 Analyse de population d'Hirshfeld

Dans l'approche d'Hirshfeld [32], la densité électronique est répartie entre tous les atomes en chaque point : la charge atomique est donc vue comme étant diffuse dans tout l'espace.

Pour ce faire, la méthode passe par la construction d'une « promolécule », où tous les atomes sont placés comme dans la structure du système étudié, mais ne sont pas considérés comme liés. La densité électronique résultante est la somme des densités électroniques des atomes isolés :

$$\rho_{promolécule}(\vec{r}) = \sum_{A} \rho_A^{atomique}(\vec{r}) \tag{2.33}$$

De cette façon, on peut simplement définir des poids $w_A(\vec{r})$ sommant à 1 en chaque point de l'espace et représentant la fraction que représente la densité atomique $\rho_A^{atomique}$ dans la densité électronique promoléculaire $\rho_{promolécule}$:

$$w_A(\vec{r}) = \frac{\rho_A^{atomique}(\vec{r})}{\rho_{promolécule}(\vec{r})} \tag{2.34}$$

Ces poids servent ensuite à partitionner la densité électronique moléculaire réelle $\rho(\vec{r})$ en chaque point et à calculer la charge atomique Hirshfeld :

$$q_{Hirshfeld} = Z_A - \int w_A(\vec{r})\rho(\vec{r})d^3r \tag{2.35}$$

Les charges obtenues de cette façon sont généralement beaucoup plus faibles en valeur absolue que celles des autres méthodes [33].

L'inconvénient de la méthode est que les charges obtenues dépendent fortement de la densité promoléculaire initiale. Habituellement, la densité atomique des atomes neutres et isolés est utilisée, ce qui pose des problèmes de choix lors de l'étude de molécules chargées.

La méthode itérative Hirshfeld-I proposée par Bultinck [33] permet de contourner ce problème en utilisant un cycle auto-cohérent pour déterminer la densité électronique promoléculaire initiale. La densité électronique promoléculaire est recalculée avec des populations non-entières par interpolation linéaire entre les densités aux populations entières immédiatement supérieures et inférieures à chaque itération, et le cycle s'arrête lorsque la variation des populations atomiques devient inférieure au critère de convergence.

2.4.3.2 Théorie quantique AIM

La théorie quantique AIM [34,35] (QTAIM, « *Quantum Theory of Atoms In Molecules* », théorie quantique des atomes dans les molécules) a été développée par Bader et propose une méthode de partition non arbitraire de la densité électronique en bassins atomiques, chaque bassin correspondant à un atome. Elle se base sur une analyse topologique des gradients de la densité électronique, et est valide en tout point de la PES.

Gradients et champ de gradients

Le gradient de la densité électronique est défini en chaque point $\vec{r}$ de l'espace par un vecteur $\vec{\nabla}\rho(\vec{r})$ pointant dans le sens de l'augmentation maximale de $\rho(\vec{r})$. On peut tracer des chemins de gradient à partir d'un point $\vec{r}$ en avançant d'un pas infinitésimal dans la direction de $\vec{\nabla}\rho(\vec{r})$, en recalculant le gradient en ce nouveau point, et en répétant le processus. Ainsi, les vecteurs gradient sont en tout point tangents aux chemins de gradient. Ces chemins aboutissent en des maxima locaux de la densité. L'ensemble des chemins de gradient est appelé champ de vecteur gradient.

Points critiques de la densité électronique

Sur ces chemins de gradient, on peut définir des points remarquables appelés points critiques là où le gradient s'annule :

$$\vec{\nabla}\rho(\vec{r}) = \vec{0} \tag{2.36}$$

Les caractéristiques de ce point critique peuvent être obtenues à partir de la matrice Hessienne déjà évoquée en section 2.1 (il s'agit de la matrice des dérivées secondes pour toutes les variables d'une fonction à plusieurs variables). Pour la densité électronique en coordonnées cartésiennes, on a :

$$\mathcal{H}_\rho = \begin{pmatrix} \frac{\partial^2\rho(\vec{r})}{\partial x^2} & \frac{\partial^2\rho(\vec{r})}{\partial x\partial y} & \frac{\partial^2\rho(\vec{r})}{\partial x\partial z} \\ \frac{\partial^2\rho(\vec{r})}{\partial y\partial x} & \frac{\partial^2\rho(\vec{r})}{\partial y^2} & \frac{\partial^2\rho(\vec{r})}{\partial y\partial z} \\ \frac{\partial^2\rho(\vec{r})}{\partial z\partial x} & \frac{\partial^2\rho(\vec{r})}{\partial z\partial y} & \frac{\partial^2\rho(\vec{r})}{\partial z^2} \end{pmatrix} \tag{2.37}$$

En diagonalisant $\mathcal{H}_\rho$ dans une nouvelle base orthonormée (x', y', z'), on obtient :

$$\mathcal{H}_\rho = \begin{pmatrix} \frac{\partial^2\rho(\vec{r})}{\partial x'^2} & 0 & 0 \\ 0 & \frac{\partial^2\rho(\vec{r})}{\partial y'^2} & 0 \\ 0 & 0 & \frac{\partial^2\rho(\vec{r})}{\partial z'^2} \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \tag{2.38}$$

Cela équivaut à une rotation du système et permet d'obtenir les valeurs propres de la Hessienne λ_1, λ_2 et λ_3 correspondant aux courbures de $\rho(\vec{r})$ au point critique, qui permettent de catégoriser les types de points critiques. On utilise pour cela le rang, c'est-à-dire le nombre de valeurs propres de $\mathcal{H}_\rho$ non nulles au point critique, et la signature, qui est la somme algébrique des signes de ces valeurs propres.

On obtient ainsi quatre catégories de points critiques de la densité électronique avec trois valeurs propres non nulles :

Tableau 1 - Classement des différents points critiques

Type (rang, signature)	Catégorie de point critique	Courbures
(3,-3)	Attracteur nucléaire ou non nucléaire	3 négatives : maximum local
(3,-1)	Point critique de liaison	2 négatives, 1 positive
(3,+1)	Point critique d'anneau	1 négative, 2 positives
(3,+3)	Point critique de cage	3 positives : minimum local

Ainsi, les points critiques de type (3,-3), dits attracteurs, sont des maxima locaux de la densité électronique et correspondent souvent aux noyaux (on les désigne NACP, « *Nuclear Attractor Critical*

Point »), bien que ce ne soit pas toujours le cas. On parle dans ce cas de NNACP [36] (« *Non-Nuclear Attractor Critical Point* »).

Les points critiques de type (3,-1) sont appelés points critiques de liaison (BCP, « *Bond Critical Point* »). Ils correspondent à un minimum de densité sur le chemin de vecteur gradient entre deux atomes liés et à un maximum de densité dans le plan normal à ce chemin. Ils permettent de caractériser la nature des liaisons : en calculant la densité électronique et son Laplacien $\nabla^2\rho(\vec{r})$ (la somme des valeurs propres de la Hessienne), on peut savoir si une liaison est plutôt covalente ($\rho(\vec{r})$ élevée, $\nabla^2\rho(\vec{r}) < 0$) ou ionique ($\rho(\vec{r})$ faible, $\nabla^2\rho(\vec{r}) > 0$). Un BCP se trouve donc toujours entre deux NACP, et l'union d'un BCP et des deux chemins de gradient le reliant à ces deux NACP est appelée chemin de liaison ; ce sont les liaisons chimiques du point de vue QTAIM.

Les points critiques de type (3,+1) dits d'anneau (RCP, « *Ring Critical Point* ») sont caractéristiques des cycles (aromatiques, entre autres). Ils correspondent à un minimum de densité électronique au niveau du centre du cycle et à un maximum dans la direction perpendiculaire à celui-ci. Les RCP se trouvent toujours entre plusieurs BCP. Enfin, les points critiques (3,+3) dits de cage (CCP, « *Cage Critical Point* ») sont beaucoup plus rares puisqu'ils doivent se situer entre plusieurs RCP, donc entre plusieurs cycles.

En QTAIM, l'ensemble des points critiques forme le graphe moléculaire, qui représente la connectivité topologique des atomes.

Bassins atomiques et atomes QTAIM

L'ensemble des chemins de gradient aboutissant au même attracteur définit une région de l'espace appelée bassin atomique Ω_A. Dans la théorie QTAIM, un atome est donc simplement l'union d'un bassin atomique et de l'attracteur associé. En quelque sorte, si l'on veut comparer la théorie QTAIM à la méthode d'Hirshfeld, en QTAIM, la fonction $w_A(\vec{r})$ vaut 1 si l'on se trouve dans Ω_A et 0 partout ailleurs.

Cependant, tous les chemins de gradient n'aboutissent pas à un attracteur : certains se terminent à un BCP et permettent ainsi de définir la surface interatomique (IAS, « *InterAtomic Surface* ») qui sépare

deux bassins atomiques. À chaque point $\vec{r}$ de ces IAS, le flux de gradient perpendiculaire à la surface interatomique est nul :

$$\vec{\nabla}\rho(\vec{r}) \cdot \vec{n}(\vec{r}) = 0 \tag{2.39}$$

avec $\vec{n}(\vec{r})$ un vecteur normal à l'IAS en $\vec{r}$.

L'espace est donc partitionné en différents bassins ne se recouvrant pas et définis de façon unique et non arbitraire. En théorie, ces bassins s'étendent jusqu'à l'infini (puisque les chemins de gradient commencent à l'infini). La Figure 2.7 ci-dessous montre un exemple de bassins atomiques pour un système modèle.

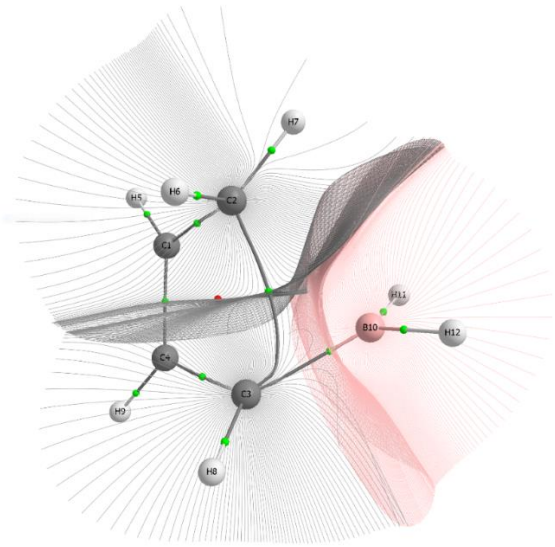


Figure 2.7 – Représentation du graphe moléculaire et d'une partie des bassins atomiques et des chemins de gradient de l'état de transition du borylcyclobutène lors de l'ouverture thermique du cycle.

Les points critiques de liaison sont représentés en vert et le point critique d'anneau en rouge. Image obtenue avec le logiciel AIMAll.

<u>*Propriétés atomiques*</u>

Grâce à cette définition des bassins atomiques, il est possible d'obtenir des propriétés atomiques en intégrant des fonctions dépendant de $\vec{r}$ sur le volume de chaque bassin :

$$P(A) = \int_{\Omega_A} p(\vec{r}) d^3r \tag{2.40}$$

Dans ce cadre, la population électronique d'un atome N_{Ω_A} est obtenue en intégrant la densité électronique sur le bassin Ω_A :

$$N_{\Omega_A} = \int_{\Omega_A} \rho(\vec{r}) d^3 r \tag{2.41}$$

Il s'ensuit que la charge QTAIM est calculée par :

$$q_{QTAIM} = Z_A - N_{\Omega_A} \tag{2.42}$$

Par cette méthode, on peut obtenir des valeurs atomiques (« condensées ») de descripteurs locaux de CDFT, comme par exemple une fonction de Fukui atomique :

$$f^+(A) = \int_{\Omega_A} f^+(\vec{r}) d^3 r \tag{2.43}$$

Cela permet de pouvoir comparer directement différents sites de façon plus parlante que les valeurs locales. Il est également possible de sommer ces descripteurs atomiques pour pouvoir comparer des groupes fonctionnels.

Indices de localisation et de délocalisation électroniques

Il existe deux autres grandeurs permettant de caractériser les propriétés électroniques des atomes dans la théorie QTAIM : les indices de localisation électronique (LI) et de délocalisation électronique (DI) [37]. Ceux-ci sont calculés à l'aide des matrices de densité réduite (RDM, « *Reduced Density Matrices* ») : en utilisant la convention de McWeeny [38], les RDM du premier et second ordre ($\boldsymbol{\rho_1}$ et $\boldsymbol{\rho_2}$) sont définies par :

$$\boldsymbol{\rho_1}(\vec{r}_1; \vec{r}_1') = N \int \psi_{el}(\vec{x}_1, \dots \vec{x}_N) \psi_{el}^*(\vec{x}_1', \dots \vec{x}_N') d\sigma_1 d\sigma_1' d^4 x_2 \dots d^4 x_N \tag{2.44}$$

$$\boldsymbol{\rho_2}(\vec{r}_1, \vec{r}_2; \vec{r}_1', \vec{r}_2') = N(N-1) \int \psi_{el}(\vec{x}_1, \dots \vec{x}_N) \psi_{el}^*(\vec{x}_1', \dots \vec{x}_N') d\sigma_1 d\sigma_2 d\sigma_1' d\sigma_2' d^4 x_3 \dots d^4 x_N \tag{2.45}$$

où les vecteurs $\vec{x}_i$ comprennent la position des électrons $\vec{r}_i$ et leur spin σ_i : $\vec{x}_i = (\vec{r}_i, \sigma_i)$.

À partir des éléments diagonaux de la RDM du second ordre, il est possible de définir la fonction de paire du second ordre $\rho_2(\vec{r}_1, \vec{r}_2)$ qui correspond à la probabilité de trouver une paire d'électrons en $d^3\vec{r}_1$ et $d^3\vec{r}_2$.

Cette fonction peut se partitionner naturellement en trois termes : un terme coulombien, un terme d'échange et un terme de corrélation. Le premier terme s'exprime par $\rho(\vec{r}_1)\rho(\vec{r}_2)$ et les deux autres peuvent être rassemblés en une densité d'échange-corrélation $\rho_{xc}(\vec{r}_1,\vec{r}_2)$ égale à $\rho(\vec{r}_1)\rho(\vec{r}_2) - \rho_2(\vec{r}_1,\vec{r}_2)$. La moyenne de cette densité sur une paire de bassins donne le nombre de paires d'électrons partagés entre les deux atomes en question.

On peut alors définir le LI et le DI : le LI correspond au nombre moyen de paires d'électrons situées sur un atome A :

$$LI(A,A) = \int_{\Omega_A} \rho_{xc}(\vec{r}_1,\vec{r}_2) d^3r_1 d^3r_2 \tag{2.46}$$

Le DI correspond au nombre de paires d'électrons délocalisées entre deux atomes A et B, ce qui revient à un indice de liaison QTAIM :

$$DI(A,B) = 2\int_{\Omega_A}\int_{\Omega_B} \rho_{xc}(\vec{r}_1,\vec{r}_2) d^3r_1 d^3r_2 \tag{2.47}$$

La somme des LI et des DI d'une molécule entière redonne le nombre d'électrons total :

$$N = \sum_A LI(A,A) + \frac{1}{2}\sum_A \sum_{B\neq A} DI(A,B) \tag{2.48}$$

Décomposition énergétique IQA

La méthode des atomes quantiques en interaction (IQA, « *Interacting Quantum Atoms* ») est une méthode de décomposition énergétique proposée par Blanco et al. en 2005 [39]. Elle peut en principe être utilisée avec n'importe quelle méthode de partition spatiale (comme celle d'Hirshfeld vue précédemment), mais seule son application à la théorie QTAIM sera examinée ici.

La méthode IQA utilise donc la partition QTAIM de l'espace pour décomposer de façon exacte l'énergie moléculaire totale en une somme de termes définis de façon non arbitraire. Ces termes sont divisés en deux catégories : les termes intra-atomiques et les termes inter-atomiques. L'énergie totale s'écrit donc comme deux sommes :

$$E_{totale} = \sum_{A} E_{intra}(A) + \frac{1}{2}\sum_{A}\sum_{B\neq A} E_{inter}(A,B) \tag{2.49}$$

où A et B sont deux atomes du système.

On peut continuer à décomposer chacun des termes intra-atomiques et inter-atomiques. Pour le terme intra-atomique, on a :

$$E_{intra}(A) = T(A) + V_{en}(A) + V_{ee}(A,A) \tag{2.50}$$

où $T(A)$ est l'énergie cinétique des électrons dans le bassin atomique de A, $V_{en}(A)$ est l'énergie d'interaction électrostatique entre les électrons du bassin et le noyau de A, et $V_{ee}(A)$ est l'énergie d'interaction électrostatique entre les électrons dans le bassin. $E_{intra}(A)$ est toujours un terme stabilisant.

$T(A)$ est calculée comme :

$$T(A) = -\frac{1}{2}\int_{\Omega_A} \nabla^2_{\vec{r}_1} \boldsymbol{\rho_1}(\vec{r}_1;\vec{r}_1')|_{\vec{r}_1=\vec{r}_1'} d^3r_1 \tag{2.51}$$

Tandis que $V_{en}(A)$ s'exprime par :

$$V_{en}(A) = -\int_{\Omega_A} \frac{Z_A\rho(\vec{r})}{\|\vec{r}-\vec{r}_A\|} d^3r \tag{2.52}$$

où $\vec{r}_A$ est la position du noyau dans le bassin A.

Quant à $V_{ee}(A,A)$, elle peut à nouveau se décomposer en deux termes comme proposé par Pendás et al. en 2004 [40] :

$$V_{ee}(A,A) = V_{ee,cl}(A,A) + V_{ee,xc}(A,A) \tag{2.53}$$

$V_{ee,cl}(A,A)$ représente la partie coulombienne (classique) de l'interaction biélectronique :

$$V_{ee,cl}(A,A) = \frac{1}{2}\int_{\Omega_A} \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d^3r_1 d^3r_2 \tag{2.54}$$

où r_{12} est la distance inter-électrons.

$V_{ee,xc}(A,A)$ correspond à la contribution d'échange-corrélation de cette même interaction. Pour une fonction d'onde monodéterminantale, on a :

$$V_{ee,xc}(A,A) = -\sum_i \sum_{j\neq i} \int_{\Omega_A} \frac{\varphi_i(\vec{r}_1)\varphi_j(\vec{r}_1)\varphi_i(\vec{r}_2)\varphi_j(\vec{r}_2)}{r_{12}} d^3r_1 d^3r_2 \tag{2.55}$$

avec les différentes orbitales moléculaires φ_i et φ_j.

Similairement à l'énergie intra-atomique, on peut décomposer l'énergie inter-atomique $E_{inter}(A,B)$ selon :

$$E_{inter}(A,B) = V_{nn}(A,B) + V_{en}(A,B) + V_{ne}(A,B) + V_{ee}(A,B) \tag{2.56}$$

$V_{nn}(A,B)$ est l'énergie de répulsion électrostatique noyau-noyau, $V_{en}(A,B)$ correspond à l'énergie d'interaction des électrons de A avec le noyau de B tandis que $V_{ne}(A,B)$ correspond à l'interaction opposée (noyau de A avec électrons de B), et $V_{ee}(A,B)$ est l'énergie d'interaction électron-électron.

$V_{nn}(A,B)$ est simplement définie par :

$$V_{nn}(A,B) = \frac{Z_A Z_B}{\|\vec{R}_A - \vec{R}_B\|} \tag{2.57}$$

$V_{en}(A,B)$ et $V_{ne}(A,B)$ sont définies très similairement ; pour $V_{en}(A,B)$, on a :

$$V_{en}(A,B) = -\int_{\Omega_A} \frac{Z_B \rho(\vec{r})}{\|\vec{r} - \vec{R}_B\|} d^3r \tag{2.58}$$

Pour obtenir $V_{ne}(A,B)$, il suffit d'échanger les indices A et B.

Enfin, $V_{ee}(A,B)$ peut se décomposer de la même façon que le terme intra-atomique correspondant :

$$V_{ee}(A,B) = V_{ee,cl}(A,B) + V_{ee,xc}(A,B) \tag{2.59}$$

avec $V_{ee,cl}(A,B)$ l'énergie d'interaction coulombienne entre A et B et $V_{ee,xc}(A,B)$ l'énergie d'échange-corrélation. $V_{ee,cl}(A,B)$ s'écrit :

$$V_{ee,cl}(A,B) = \int_{\Omega_A} \int_{\Omega_B} \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d^3r_1 d^3r_2 \tag{2.60}$$

alors que $V_{ee,xc}(A,B)$ s'exprime ainsi pour une fonction d'onde monodéterminantale :

$$V_{ee,xc}(A,B) = -2\sum_i \sum_j \int_{\Omega_A} \int_{\Omega_B} \frac{\varphi_i(\vec{r}_1)\varphi_j(\vec{r}_1)\varphi_i(\vec{r}_2)\varphi_j(\vec{r}_2)}{r_{12}} d^3r_1 d^3r_2 \tag{2.61}$$

Ce dernier terme est directement relié au DI défini précédemment, on a ainsi :

$$V_{ee,xc}(A,B) \approx -\frac{DI(A,B)}{2R_{\mathrm{AB}}} \tag{2.62}$$

On peut enfin regrouper les différents termes inter-atomiques d'interaction coulombienne :

$$E_{inter}(A,B) = V_{nn}(A,B) + V_{en}(A,B) + V_{ne}(A,B) + V_{ee,cl}(A,B) + V_{ee,xc}(A,B) \tag{2.63}$$

$$= V_{cl}(A,B) + V_{ee,xc}(A,B) \tag{2.64}$$

Le terme d'échange-corrélation représente ainsi la covalence dans l'interaction $A - B$.

La méthode IQA permet ainsi de décomposer l'énergie d'un système à M atomes en M termes intra-atomiques et $\frac{M(M-1)}{2}$ termes inter-atomiques, soit $\frac{M(M+1)}{2}$ termes au total.

Il faut noter que l'application de la méthode IQA est en pratique différente selon que les calculs se basent sur la fonction d'onde ou la densité électronique, les orbitales et définitions des termes n'étant pas identiques (par exemple, dans le cadre de l'approche de Kohn-Sham en DFT, la RDM du second ordre ne peut être qu'approximée par celle du système fictif sans interactions).

En DFT, en particulier, lorsque des fonctionnelles LDA, GGA ou méta-GGA (voir chapitre 1, section 1.5.4) sont utilisées, la définition de l'échange pose problème lors de la partition de l'énergie. En effet, dans le cas des fonctionnelles GGA, l'échange $E_X^{GGA}[\rho]$ est purement local d'un point de vue mathématique (il ne dépend que de $\rho(r)$ et $\nabla\rho(r)$), ce qui conduit à le classer entièrement dans le terme intra-atomique $V_{ee,xc}(A,A)$; il n'y a alors aucun terme d'échange dans $V_{ee,xc}(A,B)$! Une solution peut être d'utiliser l'échange exact non-local $E_{XX}[\{\phi_i^{KS}\}]$ au lieu de l'échange DFT et de modifier en conséquence l'expression de $V_{ee,xc}(A,A)$ comme proposé par Maxwell et al. [41] ou Francisco et al. [42]. On peut également mentionner l'extension d'IQA à la méthode MP2 par Popelier et al. [43] et CC par Rocha-Rinza et al. [44].

Depuis l'article fondateur de la méthode IQA, ses applications se sont multipliées ces dernières années, en particulier pour étudier la corrélation électronique [45], la relation entre $V_{ee,xc}(A,B)$ et le DI [46], la nature de liaisons chimiques [47,48], les interactions non-covalentes dont les liaisons hydrogène [49] et halogène [50–52], les liaisons métal-métal [53], la réactivité [54] et notamment la catalyse [55], les états excités [56], les effets stériques [57] et enfin l'utilisation de données IQA en *machine learning* pour la conception de nouveaux champs de force [58].

2.5 Méthode REG

2.5.1 Objectif

La méthode REG (*Relative Energy Gradient* en anglais, gradient d'énergie relative) est une méthode d'analyse des descripteurs le long d'un profil d'énergie (comme un profil de réaction). Elle a été développée à partir de 2017 [59] dans le groupe du Pr Popelier à Manchester, d'abord au sein du logiciel ANANKE puis comme une librairie Python indépendante (REG.py) [60].

Cette méthode vise à comprendre l'origine des variations de l'énergie le long du profil en établissant des corrélations entre les descripteurs calculés et l'énergie totale du système, ce qui permet de trouver les descripteurs les plus pertinents. En effet, si l'on peut trouver des descripteurs variant comme l'énergie totale, on peut considérablement simplifier l'analyse du système en se ramenant à un faible nombre de descripteurs pertinents. On a déjà vu précédemment qu'un grand nombre de descripteurs pouvaient être calculés même pour des systèmes de taille réduite, et ce nombre augmente quadratiquement lorsque la taille du système augmente. Une méthode permettant de trouver automatiquement un nombre réduit de descripteurs d'intérêt parmi l'intégralité des descripteurs calculés est donc très intéressante. Ainsi, on évite de se reposer uniquement sur l'intuition chimique et de risquer de passer à côté de descripteurs pertinents.

2.5.2 Principe

Dans la méthode REG, les descripteurs sont évalués à différents points du profil réactionnel. Ce profil est découpé en segments afin d'avoir des variations monotones de l'énergie sur chacun des segments, ce qui permet d'avoir par la suite des régressions linéaires de bonne qualité. Cela permet aussi de déterminer les descripteurs responsables de différents évènements ou phénomènes successifs, par exemple la déstabilisation d'un système au cours d'une réaction en allant des réactifs jusqu'à l'état de transition, puis sa stabilisation après l'état de transition. La Figure 2.8 ci-dessous représente un exemple de découpage de profil simple en utilisant la force de réaction définie dans la section 2.1.

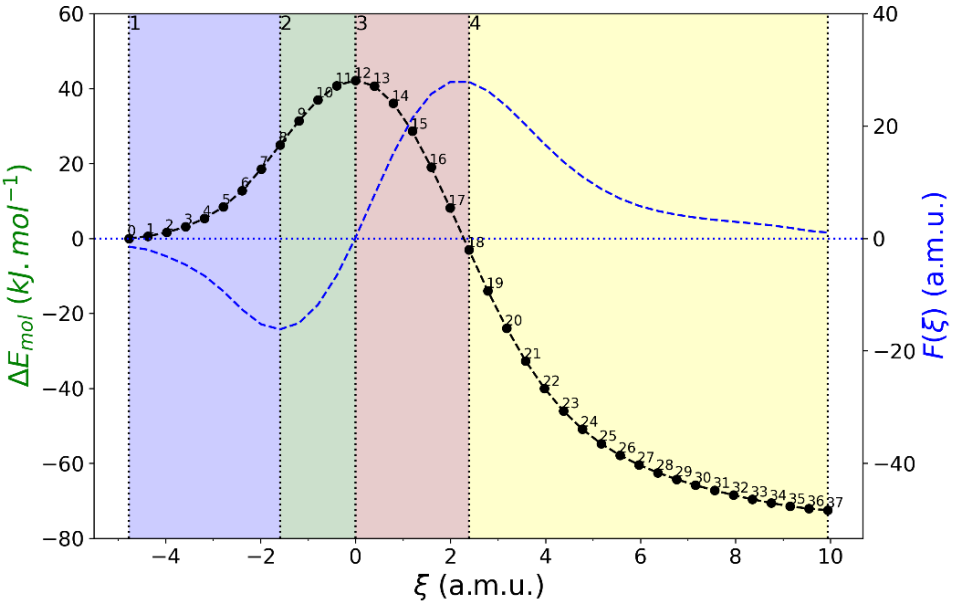


Figure 2.8 - Exemple de découpage du profil réactionnel de l'ouverture thermique du borylcyclobutène.

L'énergie relative (ΔE_{mol}) est représentée en noir et la force de réaction (F) en bleu.

Il faut prêter une attention particulière au découpage : en effet, des segments trop petits peuvent conduire à des R^2 artificiellement élevés tandis que de grands segments les abaisseront plutôt.

Il est possible de visualiser la répartition de ces R^2 grâce à des diagrammes dits « en violon », qui représentent la densité des termes de R^2. Un exemple d'un tel diagramme est donné en Figure 2.9 ci-dessous.

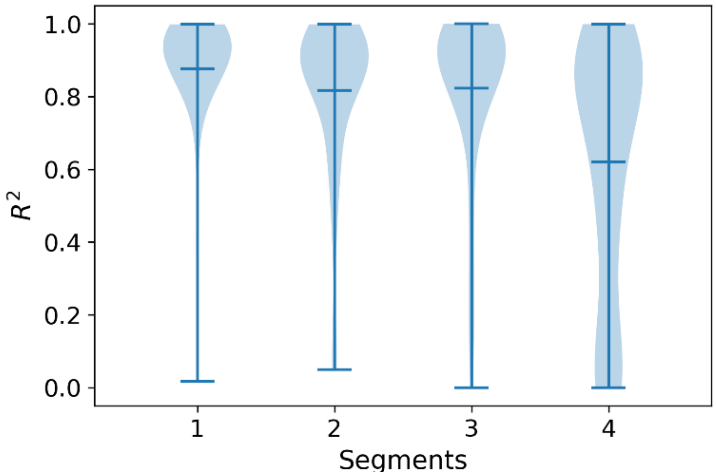


Figure 2.9 – Diagrammes en violon représentant la densité des coefficients de détermination des régressions des termes d'une analyse REG sur chaque segment

La barre horizontale indique la moyenne des R². Plus le lobe (et conséquemment, la moyenne) du diagramme est décalé vers 1, plus l'ensemble des régressions sont de meilleure qualité.

L'analyse REG est ensuite conduite sur chaque segment : pour chaque descripteur, une régression linéaire est établie entre la variation du descripteur en question et la variation de l'énergie totale (équation 2.65). Cela revient à comparer les gradients des descripteurs (ΔD_i) et de l'énergie (ΔE_{mol}), d'où le nom de la méthode.

$$\Delta D_i(\xi) = m_i^{REG} \cdot \Delta E_{mol}(\xi) + c_i \tag{2.65}$$

Deux paramètres importants caractérisent ces régressions :

- La pente, ou coefficient REG (m_i^{REG}), qui va quantifier la variation relative du descripteur et de l'énergie totale. Son signe indique si les deux grandeurs varient de la même façon ou à l'opposé l'une de l'autre, et sa valeur indique si le descripteur varie plus ou moins vite que l'énergie (cf. Figure 2.10 ci-dessous).
- Le coefficient de détermination (R^2), qui va évaluer la correspondance des profils du descripteur et de l'énergie totale. Plus ce coefficient est proche de l'unité, plus les profils se ressemblent.

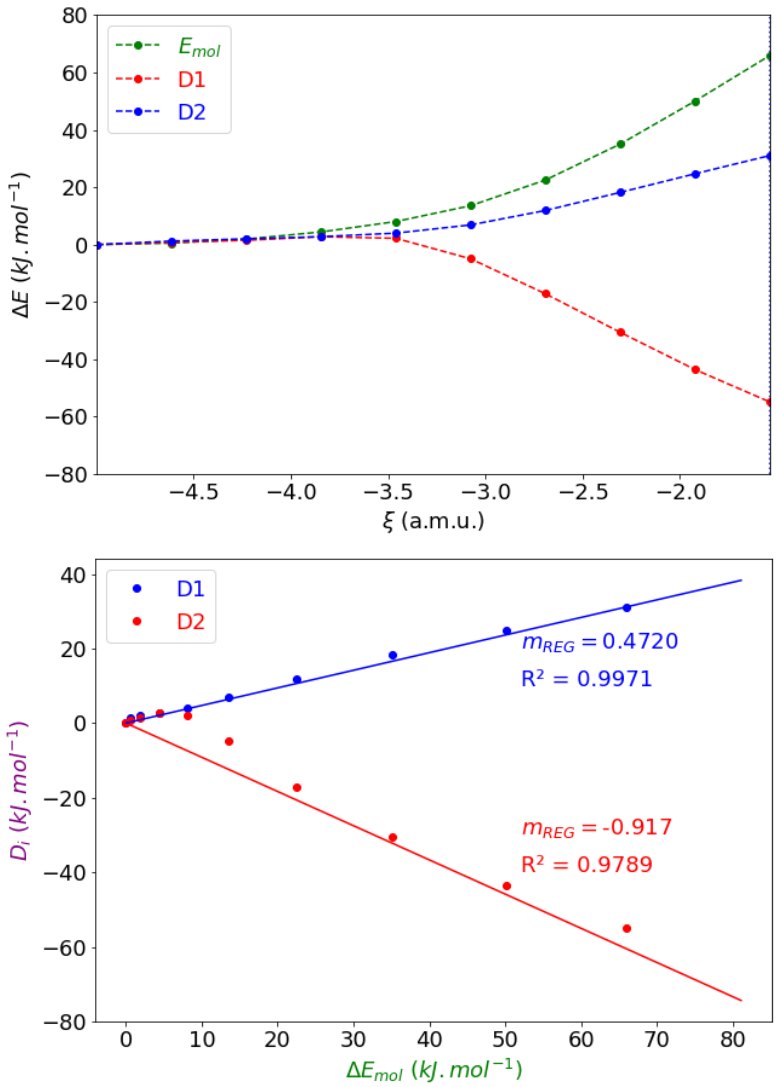


Figure 2.10 - Évolution de l'énergie totale et de deux termes IQA (haut) et régressions linéaires de ces termes par rapport à l'énergie totale (bas).

On peut observer que le descripteur D1 (en bleu) croît, comme l'énergie totale, et a donc un coefficient REG positif, tandis que le descripteur D2 (en rouge) décroît et a un coefficient REG négatif.

Il va ensuite de soi que ces deux paramètres doivent être simultanément élevés pour que le descripteur soit considéré comme pertinent ; un descripteur dont la régression présenterait un coefficient REG élevé mais un faible coefficient de détermination aurait une grande variation mais varierait bien différemment de l'énergie, tandis qu'un descripteur dont la régression comporterait un coefficient REG faible et un coefficient de détermination proche de 1 varierait très faiblement mais de la même façon que l'énergie totale. On ne retiendra donc par la suite que des descripteurs pour lesquels on obtient à la fois un coefficient REG et un coefficient de détermination élevés.

La Figure 2.11 ci-dessous est un exemple de visualisation des résultats d'une analyse REG, en traçant les termes selon leur coefficient de détermination (en abscisse, ici restreint aux R^2 supérieurs à 0,90) et leur coefficient REG (en ordonnée) : les termes les plus pertinents ont les coefficients REG les plus élevés dans un intervalle de R^2 élevé (typiquement, $R^2 > 0.9$).

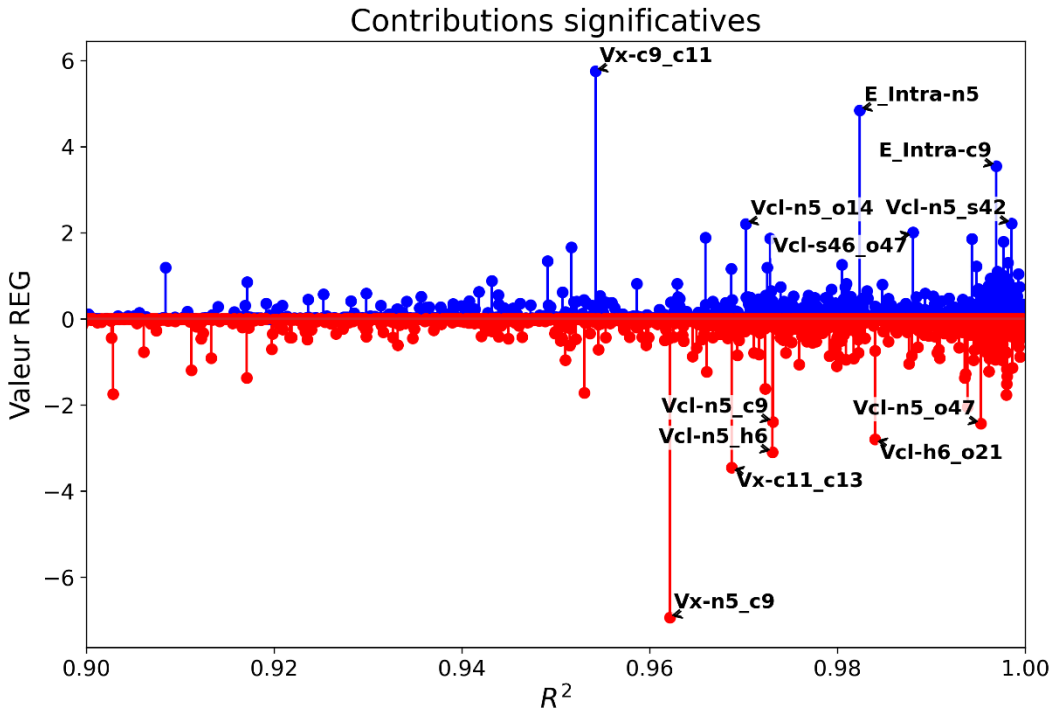


Figure 2.11 – Graphique à tiges permettant la visualisation d'une analyse REG-IQA restreinte aux R^2 supérieurs à 0,90, pour la réaction de la méthylamine avec l'acrylate de méthyle en solvant explicite (DMSO).

Les termes de REG positif varient de la même façon que l'énergie et les termes de REG négatifs varient de façon opposée ; si l'on se situe sur un segment correspondant à une barrière d'activation, les termes de REG positif déstabilisent le système (comme l'énergie augmente jusqu'à l'état de transition), et les termes de REG négatif stabilisent le système. Sur un segment correspondant à la stabilisation du système jusqu'au produit, il faut inverser les catégories puisque l'énergie diminue.

Une autre fonctionnalité inhérente à REG est le groupement des termes : puisque les termes de la méthode IQA sont tous additifs (on retrouve l'énergie moléculaire totale en sommant tous les termes), on peut grouper des termes entre eux, par exemple les termes $V_{cl}(A,B)$ et $V_{ee,xc}(A,B)$ pour une paire d'atomes, mais également des termes de même nature pour différents atomes, formant ainsi des groupes.

Cela peut s'avérer pertinent en groupant ensemble les atomes d'un substituant, ou des hydrogènes avec le carbone qui les porte, par exemple. Dans ce cas, les termes intra-atomiques des atomes groupés sont tous sommés pour donner un terme intra-atomique unique pour le groupe :

$$E_{intra}(G) = \sum_{A \in G} E_{intra}(A) \tag{2.66}$$

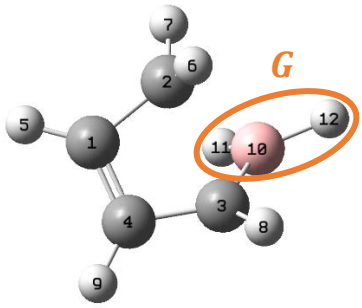
Les termes d'interaction entre les atomes du groupe sont également sommés :

$$E_{inter}(G,G) = \frac{1}{2} \sum_{A \in G} \sum_{(B \neq A) \in G} E_{inter}(A,B) \tag{2.67}$$

Et les termes d'interaction des atomes extérieurs A' avec chaque atome du groupe sont également sommés :

$$E_{inter}(A',G) = \sum_{A \in G} E_{inter}(A',A) \tag{2.68}$$

En pratique, si l'on prend le borylcyclobutène et qu'on souhaite grouper les trois atomes du substituant boryle ensemble, donc le bore 10 et les hydrogènes 11 et 12, on a les équations suivantes :



$$E_{intra}(G) = E_{intra}(B_{10}) + E_{intra}(H_{11}) + E_{intra}(H_{12})$$

$$E_{inter}(G,G) = E_{inter}(B_{10},H_{11}) + E_{inter}(H_{11},H_{12}) + E_{inter}(B_{10},H_{12})$$

Pour tout atome A' n'appartenant pas à G :

$$E_{inter}(A',G) = E_{inter}(A',B_{10}) + E_{inter}(A',H_{11}) + E_{inter}(A',H_{12})$$

D'un point de vue pratique, l'analyse est réalisée par une librairie Python organisée en cinq fichiers :

- Un script permet de lancer l'analyse en faisant appel à des procédures des quatre autres fichiers ;
- Gaussian_utils.py sert à lire des informations dans des fichiers résultats Gaussian ;

- Aimall_utils.py sert à récupérer et ordonner les données des calculs IQA (réalisés avec le logiciel AIMAll par le groupe de Popelier et le nôtre) ;
- REG.py réalise l'analyse REG proprement dite et produit les fichiers de résultats ;
- REG_vis.py produit des images permettant une visualisation simple des résultats.

Les étapes de la procédure sont détaillées sur la Figure 2.12 ci-dessous :

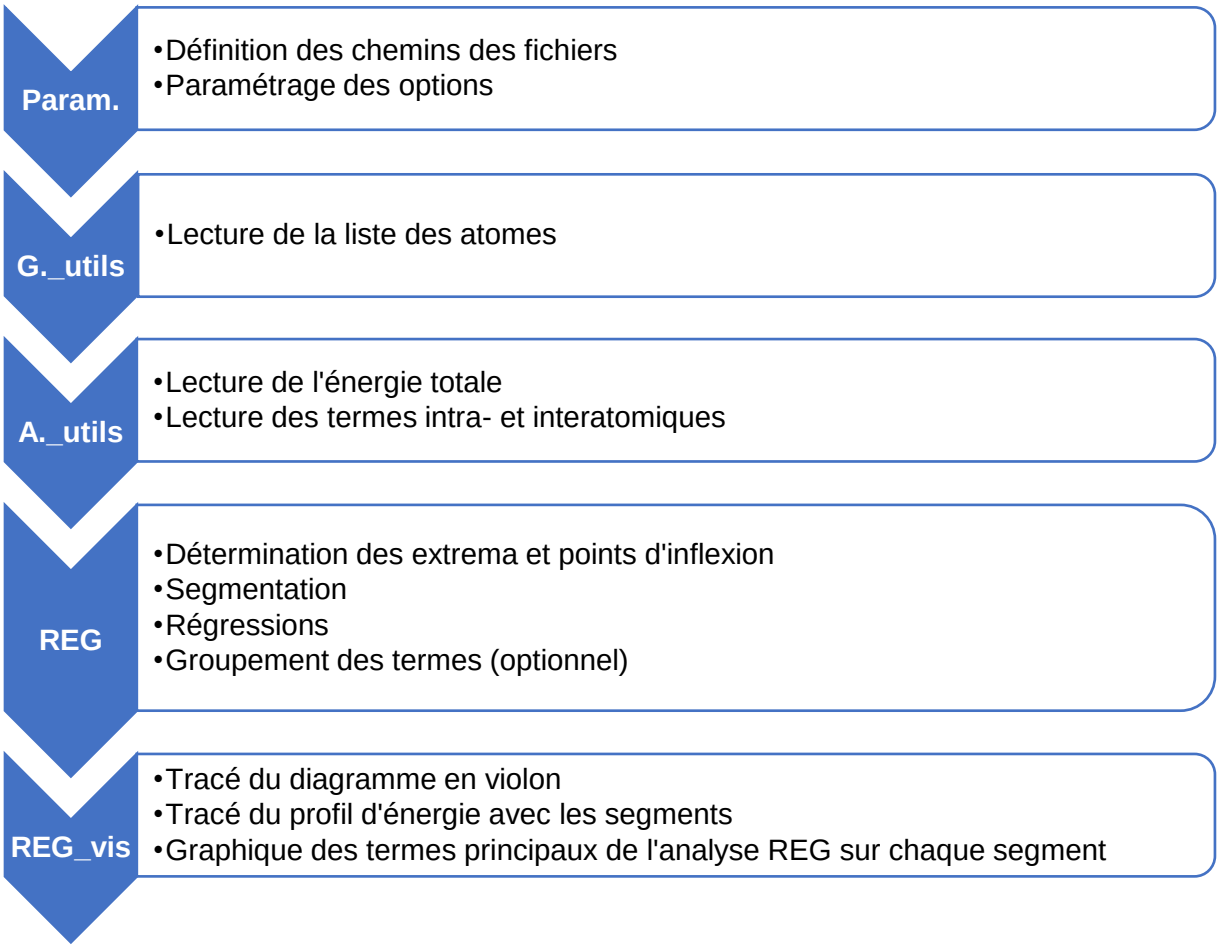


Figure 2.12 - Schéma du fonctionnement de la librairie REG.py

2.5.3 Applications

La méthode a été appliquée avec succès à l'étude du dimère d'eau [59], de l'effet *gauche* du fluor [61], de l'hydrolyse des peptides par la protéase HIV-1 [62], de la barrière de rotation du biphényle [63], des réseaux de liaisons hydrogène complémentaires [64], de l'inversion de l'azote chez les aziridines [65], des liaisons halogène [66], des transferts de proton dans des enzymes [67], des liens entre énergie de

déformation et énergie stérique [68], d'une réaction d'abstraction d'hydrogène [69], de réactions de S_N2 modèles [70], du polymorphisme de l'acide succinique [71] … À chaque fois, la méthode REG a permis de confirmer des explications connues, d'infirmer certaines hypothèses dans des cas sujets à controverse et/ou de mettre en évidence de nouveaux effets.

Bibliographie du chapitre 2

1. P. Politzer, A. Toro-Labbé, S. Gutiérrez-Oliva, B. Herrera, P. Jaque, M. C. Concha & J. S. Murray. *J. Chem. Sci.* **117**, 467–472 (2005).
2. B. Herrera & A. Toro-Labbé. *J. Phys. Chem. A* **111**, 5921–5926 (2007).
3. P. Flores-Morales, S. Gutiérrez-Oliva, E. Silva & A. Toro-Labbé. *J. Mol. Struct. THEOCHEM* **943**, 121–126 (2010).
4. L. Onsager. *J. Am. Chem. Soc.* **58**, 1486–1493 (1936).
5. S. Miertuš, E. Scrocco & J. Tomasi. *Chem. Phys.* **55**, 117–129 (1981).
6. J. L. Pascual-ahuir, E. Silla & I. Tuñon. *J. Comput. Chem.* **15**, 1127–1138 (1994).
7. J. Tomasi, B. Mennucci & R. Cammi. *Chem. Rev.* **105**, 2999–3093 (2005).
8. M. J. Field, P. A. Bash & M. Karplus. *J. Comput. Chem.* **11**, 700–733 (1990).
9. M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber & K. Morokuma. *J. Phys. Chem.* **100**, 19357–19363 (1996).
10. R. G. Parr, R. A. Donnelly, M. Levy & W. E. Palke. *J. Chem. Phys.* **68**, 3801–3807 (1978).
11. R. G. Parr & R. G. Pearson. *J. Am. Chem. Soc.* **105**, 7512–7516 (1983).
12. P. Geerlings, E. Chamorro, P. K. Chattaraj, F. De Proft, J. L. Gázquez, S. Liu, C. Morell, A. Toro-Labbé, A. Vela & P. Ayers. *Theor. Chem. Acc.* **139**, 1–18 (2020).
13. A. T. Maynard, M. Huang, W. G. Rice & D. G. Covell. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 11578–11583 (1998).
14. R. G. Parr, L. V. Szentpály & S. Liu. *J. Am. Chem. Soc.* **121**, 1922–1924 (1999).
15. L. R. Domingo & P. Pérez. *Org. Biomol. Chem.* **9**, 7168–7175 (2011).
16. R. G. Parr & W. Yang. *J. Am. Chem. Soc.* **106**, 4049–4050 (1984).
17. C. Morell, A. Grand & A. Toro-Labbé. *J. Phys. Chem. A* **109**, 205–212 (2005).
18. W. Yang & R. G. Parr. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 6723–6726 (1985).
19. L. Meneses, W. Tiznado, R. Contreras & P. Fuentealba. *Chem. Phys. Lett.* **383**, 181–187 (2004).
20. T. Gál, P. Geerlings, F. De Proft & M. Torrent-Sucarrat. *Phys. Chem. Chem. Phys.* **13**, 15003–15015 (2011).
21. R. G. Pearson. *J. Chem. Educ.* **45**, 581–587 (1968).
22. R. G. Pearson. *J. Chem. Educ.* **64**, 561–567 (1987).
23. E. Chamorro, P. K. Chattaraj & P. Fuentealba. *J. Phys. Chem. A* **107**, 7068–7072 (2003).
24. R. G. Parr & Y. Weitao. *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, 1989).
25. R. A. Miranda-Quintana, F. Heidar-Zadeh & P. W. Ayers. *J. Phys. Chem. Lett.* **9**, 4344–4348 (2018).
26. P. P. Zhou, S. Liu, P. W. Ayers & R. Q. Zhang. *J. Chem. Phys.* **147**, 134303 (2017).
27. R. S. Mulliken. *J. Chem. Phys.* **23**, 1833–1840 (1955).

28. F. Martin & H. Zipse. *J. Comput. Chem.* **26**, 97–105 (2005).

29. A. E. Reed, R. B. Weinstock & F. Weinhold. *J. Chem. Phys.* **83**, 735–746 (1985).

30. U. C. Singh & P. A. Kollman. *J. Comput. Chem.* **5**, 129–145 (1984).

31. C. M. Breneman & K. B. Wiberg. *J. Comput. Chem.* **11**, 361–373 (1990).

32. F. L. Hirshfeld. *Theor. Chim. Acta* **44**, 129–138 (1977).

33. P. Bultinck, C. Van Alsenoy, P. W. Ayers & R. Carbó-Dorca. *J. Chem. Phys.* **126**, 144111 (2007).

34. R. F. W. Bader. *Atoms in molecules : a quantum theory* (Oxford University Press, 1990).

35. P. L. A. Popelier. *Atoms in Molecules: An Introduction* (Pearson Education, 2000).

36. C. Gatti, P. Fantucci & G. Pacchioni. *Theor. Chim. Acta* **72**, 433–458 (1987).

37. X. Fradera, M. A. Austen & R. F. W. Bader. *J. Phys. Chem. A* **103**, 304–314 (1999).

38. R. McWeeny. *Methods of molecular quantum mechanics* (Academic Press, 1992).

39. M. A. Blanco, A. M. Pendás & E. Francisco. *J. Chem. Theory Comput.* **1**, 1096–1109 (2005).

40. A. Martín Pendás, M. A. Blanco & E. Francisco. *J. Chem. Phys.* **120**, 4581–4592 (2004).

41. P. Maxwell, Á. M. Pendás & P. L. A. Popelier. *Phys. Chem. Chem. Phys.* **18**, 20986–21000 (2016).

42. E. Francisco, J. L. Casals-Sainz, T. Rocha-Rinza & A. Martín Pendás. *Theor. Chem. Acc.* **135**, 1–8 (2016).

43. V. Tognetti, A. F. Silva, M. A. Vincent, L. Joubert & P. L. A. Popelier. *J. Phys. Chem. A* **122**, 7748–7756 (2018).

44. R. Chávez-Calvillo, M. García-Revilla, E. Francisco, Á. Martín Pendás & T. Rocha-Rinza. *Comput. Theor. Chem.* **1053**, 90–95 (2015).

45. I. Ruiz, E. Matito, F. J. Holguín-Gallego, E. Francisco, Á. Martín Pendás & T. Rocha-Rinza. *Theor. Chem. Acc.* **135**, 1–11 (2016).

46. J. Jara-Cortés & J. Hernández-Trujillo. *J. Comput. Chem.* **39**, 1103–1111 (2018).

47. A. Martín Pendás, E. Francisco, M. A. Blanco & C. Gatti. *Chem. – A Eur. J.* **13**, 9362–9371 (2007).

48. V. Tognetti & L. Joubert. *J. Chem. Phys.* **138**, 024102 (2013).

49. A. Martín Pendás, M. A. Blanco & E. Francisco. *J. Chem. Phys.* **125**, 184112 (2006).

50. O. A. Syzgantseva, V. Tognetti & L. Joubert. *J. Phys. Chem. A* **117**, 8969–8980 (2013).

51. M. Yahia-Ouahmed, V. Tognetti & L. Joubert. *Comput. Theor. Chem.* **1053**, 254–262 (2015).

52. M. Yahia-Ouahmed, V. Tognetti & L. Joubert. *Theor. Chem. Acc.* **135**, 1–15 (2016).

53. D. Tiana, E. Francisco, M. A. Blanco, P. MacChi, A. Sironi & A. Martín Pendás. *J. Chem. Theory Comput.* **6**, 1064–1074 (2010).

54. V. Tognetti, S. Bouzbouz & L. Joubert. *J. Mol. Model.* **23**, 1–11 (2017).

55. E. Romero-Montalvo, J. M. Guevara-Vela, W. E. Vallejo Narváez, A. Costales, Á. M. Pendás, M. Hernández-Rodríguez & T. Rocha-Rinza. *Chem. Commun.* **53**, 3516–3519 (2017).

56. D. Ferro-Costas, Á. M. Pendás, L. González & R. A. Mosquera. *Phys. Chem. Chem. Phys.* **16**, 9249–9258 (2014).

57. J. Dillen. *Int. J. Quantum Chem.* **113**, 2143–2153 (2013).

58. B. C. B. Symons, M. K. Bane & P. L. A. Popelier. *J. Chem. Theory Comput.* **17**, 7043–7055 (2021).

59. J. C. R. Thacker & P. L. A. Popelier. *Theor. Chem. Acc.* **136**, 86 (2017).

60. Page Github de REG : https://github.com/FabioFalcioni/REG.py.

61. J. C. R. Thacker & P. L. A. Popelier. *J. Phys. Chem. A* **122**, 1439–1450 (2018).

62. J. C. R. Thacker, M. A. Vincent & P. L. A. Popelier. *Chem. - A Eur. J.* **24**, 11200–11210 (2018).

63. P. L. A. Popelier, P. I. Maxwell, J. C. R. Thacker & I. Alkorta. *Theor. Chem. Acc.* **138**, 12 (2019).

64. O. J. Backhouse, J. C. R. Thacker & P. L. A. Popelier. *ChemPhysChem* **20**, 555–564 (2019).

65. I. Alkorta, J. Elguero & P. L. A. Popelier. *Chem. Phys. Lett.* **758**, 137927 (2020).

66. N. Orangi, K. Eskandari, J. C. R. Thacker & P. L. A. Popelier. *ChemPhysChem* **20**, 1922–1930 (2019).

67. A. J. Robertson, A. L. Wilson, M. J. Burn, M. J. Cliff, P. L. A. Popelier & J. P. Waltho. *ACS Catal.* **11**, 12840–12849 (2021).

68. B. C. B. Symons, D. J. Williamson, C. M. Brooks, A. L. Wilson & P. L. A. Popelier. *ChemistryOpen* **8**, 560–570 (2019).

69. M. Petković, Đ. Nakarada & M. Etinski. *J. Comput. Chem.* **39**, 1868–1877 (2018).

70. I. Alkorta, J. C. R. Thacker & P. L. A. Popelier. *J. Comput. Chem.* **39**, 546–556 (2018).

71. A. S. De la Vega, L. J. Duarte, A. F. Silva, J. M. Skelton, T. Rocha-Rinza & P. L. A. Popelier. *Phys. Chem. Chem. Phys.* **24**, 11278–11294 (2022).

Chapitre 3 – Étude de la torquosélectivité de l'ouverture thermique de cyclobutènes substitués

3.1 État de l'art 90

3.2 Ajouts de nouvelles procédures pour l'analyse REG-IQA et REGdiff 92

3.3 Implémentations pour l'analyse REG-NBO et REG-NBOdiff 103

3.4 Résultats 107

3.4.1 Systèmes et détails de calcul 107

3.4.2 Étude des profils d'énergie 108

3.4.3 Étude REG-NBOdiff de la différence des profils 110

3.4.4 Étude REG-IQAdiff de la différence des profils 120

3.1 État de l'art

Comme mentionné dans l'introduction générale, les cycles insaturés peuvent s'ouvrir sous l'action de la chaleur ou de la lumière [1]. Si l'on considère l'ouverture thermique de cyclobutènes, le produit obtenu est un diène, et les éventuels substituants de part et d'autre de la liaison rompue tournent en raison du changement d'hybridation des atomes de carbone qui les portent. Le processus de rotation est conrotatoire, en accord avec les règles de Woodward-Hoffmann [2] pour les réactions électrocycliques thermiques à des systèmes ouverts à 4n électrons π. Cette conrotation peut se faire vers l'intérieur ou l'extérieur du cycle, ce qui forme respectivement un diène E ou un diène Z comme indiqué sur la Figure 3.13 ci-dessous. Cette réaction est élémentaire et sous contrôle cinétique : l'état de transition se formant le plus vite (c'est-à-dire de plus basse énergie) mène au produit majoritaire.

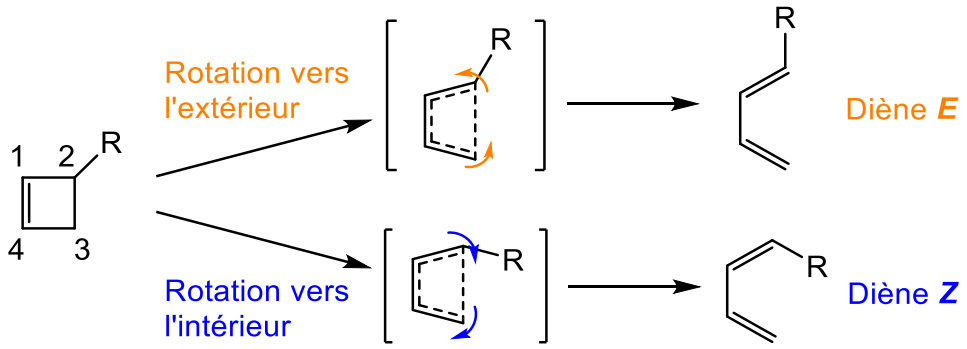


Figure 3.13 - Mécanisme de l'ouverture thermique de cyclobutènes substitués

Il a été observé expérimentalement que certains substituants favorisaient l'un ou l'autre mode de rotation ; cette préférence a été appelée torquosélectivité [3]. On pourrait a priori supposer que les effets stériques seraient les seules forces en présence et favoriseraient une rotation des substituants les plus encombrants vers l'extérieur du cycle. Cependant, les études précédentes [4–7] ont montré que les effets électroniques prennent le pas sur les effets stériques : ainsi, les substituants électroattracteurs favorisent la rotation vers l'intérieur du cycle tandis que les substituants électrodonneurs favorisent la rotation vers l'extérieur du cycle.

L'explication proposée dans la littérature [4,8] est de nature orbitalaire :

- Les substituants électroattracteurs possèdent généralement une orbitale localisée vacante suffisamment basse en énergie sur l'atome relié au cycle, par exemple une lacune électronique pour les boranes, une orbitale π* pour les dérivés carbonylés, ou une orbitale σ* pour les dérivés silylés. Lors de l'ouverture du cycle, l'orbitale σ occupée de la liaison qui se rompt peut interagir de façon stabilisante avec cette orbitale vacante, et la stabilisation sera d'autant plus forte si la rotation du substituant se fait vers l'intérieur du cycle, favorisant ainsi la formation du diène Z (à gauche sur la Figure 3.14 ci-dessous).
- En revanche, les substituants électrodonneurs possèdent une orbitale localisée occupée, comme une paire libre ou une orbitale σ. L'interaction entre cette orbitale occupée et celle de la liaison qui se rompt est cette fois déstabilisante puisque les deux paires d'électrons se repoussent. La déstabilisation est minimisée si la rotation du substituant se fait vers l'extérieur du cycle, favorisant ainsi la formation du diène E (à droite sur la Figure 3.14 ci-dessous).

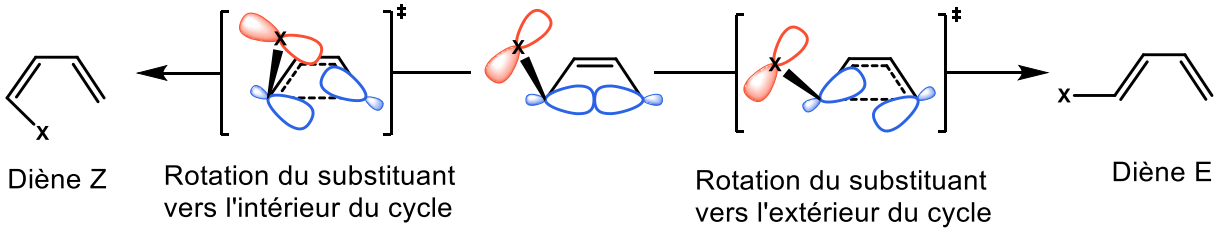


Figure 3.14 - Explication orbitalaire de l'ouverture de cyclobutènes substitués.

Gauche : conrotation du substituant vers l'intérieur du cycle. Droite : conrotation du substituant vers l'extérieur du cycle.

Cette explication a été vérifiée à de nombreuses reprises [8–13] mais présente des inconvénients : elle ne s'appuie que sur deux orbitales localisées, et on peut donc se demander si d'autres effets peuvent avoir été négligés. Des études plus complètes [14] ne se sont intéressées qu'à une sélection d'une dizaine d'interactions entre orbitales en considérant les interactions les plus fortes. D'autres études se sont consacrées aux effets de solvant [15] et à l'influence des substituants plus éloignés (situés de part et d'autre de la double liaison) [16] sur la torquosélectivité.

L'objet de notre étude est justement d'utiliser l'analyse REG pour comprendre, sans biais ni recours préalable à l'intuition chimique, de quels termes, tant en IQA qu'en NBO, provient la différence entre les deux chemins, afin de vérifier si l'on retrouve l'explication habituelle ou si d'autres effets interviennent. Cette approche est dénommée « REGdiff » (ou REG-IQAdiff pour la version IQA, et REG-NBOdiff pour la version NBO). Il est en effet intéressant d'étudier ces réactions sous les deux méthodes d'analyse simultanément pour vérifier s'il existe ou non un parallèle entre les approches IQA et NBO. En effet, ces deux méthodes sont fondamentalement différentes, puisque les termes IQA ne font intervenir que des termes mono- ou diatomiques tandis que les interactions NBO considèrent des interactions de 2 à 4 centres (interaction paire libre – lacune, ou interaction entre une orbitale liante et une orbitale antiliante, par exemple). Les représentations des systèmes dans les deux théories peuvent donc être assez éloignées mais potentiellement complémentaires.

3.2 Ajouts de nouvelles procédures pour l'analyse REG-IQA et REGdiff

Pour pouvoir comparer deux chemins aisément, l'approche la plus évidente est de soustraire les profils énergétiques et les descripteurs un à un : cela permet d'éliminer les termes dont les variations sont similaires pour mettre en évidence les termes variant différemment sur les deux chemins. Seulement, cette approche n'est pas sans poser quelques problèmes : les deux calculs d'IRC n'ont pas forcément la même longueur, le même nombre de points, le même positionnement des points…

Une solution pour résoudre ce problème est de normaliser les deux profils pour les ramener à une abscisse commune puis de les rééchantillonner pour avoir des points de même abscisse [17,18]. À cette fin, de nouvelles procédures ont été implémentées dans la librairie REG.py au cours de cette thèse. Les modifications et ajouts sont détaillés ci-dessous.

Le script général doit lire deux jeux de données Gaussian (profils d'IRC) et AIMAll (décomposition IQA), un pour chaque chemin, puis déterminer la segmentation des profils par l'intermédiaire de la force de réaction. Ensuite, une nouvelle procédure (`reg.crit_trim`) vérifie si les deux chemins ont le même nombre de segments, puisque la normalisation, les régressions, les rééchantillonnages et l'analyse sont

réalisés pour chaque segment. Les profils des deux chemins sont généralement très similaires et les segments supplémentaires pouvant apparaître sont généralement dus à de faibles irrégularités à l'une des extrémités du profil. En cas de désaccord entre les deux chemins, l'utilisateur peut choisir le(s) point(s) de séparation des segments à retirer du profil ayant trop de segments.

Ensuite, la coordonnée de réaction est normalisée sur chaque segment (procédure `reg.normalize`) afin que les segments de chaque profil soient compris dans les mêmes intervalles de coordonnée de réaction renormalisée. Dans notre cas, nous avons choisi d'imposer une coordonnée nulle pour le premier point du premier segment et une longueur de 1 pour chaque segment. Cela revient à imposer des valeurs entières aux bornes de chaque segment ($N-1$ et N pour le $N^{\text{ième}}$ segment) et à positionner les points du segment entre ces bornes en respectant leurs positions relatives par rapport aux bornes grâce à une transformation linéaire. L'équation utilisée à cette fin est la suivante :

$$\tilde{\xi}_k^{(N)} = \frac{\xi_k^{(N)} - \xi_0^{(N)}}{\xi_n^{(N)} - \xi_0^{(N)}} + N - 1 \tag{3.1}$$

où $\tilde{\xi}_k^{(N)}$ est la coordonnée normalisée du $k^{\text{ième}}$ point du $N^{\text{ième}}$ segment, $\xi_k^{(N)}$ est la coordonnée non-normalisée de ce point, et $\xi_0^{(N)}$ et $\xi_n^{(N)}$ sont respectivement les coordonnées du premier et dernier point du $N^{\text{ième}}$ segment. Le terme $N-1$ sert simplement à assurer que les coordonnées de chaque segment sont comprises dans l'intervalle $[N-1, N]$ pour que la coordonnée soit bien incrémentée de 1 à chaque segment, et le -1 impose que le premier point du premier segment ($N=1$) a une coordonnée nulle.

L'application de cette méthode est illustrée dans la Figure 3.15 ci-après.

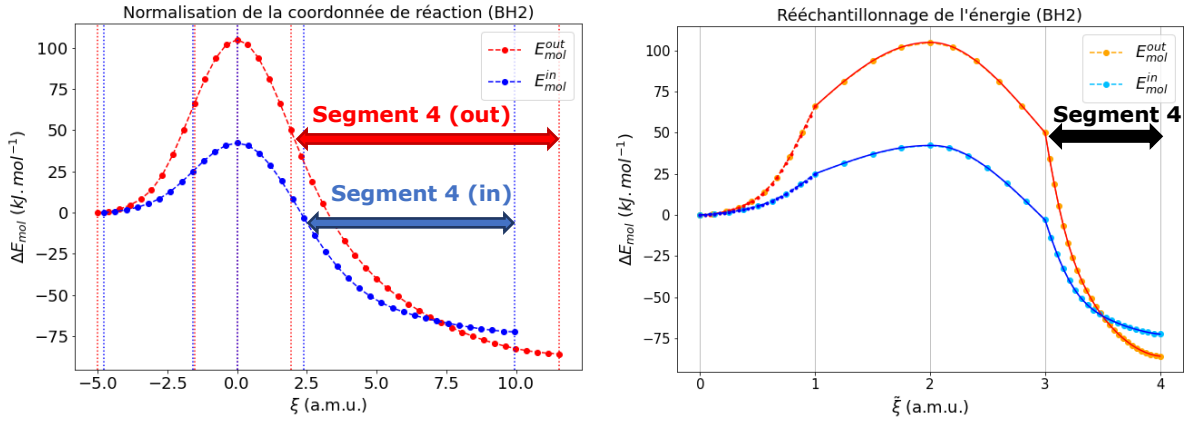


Figure 3.15 - Normalisation de la coordonnée de réaction des profils de réaction de l'ouverture thermique du borylcyclobutène.

Gauche : profils d'origine / Droite : profils renormalisés. Sur les deux graphiques, les barres verticales indiquent les limites de chaque segment.

On peut ainsi voir que les segments, qui ont à l'origine des tailles différentes pour chaque profil (ouverture « in » ou « out »), en particulier le quatrième (indiqué par les flèches bleue et rouge sur la figure de gauche), ont tous la même longueur et la même position après normalisation, comme par exemple le quatrième segment, indiqué par une flèche noire sur la figure de droite.

L'étape suivante consiste à faire une régression polynômiale des profils pour pouvoir les rééchantillonner, ce qui est assuré par la procédure `reg.fit`. Pour chaque profil énergétique et chaque descripteur, une régression polynômiale (cf. équation 3.2, le descripteur D est approximé par un polynôme P) est réalisée sur chaque segment en utilisant la méthode des moindres carrés. Par défaut, des polynômes d'ordre $n = 4$ sont utilisés.

$$D(\xi) \cong P(\xi) = c_0 + c_1\xi + c_2\xi^2 + \cdots + c_n\xi^n \tag{3.2}$$

La qualité de la régression est vérifiée avec un seuil minimal de 10^{-5} sur l'écart quadratique moyen normalisé (« $NRMSE$ », équation 3.3), qui n'est pas calculé de la façon habituelle (« $RMSE$ », équation 3.4) mais divisé par le nombre de points et l'écart entre les valeurs minimales et maximales du terme sur le segment :

$$NRMSE = \frac{\sqrt{\sum_{i=0}^{n}\left(P(\tilde{\xi}_i) - D(\tilde{\xi}_i)\right)^2}}{n\left|\max\left(D(\tilde{\xi})\right) - \min\left(D(\tilde{\xi})\right)\right|} \tag{3.3}$$

$$RMSE = \sqrt{\sum_{i=0}^{n}\left(P(\tilde{\xi}_i) - D(\tilde{\xi}_i)\right)^2} \tag{3.4}$$

En effet, des termes de valeur faible donnent lieu à des écarts quadratiques moyens également faibles, même si la régression est médiocre comme illustré sur la Figure 3.16. Cette modification permet de contourner le problème.

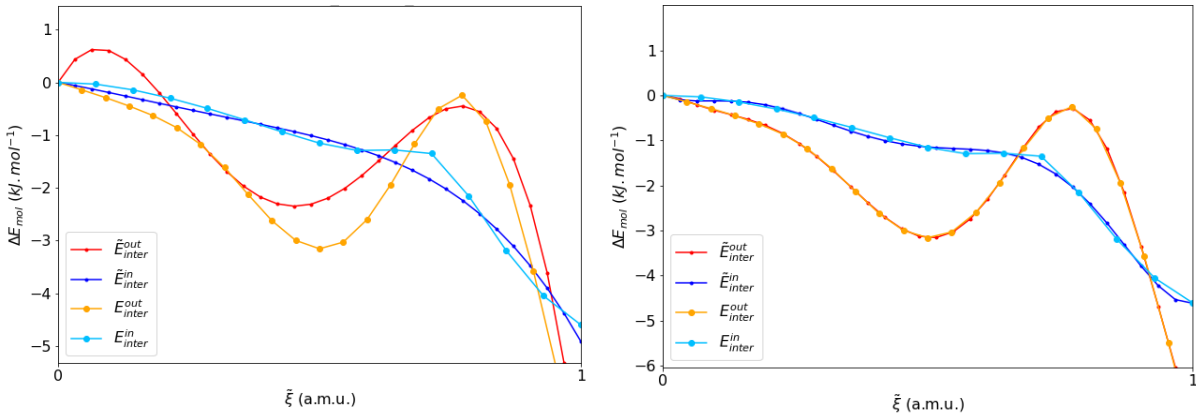


Figure 3.16 - Régressions polynômiales d'un terme IQA pour le difluorosilylcyclobutène sans et avec modification de correction.

Gauche : termes (cyan et orange) et régressions (bleu marine et rouge) obtenues avec le critère de RMSE < 10^{-5}. Droite : termes et régressions obtenues avec le critère de NRMSE < 10^{-5}.

On peut voir sur la figure de gauche que la régression (en rouge) du terme « out » (en orange) est assez mauvaise, tandis que la régression (en bleu foncé) du terme « in » (en cyan) est plutôt bonne en moyenne mais ne reproduit pas les fluctuations. Dans ce cas, les erreurs n'ont pas été détectées, le RMSE classique ne suffit pas. La figure de droite, quant à elle, présente les mêmes termes et leurs régressions obtenues avec le NRMSE, sans changer la valeur seuil : les régressions sont cette fois qualitativement bien meilleures, les erreurs ont pu être corrigées. Si la même erreur se répétait sur plusieurs termes, les régressions linéaires des différences des termes avec l'énergie totale pourraient être faussées et le classement REG bouleversé.

En cas de mauvaise qualité de régression, l'ordre du polynôme est incrémenté jusqu'à la limite de la moitié du nombre des points du segment. Cette limite sert à éviter l'apparition du phénomène de Runge : si l'ordre est augmenté jusqu'au nombre de points du segment, le polynôme passera forcément par tous les points (le polynôme de régression est alors le polynôme interpolateur de Lagrange) mais peut présenter d'importantes oscillations, notamment aux bords de l'intervalle, qui s'éloignent de la courbe souhaitée. C'est un problème classique des régressions polynômiales.

La régression obtenue est ensuite échantillonnée en des points régulièrement espacés et avec un nombre de points identique pour tous les segments (ici pris comme le nombre de points du segment en comportant le plus) :

$$\tilde{\xi}_i = \tilde{\xi}_0 + i\frac{\tilde{\xi}_n - \tilde{\xi}_0}{n}, i \in [0\,;n] \tag{3.5}$$

Enfin, les profils obtenus sont soustraits. Plusieurs procédures (`reg_vis.plot_energy`, `reg_vis.plot_renorm`, `reg_vis.plot_fit`, `reg_vis.plot_substract`) produisant des graphiques permettent de contrôler les différentes étapes de la méthode : lecture des énergies, renormalisation, fit et soustraction. Le résultat de la dernière procédure est illustré sur la Figure 3.17 ci-dessous.

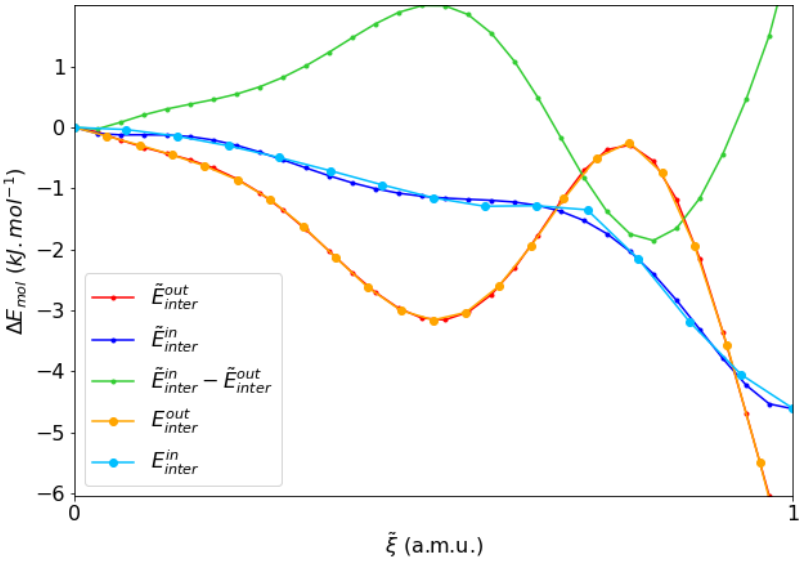


Figure 3.17 - Régressions polynômiales et différences d'un terme IQA pour le difluorosilylcyclobutène.

Termes (cyan et orange), régressions (bleu marine et rouge) et différence des deux régressions (vert) obtenues avec le critère de NRMSE < 10^{-5}.

Ensuite, la méthode REG habituelle reprend, avec cinq jeux de données (les termes IQA des deux profils, leurs courbes rééchantillonnées et la différence de ces régressions) : les signes des termes sont déterminés, les termes sont divisés en segments, les régressions linéaires des termes par rapport à l'énergie totale sont calculées et ceux-ci sont triés sur chaque segment par coefficient REG, puis les différents tableaux et graphiques de visualisation sont produits.

Cependant, dans certains cas plus difficiles, cette procédure ne suffit pas. En effet, lorsque les profils d'origine se croisent ou évoluent à des vitesses différentes, la différence des profils peut présenter des fluctuations qui n'apparaissent pas dans les profils de départ, ce qui fait que la variation de la différence n'est pas monotone. C'est notamment le cas pour le chlorocyclobutène, dont les profils sont représentés sur la Figure 3.18.

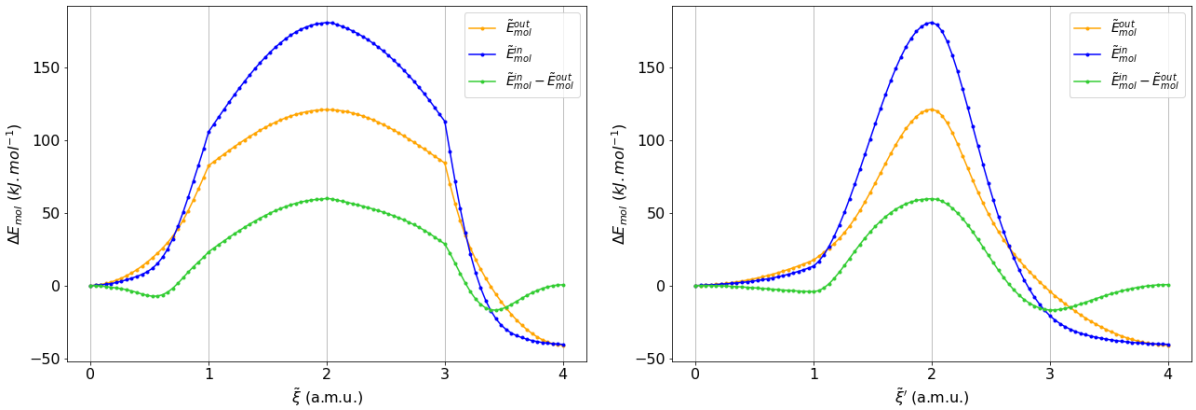


Figure 3.18 - Régressions rééchantillonnées (rouge et bleu) et différence des régressions (vert) des profils d'énergie de l'ouverture thermique du chlorocyclobutène.

Gauche : courbes obtenues en considérant les extrema et points d'inflexion des profils d'origine / Droite : courbes obtenues en considérant les extrema de la différence des profils.

Sur le graphique de gauche, la différence présente deux minima bien marqués au milieu du premier et du dernier segment. L'étude de cette différence donne lieu à des corrélations assez mauvaises puisque ses variations ne sont pas monotones sur chaque segment. Le graphique de droite montre les segments obtenus avec les extrema de la différence : cette fois, les variations d'énergie de la différence sont bien monotones et les corrélations sont bien meilleures.

Il peut alors être intéressant d'étudier le profil des différences en ne se basant pas sur les extrema et points d'inflexion des profils d'origine mais sur ceux du profil de différences. Selon le cas, on peut se restreindre aux extrema pour rester avec 4 segments.

Cependant, ce changement de procédure n'est pas anodin, puisque modifier les points de séparation des segments revient à changer la définition des segments (sans quoi on perd les segments de taille égale et avec le même nombre de points). Il faut donc trouver les points des profils d'origine correspondant le plus aux extrema de la différence : comme celle-ci est obtenue après rééchantillonnage, les points de la différence ne sont généralement pas les mêmes que ceux des profils d'origine.

En partant de $\tilde{\xi}_e^{(N)}$, la coordonnée normalisée d'un extrema de la différence d'énergie sur le $N^{\text{ème}}$ segment, on a la coordonnée correspondante sur un des profils d'origine :

$$\xi_e^{(N)} = \xi_0^{(N)} + (\tilde{\xi}_e^{(N)} - \tilde{\xi}_0^{(N)})(\xi_n^{(N)} - \xi_0^{(N)}) \tag{3.6}$$

avec $\xi_0^{(N)}$ la coordonnée d'origine du premier point du $N^{\text{ème}}$ segment, $\tilde{\xi}_0^{(N)}$ la coordonnée normalisée de ce même point et $\xi_n^{(N)}$ la coordonnée d'origine du dernier point du $N^{\text{ème}}$ segment.

Mais cette coordonnée intermédiaire $\xi_e^{(N)}$ ne correspond pas forcément à celle d'un point du profil d'origine. Il faut donc trouver la coordonnée du point le plus proche :

$$\xi_p^{(N)} = \min\{\left|\xi_e^{(N)} - \xi_i^{(N)}\right|, i \in [0, n]\} \tag{3.7}$$

La procédure `reg.norm_diff` se charge de ces deux étapes. Puis la procédure de normalisation / régression polynômiale / échantillonnage est répétée pour obtenir le nouveau profil de différence avec de nouveaux segments. La normalisation repart de la coordonnée de réaction d'origine en changeant les points de séparation des segments (équation 3.8), et n'utilise donc pas comme point de départ la coordonnée normalisée obtenue avec les extrema et points d'inflexion des profils d'origine. En effet, faire deux normalisations successives fausserait les résultats puisque les segments sont normalisés différemment les uns des autres.

$$\tilde{\xi}'^{(N)}_k = \frac{\xi^{(N)}_k - \xi'^{(N)}_0}{\xi'^{(N)}_n - \xi'^{(N)}_0} + N - 1 \tag{3.8}$$

On obtient ainsi une seconde coordonnée normalisée $\tilde{\xi}'$ en changeant les bornes des segments, à savoir $\xi'^{(N)}_0$ et $\xi'^{(N)}_n$, grâce à l'équation 3.8. La procédure est illustrée sur la Figure 3.19 ci-dessous.

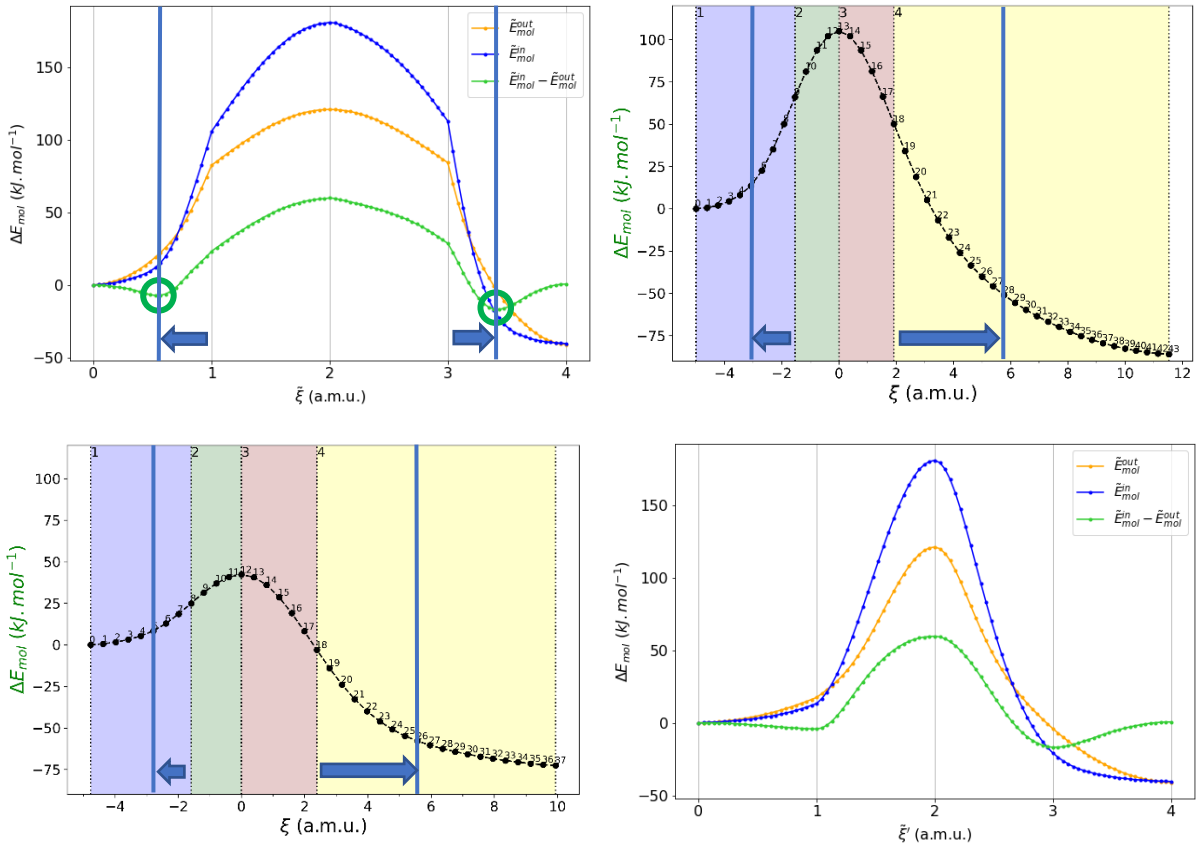


Figure 3.19 - Schéma d'illustration de la procédure de changements de segmentation

Haut à gauche : régressions des profils normalisés et différence ; les cercles verts indiquent les minima de la différence / Haut à droite : changement de segmentation du profil « out » en trouvant les points correspondant aux minima de la différence des profils via les équations 3.6 et 3.7 / Bas à gauche : même chose pour le profil « in » / Bas à droite : régressions des profils normalisés et différence obtenus après changement de segmentation et renormalisation des profils « in » et « out » suivant l'équation 3.8.

Le principal inconvénient de cette procédure est la grande sensibilité de celle-ci aux variations d'énergie de la différence, qui sont elles-mêmes sensibles à la qualité des régressions : même en ne se basant que sur les extrema de la différence, de faibles fluctuations dans les régressions peuvent donner lieu à des extrema supplémentaires de la différence qu'il faut trier. On peut aussi argumenter sur le fait que se

reposer uniquement sur la différence peut faire perdre de vue l'évolution des profils d'origine : les segments obtenus de cette façon ont-ils un sens chimique ?

Enfin, des changements ont été apportés aux procédures de visualisation. Comme décrit dans le chapitre 2, celles-ci génèrent des images, en particulier le graphique des termes de R^2 et coefficients REG les plus élevés en valeur absolue sur chaque segment. À l'origine, ces termes étaient simplement colorés suivant le signe de leur coefficient REG, information qui se trouve déjà sur le graphique. Comme il est intéressant de savoir si un terme est stabilisant ou non, nous avons décidé de changer la couleur des termes pour inclure les informations concernant leur signe sur ce graphique, avec des couleurs différentes selon que le terme est stabilisant (signe négatif), déstabilisant (signe positif) ou change de signe sur le segment.

On rappelle qu'en IQA, les termes intra-atomiques ($E_{intra}(A)$) sont toujours stabilisants tandis que les termes inter-atomiques ($E_{inter}(A,B)$) peuvent être stabilisants ou déstabilisants ; le terme d'interaction coulombienne $V_{cl}(A,B)$ peut également avoir n'importe quel signe tandis que le terme d'échange-corrélation $V_{ee,xc}(A,B)$ est toujours stabilisant. En NBO, les termes E(2) sont tous stabilisants et les termes E(4) tous déstabilisants.

Pour l'étude de différences de termes, il est possible de tracer le graphique avec des couleurs correspondant au signe de la différence des termes, ou bien avec une palette étendue indiquant le signe des termes d'origine. Sur la Figure 3.20, il s'agit de la seconde option. Ainsi,

- « Toujours stabilisant » indique que le terme est stabilisant (énergie de signe négatif) sur les deux chemins,
- « Toujours déstabilisant » qu'il est déstabilisant (énergie de signe positif) sur les deux chemins,
- « Changement de signe » que le terme change de signe sur les deux chemins,
- « Stabilisant / changement de signe » que le terme est stabilisant sur un chemin et change de signe sur l'autre (sans préciser lequel),
- « Déstabilisant / changement de signe » que le terme est déstabilisant sur un chemin et change de signe sur l'autre (sans préciser lequel).

Cependant, lorsqu'un terme est indiqué comme ayant des signes différents sur les deux chemins, le graphique n'indique pas quel signe correspond à quel chemin.

La Figure 3.20 ci-dessous représente un graphique d'analyse REG de différences de termes IQA utilisant cette légende (relative aux signes des termes d'origine).

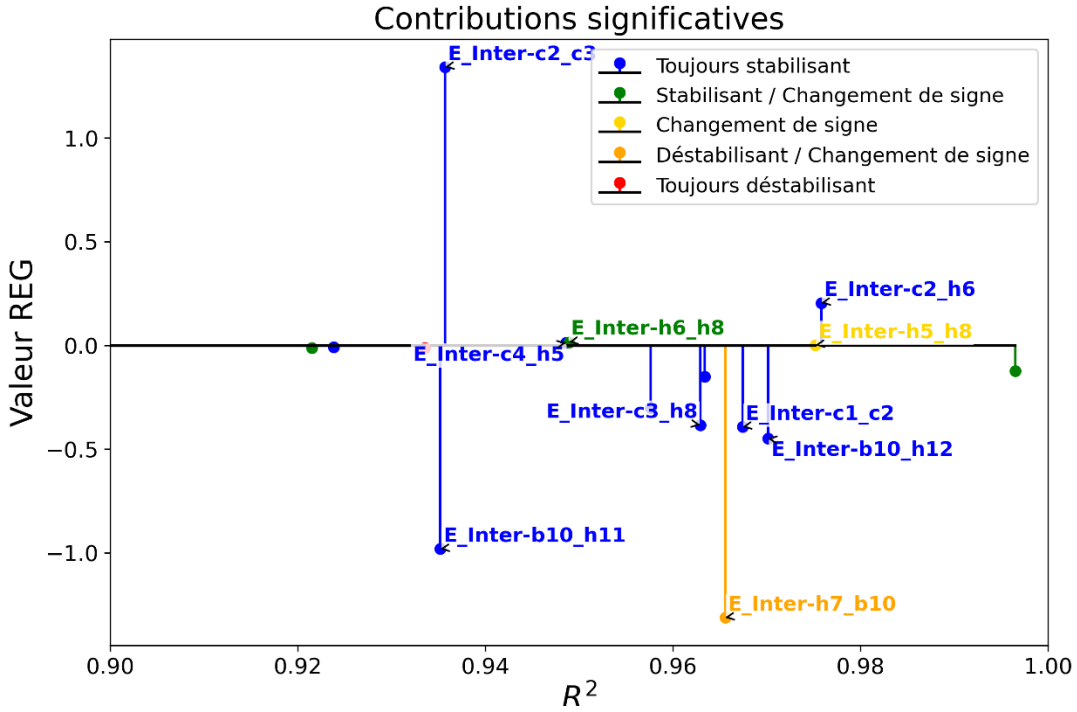


Figure 3.20 - Tracé de l'analyse REG-IQAdiff de la différence des termes IQA principaux sur le 1er segment de la réaction d'ouverture du borylcyclobutène.

Le fonctionnement de la librairie modifiée est représenté sur le schéma en Figure 3.21 ci-dessus, avec les modifications et implémentations effectuées au cours de la thèse en orange.

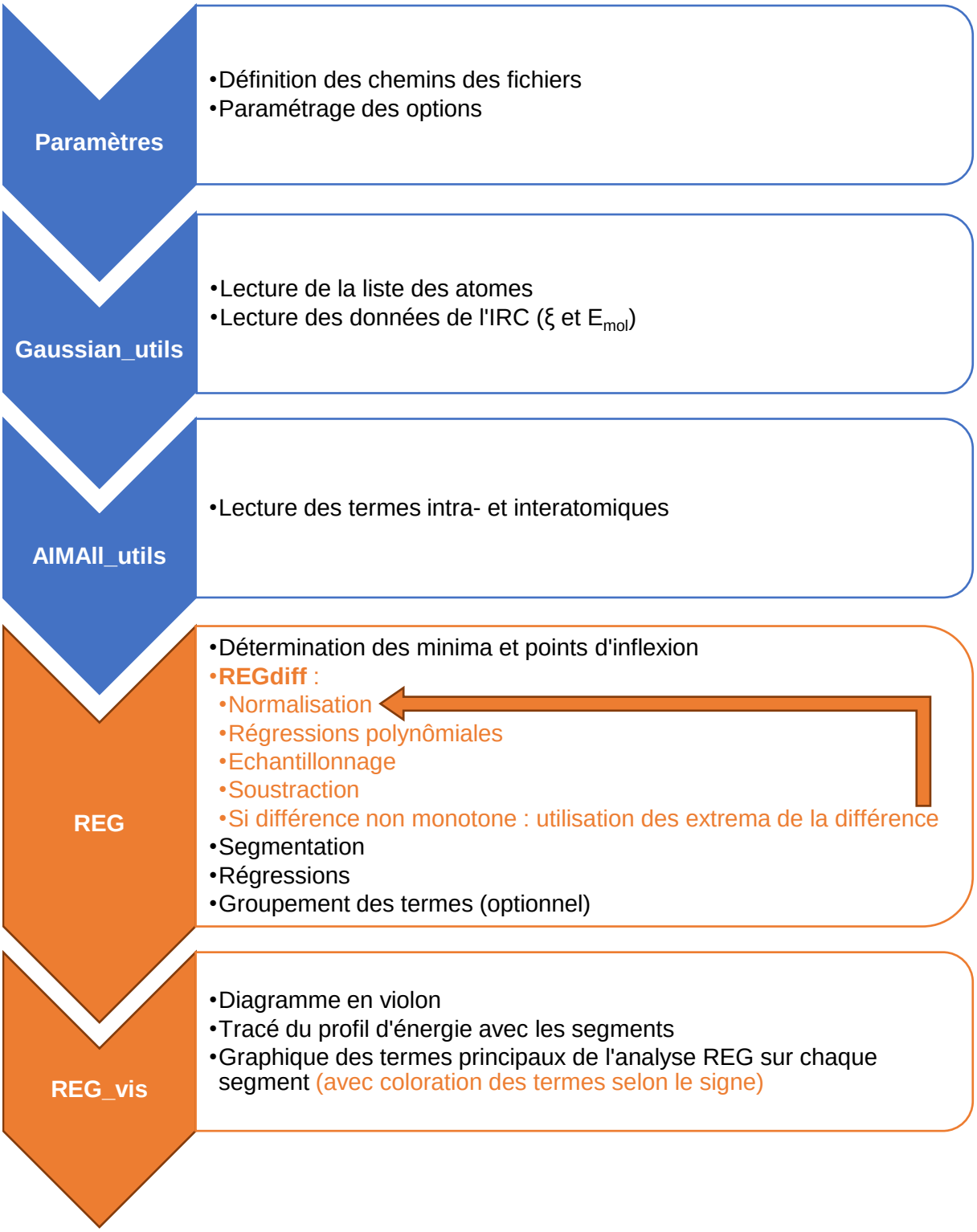


Figure 3.21 – Schéma du fonctionnement de la librairie REG.py modifiée (analyse REGdiff)

3.3 Implémentations pour l'analyse REG-NBO et REG-NBOdiff

La méthode REG étant initialement prévue uniquement pour l'analyse IQA, il a fallu implémenter de nouvelles procédures pour pouvoir lire et traiter les termes d'interaction NBO E(2) et E(4) décrits dans le chapitre 2. À terme, ces implémentations pourraient notamment permettre de rendre la méthode accessible à un plus grand public. En effet, la méthode IQA est assez marginale, reste coûteuse en temps de calcul, et nécessite des logiciels spécialisés. L'analyse NBO, quant à elle, est très rapide et incluse dans plusieurs logiciels usuels de chimie quantique comme Gaussian, ADF, Orca, Q-Chem et bien d'autres, et est utilisée dans plusieurs milliers d'articles par an.

Cependant, l'implémentation de l'analyse NBO n'est pas triviale. La raison principale en est la variation du nombre d'interactions entre orbitales dans les fichiers de résultats NBO, qui a deux causes : le changement du nombre d'orbitales et de leur nature, et le seuil d'affichage des interactions dans les fichiers de calcul.

Au contraire des termes IQA, les orbitales NBO présentes peuvent changer au cours d'un profil réactionnel : lors de la rupture d'une liaison, il est naturel que les orbitales liantes et antiliantes relatives à cette liaison disparaissent, tout comme la paire libre d'un atome nucléophile disparaît pour donner naissance à une orbitale liante et une orbitale antiliante correspondante lors d'une attaque nucléophile.

Par exemple, lors de l'ouverture du borylcyclobutène (illustrée en Figure 3.22), les orbitales liantes et antiliantes de la liaison entre les carbones 2 et 3 disparaissent juste après l'état de transition, une orbitale à trois centres (carbones 2, 3 et bore) se forme transitoirement, puis la seconde orbitale liante entre les carbones 1 et 4 disparaît tandis qu'apparaissent les secondes orbitales liantes entre les carbones 1 et 2 d'une part et 3 et 4 d'autre part. Il y a toujours 18 NBO occupées (en prenant un critère de plus d'1,8 électron par orbitale) dans le système, mais la nature de celles-ci est susceptible d'évoluer.

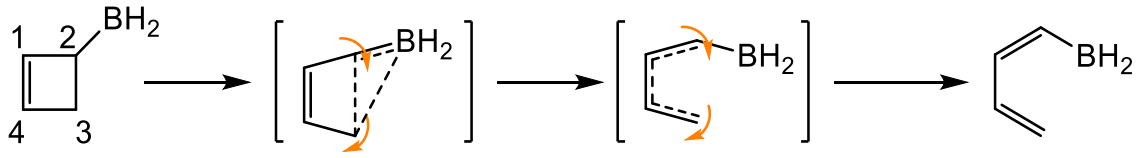


Figure 3.22 – Schéma de Lewis des étapes de l'ouverture du borylcyclobutène du point de vue NBO

Ces changements d'orbitales provoquent souvent des modifications dans la numérotation des orbitales, ce qui a posé des problèmes lors de l'implémentation. En effet, si l'on trie toutes les interactions pour garder celles qui restent présentes sur un segment entier, la prise en compte des numéros des orbitales dans les interactions fait que dès qu'une orbitale du système disparaît, les numéros de toutes les autres peuvent être décalés et la procédure peut incorrectement conclure que ces interactions n'existent pas en tout point. Elles peuvent donc être retirées de l'analyse, même si les orbitales sont toujours présentes.

Autre point de difficulté, les interactions entre orbitales, même si les orbitales sont bien définies, ne sont affichées dans les fichiers de résultats NBO que si leur valeur dépasse un certain seuil (par défaut défini à 0,5 kcal.mol^{-1} dans Gaussian). Ce seuil peut être abaissé en utilisant le programme externe NBO7 développé par Weinhold et al. [19] mais certaines interactions, de valeur trop faible, ne sont jamais affichées. Il faut donc vérifier dans chaque segment que les interactions sont bien présentes à chaque point étudié.

Ainsi, les interactions entre orbitales peuvent apparaître ou disparaître au cours de la réaction ; si l'on veut les étudier le long d'un segment et qu'elles n'existent pas en tout point du segment, la comparaison avec les termes présents sur tout le segment n'est pas possible, à moins de différencier les analyses pour chaque terme présentant ces irrégularités, ce qui est difficile lorsqu'on étudie des systèmes complexes.

Nous avons donc préféré retirer ces termes des segments où ils n'étaient pas définis en tout point. Après lecture par la nouvelle procédure `nbo_utils.nbo_property_from_gx_file`, ce tri est géré par une la procédure `nbo_utils.nbo_trim`. De la même façon, lorsque l'on compare deux profils, les termes peuvent également ne pas être présents sur les deux profils. Il faut donc aussi les trier avec la procédure `nbo_utils.nbo_trimdiff`.

Cela a à son tour provoqué de nouveaux problèmes, puisque cela signifie que le nombre de termes peut varier d'un segment à l'autre. Les valeurs des termes et les résultats des régressions sont stockés dans des tableaux multi-dimensionnels en utilisant la librairie Python NumPy, et ceux-ci ne sont pas conçus pour gérer facilement des objets de différentes tailles dans un même tableau. Le souci ne se pose pas avec les interactions IQA puisque leur nombre reste constant sur tout le profil. Ce problème a été résolu

en alternant entre tableaux NumPy pour chacun des segments et listes Python standard pour le tout. Les opérations sur les listes sont cependant différentes et plus complexes que pour les tableaux ; par exemple, sommer un à un les éléments de deux listes requiert une boucle parcourant les deux listes, tandis pour des tableaux, il suffit d'ajouter les deux tableaux avec un signe +.

Certains termes NBO se sont avérés être constants sur un segment entier du fait de la précision des résultats, arrondis à 0,01 kcal.mol^{-1}, ce qui se traduit par un coefficient REG nul et un coefficient de détermination R^2 non-défini lors des régressions avec l'énergie. Cela posait notamment problème lors de la visualisation. Les termes concernés sont donc également retirés.

De nouvelles procédures ont été développées pour la présentation des résultats, sans oublier de gérer le cas où un utilisateur se servirait du programme NBO3 inclus dans Gaussian ; celui-ci ne calcule que les termes E(2) alors que le programme NBO7 peut également calculer les termes E(4).

Le fonctionnement de la librairie modifiée est représenté sur le schéma en Figure 3.23 ci-après, avec les modifications en orange.

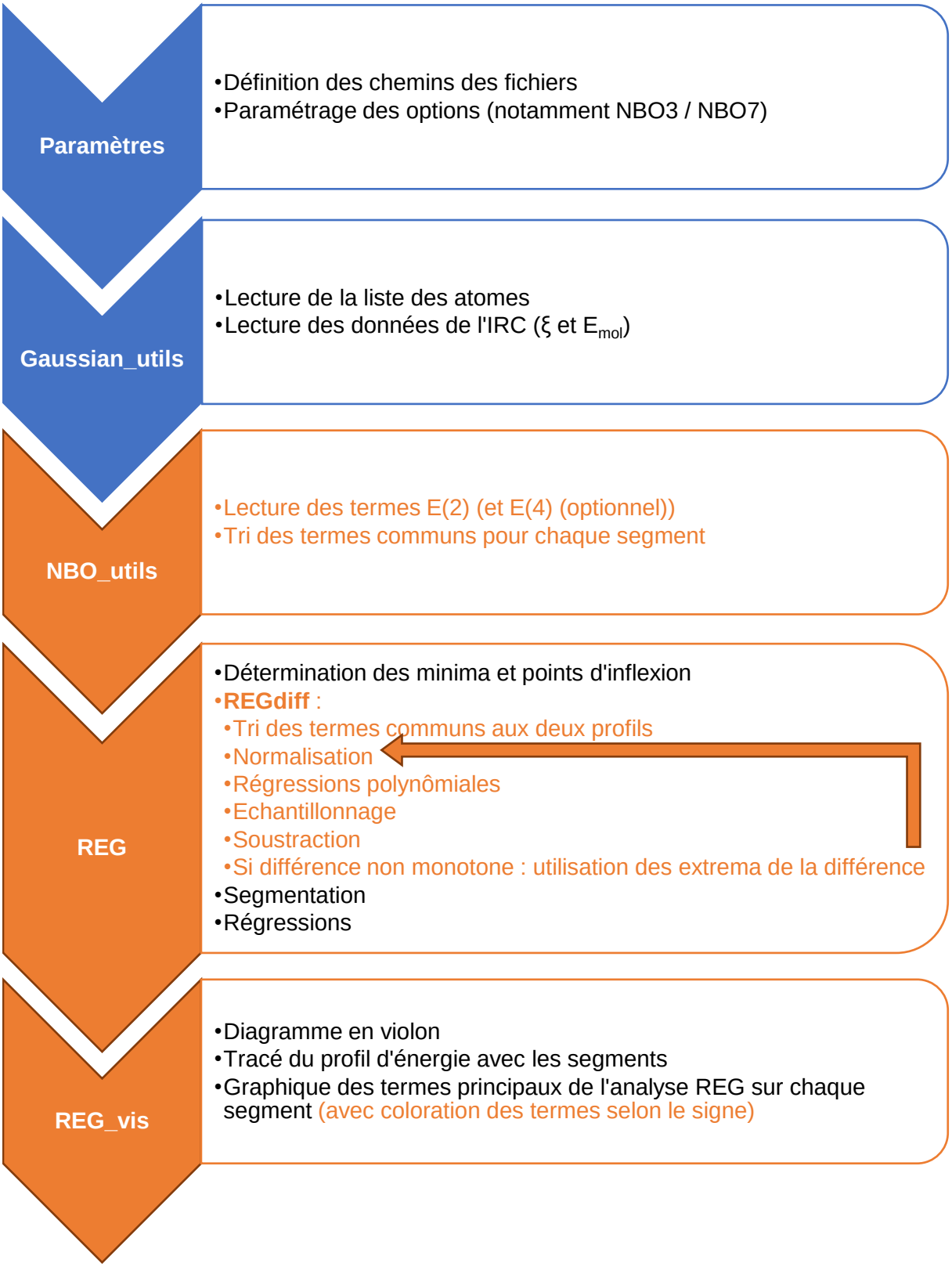


Figure 3.23 - Schéma du fonctionnement de la librairie REG.py modifiée (analyse NBO et NBOdiff)

3.4 Résultats

3.4.1 Systèmes et détails de calcul

Dans cette étude, nous avons considéré 16 cyclobutènes substitués sur la position 2 de sorte à couvrir une large gamme de torquosélectivités en nous appuyant sur les résultats de la littérature (voir Figure 3.24 ci-dessous) :

- 7 cyclobutènes favorisant la rotation vers l'intérieur du cycle (dite « in ») : $-BH_2$, $-BMe_2$, $-BCl_2$, $-SiF_3$, -CHO, -NO, $-SiH_3$.
- 9 cyclobutènes favorisant la rotation vers l'extérieur du cycle (dite « out ») : $-PH_2$, -CN, $-NO_2$, $-CH_3$, $-NH_3^+$, $-BH_3^-$, -Cl, -F, $-NH_2$.

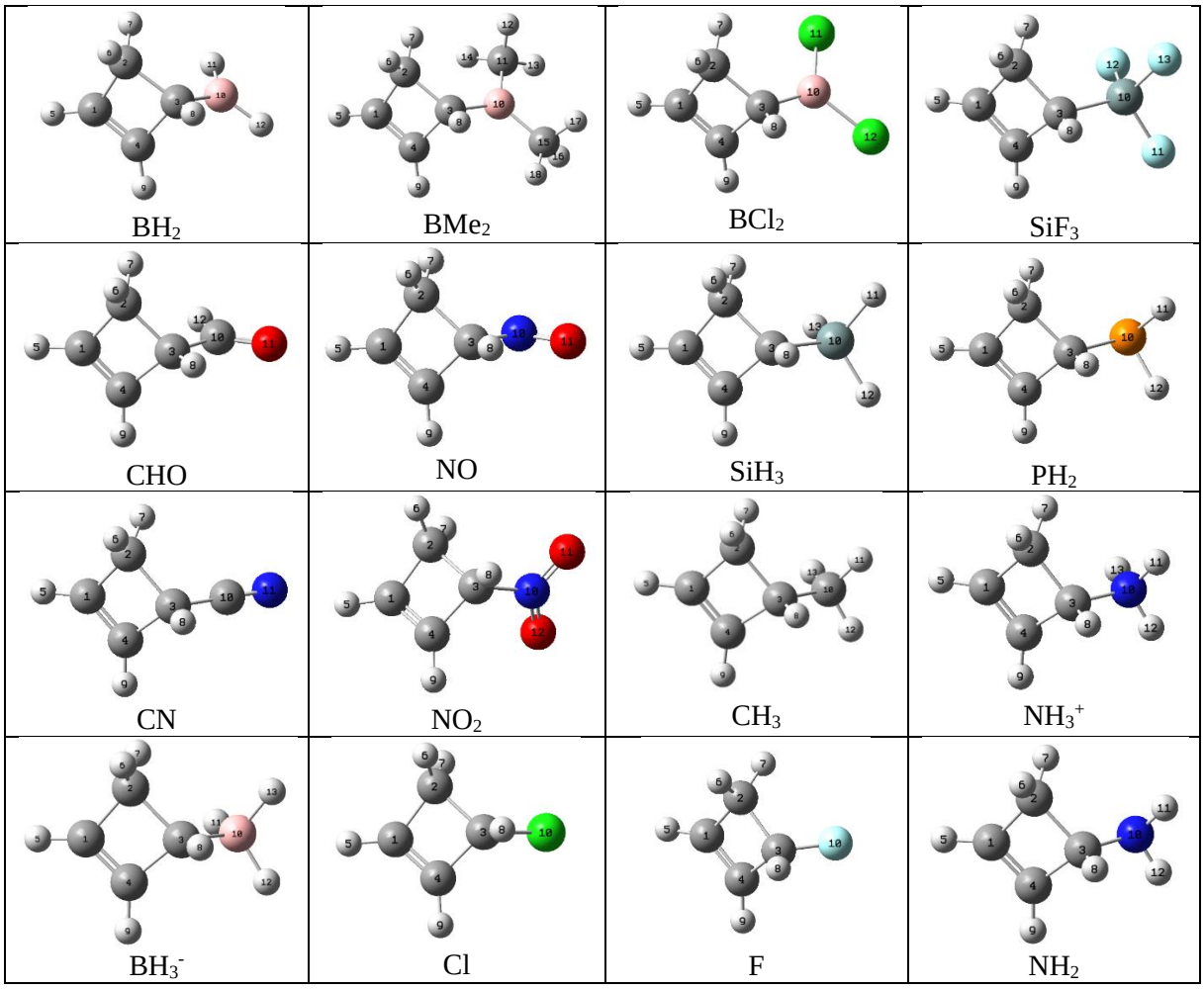


Figure 3.24 - Représentation "boules-bâtons" des systèmes étudiés.

De gauche à droite et de haut en bas : $-BH_2$, $-BMe_2$, $-BCl_2$, $-SiF_3$, -CHO, -NO, $-SiH_3$, $-PH_2$, -CN, $-NO_2$, $-CH_3$, $-NH_3^+$, $-BH_3^-$, -Cl, -F, $-NH_2$.

Nous avons utilisé la fonctionnelle hybride B3LYP [20] avec la correction de dispersion D3 [21] et la base de Pople 6-311++G(2d,2p). Les logiciels utilisés dans l'étude sont Gaussian 09 révision D.01 [22] pour les optimisations de géométrie, scans relaxés, calculs d'état de transition, de fréquences et d'IRC, AIMAll 19.10.12 [23] pour les calculs IQA et NBO7 [19] pour les calculs NBO.

Les effets de solvant n'ont pas été considérés ; en effet, AIMAll ne peut pas réaliser de calculs IQA en prenant en compte les modèles de solvant implicites (comme le modèle PCM) et il aurait donc fallu utiliser un modèle de solvant explicite pour ce faire, compliquant nettement les calculs. Cette approche est toutefois explorée pour l'addition d'aza-Michael dans le chapitre 5.

L'utilisation d'AIMAll a également conditionné l'utilisation de la fonctionnelle B3LYP puisque seules quelques fonctionnelles de DFT (B3LYP, M06-2X, LSDA, PBE et PBE0) sont explicitement implémentées dans AIMAll pour les calculs IQA. Nous avons choisi cette fonctionnelle en particulier puisqu'elle est utilisée dans plusieurs études précédentes sur la torquosélectivité.

Pour chaque système, la géométrie du réactif de départ a été optimisée, puis un scan relaxé d'élongation de la liaison C_2-C_3 a été réalisé pour trouver un des états de transition. L'autre état de transition a été obtenu en inversant les positions des substituants sur le C_2. Les calculs d'IRC ont été réalisés avec un pas de 0,05 bohr dans la majorité des cas (0,08 bohr en cas de problème de convergence de l'IRC) avec l'algorithme HPC. Les calculs de fréquences vibrationnelles ont permis de vérifier qu'aucune des structures optimisées des réactifs et produits ne comportaient de fréquence imaginaire tandis que les structures des états de transition en possèdent exactement une. Tous les 5 points de l'IRC, un calcul AIMAll complet (calcul de tous les termes) et un calcul NBO des termes E(2) et E(4) ont été réalisés.

3.4.2 Étude des profils d'énergie

Puisque la réaction est sous contrôle cinétique, l'état de transition de plus basse énergie mène au produit favorisé, même si celui-ci est moins stable que le produit minoritaire comme sur la Figure 3.25 : dans cet exemple, l'état de transition correspondant à la rotation « in » (profil bleu) est plus bas en énergie

que l'état de transition correspondant à la rotation « out » (profil orange), même si le produit « in » est moins stable que le produit « out ».

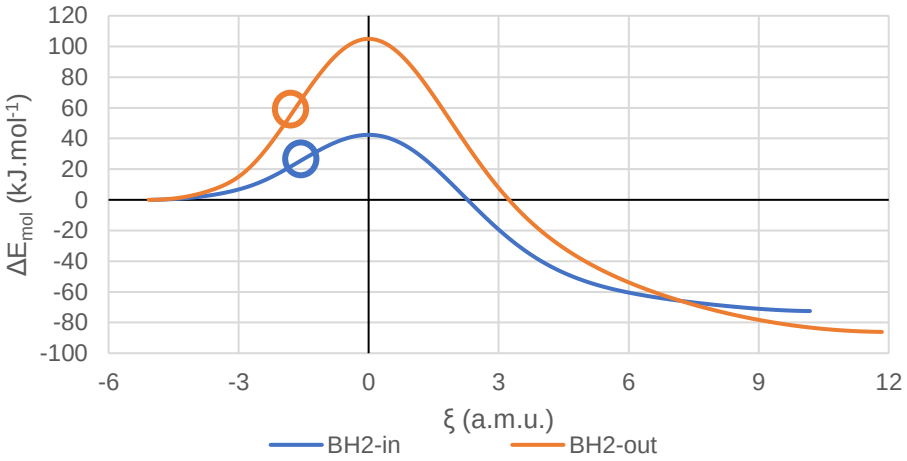


Figure 3.25 - Profils d'énergie de l'ouverture thermique du borylcyclobutène

La référence des énergies est placée au réactif. Les cercles indiquent le premier point d'inflexion de chaque profil, qui sépare les deux premiers segments.

Il suffit donc d'étudier le début de la réaction, du réactif aux états de transition, pour comprendre l'origine de la torquosélectivité. On peut également constater sur la Figure 3.26 qu'il existe une corrélation entre la barrière d'activation et la différence d'énergie entre le réactif et le premier point d'inflexion du profil, qui délimite le premier segment.

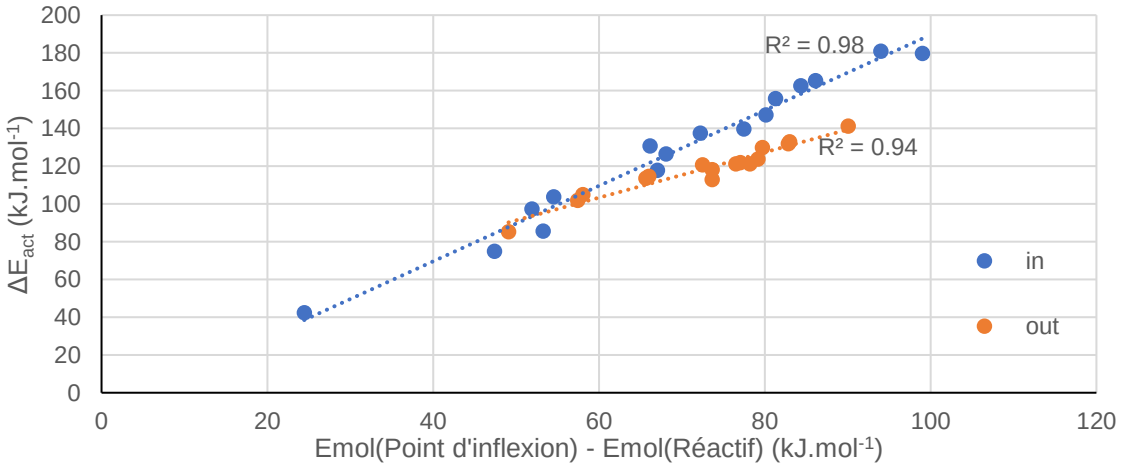


Figure 3.26 - Corrélation entre la barrière d'activation ΔE_{act} et la différence d'énergie entre les réactifs et le premier point d'inflexion du profil d'énergie pour toutes les réactions de tous les systèmes

On peut donc étudier seulement le premier segment du profil, qui va du réactif au premier point d'inflexion du profil, pour comprendre la différence de stabilité des états de transition.

Le Tableau 2 rassemble les différences de barrières d'activation en énergie et énergie libre entre les chemins « in » et « out » issus de la littérature et de ces travaux.

Tableau 2 - Différences d'enthalpies d'activation entre les chemins "in" et "out" en kJ.mol^{-1}

(kJ.mol^{-1})	**Données de** [24]		**Données de** [6]	**Mes travaux**
Substituant	B3LYP/ 6-31G(d)	B3LYP/ 6-311++G(3d,3p)	MP2/6-31G(d) // 3-21G	B3LYP-D3/ 6-311++G(2d,2p)
BH_2	-66,5	-62,8	-78,2	-61,2
BMe_2				-37,5
BCl_2				-28,0
SiF_3	-17,2	-15,9		-16,1
CHO	-16,3	-15,9	-19,7	-16,0
NO			-19,7	-13,8
SiH_3	-6,7	-6,3		-5,2
H	0	0	0	0
PH_2	17,2	17,6		19,2
CN			17,6	18,9
NO_2			22,2	23,6
CH_3	24,7	27,2		27,6
NH_3^+	26,4	25,9		26,6
Cl				58,7
F			63,2	59,9
NH_2	61,5	66,9		78,6
BH_3^-	146,4			28,7

Ces résultats sont en bon accord avec les résultats des études précédentes, excepté BH_3^- dont la différence est donnée beaucoup plus élevée (de l'ordre de 146 kJ.mol^{-1}) [24], plus encore que NH_2, ce que nous supposons être une erreur dans l'article d'origine après avoir refait les calculs avec le même protocole et avoir trouvé une valeur très proche (30,6 kJ.mol^{-1}) de celle présentée plus haut.

3.4.3 Étude REG-NBOdiff de la différence des profils

Nous allons donc étudier les différents systèmes par l'analyse REG-NBOdiff en partant des systèmes extrêmes de chaque catégorie, le borylcyclobutène (-BH_2) et l'aminocyclobutène (-NH_2). On pourra ensuite vérifier si les résultats obtenus sur ces systèmes pourront être étendus au reste de leur catégorie.

Pour chaque composé, le profil « out » a été soustrait au profil « in ». Cela mène à un profil globalement négatif pour les composés favorisant le chemin « in » : celui-ci étant d'énergie plus faible que le chemin « out » pour ces composés, la différence « in » - « out » sera négative ; pour les composés favorisant le chemin « out », c'est l'inverse, puisque le chemin « in » est cette fois d'énergie supérieure au chemin « out ». Les différences de termes variant comme la différence d'énergie totale auront un coefficient REG positif, quelle que soit la variation de la différence d'énergie totale.

3.4.3.1 Cas extrêmes

Commençons donc par le borylcyclobutène, qui favorise largement le chemin « in » (ΔE_{act} = -66,3 kJ.mol^{-1}). Les deux profils normalisés et leur différence sont visibles sur la Figure 3.27 ci-dessous.

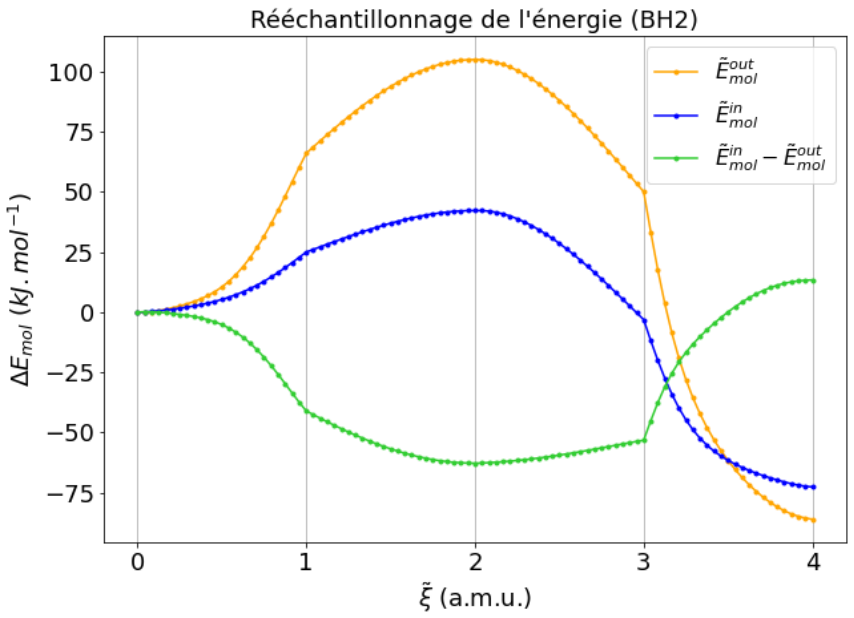


Figure 3.27 - Profils normalisés selon l'équation 3.1 et différence des profils « in » et « out » pour l'ouverture thermique du borylcyclobutène

Ici, la différence entre les chemins ne fait que croître (en valeur absolue) jusqu'à l'état de transition et décroît ensuite. Le résultat de l'analyse REG-NBOdiff sur le premier segment est présenté en Figure 3.28 ci-après.

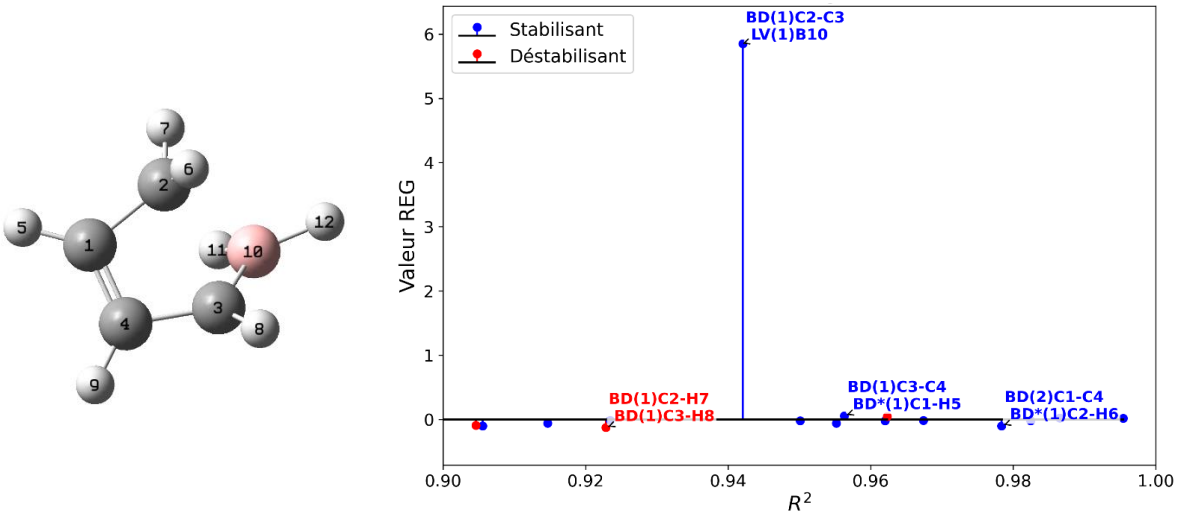


Figure 3.28 - Termes significatifs (R^2>0,90) issus de l'analyse REG-NBOdiff pour le premier segment de la réaction d'ouverture thermique du borylcyclobutène

On observe qu'un terme favorisant le chemin « in » (REG positif) ressort très largement sur les deux premiers segments : il s'agit d'une interaction E(2) entre la NBO liante des carbones 2 et 3 et la NBO vacante sur le bore. Ce terme est donc plus stabilisant sur le chemin « in » que « out », ce qui est cohérent avec l'explication orbitalaire habituelle, qui est ainsi confirmée pour ce système. On obtient le même résultat sur le second segment (Figure 3.29), le même terme domine largement les autres.

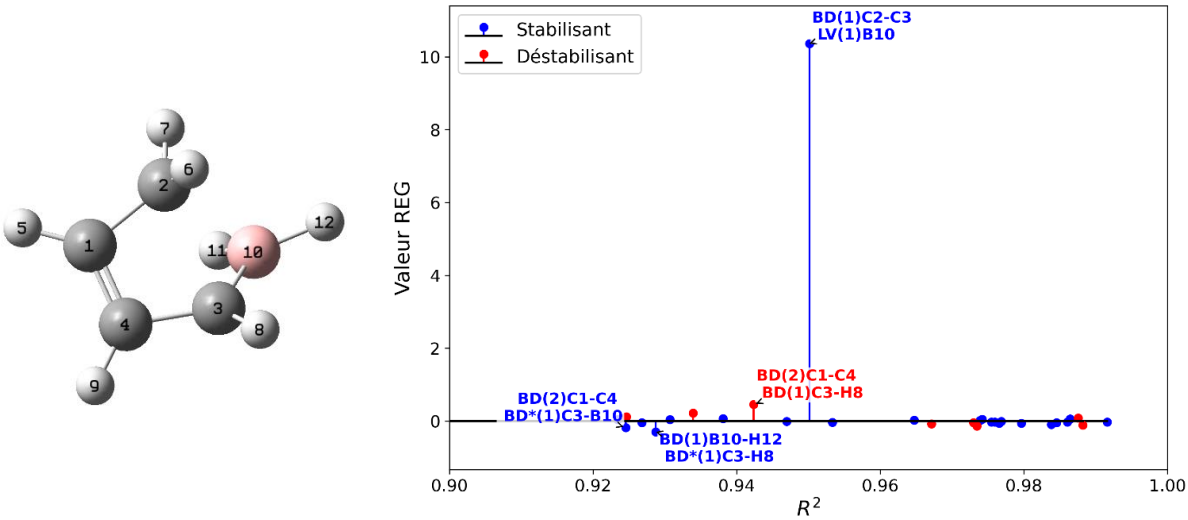


Figure 3.29 - Termes significatifs (R^2>0,90) issus de l'analyse REG-NBOdiff pour le second segment de la réaction d'ouverture thermique du borylcyclobutène

Quant à l'aminocyclobutène, qui favorise le chemin « out » (ΔE_{act} = 78,6 kJ.mol^{-1}), on observe que la différence d'énergie croît légèrement d'abord en faveur du chemin « in » sur le premier segment, puis change totalement de direction pour repasser en faveur du chemin « out ». Ce profil était bien adapté à une normalisation utilisant l'équation 3.7 comme on peut le voir sur la Figure 3.30.

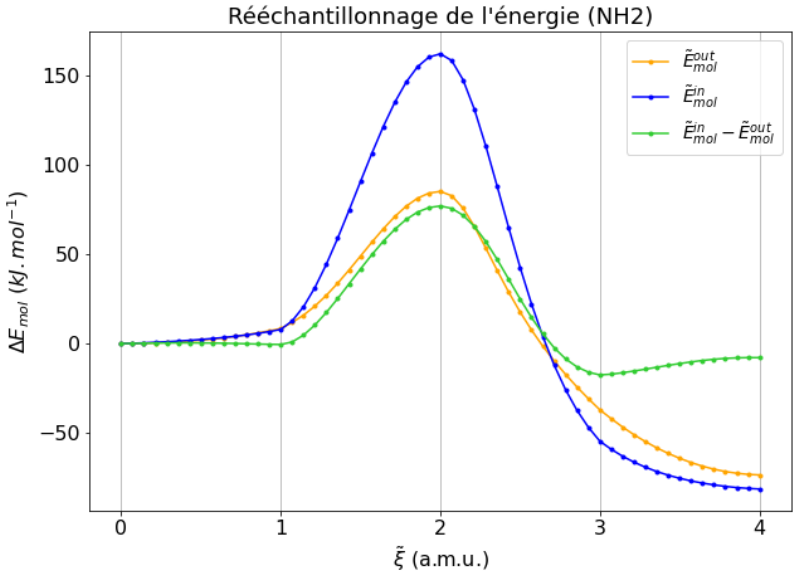


Figure 3.30 - Profils normalisés selon l'équation 3.7 et différence des profils « in » et « out » pour l'ouverture thermique de l'aminocyclobutène

Dans ce cas, la variation de la différence d'énergie sur le premier segment est très faible et ne favorise pas le bon chemin. Cela donne lieu à des corrélations assez mauvaises (aucun terme n'a un R^2 supérieur à 0,90). En revanche, les corrélations sont bien meilleures sur le second segment (cf. Figure 3.31).

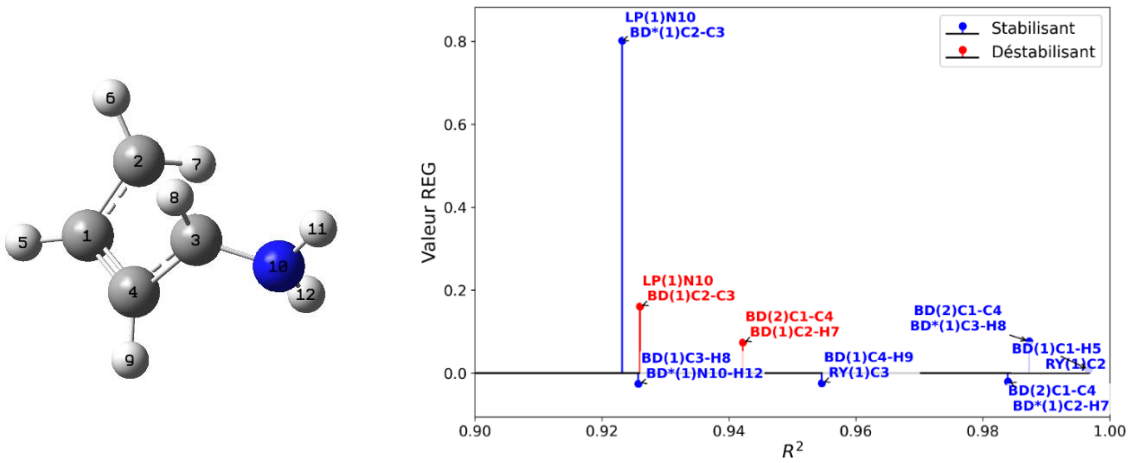


Figure 3.31 - Termes significatifs (R^2>0,90) issus de l'analyse REG-NBOdiff pour le second segment de la réaction d'ouverture thermique de l'aminocyclobutène

On constate que le terme favorisant le plus le chemin « out » est, à l'instar des systèmes précédents, un terme E(2) (stabilisant), cette fois entre la NBO antiliante entre les C_2 et C_3 et la NBO de la paire libre de l'azote. On trouve également l'interaction E(4) (déstabilisante) entre la NBO liante entre les C_2 et C_3 et la NBO de la paire libre de l'azote, qui est le terme attendu par l'explication orbitalaire, mais son coefficient REG est bien plus faible, presque 4 fois moins que le terme E(2). Ils contribuent donc tous deux à favoriser le chemin « out » mais le terme E(2) est nettement plus important du point de vue REG.

Cela montre un effet jusque-là inaperçu, qui va certes dans le même sens que l'effet habituellement mis en avant, mais qui est prépondérant et a du sens chimiquement : la paire libre de l'azote permet de faciliter la rupture de la liaison C_2-C_3.

Peut-on retrouver les mêmes termes que les systèmes extrêmes dans d'autres systèmes ?

3.4.3.2 Systèmes ressemblant aux systèmes extrêmes

Pour les autres systèmes favorisant le chemin « in », on retrouve, comme pour le borylcyclobutène, l'interaction stabilisante entre la NBO liante entre les C_2 et C_3 et une NBO vacante, qui peut être une lacune électronique pour les boranes ou une π^* pour le carbonyle et le nitrosyle. Cependant, ce terme domine moins que pour le borylcyclobutène et les profils sont moins évidents à étudier.

Par exemple, pour le carbonylcyclobutène (ΔE_{act} = -18 kJ.mol^{-1}), la différence d'énergie décroît en valeur absolue sur le second segment (voir Figure 3.32).

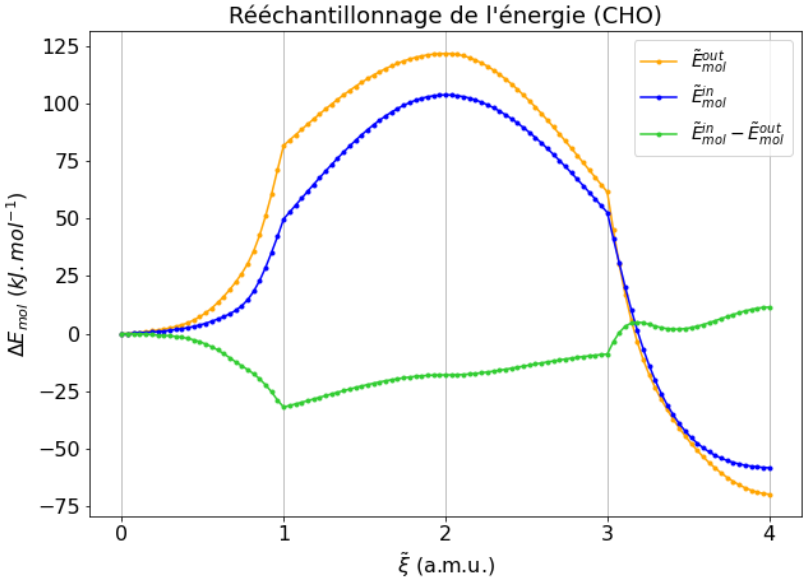


Figure 3.32 - Profils normalisés selon l'équation 3.1 et différences des profils « in » et « out » pour l'ouverture thermique du carbonylcyclobutène

Cela provoque une inversion du signe REG du terme précédemment cité sur ce segment (voir Figure 3.33 ci-dessous).

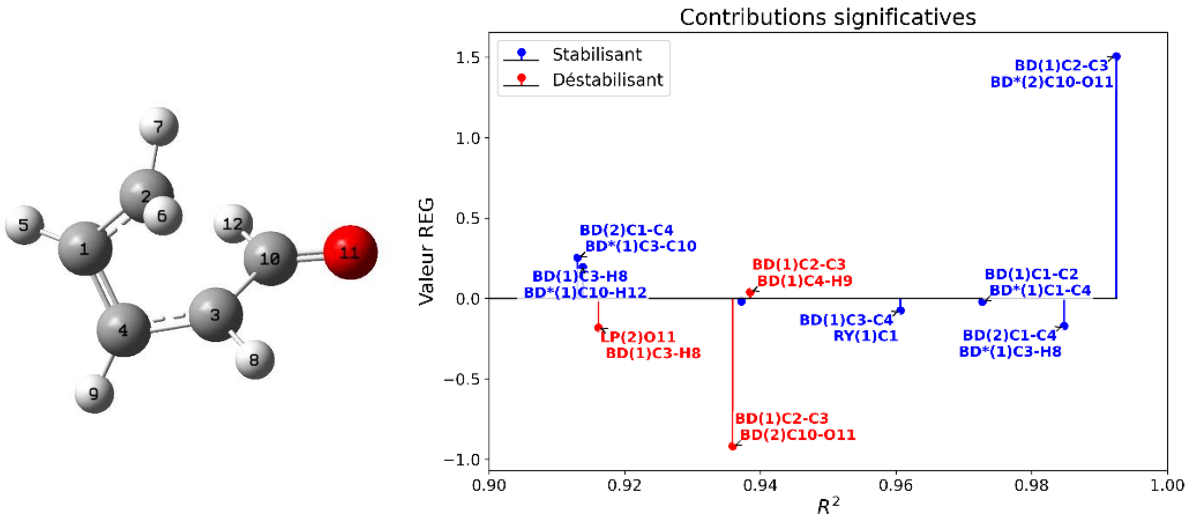


Figure 3.33 - Termes significatifs (R^2>0,90) issus de l'analyse REG-NBOdiff sur le premier segment de l'ouverture thermique du carbonylcyclobutène

En effet, ce terme continue à devenir plus stabilisant, à l'encontre de la différence d'énergie totale. Il est également accompagné par son analogue E(4) déstabilisant, l'interaction entre la NBO liante C_2-C_3 et la NBO liante π de la double liaison (C=O ou N=O selon le cas). Cette interaction a un coefficient REG

négatif, indiquant qu'elle est plus déstabilisante sur le chemin « out » que « in » : ce terme joue également en la faveur du chemin « in ». De plus, son coefficient REG est inférieur mais du même ordre de grandeur que celui du terme E(2) correspondant. En revanche, ce terme n'apparaît plus sur le deuxième segment.

On observe la même chose pour le nitrosocyclobutène (ΔE_{act} = -15,7 kJ.mol^{-1}) sur la Figure 3.34 et la Figure 3.35 ci-après.

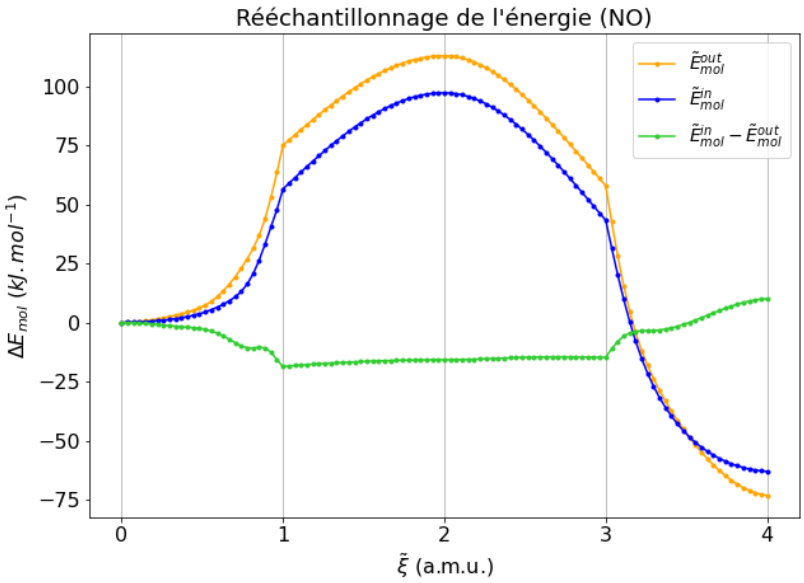


Figure 3.34 - Profils normalisés selon l'équation 3.1 et différences des profils « in » et « out » pour l'ouverture thermique du nitrosocyclobutène

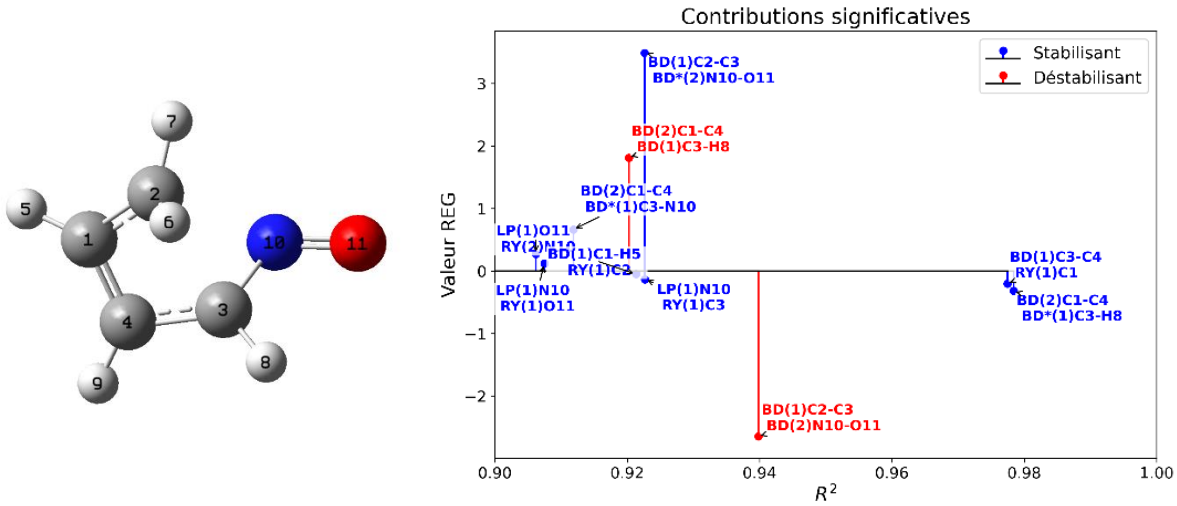


Figure 3.35 - Termes significatifs (R^2>0,90) issus de l'analyse REG-NBOdiff sur le premier segment de l'ouverture thermique du nitrosocyclobutène

Parmi les autres systèmes favorisant le chemin « in » pour lesquels cette explication est valide, on peut citer le diméthylborylcyclobutène (ΔE_{act} = -37,5 kJ.mol^{-1}). Contrairement aux systèmes précédemment cités, il ne favorise pas le chemin « in » sur le premier segment mais marque un rebroussement à la séparation entre les deux premiers segments.

Quant aux systèmes favorisant le chemin « out », le méthylcyclobutène (ΔE_{act} = 27,6 kJ.mol^{-1}) et le cyanocyclobutène (ΔE_{act} = 18,9 kJ.mol^{-1}) montrent le même terme E(2) significatif que l'aminocyclobutène. Il faut noter que pour tous les systèmes favorisant le chemin « out », la différence des profils favorise légèrement le chemin « in » sur le premier segment puis le chemin « out » sur le second comme sur la Figure 3.36 ci-dessous.

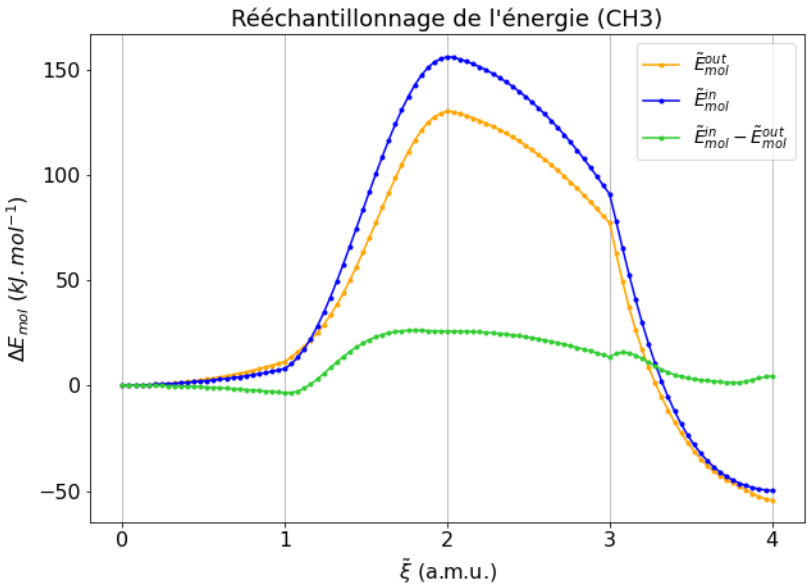


Figure 3.36 - Profils normalisés selon l'équation 3.7 et différences des profils « in » et « out » pour l'ouverture thermique du méthylcyclobutène

Pour le méthylcyclobutène (voir Figure 3.37 ci-après), on observe sur le second segment la présence de deux interactions E(2) inverses, entre une NBO d'une liaison C-H du méthyle et une NBO de la liaison C_2-C_3. L'une d'elles favorise le chemin « out » (la NBO liante C-H avec l'antiliante C_2-C_3, le même terme que pour l'aminocyclobutène) et l'autre (la NBO antiliante C-H avec la liante C_2-C_3, un terme équivalent à celui qui ressort pour le borylcyclobutène) le chemin « in ».

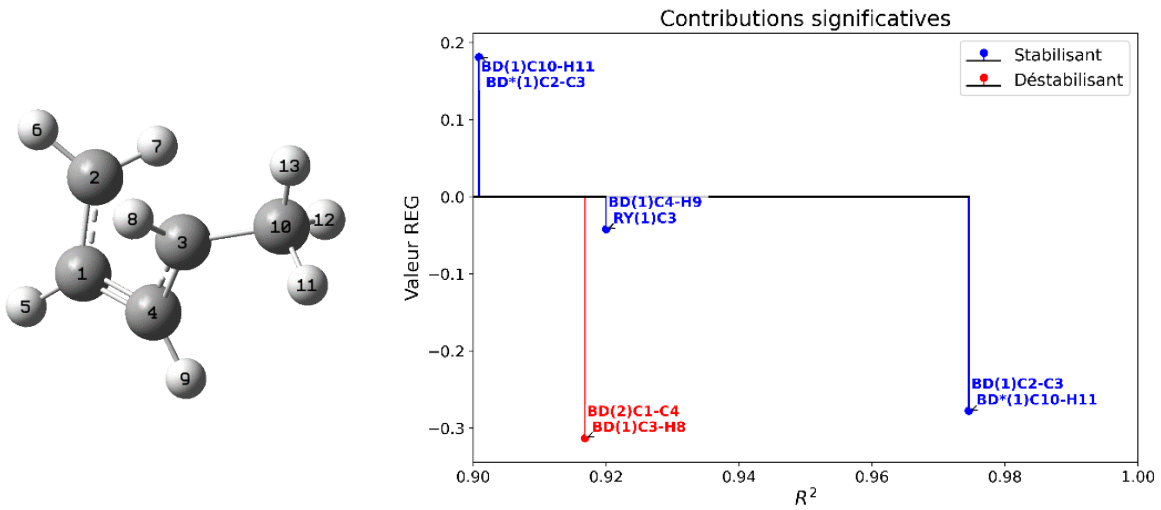


Figure 3.37 - Termes significatifs (R^2>0,90) issus de l'analyse REG-NBOdiff sur le second segment de l'ouverture thermique du méthylcyclobutène

On peut donc conclure pour ces quelques systèmes que l'explication orbitalaire habituelle est confirmée, bien que d'autres effets soient également présents et jouent sur la torquosélectivité.

3.4.3.3 Systèmes différant des systèmes extrêmes

Cependant, ces explications ne fonctionnent pas pour tous les systèmes, que ce soit ceux qui favorisent le chemin « in » ou le chemin « out ». Pour la première catégorie, les systèmes silylés (silyl- et trifluorosilylcyclobutène) et le dichloroborylcyclobutène ne présentent pas de terme mettant en jeu une orbitale de la liaison C_2-C_3.

Le silylcyclobutène (-SiH_3, ΔE_{act} = -5,2 kJ.mol^{-1}) a un profil similaire au nitrosocyclobutène, avec un rebroussement de la différence des profils énergétique à la division entre les deux premiers segments. Cependant, sur le premier segment, très peu de termes possèdent une corrélation assez bonne. Sur le second segment, le seul terme favorisant le chemin « in », impliquant le substituant et ayant une bonne corrélation est l'interaction E(2) entre la seconde NBO antiliante entre les C_1 et C_4 (la double liaison dans le cycle) et la NBO liante entre le C_3 et l'atome de silicium.

Pour le trifluorosilylcyclobutène (-SiF_3, ΔE_{act} = -16,1 kJ.mol^{-1}), on trouve un profil assez différent : le premier segment est assez plat puis favorise le chemin « in », puis le profil redevient plat. Sur le premier

segment, le terme majeur est l'interaction E(4) entre une paire libre d'un atome de fluor et la NBO liante entre le C_3 et l'atome de silicium. Sur le second, les interactions NBO ne sont pas bien corrélées.

Pour le dichloroborylcyclobutène (-BCl_2, ΔE_{act} = -28,0 kJ.mol^{-1}), le premier segment favorise légèrement le chemin « out » (alors que ce système présente une sélectivité plus élevée !) puis marque un rebroussement, comme pour les systèmes favorisant le chemin « out ». On trouve bien un terme favorisant le chemin « in » impliquant une orbitale de la liaison C_2-C_3, bien que la seconde NBO soit celle de la liaison C_3-bore.

La majorité des composés favorisant le chemin « out » ne présente pas les mêmes termes que l'aminocyclobutène : même le composé le plus proche, l'ion cyclobutylammonium (-NH_3^+, ΔE_{act} = 26,6 kJ.mol^{-1}) ne présente que le terme E(4) de l'explication habituelle.

Pour les autres composés, aucun de ces termes n'est présent. Par exemple, le nitrocyclobutène (-NO_2, ΔE_{act} = 23,4 kJ.mol^{-1}) ne montre aucune interaction entre une NBO des carbones 2 et 3 et une NBO du substituant sur le second segment (cf. Figure 3.38).

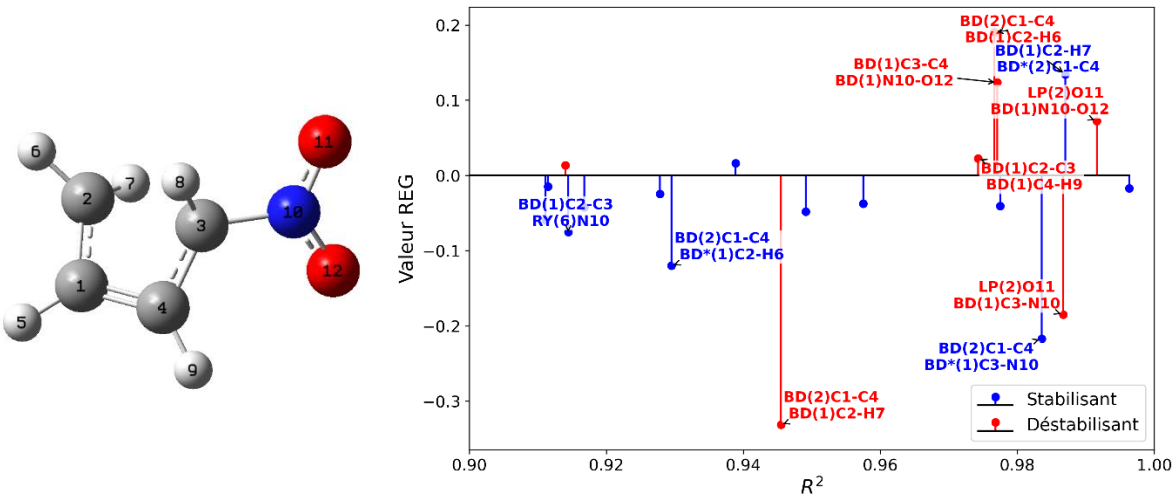


Figure 3.38 - Termes significatifs (R^2>0,90) issus de l'analyse REG-NBOdiff pour le second segment de la réaction d'ouverture thermique du nitrocyclobutène (normalisation via l'équation 3.7)

Pour ce composé, les termes les plus importants favorisant le chemin « out » sont en majorité des termes d'interaction à l'intérieur du cycle ou à l'intérieur du substituant. Seul le terme E(4) d'interaction entre la NBO de la liaison C_3-C_4 et la NBO liante entre l'azote et O_{12} relie le cycle et le substituant. En

revanche, on trouve deux termes favorisant le chemin « in », un terme E(2) (interaction de la NBO de la liaison double C_1-C_4 avec la NBO antiliante entre le C_3 et l'azote) et un terme E(4) (interaction d'une NBO paire libre de O_{11} et de la NBO de la liaison entre le C_3 et l'azote).

On retrouve des termes impliquant le substituant et la liaison C_3-C_4 pour le chlorocyclobutène (-Cl, ΔE_{act} = 58,7 kJ.mol^{-1}). En revanche, pour le fluorocyclobutène (-F, ΔE_{act} = 59,9 kJ.mol^{-1}), qu'on pourrait supposer être similaire, on ne trouve aucun terme impliquant le substituant favorisant le chemin « out ». Pour le phosphinocyclobutène (-PH_2, ΔE_{act} = 19,2 kJ.mol^{-1}) et le borohydrure de cyclobutyle (-BH_3^-, ΔE_{act} = 28,7 kJ.mol^{-1}), les corrélations des termes sont trop mauvaises et aucun terme n'a un R^2 supérieur à 0,90.

Il semble donc que l'on arrive bien à expliquer la torquosélectivité d'une partie des systèmes grâce à l'approche REG-NBOdiff, bien que l'approche montre également que le reste des systèmes a une réactivité différente qui ne peut pas être rationalisée par seulement un ou deux termes. Pour ces systèmes, il est probable que la torquosélectivité soit multifactorielle et que plusieurs petits facteurs mis bout-à-bout puissent l'expliquer. Dans ce but, on pourrait imaginer utiliser des régressions bi-, tri- ou multilinéaires, ce qui pourrait être une extension future de REG.

3.4.4 Étude REG-IQAdiff de la différence des profils

3.4.4.1 Systèmes extrêmes

Qu'en est-il du côté de l'analyse REG-IQAdiff ? Reprenons le cas du borylcyclobutène, dont l'analyse REG-IQAdiff du premier segment est représentée par la Figure 3.39 ci-après.

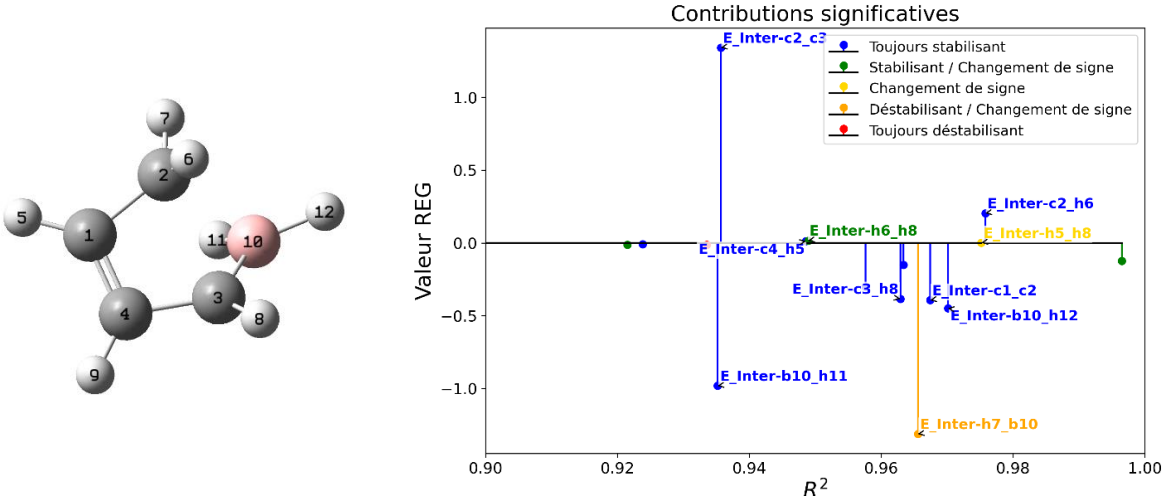


Figure 3.39 – Termes significatifs (R^2>0,90) issus de l'analyse REG-IQAdiff pour le premier segment de la réaction d'ouverture thermique du borylcyclobutène

On constate qu'il n'y a pas de terme impliquant le substituant favorisant le chemin « in », seulement un terme d'interaction entre H_7 et le bore favorisant le chemin « out ». Une analyse plus détaillée indique que le terme de plus haut coefficient REG favorisant le chemin « in » est un terme d'interaction entre C_2 et le bore, mais qu'il possède un R^2 de 0,85. Ce terme et l'interaction C_2-C_3 sont analogues à l'interaction NBO entre la liaison C_2-C_3 et la lacune du bore, ce qui est déjà un bon signe quant à un possible parallèle entre les analyses NBO et IQA, même s'ils ne dominent pas les autres de façon aussi écrasante qu'en NBO. Le terme d'interaction C_2-B a en revanche une meilleure corrélation sur le second segment (cf. Figure 3.40).

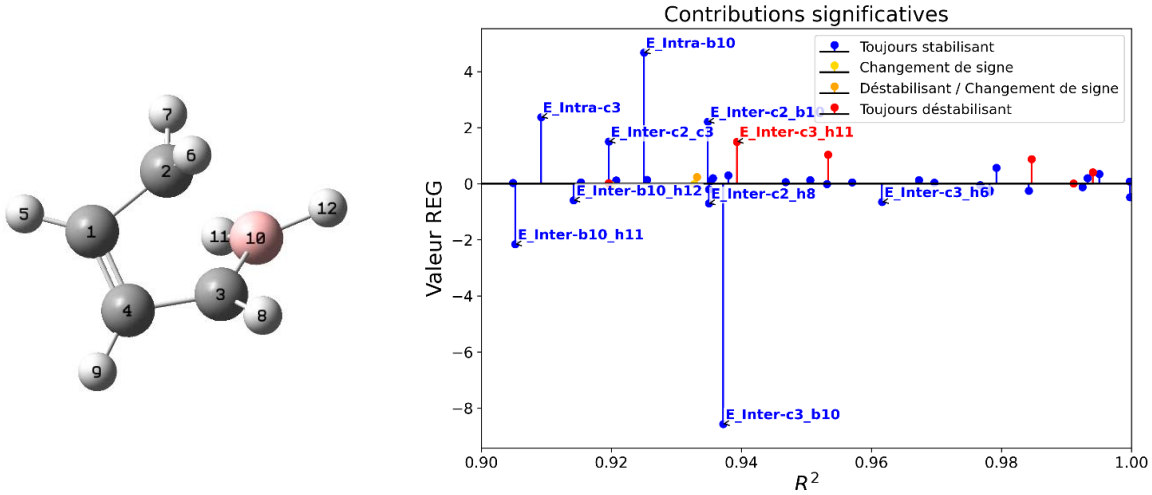
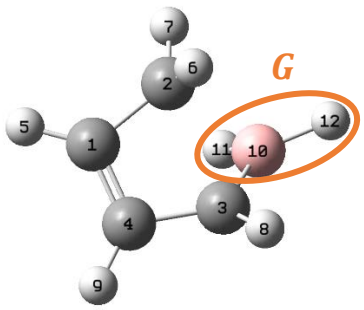


Figure 3.40 - Termes significatifs (R^2>0,90) issus de l'analyse REG-IQAdiff pour le second segment de la réaction d'ouverture thermique du borylcyclobutène

Cependant, ce n'est pas le terme le plus important, et le terme d'interaction C_3-B (qui favorise le chemin « out », cette fois) est nettement plus important, avec un coefficient REG presque cinq fois supérieur en valeur absolue.

On peut se demander si ces résultats ne peuvent pas être améliorés en groupant ensemble les termes du substituant : en effet, peut-être que tous les atomes du substituant jouent un rôle plutôt qu'uniquement le bore. Les termes IQA étant additifs, il est possible de simplement sommer les termes des différents atomes. On rappelle (cf. chapitre 2 section 2.5.2) l'exemple du groupement des atomes du substituant boryle (G1 = B_{10} + H_{11} + H_{12}) :

	$E_{intra}(G) = E_{intra}(B_{10}) + E_{intra}(H_{11}) + E_{intra}(H_{12})$ $E_{inter}(G,G) = E_{inter}(B_{10},H_{11}) + E_{inter}(H_{11},H_{12})$ $+ E_{inter}(B_{10},H_{12})$ Pour tout atome A n'appartenant pas à G : $E_{inter}(A,G) = E_{inter}(A,B_{10}) + E_{inter}(A,H_{11})$ $+ E_{inter}(A,H_{12})$

Ce groupement donne le résultat illustré en Figure 3.41 ci-dessous.

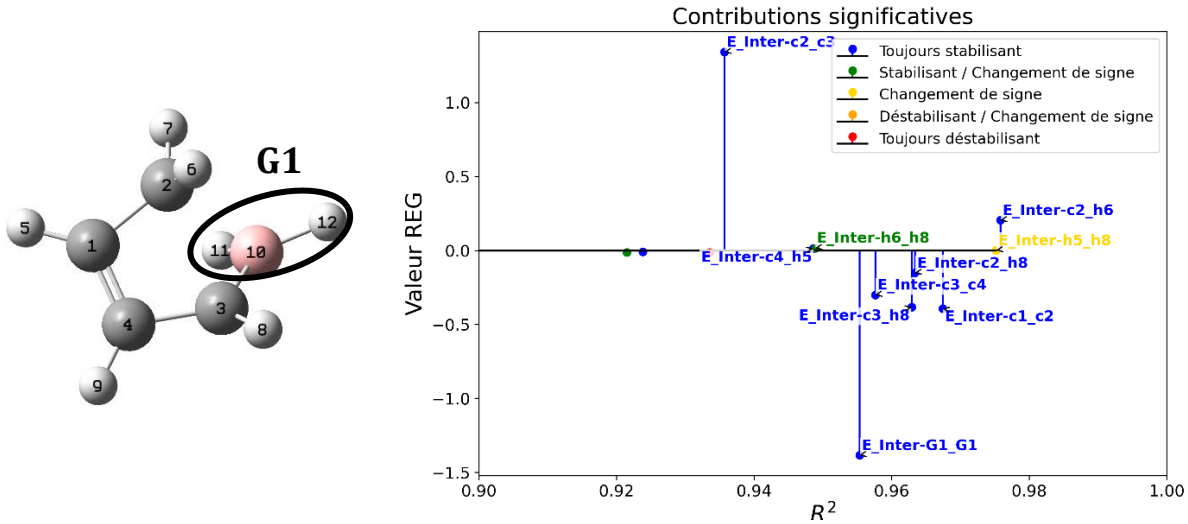


Figure 3.41 - Termes significatifs (R^2>0,90) issus de l'analyse REG-IQAdiff pour le premier segment de la réaction d'ouverture thermique du borylcyclobutène (G1 = BH_2). Normalisation des segments suivant l'équation 3.1.

On retrouve le terme d'interaction C_2-C_3 favorisant le chemin « in », et cette fois, un terme d'interaction à l'intérieur du groupe G1 qui favorise le chemin « out ». Le terme d'interaction C_2-G1 est toujours hors du cadre avec un R^2 de 0,87, mais favorise toujours le chemin « in » et reste le terme de coefficient REG le plus important.

Ces termes (C_2-C_3, C_2-G1) peuvent être rapprochés de l'interaction principale en REG-NBOdiff, entre la NBO liante C_2-C_3 et la lacune du bore.

Quant à l'aminocyclobutène, les termes d'interaction C_3-G1, G1-G1 et C_2-C_3 et le terme intra-atomique du groupe G1 ressortent sur le second segment : les trois premiers favorisent le chemin « out » et le dernier le chemin « in ». En revanche, de façon surprenante, le terme d'interaction C_2-G1 a un coefficient REG très faible et favorise le chemin « out » comme illustré en Figure 3.42.

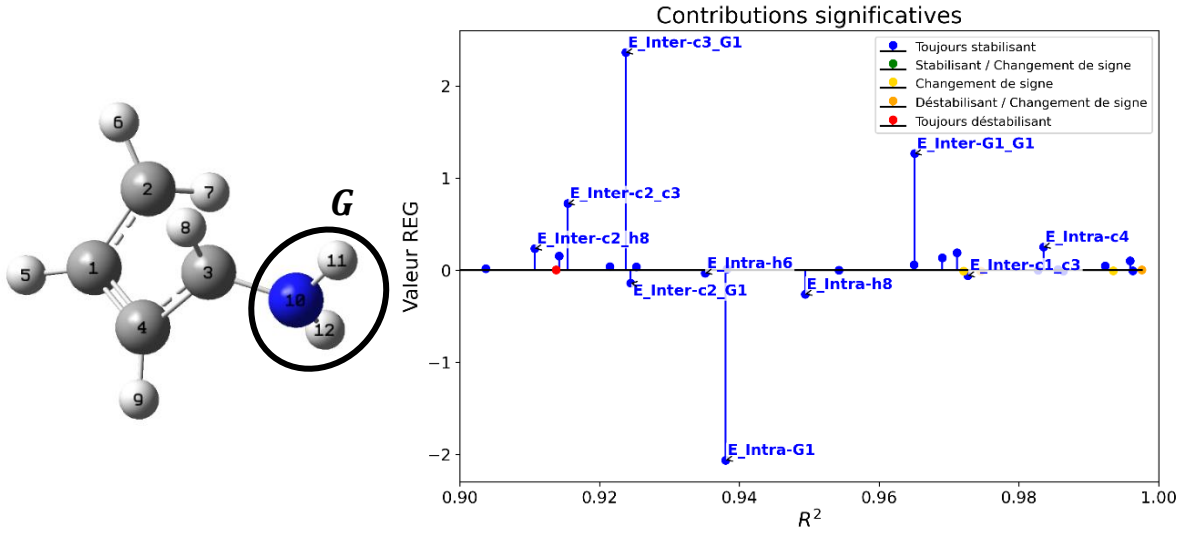


Figure 3.42 - Termes significatifs (R^2>0,90) issus de l'analyse REG-IQAdiff pour le second segment de la réaction d'ouverture thermique de l'aminocyclobutène (G1 = NH_2). Normalisation des segments suivant l'équation 3.8.

On trouve à nouveau des termes IQA proches des termes significatifs en REG-NBOdiff (bien que cette fois l'interaction C_2-G1 ne soit pas prépondérante).

Voyons désormais si cette explication se retrouve pour les autres systèmes.

3.4.4.2 Systèmes ressemblant aux cas extrêmes

Pour certains autres composés favorisant le chemin « in », comme le carbonylcyclobutène, on retrouve le même terme que le borylcyclobutène en groupant les atomes du substituant (cf. Figure 3.43).

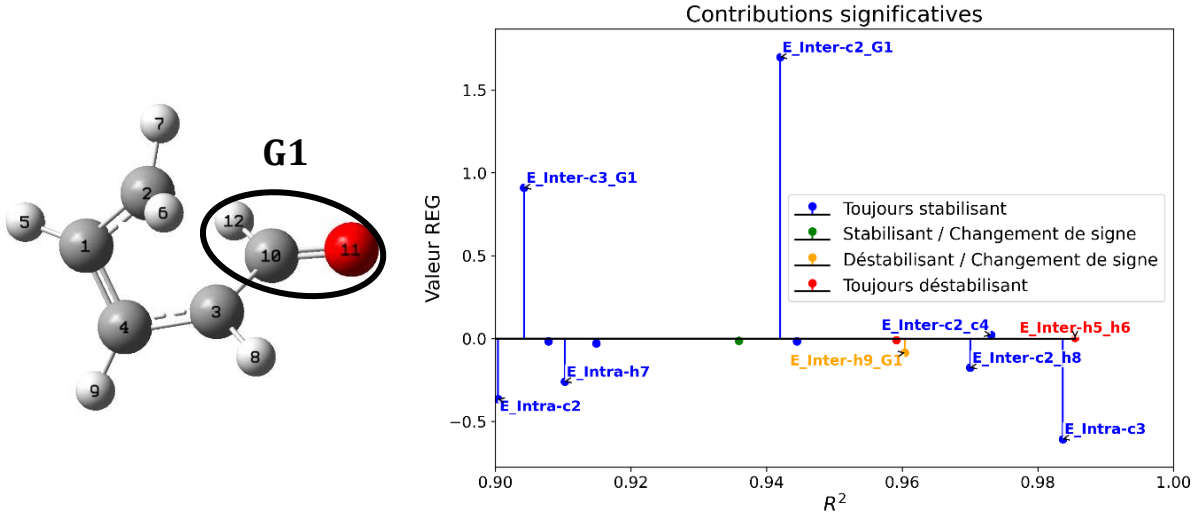


Figure 3.43 - Termes significatifs (R^2>0,90) issus de l'analyse REG-IQAdiff pour le premier segment de la réaction d'ouverture thermique du carbonylcyclobutène (G1 = CHO). Normalisation des segments suivant l'équation 3.1.

Cette fois, le terme Inter-C_2-G1 domine largement plus et a une meilleure corrélation, en plus de favoriser le chemin « in ». On observe également que le terme Inter-C_3-G1 favorise le chemin « in ». En revanche, pas de terme d'interaction C_2-C_3.

Même chose pour le nitrosocyclobutène, avec seulement le terme Inter-C_2-G1 (cf. Figure 3.44).

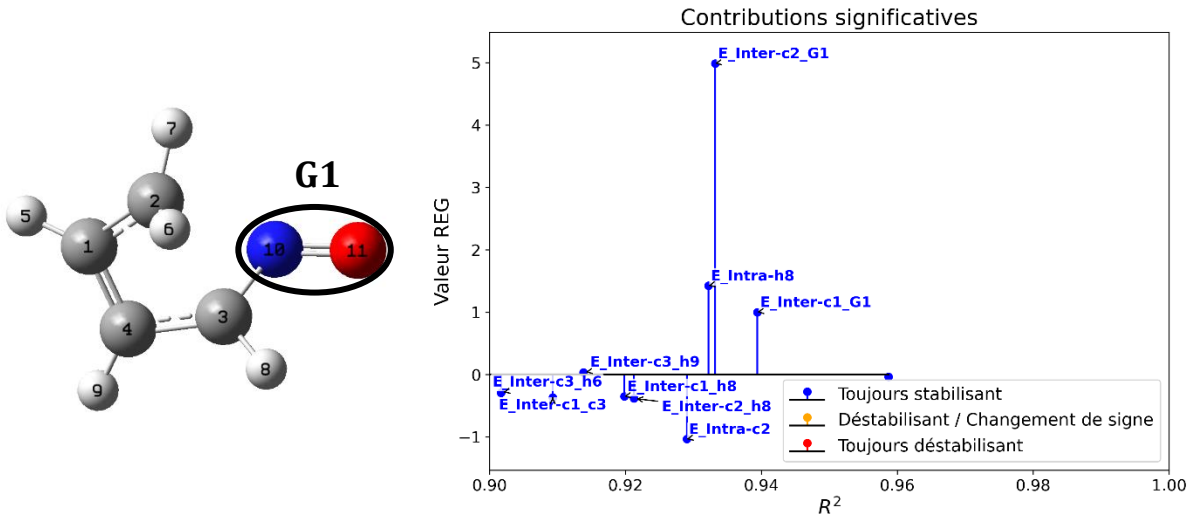


Figure 3.44 - Termes significatifs (R^2>0,90) issus de l'analyse REG-IQAdiff pour le premier segment de la réaction d'ouverture thermique du nitrosocyclobutène (G1 = NO). Normalisation des segments suivant l'équation 3.1.

On trouve également ce terme pour le diméthylborylcyclobutène. L'explication reste donc à peu près la même que pour le borylcyclobutène pour ces trois composés.

Du côté des composés favorisant le chemin « out », on peut trouver certains des termes de l'aminocyclobutène pour plusieurs composés, dont le chlorocyclobutène (voir Figure 3.45).

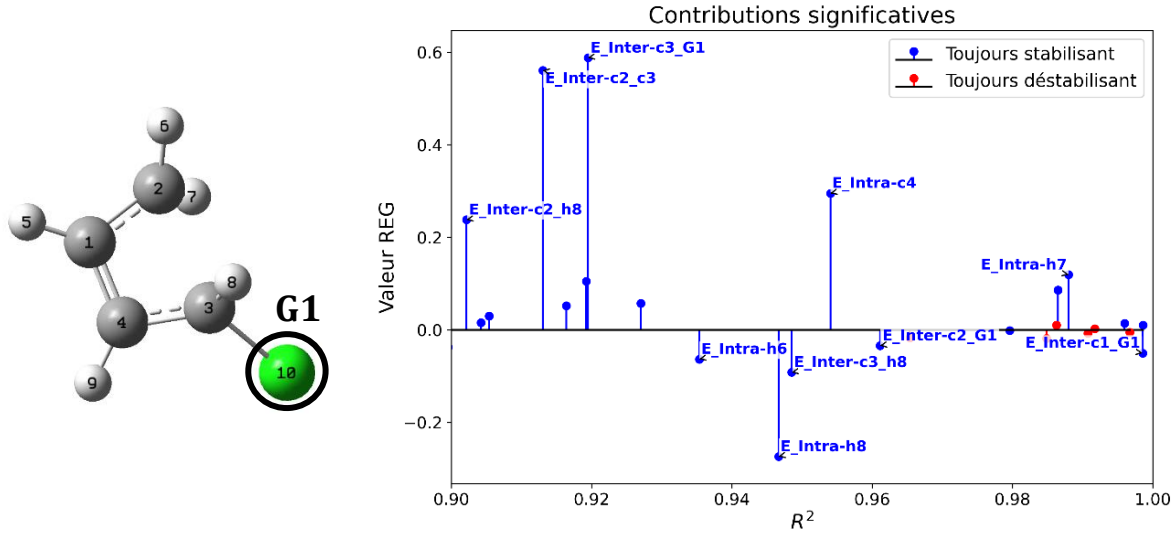


Figure 3.45 - Termes significatifs (R^2>0,90) issus de l'analyse REG-IQAdiff pour le second segment de la réaction d'ouverture thermique du chlorocyclobutène (G1 = Cl). Normalisation des segments suivant l'équation 3.8.

Les termes C_2-C_3 et C_3-G1 ont à peu près la même valeur de coefficient REG tandis que le terme C2-G1 a un coefficient REG faible comme pour l'aminocyclobutène.

On retrouve la même chose pour le borohydrure de cyclobutène avec le terme C_3-G1 ; le terme C_2-C_3 a un coefficient REG plus faible mais favorise tout de même le chemin « out ».

En revanche, on ne trouve plus que le terme C_2-C_3 pour le nitrocyclobutène (cf. Figure 3.46).

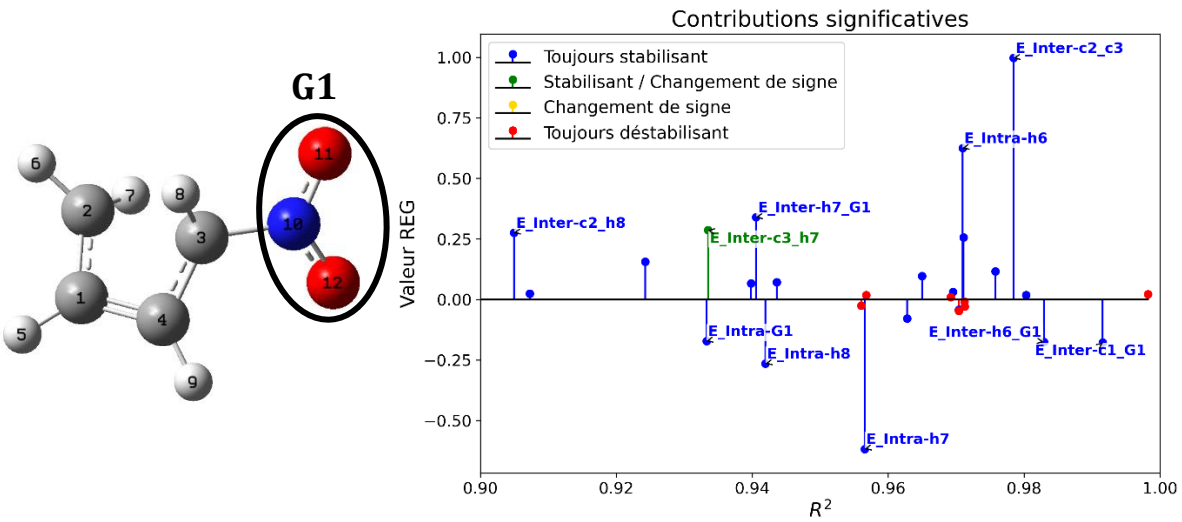


Figure 3.46 - Termes significatifs (R^2>0,90) issus de l'analyse REG-IQAdiff pour le second segment de la réaction d'ouverture thermique du nitrocyclobutène (G1 = NO_2). Normalisation des segments suivant l'équation 3.8.

Le terme C_3-G1 favorise en fait bien le chemin « out » mais n'a pas une bonne corrélation (R^2 = 0,87).

Les termes C_2-C_3 et C_3-G1 peuvent donc expliquer la torquosélectivité de certains composés favorisant le chemin « out » avec différents degrés de certitude en raison de leurs corrélations variables avec l'énergie totale. Le terme C_2-G1 n'est en revanche jamais discriminant.

3.4.4.3 Systèmes différant des cas extrêmes

En revanche, pour la majorité des composés, les termes significatifs ne sont pas présents ou favorisent le mauvais chemin.

Pour les composés favorisant le chemin « in », seul le terme Intra-G1 favorise le bon chemin pour dichloroboryl-, le silyl- et le trifluorosilylcyclobutène. Les termes C_2-C_3 et C_3-G1 favorisent, eux, le

chemin « out ». On pourrait supposer que cela explique la sélectivité moindre de ces systèmes, bien que le carbonyl- et le nitrosocyclobutène aient une sélectivité inférieure à deux de ces systèmes.

Pour les composés favorisant le chemin « out », la situation est plus complexe. Le phosphinocyclobutène ($-PH_2$), le cyanocyclobutène (-CN) et l'ion cyclobutylammonium ($-NH_3^+$) ne présentent aucun terme du substituant favorisant le chemin « out ». Pour le méthylcyclobutène, seul le terme Intra-G1 favorise le chemin « out » comme illustré sur la Figure 3.47.

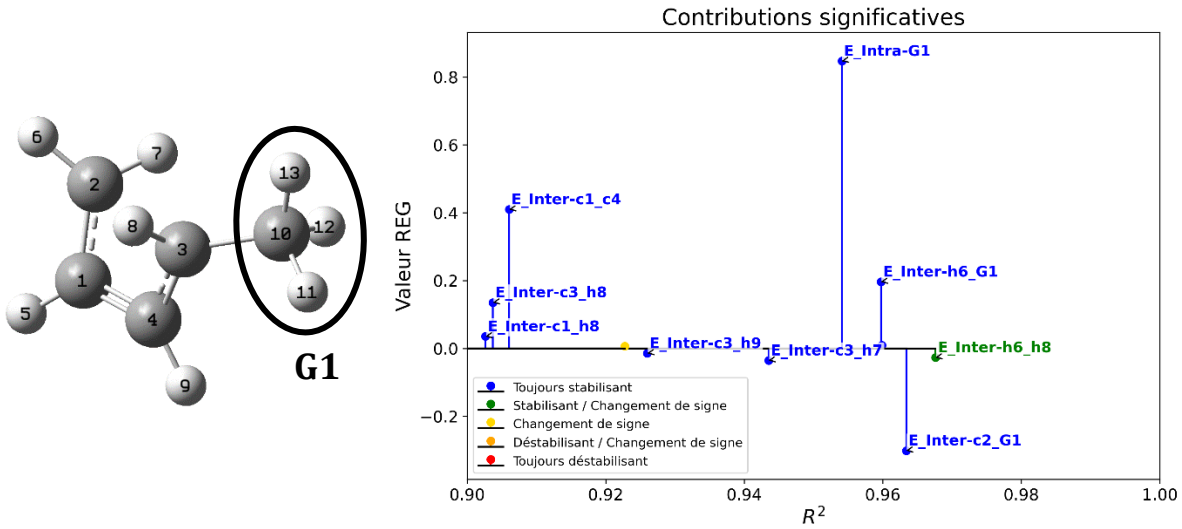


Figure 3.47 - Termes significatifs (R^2>0,90) issus de l'analyse REG-IQAdiff pour le second segment de la réaction d'ouverture thermique du méthylcyclobutène (G1 = CH_3). Normalisation des segments suivant l'équation 3.8.

Le terme C3-G1 a quant à lui une mauvaise corrélation (R^2 = 0,71) tandis que les termes C_2-G1 et C_2-C_3 favorisent tous deux le mauvais chemin,

Enfin, pour le fluorocyclobutène, seul le terme C_2-C_3 favorise le chemin « out ».

On a ainsi pu établir plusieurs parallèles entre l'analyse NBO et IQA, bien que les résultats ne soient pas simples à interpréter pour tous les systèmes. Selon les cas, les termes d'interaction IQA C_2-C_3, C_2-G1 et/ou C_3-G1 peuvent être rapprochés des termes d'interaction NBO entre les NBO liante ou antiliante entre les C_2-C_3 et les NBO vacantes ou occupées sur le substituant. La piste de régressions multilinéaires est envisagée pour mettre en lumière des effets agissant de concert ou de façon antagoniste pour les systèmes plus complexes.

Bibliographie du chapitre 3

1. E. N. Marvell. *Thermal electrocyclic reactions* (Academic Press, 1980).
2. R. B. Woodward & R. Hoffmann. *J. Am. Chem. Soc.* **87**, 395–397 (1965).
3. W. Kirmse, N. G. Rondan & K. N. Houk. *J. Am. Chem. Soc.* **106**, 7989–7991 (1984).
4. K. Rudolf, D. C. Spellmeyer & K. N. Houk. *J. Org. Chem.* **52**, 3708–3710 (1987).
5. C. W. Jefford, G. Bernardinelli, Y. Wang, D. C. Spellmeyer, A. Buda & K. N. Houk. *J. Am. Chem. Soc.* **114**, 1157–1165 (1992).
6. S. Niwayama, E. A. Kallel, D. C. Spellmeyer, C. Sheu & K. N. Houk. *J. Org. Chem.* **61**, 2813–2825 (1996).
7. M. Murakami, I. Usui, M. Hasegawa & T. Matsuda. *J. Am. Chem. Soc.* **127**, 1366–1367 (2005).
8. S. Niwayama. *J. Org. Chem.* **61**, 640–646 (1996).
9. M. Murakami, Y. Miyamoto & Y. Ito. *Angew. Chemie Int. Ed.* **40**, 189–190 (2001).
10. M. Murakami & M. Hasegawa. *Angew. Chemie Int. Ed.* **43**, 4874–4876 (2004).
11. M. Hasegawa & M. Murakami. *J. Org. Chem.* **72**, 3764–3769 (2007).
12. M. Hasegawa, I. Usui, S. Konno & M. Murakami. *Org. Biomol. Chem.* **8**, 4169–4175 (2010).
13. L. Akilandeswari & C. Prathipa. *J. Chem. Sci.* **127**, 1505–1511 (2015).
14. A. Saal & L. Merzoud. *J. Comput. Methods Sci. Eng.* **16**, 437–447 (2016).
15. K. Saadat, R. Villar López, A. Shiri, O. Nieto Faza & C. Silva López. *Org. Biomol. Chem.* **18**, 6287–6296 (2020).
16. O. Larrañaga & A. Cózar. *ChemPhysChem* **21**, 1805–1813 (2020).
17. G. I. Cárdenas-Jirón, S. Gutiérrez-Oliva, J. Melin & A. Toro-Labbé. *J. Phys. Chem. A* **101**, 4621–4627 (1997).
18. A. Toro-Labbé. *J. Phys. Chem. A* **103**, 4398–4403 (1999).
19. ***NBO 7.0.*** E. D. Glendening, J, K. Badenhoop, A. E. Reed, J. E. Carpenter, J. A. Bohmann, C. M. Morales, P. Karafiloglou, C. R. Landis, and F. Weinhold, Theoretical Chemistry Institute, University of Wisconsin, Madison WI (2018).
20. A. D. Becke. *J. Chem. Phys.* **98**, 5648–5652 (1993).
21. S. Grimme, J. Antony, S. Ehrlich & H. Krieg. *J. Chem. Phys.* **132**, 154104 (2010).
22. ***Gaussian 09 revision D.01***. 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, and D. J. Fox, Gaussian, Inc., Wallingford CT (2013).

23. ***AIMAll (Version 19.10.12)***, Todd A. Keith, TK Gristmill Software, Overland Park KS, aim.tkgristmill.com (2019).

24. P. S. Lee, X. Zhang & K. N. Houk. *J. Am. Chem. Soc.* **125**, 5072–5079 (2003).

Chapitre 4 – Étude théorique des additions d'aza-Michael

4.1 Introduction 132

4.2 Article 137

4.3 Conclusion 165

4.1 Introduction

Parmi les grands types de réactions de la chimie organique, les réactions entre nucléophiles et électrophiles occupent une place prépondérante. Pour comprendre les facteurs physico-chimiques gouvernant ces réactions, on peut par exemple évaluer la réactivité de chaque réactif, ce qui revient à mesurer leur pouvoir nucléophile et électrophile. On peut également s'intéresser aux variations des propriétés des réactifs et de leurs interactions au cours de la réaction : charges atomiques, descripteurs de DFT conceptuelle, termes IQA... La compréhension de ces facteurs pourrait *in fine* permettre de concevoir *in silico* des molécules de sorte à obtenir des propriétés désirées.

L'électrophilie, tout comme la nucléophilie, sont des termes proposés par Ingold en 1934 [1] pour qualifier les molécules ayant tendance à accepter ou donner des électrons.

On peut ainsi définir l'électrophilie, ou pouvoir électrophile, comme la réactivité relative d'un réactif électrophile [2]. La nucléophilie se définit similairement comme la réactivité relative d'un réactif nucléophile. Ces concepts à l'origine empirique ont par la suite été mesurés expérimentalement de différentes façons, notamment grâce aux potentiels d'oxydo-réduction et aux constantes de vitesse.

Plusieurs modèles ont été proposés, entre autres celui de Swain et Scott [3] pour les réactions entre un nucléophile et un électrophile halogéné, celui de Ritchie [4] pour des nucléophiles et cations organiques stables et celui de Mayr pour tout type de nucléophile et d'électrophile organique.

Le modèle de Mayr proposé en 1994 [5] repose sur l'équation suivante :

$$\log k_{20°C} = s_N(E + N) \tag{4.1}$$

Dans cette équation, $k_{20°C}$ est la constante de vitesse d'une réaction à 20 °C entre un électrophile et un nucléophile donnés, s_N et N sont respectivement la sensibilité spécifique et le paramètre de nucléophilie du nucléophile, et E est le paramètre d'électrophilie de l'électrophile.

Il est ainsi possible de classer quantitativement des électrophiles et des nucléophiles sur une échelle de réactivité.

Les références utilisées pour construire cette échelle sont un ensemble de 29 électrophiles permettant de couvrir une large gamme de réactivités, avec l'ion bis(p-méthoxyphényl)carbénium possédant une électrophilie E définie comme égale à 0 et l'allyltriméthylsilane possédant une sensibilité s_N définie comme égale à 1.

Les valeurs de ces paramètres pour près de 1 600 composés de diverses familles sont disponibles sur une base de données en ligne [6]. Il a été observé expérimentalement que l'on peut ainsi prédire la faisabilité de la réaction entre un électrophile et un nucléophile donné en additionnant leurs paramètres E et N. Si la somme $E + N$ a une valeur supérieure à -5, la réaction est dite faisable [7].

Cependant, cette approche possède plusieurs limites : premièrement, une réaction dite faisable ne garantit pas que le produit formé sera forcément stable ni qu'il ne se réarrangera pas ensuite. Deuxièmement, l'échelle n'est applicable que pour la formation d'une seule liaison σ : il est alors possible que la réaction soit beaucoup plus rapide et suive un autre mécanisme, par exemple en passant par un mécanisme concerté.

Il est tentant de relier cette échelle de réactivité expérimentale aux descripteurs de la DFT conceptuelle définis dans le chapitre 2. On peut dans un premier temps envisager d'étudier les propriétés intrinsèques des réactifs ; un lien entre l'électrophilie et ces propriétés permettrait la prédiction aisée de la réactivité de composés organiques sans avoir à mener des études cinétiques théoriques ou expérimentales et ainsi de faciliter la tâche des chimistes de synthèse.

Des articles [8–10] se sont déjà attachés à la prédiction de l'électrophilie des cations benzhydrilium, les références de l'échelle de Mayr, mais peu d'études portent sur les autres familles de composés. Notre groupe s'est donc intéressé à la prédiction de l'électrophilie d'accepteurs de Michael, définis dans l'introduction de cette thèse comme des composés α,β-insaturés possédant un groupement électroattracteur conjugué. 114 composés de cette famille sont désormais répertoriés sur la base de données de Mayr.

Au cours de la thèse du Dr Guillaume Hoffmann, trois études de notre groupe sur les additions de Michael ont été publiées. La première étude [11] portait sur un échantillon représentatif de 35

électrophiles choisis dans la base de données de Mayr de sorte à représenter une large gamme d'électrophilies et de familles d'accepteurs. Les corrélations entre les descripteurs de réactivité de ces composés isolés et l'électrophilie de Mayr ont été testées. Sa conclusion était qu'un seul descripteur ne pouvait pas prédire l'électrophilie de l'intégralité des composés, mais pouvait suffire pour une famille restreinte. Une régression bilinéaire parvenait à un résultat plus satisfaisant (R^2 = 0,87) pour les 35 composés.

Un second article [12] visait à améliorer les résultats obtenus en étudiant un ensemble plus restreint de 11 composés parmi les 35 précédemment étudiés par dynamique moléculaire semi-empirique afin d'observer les effets des variations de structure sur les descripteurs et parvient à un R^2 de 0,97, tandis qu'un troisième article [13] se penchait sur l'utilisation des méthodes de *machine learning* avec un ensemble de 111 électrophiles pour améliorer le modèle au-delà des régressions linéaires classiques et arrivait à un R^2 de 0,98 sans prendre en compte d'effets dynamiques.

Le présent chapitre et le suivant poursuivent l'étude des additions d'aza-Michael en s'intéressant cette fois au mécanisme et aux descripteurs le long du chemin réactionnel. Ces chapitres sont présentés sous forme de deux articles (soumis mais pas encore publiés).

Le premier article est consacré à l'étude de l'addition d'aza-Michael de la méthylamine à l'ensemble des 11 accepteurs de Michael choisis pour le second article susmentionné ; nous nous sommes intéressés à l'étude de descripteurs intrinsèques globaux et locaux afin de suivre les variations de ces propriétés le long du chemin réactionnel et de comprendre le déroulement de la réaction. Plus précisément, les multiples chemins du mécanisme complet sont détaillés, les constantes de vitesse des différentes étapes sont calculées et mises en relation avec l'électrophilie de Mayr des accepteurs, et les transferts de charge et les variations de plusieurs descripteurs de DFT conceptuelle le long du chemin de réaction sont étudiés pour l'étape d'addition afin de vérifier la validité de plusieurs principes de réactivité.

Le second article se consacre quant à lui à l'étude REG-IQA des interactions interatomiques entre réactifs et avec le solvant au cours de la première étape de la réaction. En raison de l'utilisation d'un modèle de solvant explicite, seuls 6 systèmes parmi les 11 du premier article ont pu être considérés pour

cet article. Les différents évènements de l'étape d'addition sont décrits, les interactions avec le solvant sont examinées, et un lien entre les termes IQA et l'électrophilie de Mayr via l'analyse REG est présenté.

Bibliographie

1. C. K. Ingold. *Chem. Rev.* **15**, 225–274 (1934).
2. A. D. Mc Naught & A. Wilkinson. *The IUPAC Compendium of Chemical Terminology The IUPAC Compendium of Chemical Terminology* (Blackwell Scientific Publicatons, 2019).
3. C. Gardner Swain & C. B. Scott. *J. Am. Chem. Soc.* **75**, 141–147 (1953).
4. C. D. Ritchie. *Acc. Chem. Res.* **5**, 348–354 (1972).
5. H. Mayr & M. Patz. *Angew. Chemie Int. Ed. English* **33**, 938–957 (1994).
6. Mayr's Database Of Reactivity Parameters: https://www.cup.lmu.de/oc/mayr/reaktionsdatenbank/fe/.
7. R. Lucius, R. Loos & H. Mayr. *Angew. Chem. Int. Ed* **41**, 91–95 (2002).
8. P. Pérez, A. Toro-Labbé, A. Aizman & R. Contreras. *J. Org. Chem.* **67**, 4747–4752 (2002).
9. V. Tognetti, C. Morell & L. Joubert. *Chem. Phys. Lett.* **635**, 111–115 (2015).
10. S. Liu, C. Rong & T. Lu. *J. Phys. Chem. A* **118**, 3698–3704 (2014).
11. G. Hoffmann, V. Tognetti & L. Joubert. *J. Mol. Model.* **24**, 281 (2018).
12. G. Hoffmann, V. Tognetti & L. Joubert. *Chem. Phys. Lett.* **724**, 24–28 (2019).
13. G. Hoffmann, M. Balcilar, V. Tognetti, P. Héroux, B. Gaüzère, S. Adam & L. Joubert. *J. Comput. Chem.* **41**, 2124–2136 (2020).

A theoretical study on aza-Michael additions

Aël CADOR[a], Guillaume HOFFMANN[b], Vincent TOGNETTI[a],* Laurent JOUBERT[a]*

[a] : Normandy Univ., COBRA UMR 6014 & FR 3038, Université de Rouen, INSA Rouen, CNRS, 1 rue Tesnière 76821 Mont St Aignan Cedex, France

[b] : MONARIS, UMR CNRS-Sorbonne Université 8233, 4 Place Jussieu, 75005 Paris, France

Abstract:

In this paper, we investigate aza-Michael additions, which are key-reactions in organic synthesis, from the theoretical point of view on a representative database collecting weak, medium and strong electrophiles. The full mechanism is studied in detail, focusing on the various elementary steps (creation of the C-N bond, proton transfer) and on the catalytic role of the amine reagent, through a comprehensive sampling of the potential energy surface. The computed activations barriers are then discussed in relation with the available experimental electrophilicity values. Furthermore, in order to unravel the main physico-chemical factors governing the reaction, charge transfers and descriptors from conceptual density functional theory are followed along the reaction path, casting light on the reactivity principles that are verified, and finally affording a unified view of these important chemical processes.

Keywords: Michael additions, reaction mechanisms, reactivity principles, conceptual DFT, quantum theory of atoms-in-molecules

Corresponding authors: vincent.tognetti@univ-rouen.fr, laurent.joubert@univ-rouen.fr

1. Introduction

The Michael reaction is one of the oldest and most widely used reactions in organic chemistry[1,2]. It notably allows the formation of a carbon-carbon bond without resorting to the use of metallic catalysts. As valuable variants enlarging the initial scope, many hetero-Michael additions[3,4] have been developed over the years, notably the aza-Michael reaction (aMR)[5], which links a nitrogen atom to the electrophilic carbon. This naturally includes amines, but heterocycles[6], amides[7], azides[8], oximes[9] have also been shown to react in aMR. Asymmetric versions of the aMR have been designed as well[10], affording a versatile toolbox in organic synthesis.

Having a better understanding of how this reaction proceeds in details can not only allow its extension to other systems, the prediction of its regioselectivity and a rationale for the choice of the optimal reagents, but it can also provide deeper chemical insight useful for other chemical transformations as well.

From the experimental viewpoint, Mayr's electrophilicity scale[11], which classifies electrophiles and nucleophiles according to their reaction rate, affords an empirical way to predict the feasibility of reactions between a given electrophile and a given nucleophile. The resulting parameters have been published in the form of an online, freely accessible database[12], currently comprising parameters for more than 1,600 compounds. Noteworthy, several electrophiles in this database are suitable for Michael reactions and thus are coined *Michael acceptors* in Mayr's website.

In a series of recent papers[13–15], we have reported on a reliable theoretical protocol to predict Mayr's electrophilicity through density functional theory (DFT)[16,17] calculations and, more particularly, through the use of conceptual DFT (CDFT)[18–21] and Bader's quantum theory of atoms-in-molecules (QTAIM)[22,23] descriptors. Such work was actually motivated by a double goal: identifying rationalizing tools and establishing a systematic theoretical method suitable also to molecules for which Mayr's experimental protocol has not yet been applied.

In the present study, we intend to complement our former analysis by investigating the ***mechanistic*** aspects of Michael additions, in contrast with our previous papers in which only the ***intrinsic*** properties

of reactants were scrutinized. One can mention an inspiring similar work (but mainly within the complementary natural bonding orbital (NBO) point of view) on thio-Michael reactions by Giraldo et al[24].

To this aim, this paper will be divided as follows: in section 2, the basics of CDFT will be briefly reviewed and the associated reactivity principles will be presented in a self-explanatory fashion, while section 3 will describe the computational protocol. Section 4 will then present the main results of our calculations.

First, section 4.1 will describe the full reaction mechanism through the comprehensive exploration of the potential energy surface (PES) for 11 selected representative aza-Michael acceptors, with a particular emphasis on the competing pathways and on the possible intermediates and transition states they involve. Reaction rates can afterwards be estimated, and will be related to Mayr's experimental values in section 4.2. Then we will focus only on the most favoured reaction pathway and will follow in 4.3 and 4.4 the variations of key-CDFT descriptors, in the spirit of refs[25–30], along its first step (the C-N bond formation) to unravel the nature of the reaction and to determine to what extent well-known reactivity principles apply or not for these reactions.

Indeed, although some studies have already discussed the applicability and criticised the limits of such principles[31–34], there is still a need, from our point of view, to understand why some principles are applicable to certain families of reactions and why others are not. Since they have not been yet applied to aza-Michael reactions thus far, their applicability to aMR will be assessed.

The paper will finally end with conclusions and perspectives.

2. Conceptual DFT descriptors and reactivity principles

2.1. CDFT descriptors

As already mentioned, CDFT will be invoked because it has arisen as a rigorous theory of chemical reactivity, bridging empirical chemical concepts and quantum chemistry[20,21,35]. In its canonical

formulation, its starting point is the electronic energy derivatives with respect to the number of electrons N and/or the external potential $v(\mathbf{r})$ generated by the nuclei, from which one can straightforwardly identify meaningful chemical concepts such as electronegativity χ, chemical hardness η, or chemical softness S. These quantities can be subsequently used to describe and compare chemical systems, hence their name of descriptors.

CDFT descriptors can be mainly divided into three categories: global, local and non-local ones. Global descriptors have a single value for a given system and do not depend on spatial coordinates. They characterize a molecule as a whole and are useful to quickly compare two molecules. For instance, the chemical potential[18] μ (also the opposite of the CDFT electronegativity χ) is defined as the first derivative of the total electronic energy of the system (in the so-called canonical representation, $E_e[N, v(\mathbf{r})]$) with respect to N at constant $v(\mathbf{r})$, while the chemical hardness[19] η, modelling the resistance to electron transfer, is the second derivative and the reciprocal of molecular softness S:

$$\mu = \left(\frac{\partial E_e}{\partial N}\right)_{v(\mathbf{r})} = -\chi, \quad \eta = \left(\frac{\partial^2 E_e}{\partial N^2}\right)_{v(\mathbf{r})} = 1/S. \tag{1}$$

Parr's electrophilicity index[36–38] ω is then defined by:

$$\omega = \frac{\mu^2}{2\eta}. \tag{2}$$

Contrariwise, local descriptors depend on real space coordinates, allowing discriminating different reactive sites in the molecule, as epitomized by the electron density ρ itself (which is the functional derivative of E_e with respect to $v(\mathbf{r})$) and the so-called Fukui functions defined by:

$$f^+(\mathbf{r}) = \left(\frac{\partial \rho(\mathbf{r})}{\partial N}\right)_{\partial N>0}, \; f^-(\mathbf{r}) = \left(\frac{\partial \rho(\mathbf{r})}{\partial N}\right)_{\partial N<0}. \tag{3}$$

The dual descriptor[39] $f^2(\mathbf{r})$, which compares electrophilic and nucleophilic behaviours at each point, then reads:

$$f^2(\mathbf{r}) = \left(\frac{\delta^3 E_e}{\partial N^2 \delta v(\mathbf{r})}\right) = \left(\frac{\partial^2 \rho(\mathbf{r})}{\partial N^2}\right)_{v(\mathbf{r})} \approx f^+(\mathbf{r}) - f^-(\mathbf{r}). \tag{4}$$

Fukui functions can also be used to define the local softness within the grand-canonical ensemble corresponding representation, according to:

$$s^{\pm}(\mathbf{r}) = \left(\frac{\partial \rho(\mathbf{r})}{\partial \mu}\right)_{v(\mathbf{r})} = S f^{\pm}(\mathbf{r}), \tag{5}$$

which integrates (over whole) space to the molecular softness *S*. A local kernel for hardness can be similarly defined so that it integrates to η. However, there is no unique solution. We will consider here Meneses et al.s' intuitive (η_M)[40] and Gál's (η_G)[41–43] widespread proposals:

$$\begin{cases} \eta_M(\mathbf{r}) = I f^{-}(\mathbf{r}) - A f^{+}(\mathbf{r}) \\ \eta_G(\mathbf{r}) = \eta\, f(\mathbf{r}) + \mu f^2(\mathbf{r}) \end{cases}, \tag{6}$$

where *I* and *A* stand for the vertical ionization potential and electron affinity, respectively, and f is here taken as the average of f^+ and f^-.

Finally, non-local descriptors, such as the linear response kernel[44], depend on at least two sets of spatial coordinates, but they have not been considered in this study.

In order to "coarse-grain" the information encoded in local descriptors, condensation is often used. When an atomic partition is used, atomic values are then obtained. We have chosen here, as in our previous papers, Bader's QTAIM partitioning of the electron density. Schematically, for any local property *p*, an atomic value is obtained by integration over the corresponding atomic basin Ω:

$$p(\Omega) = \int_{\Omega} p(\mathbf{r}) d^3 r. \tag{7}$$

2.2. Reactivity principles

Such reactivity descriptors have been used to infer important reactivity rules. They can be categorized into two types. The first one corresponds to rules predicting which sites will react. For instance, Pearson's celebrated Hard and Soft Acids and Bases (HSAB) principle[19,45,46] states that "Hard (unpolarizable) acids bind strongly to hard bases, and soft (polarizable) acids bind strongly to soft

bases". It was initially supported by experimental data until a link with theoretical models was established once the chemical hardness and softness were unambiguously defined by CDFT[19].

The principle colloquially known as "|Δμ| big is good" (DMB), proposed by Parr and Yang in 1989[47], is another example of this first category. It translates the fact that, given two reactive moieties on a reactant, another reactant will preferentially approach one of the moieties so that the variation of its chemical potential is maximised. It has been a conjecture for a few decades until an analytical proof of it and links with other reactivity principles were given recently by Miranda-Quintana and Ayers[48–51].

Finally, Ayers and coworkers proposed in 2017[52] the Maximal Fukui Sum (MFS) principle: atoms A and B for which the sum of the atomic electrophilic $f^{+}(A)$ and nucleophilic Fukui function $f^{-}(B)$ reaches its maximum are the atoms that will bond when the electrophile bearing atom A meets the nucleophile carrying atom B.

The secondary category of reactivity principles gathers rules that govern the variation of CDFT descriptors along a reaction path. For instance, the Maximal hardness principle (MHP) was also formulated by Pearson[53]: "There seems to be a rule of nature that molecules arrange themselves so as to be as hard as possible." This in particular implies that the product of a given reaction should have a higher hardness η than the reactant[34,54,55], and that hardness is expected to increase during the formation of a more stable product.

Similarly, the Minimum Electrophilicity Principle (MEP, an acronym that should not be confounded with the same also used for the molecular electrostatic potential) was formulated by Chattaraj et al. in 2007[56] and postulates that the electrophilicity index ω reaches a minimum at equilibrium structures.

2.3. Reaction pathway analysis

For a single reaction step described by its intrinsic reaction coordinate (IRC)[57] ξ, the reaction force is defined according to Toro-Labbé and co-workers[27,28,30,58] by:

$$F(\xi) = -\frac{dE(\xi)}{d\xi}, \tag{8}$$

where E denotes the total molecular energy ($E = E_e + E_{NN}$, where E_{NN} represents the nuclei repulsion energy). F vanishes at the reactant, product, and transitions state (TS) points, it is negative between the reactant and the TS, and positive between the TS and the product. It reaches a minimal value at ξ_1 before the TS and a maximum at ξ_2 after the TS, the energy profile $E(\xi)$ showing typical inflexion points at ξ_1 and ξ_2..

Three regions can be subsequently defined: the "reactant zone" between the reactants and ξ_1 linked to the "preparation" stage, the "TS zone" between ξ_1 and ξ_2 where most of the electronic activity is expected to occur, and the "product zone" after ξ_2, with final structural relaxations.

3. Computational details

Firstly, the Balloon software[59,60] was employed to search for the most stable conformer for all molecules with the MMFF94 force field. Full geometrical optimizations were then carried out using the Gaussian 09 software[61] with the dispersion-corrected ωB97X-D range-separated hybrid exchange-correlation functional[62] in conjunction with the 6-311++G(2d,2p) triple-ζ Pople basis set, with the so-called 'ultrafine' integration grid. The IRC step size was set to 0.05 bohr for all compounds.

We used the Gibbs energy formula proposed by Chen et al.[63,64] in which the translation and rotational entropy values are scaled by a two-thirds factor in order to correct the overestimation of these quantities in solution due to the use of ideal gas formulas:

$$G^0 = H^0 - T^0\left(S_{vib} + \frac{2}{3}S_{trans} + \frac{2}{3}S_{rot}\right). \tag{9}$$

Rate constants were evaluated using Eyring's simple formula with a transmission factor equal to 1 where $\Delta G^{\ddagger}$ denotes the corresponding Gibbs activation energy:

$$k = \frac{k_B T}{h} e^{-\frac{\Delta G^{\ddagger}}{RT}}. \tag{10}$$

Solvent effects, namely dimethylsulfoxide (DMSO) (the solvent of choice for Mayr's kinetic measurements with Michael acceptors), were implicitly described using a polarizable continuum model in the integral equation formalism (IEF-PCM),[65] an approach that is reliable since aMRs show the computational advantage to proceed only through neutral species (which is not the case of the original Michael reaction in which ionic species are involved, making accurate solvation modelling more involved).

All reagents, products and intermediates were verified to have no imaginary frequencies, while TSs feature only one, corresponding to the normal mode associated to the intrinsic reaction coordinate.

Briefly speaking, CDFT descriptors can be usually computed using either the Finite Difference Linear (FDL) approximation or the Frontier Molecular Orbitals (FMO) approach[66]. FDL requires three calculations performed at the N-electron frozen geometry with $N\pm1$ (the quadratically convergent (QC) algorithm should be sometimes used to properly converge the corresponding self-consistent field (SCF) energies) and N electrons. Conversely, the FMO approach only requires one calculation and uses the properties of the highest occupied molecular orbital (HOMO) and of the lowest unoccupied (LUMO) one to compute the relevant descriptors.

While FDL presents the theoretical advantage to take into account all molecular orbital relaxations during the loss of capture of electrons, it can also induce computational artefacts, in particular in case of negative electron affinities A. Indeed, such values indicate that the additional electron cannot bind to the studied molecule. However, due to the fact that finite basis sets are used, the expelled electron is not properly described and tends to locate to unphysical region spaces (see for instance Figure 3 in ref. 67[67]). Such issue has been addressed in several ways[68,69]. We have here chosen to apply Tozer and De Proft's procedure[70]:

$$A \approx -(\varepsilon_{HOMO} + \varepsilon_{LUMO}) - I, \tag{11}$$

where ε_{HOMO} (resp. ε_{LUMO}) denotes the energy of the highest occupied (resp. lowest unoccupied) molecular Kohn-Sham orbital of the unperturbed system, while I is valuated through the usual FDL approach ($I = E(N-1) - E(N)$). Then, the hardness will be computed as the difference between I and

A, while the electron density of the HOMO (resp. LUMO) will be used as effective approximations for f^- (resp. f^+).

QTAIM calculations were performed with the AIMAll software[71] with default parameters, while homemade scripts were used to extract and calculate all atomic and molecular descriptors, as described more in details in our previous studies[13–15].

4. Results and discussion

4.1. Reaction mechanism

For this mechanistic study, the chosen and fixed nucleophile was methylamine, and a set of 11 Michael acceptors from Mayr's electrophile database was used as electrophiles (see Figure 1). This set is a representative sample of a set of 35 Michael acceptors already considered in previous papers[13,14] that gathers weak, medium and strong electrophiles.

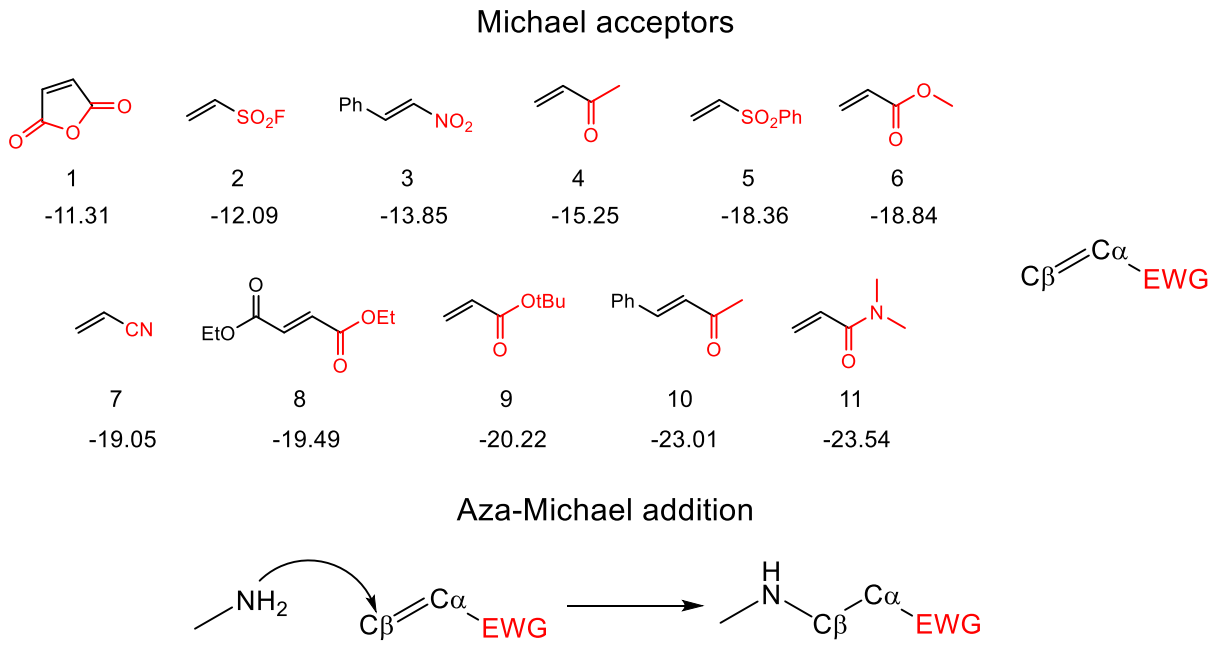


Figure 1. *The 11 studied systems with their respective Mayr electrophilicity values (arbitrary unit), decreasing from* ***1*** *to* ***11****, and the studied aza-Michael reaction.*

Our investigation preliminarily relies on a seminal study by Desmet et al.[72] who investigated in great detail the mechanism of aza-Michael additions of primary and secondary amines to acrylates. While they aimed at accurately determining kinetic parameters from a complex and rigorous pipeline through the numerical solvation of kinetic equations (including diffusion processes), we will mainly focus here on the potential energy surface analysis for the 11 chosen systems.

As evidenced in Figure 2, the reaction starts with the formation of a non-covalent complex (**I_1**) resulting from the noncovalent interaction between the methylamine and the Michael acceptor (**MA**), followed by the nucleophilic attack itself (red arrow) of the nitrogen-bearing compound on **MA**, creating the C-N bond (Michael **Adduct)**. Then a proton transfer (orange arrows) should occur between the nitrogen atom and the carbon atom next to the electrophilic site, affording the final stable **Product**. Several pathways exist for such a transformation since it can proceed either with the assistance of a second amine molecule (represented in green in Figure 2) acting as a catalytic relay, or without it.

The uncatalyzed case can be a one-step process through transition state **TS_{AP}** or a two-step one involving intermediate **I_2** where the proton shifts from N to the heteroatom in **MA**. While **TS_{AP}** was located for all systems, **I_2** can be formed only for compounds **4**, **6**, **8-11**. The catalytic counterpart is initiated by the formation of intermediate **I_3** that features a 6-membered and a 4-membered rings maintained by intermolecular hydrogen bonds. Through one or multiple steps from **I_3**, the non-covalent complex **I_4** between the catalytic amine and the final **Product** can be formed, the final elimination of methylamine ending the process.

It can be noticed that, in the catalyzed case (at variance with the uncatalyzed one), the proton (represented in green in Figure 2) recovered by the carbon atom does not come from the amine involved in the addition, but from the second amine, which nevertheless remains a catalytic species since it abstracts the proton (depicted in orange) from the first amine molecule. The **I_3-I_4** transformation can occur in one step through **TS_{34}** in the case of compounds **4-7**, **9** and **11**. It can be a three-step mechanism involving either intermediates **I_5** and **I_6** (cases of **1-3**, **8** and **10**), or intermediates **I_7**, **I_8** and **I_4** (for **4**, **6** and **8-11**). It follows that for several compounds, different competitive catalyzed pathways should be considered.

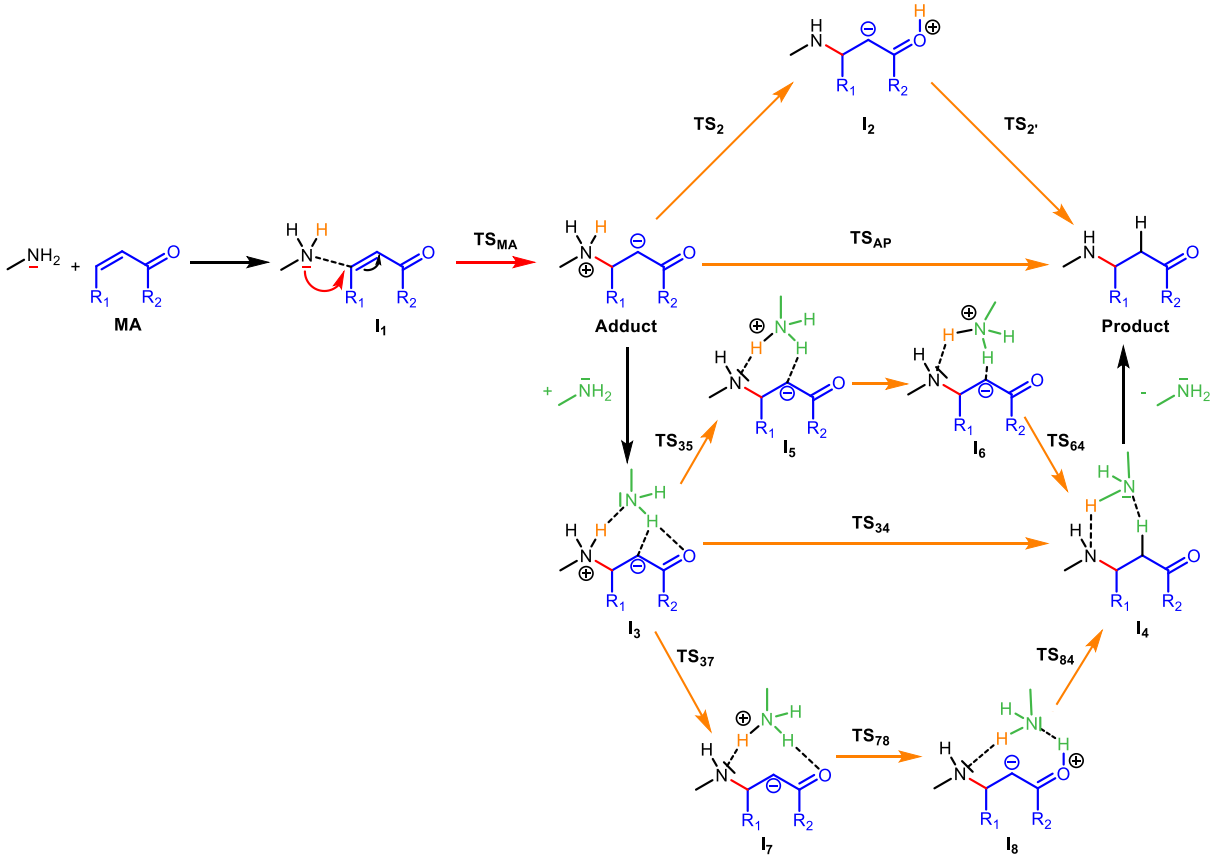


Figure 2. *Full mechanism of the aza-Michael reaction, with various possible pathways for proton transfer.*

Now, from a quantitative point of view, the amine addition (between $\mathbf{I_1}$ and **Adduct**) features a relatively high activation barrier (9.6 – 20.7 kcal mol^{-1} in terms of standard Gibbs energies with respect to $\mathbf{I_1}$), but which remains in average (15.1 kcal mol^{-1}) significantly lower than that related to $\mathbf{TS_{AP}}$ (22.4 kcal mol^{-1}, within the 16.5 – 31.0 kcal mol^{-1} range) and that for $\mathbf{TS_{2'}}$ (in the 38.8 – 50.4 kcal mol^{-1}). These results clearly prove that the unassisted path is irremediably precluded from kinetic reasons.

Conversely, as already mentioned, the catalyzed way, much more affordable, can proceed by many ways. Figure 3 thus represents the favoured pathways for all of the 11 Michael acceptors. For each of them, all possible reaction profiles are depicted in Figures S1-S11 in the supplementary information file (SIF). The first important observation is that the reaction is overall exergonic (and thus thermodynamically favoured) since the Gibbs reaction energy is within the [-9.9, -2.2] kcal mol^{-1} range

(-6.4 in average). Secondly, the transition state for the C-N bond formation (**TS$_{MA}$**) is not always the highest point (**HP**) in the reaction profiles. Indeed, it is the HP for compounds **1**, **2**, **5** and **8**, but the HP is the last TS for all other cases (either **TS$_{64}$** or **TS$_{34}$**, both related to proton transfer). Simplified profiles depicting only the consecutive **HP**s and lowest points are represented in Figure S14.

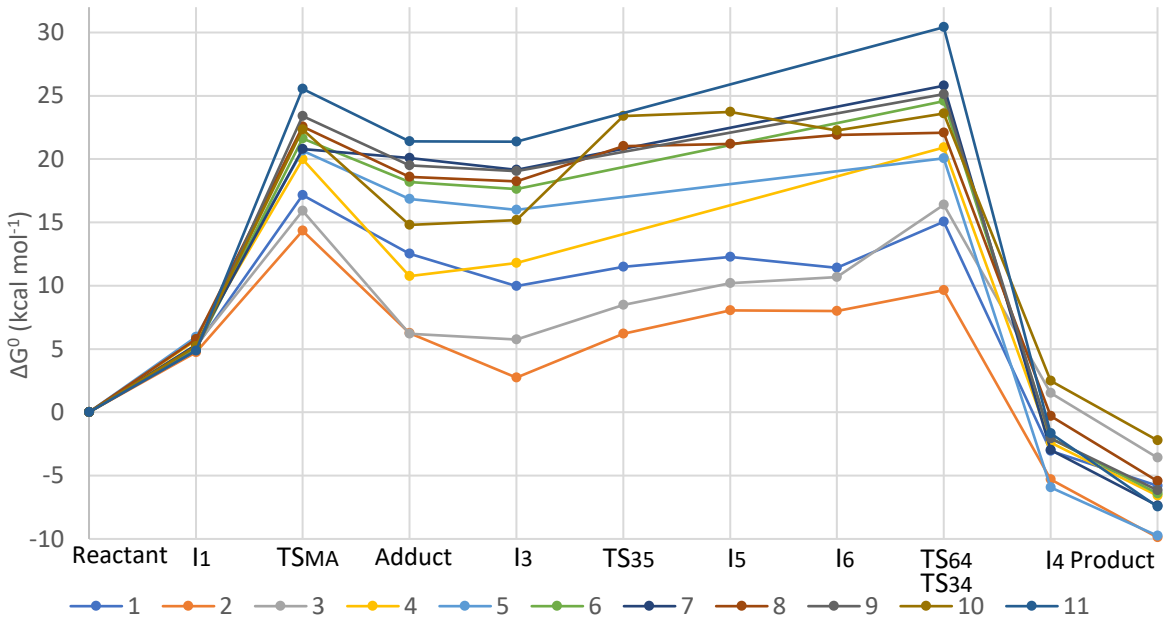


Figure 3. *Most favoured reaction pathways (in terms of Gibbs energy) for the 11 studied systems. ωB97X-D/6-311++G(2d,2p) IEF-PCM (DMSO) level of theory.*

We now look into more details in the obtained numbers. Regarding the full C-N bond formation process, the total Gibbs activation barrier is actually the sum of two barriers: one from the isolated reactants to **I$_1$** and that between **I$_1$** and **TS$_{MA}$**. The first one is almost independent of the electrophilic species (the standard deviation (SD) being equal to 0.4 kcal mol^{-1}), amounts to about 5 kcal mol^{-1}, and cannot thus be neglected. This barrier in Gibbs energy is clearly of entropic nature, since $\Delta H°$(**I$_1$**)-$\Delta H°$(**reactants**) is mostly slightly negative (average value equal to -1.8 kcal mol^{-1}). Conversely, the second barrier is mainly enthalpy-driven as revealed by the high coefficient of determination R^2 value (0.95) between $\Delta G°$(**TS$_{MA}$**)-$\Delta G°$(**I$_1$**) and $\Delta H°$(**TS$_{MA}$**)-$\Delta H°$(**I$_1$**), and is strongly system-dependent (the standard deviation for $\Delta G°$(**TS$_{MA}$**)-$\Delta H°$(**I$_1$**) equals 3.4 kcal mol^{-1}).

Another salient feature is that the C-N bond formation step is by itself reversible (while obviously the full reaction is irreversible thanks to the subsequent proton transfers that considerably stabilize the Michael adduct). Indeed, **Adduct** lies above $\mathbf{I_1}$ in all cases, the $\Delta G°$(**Adduct**)-$\Delta G°$($\mathbf{I_1}$) difference belonging to the [0.9, 16.5] kcal mol^{-1} range (9.8 in average). This reverse Gibbs activation energy remains low since it was always found lower than 9.8 kcal mol^{-1}, with an average value equal to 5.4 kcal mol^{-1}, and is particularly small for compounds **9** and **11** (about 1 kcal mol^{-1}). However, the reverse barrier is not at all correlated to the forward barrier (R^2 = 0.27), suggesting that $\mathbf{I_1}$ and **Adduct** have very different reactivity properties.

One may also wonder whether the stability of **Adduct** is related to its formation barrier. In other words: does the Bell-Evans-Polanyi principle (BEPP)[73,74] hold for this step ? R^2 value between $\Delta G°$($\mathbf{TS_{MA}}$)-$\Delta G°$($\mathbf{I_1}$) and $\Delta G°$(**Adduct**)-$\Delta G°$($\mathbf{I_1}$) equals 0.63, unvalidating BEPP. This BEPP is not more valid for the full transformation (R^2 < 0.01 between $\Delta G°$(**Product**)-$\Delta G°$(**reactants**) and $\Delta G°$(**HP**)-$\Delta G°$(**reactants**)).

Importantly, the proton transfers leading to $\mathbf{I_4}$ can be considered as irreversible, with only small barriers (3.8 – 10.6 kcal.mol^{-1}) for the most favoured pathways, so that once the product is formed, it remains stable.

The pathway proceeding through a direct N-C transfer and with the assistance of a second amine molecule is clearly favoured for all compounds (barriers: 3.8 – 10.6 kcal.mol^{-1}), while the assisted indirect (through intermediates $\mathbf{I_7}$ and $\mathbf{I_8}$, only path involving enol in the catalyzed version) transfer has very small barriers (sometimes negative) towards the formation of the protonated ketone (-0.7 – 1.9 kcal.mol^{-1}) but much higher to form the final product (7.2 – 16.4 kcal.mol^{-1}). The unassisted transfers have even higher barriers (16.5 – 31.0 kcal.mol^{-1} for the direct transfer and -1.4 – 0.6 kcal.mol^{-1} / 38.8 – 58.6 kcal.mol^{-1} for the indirect transfer).

Finally, one may also wonder whether the amine can also catalyze the formation of the C-N bond through an “electrophilic assistance” process. Indeed, it is well known that the complexation of a Lewis acid or a cation to the oxygen of a C=O group enhances its electrophilicity. Figure S13 shows the

corresponding mechanism in which one amine molecule will add to the Cβ atom while a second one binds to the oxygen of the Michael acceptor **1**. The mechanism without the intervention of this additional amine molecule is available in Figure S14.

This "double" complex **I1** is about twice more stable with respect to the isolated reactants than in the case of a unique amine molecule in terms of SCF energy (-6.8 vs. -3.3 kcal mol^{-1}), but is twice less stable in terms of Gibbs energy (11.8 vs. 5.0 kcal mol^{-1}), while the activation barrier from **I1** to **TS1** is almost identical in terms of SCF energy (14.1 kcal mol^{-1}) and very similar in terms of Gibbs energy (15.3 vs. 16.6 kcal mol^{-1}). As a consequence, the total Gibbs activation barrier from the reactants to **TS1** is considerably higher with two amine molecules (27.1 vs. 21.6 kcal/mol): entropy thus precludes any efficient catalytic effect for the C-N bond formation.

4.2. Kinetic constants and Mayr's electrophilicity values

Now that we disclosed the full mechanism, links between it and experimental data can be considered. To this aim, we recall that Mayr's electrophilicity experimental values E_{Mayr} are based on rate constant measurements according to:

$$log_{10}(k_{eff}) = s_N(N_{Mayr} + E_{Mayr}), \tag{12}$$

where k_{eff} denotes the pseudo first-order rate constant of the reaction, and s_N and N_{Mayr} are parameters related to the nucleophile. In this study, the same nucleophile was considered for all electrophiles, so that we may expect

$$E_{Mayr} = \alpha\, log_{10}(k_{eff}) + \beta \tag{13}$$

with $\alpha > 0$ and β pure constants. In principle, rate constants can be computed from the reaction profiles by invoking Arrhenius law (or more advanced theories) for each elementary step. However, we will use here a simplified approach. Each favoured profile will be "converted" into a two-step process according to the following scheme (see ref 75[75] for a similar approach in an experimental context):

$$\mathbf{MeNH_2 + MA} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathbf{I},\ \mathbf{MeNH_2 + I} \xrightarrow{k_2} \mathbf{MeNH_2 + Product}\,, \tag{14}$$

where **I** denotes the lowest Gibbs energy intermediate between $\mathbf{I_1}$ and $\mathbf{I_5}$. It thus corresponds to **I3** for all compounds except for **4** and **10** (**Adduct** in these two last cases). The k_1 and k_{-1} rate constants are linked to the addition process, while k_2 will be associated to the highest energy transition state (associated to proton transfer) after **I**, which is $\mathbf{TS_{64}}$ in all cases. These simplified profiles are depicted in Figure S14 in the SIF. Then, in the case of large amine excess and assuming the quasi-steady state approximation for **I**, $log_{10}(k_{eff})$ can be put in the following form (with $\gamma > 0$, see derivation in SI page S11):

$$log_{10}(k_{eff}) = \ log_{10}(k_1) - \gamma \frac{k_{-1}}{k_2} + \delta. \tag{15}$$

In case the formation of the C-N bond is the rate-determining step (RDS), equation 12 becomes as expected $log_{10}(k_{eff}) \approx log(k_1)$. Taking into account the reversibility of this addition step and the existence of an activation barrier for the proton transfer obviously decreases the rate constant of the full process ($\gamma > 0$). Plugging equations 12 and 14 into equation 9 directly gives (with A and B positive values):

$$E_{Mayr} = A\ \ log_{10}(k_1) - B\frac{k_{-1}}{k_2} + C, \tag{16}$$

where the three constants will also be used to 'absorb' all simplifications in our model (restriction to two steps, possible DFT inaccuracies, crudeness of Eyring's formula, diffusion processes not taken into account…). A bilinear regression provided A = 1.485, B = 0.00017, C = -14.39 with R^2 = 0.85. The mean absolute error (MAE) for this model is equal to 1.13, which can be compared to the value obtained using a monolinear regression with only $log_{10}(k_1)$, 1.21, and to the bilinear model (based on reactivity descriptors computed on the isolated MA) reported in our previous work[13] (1.25 for this set of 11 compounds).

4.3. Following charge transfers along the addition step

In the following sections, we will mainly focus on the addition step between $\mathbf{I_1}$ and **Adduct**, and will follow the variations of electronic descriptors to better understand the nature of the C-N bond formation, which, as pointed out in the previous section, is the step that has the highest influence on the rate constant, and thus on the electrophilicity value.

To this aim, we first computed five types of charges along the IRC path: QTAIM[22], Mulliken[76], Hirshfeld[77], charges fitted on the molecular electrostatic potential using the CHelpG procedure[78,79] and NPA ones[80]. As the Michael addition is usually described as the nucleophilic addition of nitrogen on the electrophilic reactive carbon, the atomic charge on the nitrogen is expected to increase (corresponding to a loss of electrons) while that on the attacked carbon would decrease (representing a gain of electrons). Such an analysis can also be performed at a 'regional' level by summing all charges on the amine moiety (which will be the opposite, at any point on the PES, of the cumulative charge on the Michael acceptor fragment due to total charge conservation).

We first have a look at the net charge variation Δq defined at the difference between the given atomic charge in **Adduct** and that in the isolated reactants, as shown in Figure 4. As intuitively expected, all charge definitions show a positive variation for the nitrogen atom and a negative variation for the Michael-acceptor moiety, so that this last one can be safely considered as the electrophile.

The situation is much more contrasted when decomposing this global charge transfer at the atomic level. Indeed, a large discrepancy is observed depending on the charge definition. Indeed, CHelpG Δq values revealed highly positive for carbon Cβ (around 0.6 e), which are also positive, but much smaller (around 0.1 – 0.3 e) within QTAIM, Hirshfeld and NPA, while Mulliken's definition affords both negative and positive values. The other schemes thus describe Cβ as globally nucleophilic.

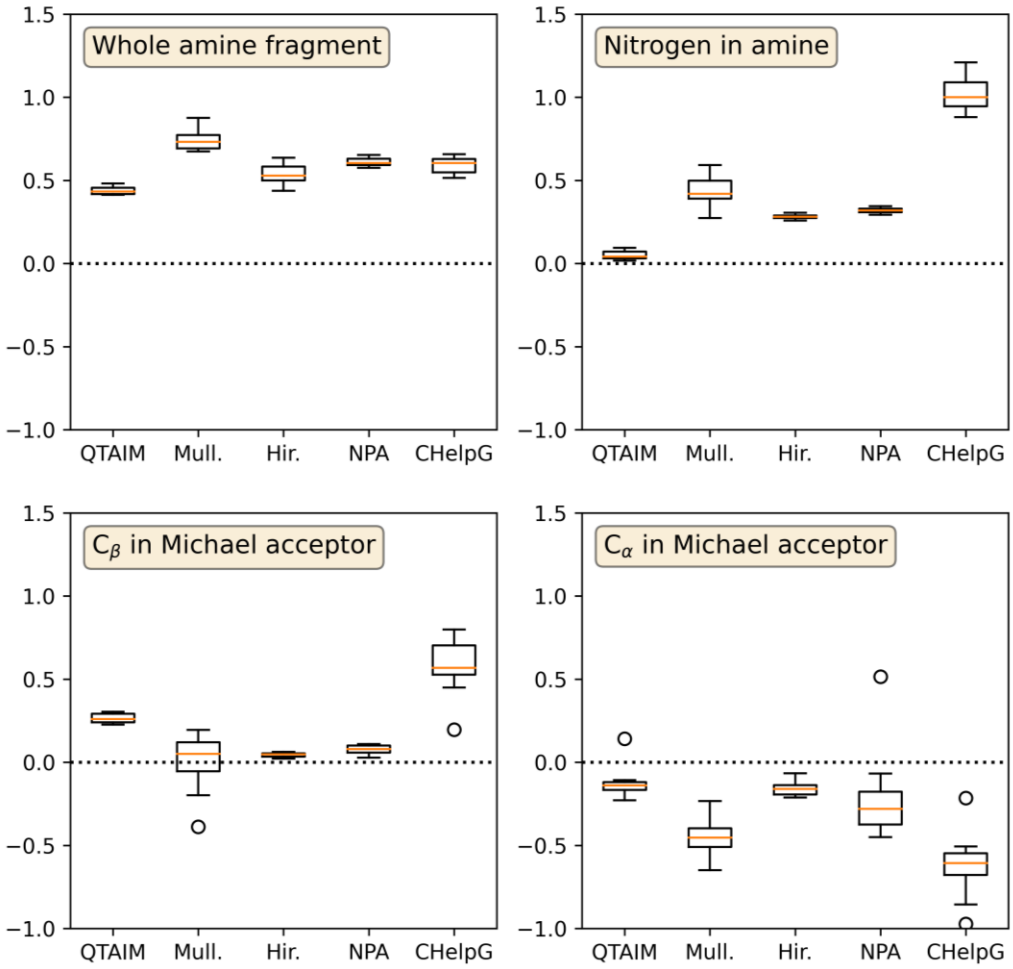


Figure 4. *Box plots for the variation of atomic charges (in e) in the aza-Michael addition from data for the 11 selected systems. Circles correspond to outlier values. Various atomic charge definitions are used. ωB97X-D/6-311++G(2d,2p) IEF-PCM (DMSO) level of theory.*

This counterintuitive result can be rationalized by invoking the electron delocalization along the conjugated system at stake during the charge transfer. Indeed, while nitrogen transfers part of its electrons to Cβ, the electron pair representing the π-electrons in the Cβ=Cα double bond is also shifting (see arrows in Figure 2). This flux is actually reflected in the Cα charge variation.

These variations can be further analyzed by using the reaction force decomposition as shown in Figure 5 for the QTAIM and NPA charge schemes, which collects the amount of charge transfer from amine to **MA** that is fulfilled in each region (the reactant, the TS, and the **Adduct** zones). Noteworthy, all systems exhibited the same trend: most of the charge transfer (63 ± 7%) occurs in the TS zone, with a third (33 ± 7%) occurring in the reactant zone and a very small amount, sometimes negative, in the product zone, due to small charge rearrangements in the latter part. These observations are in full agreement with the usual interpretation of the reaction force that states that electronic activity mostly occurs in the TS region.

All these charge transfer data could be used as descriptors for the prediction of experimental electrophilicities. Unfortunately, the correlation was unsatisfactory for all of the studied charge systems, the R^2 values staying below 0.60 (see Figures S15-19 in SI). One then may wonder whether other chemical descriptors might be more relevant. For this purpose, next section will investigate the variation of quantities derived from CDFT and how they illustrate known reactivity principles.

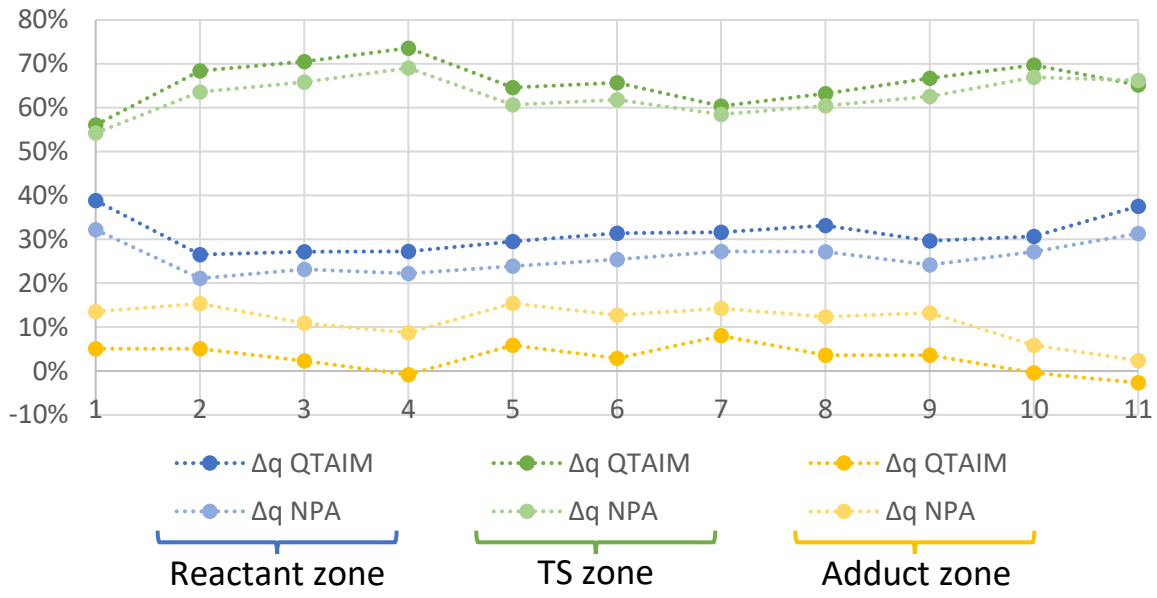


Figure 5. *Percentages of charge transfers occurring between methylamine and each Michael acceptor in each region defined by the reaction force profile. ωB97X-D/6-311++G(2d,2p) IEF-PCM (DMSO) level of theory.*

4.3. CDFT reactivity principles for aza-Michael additions

In this section, we examine to what extent the reactivity principles presented in 2.2 are verified for the considered reaction. Our first point will deal with the regioselectivity: can the Cβ-N bond formation be accounted by atomic descriptors whose values are gathered in Tables S2-S13? Let us recall that, within the retained approach, the atomic softness is evaluated by $s^+ = Sf^+$ for an electrophilic site and by $s^- = Sf^-$ for a nucleophilic one, while the atomic hardness is evaluated through two different kernels, η_G or η_M.

Regarding methylamine, nitrogen is by very far the softest atom and the most electrophilic one (largest and only non-negligible f^- or s^- value). The situation is less obvious for the atomic hardness: nitrogen shows the largest values according to η_G, while all atomic values are similar from the η_M point of view, so that this last descriptor is not discriminating.

We now turn our attention to the Michael acceptor properties. For all systems except compound **3** (for which $s^+(C_\alpha)$ and $s^+(C_\beta)$ respectively rank in third and fourth positions), the maximal value for s^+ is reached for carbon β even if several scenarii are possible. For instance, for compounds **1** and **8**, the s^+ values for carbons α and β are almost equal, while $s^+(C_\beta)/s^+(C_\alpha)$ lies within the [1.5, 2.1] range for **2**, **4**-**7**, **10**-**11**, but it amounts to almost 10 for compound **9**. Conversely, the maximal value for the atomic dual descriptor is obtained for carbon α in case of **10**, for carbon β in **4**, **5**, **8**, **9**, **11**, but for other atoms than Cα and Cβ in the remaining cases, evidencing that $f^{(2)}$ cannot be safely used to infer the reactivity in aMRs.

A similar analysis can be performed for atomic hardness. For instance, the $\eta_G(A)$ maximal atomic value corresponds to Cα for compounds **2**, **6**, **7** and **10**, and to atoms that are not of carbon type for the other ones. Contrastingly, information provided by $\eta_M(A)$ resembles to that observed for $s^+(C_\alpha)$. Indeed, for compounds **1** and **8**, $\eta_M(C_\alpha) \approx \eta_M(C_\beta)$ corresponding to the highest values located in the molecule,

$\eta_M(C_\beta)/\eta_M(C_\alpha)$ lies in the [1.7, 2.6] range for **2**, **4-7**, **10-11**, with $\eta_M(C_\beta)$ corresponding to the maximal value within each reactant, while $\eta_M(C_\beta)$ overdominates in **9**.

One may wonder whether the similarity of trends between η_M and s^+ can be rationalized. In the case where $I \gg A$, one has $\eta_M(\mathbf{r}) \approx If^+(\mathbf{r})$ and $s^+(\mathbf{r}) \approx f^+(\mathbf{r})/I$, so that both functions are obviously proportional. A linear regression between all atomic $s^+(A)$ and $\eta_M(A)$ values for the 11 systems gave a model with R^2 = 0.81. The fact that the correlation is moderate is due to the fact that the I/A ratio is within the [2.9, 6.4] range: *I* effects dominate, but *A* cannot be fully neglected.

We can now come back to the principles enunciated in section 2.2. We have shown that the nitrogen atom in methylamine has the maximal atomic f^- values and that, in almost all cases, f^+ is maximal for carbon β in the Michael acceptor. This implies that the Maximal Fukui Sum principle can be considered as useful to predict Michael addition sites. As the atomic softnesses are directly proportional to these descriptors (and as the same amine is common for all studied reactions), we can also conclude that the softest electrophilic site reacts with the softest nucleophilic site.

Regarding atomic hardness, no useful insight was gained from η_G, while carbon β was found to be the hardest site from η_M (please note that an atom might be at the same time hard and soft because while $S = \sum_A s^\pm(A) = \frac{1}{\eta} = 1/\sum_A \eta(A)$ by definition, there is no mathematical reason that $s^\pm(A) = 1/\eta(A)$). Nevertheless, the nitrogen atom in methylamine is not the hardest one from the η_M point of view.

It follows from these findings that the HSAB principle is verified only from the softness viewpoint. However, as already noticed in ref. 13, the $s^\pm(C_\beta)$ atomic softness is not at all correlated to experimental electrophilicity values (R^2 = 0.03), so that these principles can only be applied here for qualitative purposes dealing with regioselectivity issues.

We now inspect the relevance of CDFT principles describing the evolution of key-features with respect to the reaction coordinate between $\mathbf{I_1}$ and **Adduct**. Following CDFT descriptors along a reaction path has actually a long tradition, notably fostered by the seminal works by Toro-Labbé and Russo[81,82] and by several other groups[83,84].

First, it should be recalled that CDFT provides information only on electronic behavior, so that it may give insight on ΔE profile but not on $\Delta G°$ ones since CDFT is speechless about vibrational and entropic effects.

We found that E(**Adduct**) > E(**I$_1$**) for compounds **1**, **5-11**, while E(**Adduct**) < E(**I$_1$**) only for **2-4**. One of the interpretations of MHP is that hardness is higher for more stable molecules. Hence, one may expect **I$_1$** to be harder than **Adduct** in cases **1**, **5-11**. Intuitively, if one assumes the validity of MHP, hardness is thus conjectured to decrease from **I$_1$** to **Adduct** for these eight reactions. However, as shown in the top left part of Figure 6, the hardness profiles feature no clear trend, invalidating MHP.

The top right panel in Figure 6 now plots the electrophilicity index profiles for five selected systems (see Figure S22 for all systems). There is an obvious considerable decrease of ω between **I$_1$** and **TS$_{MA}$**. After **TS$_{MA}$**, the curves are extremely flat, so that the value of ω at the transition state can be considered as almost minimal: in other words, MEP can be viewed as valid for this step.

The bottom left graph in Figure 6 shows the variations of the electronic chemical potential μ for the same selected Michael acceptors (see Figure S20 for all systems). Once more, the curves are almost fully flat after **TS$_{MA}$**. Interestingly, they are all increasing in the first part of this reaction step, as also does the molecular energy. One may thus wonder, in the spirit of the Relative Energy Gradient (REG) analysis[85,86], if $\Delta\mu$ and ΔE are correlated along the backward part of the IRC profile. As proven by the bottom right part in Figure 6, this is actually the case since high R^2 values are obtained (more than 0.98 ; see Table S15 for all systems).

As the energy variation is linked to the reaction force (RF) and as the μ variations are governed by the reaction electronic flux (REF)[87,88], these observations suggest that RF and REF are intimately linked in the first part of this C-N bond creation. Further studies are however needed to fully disclose such relationships.

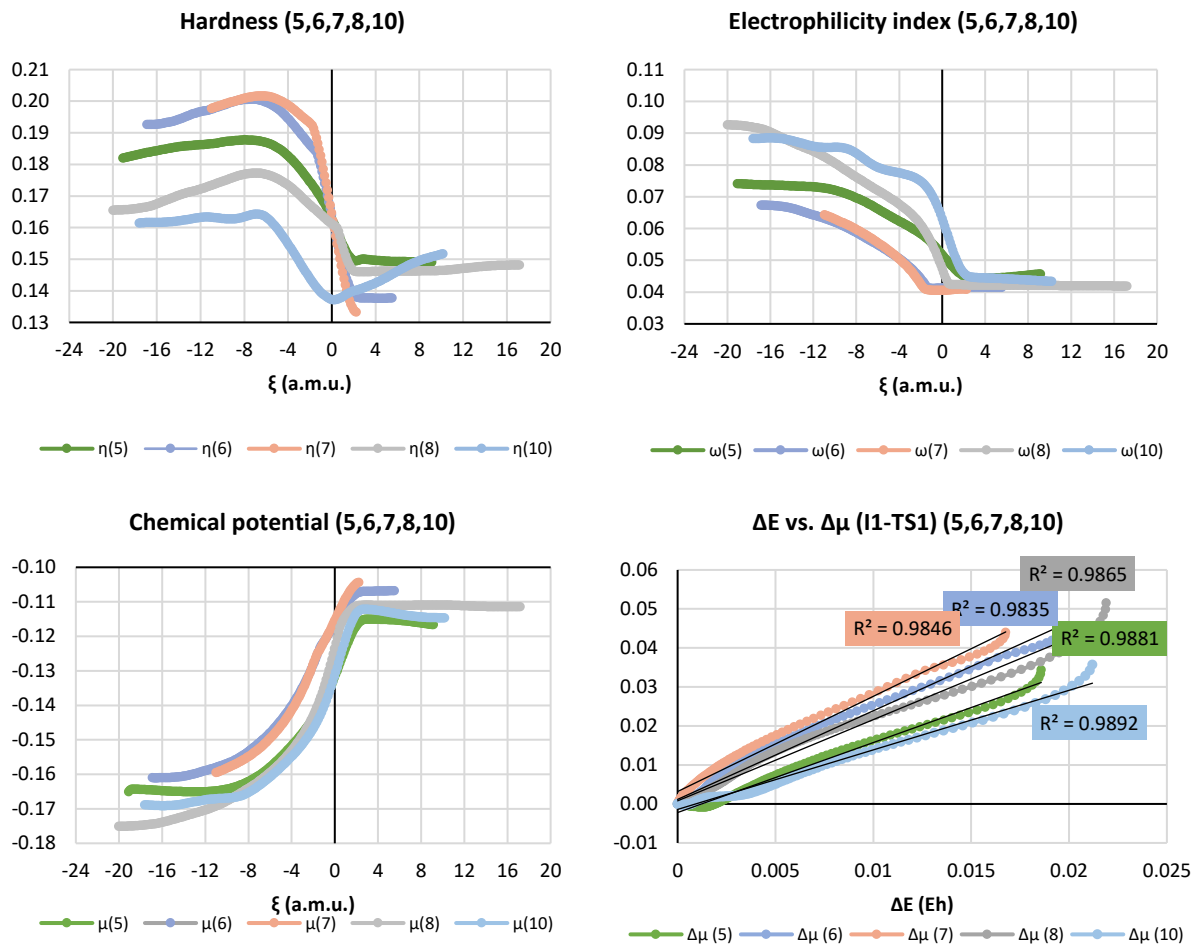


***Figure 6**. Variations of molecular hardness (top left), electrophilicity index (top right), electronic chemical potential (bottom left), and correlations between ΔE (y-axis) and Δμ (x-axis) for five selected systems. ωB97X-D/6-311++G(2d,2p) IEF-PCM (DMSO) level of theory. All values in atomic units.*

Finally, variations for atomic Fukui and atomic dual descriptors were scrutinized, but no clear trend was noticed. The only valuable results are that $f^{-}(N), f^{+}(C_{\beta}), f^{(2)}(N)$ and $f^{(2)}(C_{\beta})$ all have very small values at **Adduct** (see Table S14). This is not surprising since, once the C-N bond is formed, it is not anymore reactive. On the other hand, $f^{(2)}(C_{\alpha})$ at **Adduct** takes significant negative values, between -0.57 and -0.36 (see Table S14). This is indeed in agreement with the full reaction mechanism: C_{α} should be electrophilic in **Adduct** in order to capture the stabilizing proton.

4. Conclusion

In this paper, we have investigated aza-Michael additions involving 11 representative electrophiles. First, the full mechanism was determined with a particular emphasis on the role of the reversibility of the C-N bond formation step and of the proton transfer catalyzed by the amine that allows for full conversion. Relationships with Mayr's electrophilicity scale were then discussed, as well as the charge transfers during the addition step. We notably evidenced that a full electronic redistribution occurs over several atoms, so that the full backbone should be taken into account to characterize the electrophilicity of each species. However, we showed that the regioselectivity can be predicted using atomic hardnesses even if hardness profiles along the addition step are complex. Contrarily, the electronic chemical potential has been shown to increase before the transition state, while the electrophilicity index was found to concomitantly decrease, the variations of both descriptors being vanishing after the transition state. Finally, the chemical potential and the molecular profiles were shown to be correlated in the first part of this addition, highlighting once more the chemical relevance of descriptors stemming from conceptual density functional theory for the understanding of chemical reactions.

Acknowledgements

This work has been partially supported by University of Rouen Normandy, INSA Rouen Normandy, the "Centre National de la Recherche Scientifique" (CNRS), the European Regional Development Fund (ERDF), Labex SynOrg (ANR-11-LABX-0029), Carnot Institut I2C, the graduate school for research XL-Chem (ANR-18-EURE-0020 XL CHEM), and the "Région Normandie". The authors would like also to gratefully acknowledge the "Centre Régional Informatique et d'Applications Numériques de Normandie" (CRIANN) for computing resources.

This paper is dedicated to Prof. Pratim K. Chattaraj who made outstanding contributions to conceptual DFT and, more generally, to the fundamental understanding of chemical reactivity.

Supplementary information

Additional figures showing all reaction pathways and variations of descriptors along them, Cartesian coordinates for all optimized geometries, tables with computed descriptors values.

References

1. A. Michael. *J. für Prakt. Chemie* **35**, 349–356 (1887).
2. E. D. Bergmann, D. Ginsburg & R. Pappo. in *Organic Reactions* vol. 10 179–556 (John Wiley & Sons, Inc., 1959).
3. C. F. Nising & S. Brase. *Chem. Soc. Rev.* **41**, 988–999 (2012).
4. D. Enders, A. Saint-Dizier, M. Lannou & A. Lenzen. *European J. Org. Chem.* **2006**, 29–49 (2006).
5. A. Y. Rulev. *Russ. Chem. Rev.* **80**, 197–218 (2011).
6. M. Kawatsura, S. Aburatani & J. Uenishi. *Tetrahedron* **63**, 4172–4177 (2007).
7. A. Kamimura, Y. Nagata, A. Kadowaki, K. Uchida & H. Uno. *Tetrahedron* **63**, 11856–11861 (2007).
8. P. Lakshmipathi & A. V. Rama Rao. *Tetrahedron Lett.* **38**, 2551–2552 (1997).
9. K. Nakama, S. Seki & S. Kanemasa. *Tetrahedron Lett.* **42**, 6719–6722 (2001).
10. P. R. Krishna, A. Sreeshailam & R. Srinivas. *Tetrahedron* **65**, 9657–6972 (2009).
11. H. Mayr & M. Patz. *Angew. Chemie Int. Ed. English* **33**, 938–957 (1994).
12. Mayr's Database Of Reactivity Parameters: https://www.cup.lmu.de/oc/mayr/reaktionsdatenbank/fe/.
13. G. Hoffmann, V. Tognetti & L. Joubert. *J. Mol. Model.* **24**, 281 (2018).
14. G. Hoffmann, V. Tognetti & L. Joubert. *Chem. Phys. Lett.* **724**, 24–28 (2019).
15. G. Hoffmann, M. Balcilar, V. Tognetti, P. Héroux, B. Gaüzère, S. Adam & L. Joubert. *J. Comput. Chem.* **41**, 2124–2136 (2020).
16. P. Hohenberg & W. Kohn. *Phys. Rev.* **136**, B864–B871 (1964).
17. W. Kohn & L. J. Sham. *Phys. Rev.* **140**, A1133–A1138 (1965).
18. R. G. Parr, R. A. Donnelly, M. Levy & W. E. Palke. *J. Chem. Phys.* **68**, 3801–3807 (1978).
19. R. G. Parr & R. G. Pearson. *J. Am. Chem. Soc.* **105**, 7512–7516 (1983).
20. H. Chermette. *J. Comput. Chem.* **20**, 129–154 (1999).
21. P. Geerlings, F. De Proft & W. Langenaeker. *Chem. Rev.* **103**, 1793–1874 (2003).
22. R. F. W. Bader. *Atoms in molecules : a quantum theory* (Oxford University Press, 1990).
23. P. L. A. Popelier. *Atoms in Molecules: An Introduction* (Pearson Education, 2000).
24. C. Giraldo, S. Gómez, F. Weinhold & A. Restrepo. *ChemPhysChem* **17**, 2022–2034 (2016).
25. E. Chamorro, P. K. Chattaraj & P. Fuentealba. *J. Phys. Chem. A* **107**, 7068–7072 (2003).
26. A. Hocquet, A. Toro-Labbé & H. Chermette. *J. Mol. Struct. THEOCHEM* **686**, 213–218 (2004).
27. B. Herrera & A. Toro-Labbé. *J. Phys. Chem. A* **111**, 5921–5926 (2007).
28. V. Labet, C. Morell, A. Grand & A. Toro-Labbé. *J. Phys. Chem. A* **112**, 11487–11494 (2008).
29. P. Flores-Morales, S. Gutiérrez-Oliva, E. Silva & A. Toro-Labbé. *J. Mol. Struct. THEOCHEM*

943, 121–126 (2010).

30. S. Giri, E. Echegaray, P. W. Ayers, A. S. Nuñez, F. Lund & A. Toro-Labbé. *J. Phys. Chem. A* **116**, 10015–10026 (2012).

31. S. Pan, M. Solà & P. K. Chattaraj. *J. Phys. Chem. A* **117**, 1843–1852 (2013).

32. R. A. Miranda-Quintana, P. W. Ayers, F. Heidar-Zadeh, R. A. Miranda-Quintana, P. W. Ayers & F. Heidar-Zadeh. *ChemistrySelect* **6**, 96–100 (2021).

33. C. Morell, V. Labet, A. Grand & H. Chermette. *Phys. Chem. Chem. Phys.* **11**, 3417–3423 (2009).

34. P. K. Chattaraj, A. Cedillo & R. G. Parr. *Chem. Phys.* **204**, 429–437 (1996).

35. P. Geerlings, E. Chamorro, P. K. Chattaraj, F. De Proft, J. L. Gázquez, S. Liu, C. Morell, A. Toro-Labbé, A. Vela & P. Ayers. *Theor. Chem. Acc.* **139**, 1–18 (2020).

36. R. G. Parr, L. V. Szentpály & S. Liu. *J. Am. Chem. Soc.* **121**, 1922–1924 (1999).

37. P. K. Chattaraj, U. Sarkar & D. R. Roy. *Chem. Rev.* **106**, 2065–2091 (2006).

38. G. Hoffmann, V. Tognetti & L. Joubert. *J. Phys. Chem. A* **124**, 2090–2101 (2020).

39. C. Morell, A. Grand & A. Toro-Labbé. *J. Phys. Chem. A* **109**, 205–212 (2005).

40. L. Meneses, W. Tiznado, R. Contreras & P. Fuentealba. *Chem. Phys. Lett.* **383**, 181–187 (2004).

41. T. Gál, P. Geerlings, F. De Proft & M. Torrent-Sucarrat. *Phys. Chem. Chem. Phys.* **13**, 15003–15015 (2011).

42. C. A. Polanco-Ramírez, M. Franco-Pérez, J. Carmona-Espíndola, J. L. Gázquez & P. W. Ayers. *Phys. Chem. Chem. Phys.* **19**, 12355–12364 (2017).

43. F. Guégan, W. Lamine, H. Chermette & C. Morell. *Phys. Chem. Chem. Phys.* **20**, 9006–9010 (2018).

44. P. Geerlings, S. Fias, Z. Boisdenghien & F. De Proft. *Chemical Society Reviews* vol. 43 4989–5008 (2014).

45. R. G. Pearson. *J. Am. Chem. Soc.* **85**, 3533–3539 (1963).

46. R. G. Pearson. *J. Chem. Educ.* **64**, 561–567 (1987).

47. R. G. Parr & Y. Weitao. *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, 1989).

48. R. A. Miranda-Quintana & P. W. Ayers. *Theor. Chem. Acc.* **137**, 177 (2018).

49. R. A. Miranda-Quintana, F. Heidar-Zadeh & P. W. Ayers. *J. Phys. Chem. Lett.* **9**, 4344–4348 (2018).

50. R. A. Miranda-Quintana & P. W. Ayers. *Theor. Chem. Acc.* **138**, 44 (2019).

51. R. A. Miranda-Quintana, P. W. Ayers & F. Heidar-Zadeh. *ChemistrySelect* **6**, 96–100 (2021).

52. P. P. Zhou, S. Liu, P. W. Ayers & R. Q. Zhang. *J. Chem. Phys.* **147**, 134303 (2017).

53. R. G. Pearson. *Acc. Chem. Res.* **26**, 250–255 (1993).

54. V. I. Minkin. *Pure Appl. Chem.* **71**, 1919–1981 (1999).

55. R. G. Parr & P. K. Chattaraj. *J. Am. Chem. Soc.* **113**, 1854–1855 (1991).

56. P. K. Chattaraj & S. Giri. *Indian J. Phys.* **81**, 871–879 (2007).

57. K. Fukui. *Acc. Chem. Res.* **14**, 363–368 (1981).

58. P. Politzer, A. Toro-Labbé, S. Gutiérrez-Oliva, B. Herrera, P. Jaque, M. C. Concha & J. S. Murray. *J. Chem. Sci.* **117**, 467–472 (2005).

59. M. J. Vainio & M. S. Johnson. *J. Chem. Inf. Model.* **47**, 2462–2474 (2007).

60. J. S. Puranen, M. J. Vainio & M. S. Johnson. *J. Comput. Chem.* **31**, 1722–1732 (2009).

61. ***Gaussian 09 revision D.01***. 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, and D. J. Fox, Gaussian, Inc., Wallingford CT (2013).

62. J. Da Chai & M. Head-Gordon. *Phys. Chem. Chem. Phys.* **10**, 6615–6620 (2008).

63. P. Chen, B. A. Dougan, X. Zhang, Y. D. Wu & Z. L. Xue. *Polyhedron* **58**, 30–38 (2013).

64. L. A. Jouanno, V. Di Mascio, V. Tognetti, L. Joubert, C. Sabot & P. Y. Renard. *J. Org. Chem.* **79**, 1303–1319 (2014).

65. J. Tomasi, B. Mennucci & R. Cammi. *Chem. Rev.* **105**, 2999–3093 (2005).

66. F. Zielinski, V. Tognetti & L. Joubert. *Chem. Phys. Lett.* **527**, 67–72 (2012).

67. V. Tognetti, C. Morell & L. Joubert. *J. Comput. Chem.* **36**, 649–659 (2015).

68. C. Cárdenas, P. Ayers, F. De Proft, D. J. Tozer & P. Geerlings. *Phys. Chem. Chem. Phys.* **13**, 2285–2293 (2011).

69. J. Carmona-Espíndola, J. L. Gázquez, A. Vela & S. B. Trickey. *J. Phys. Chem. A* **124**, 1334–1342 (2020).

70. D. J. Tozer & F. De Proft. *J. Phys. Chem. A* **109**, 8923–8929 (2005).

71. ***AIMAll (Version 19.10.12)***, Todd A. Keith, TK Gristmill Software, Overland Park KS, aim.tkgristmill.com (2019).

72. G. B. Desmet, D. R. D'Hooge, P. S. Omurtag, P. Espeel, G. B. Marin, F. E. Du Prez & M. F. Reyniers. *J. Org. Chem.* **81**, 12291–12302 (2016).

73. R. P. Bell. *Proc. R. Soc. London. Ser. A - Math. Phys. Sci.* **154**, 414–429 (1936).

74. M. G. Evans & M. Polanyi. *Trans. Faraday Soc.* **32**, 1333–1360 (1936).

75. A. F. Popov, I. F. Perepichka & L. I. Kostenko. *J. Chem. Soc. Perkin Trans. 2* **2**, 395–400 (1989).

76. R. S. Mulliken. *J. Chem. Phys.* **23**, 1833–1840 (1955).

77. F. L. Hirshfeld. *Theor. Chim. Acta* **44**, 129–138 (1977).

78. L. E. Chirlian & M. M. Francl. *J. Comput. Chem.* **8**, 894–905 (1987).

79. E. Sigfridsson & U. Ryde. *J. Comput. Chem.* **19**, 377–395 (1998).

80. A. E. Reed, R. B. Weinstock & F. Weinhold. *J. Chem. Phys.* **83**, 735–746 (1985).

81. T. Mineva, E. Sicilia & N. Russo. *J. Am. Chem. Soc.* **120**, 9053–9058 (1998).

82. E. Rincón, P. Jaque & A. Toro-Labbé. *J. Phys. Chem. A* **110**, 9478–9485 (2006).

83. 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* **111**, 9130–9138 (2007).

84. M. L. Bonnet & V. Tognetti. *Chem. Phys. Lett.* **511**, 427–433 (2011).

85. J. C. R. Thacker & P. L. A. Popelier. *Theor. Chem. Acc.* **136**, 86 (2017).

86. I. Alkorta, J. C. R. Thacker & P. L. A. Popelier. *J. Comput. Chem.* **39**, 546–556 (2018).

87. E. Echegaray & A. Toro-Labbé. *J. Phys. Chem. A* **112**, 11801–11807 (2008).

88. C. Morell, V. Tognetti, E. Bignon, E. Dumont, N. Hernandez-Haro, B. Herrera, A. Grand, S. Gutiérrez-Oliva, L. Joubert, A. Toro-Labbé & H. Chermette. *Theor Chem Acc* **134**, 133 (2015).

4.3 Conclusion

Dans cet article, nous avons étudié le mécanisme complet d'une addition d'aza-Michael. Nous avons montré que la première étape (addition) était réversible et cinétiquement déterminante pour tous les composés (voir **Figure 3**), et que la seconde étape (transfert de proton) pouvait passer par plusieurs chemins (voir **Figure 2**), en accord avec les travaux de Pardo et Desmet. L'étude de Desmet et al. indiquait également qu'il était possible au produit de la réaction de subir une seconde addition, ce que nous n'avons pas considéré. Cela n'aurait pas affecté les barrières calculées ni le déroulé de la première addition, mais cela aurait pu avoir une influence sur les constantes cinétiques prenant en compte les deux additions, et peut-être améliorer les corrélations. Cependant, l'objectif de l'article était plutôt d'étudier la réaction de plusieurs accepteurs avec le même donneur plutôt que de réaliser une étude cinétique détaillée d'une seule réaction.

La catalyse de l'addition par une molécule de méthylamine a également été considérée, mais la barrière d'activation de l'étape d'activation catalysée étant supérieure à la version non catalysée, l'étude de cette voie a été abandonnée.

La majeure partie du transfert de charge a lieu dans la zone de l'état de transition (voir **Figure 5**), en accord avec les observations de Toro-Labbé et al. : le maximum d'activité électronique se produit autour de l'état de transition.

Il est particulièrement intéressant de constater que le C_β ne gagne pas d'électrons mais en perd au cours de l'addition et ne peut donc pas être considéré comme électrophile. La charge transférée de l'amine vers l'accepteur est en fait délocalisée dans le système conjugué et est principalement récupérée par le C_α (voir **Figure 4**). Cela montre que parfois, considérer un seul atome comme le site électrophile ne suffit pas et qu'il faut prendre en compte un groupe d'atomes pour comprendre la réactivité d'une molécule. Il convient également de distinguer site électrophile et site réactif : l'électrophilie est une caractéristique intrinsèque du réactif et caractérise les mouvements d'électrons dans celui-ci au cours de la réaction. Cependant, l'attaque se produit bien sur le C_β. Il est peut-être possible de l'expliquer en étudiant les interactions entre les deux réactifs, ce qui est l'objet du second article, consacré à l'analyse REG-IQA de la réaction d'aza-Michael.

L’étude pourrait être étendue à d’autres systèmes, par exemple les 35 composés du premier article de notre groupe, pour vérifier si la bonne corrélation du potentiel chimique avec l’énergie totale sur la première partie de l’étape d’addition (voir **Figure 6**, bas à droite) est valable pour d’autres systèmes.

Chapitre 5 – Étude théorique des additions d'aza-Michael en solvant explicite

5.1 Introduction 167

5.2 Article 168

5.3 Conclusion 212

5.1 Introduction

Dans la lignée du premier article sur la réaction d'aza-Michael, nous avons souhaité continuer notre étude de cette réaction via une analyse topologique et IQA, qui nous ont entre autres permis de nous intéresser aux effets de solvant. Ceux-ci jouent un rôle important dans cette réaction, puisqu'elle est difficile dans les solvants apolaires et facilitée dans les solvants polaires protiques qui peuvent former des liaisons hydrogène avec les réactifs et ainsi augmenter la nucléophilie du donneur et l'électrophilie de l'accepteur.

Cela était aussi l'occasion d'élaborer un protocole de calcul IQA pour des grands systèmes, puisqu'à notre connaissance, aucun système de cette taille (de l'ordre de 200 atomes) ni comportant un solvant explicite n'a été considéré en IQA. Notre stratégie s'est basée sur des calculs divisés en plusieurs parties afin de pouvoir réduire les temps de calcul et éviter de calculer les interactions non désirées (ici solvant-solvant). C'était également une opportunité pour appliquer REG à un grand système et développer la méthode afin de pouvoir traiter des calculs où seule une partie des termes interatomiques IQA est calculée.

The aza-Michael addition in explicit solvent: a Relative Energy Gradient – Interacting Quantum Atoms study

Aël CADOR[a], Vincent TOGNETTI[a],* Laurent JOUBERT[a]*, Paul L.A. Popelier[b]*

[a] Normandy Univ., COBRA UMR 6014 & FR 3038, Université de Rouen, INSA Rouen, CNRS, 1 rue Tesnière 76821 Mont St Aignan Cedex, France

[b] Manchester Institute of Biotechnology (MIB), Manchester M1 7DN, Great Britain; Department of Chemistry, University of Manchester, Manchester M13 9PL, Great Britain

Abstract:

In this paper, we investigate aza-Michael additions, which are key-reactions in organic synthesis, from the theoretical point of view on several examples ranging from weak to strong electrophiles in explicit solvent (dimethylsulfoxide). Using the REG-IQA (Relative Energy Gradient – Interacting Quantum Atoms) method, we focus on the (rate-limiting) addition step in order to unravel the different events taking place in this step, understand the influence of solvent on the reaction and attempt to predict the Mayr electrophilicity. A first link between the REG-IQA analysis and experimental values is reported.

Keywords: Michael additions, reaction mechanisms, Mayr electrophilicity, quantum theory of atoms-in-molecules, interacting quantum atoms, explicit solvation

Corresponding authors: vincent.tognetti@univ-rouen.fr, laurent.joubert@univ-rouen.fr, paul.popelier@manchester.ac.uk

1. Introduction

Since its discovery at the end of the 19th century, the Michael reaction has imposed itself as one of the most widely used reactions in organic chemistry [1,2], since it can allow for the formation of a carbon-carbon bond without having to resort to using metallic catalysts. Over the years, many variants of interest have been developed to link heteroatoms to carbon atoms [3,4]. Most notable among them is the aza-Michael addition (aMR) [5], which can form a nitrogen-carbon bond. The scope of the reaction includes amines, but heterocycles [6], amides [7], azides [8], oximes [9] have also been shown to react in aMR. Recently, asymmetric versions have been developed as well [10], making it a powerful tool in organic synthesis.

The solvent influence on this reaction has also been well-characterized experimentally. The reaction rates are enhanced in protic solvents such as water due to the formation of hydrogen bonds which enhance the electrophilicity of the acceptor and the nucleophilicity of the donor, while the reaction is slower in polar aprotic and apolar solvents [5]. The reaction can proceed by using the nucleophile as the solvent, for example cyclohexylamine [11].

The simplified mechanism of this important reaction is shown in Figure 1.

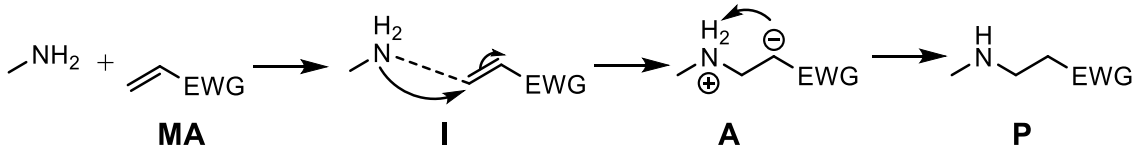


Figure 1. Simplified mechanism of the aza-Michael addition.

Understanding the detailed course of this reaction for a subset of compounds can allow for the extension of its scope, but also for the prediction of its selectivity and a rationale for the choice of suited reagents, and even greater chemical insight which can be applied to other reactions.

From a chemical point of view, the feasibility of reactions between electrophiles and nucleophiles can be empirically predicted using Mayr's electrophilicity scale [12], designed using kinetic constants of electrophile-nucleophile reactions. This scale allows to classify reactants according to their relative electrophilicities and nucleophilicities. These parameters are available online on a freely accessible

database [13], currently including more than 1,600 compounds. Among these, 114 electrophiles are suited for Michael reactions, hence their categorization as Michael acceptors in the database.

Through several recent papers [14–16], our group has reported on a robust theoretical protocol to predict Mayr's electrophilicity through a three-fold strategy relying on density functional theory (DFT) [17,18] calculations and, more particularly, on conceptual DFT (CDFT) [19–22] and Bader's quantum theory of atoms-in-molecules (QTAIM) [23,24] descriptors. The aim of this work is actually the design of a theoretical approach suited for all types of molecules, especially those which are not yet included in Mayr's electrophilicity scale.

We have also recently reported a study of the aza-Michael addition (*see chapter 4*) to complement our prior studies by investigating the mechanistic aspects of the reaction rather than focusing on the intrinsic properties of reactants. This study continues further in this direction by reviewing the reaction under the prism of the Relative Energy Gradient – Interacting Quantum Atoms (REG-IQA) analysis. The focus here will be on the use of the IQA method to determine the driving force of the reaction, the interactions giving rise and opposing the activation barrier, particularly in order to know why the reaction may be reversible, the solvent influence and lastly the prediction of the experimental Mayr electrophilicity. To this end, the REG-IQA analysis will be used to retrieve the main events of the reaction as well as the significant solute-solvent interactions. A new protocol will be finally reported to find suitable IQA terms able to predict activation barriers or Mayr electrophilicities.

To this aim, this paper will be divided as follows: in section 2, the basics of IQA and REG will be briefly reviewed, while section 3 will describe the computational protocol. Section 4 will then present the main results of our calculations.

First, section 4.1 will describe the activation barriers and mechanism of the addition step while section 4.2 will detail the reaction events taking place during the addition step using REG-IQA, QTAIM charges and structural parameters. Section 4.3 will then focus on the influence of solvent on the addition step, notably by studying the variation of the atomic charges and IQA terms in both gas phase and solvent. Finally, section 4.4 will present an attempt at finding an IQA term which would be able to predict the

activation barriers and/or the Mayr electrophilicity. The paper will finally end with conclusions and perspectives.

2. Theoretical background

2.1. Interacting Quantum Atoms

The IQA method [25] is a partitioning scheme of the total molecular energy E_{mol} relying on the QTAIM atomic basins to compute intra-atomic and inter-atomic energy terms:

$$E_{mol} = \sum_{A} E_{intra}(A) + \frac{1}{2}\sum_{A}\sum_{B\neq A} E_{inter}(A,B) \tag{1}$$

where the sums run over every atom A and pair of atoms (A,B) in the system.

The intra-atomic terms $E_{intra}(A)$ include the kinetic energy $T(A)$ of electrons, the nucleus-electron $V_{en}(A)$ and the electron-electron $V_{ee}(A,A)$ repulsion energies inside the atomic basin of A:

$$E_{intra}(A) = T(A) + V_{en}(A) + V_{ee}(A,A) \tag{2}$$

The inter-atomic terms $E_{inter}(A,B)$ include the interaction energies of the nuclei of atoms A and B ($V_{nn}(A,B)$), of the electrons of A and the nucleus of B ($V_{en}(A,B)$), of the nucleus of A and the electrons of B ($V_{ne}(A,B)$), and of the electrons of A and B ($V_{ee}(A,B)$):

$$E_{inter}(A,B) = V_{nn}(A,B) + V_{en}(A,B) + V_{ne}(A,B) + V_{ee}(A,B) \tag{3}$$

The latter term can be divided into coulombic (also called “classical”) and exchange-correlation contributions:

$$V_{ee}(A,B) = V_{ee,cl}(A,B) + V_{ee,xc}(A,B) \tag{4}$$

which allows to rewrite $E_{inter}(A,B)$ as the sum:

$$E_{inter}(A,B) = V_{cl}(A,B) + V_{ee,xc}(A,B) \tag{5}$$

where $V_{cl}(A,B)$ includes all electrostatic interactions between atoms A and B.

More details about the IQA scheme and its applications can be found in dedicated works such as [23,26].

2.2 Relative Energy Gradient

One major downfall of the application of IQA to large systems is the quadratic increase of the number of terms (and subsequent computational cost): for a N-atom system, using a decomposition such as:

$$E_{mol} = \sum_{A} E_{intra}(A) + \frac{1}{2}\sum_{A}\sum_{B\neq A}(V_{cl}(A,B) + V_{ee,xc}(A,B)) \tag{6}$$

already leads to the computation of N^2 terms, and the situation only gets worse by partitioning the energy further. The study of all terms and the selection of relevant terms become impractical; this calls for a method which would be able to circumvent such issues.

The Relative Energy Gradient (REG) [27] is a method developed specifically to solve this issue for IQA ; it is able to automatically rank terms according to their ability to describe the variation of the total energy of the system. This variation can be due to the change of one selected coordinate in the system, such as in a relaxed Potential Energy Surface (PES) scan, or the change of the intrinsic reaction coordinate when the minimal energy path is followed. The coordinate governing this change (bond distance, angle, dihedral or IRC) can then be called a control coordinate.

REG thus allows the selection of a small subset of terms correctly describing the variation of the total energy along the control coordinate. To this end, the energy profile is divided into segments to ensure a monotonous variation of the energy on each segment. We have used the reaction force to define the limits of each segment (see section 2.3).

The REG analysis then proceeds by performing least-squares linear regressions between the total energy and each IQA term according to equation (7):

$$E_i^{IQA} = m_i^{REG} \times E_{mol} + c_i \tag{7}$$

where E_i^{IQA} is any IQA term, m_i^{REG} is the slope (or "REG value") of the regression line and c_i the intercept. The terms are ranked according to the magnitude of their REG coefficient: a REG value above 1 means that the term of interest contributes more to the behavior of the system on a given segment than the total system does. On the other hand, a REG value below 1 indicates that the term contributes less to this behavior than the total system. A positive REG value means that the term has the same variation

trend as the total energy of the system, while a negative REG value indicates that the term opposes the energy variation of the system.

The coefficient of determination R^2 must be also taken into account to assess the quality of the regressions. As such, in this study, only terms whose regression has a R^2 coefficient above 0.9 are considered.

Due to the additive nature of the IQA terms (summing all IQA terms recovers the total energy), the REG procedure can consider terms of single atoms as well as grouped atoms; this can be useful to consider a whole substituent at once instead of each separate atom. In this case, the intra-atomic term of the group is simply equal to the sum of the intra-atomic terms of all atoms A of the group G:

$$E_{intra}(G) = \sum_{A \in G} E_{intra}(A) \tag{8}$$

The global interatomic term of the group is the sum of the interatomic terms between all atoms inside the group:

$$E_{inter}(G, G) = \frac{1}{2} \sum_{A \in G} \sum_{(B \neq A) \in G} E_{inter}(A, B) \tag{9}$$

Finally, the interatomic terms of the group with the other atoms of the system consist in sums of the interatomic terms of each atom of the group with the given atom outside of the group $A'_{outside}$:

$$E_{inter}(A'_{outside}, G) = \sum_{A \in G} E_{inter}(A'_{outside}, A) \tag{10}$$

2.3. Reaction pathway analysis

For a single reaction step described by its intrinsic reaction coordinate (IRC)[28] ξ, the reaction force is defined according to Toro-Labbé and co-workers[29–32] by:

$$F(\xi) = -\frac{dE_{mol}(\xi)}{d\xi} \tag{11}$$

F vanishes at the reactant, product, and transitions state (TS) points, it is negative between the reactant and the TS, and positive between the TS and the product. It reaches a minimal value at ξ_1 before the TS and a maximum at ξ_2 after the TS, the energy profile $E_{mol}(\xi)$ showing typical inflexion points at ξ_1 and ξ_2..

Three regions can be subsequently defined: the "reactant zone" between the reactants and ξ_1 linked to the "preparation" stage, the "TS zone" between ξ_1 and ξ_2 where most of the electronic activity is expected to occur, and the "product zone" after ξ_2, with final structural relaxations.

The TS can also be considered as a dividing point, affording four segments: segment 1 corresponds to the reactant zone (reactant – ξ_1), segment 2 corresponds to the first part of the TS zone (ξ_1 – TS), segment 3 corresponds to the second part of the TS zone (TS – ξ_2) and segment 4 to the product zone (ξ_2 – product).

3. Computational details

Firstly, the Balloon software[33,34] was employed to search for the most stable conformer for all molecules with the MMFF94 force field. The geometries of the isolated reaction adducts were optimized using the Gaussian 16 software [35], then encapsulated in an equilibrated box of 1,024 DMSO molecules [36] using the *gmx solvate* utility from the GROMACS 5.1.4 distribution [37]. The furthest molecules were removed in order to keep only the first solvation shell (18 DMSO molecules for all compounds) using a cutoff radius of 6 Å. Full geometrical optimizations were then carried out using the dispersion-corrected B3LYP-D3 hybrid exchange-correlation functional [38,39] and the 6-31+G(d) basis set [40] with Gaussian 16, using the 'ultrafine' integration grid. The IRC step size was set to 0.05 bohr for compound **6**, 0.08 bohr for compounds **1**, **2**, **3** and **5**, 0.1 bohr for compound **4**; the default (Hessian-based Predictor-Corrector integrator [41,42]) IRC algorithm was used for all compounds, except **4** for which the local quadratic approximation algorithm [43,44] had to be used.

We used the Gibbs energy formula proposed by Chen et al. [45,46] in which the translation and rotational entropy values are scaled by a two-thirds factor in order to correct the overestimation of these quantities in solution due to the use of ideal gas formulas:

$$G^0 = H^0 - T^0\left(S_{vib} + \frac{2}{3}S_{trans} + \frac{2}{3}S_{rot}\right) \tag{12}$$

All reagents, products and intermediates were verified to have no imaginary frequencies, while TSs feature only one, corresponding to the normal mode associated to the intrinsic reaction coordinate.

QTAIM calculations were achieved for a selection of IRC points with the AIMAll 19.10.12 software [47] using the 'B3LYP model' tag [48,49] to ensure correct energy decompositions. The Promega5 integration method was used to correct integration errors. We restricted our study to the solute-solute and solute-solvent interaction terms after showing on smaller systems (see Figure SI-X) that the solvent-solvent interactions are not relevant to understand the reaction profile. In order to avoid calculating the solvent-solvent interactions, we split the IQA calculations in two stages: first the calculation of the IQA terms of the solute, then the calculations of the IQA terms of each solvent molecule and of their interaction with the solute (see Figure 2).

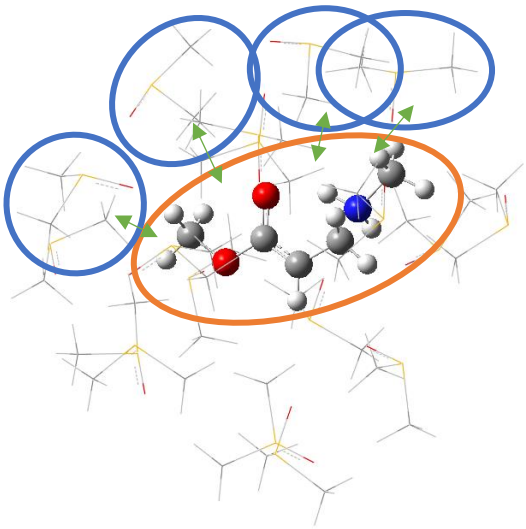


Figure 2. View of the split-calculation strategy with the solutes circled in red, some solvent molecules circled in blue and each arrow denoting one of the solute-solvent IQA calculations.

This approach allowed us to cut down on the computational time and to parallelize the calculations (since each solute-solvent calculation is done in a different job), even if only one configuration can be studied for each compound. The IRC points for these calculations were chosen so as to keep 10 points for each segment on average, with evenly spaced points on each segment. One may object that if there is a large discrepancy of segment size, this leads to an over-representation of the smallest segments (the density of points will be higher), but a smaller segment also indicates a greater variation of the energy, and using less points may cause a poor representation of the energy profile on these segments.

We studied three of the six systems (compounds **3**, **4** and **6**) using this split-calculation approach, while we calculated only a small selection of terms based on our first results for the last three (compounds **1**, **2** and **5**) during the search for a term that would be well-correlated to the activation barriers and/or the Mayr electrophilicity. Hence, sections 4.2 and 4.3 only deal with compounds **3**, **4** and **6**.

The IQA terms were then studied using the REG.py Python library and modified and expanded by our groups, using either E_{intra} and E_{inter} terms, E_{intra}, V_{cl}, V_x and/or grouped solvent terms.

4. Results and discussion

4.1. Reaction mechanism

For this mechanistic study, the chosen unique nucleophile was methylamine, and a set of 6 Michael acceptors from Mayr's electrophile database was used as electrophiles (see Figure 3).

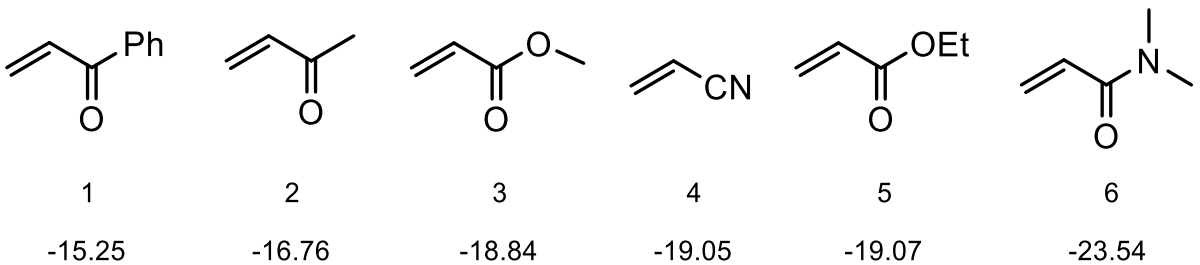


Figure 3. Chemical structures of the six studied electrophiles with their respective Mayr electrophilicities.

This set is a representative sample of a set of 35 Michael acceptors already considered in previous papers[14,15] that gathers weak, medium and strong electrophiles. Our motivations for choosing these systems are that they are the smallest compounds of the set (thus reducing calculation times), five of them share the same enone backbone (and thus share the same mechanism) while compound **4** has a cyano group, hindering its ability to form a hydrogen bond with the nucleophile.

The atom notation used in this article is shown on Figure 4. The solvent atoms which can interact with the solutes are labelled as Z_{DMSO}, with Z replaced by the appropriate element symbol.

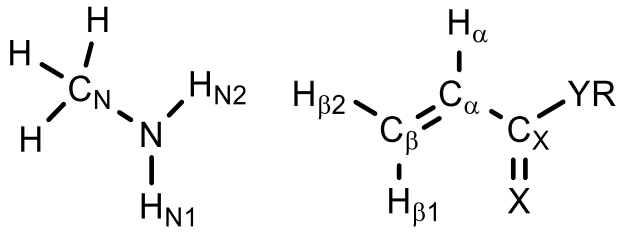


Figure 4. Notation of the relevant atoms of the reactants in this study.

As shown by previous papers [50], the aza-Michael reaction consists in two main steps: the rate-limiting, reversible (in general) 1,4-addition of the nucleophile to the C_β, and the fast, irreversible proton transfer from the nitrogen atom of the nucleophile to the C_α of the electrophile. The addition can only proceed by one pathway while several paths are in competition for the proton transfer. Due to the difference in kinetics, the complexity of the second step and the large computational time a thorough analysis of proton transfer would require, we focused our efforts on the addition step. The reaction was studied in both gas phase and explicit DMSO solution.

As shown on Figure 1, the addition step starts with the formation of a non-covalent complex (**I**) between the methylamine and the Michael acceptor (**MA**). It is then followed by the nucleophilic attack of the nitrogen atom of the methylamine on the C_β of the electrophile, which forms the $C_\beta - N$ bond and leads to the adduct (**A**). The $C_\beta = C_\alpha$ bond becomes a single bond in the process.

The TS of the addition step (between **I** and **A**) is taken as the reference point for the computation of relative energies due to the difficulty of determining a correct solvated structure for the isolated reactants while retaining the same number of solvent molecules.

Like in previous studies, this step is found to be endothermic for most systems (see Table 1), with a high direct activation barrier (equal to the energy difference between I and TS, and not with respect to the isolated reactants) in terms of ΔE_{mol} ranging 52.5 – 74.0 kJ mol^{-1} (average: 68.2 kJ mol^{-1}) for the 6 systems in gas phase and 33.3 – 80.7 kJ mol^{-1} (average: 48.1 kJ mol^{-1}) in DMSO. The barriers are on average noticeably lower in solvent, except for compound **6**, which features an activation barrier slightly higher in DMSO (by 6.7 kJ mol^{-1}) than in gas phase. This confirms that the solvent molecules stabilize the system, and will be detailed in section 4.3.

However, the reaction is exothermic and almost anergonic for the more electrophilic systems **1** and **2**. The reverse barriers are sometimes higher than the direct ones: in gas phase, this can happen when the adduct is unstable and undergoes an enolization (system **1**, **2** and **6**).

This variable mechanism in gas phase, as well as the fact that the reaction is impossible for system **4** in these conditions, further underlines the need of considering the reaction in solvent rather than in gas phase.

		1	**2**	**3**	**4**	**5**	**6**
E_{Mayr}		-15.25	-16.76	-18.84	-19.05	-19.07	-23.54
Gas phase	ΔE_{act}	52.5	58.4	77.5	N/A	78.5	74.0
	ΔG^0_{act}	70.5	75.5	94.7	N/A	95.5	94.0
	$\Delta E_{act, rev}$	76.2	81.2	7.4	N/A	7.6	50.1
	$\Delta G^0_{act, rev}$	69.3	72.0	5.0	N/A	5.5	41.9
Explicit solvent	ΔE_{act}	33.3	27.5	45.1	36.4	65.6	80.7
	ΔG^0_{act}	44.1	39.3	45.8	48.0	62.3	75.9
	$\Delta E_{act, rev}$	57.4	49.8	27.7	15.5	27.7	16.5
	$\Delta G^0_{act, rev}$	43.7	39.4	12.7	9.7	15.0	15.6

Table 1. Energy and free energy activation barriers (in kJ mol^{-1}) in gas phase and explicit solvent for the six studied systems.

The corresponding IRC profiles are shown on Figure 6.

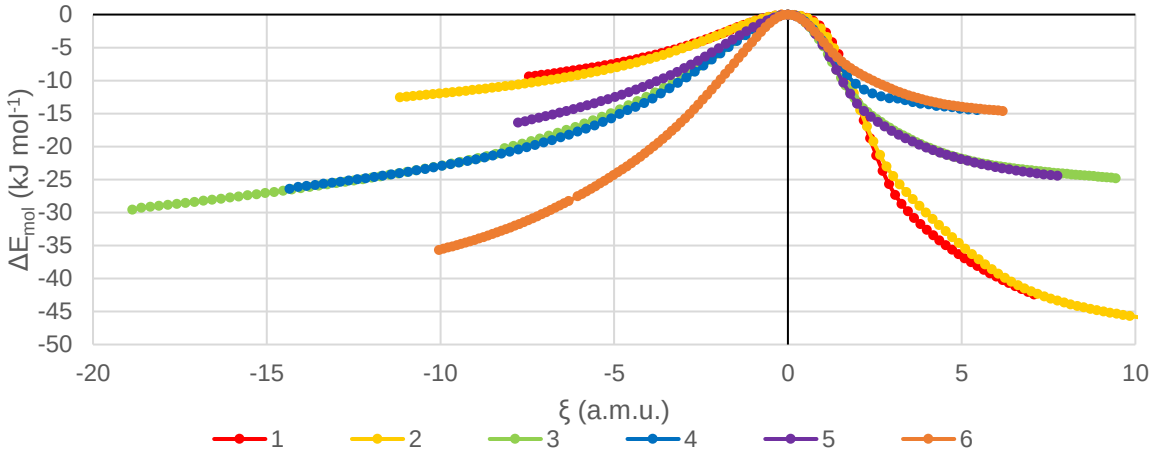


Figure 6. Explicit solvent IRC profiles for the six studied systems.

4.2. REG-IQA analysis of the reaction

Now that we have described the mechanism of the reaction, we can go further in detail in the reaction events shown by the REG analysis.

In this section, only solute terms (under the influence of the solvent) are discussed. Only systems **3**, **4** and **6** are considered since they are the only systems for which all of these terms have been calculated; refer to section 4.3 for more details on solvent interactions.

System **3** (methyl acrylate) has been first chosen as an illustrative example; REG analysis stem plots of each segment are given below (graphs A of Figures 8 and 10-12). The *Stabilising / Sign change / Destabilising* captions in the plot keys indicate the sign of each IQA energy term E_i^{IQA}, respectively always negative (E_i^{IQA}<0), changing from negative to positive (or vice-versa) or always positive (E_i^{IQA}>0). The similarities and differences with systems **4** (acrylonitrile) and **6** (N,N-dimethylacrylamide) will be discussed after system **3**.

The segmentation chosen for system **3** is shown on Figure 7, with the extrema of the reaction force (ξ_1 and ξ_2) very close to the TS.

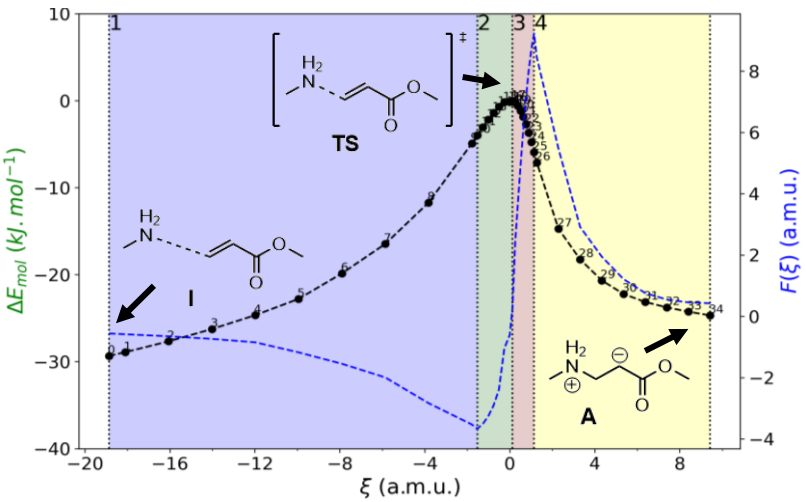


Figure 7. Energy profile (in black), reaction force (in blue), chosen IRC points and segments (compound 3). The origin of energies is located at the TS.

For each segment, the ten terms with the largest absolute REG values (five most positive and five most negative) and with R^2 above 0.90 are labelled on the stem plot. These terms will be systematically discussed from largest to smallest REG coefficient in absolute value, with graphs representing E_i^{IQA} versus ξ and E_{mol} (graphs B and C on Figures 8 and 10-12).

Segment 1

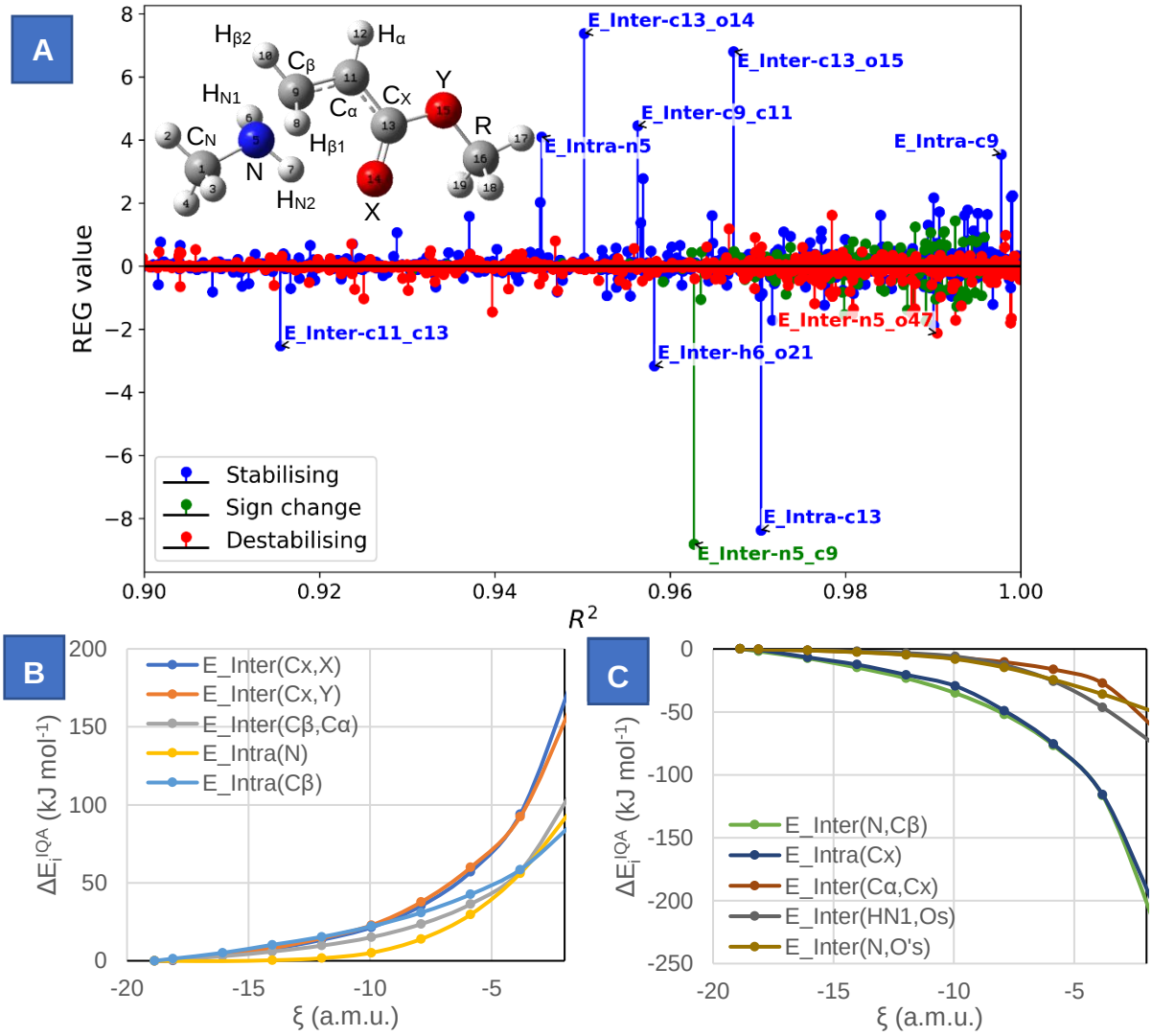


Figure 8. REG-IQA analysis results for system 3 (segment 1). A: Stem plot of the REG-IQA analysis results. B: Variations of main positive-REG IQA terms. C: Variations of main negative-REG IQA terms.

Positive REGs:

The terms with the largest positive REG coefficients on this segment (thus becoming less stabilizing) are related to the C_X atom ($E_{inter}(C_X, X)$ and $E_{inter}(C_X, Y)$), the $C_\beta - C_\alpha$ double bond ($E_{inter}(C_\beta, C_\alpha)$) and the two atoms which are bonding ($E_{intra}(N)$ and $E_{intra}(C_\beta)$). While the latter three terms can be expected since they are linked to the weakening of the $C_\beta - C_\alpha$ bond and the preparation of reactants (as E_{intra} depends on the charge and atomic volume of atoms), the first two are rather surprising since

they involve the EWG and the substituent beyond, which are (at first glance) not directly involved in the mechanism.

- Analysis of $E_{inter}(C_X, X)$ and $E_{inter}(C_X, Y)$

Both terms are dominated by the V_{cl} component ($\frac{V_x}{V_{cl}}$ is almost constant at a value around 0.34 at any point of segment 1), which is very well-correlated to the monopole-monopole electrostatic interaction $\frac{q_A q_B}{R_{AB}}$ (R^2>0.998). This component represents a constant value of 76% and 84% of V_{cl} for each term. Both V_{cl} and V_x components are stabilizing on the entire segment but become less stabilizing. Both E_{inter} terms change at the same rate (see graph D of Figure 8). All of this indicates a slight, simultaneous weakening of the $C_X - X$ and $C_X - Y$ bonds, which is reflected in the structural parameters: both interatomic distances increase slowly on segment 1 while the QTAIM charges remain almost constant, leading to weaker coulombic interactions.

- Analysis of $E_{inter}(C_\beta, C_\alpha)$

This term corresponds to the weakening of the $C_\beta - C_\alpha$ bond. V_x and V_{cl} have opposing variations, V_x dominates over V_{cl} ($\frac{V_x}{V_{cl}}$ ratio ~ 8.5), is very well-correlated to the delocalization index DI_{AB} divided by the internuclear distance R_{AB}, $\frac{DI_{AB}}{2R_{AB}}$ (R^2>0.999) and accounts for most of the variation of the E_{inter} term, showing a loss of covalency that cannot be counterbalanced by the coulombic attraction. The interatomic distance starts to increase at the end of this segment.

- Analysis of $E_{intra}(N)$ and $E_{intra}(C_\beta)$

The $E_{intra}(N)$ and $E_{intra}(C_\beta)$ terms are related to the preparation of reactants, which is the charge transfer between these atoms and the deformation of their electronic densities as they grow closer in order to form the $C_\beta - N$ bond. Both terms have similar REG coefficients (4.09 and 3.54). The variation of the $E_{intra}(C_\beta)$ term is well-correlated to that of the C_β QTAIM charge (R^2=0.98, see left graph of Figure 9), as the C_β loses electrons at a steady rate throughout the segment, which unexpectedly means it is not electrophilic. This was already discussed in our previous article (*see chapter 4*). This implies that the variation of $E_{intra}(C_\beta)$ is mostly due to charge transfer.

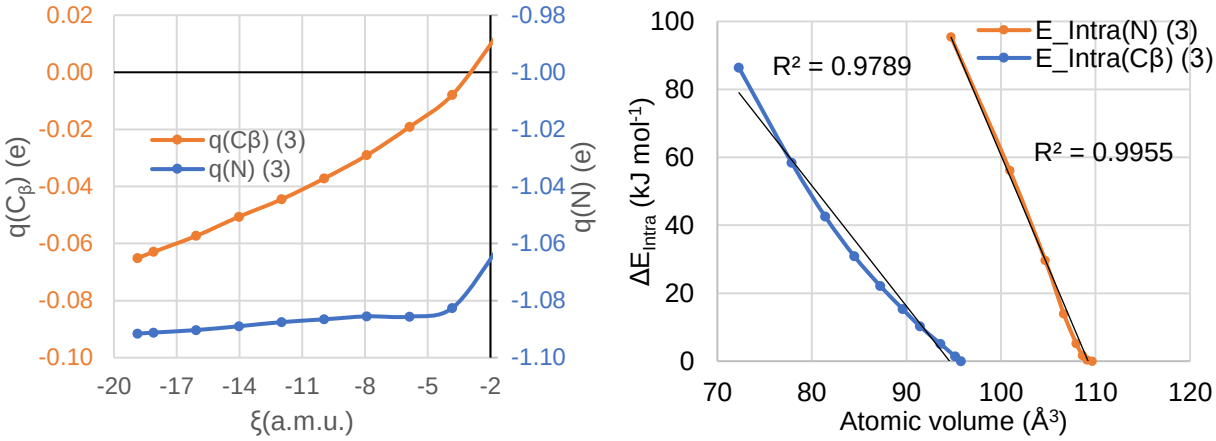


Figure 9. Evolution of parameters related to N and C_β for system 3 (segment 1). Left: QTAIM charges of N and C_β along the IRC. Right: $E_{intra}(N)$ vs. N atomic volume.

On the other hand, this good correlation does not hold for N (R^2=0.88). The N charge is almost constant on the entire segment (see left graph of Figure 9), owing to a compensation of the electron transferred to C_β by charge transfer from the methyl substituent. Charge transfer alone is thus unable to explain the variation of $E_{intra}(N)$. Previous studies [51] have shown that in this case, the E_{intra} variation can be linked to the change of the atomic volume, so that E_{intra} effectively quantifies the steric interactions through the atomic volume deformation: this is confirmed by the excellent correlation of $E_{intra}(N)$ with the atomic volume of N in Figure 9. This correlation also holds for C_β (though to a lesser degree); both atoms shrink on the entire segment as they come closer and change hybridization (sp^2 to sp^3 for C_β). This could be evidenced by a NBO study.

Negative REGs:

The terms with the largest negative REG coefficients on this segment (becoming more stabilizing) correspond to the $N - C_\beta$ bond that will soon form ($E_{inter}(N, C_\beta)$), the C_X atom ($E_{intra}(C_X)$), two interactions with the solvent ($E_{inter}(H_{N2}, O_{DMSO})$ and $E_{inter}(N, O'_{DMSO})$) and the $C_\alpha - C_X$ single bond ($E_{inter}(C_\alpha, C_X)$). The two solvent interactions will be discussed more in detail in section 4.3.

- Analysis of $E_{inter}(N, C_\beta)$

This term indicates the formation of the $N - C_\beta$ bond. For this system, the term is slightly destabilizing at first with both atomic charges being negative, then changes sign as the C_β charge changes sign from negative to positive. V_{cl} dominates at first but is overtaken by V_x in the middle of the segment, with a $\frac{V_x}{V_{cl}}$ ratio of -0.5 at the start of the segment and +10.0 at the end. V_{cl} correlates well with $\frac{q_A q_B}{R_{AB}}$ (R^2=0.975), which represents 50 to 80% of V_{cl}, while V_x is very well-correlated to $\frac{DI_{AB}}{2R_{AB}}$ (R^2>0.999). The $N - C_\beta$ distance continuously decreases from 3.1 to 2.2 Å on this segment as the reactants approach to form the bond.

A BCP (Bond Critical Point) also appears between both atoms near the middle of the segment, as the interatomic distance reaches 2.9 Å. It is interesting to note that the evolution of the electronic density at the BCP is very well-correlated to that of the $E_{inter}(N, C_\beta)$ (R^2>0.999), with a relationship approximately equal to $E_{inter}(N, C_\beta) \sim -2\rho_{BCP,N-C_\beta}$.

- Analysis of $E_{intra}(C_X)$

This term shows a variation that is almost "synchronous" to that of $E_{inter}(N, C_\beta)$, with almost the same values (see graph C of Figure 8). This means that the approach of the reactants coincides with a simultaneous charge transfer to C_X, which continuously gains electrons over the first segment. This takes place at the same time as the stretch of the two C-O bonds with oxygens X and Y.

- Analysis of $E_{inter}(C_\alpha, C_X)$

This term indicates the strengthening of the $C_\alpha - C_X$ bond. Like the $C_\beta - C_\alpha$ term, V_x dominates over V_{cl} ($\frac{V_x}{V_{cl}}$ ratio ~ 6) and the two components have opposite variations, with V_{cl} becoming more destabilizing and V_x becoming more stabilizing, indicating a higher covalency and thus a stronger bond. This can be seen in the excellent correlation of V_x with $\frac{DI_{AB}}{2R_{AB}}$ (R^2>0.999). Unlike the other main terms, V_{cl} does not correlate at all with $\frac{q_A q_B}{R_{AB}}$ (R^2=0.26); the monopole-monopole interaction even has a sign opposite to V_{cl}, meaning that other contributions dominate in V_{cl}.

This change is reflected in the $C_\alpha - C_X$ distance which decreases slightly from 1.47 to 1.44 Å over the segment.

Differences with other systems:

As for the significant terms of **4** and **6**: system **4** also features a $E_{inter}(C_X, X)$ term (REG=3.31), while system **6** shows both $E_{inter}(C_X, X)$ and $E_{inter}(C_X, Y)$ on this segment, with large positive REG coefficients (REGs=5.56 and 4.99). The difference is only due to the fact that the YR substituent and by extension the $E_{inter}(C_X, Y)$ term do not exist in system **4**.

Both systems **4** and **6** also feature $E_{intra}(N)$, $E_{inter}(C_\beta, C_\alpha)$, and $E_{intra}(C_\beta)$ with large positive REG coefficients (5.27, 4.33, 2.80 for **4** and 4.05, 3.75, 2.53 for **6**).

The $E_{inter}(N, C_\beta)$, $E_{inter}(C_\alpha, C_X)$ and $E_{intra}(C_X)$ terms are also significant for systems **4** and **6** (REGs = -8.97, -2.11, -2.62 for **4** and -7.71, -5.20, -2.15 for **6**). $E_{inter}(N, C_\beta)$ is always stabilizing for these two systems.

It must be noted that due to the differing nature of the substituent, the C_X charge has quite different values in each compound: around 1.55 e in **3**, 1.40 e in **6** and 0.95 e in **4**. This is consistent with the different electronegativities of the heteroatoms to which C_X is bonded (two oxygens in **3**, one oxygen and one nitrogen in **6** and only one nitrogen in **4**). This is also true, though at a lesser extent, for C_α.

As seen before, most of the previously discussed terms are also observed for the other systems. The notable differences include a positive-REG interaction between N and a solvent hydrogen atom as well as a second hydrogen bond with the solvent for system **4.**

Surprisingly, an interaction between one of the methylamine hydrogens (H_{N2}) and X is significant for systems **4** and **6**, showing that this interaction is becoming more stabilizing (negative REG coefficients: -2.10 and -2.21). We could have expected this term to be significant for **3** rather than **4**, since the X atom is much farther from H_{N2} in **4** than in **3** and **6** (2.56 – 2.65 Å for **3** and **6**, 4.57 Å for **4** at the end of

segment 1). The term actually has a low correlation (R^2 = 0.52) for system **3** due to a non-monotonous variation (possibly linked to the non-monotonous variation of the interatomic distance). However, this interaction is not (yet) a hydrogen bond due to the long distance between these atoms.

Overall, the terms in this segment show evidence of the system preparing for the reaction proper, with interactions of the entire conjugated system coming into play as well as intra-atomic terms of atoms which undergo charge transfer. The bonds have not started to significantly stretch or shrink, except the $N - C_\beta$ which is almost 1 Å shorter at the end of segment 1 than at the start.

Segment 2

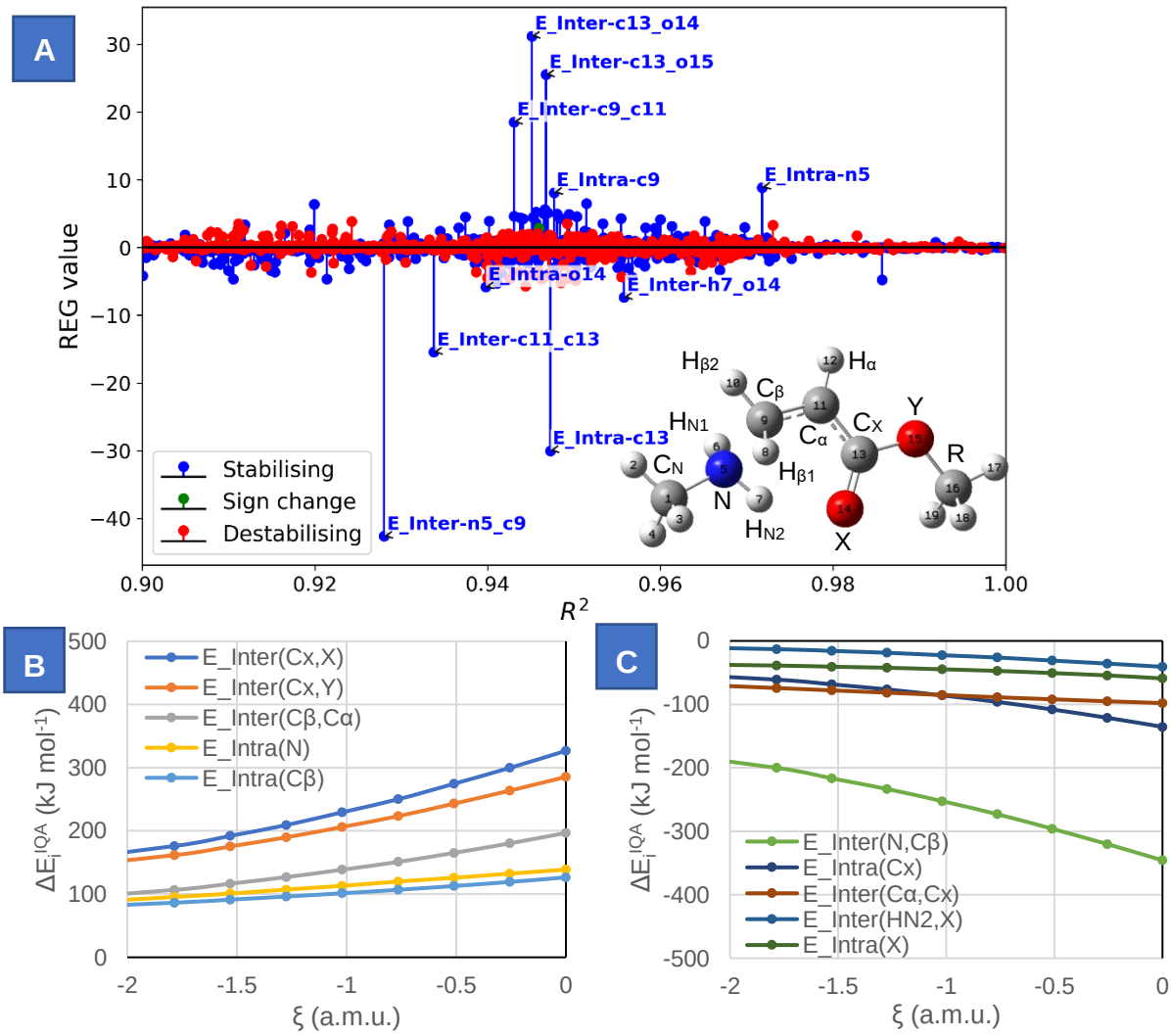


Figure 10. REG-IQA analysis results for system 3 (segment 2). A: Stem plot of the REG-IQA analysis results. B: Variations of main positive-REG IQA terms. C: Variations of main negative-REG IQA terms.

Positive REGs:

The terms with the largest positive REG coefficients (thus becoming less stabilizing) on this segment are the same as on segment 1, with the same order: $E_{inter}(C_X, X)$, $E_{inter}(C_X, Y)$, $E_{inter}(C_\beta, C_\alpha)$, $E_{intra}(N)$ and $E_{intra}(C_\beta)$.

It then follows that we can have the same analysis as on segment 1, except that the variations of the terms, charges and distances are larger, particularly the N and C_α charges. The $C_X - X$, $C_X - Y$, $C_\beta - C_\alpha$ bonds all weaken while N and C_β undergo charge transfer in order to form the bond.

Negative REGs:

The terms with the largest negative REG coefficients on this segment (hence becoming more stabilizing) are quite different from segment 1: $E_{inter}(N, C_\beta)$, $E_{intra}(C_X)$, $E_{inter}(C_\alpha, C_X)$, $E_{inter}(H_{N2}, X)$ and $E_{intra}(X)$. The two latter interactions are respectively related to the intramolecular hydrogen bond and to the carbonyl oxygen X. $E_{inter}(H_{N2}, X)$ was common to the two other systems so it is not as surprising to have it on this segment. $E_{inter}(H_{N2}, O_S)$ falls at the 17th place of the REG ranking.

The three terms common to segments 1 and 2 follow the same evolution on segment 2, with a notable acceleration of their variations. We will now focus on the new terms appearing in this segment.

- Analysis of $E_{inter}(H_{N2}, X)$

This time, the interaction between H_{N2} (the hydrogen borne by N engaging in an intramolecular hydrogen bond) and X is significant for system **3**. The V_{cl} component makes up more than 90% of the $E_{inter}(H_{N2}, X)$ term, as expected from a non-covalent, electrostatics-driven bond. Like previous terms, V_{cl} is also perfectly correlated to the corresponding monopole-monopole interaction ($R^2 > 0.999$ on segment 2). At the end of this segment (the TS), the $H_{N2} - X$ distance reaches 2.5 Å, which is the usual threshold distance to define a H-X interaction as a hydrogen bond.

- Analysis of $E_{intra}(X)$

The $E_{intra}(X)$ term is linked to the $E_{intra}(C_X)$ and $E_{inter}(C_X, X)$ terms seen previously, with the charge transfer going on around the EWG. The term variation is very well-correlated to the QTAIM charge variation of the corresponding atom (R^2=1.00), with a slight decrease of 0.07 e. We can then conclude that the charge transferred from N and C_β ends up (partly) on X even if the $C_X - X$ bond is weakened.

Overall, the $N - C_\beta$ bond keeps on forming (the $N - C_\beta$ passes the 2 Å threshold at the TS), the $C_\beta - C_\alpha$, C_X, X and C_X, Y bonds are still weakening and the $C_\alpha - C_X$ strengthens.

Differences with other systems:

In the other systems, almost exactly the same positive-REG terms are significant: $E_{inter}(C_X, X)$, $E_{inter}(C_\beta, C_\alpha)$, $E_{intra}(N)$ and $E_{intra}(C_\beta)$. Of course, in **4,** $E_{inter}(C_X, Y)$ is replaced by another term (an interaction between N and a DMSO sulfur).

The situation is different for the negative-REG terms: only $E_{inter}(N, C_\beta)$, $E_{intra}(C_X)$, and $E_{inter}(C_\alpha, C_X)$ are common to all three systems. $E_{inter}(H_{N2}, X)$ is only shared by **6** and $E_{intra}(X)$ is only significant for **3**. In **4**, the remaining significant terms are the solvent hydrogen bond, $E_{inter}(H_{N1}, O_{DMSO})$ and $E_{inter}(N, O'_{DMSO})$, and for **6**, $E_{intra}(Y)$.

Segment 3

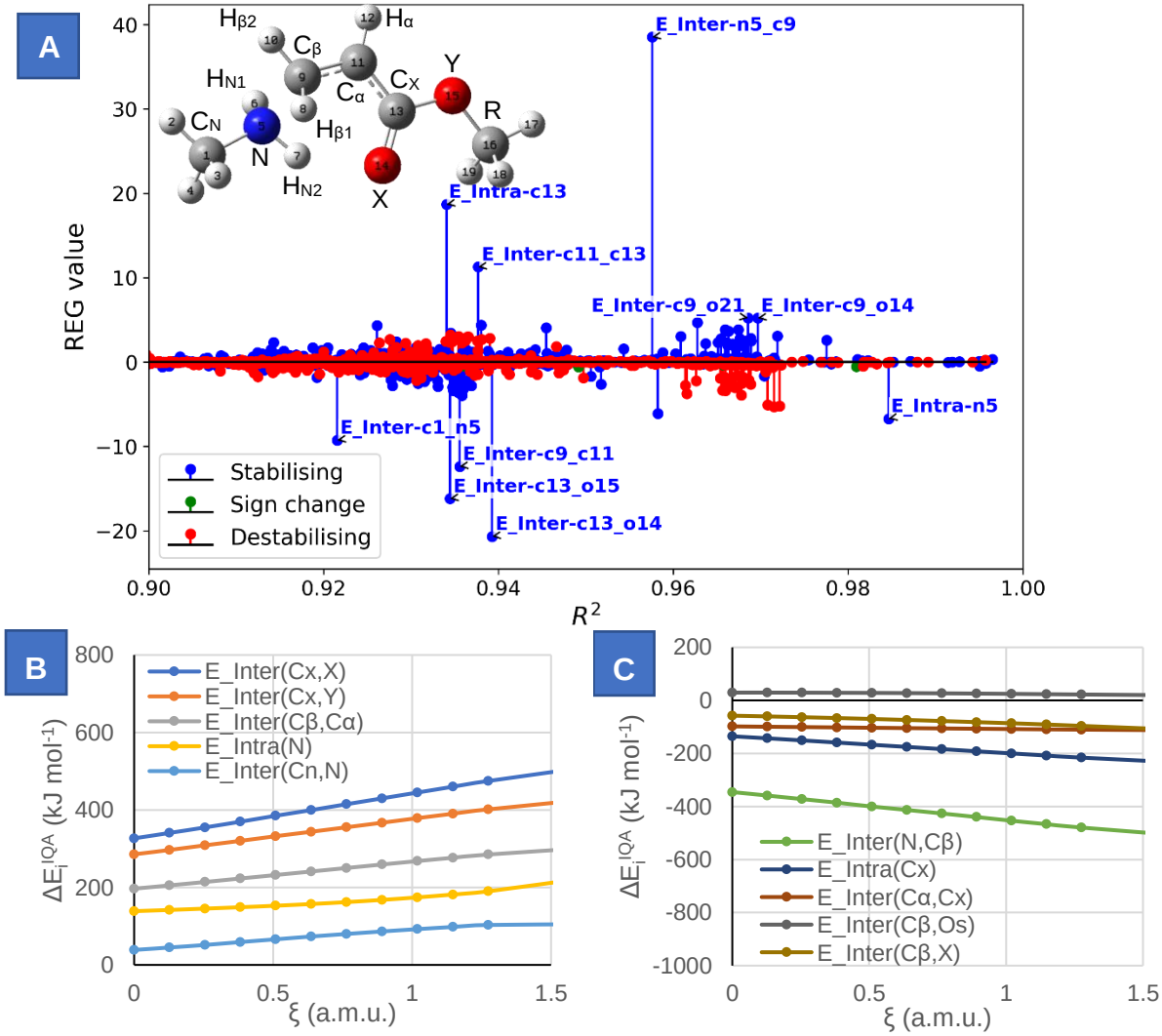


Figure 11. REG-IQA analysis results for system 3 (segment 3). A: Stem plot of the REG-IQA analysis results. B: Variations of main positive-REG IQA terms. C: Variations of main negative-REG IQA terms.

On this segment, the total energy decreases instead of increasing, so all positive-REG terms become more stabilizing as well, and all negative-REG terms become less stabilizing. This essentially means that the REG coefficients of most terms have changed sign from previous segment.

Positive REGs:

Much like the first two segments, the terms becoming more stabilizing (so with the largest positive REG coefficients) on this segment correspond to the $N - C_\beta$ bond ($E_{inter}(N, C_\beta)$), the C_X atom ($E_{intra}(C_X)$),

the $C_\alpha - C_X$ bond ($E_{inter}(C_\alpha, C_X)$), one interaction with the solvent ($E_{inter}(C_\beta, O_{DMSO})$) and one long-distance term ($E_{inter}(C_\beta, X)$). The solvent interaction will be discussed more in detail in section 4.3.

- Analysis of $E_{inter}(C_\beta, X)$

Both components have opposite variations: V_{cl} becomes more stabilizing while V_x destabilizes the system. V_{cl} dominates this term on segment 3, with an ever-decreasing $\frac{V_x}{V_{cl}}$ ratio from 0.51 to 0.23, which is understandable given the unchanging distance between both atoms (2.8 Å), the increase of the C_β charge (from 0.05 to 0.12 e) and the X charge which is the largest negative charge of the system (-1.28 e). It is very well-correlated to $\frac{q_A q_B}{R_{AB}}$ (R^2>0.999), which represents the quasi-totality of V_{cl} (93-100%). This long-distance term is not the most intuitive one.

Negative REGs:

Like in the first two segments, the terms becoming less stabilizing (with the largest negative REG coefficients) on this segment are related to the X atom ($E_{inter}(C_X, X)$ and $E_{inter}(C_X, Y)$), the $C_\beta - C_\alpha$ double bond ($E_{inter}(C_\beta, C_\alpha)$) and N ($E_{intra}(N)$). It must be noted that the N charge reaches a maximum at the end of this segment (-1.00 e), showing that electronic reorganization is starting and brings back electrons to N.

A new term replaces the $E_{intra}(C_\beta)$: the $E_{inter}(C_N, N)$ term, corresponding to the $C_N - N$ bond in methylamine. It has a negative REG coefficient (REG = -8.36 to -11.27, R^2 = 0.92 to 0.95), and thus becomes less stabilizing. This can be linked to a weakening of the $C_N - N$ bond and to electronic reorganization already starting in the methylamine. Indeed, the $C_N - N$ distance does increase a bit (by 0.03 Å on the entire profile), leading to a decrease in the exchange term between both atoms, and both QTAIM charges decrease in absolute values, leading to a decrease of the stabilizing coulombic interaction between both atoms.

Differences with other systems:

In the other systems, much like the first two segments, the common terms with the largest positive REG coefficients are $E_{inter}(N, C_\beta)$, $E_{intra}(C_X)$, and $E_{inter}(C_\alpha, C_X)$. For **4**, the last interactions are two interactions with the solvent ($E_{inter}(C_\beta, O_{DMSO})$; for **6**, the long-distance term ($E_{inter}(C_\beta, X)$) is also common with **3**, and the last term is $E_{intra}(Y)$. The solvent interaction will be discussed more in detail in section 4.3.

Like in the first two segments, the terms with the largest negative REG coefficients on this segment are $E_{inter}(C_X, X)$, $E_{inter}(C_\beta, C_\alpha)$, $E_{intra}(N)$ and $E_{inter}(C_N, N)$. For **4**, $E_{inter}(C_X, Y)$ is replaced by $E_{intra}(C_\beta)$.

Segment 4

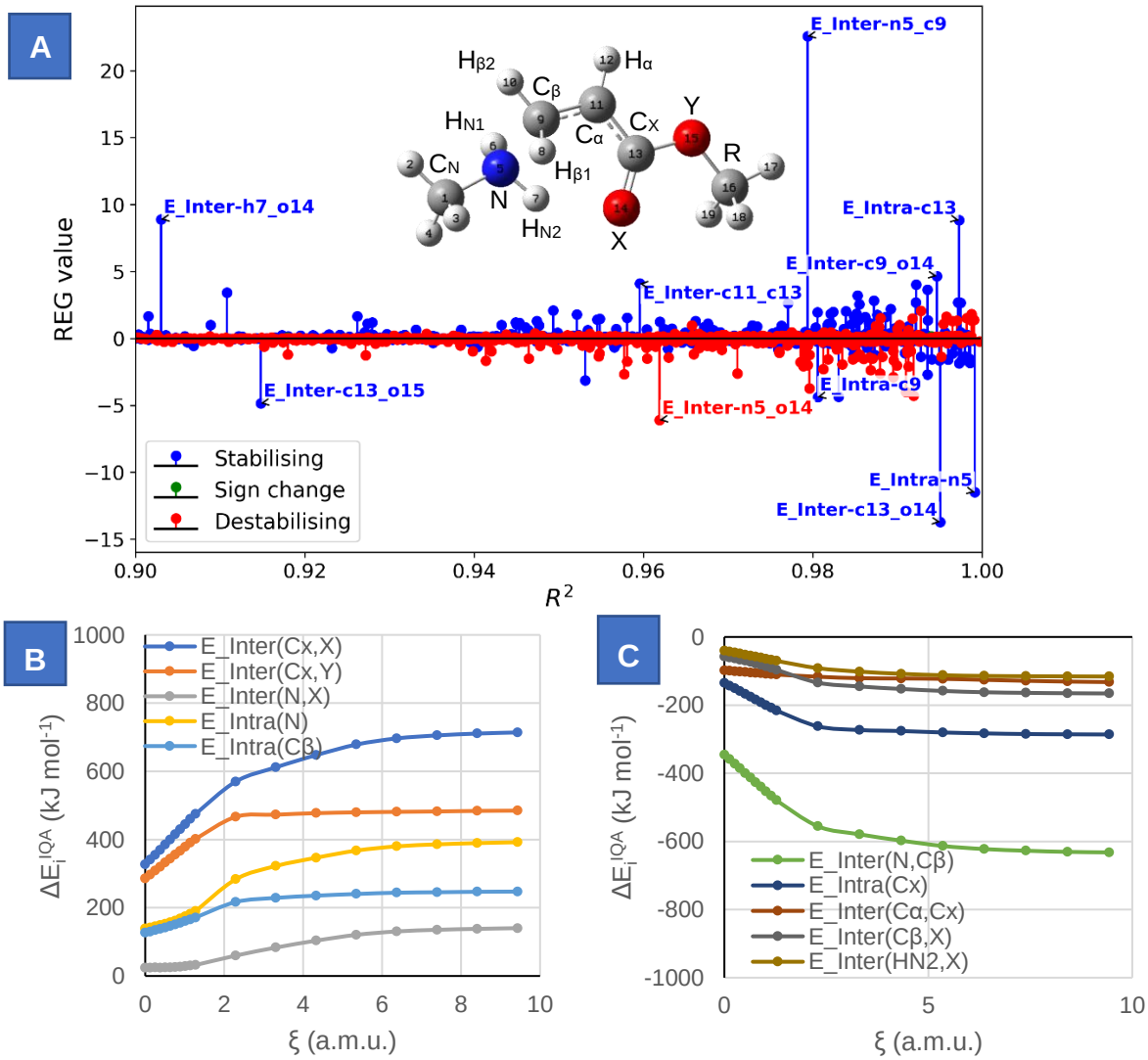


Figure 12. REG-IQA analysis results for system 3 (segment 4). A: Stem plot of the REG-IQA analysis results. B: Variations of main positive-REG IQA terms. C: Variations of main negative-REG IQA terms.

Positive REGs:

Once again, the terms with the largest positive REG coefficients on this segment (becoming more stabilizing) correspond to the $N - C_\beta$ bond ($E_{inter}(N, C_\beta)$), the C_X atom ($E_{intra}(C_X)$), the intramolecular hydrogen bond ($E_{inter}(H_{N2}, X)$), the $C_\alpha - C_X$ bond ($E_{inter}(C_\alpha, C_X)$) and one long-distance term ($E_{inter}(C_\beta, X)$). The solvent interaction will be discussed more in detail in section 4.3.

Negative REGs:

Like in the other segments, the terms with the largest negative REG coefficients (thus becoming more stabilizing) on this segment are related to the X atom ($E_{inter}(C_X, X)$ and $E_{inter}(C_X, Y)$) and the bonding atoms ($E_{intra}(N)$ and $E_{intra}(C_\beta)$) as well as a destabilizing long-distance interaction, $E_{inter}(N, X)$.

- Analysis of $E_{inter}(N, X)$.

This is a destabilizing term, and the first significant destabilizing solute-only term in this analysis. It mainly corresponds to the repulsive interaction between the two largest negative charges of the system, at -1.0 and -1.3 e, as V_{cl} completely dominates V_x, with a $\frac{V_x}{V_{cl}}$ ratio as small as -0.02 to -0.04. Once again, V_{cl} correlates perfectly with $\frac{q_A q_B}{R_{AB}}$ (R^2>0.999), which represents the quasi-totality of V_{cl} (at least 99%). This is another surprising long-distance term (the $N - X$ distance goes from 3.2 to 2.8 Å on this segment), but that has physical meaning as the two largest negative charges come closer to each other at the end of the reaction.

The $N - C_\beta$ distance reaches a plateau around 1.55 Å while the N and C_β charges decrease (almost returning to its original value) and increase, respectively, before slowly coming to a stop. The $H_{N2} - X$ distance grows shorter and reaches a plateau, around 2 Å.

Differences with other systems:

This time, the common terms with the largest positive REG coefficients on this segment are only $E_{inter}(N, C_\beta)$ and $E_{inter}(C_\beta, X)$. The $E_{intra}(C_X)$ and $E_{inter}(C_\alpha, C_X)$ terms are common to **6**, and the last term for this system is $E_{inter}(H_{N2}, X)$. For **4**, we find the solvent hydrogen bond $E_{inter}(H_{N1}, O_{DMSO})$ and two interactions of the C_β with the solvent, $E_{inter}(C_\beta, O_{DMSO})$ and $E_{inter}(C_\beta, O'_{DMSO})$. The solvent interaction will be discussed more in detail in section 4.3.

Like in the other segments, the terms with the largest negative REG coefficients on this segment are $E_{inter}(C_X, X)$, $E_{intra}(N)$ and $E_{intra}(C_\beta)$. For **4**, the remaining terms are $E_{inter}(C_\beta, C_\alpha)$ and a rather

surprising term, $E_{inter}(H_{N1}, C_\beta)$. For **6**, the destabilizing long-distance interaction $E_{inter}(N, X)$ is also significant, as well as the $E_{inter}(C_X, Y)$.

In conclusion to this part, intuitively, we would have guessed that there are four major events: the preparation of reactants in the first segment (reactant zone), the formation of the $N - C_\beta$ bond and the weakening of the $C_\beta = C_\alpha$ bond, both taking place in the two central segments (TS zone), followed by electronic reorganization in the last segment (product zone).

The IQA picture is a bit different: we do retrieve these four events through the $E_{intra}(N)$, $E_{intra}(C_\beta)$ (reactant preparation), $E_{inter}(N, C_\beta)$ (bond formation), $E_{inter}(C_\beta, C_\alpha)$(bond weakening), $E_{intra}(C_X)$ and $E_{intra}(X)$ (electronic reorganization).

However, several other interactions are consistently significant. Among the terms stabilizing the system, the $E_{inter}(C_\alpha, C_X)$ term, linked to a bond reinforcement, is present on all segments. The $E_{inter}(C_X, X)$ and $E_{inter}(C_X, Y)$ terms, related to a weakening of the corresponding bonds, even always have the largest REG coefficients, meaning the Y substituent cannot be dismissed as an inert substituent.

A variety of other interactions, mostly stabilizing, can also be seen in the analysis, notably the intra- or intermolecular hydrogen bonds, $C_N - N$ interactions, long-range interactions between the bonding atoms and the EWG… This means that more of the system is involved in the reaction than we would have thought using the mechanism drawn in Figure 5 in section 4.1, with charge transfer happening in the methylamine, in the entire conjugated system and even with the substituent.

We can then deduce the following mechanism shown on Figure 13 using IQA:

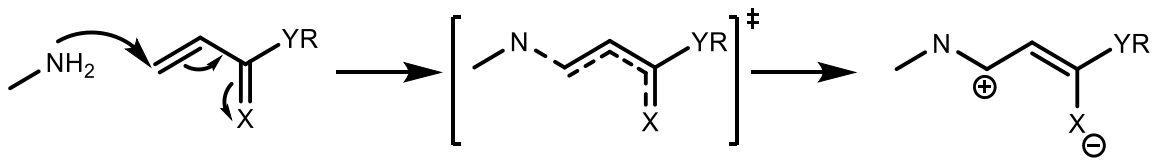


Figure 13. Scheme of the aza-Michael reaction mechanism from the IQA point of view.

This is quite different from the expected mechanism where the positive charge would rather end up on N and the negative charge on C_α. If we include all QTAIM charges, most of the important atoms are clearly positively or negatively charged either at the start or at the end of the reaction.

We can use this analysis to define the different events from the QTAIM and IQA points of view: in this framework, the reactant preparation is associated with a destabilization of the E_{intra} terms of two unbonded atoms between which no BCP exists or which are sufficiently far away (more than 2 Å). The covalent bond formation is associated to a strong stabilization of the E_{inter} term and, most importantly, of the V_x component (which must dominate over the V_{cl}), with the apparition of a BCP, while the hydrogen bond formation is characterized by a domination of the V_{cl} component instead. The bond weakening or strengthening is then linked to a decrease or increase of the E_{inter} term, respectively, without any appearance or disappearance of the associated BCP. The electronic reorganization can be loosely viewed as changes in the IQA terms of the system, notably E_{intra} terms, without any large changes in the interatomic distances.

4.3 Solvent influence

Now that we have a better understanding of the reaction mechanism, we can turn to the influence of the solvent. As we have seen before, the activation barriers are noticeably lower in DMSO than in gas phase, and the reaction of **4** is simply impossible in gas phase, meaning that the solvent considerably stabilizes the system.

In order to understand the solvent influence, we are going to study first the QTAIM charges, looking for eventual solute-solvent charge transfer, and then the IQA terms, as the E_{intra} terms can show polarization effects and the E_{inter} terms are more obviously markers of interactions between each single atom, entire molecules or the whole solvent.

Charges

The analysis of the charges of solvent molecules proves that there is no net charge transfer between the solvent and the solute: as shown on Figure SI-X, the charge variations of each solvent molecules remain close to 0.01 e, while the charge variation of each reactive fragment reaches 0.35 e by the end of the reaction. If any single solvent atom charge varies significantly, the rest of the molecule to which it belongs cancels it out.

Decomposition of solute and solvent interactions

Let us now consider the IQA terms. In order to have a first broad look at the solvent terms, we can break them down into categories. Indeed, we can write the total energy as:

$$E_{mol} = E_{solute} + E_{solute-solvent} + E_{solvent} \tag{13}$$

with E_{solute}, the sum of all terms involving only the solute, $E_{solute-solvent}$, the sum of interaction terms between the solute and the solvent, and $E_{solvent}$, the sum of all terms involving only the solvent. The first and third terms are respectively equal to:

$$E_{solute} = E_{intra-solute} + E_{inter-solute} \tag{14}$$

$$E_{solvent} = E_{intra-solvent} + E_{inter-solvent} + E_{solvent-solvent} \tag{15}$$

We have computed most of these quantities as sums of IQA terms: the intra- and inter-atomic solute terms $E_{intra-solute}$ and $E_{inter-solute}$, the intra-atomic solvent terms $E_{intra-solvent}$, and the interactions between atoms of the same solvent molecule $E_{inter-solvent}$. The only missing term is the solvent-solvent interactions $E_{solvent-solvent}$. The complete energy formula is then:

$$E_{mol} = E_{intra-solute} + E_{inter-solute} + E_{solute-solvent} + E_{intra-solvent} + E_{inter-solvent} + E_{solvent-solvent} \tag{16}$$

Thus, we simply have to subtract all terms that are available to us from the total energy to access the missing terms:

$$E_{solvent-solvent} = E_{mol} - (E_{solute} + E_{solute-solvent} + E_{intra-solvent} + E_{inter-solvent}) \tag{17}$$

Obviously, this is something we could not have seen with an implicit solvent model, which would only have afforded a single, indivisible term.

As shown on Figure 14, E_{solute} features an energy maximum near the TS and a slight decrease of energy afterward (on the scale of ~10 kJ mol^{-1}) for compounds **3** and **6**, while for **4**, this gives an ever-increasing profile, underlining the necessity of solvent terms to make the reaction possible for this compound.

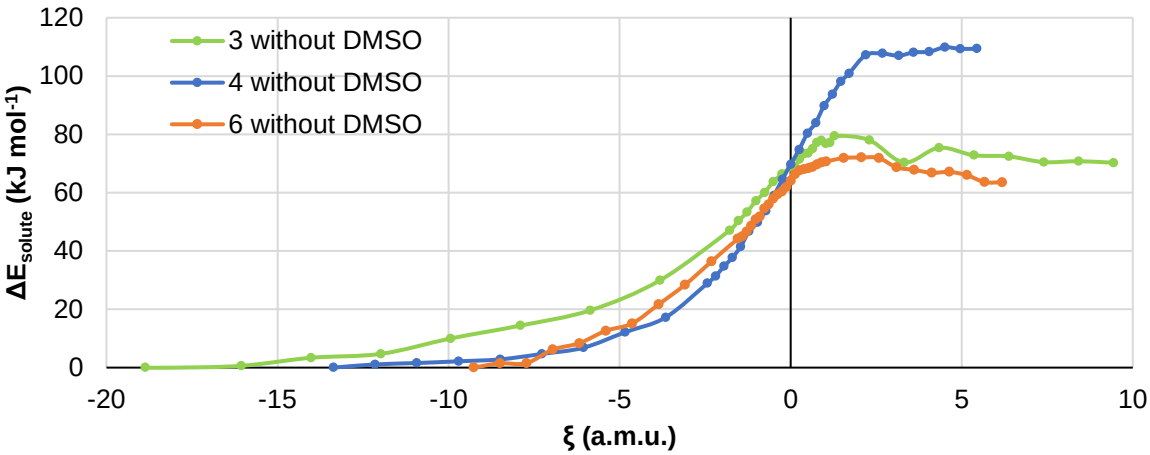


Figure 14. Evolution of E_{solute} along the IRC for systems 3, 4 and 6.

The evolution of the other quantities defined above is quite different for the three systems, as shown on Figure 15:

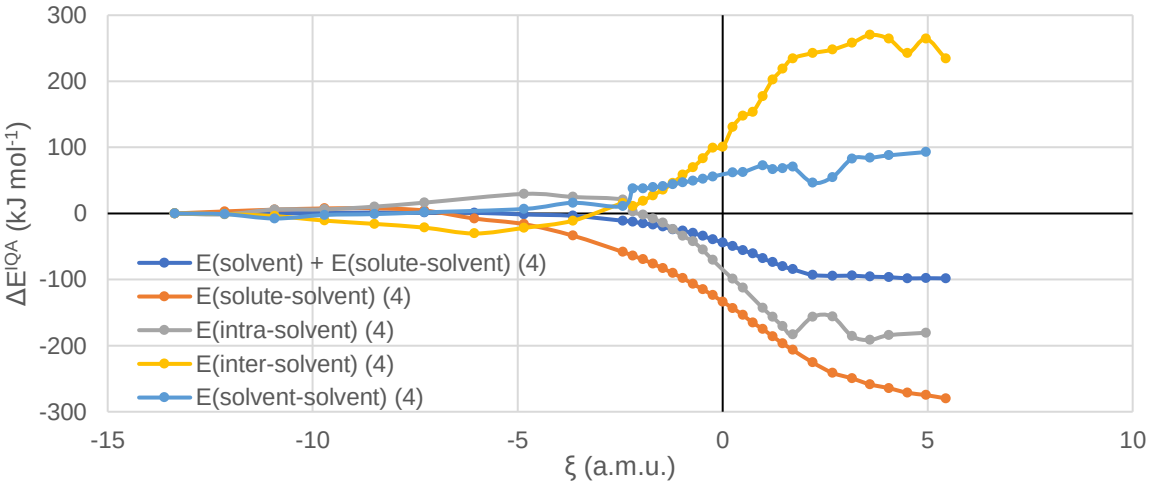


Figure 15. Variation of solvent terms ($E_{solvent} + E_{solute-solvent}$, $E_{solute-solvent}$, $E_{intra-solvent}$, $E_{inter-solvent}$, $E_{solvent-solvent}$) of compound 4.

The sum of solvent terms $E_{solvent} + E_{solute-solvent}$ always becomes slightly more stabilizing, a result of a balance of the intra-solvent $E_{intra-solvent}$ and solute-solvent $E_{solute-solvent}$ terms which become more stabilizing, and the interactions inside solvent molecules $E_{inter-solvent}$ which become less stabilizing. That means that to understand how the solvent stabilizes the system, we must look into more detail at the $E_{solute-solvent}$ terms.

The interactions between solvent molecules ($E_{solvent-solvent}$) can become more stabilizing (**3** and **6**) or destabilizing (**4**). For this system only, the solute-solvent interactions $E_{solute-solvent}$ exceed the intra-solvent terms $E_{intra-solvent}$, being three times more stabilizing than the same term in **3** and **6**, leading to the most stabilizing global solvent term of all three compounds.

The increased destabilization between solvents for **4** can be understood as a consequence of the small size of **4**: with the same number of solvent molecules, these are bound to be closer or to have more interactions with each other rather than with the solute. It can also be a consequence of the limited size of our system: in a system taking into account additional spheres of solvation, the solvent molecules of the first sphere could be stabilized by the solvent molecules of the second sphere, and so on. This is an additional argument as to why we can avoid calculating the IQA solvent-solvent terms in our case: if these terms are bound to be incorrect due to the limited size of the system, we would have increased calculation times and wrong conclusions.

Just as the solute-only terms studied above, most terms for **3** and **6** feature an extremum shortly after the TS while the terms of **4** follow an almost monotonous variation. It is interesting to see that even without any extrema appearing in any of these terms, the solvent terms are able to overcome the solute terms and stabilize the system around and after the TS.

It is still difficult to understand why **6** features a higher activation barrier in solvent than in gas phase, since its solvent terms show a stabilization comparable to that of compound **3**.

Summed solute-solvent interactions

If we now turn our attention to the sum of all IQA interactions between the solute and the solvent ($E_{solute-solvent}$), as seen above, this sum is always stabilizing and this stabilization increases strongly over the course of the reaction, especially in the TS zone, but not in the same way for the three systems (see Figure 16).

The compound for which this effect is the strongest is compound **4**: this is in part due to its smaller number of atoms, as the carbon and hydrogen atoms of substituents of **3** and **6** often interact with the solvent in a slightly stabilizing at best or destabilizing manner, and the fact that the second methylamine hydrogen (H_{N2}) atom is "free", i.e. not engaged in an hydrogen bond with the electron-withdrawing group, bringing in additional solvent stabilization through a hydrogen bond with the solvent.

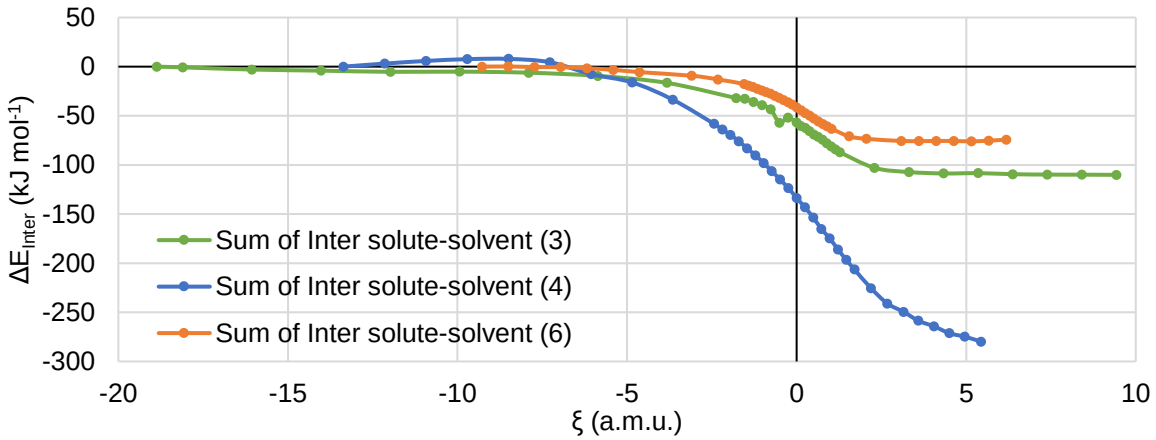


Figure 16. Variation of $E_{solute-solvent}$ for compounds 3, 4 and 6.

It must be noted that the atoms whose interactions with the solvent is the most stabilizing are indeed the methylamine free hydrogen(s) H_{N1} (and H_{N2} for **4**) and the heteroatoms X and Y of the electron-withdrawing group (both ester oxygens for **3**, the nitrogen for **4** and the oxygen and nitrogen of the amide for **6**). On the other hand, N always interacts with the solvent in a destabilizing manner, which is a consequence of its large negative charge: this leads to strong repulsions with the solvent heteroatoms.

Can we see which specific solvent molecules stabilize the system the most? To this end, we can study the summed IQA terms of each solvent molecule, $E_{inter}(S, G)$ (with S being the group of all solute atoms and G the group of the atoms of a given solvent molecule).

As shown on Figure 17, for **4**, each $E_{inter}(S,G)$ term can become stabilizing or destabilizing with no clear trend except the most stabilizing group which is almost three times more stabilizing than the second most stabilizing, and represents half of the total solute-solvent stabilization. For the other two systems, no such term is seen and all solvent terms remain in the same range. Again, we see that compound **4** is quite different from the others.

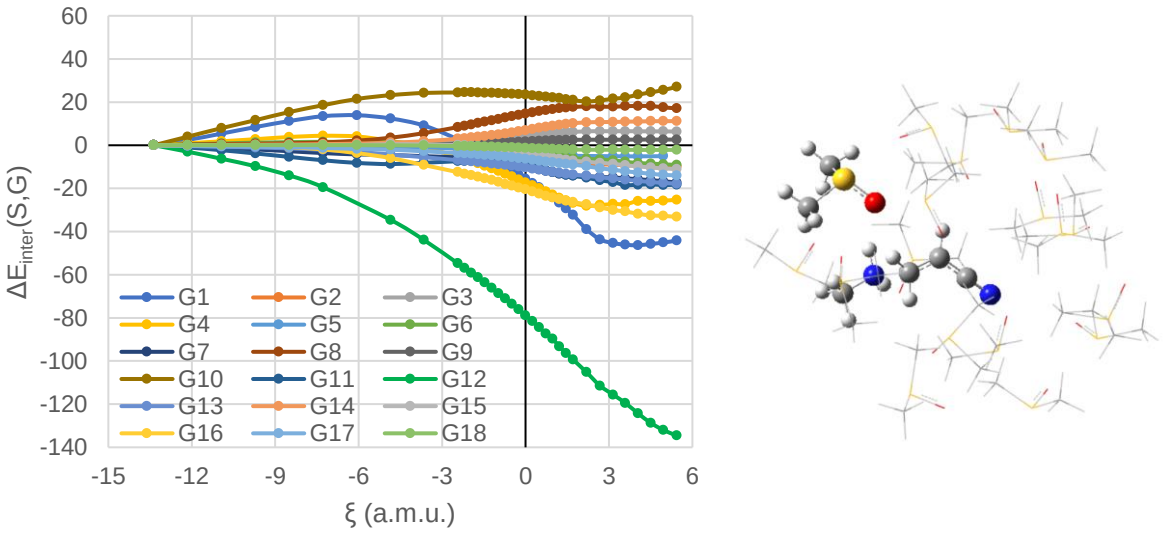


Figure 17. Left: Variation of the summed solute-solvent E_{inter} terms of each solvent molecule over the IRC for system 4. Right: Structure of the TS of system 4, with the solutes and the most stabilizing solvent molecule in ball-and-stick representation.

The variations are often smaller in amplitude compared to the variation of the summed IQA terms of the reactants, usually never reaching more than 60 kJ mol^{-1} in absolute value while the main IQA terms typically reach variations above 100 kJ mol^{-1} and up to 1 000 kJ mol^{-1} for the $E_{inter}(N, C_\beta)$.

<u>4 gas/solvent</u>

Finally, can we explain why the reaction is not possible for **4** in gas phase?

The unfeasibility of the reaction in gas phase is evidenced by a relaxed scan over the $C_\beta - N$ distance in gas phase showing an ever-increasing energy (top curve in Figure 18). Furthermore, performing single-point PCM calculations using the frozen geometries of the gas phase relaxed scan of compound **4** "flattens" the profile, although not to the point where a TS would appear; the destabilization is still reduced by 60% (bottom curve in Figure 18).

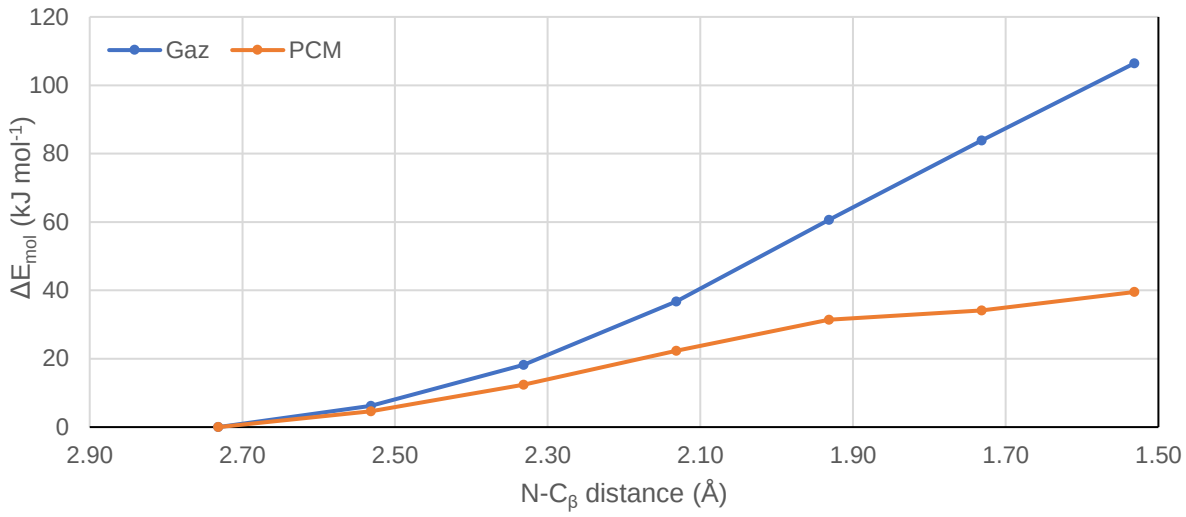


Figure 18. Energy profiles of gas phase relaxed scan and PCM single-points using the gas phase frozen geometry for compound 4.

As we have seen previously on Figure 17, the $E_{solute-only}$ term becomes far more destabilizing for 4 than for the other systems, with no energy maxima, in the same way as the two scans on Figure 22. We have also seen on Figure 19 that the $E_{solute-solvent}$ terms are considerably more stabilizing in **4** than in the other systems, since system **4** forms more stabilizing solute-solvent interactions. The balance of these two contributions brings back the energy profile of **4** to the same level as **3** (both Reverse IRC profiles coincide on Figure 6).

We may also wonder if the lack of a stabilizing hydrogen bond between one of the methylamine hydrogens (H_{N2}) and the EWG (a cyano group, in this case) is also to blame for this.

As shown on Figure 19, in explicit DMSO, the $E_{inter}(H_{N2}, X)$ does indeed become considerably less stabilizing for **4** over the IRC (down to -94 kJ mol^{-1}) than for **3** and **6** (around -250 kJ mol^{-1}), partly due to the longer interatomic distance (always above 4 Å for **7**, and between 2 and 3 Å for **3** and **6**). The combination of these different factors is probably what makes the reaction of **4** unfeasible in gas phase.

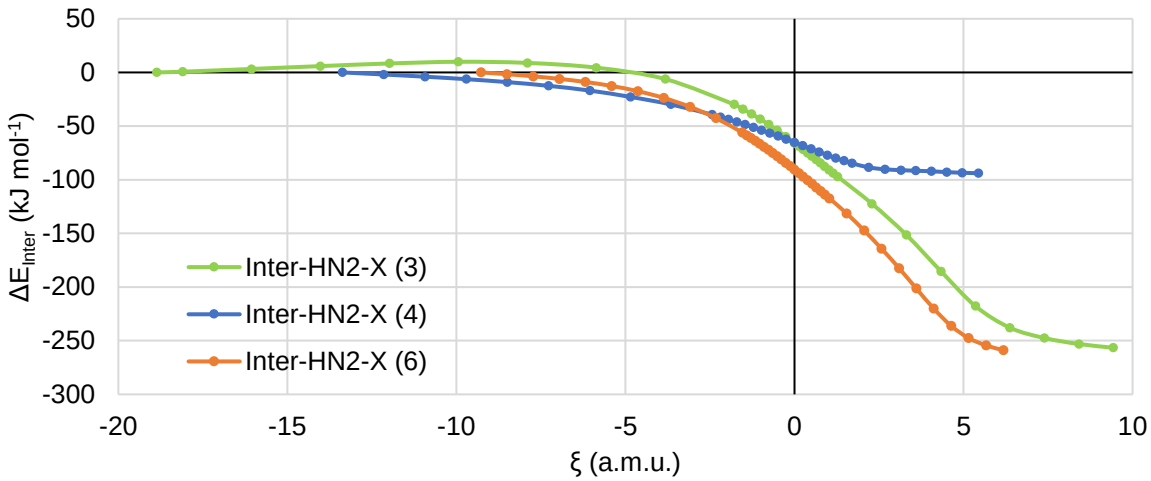


Figure 19. Evolution of $E_{inter}(H_{N2}, X)$ in explicit solvent for systems 3, 4 and 6.

4.4 Link between REG values, activation barriers and Mayr's electrophilicity

Now that we disclosed the full mechanism, links between it and experimental data can be considered. To this aim, we recall that Mayr's electrophilicity experimental values E_{Mayr} are based on rate constant measurements according to:

$$log_{10}(k_{eff}) = s_N(N_{Mayr} + E_{Mayr}) \tag{15}$$

where k_{eff} denotes the pseudo first-order rate constant of the reaction, and s_N and N_{Mayr} are parameters related to the nucleophile. Here, we consider only one nucleophile, so this equation can be rewritten as:

$$log_{10}(k_{eff}) = AE_{Mayr} + B \tag{16}$$

with constants $A = s_N$ and $B = s_N N_{Mayr}$. On the other hand, k_{eff} is linked to the activation energy of the reaction E_{act} by invoking Arrhenius' law:

$$log_{10}(k_{eff}) = CE_{act} + D \tag{17}$$

with C and D pure constants at a given temperature.

It follows that the Mayr electrophilicity is linked to the activation energy, therefore understanding the activation barrier equals understanding the Mayr electrophilicity.

Thus, we may wonder if there is a correlation between the activation barriers and the Mayr electrophilicity. It appears that they do follow the trend but without any actual correlations, reaching R^2

= 0.76 at best for the free energy activation barrier in explicit DMSO (see Figure 20). However, the explicit solvent model performs better than the gas phase model.

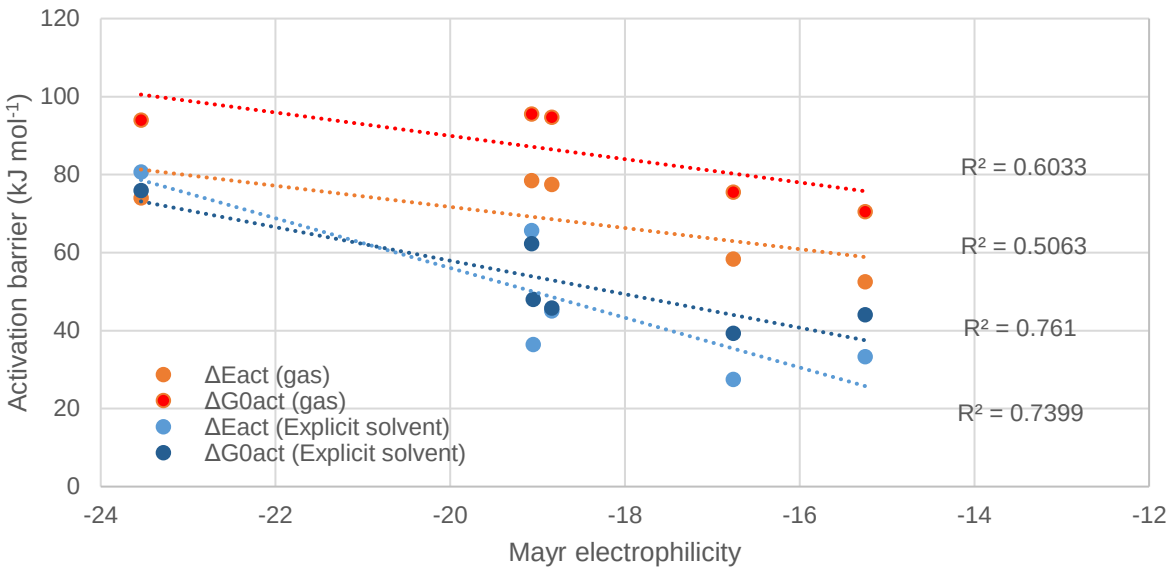


Figure 20. Correlations between the Mayr electrophilicity and the energy and free energy activation barriers in both gas phase and explicit solvent.

This discrepancy can be explained by two factors: one is the restricted and non-representative sampling of solvent configurations of our study, which could only afford considering one solvent configuration for each system without any control on the position of solvent molecules between systems. Thus, we might not necessarily have found the minimal energy path followed by the reaction for these systems, and even not the same path for all systems. The other factor is the fact that our ΔG^0_{act} computation does not take into account the additional barrier between the isolated reactants and the non-covalent intermediate **I**, due to the difficulty of determining a correct reference point for the isolated reactants. This issue does not affect the total energy activation barrier as the formation of non-covalent complexes always leads to a stabilization of the system in E_{mol}.

Since the barriers in explicit solvent are not well-correlated to E_{Mayr}, our search here will focus on finding IQA terms which could be correlated to E_{Mayr}. Given that we showed previously that the activation barriers obtained in explicit solvent are poorly correlated to E_{Mayr}, we can also search for IQA terms that could retrieve these barriers.

To this end, we have performed linear regressions to see which terms reproduce best the energy profile before the TS; that is, the same approach as REG, but in reverse:

$$E_{total} = A_i \cdot E_i(A) + c'_i \quad (17)$$

If the c'_i coefficient equals zero, then $A_i = 1/m_i^{REG}$. To evaluate the quality of these regressions, the determination coefficient is used (like in the REG approach) as well as the MAE (Mean Absolute Error). The terms are ranked according to their MAE; terms with low REG coefficients (below 0.1) are not considered since they only show a small variation.

Since we have all relevant terms only for compounds **3**, **4** and **6**, they are used to sort out the relevant terms, while compounds **1**, **2** and **5** are used to ensure the correlations remain consistent for the complete set.

First, no terms with positive A_i coefficients have been found using this approach because none of the smallest-MAE terms followed the electrophilicity trend.

We then investigated negative-A_i terms (see Table SI-X for detailed rankings): $E_{inter}(H_{\beta 1}, H_{\beta 2})$ (the interaction between both hydrogens borne by the C_β) is a good candidate with low MAEs (maximum value is 1.15 kJ mol^{-1} for **7**) and good R^2 (lowest is 0.98 for 7).

For the complete set, it is able to retrieve the energy activation barriers with a good correlation (R^2=0.86) and the free energy activation barriers with an even better correlation (R^2=0.90), as shown on Figure 21:

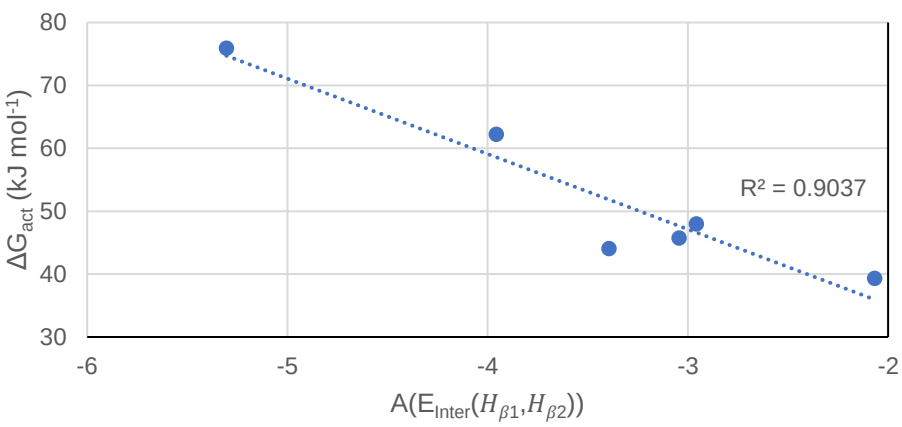


Figure 21. Correlation between the free energy barriers in explicit solvent and the A_i coefficient of $E_{inter}(H_{\beta 1}, H_{\beta 2})$.

This term can be understood as a probe to show the change of hybridization of C_β, since the two hydrogens go from a trigonal, planar geometry to a tetrahedral one and thus undergo a change in the $H_{\beta 1} - C_\beta - H_{\beta 2}$ angle as well as their distances. However, while we could expect the interaction between these two atoms to become less stabilizing due to increased electrostatic repulsion and steric hindrance, this is not the case since the A_i coefficient (and by extension the REG coefficient) is negative, indicating that this term opposes the activation barrier. The charges of both hydrogens are slightly positive at the start of the reaction and decrease until the TS (see Figure 22).

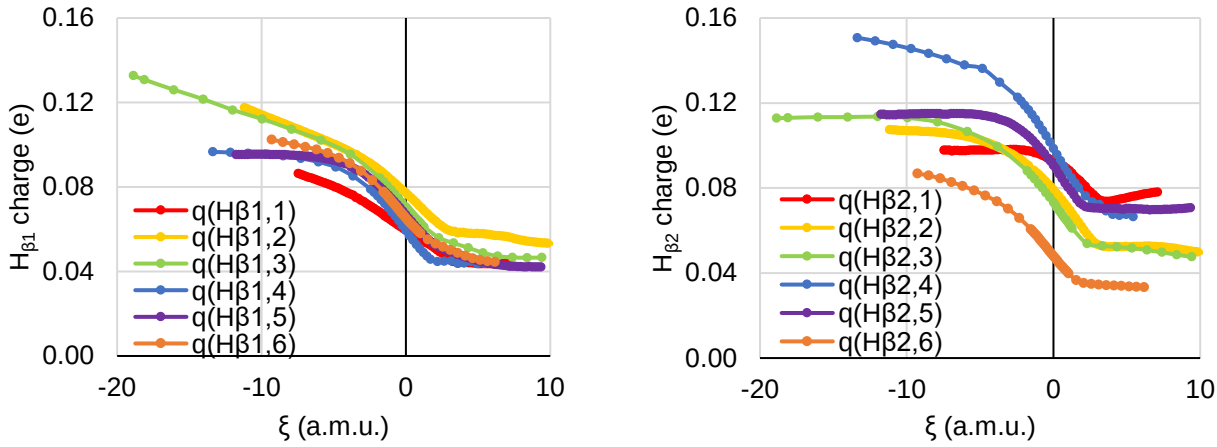


Figure 22. Variation of the QTAIM charges of $H_{\beta 1}$ (left) and $H_{\beta 2}$ (right) along the IRC.

However, it fails quite badly with the Mayr electrophilicities (R^2 = 0.59) due to the deviant values obtained for **1** and **5** (see Figure 23). In particular, the term of **5** shows a parabolic profile instead of the linear profiles obtained for other compounds.

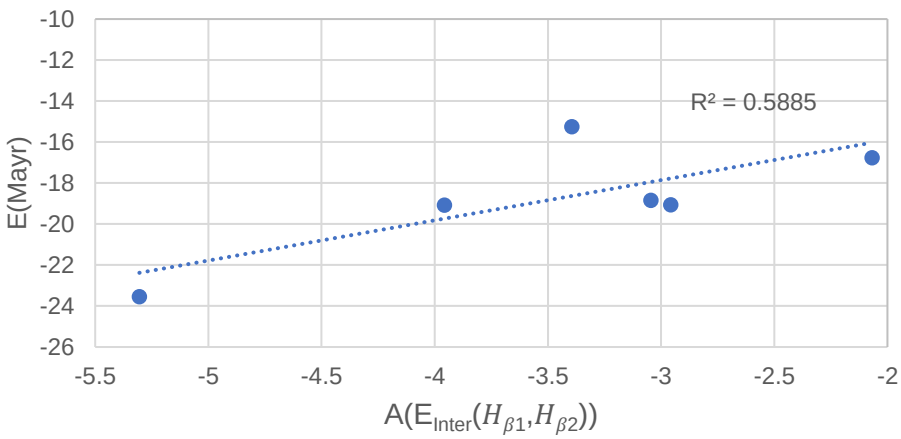


Figure 23. Correlation between E_{Mayr} and the A_i coefficient of $E_{inter}(H_{\beta 1}, H_{\beta 2})$.

Another low-MAE term, $E_{intra}(H_\alpha)$ (the intra-atomic term of the hydrogen borne by C_α), is able to retrieve the electrophilicity values with a satisfying accuracy, even if the correlation is not very good for compound **3** (MAE = 2.99 kJ mol^{-1} and R^2 = 0.89): as shown on Figure 24, the overall correlation is now R^2 = 0.91.

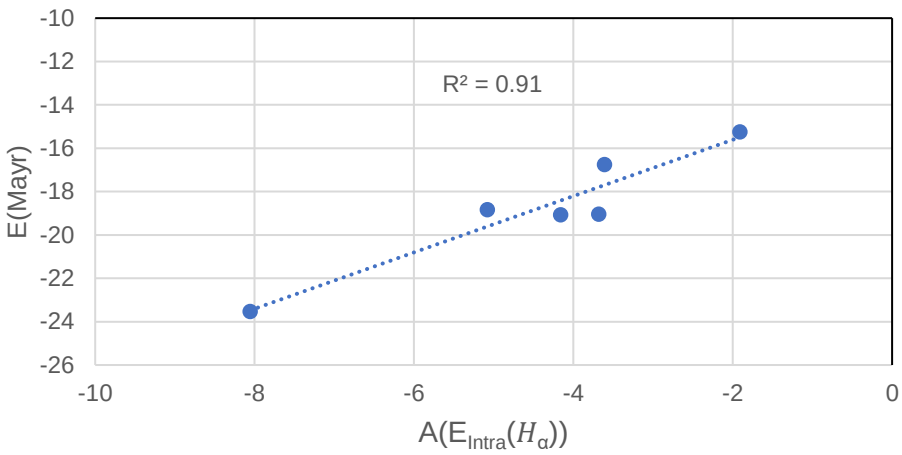


Figure 24. Correlation between E_{Mayr} and the A_i coefficient of $E_{intra}(H_\alpha)$.

On the other hand, $E_{intra}(H_\alpha)$ cannot predict the activation barriers, with R^2=0.66 and R^2=0.63 for the total energy and the free energy activation barriers, respectively.

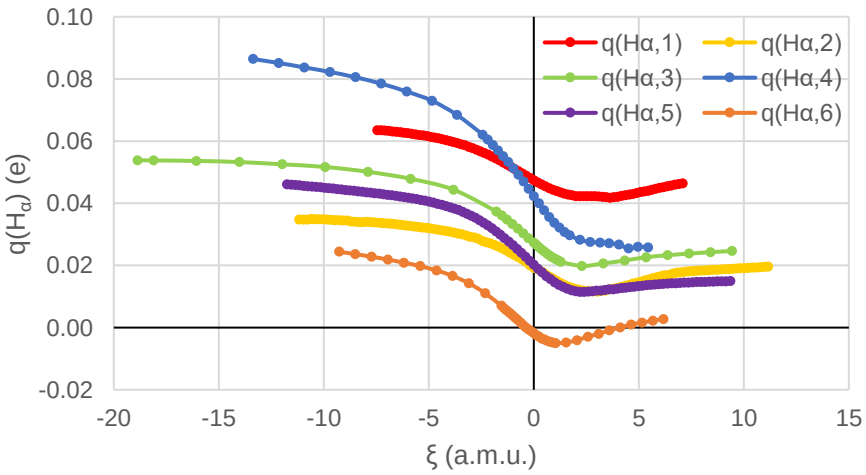


Figure 25. Variation of the QTAIM charge of H_α along the IRC.

The charge of H_α decreases during the first part of the reaction (see Figure 25), marking a sharp decrease around the TS and then either increases or decreases slightly. This behavior is quite reminiscent of the

total energy profile, even if the minimum is slightly off compared to the TS. The positions of the charge curves do not follow the electrophilicity order.

This term has still the "advantage" of being more applicable than $E_{inter}(H_{\beta 1}, H_{\beta 2})$: instead of limiting its predictive power to doubly-unsubstituted C_β systems, it can be applied to all systems bearing only one substituent (the electron-withdrawing group facilitating the reaction) on the C_α. Its interpretation, however, remains less clear: as an unreactive atom positioned away from the flow of electrons, its role is difficult to determine.

4. Conclusion

In this paper, we have investigated aza-Michael additions involving methylamine and 6 representative electrophiles in explicit solvent. The addition step of the reaction has been scrutinized in detail using the REG-IQA analysis.

The various events of the addition step, namely the preparation of reactants, the subsequent formation, weakening and strengthening of bonds, and the electronic reorganization of the systems, have been retrieved through the REG-IQA analysis. This method also shows the relevance of *a priori* insignificant terms such as the interactions of substituents beyond the EWG of the acceptor.

The influence of the solvent is described through the variation of both intra-atomic solute IQA terms and global solute-solvent interactions as well as hydrogen bonds between the solute and the solvent, stabilizing the systems and lowering the activation barriers. The presence of the solvent even enables the reaction of one of the electrophiles, which is otherwise impossible.

A first link between REG coefficients, activation barriers and experimental data from Mayr's electrophilicity scale is reported, showing how hydrogen atoms bonded to reactive atoms can act as probes of the reaction.

Thus, we have been able to show the applicability of REG and the advantages of its use in reactivity, especially for such big systems, where selecting relevant interactions while resorting only to chemical

intuition is challenging. However, the applications of REG are not limited to IQA, and one may generalize it to any energy decomposition scheme, or even descriptors based on other theories like NBO interactions.

Acknowledgements

This work has been partially supported by University of Rouen Normandy, INSA Rouen Normandy, the "Centre National de la Recherche Scientifique" (CNRS), the European Regional Development Fund (ERDF), Labex SynOrg (ANR-11-LABX-0029), Carnot Institut I2C, the graduate school for research XL-Chem (ANR-18-EURE-0020 XL CHEM), and the "Région Normandie". The authors would like also to gratefully acknowledge the "Centre Régional Informatique et d'Applications Numériques de Normandie" (CRIANN) for computing resources.

Supplementary information

Additional figures showing the REG study of the model system, all reaction pathways, the variation of some structural and charge parameters along the IRC, most relevant IQA descriptors according to REG analysis, and Cartesian coordinates for all optimized geometries.

References

1. A. Michael. *J. für Prakt. Chemie* **35**, 349–356 (1887).
2. E. D. Bergmann, D. Ginsburg & R. Pappo. in *Organic Reactions* vol. 10 179–556 (John Wiley & Sons, Inc., 1959).
3. C. F. Nising & S. Brase. *Chem. Soc. Rev.* **41**, 988–999 (2012).
4. D. Enders, A. Saint-Dizier, M. Lannou & A. Lenzen. *European J. Org. Chem.* **2006**, 29–49 (2006).
5. A. Y. Rulev. *Russ. Chem. Rev.* **80**, 197–218 (2011).
6. M. Kawatsura, S. Aburatani & J. Uenishi. *Tetrahedron* **63**, 4172–4177 (2007).
7. A. Kamimura, Y. Nagata, A. Kadowaki, K. Uchida & H. Uno. *Tetrahedron* **63**, 11856–11861 (2007).
8. P. Lakshmipathi & A. V. Rama Rao. *Tetrahedron Lett.* **38**, 2551–2552 (1997).
9. K. Nakama, S. Seki & S. Kanemasa. *Tetrahedron Lett.* **42**, 6719–6722 (2001).
10. P. R. Krishna, A. Sreeshailam & R. Srinivas. *Tetrahedron* **65**, 9657–6972 (2009).
11. M. Bláha, O. Trhlíková, J. Podešva, S. Abbrent, M. Steinhart, J. Dybal & M. Dušková-Smrčková. *Tetrahedron* **74**, 58–67 (2018).
12. H. Mayr & M. Patz. *Angew. Chemie Int. Ed. English* **33**, 938–957 (1994).
13. Mayr's Database Of Reactivity Parameters: https://www.cup.lmu.de/oc/mayr/reaktionsdatenbank/fe/.
14. G. Hoffmann, V. Tognetti & L. Joubert. *J. Mol. Model.* **24**, 281 (2018).
15. G. Hoffmann, V. Tognetti & L. Joubert. *Chem. Phys. Lett.* **724**, 24–28 (2019).
16. G. Hoffmann, M. Balcilar, V. Tognetti, P. Héroux, B. Gaüzère, S. Adam & L. Joubert. *J. Comput. Chem.* **41**, 2124–2136 (2020).
17. P. Hohenberg & W. Kohn. *Phys. Rev.* **136**, B864–B871 (1964).
18. W. Kohn & L. J. Sham. *Phys. Rev.* **140**, A1133–A1138 (1965).
19. R. G. Parr, R. A. Donnelly, M. Levy & W. E. Palke. *J. Chem. Phys.* **68**, 3801–3807 (1978).
20. R. G. Parr & R. G. Pearson. *J. Am. Chem. Soc.* **105**, 7512–7516 (1983).
21. H. Chermette. *J. Comput. Chem.* **20**, 129–154 (1999).
22. P. Geerlings, F. De Proft & W. Langenaeker. *Chem. Rev.* **103**, 1793–1874 (2003).
23. R. F. W. Bader. *Atoms in molecules : a quantum theory* (Oxford University Press, 1990).
24. P. L. A. Popelier. *Atoms in Molecules: An Introduction* (Pearson Education, 2000).
25. M. A. Blanco, A. M. Pendás & E. Francisco. *J. Chem. Theory Comput.* **1**, 1096–1109 (2005).
26. J. M. Guevara-Vela, E. Francisco, T. Rocha-Rinza & Á. M. Pendás. *Molecules* **25**, 4028 (2020).
27. J. C. R. Thacker & P. L. A. Popelier. *Theor. Chem. Acc.* **136**, 86 (2017).
28. K. Fukui. *Acc. Chem. Res.* **14**, 363–368 (1981).

29. B. Herrera & A. Toro-Labbé. *J. Phys. Chem. A* **111**, 5921–5926 (2007).

30. V. Labet, C. Morell, A. Grand & A. Toro-Labbé. *J. Phys. Chem. A* **112**, 11487–11494 (2008).

31. S. Giri, E. Echegaray, P. W. Ayers, A. S. Nuñez, F. Lund & A. Toro-Labbé. *J. Phys. Chem. A* **116**, 10015–10026 (2012).

32. P. Politzer, A. Toro-Labbé, S. Gutiérrez-Oliva, B. Herrera, P. Jaque, M. C. Concha & J. S. Murray. *J. Chem. Sci.* **117**, 467–472 (2005).

33. M. J. Vainio & M. S. Johnson. *J. Chem. Inf. Model.* **47**, 2462–2474 (2007).

34. J. S. Puranen, M. J. Vainio & M. S. Johnson. *J. Comput. Chem.* **31**, 1722–1732 (2009).

35. ***Gaussian 16, Revision C.01***. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, (2016).

36. D. P. Geerke, C. Oostenbrink, N. F. Vegt & W. F. Gunsteren. *J. Phys. Chem. B 2004, 108, 4, 1436–1445* (https://atb.uq.edu.au/system_download.py?boxid=67) (2004).

37. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess & E. Lindah. *SoftwareX* **1–2**, 19–25 (2015).

38. A. D. Becke. *J. Chem. Phys.* **98**, 5648–5652 (1993).

39. C. Lee, W. Yang & R. G. Parr. *Phys. Rev. B* **37**, 785–789 (1988).

40. R. Ditchfield, W. J. Hehre & J. A. Pople. *J. Chem. Phys.* **54**, 720–723 (1971).

41. H. P. Hratchian & H. B. Schlegel. *J. Chem. Phys.* **120**, 9918–9924 (2004).

42. H. P. Hratchian & H. B. Schlegel. *J. Chem. Theory Comput.* **1**, 61–69 (2005).

43. M. Page & J. W. McIver. *J. Chem. Phys.* **88**, 922–935 (1988).

44. M. Page, C. Doubleday & J. W. McIver. *J. Chem. Phys.* **93**, 5634–5642 (1990).

45. P. Chen, B. A. Dougan, X. Zhang, Y. D. Wu & Z. L. Xue. *Polyhedron* **58**, 30–38 (2013).

46. L. A. Jouanno, V. Di Mascio, V. Tognetti, L. Joubert, C. Sabot & P. Y. Renard. *J. Org. Chem.* **79**, 1303–1319 (2014).

47. ***AIMAll (Version 19.10.12)***, Todd A. Keith, TK Gristmill Software, Overland Park KS, aim.tkgristmill.com (2019).

48. P. Maxwell, Á. M. Pendás & P. L. A. Popelier. *Phys. Chem. Chem. Phys.* **18**, 20986–21000 (2016).

49. V. Tognetti & L. Joubert. *ChemPhysChem* **18**, 2675–2687 (2017).

50. G. B. Desmet, D. R. D'Hooge, P. S. Omurtag, P. Espeel, G. B. Marin, F. E. Du Prez & M. F. Reyniers. *J. Org. Chem.* **81**, 12291–12302 (2016).

51. P. L. A. Popelier, P. I. Maxwell, J. C. R. Thacker & I. Alkorta. *Theor. Chem. Acc.* **138**, 12 (2019).

5.3 Conclusion

Dans cet article, nous avons étudié la réaction d'aza-Michael de la méthylamine avec six accepteurs de Michael en solvant explicite en nous concentrant sur l'étape d'addition. Nous avons pu montrer (**Table 1**) que les barrières d'activation sont plus faibles en solvant explicite qu'en phase gaz (sauf pour l'un des accepteurs, avec toutefois un faible écart), conformément aux précédentes études. Cependant, la réaction d'un des accepteurs est impossible en phase gaz ce qui peut s'expliquer par un manque de liaison hydrogène intramoléculaire stabilisante (**Figure 19**) et également par l'absence de stabilisation supplémentaire du solvant (**Figures 14** à **17**).

Nous avons également pu retrouver les évènements habituels se produisant au cours de la réaction via l'analyse REG-IQA (**Figures 8** et **10-12**), notamment la préparation des réactifs sur le premier segment, la formation de la liaison $C_\beta - N$, l'affaiblissement de la liaison $C_\beta - C_\alpha$, le renforcement de la liaison $C_\alpha - C_X$ et la formation de la liaison $H_{N2} - X$ sur les segments autour de l'état de transition et enfin la réorganisation électronique dans le dernier segment. D'autres évènements ont également été mis en évidence, notamment la formation de liaisons hydrogène avec le solvant mais aussi des transferts de charge dans la méthylamine ou impliquant le substituant au-delà du groupe électroattracteur. Cela confirme, d'une part, la nécessité d'utiliser un modèle de solvant explicite plutôt qu'implicite (qui ne donnerait qu'un terme unique et indécomposable) et d'autre part l'utilité de REG pour identifier des interactions pertinentes qui auraient été mises de côté en ne se reposant que sur l'intuition. À chaque fois, les interactions IQA significatives peuvent être rationalisées en étudiant les paramètres structuraux et les charges QTAIM.

L'influence du solvant sur la réaction ne se reflète pas au travers des charges des solvants, dont les variations restent pratiquement nulles ou très faibles. Les sommes des interactions IQA des solutés avec chaque molécule de solvant restent dans la majorité des cas légèrement stabilisantes ou déstabilisantes (**Figure 17**), à l'exception du système **4** où une des molécules se détache nettement. En les sommant toutes, en revanche, un effet clairement stabilisant apparaît (**Figure 16**). La stabilisation apportée par le solvant est également visible grâce aux liaisons hydrogène entre les solutés et le solvant. Nous avons également étudié les termes globaux des solvants accessibles via une décomposition simple de l'énergie

(**Figure 15**). Il s'avère que les termes stabilisants sont les interactions du solvant avec le soluté et les termes intra-atomiques du solvant, tandis que les interactions entre molécules de solvant et à l'intérieur des différentes molécules de solvant sont déstabilisantes. On peut toutefois contester la validité de ce modèle, puisque les molécules de solvant ne sont pas elles-mêmes totalement solvatées (il n'y a pas de seconde couche de solvatation qui permettrait de solvater les molécules de solvant de la première couche), menant à une description fausse des interactions entre molécules de solvant. Cela reste un argument supplémentaire pour éviter de calculer toutes ces interactions en détail.

Enfin, on peut se demander si des descripteurs issus de ce modèle de solvant explicite peuvent être corrélés à des grandeurs d'intérêt comme la barrière d'activation ou l'électrophilie de Mayr ; après tout, les termes énergétiques IQA contribuent à la barrière et l'analyse REG peut mettre en évidence les termes les plus corrélés et dont la variation est la plus importante. Les barrières elles-mêmes ne sont pas corrélées à l'électrophilie de Mayr (avec des R^2 inférieurs à 0,80).

Après avoir établi quels étaient les descripteurs reproduisant le mieux les profils énergétiques sur la base de leur coefficient de détermination et de leur MAE, les corrélations des coefficients REG de plusieurs descripteurs avec les grandeurs d'intérêt ont été testées. Il s'avère que l'interaction entre les deux atomes d'hydrogène portés par le carbone attaqué corrèlent bien avec les barrières d'activation (R^2 = 0,86 – 0,90, **Figure 21**), ce qui peut se comprendre par le fait que ces deux atomes se rapprochent lors de l'attaque nucléophile et du changement d'hybridation du carbone. Le terme intra-atomique de l'hydrogène du carbone voisin, quant à lui, présente une corrélation similaire avec l'électrophilie de Mayr (R^2 = 0,91, **Figure 24**), avec une évolution de charges proche de celle du profil énergétique total. Ces atomes d'hydrogène peuvent donc être vus comme des « sondes » chimiques du système en n'étant pas directement affectés par les transformations chimiques mais en étant tout de même liés aux atomes réactifs. Encore une fois, ce sont des termes qui n'auraient pas pu être identifiés sans l'analyse REG. Les corrélations obtenues sont également meilleures que celles obtenues lors d'études précédentes avec des régressions linéaires (voire même meilleures que certaines des régressions multilinéaires).

Cependant, notre méthode n'est pas sans inconvénients :

- Coût computationnel élevé, malgré une réduction significative du temps de calcul par division des calculs (limitant ainsi le nombre de systèmes étudiables en un temps raisonnable) :
    - La recherche de structures stables, d'états de transition et de chemins de réaction avec autant d'atomes constituent un goulet d'étranglement. Il est bien entendu possible d'étudier plusieurs structures en parallèle, mais cela ne réduit pas le temps de calcul nécessaire pour chacune. On pourrait aussi réduire le niveau de théorie, mais au risque de mal décrire les interactions intermoléculaires et les effets de polarisation.
    - Lors des calculs IQA, la méthode nécessite d'avoir plusieurs dizaines de nœuds de calcul à disposition pendant plusieurs jours ou semaines pour obtenir des résultats pour un seul système. Il est possible de réduire le nombre de points utilisés, notamment en utilisant des algorithmes permettant de trouver les points optimaux pour décrire un profil avec un nombre de points réduit.
- Seule une configuration de solvant donnée peut être étudiée, à moins de refaire toute l'analyse pour un système donné.
- Absence de contrôle de la configuration du solvant entre les différents composés, ce qui peut expliquer certains des résultats intrigants que nous avons pu avoir (comme la barrière plus haute en solvant qu'en gaz pour le composé **6**). On peut toutefois se rassurer en voyant que les interactions majeures avec le solvant restent proches entre les différents composés.
- Mauvaise description des interactions entre molécules de solvant : comme mentionné précédemment, les molécules de solvant ne sont pas bien solvatées, ce qui aboutit à des termes d'interaction déstabilisants entre celles-ci.
- Difficulté de définition des réactifs séparés solvatés : il faudrait pour cela « diviser » les molécules de solvant entre les deux réactifs, ce qui aboutirait sans doute à une mauvaise solvatation de chacun d'eux. On pourrait également augmenter le nombre de molécules de solvant utilisées pour cela (ou utiliser le même nombre de molécules de solvant que pour les deux réactifs), mais au risque de perdre la référence des énergies.

Pour aller plus loin et résoudre les quatre derniers problèmes, on pourrait envisager des calculs de dynamique moléculaire pour avoir une représentation plus complète des différentes structures et chemins possibles. Cependant, une étude IQA complète de telles simulations serait beaucoup trop longue à moins de restreindre les atomes considérés et d'avoir une puissance de calcul disponible nettement supérieure et des algorithmes de calcul considérablement plus efficaces. L'analyse REG ne constituerait en revanche pas un goulet d'étranglement puisqu'elle ne prend que quelques minutes même pour nos systèmes.

On peut également envisager d'utiliser l'analyse REG avec d'autres types de descripteurs, par exemple les descripteurs NBO comme mentionné dans le chapitre 3, ou d'autres descripteurs issus de décompositions énergétiques.

Conclusion générale

En conclusion, dans les précédents chapitres, nous avons montré les développements de la méthode REG effectués au cours de cette thèse. Grâce à ces nouveautés, la méthodologie REG, auparavant restreinte à l'analyse IQA d'une réaction unique pour un système de taille réduite, a pu être étendue pour pouvoir effectuer des analyses REG-NBO et comparer directement deux réactions par soustraction de profils réactionnels. De nombreuses autres modifications mineures ont permis de faciliter la lecture des résultats graphiques produits par l'analyse.

Nous avons également pu proposer une méthodologie de calculs IQA pour l'étude de systèmes solvatés de grande taille. Ces développements ont permis de nouvelles applications de REG en réactivité, notamment au travers de l'étude de réactions compétitives (torquosélectivité) et des effets de solvant (réaction d'aza-Michael). Dans chacun de ces cas, REG a pu confirmer les résultats de la littérature ou les compléter en mettant en évidence de nouveaux effets (parfois même prépondérants) que l'intuition chimique n'aurait pas considérés.

Plus précisément, dans le chapitre 3, nous avons détaillé le développement de nouvelles routines pour comparer deux réactions du même système par soustraction des profils (analyse « REGdiff ») ainsi que pour réaliser l'analyse REG-NBO des réactions en complément de l'analyse REG-IQA. La soustraction des profils n'est pas triviale et nécessite plusieurs étapes de traitement numérique, notamment pour normaliser les profils et les rééchantillonner de sorte à retrouver les mêmes points. L'analyse REG-NBO a également nécessité de résoudre quelques problèmes en raison de la variabilité du nombre de termes sur chaque segment, problème qui ne se pose pas pour l'analyse REG-IQA.

Ces développements ont permis d'étudier la torquosélectivité de la réaction d'ouverture thermique de divers cyclobutènes substitués et de confirmer les explications précédentes ou d'apporter des renseignements supplémentaires. On retrouve notamment pour certains systèmes l'explication habituelle (un seul terme E(2) ou E(4) en NBO pour les systèmes favorisant respectivement le chemin « in » ou « out »), alors que pour d'autres, le terme habituellement mis en avant est supplanté par une

autre interaction. Enfin, pour certains systèmes, la situation est plus complexe et considérer seulement un ou deux termes ne suffit pas à expliquer la torquosélectivité. Des parallèles entre les résultats NBO et IQA ont pu être établis dans certains cas, malgré les différences inhérentes à ces deux théories.

Dans le chapitre 4, nous avons étudié la réaction d'aza-Michael au travers de deux articles, le premier utilisant un modèle de solvant implicite et le second en solvant explicite. Le premier article se concentre sur les aspects mécanistiques et cinétiques, où nous avons pu déterminer que lors de l'étape de transfert de proton, le chemin privilégié par tous les systèmes est le transfert direct assisté par une seconde molécule d'amine. L'étude des descripteurs de DFT conceptuelle a permis de montrer que l'évolution du potentiel chimique est corrélée avec celle du profil énergétique au début de l'addition. L'utilisation des constantes cinétiques a permis une légère amélioration de notre modèle de prédiction de l'électrophilie de Mayr.

Nous nous sommes toutefois demandés comment le solvant agissait sur la réaction et si le modèle ne pouvait pas être amélioré en considérant un modèle de solvant explicite plutôt qu'implicite, ce que nous avons fait dans le second article en nous restreignant à l'étape d'addition.

Dans un premier temps, nous avons détaillé la méthodologie utilisée pour réaliser des calculs IQA sur de grands systèmes (environ 200 atomes). Celle-ci consiste à ne considérer qu'une partie des interactions du système (ici, seulement les interactions soluté-soluté et soluté-solvant) et donc à diviser les calculs pour ne calculer que les interactions désirées. Cela a permis de paralléliser les calculs et de gagner en temps de calcul.

Nous avons ensuite montré que l'analyse REG-IQA permettait non seulement de retrouver les évènements attendus de la réaction, tels que la formation de la liaison N-C, l'affaiblissement de la double liaison C=C et le renforcement de la liaison C-C adjacente, mais également que d'autres interactions étaient importantes voire même prépondérantes, notamment les interactions concernant le substituant électroattracteur et les atomes au-delà.

L'analyse a aussi mis en évidence quels sont les termes de solvatation qui stabilisent ou déstabilisent le système, et la présence d'interactions entre le soluté et le solvant (pour la plupart stabilisantes, comme

des liaisons hydrogène qui aident les réactifs à s'approcher). L'influence du solvant peut également être indirecte, comme le prouvent les variations des termes intra-atomiques plus stabilisants en phase solvant que gaz, en accord avec les barrières d'activation plus basses en solvant. Les effets de solvant ont ainsi pu être quantifiés et ont permis d'expliquer pourquoi la réaction d'un des composés considérés n'est pas possible en phase gaz.

Enfin, l'analyse REG a permis de trouver des descripteurs dont le coefficient REG (et par extension leur variation) corrèle bien avec les barrières d'activation et/ou l'électrophilie de Mayr, ce qui est le premier lien entre l'analyse REG et une donnée expérimentale. Les atomes d'hydrogène impliqués dans les termes IQA en question peuvent donc être vus comme des « sondes » de la réaction en ne réagissant pas directement mais en étant sujets aux modifications du système provoqués par la réaction.

En perspectives de ces travaux, nous pourrions prolonger les études réalisées dans cette thèse : pour la torquosélectivité, cela pourrait notamment être réalisé en utilisant des régressions bi-, tri- ou plus généralement multilinéaires, mais également par des méthodes non linéaires comme le *machine learning* pour trouver des groupes de termes expliquant la torquosélectivité pour les systèmes complexes.

De même, pour la réaction d'aza-Michael en solvant explicite, il faudrait étendre la méthodologie à d'autres systèmes et considérer d'autres configurations de solvant, soit en changeant simplement leur position, soit en modifiant le nombre de molécules de solvant, afin de confirmer les résultats obtenus. Cela pourrait passer par des calculs de dynamique moléculaire qui permettent d'étudier rapidement et à faible coût computationnel un grand nombre de configurations.

Plus globalement, le principe de REG étant simplement de réaliser des régressions linéaires entre des descripteurs uniques et l'énergie totale le long d'une coordonnée quelconque, la méthode peut être généralisée aux autres modèles de décomposition énergétiques comme les partitions EDA, Hirshfeld, SAPT ou « fuzzy atoms »… voire à d'autres approches, comme les charges (en comparant les charges atomiques à la charge totale), les descripteurs issus de la CDFT, comme évoqué dans le premier article, etc. Bien évidemment, cela signifie qu'il faudrait développer de nouvelles routines pour pouvoir lire les

données issues de ces analyses, mais le cœur de la méthode resterait le même, et cela permettrait de rendre la méthode plus versatile et de toucher un plus large public au sein de la communauté des chimistes théoriciens, en l'appliquant notamment pour l'étude de la liaison chimique, des changements d'état, en réactivité... Il faut toutefois prendre garde aux problèmes inhérents à la méthode tels que la segmentation, dont le choix demeure arbitraire, ou la répartition des points qui peut altérer les résultats obtenus.

La suite du développement de la librairie de sorte à la rendre plus simple d'utilisation, même par des utilisateurs ne connaissant pas Python, est également en cours, notamment pour doter la librairie d'une interface graphique permettant de régler facilement les différentes options et de visualiser directement les résultats. En effet, bien que la librairie soit d'ores et déjà disponible en ligne, il faut une certaine expérience en programmation avec Python pour l'utiliser de façon satisfaisante.

Enfin, le calcul IQA reste un goulet d'étranglement en termes de temps de calcul, malgré la méthodologie que nous avons employée. Une thèse est prévue dans le but d'accélérer les temps de calcul IQA par *machine learning* ainsi que prédire les profils réactionnels à moindre coût.

Résumé : Ces travaux ont pour objet le développement et l'application de la méthode REG en réactivité. Cette nouvelle méthode d'interprétation permet de sélectionner automatiquement et sans biais les descripteurs les plus pertinents (issus d'une décomposition énergétique) permettant d'expliquer la variation de l'énergie totale d'une molécule le long d'un profil réactionnel. Deux cas sont présentés : l'étude de l'ouverture thermique sélective de cyclobutènes substitués afin de comparer deux chemins d'ouverture en compétition ainsi que l'analyse de réactions d'aza-Michael pour comprendre les effets de solvant. Dans le premier cas, les deux chemins ont été comparés par soustraction des profils et étudiés simultanément par analyse REG-NBO et REG-IQA, ce qui a permis de confirmer ou de compléter l'explication orbitalaire habituelle. Dans le second cas, après une étude mécanistique et cinétique de la réaction en solvant implicite, l'étape d'addition a été examinée en détail par analyse REG-IQA avec un modèle de solvant explicite. Nous avons montré que les interactions prépondérantes impliquaient le substituant a priori non réactif et que le solvant stabilisait le système de façon directe et indirecte.

Mots-clés : chimie théorique, DFT, réactivité, torquosélectivité, aza-Michael, IQA, NBO, REG

Abstract: This work aims at developing and applying the REG method in reactivity. This new interpretative method allows for an automatic and unbiased selection of the most relevant descriptors (originating from energy decompositions) linked to the variation of the total energy of a molecule along a reaction profile. Two cases are presented: the study of the selective thermal ring-opening of substituted cyclobutenes in order to compare two competing pathways, and the analysis of the aza-Michael reaction to understand solvent effects. In the first case, the two pathways have been compared through a subtraction of the profiles and simultaneously studied by REG-IQA and REG-NBO analysis, which has enabled us to confirm or complete the usual orbital explanation. In the second case, after a mechanistic and kinetic study of the reaction in implicit solvent, the addition step has been scrutinized in detail by REG-IQA analysis using an explicit solvent model. We have thus shown that the dominating interactions involve the (a priori unreactive) substituent and that the solvent stabilizes the system in direct and indirect ways.

Keywords: theoretical chemistry, DFT, reactivity, torquoselectivity, aza-Michael, IQA, NBO, REG