

Développement de modèles QSPR pour la prédiction des propriétés d'explosibilité des composés nitroaromatiques

Guillaume Fayet

▶ **To cite this version:**

Guillaume Fayet. Développement de modèles QSPR pour la prédiction des propriétés d'explosibilité des composés nitroaromatiques. Chimie. Chimie ParisTech, 2010. Français. NNT : . pastel-00006157

HAL Id: pastel-00006157
https://pastel.hal.science/pastel-00006157

Submitted on 9 Sep 2010

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.

THESE DE DOCTORAT DE
L'UNIVERSITE PIERRE ET MARIE CURIE

Présentée le 30 mars 2010 par

Guillaume FAYET

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE
Ecole Doctorale Chimie Physique et Analytique de Paris VI

Spécialité Chimie Informatique et Théorique

DEVELOPPEMENT DE MODELES QSPR POUR LA PREDICTION DES PROPRIETES D'EXPLOSIBILITE DES COMPOSES NITROAROMATIQUES

devant le jury composé de :

Pr. Henry Chermette	Rapporteur
Dr. Didier Mathieu	Rapporteur
Dr. Patricia Rotureau	Examinatrice
Pr. Alain Fuchs	Examinateur
Pr. M. Esmaïl Alikhani	Examinateur
Dr. Alberto Del Rio	Examinateur
Pr. Laurent Joubert	Examinateur
Pr. Carlo Adamo	Directeur de thèse

SOMMAIRE

Remerciements **9**

Glossaire **11**

Liste des figures **13**

Liste des tableaux **17**

Introduction générale **21**

CHAPITRE 1 : Contexte et objectifs **23**

1- Le règlement REACH **24**

1.1- Perception sociétale de la chimie 24

1.2- Vers un règlement européen unifié 25

1.3- Objectifs et champs d'application 25

1.4- Les procédures de REACH 27

1.5- Le règlement CLP 28

1.6- Les méthodes alternatives dans REACH 30

2- Explosibilité **32**

2.1- Définitions 32

2.2- Evaluation expérimentale des dangers d'explosion 36

2.3- Modélisation et explosibilité 40

3- Projet REPLACE pour le développement de méthodes alternatives **41**

CHAPITRE 2 : Bases théoriques **45**

1- Chimie quantique **46**

1.1- Bases de la chimie quantique 46

1.1.1- Equation de Schrödinger 46

1.1.2- Approximation de Born-Oppenheimer 46

1.1.3- Approximation orbitalaire et fonctions de base 47

1.1.4- Principe variationnel 49

1.2- Méthode Hartree-Fock 49

1.2.1- Equations Hartree-Fock 50

1.2.2- Equations de Roothaan-Hall 51

1.2.3- Cas des systèmes à couches ouvertes 51

1.2.4- Limites de la méthode Hartree-Fock 52

1.3- Méthodes semi-empiriques 52

1.4- Théorie de la Fonctionnelle de la Densité 54

1.4.1. Théorèmes de Hohenberg-Kohn 54

1.4.2. Approche Kohn-Sham 55

1.4.3. Fonctionnelles d'échange-corrélation 56

2- Quantitative Structures-Property Relationships 57
2.1- Principe 58
2.2- Importance des bases de données 59
2.3- Descripteurs moléculaires 61
2.3.1- Descripteurs constitutionnels 61
2.3.2- Descripteurs topologiques 61
2.3.3- Descripteurs géométriques 63
2.3.4- Descripteurs quantiques 64
2.4- Méthodes d'analyse de données 68
2.4.1- Régression multi-linéaire 69
2.4.2- Analyse par composantes principales 71
2.4.3- Régression aux moindres carrés partiels 73
2.4.4- Arbre de décision 74
2.4.5- Sélection des descripteurs 75
2.4.6- Validation 76
2.4.7- Domaines d'applicabilité 78
2.4.8- Interprétation des modèles 80

CHAPITRE 3 : Prédiction des propriétés explosives des composés nitroaromatiques 81
1- Méthodologie 82
1.1- Calcul des descripteurs moléculaires 82
1.2- Outils d'analyse de données 82
2- Stabilité thermique 83
2.1- Généralités 83
2.2- Travaux antérieurs 85
2.3- Chaleur de décomposition 86
2.3.1- Intérêt des descripteurs quantiques 86
2.3.2- Utilisation d'un jeu de descripteurs étendu 95
2.3.3- Modèle QSPR pour une base de données étendue 100
2.3.4- Utilisation d'autres outils de traitement de données 109
2.4- Température de décomposition 123
2.5- Bilan 126
3- Sensibilité à la décharge électrique 128
3.1- Généralités 128
3.2- Travaux antérieurs 129
3.3- Analyse QSPR 131
3.3.1- Développement du modèle (jeu d'entraînement) 131
3.3.2- Validation externe et domaine d'applicabilité 134
3.3.3- Analyse par composantes principales 135
3.4- Bilan 136
4- Sensibilité à l'impact 138
4.1- Généralités 138
4.2- Travaux antérieurs 139
4.2.1- Balance en oxygène 139
4.2.2- Autres descripteurs moléculaires 140
4.2.3- Régressions multi-linaires (MLR) 141
4.2.4- Autres approches multi-variables 142
4.3- Analyse QSPR 142
4.3.1. Base de données complète 144
4.3.2. Molécules sans liaison C-H alpha 144

4.3.3. Molécules avec liaison C-H alpha 146
4.3.4. Analyse par composantes principales 147
4.4- Bilan 149
5- Conclusion 149

CHAPITRE 4 : Mécanismes de décomposition des composés nitroaromatiques 153
1- Généralités 154
2- Méthodologie 156
3- Décomposition des nitrobenzènes substitués 156
3.1- Décomposition du nitrobenzène 156
3.2- Influence des substituants en position para 158
3.3- Influence des substituants en position méta 162
3.4- Influence des substituants en position ortho 164
4- Décomposition des *o*-nitrotoluènes substitués 166
4.1- Décomposition de l'*o*-nitrotoluène 166
4.2- Influence des substituants sur le carbone 3 169
4.3- Influence des substituants sur le carbone 4 171
4.4- Influence des substituants sur le carbone 5 173
4.5- Influence des substituants sur le carbone 6 176
5- Conclusion 178

Conclusion générale 181

Bibliographie 185

Annexes 193
ANNEXE I : Procédure de classification des substances, mélanges et objets explosibles dans le Manuel d'épreuves et de critères relatif à la réglementation du Transport des Marchandises Dangereuses (5ème édition révisée, 2009) 195
ANNEXE II : Liste des publications issues de ces travaux de thèse 199

Résumé 201

Abstract 203

REMERCIEMENTS

La thèse est une aventure dans laquelle l'entourage prend une part importante. Remercier comme il se doit toutes les personnes qui m'ont accueilli, aidé et soutenu au cours de ces trois années n'est pas chose aisée. Je vais faire de mon mieux pour n'oublier personne.

En premier lieu, je tiens à remercier Messieurs Henry Chermette et Didier Mathieu d'avoir accepté d'être les rapporteurs de ce manuscrit ainsi qu'aux Professeurs Fuchs et Alikhani d'avoir accepté de faire partie de mon jury de thèse.

Je tiens bien évidemment à remercier l'INERIS pour le financement et l'accueil qui m'ont été fournis pour cette thèse. Ces remerciements s'adressent en particulier à Patricia Rotureau pour avoir proposé cette thématique et accompagné ma thèse au sein des Directions de la Certification puis des Risques Accidentels. Merci également à toutes les personnes avec qui j'ai pu travailler là-bas, en particulier Guy Marlair pour ces remarques et conseils toujours pertinents.

Un grand merci également à toute l'équipe de Modélisation des Systèmes Complexes de l'ENSCP et en particulier au Pr. Carlo Adamo, mon directeur de thèse. Ses avis et recommandations toujours avisés ont fortement contribué au succès de ces travaux. Merci bien entendu à Laurent Joubert, nouvellement nommé Professeur à Rouen, qui a pris une part importante à l'encadrement de ma thèse. Egalement un grand merci à Ilaria pour tous ses conseils et surtout m'avoir supporté pendant ces trois ans…

Je ne suis bien entendu pas près d'oublier la joyeuse troupe des thésards et post-docs qui se sont succédé à mes côtés. Il y a tout d'abord eu Fred dit "Senior", le rigoureux mais fourbissime Vincent, compagnon d'infortune pour un petit Lagavulin de fin de journée, le jeune padawan (Cyril), mais aussi Super Tangui, le gourmant Roberto, Antonio le Sicilien. Un petit ciao particulier pour LA fille, Stefania, qui partagea mes trajets Paris-Verneuil pendant un an et demi.

Un grand merci bien entendu à mes "petits" stagiaires Sylvain et Nicolas qui ont contribué aussi à ces travaux. Enfin, je n'oublierai pas les p'tits nouveaux (Adrian, Eric, Thomas) et les voyageurs égarés (Ciro, Denis).

J'ai une petite pensée pour M. Berthier, qui nous a quittés en fin d'année dernière. Je pense que je n'oublierai pas les discussions que nous avons pu avoir, avec lui ainsi que M. Savinelli, qu'elles soient scientifiques, linguistiques, historiques voire touristiques et même rugbistiques…

Je tiens à remercier également Alberto Del Rio d'avoir permis ma venue au Département des Sciences Pharmaceutiques de l'Università di Modena e Reggio Emilia. Sa disponibilité et sa motivation ont permis à cette courte période d'être très productive et enrichissante pour moi. Merci d'ailleurs à tous ces collaborateurs très sympathiques et en particulier à Giuseppe qui a depuis rejoints l'équipe MSC au début de l'année.

Enfin, *Last but not Least*, je tiens à remercier mes proches, mes amis et ma famille, tout particulièrement mes parents qui m'ont permis d'en arriver là !

Merci à tous !!

Grazie a tutti !!

GLOSSAIRE

REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) est le règlement européen concernant l'enregistrement, l'évaluation et l'autorisation des substances chimique, entré en vigueur le 1er juin 2007.

L'ECHA (European chemical Agency) est l'agence européenne créée pour coordonner la mise en place du nouveau règlement REACH au sein de l'Union Européenne.

L'ECVAM (European Centre for the Validation of Alternative Methods) est le centre européen créé pour coordonner les démarches de validation des méthodes alternatives à l'expérience.

Le GHS (Globally Harmonized System for classification and labelling of chemical products), adopté par les Nations Unies en juillet 2003, vise à harmoniser sur le plan international les critères de classification et à communiquer, au moyen de l'étiquetage, les dangers liés aux produits chimiques afin d'en garantir la sécurité d'emploi. L'objectif est de faciliter les échanges internationaux de substances chimiques et, en même temps, d'améliorer la protection des personnes et de l'environnement au niveau international. Le 16 décembre 2008, le SGH a été adopté par l'Union européenne via le règlement CE n° 1272/2008 appelé **« règlement CLP » (Classification, Labelling and Packaging of chemical substances and mixtures)**.

La sensibilité d'un explosif est l'énergie nécessaire pour le faire réagir. Cette dernière peut être apportée sous la forme d'une décharge électrique, de chaleur ou encore d'un choc.

La stabilité thermique caractérise le comportement d'un matériau sous l'effet de la température.

Un groupement explosophore est un groupement dont la présence est indicateur de propriétés explosives (ex : NO_2)

La Théorie de la Fonctionnelle de la Densité (DFT) est une méthode de calcul basée sur des concepts de mécanique quantique, utilisée en chimie et physique pour accéder à des propriétés à l'échelle moléculaire.

Une méthode de criblage (*screening*) est une méthode permettant d'effectuer un tri au sein d'une grande quantité de composés afin d'extraire ceux susceptibles de présenter les plus forts potentiels en vue d'une application donnée.

Une relation quantitative structure-activité ou propriété (QSAR ou QSPR) est une méthodologie consistant à mettre en évidence une relation quantitative entre une grandeur macroscopique (activité biologique, propriété physico-chimique) et la structure moléculaire des composés étudiés.

Un descripteur est une grandeur ou propriété permettant de caractériser la structure (géométrique ou électronique) à l'échelle moléculaire d'un composé chimique.

Un espace chimique est défini pour une base de données de molécules en fonction des descripteurs moléculaires.

La régression multi-linéaire (MLR) est une méthode de régression permettant de relier linéairement une variable dépendante Y avec une série de variables indépendantes X_i.

La méthode *Best Multi-Linear Regression* (BMLR) est une technique implémentée dans le logiciel CodessaPro pour la mise en place de modèles multi-linéaires.

L'analyse par composantes principales (PCA) est une méthode statistique qui consiste à transformer des variables corrélées entre elles en nouvelles variables indépendantes les unes des autres (les composantes principales).

La PLS (Partial Least Squares) est la méthode de régression qui découle du même principe.

Un arbre de décision est un outil de classification (qualitatif) permettant de modéliser simplement, graphiquement et rapidement des phénomènes plus ou moins complexes.

Le jeu d'entraînement est un jeu de données employé pour mettre en place un modèle.

Le jeu de validation est un jeu de données utilisé pour estimer le pouvoir prédictif d'un modèle. Les molécules qui le composent sont différentes de celles du jeu d'entraînement.

La validation interne est la méthode de validation d'un modèle QSPR permettant d'estimer sa robustesse, c'est-à-dire l'influence des composés du jeu d'entraînement sur ses performances.

La validation externe est la méthode de validation d'un modèle QSPR permettant d'estimer son pouvoir prédictif à partir d'un jeu de validation.

Le domaine d'applicabilité (AD) est le domaine au sein duquel les prédictions réalisées avec le modèle peuvent être considérées comme fiables.

LISTE DES FIGURES

Figure 1 Calendrier prévu pour l'enregistrement des substances dans le cadre du règlement REACH (extrait de [20]) 28

Figure 2 Profil typique d'une décomposition explosive 33

Figure 3 Triangle du feu et hexagone de l'explosion (extrait de [41]) 34

Figure 4 Exemples de composés explosifs 35

Figure 5 Epreuve de Koenen (Appareillage INERIS) 37

Figure 6 Principaux modes de sensibilité des substances explosibles 37

Figure 7 Mortier balistique (appareillage INERIS) 38

Figure 8 Coupe du bloc de plomb avant et après le tir d'explosif au cours d'un test de Trauzl [35] 39

Figure 9 Echelles de modélisation 40

Figure 10 Outil de prédiction des propriétés d'explosibilité 42

Figure 11 Comparaison de la représentation d'une fonction de Slater 1s à partir de gaussiennes contractées STO-1G, STO-2G et STO-3G 48

Figure 12 Principe de la méthode QSPR 59

Figure 13 Influence d'une valeur atypique (A) sur une corrélation 60

Figure 14 Distances topologiques au sein de la molécule de nitrobenzène 62

Figure 15 Angle dièdre dirigeant l'activité catalytique des complexes bis(arylimino)pyridine du fer 63

Figure 16 Dérivées de l'énergie aux 1^{er} et 2^{nd} ordre 65

Figure 17 Représentation graphique de la régression multi-linéaire pour deux variables indépendantes x_1 et x_2 et une variable dépendante y 69

Figure 18 Principe de mise en place de deux composantes principales PC1 et PC2 dans un espace à 3 dimensions (x_1, x_2, x_3) lors d'une analyse PCA 71

Figure 19 Matrice des coordonnées dans le plan des deux premières composantes principales d'une analyse PCA 72

Figure 20 Matrice des poids dans le plan des deux premières composantes principales d'une analyse PCA 73

Figure 21 Principe de mise en place de deux composantes principales PC1' et PC2' dans un espace à 3 dimensions (x_1, x_2, x_3) par analyse PLS..... 74

Figure 22 Arbre de décision..... 74

Figure 23 Représentation géométrique du domaine d'application d'un modèle selon deux descripteurs par des approches basées sur des distances euclidiennes (à gauche) et par simple projection des éventails de valeurs (à droite)..... 79

Figure 24 Schéma de principe d'un calorimètre différentiel à balayage (DSC) 84

Figure 25 Profil caractéristique d'un pic de décomposition exothermique tel qu'observé en DSC..... 84

Figure 26 22 molécules de la base de données de Duh [158] 87

Figure 27 Affinité électronique (EA) et électronégativité (χ) en fonction de la chaleur de décomposition expérimentale (-ΔH) 90

Figure 28 Chaleurs de décomposition prédites ($-\Delta H_{calc}$) et expérimentales ($-\Delta H_{exp}$)..... 92

Figure 29 Chaleurs de décomposition (en kJ/mol) prédites (Eq. 3.15) et expérimentales... 95

Figure 30 Corrélation R^2 en fonction du nombre de descripteurs..... 96

Figure 31 Chaleurs de décompositions (en kJ/mol) prédites (AM1 et DFT) et expérimentales..... 98

Figure 32 77 molécules de la base de données de Ando [157]..... 101

Figure 33 Corrélation entre les données de Duh [158] (Base 22) et de Ando [157] (Base 77)..... 103

Figure 34 Distributions des valeurs de chaleur de décomposition de la base complète et des jeux d'entraînement et de validation. 103

Figure 35 Chaleurs de décomposition (en kJ/mol) expérimentales et prédites par le modèle (3.20) pour les molécules des jeux d'entraînement et de validation 106

Figure 36 Chaleurs de décomposition (en kJ/mol) prédites (Eq. 3.21) et expérimentales pour une série de composés nitroaromatiques non substitués en position ortho..... 108

Figure 37 Représentation du jeu complet des descripteurs dans le plan des deux premières composantes principales de l'analyse PLS..... 110

Figure 38 Représentation du jeu réduit à 44 descripteurs dans le plan des deux premières composantes principales de l'analyse PLS..... 111

Figure 39 Représentation du jeu réduit à 7 descripteurs et de la propriété (-ΔH) dans le plan des deux premières composantes principales de l'analyse PLS 112

Figure 40 Charges atomiques (calculées par analyse NPA) pour les molécules d'acide picrique ou de 2,4,6-trinitroaniline 113

Figure 41 Représentation des molécules dans le plan des deux premières composantes principales du modèle PLS 116

Figure 42 Arbre de décision pour la prédiction de la chaleur de décomposition des composés nitroaromatiques du jeu d'entraînement 118

Figure 43 Températures de décomposition (en °C) expérimentales et prédites par le modèle AM1 (Eq. 3.25) 124

Figure 44 Températures de décomposition (en °C) prédites par le modèle DFT (Eq. 3.26) et expérimentales 125

Figure 45 Appareillage permettant la détermination de la sensibilité à la décharge électrique [197] 129

Figure 46 Sensibilités électriques (en J) des 1,5- et 1,8-dinitronaphtalènes (DNN) calculées par Keshavarz [216] et expérimentales [211] 131

Figure 47 41 Molécules de la base de données 132

Figure 48 Sensibilités à la décharge électrique prédites (Eq. 3.30) et expérimentales [207,211] 135

Figure 49 Représentation des 50 molécules dans le plan des deux premières composantes principales d'une analyse PCA 136

Figure 50 Mouton de choc pour la mesure de la sensibilité à l'impact (Appareillage INERIS) 138

Figure 51 50 molécules de la base de données 143

Figure 52 Sensibilité à l'impact (en cm) expérimentales et prédites des composés nitroaromatiques de présentant pas de possibles liaisons C-H alpha 146

Figure 53 Evolution des R^2 et R^2_{cv} en fonction du nombre de descripteurs 147

Figure 54 Représentation des 50 molécules dans le plan des deux premières composantes principales d'une analyse PCA 148

Figure 55 Mécanismes de décomposition principaux du 2-nitrotoluène 155

Figure 56 Mécanismes de décomposition principaux du nitrobenzène (énergies ΔE en vert et énergies libres de Gibbs ΔG en rouge) 157

Figure 57 Aperçu des dérivés du nitrobenzène. 158

Figure 58 Profils énergétiques pour les réactions de décomposition vers NO et NO_2 de nitrobenzènes substitués en position para 160

Figure 59 Energies libres de Gibbs relatives de la dissociation C-NO_2 (ΔG_{diss}) et de la stabilité de la structure R-C_6H_4ONO (ΔG_{ONO}) en fonction des constantes de Hammett (σ) 161

Figure 60 Profils énergétiques pour les réactions de décomposition vers NO et NO_2 de nitrobenzènes substitués en position méta 163

Figure 61 Profils énergétiques pour les réactions de décomposition vers NO et NO_2 de nitrobenzènes substitués en position para 165

Figure 62 Mécanismes de décomposition principaux de l'o-nitrotoluène 167

Figure 63 Aperçu des différents points stationnaires intervenant dans les chemins de décomposition principaux de l'o-nitrotoluène 167

Figure 64 Aperçu des dérivés de l'o-nitrotoluène 168

Figure 65 Profils énergétiques pour les réactions de décomposition principales des o-nitrotoluènes substitués en position 3 170

Figure 66 Profils énergétiques pour les réactions de décomposition principales des o-nitrotoluènes substitués en position 4 172

Figure 67 Energies libres de Gibbs relatives de la dissociation C-NO_2 (ΔG) en fonction des constantes de Hammett (σ) 173

Figure 68 Profils énergétiques pour les réactions de décomposition principales des o- nitrotoluènes substitués en position 5 174

Figure 69 Structures de résonnance pour LM1 175

Figure 70 Profils énergétiques pour les réactions de décomposition principales des o-nitrotoluènes substitués en position 6 177

LISTE DES TABLEAUX

Tableau 1 Informations standard éxigées dans l'Annexe VII de REACH [17] 26

Tableau 2 –Pictogrammes et classes des dangers associés [24] 29

Tableau 3 Estimation des coûts en millions d'euros (M€) des campagnes de tests expérimentaux en fonction du recours aux méthodes alternatives [26] 31

Tableau 4 Energies de dissociation de la liaison C-NO_2 (E_{diss}) et descripteurs globaux de réactivité (en kcal/mol) calculés au niveau PBE0/6-31+G(d,p) pour 22 molécules nitroaromatiques 89

Tableau 5 Chaleurs de décomposition (en kcal/mol) expérimentales et prédites des 22 composés nitroaromatiques étudiés 91

Tableau 6 Corrélations linéaires de 14 descripteurs sélectionnés avec la chaleur de décomposition expérimentale de 22 composés nitroaromatiques 94

Tableau 7 Modèle à trois paramètres (3.16) pour la chaleur de décomposition à partir de structures calculées au niveau PBE0/6-31+G(d,p) (R^2=0,982; R^2_{cv}=0,974; F=330; s^2=963,17)..... 97

Tableau 8 Modèle à quatre paramètres (3.17) pour la chaleur de décomposition à partir de structures calculées au niveau PBE0/6-31+G(d,p) (R^2=0,989; R^2_{cv}=0,984; F=397; s^2=605,60)..... 97

Tableau 9 Modèle à quatre paramètres (3.18) pour la chaleur de décomposition à partir de structures calculées au niveau AM1 (R^2=0,984; R^2_{cv}=0,967; F=265; s^2=901,98) 99

Tableau 10 Chaleurs de décomposition (en kJ/mol) expérimentales et prédites à partir des modèles DFT et AM1 100

Tableau 11 Chaleurs de décomposition (en kJ/mol) expérimentales et prédites pour le jeu d'entraînement 104

Tableau 12 Chaleurs de décomposition (en kJ/mol) expérimentales et prédites pour le jeu de validation..... 106

Tableau 13 Analyse des domaines d'applicabilité (AD) des modèles QSPR proposés..... 107

Tableau 14 Données statistiques des modèles PLS en fonction du nombre de descripteurs 111

Tableau 15 Données statistiques pour 10 partages aléatoires entre jeux d'entraînement et de validation 114

Tableau 16 Analyse de la fiabilité du modèle multi-linéaire 116

Tableau 17 Analyse de la fiabilité du modèle par arbre de décision 119

Tableau 18 Chaleurs de décomposition (kJ/mol) expérimentales et prédites par MLR (Eq. 3.23), PLS et arbres de décision (DT) 121

Tableau 19 Températures de décomposition (en °C) expérimentales et prédites pour une série de composés nitroaromatiques 125

Tableau 20 Sensibilités à la décharge électrique (en J) obtenues à partir de différents dispositifs expérimentaux 128

Tableau 21 Sensibilités à la décharge électrique (en J) expérimentales et prédites des 26 molécules du jeu d'entraînement 133

Tableau 22 Sensibilités à la décharge électrique (en J) expérimentales et prédites des 15 molécules du jeu de validation 134

Tableau 23 Sensibilités à la décharge électrique (en J) expérimentales et prédites des 11 molécules de la base de données n'ayant pas été employées par Keshavarz dans son analyse 137

Tableau 24 Sensibilités à l'impact (en cm) expérimentales et prédites pour les composés de la base de données ne présentant pas de possibilité de liaisons C-H alpha 145

Tableau 25 Sensibilités à l'impact (en cm) expérimentales et prédites des composés de la base de données présentant des possibilités de liaisons C-H alpha 146

Tableau 26 Energies (ΔE), enthalpies (ΔH) et énergies libres de Gibbs relatives (ΔG) en kcal/mol pour les réactions de décomposition principaux du nitrobenzène au niveau PBE0/6-31+G(d,p) 157

Tableau 27 Constantes de Hammett [94] pour les substituants en position para et charges de NO_2 (en $|e^-|$) calculées à l'aide des analyses de populations naturelle (NPA) et de Mulliken 159

Tableau 28 Paramètres structuraux principaux des nitrobenzènes substitués en position para (nomenclature en Figure 57) 160

Tableau 29 Energies libres de Gibbs relatives (en kcal/mol) pour les réactions de décomposition vers NO et NO_2 de nitrobenzènes substitués en position para 160

Tableau 30 Charges de NO_2 (en $|e^-|$), calculées à l'aide des analyses de populations naturelles (NPA) et de Mulliken, et constantes de Hammett [94] pour les substituants en position méta 162

Tableau 31 Paramètres structuraux principaux des nitrobenzènes substitués en position méta. La nomenclature des atomes est indiquée en Figure 57 162

Tableau 32 Energies libres de Gibbs relatives (en kcal/mol) pour les réactions de décomposition vers NO et NO_2 de nitrobenzènes substitués en position méta 163

Tableau 33 Charges de NO_2 (en $|e^-|$), calculées à l'aide des analyses de populations naturelles (NPA) et de Mulliken 165

Tableau 34 Paramètres structuraux principaux des nitrobenzènes substitués en position ortho (nomenclature en Figure 57) 165

Tableau 35 Energies libres de Gibbs relatives (en kcal/mol) pour les réactions de décomposition vers NO et NO_2 de nitrobenzènes substitués en position ortho 166

Tableau 36 Energies (ΔE), enthalpies (ΔH) et énergies libres de Gibbs relatives (ΔG) en kcal/mol pour les réactions de décomposition principales de l'o-nitrotoluène au niveau PBE0/6-31+G(d,p) 166

Tableau 37 Charges de NO_2 (en $|e^-|$) calculées à partir de l'analyse de population naturelle (NPA) 169

Tableau 38 Paramètres structuraux principaux des o-nitrotoluènes substitués en position 3. La nomenclature des atomes est indiquée en Figure 64 169

Tableau 39 Energies libres de Gibbs relatives (en kcal/mol) pour les réactions de décomposition principales des o-nitrotoluènes substitués en position 3 170

Tableau 40 Paramètres structuraux principaux des o-nitrotoluènes substitués en position 4. La nomenclature des atomes est indiquée en Figure 64 171

Tableau 41 Energies libres de Gibbs relatives (en kcal/mol) pour les réactions de décomposition principales des o-nitrotoluènes substitués en position 4 172

Tableau 42 Paramètres structuraux principaux des o-nitrotoluènes substitués en position 5. La nomenclature des atomes est indiquée en Figure 64 173

Tableau 43 Energies libres de Gibbs relatives (en kcal/mol) pour les réactions de décomposition principales des o-nitrotoluènes substitués en position 5 174

Tableau 44 Paramètres structuraux principaux des o-nitrotoluènes substitués en position 6. La nomenclature des atomes est indiquée en Figure 64 176

Tableau 45 Energies libres de Gibbs relatives (en kcal/mol) pour les réactions de décomposition principales des o-nitrotoluènes substitués en position 6 177

INTRODUCTION GENERALE

Identifier et déterminer aussi tôt et de manière aussi fiable que possible les propriétés dangereuses des substances chimiques, quelle que soit leur utilisation finale, représente une étape initiale importante dans la maîtrise des risques industriels. Ceci est d'autant plus vrai pour les propriétés explosives intrinsèques des substances chimiques qui se classent au premier rang des dangers physico-chimiques pouvant être craints au cours du stockage, transport ou manipulation [1]. Aux vues de la complexité des phénomènes liés à ces propriétés d'explosibilité dans les cas réels, l'approche expérimentale reste, jusqu'à présent, la voie privilégiée pour l'évaluation de cette catégorie de dangers.

En rester à cette unique approche pose toutefois un réel problème pour les industriels avec les exigences qui sont maintenant les leurs dans le nouveau cadre règlementaire européen REACH, entré en vigueur en juin 2007, et le système de classification associé (CLP). Le coût d'une telle démarche, le temps nécessaire et même la disponibilité des laboratoires d'essais agréés pour la réalisation des essais rendent le processus particulièrement difficile du fait du nombre de substances concernées (plus de 143 000),. Aussi, l'utilisation de méthodes alternatives à l'expérimentation, parmi lesquelles les relations quantitatives structure activité/propriété (QSAR/QSPR) sont devenues d'un grand intérêt et sont même recommandées dans la nouvelle règlementation, afin d'obtenir les données nécessaires à l'enregistrement des substances.

Ces méthodes sont à l'heure actuelle principalement utilisées pour le criblage des propriétés toxiques, des nitroaromatiques par exemple [2]. La procédure de criblage consiste à utiliser ces méthodes pour prédire a priori les performances (ou la dangerosité) des substances pour des applications données. Ainsi, il est possible d'identifier les molécules présentant le plus fort potentiel pour la propriété attendue (recherche d'explosifs performants par exemple) ou celles présentant d'éventuels problèmes de sécurité. Les premières applications de ces méthodes, basées sur des analyses statistiques, ont concerné principalement des applications biologiques [3,4], toxicologiques [5,6] ou pharmaceutiques [7-9]. Néanmoins, l'intérêt à leur égard est grandissant ces dernières années pour la prédiction d'autres propriétés, en particulier physico-chimiques [10-12].

Dans le domaine des propriétés explosives, des méthodes reposant sur des groupes de contribution ont déjà été développées pour l'évaluation de l'énergie libérée pendant la décomposition (CHETAH) [13] et le calcul des réactions adiabatiques (CART) [14]. Grewer [15] a, quant à lui, mis en évidence l'influence de la structure chimique sur la stabilité thermique des substances chimiques. Depuis, Saraf et coll. [16] ont mis en évidence

la pertinence des outils de criblage reposant sur l'identification de relations entre la structure chimique et la stabilité thermique de composés nitroaromatiques.

C'est dans ce but que l'INERIS, en collaboration avec l'équipe de modélisation des systèmes complexes de Chimie ParisTech, a initié avec cette thèse en 2007 un projet nommé REPLACE ayant pour but le développement de méthodes prédictives pour les propriétés explosives des substances chimiques.

Il s'agit donc, dans ces travaux de thèse, de développer et d'évaluer le potentiel de tels modèles pour la prédiction des propriétés explosives de composés nitroaromatiques en utilisant une méthodologie de type QSPR reliant les propriétés expérimentales aux structures moléculaires, calculées à l'aide d'outils de chimie quantique, en particulier la Théorie de la Fonctionnelle de la Densité (DFT).

Dans le premier chapitre, le contexte de l'étude sera explicité. La règlementation REACH sera introduite afin de présenter ses principaux objectifs et ses conséquences pour l'industrie. Cette première partie se focalisera ensuite sur les propriétés explosives nécessaires à l'évaluation des substances. Enfin, les objectifs de cette thèse seront présentés ainsi que le projet INERIS au sein duquel elle intervient.

Le deuxième chapitre s'attachera, quant à lui, à introduire les différentes bases théoriques des outils employés. Ils sont de deux types. Les premiers sont les outils de chimie quantique utilisés pour le calcul des structures moléculaires ainsi que pour l'étude des mécanismes réactionnels de décomposition. Les seconds sont les outils nécessaires à la mise en place des modèles QSPR. Le principe de la méthodologie, les descripteurs moléculaires, ainsi que les outils d'analyse de données employés pour développer et évaluer les modèles, seront ainsi détaillés.

Le troisième chapitre sera dédié à la mise en place pratique des modèles prédictifs pour une famille particulière de molécules explosives, les nitroaromatiques. Les aspects méthodologiques seront, dans un premier temps, mis à l'épreuve pour prédire la chaleur de décomposition. La pertinence de différents types de descripteurs et outils d'analyses de données pour la mise en place des modèles sera testée. Des modèles seront ensuite développés pour prédire d'autres propriétés liées à l'explosibilité des substances : température de décomposition, sensibilités à la décharge électrique et à l'impact.

Enfin, un dernier chapitre sera dédié aux travaux relatifs à l'étude des mécanismes de décomposition des composés nitroaromatiques à l'aide de la DFT. Ces mécanismes complexes et variés rendent la mise en place des modèles plus difficile. Leur analyse explicite permet non seulement de mieux comprendre les mécanismes sous-jacents mais aussi d'aider à la mise en place de modèles plus fiables.

CHAPITRE 1

CONTEXTE ET OBJECTIFS

Sommaire

1- Le règlement REACH 24
1.1- Perception sociétale de la chimie 24
1.2- Vers un règlement européen unifié 25
1.3- Objectifs et champs d'application 25
1.4- Les procédures de REACH 27
1.5- Le règlement CLP 28
1.6- Les méthodes alternatives dans REACH 30
2- Explosibilité 32
2.1 - Définitions 32
2.2- Evaluation expérimentale des dangers d'explosion 36
2.3- Modélisation et explosibilité 40
3- Projet REPLACE pour le développement de méthodes alternatives 41

Le contexte de cette thèse, proposée par l'INERIS (Institut National de l'Environnement Industriel et des Risques), est lié au nouveau règlement européen REACH et aux exigences qu'il entraîne en termes de développement de méthodes alternatives à l'éxpérimentation. L'objectif est de développer et d'évaluer les performances de modèles prédictifs utilisables pour la caractérisation des propriétés des composés explosibles.

Dans ce premier chapitre, le nouveau contexte règlementaire lié aux produits chimiques sera présenté, ainsi que ses objectifs et implications. Dans un second temps, des détails supplémentaires seront donnés concernant la caractérisation des substances explosibles, qui est menée jusqu'à présent à l'aide d'outils principalement expérimentaux. Enfin, les objectifs de la présente thèse seront explicités.

1- Le règlement REACH

Le règlement REACH (pour *« Registration, Evaluation, Authorisation and restriction of CHemicals »*) [17] est entré en vigueur le 1er juin 2007 et constitue le nouveau cadre règlementaire concernant l'enregistrement des produits chimiques au sein de l'Union Européenne. Beaucoup d'informations concernant la mise en œuvre et les procédures de REACH sont disponibles en référence [18] ou sur les sites de différents instituts (INRS [19], ECHA [20]), dont l'INERIS, qui est le service national d'assistance (*helpdesk*) sur les aspects règlementaires et techniques du texte de REACH [21].

1.1- Perception sociétale de la chimie

Les produits chimiques font partie de notre quotidien, qu'ils soient d'origine naturelle ou qu'ils aient été développés par l'homme. Ils font partie intégrante de notre environnement, dans nos maisons, nos vêtements ou encore notre alimentation. Leur présence rend notre quotidien plus agréable, plus simple mais aussi plus sain. Cela dit, il n'en reste pas moins qu'ils peuvent également présenter des risques pour l'homme, les autres êtres vivants et l'environnement dans sa globalité.

Depuis la révolution industrielle du 19ème siècle, les progrès technologiques se sont accélérés, mettant en jeu des moyens de plus en plus importants, avec une aire d'influence aujourd'hui mondiale. Si l'amélioration de la qualité de vie a dominé les débats liés aux progrès issus de la chimie durant les 19ème et 20ème siècles, les préoccupations quant à leurs effets néfastes sur la santé humaine et l'environnement sont peu à peu devenues prépondérantes. Les développements menés dans le domaine militaire n'ont d'ailleurs fait que renforcer cette tendance.

Dès la première guerre mondiale, l'utilisation de gaz de combat à ouvert la voie au développement d'armes chimiques toujours plus terrifiantes d'efficacité. Les accidents chimiques de la deuxième partie du 20ème siècle ont également largement contribué aux craintes au sein de la société. Les catastrophes de Seveso (Italie) en 1976, Bhopal (Inde) en 1984 ou plus récemment l'explosion de l'usine AZF à Toulouse en 2001 sont des exemples des dangers que peuvent engendrer une mauvaise gestion des risques liés aux produits chimiques.

Au fil du temps, le terme "chimie" a ainsi été associé aux dangers de toxicité, de pollution, de réactivité incontrôlée et une certaine méfiance s'est développée au sein de l'opinion publique. Cette perception négative a pris peu à peu le pas sur les progrès qu'elle a engendrés. Dans un tel contexte, la gestion des risques chimiques est devenue une préoccupation majeure.

1.2- Vers un règlement européen unifié

Pour répondre au retour d'expérience né des accidents au sein de l'industrie chimique, nombre de textes ont été mis en place à l'échelle des états puis au niveau de l'Union Européenne (UE) dès les années 1960, en particulier avec la directive 67/548/CEE concernant la classification, l'emballage et l'étiquetage des substances dangereuses [22].

Mais, malgré les textes européens instaurés au cours du temps, des disparités entre états subsistent. Aussi, la mise en place d'un règlement unifiant les différents textes publiés a été l'occasion d'uniformiser l'application de ces règlements au sein des états membres vers un niveau élevé de protection de la santé et de l'environnement jusque là insuffisant dans certains pays. C'est en avril 1998, au cours d'une réunion des ministres européens de l'environnement des 15 états membres de l'UE à Chester (Royaume-Uni), qu'a été reconnue la nécessité de revoir la politique européenne en matière de produits chimiques.

La première étape vers un règlement global européen a ensuite été la publication du « Livre blanc » du 27 février 2001 [23]. Ce document, mis en place par les ministères de l'environnement des 15 états membres de l'UE, ainsi que par quelques 150 responsables politiques, industriels et associatifs issus d'organisations non-gouvernementales, pose les fondations du règlement REACH, tel qu'entré en application au 1er juin 2007. Au final, ce nouveau règlement unique remplace une quarantaine de textes législatifs relatifs aux substances chimiques.

1.3- Objectifs et champs d'application

Le règlement REACH vise à améliorer la connaissance des propriétés des substances chimiques qui circulent au sein de l'Union Européenne, qu'elles y soient fabriquées ou importées (à plus d'une tonne par an). Il s'agit ainsi d'assurer la maîtrise des risques liés à leur usage et le cas échéant, à réduire voire interdire l'utilisation des plus préoccupantes.

En ce sens, quatre objectifs majeurs peuvent être identifiés :

- Protection de la santé humaine et protection environnementale face aux risques potentiels des substances chimiques ;
- Compétitivité et innovation de l'industrie chimique européenne ;
- Libre circulation des substances sur le marché intérieur de l'Union Européenne ;
- Promotion des méthodes alternatives pour l'évaluation des dangers des substances.

Les champs d'application de ce règlement couvrent à la fois la production, la mise sur le marché et l'utilisation des substances elles-mêmes mais aussi des préparations dans lesquelles on les retrouve. Quelques exceptions restent cependant en marge de ce règlement : substances radioactives, intermédiaires non isolés, polymères, déchets ou encore transport des marchandises dangereuses, pour lesquels les directives en vigueur restent en application.

Les exigences en matière d'informations standard pour l'enregistrement d'une substance sont d'ordre physico-chimique, toxicologique et éco-toxicologique. Au total, la caractérisation de 21 propriétés, présentées dans le Tableau 1, est exigée pour une substance, fabriquée ou importée à plus d'une tonne par an (cf. Annexe VII du règlement REACH [17]). D'autres informations sont requises lorsque le tonnage est plus important (cf. Annexes VIII, IX et X). A ce jour, ces propriétés (physico-chimiques, toxicologiques ou éco-toxicologiques) sont déterminées au moyen d'essais en laboratoire ou à grande échelle.

7. INFORMATIONS SUR LES PROPRIETES PHYSICO-CHIMIQUES
7.1 Etat de la substance à 20°C et 101,3 kPa
7.2 Point de fusion / congélation
7.3 Point d'ébullition
7.4 Densité relative
7.5 Pression de vapeur
7.6 Tension superficielle
7.7 Hydrosolubilité
7.8 Coefficient de partage n-octanol/eau
7.9 Point d'éclair
7.10 Inflammabilité
7.11 Propriétés explosives
7.12 Température d'auto-inflammation
7.13 Propriétés comburantes
7.14 Granulométrie
8. INFORMATIONS TOXICOLOGIQUES
8.1 Irritation ou corrosion cutanée
8.2 Irritation oculaire
8.3 Sensibilisation cutanée
8.4 Mutagénicité
8.5 Toxicité aigüe
9. INFORMATIONS ECOTOXICOLOGIQUES
9.1 Toxicité aquatique
9.2 Dégradation

Tableau 1 - Informations standard éxigées dans l'Annexe VII de REACH [17]

Parmi les acteurs mis en jeu dans le règlement REACH, les industriels jouent un rôle crucial. En effet, REACH impose à l'industrie (fabricants, importateurs, utilisateurs) la responsabilité de la gestion des risques liés aux substances chimiques. Il est également de son ressort de fournir les informations de sécurité aux utilisateurs en aval afin d'assurer un emploi en toute sécurité. Les fabricants et importateurs sont donc tenus d'identifier et gérer les risques des substances chimiques produites, importées et mises sur le marché.

Pour des quantités supérieures à une tonne par an, ces derniers doivent prouver l'application de ces dispositions par la procédure dite d'enregistrement des substances auprès de l'agence européenne des produits chimiques (ECHA pour *European CHemicals Agency* [20]). Cette dernière a été mise en place en juin 2007 afin de coordonner la mise en œuvre de l'ensemble des prérogatives liées à REACH à l'échelle européenne. Située à Helsinki (Finlande), il est notamment de son ressort de vérifier la conformité des dossiers d'enregistrement. Elle se charge également d'examiner les propositions d'essais afin de limiter le recours excessif à l'expérimentation (en particulier sur les animaux). Dans cette démarche, cette agence s'appuie sur les institutions nationales au sein des états membres, puisque la mise en œuvre du règlement REACH reste de la responsabilité de ces derniers.

1.4- Les procédures de REACH

REACH est entré en vigueur au 1er juin 2007. A partir de cette date, les procédures liées à ce règlement (enregistrement, évaluation, autorisation des substances chimiques, restrictions applicables à certaines d'entre elles) se sont mises en place.

Pré-enregistrement

Une première phase de pré-enregistrement est intervenue entre le 1er juin et le 1er décembre 2008. Il s'agissait pour les fabricants et importateurs de pré-enregistrer les substances circulant déjà sur le marché. Cela a permis aux déclarants, non seulement de pouvoir enregistrer ces substances ultérieurement mais aussi d'échanger des données afin d'éviter la réalisation d'essais redondants. Au total, ce sont plus de 143 000 substances qui ont été préenregistrées par quelques 65 000 industriels.

Enregistrement et évaluation

Il s'ensuit actuellement la phase proprement dite d'enregistrement des substances existantes produites et importées à plus d'une tonne par an. Les fabricants et importateurs doivent collecter les informations concernant les propriétés des substances afin de déterminer les données de sécurité liées à ces dernières. Pour cela, un dossier d'enregistrement, soumis à l'ECHA, est constitué. Les informations à notifier dans ce dossier comprennent différents éléments tels que :

- l'identité et les coordonnées du producteur ou de l'importateur ;
- le ou les numéros d'enregistrement ;
- l'identité de la ou des substances ;
- la classification de la ou des substances ;
- une brève description de la ou des utilisations de la ou des substances ;
- la fourchette de quantité de la ou des substances.

Dans cette étape du processus, l'ECHA coordonne l'évaluation des substances, réalisée sous le contrôle des états membres eux-mêmes. Les substances existantes doivent être enregistrées, en fonction de leur tonnage, d'ici au 1er juin 2018 selon le calendrier décrit en Figure 1. Bien entendu, les substances nouvelles sont et resteront soumises à enregistrement au-delà de ce délai.

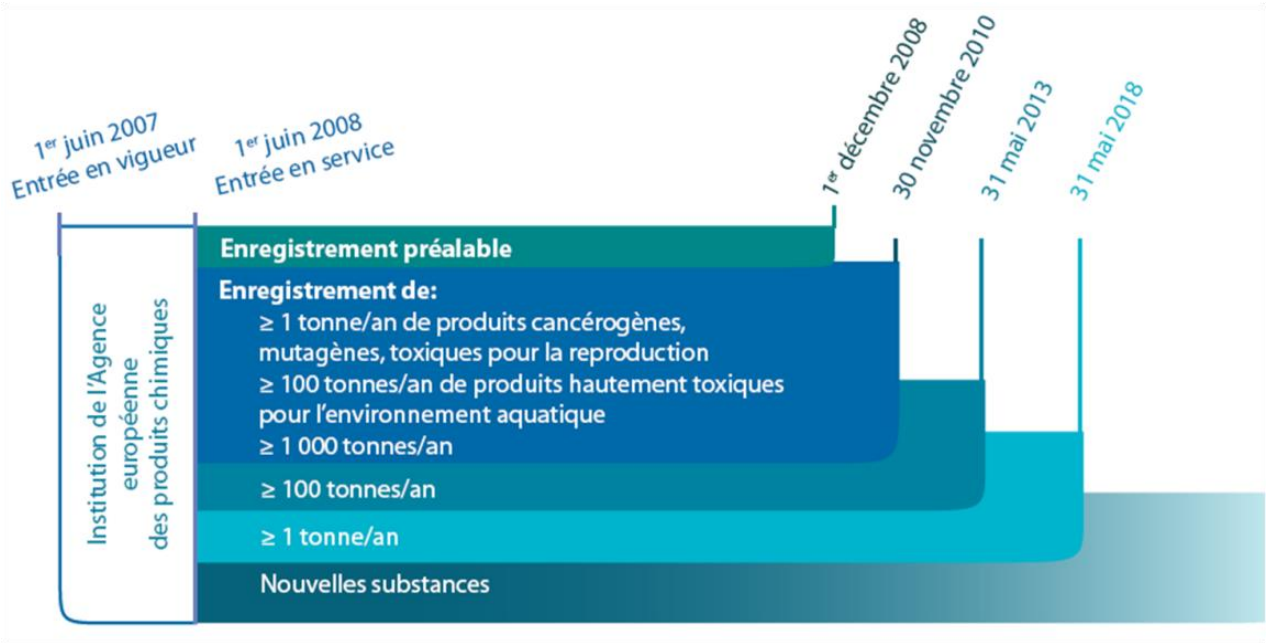


Figure 1 - Calendrier prévu pour l'enregistrement des substances dans le cadre du règlement REACH (extrait de [20])

Autorisation et restriction

Suite à son évaluation, une substance peut être considérée comme préoccupante, voire extrêmement préoccupante. Dans ce dernier cas, une demande d'autorisation est requise afin de vérifier que les risques associés à son utilisation sont maîtrisés ou que les avantages de son utilisation l'emportent sur les risques encourus. Dans ce contexte, il est d'ailleurs demandé aux industriels d'envisager des moyens de substitutions pour ces substances. L'Union Européenne peut aller jusqu'à mettre en place des mesures de restriction voire d'interdiction de la fabrication, de la mise sur le marché et de l'utilisation de certaines substances lorsque les risques pour la santé humaine et l'environnement sont considérés comme inacceptables.

1.5- Le règlement CLP

En association au règlement REACH, un nouveau système de classification, d'étiquetage et d'emballage des substances (CLP, pour *Classification, Labelling and Packaging of substances and mixtures* [24]) a également été mis en place en décembre 2008 et entré en vigueur au 20 janvier 2009. Ce dernier met en œuvre en Europe les recommandations internationales du SGH (Système Général Harmonisé) [25] en vigueur au niveau international. Il définit les nouvelles règles applicables en matière de classification, d'étiquetage et d'emballage de produits chimiques en Europe et est donc complémentaire au règlement REACH et aux méthodes d'essais qui lui sont associées. Ce système définit notamment 28 classes de dangers, auxquelles sont associés 9 pictogrammes, présentés dans le Tableau 2.

Pictogramme	Classes de dangers
	Explosibles Substances et mélanges auto-réactifs Peroxydes organiques
	Gaz inflammables Aérosols inflammables Liquides inflammables Matières solides inflammables Substances et mélanges auto-réactifs Liquides pyrophoriques Matières solides pyrophoriques Substances et mélanges auto-échauffants Substances et mélanges qui, au contact de l'eau dégagent des gaz inflammables Peroxydes organiques
	Gaz comburants Liquides comburants Matières solides comburantes
	Gaz sous pression
	Substances et mélanges corrosifs pour les métaux Corrosion / irritation cutanée Lésions oculaires graves / irritation oculaire
	Toxicité aiguë
	Toxicité aiguë Corrosion / irritation cutanée Lésions oculaires graves / irritation oculaire Toxicité spécifique pour certains organes cibles – exposition unique
	Sensibilisation respiratoire Mutagénicité sur les cellules germinales Cancérogénicité Toxicité pour la reproduction Toxicité spécifique pour certains organes cibles – exposition unique Toxicité spécifique pour certains organes cibles – exposition répétée Danger par aspiration
	Dangers pour le milieu aquatique

Tableau 2 –Pictogrammes et classes des dangers associés [24]

1.6- Les méthodes alternatives dans REACH

Comme cela a été présenté précédemment, la connaissance de nombreuses propriétés est exigée dans le cadre du règlement REACH. De plus, si le Livre Blanc estimait à 30 000 le nombre de substances concernées, plus de 143 000 substances ont finalement été préenregistrées entre juin et décembre 2008. Aussi, une caractérisation expérimentale complète dans le calendrier prévu (cf. Figure 1) est impossible pour des raisons de temps, de coût, de dangerosité de certains essais ou d'éthique (limitation des essais sur animaux vertébrés). L'utilisation de méthodes alternatives à l'expérience est donc autorisée voire recommandée par le règlement lui-même (cf. Article 13 de REACH [17]) :

1. Des informations sur les propriétés intrinsèques des substances peuvent être produites par d'autres moyens que des essais pour autant que les conditions énoncées à l'annexe XI soient respectées. En ce qui concerne la toxicité pour l'espèce humaine en particulier, les informations sont produites autant que possible par d'autres moyens que des essais sur des animaux vertébrés, par le recours à des méthodes alternatives, par exemple les méthodes in vitro ou par des modèles de relations qualitatives ou quantitatives structure-activité ou par l'exploitation de données sur des substances structurellement proches (regroupement ou références croisées).

En particulier, les relations quantitatives structure-activité (QSAR), permettant d'évaluer les propriétés à partir de la structure d'une substance, ont été identifiées comme des méthodes alternatives à encourager sous certaines conditions, comme indiqué dans l'annexe XI de REACH [17] :

1.3. Relation qualitative ou quantitative structure-activité (RSA)

Les résultats obtenus à l'aide des modèles valides de la relation qualitative ou quantitative structure-activité (R(Q)SA) peuvent indiquer la présence ou l'absence d'une certaine propriété dangereuse. Les résultats de la R(Q)SA peuvent être utilisés au lieu de l'essai lorsque les conditions suivantes sont réunies :

- *les résultats sont issus d'un modèle R(Q)SA dont la validité scientifique a été établie,*
- *la substance relève du domaine d'applicabilité du modèle R(Q)SA,*
- *les résultats conviennent pour la classification et l'étiquetage, et/ou pour l'évaluation des risques, et*
- *une description suffisante et fiable de la méthode appliquée est fournie.*

D'un point de vue économique, les essais représentent la majeure partie des coûts mis en jeu par REACH. Une étude réalisée en 2003 [26] avait estimé que l'utilisation des méthodes alternatives dans ce nouveau cadre règlementaire représenterait une réduction des coûts pouvant atteindre plus de 50 %, comme montré dans le Tableau 3.

Recours	Coût estimé
Maximum	1180 M€
Moyen	1561 M€
Minimum	2423 M€

Tableau 3 – Estimation des coûts en millions d'euros (M€) des campagnes de tests expérimentaux en fonction du recours aux méthodes alternatives [26]

Elle avait également estimé le nombre de substances concernées à 30 000 [26]. Or, 143 000 substances ont d'ores et déjà été pré-enregistrées. L'utilisation des méthodes alternatives pourrait donc engendrer une réduction des coûts encore plus importante que prévue par ce rapport.

Afin d'accompagner le développement des méthodes alternatives, un organisme dédié, l'ECVAM (*European Center for the Validation of Alternative Methods*) [27], a été mis en place à Ispra (Italie) dès 1991. Il est aujourd'hui de son ressort de coordonner les efforts concernant le développement et la validation de ces dernières au niveau de l'UE dans le cadre de REACH.

Des règles précises ont également été mises en place récemment par l'OCDE (Organisation de Coopération et de Développement Economique) [28] pour la validation des modèles QSAR [29]. La validation des modèles repose alors sur cinq grands principes [30] :

- Une propriété ciblée définie (avec un protocole expérimental identifié) ;
- Un algorithme sans équivoque ;
- Un domaine d'applicabilité défini ;
- Des mesures appropriées de la qualité d'ajustement, de robustesse et du pouvoir prédictif ;
- Si possible, une interprétation des mécanismes sous-jacents.

Jusque-là, très peu de modèles ont été mis en place suivant ces principes [27]. Pour aider au développement de tels models, l'OCDE a mis en place la *(Q)SAR Application Toolbox* qui regroupe une série d'outils aidant à la collecte d'informations sur les substances chimiques ainsi que des outils d'analyse de similarité entre molécules pour une première évaluation des propriétés à partir de la connaissance de celles de molécules similaires.

Dans ce développement, la modélisation moléculaire a un rôle important à jouer [31] puisqu'elle permet de décrire les structures moléculaires employées dans les analyses de type QSAR. De plus, elle sera particulièrement utile dans la détermination des mécanismes mis en jeu qui sont non seulement demandés dans la démarche de validation, mais font partie intégrante de la mise en place des modèles prédictifs. Certaines propriétés (ex : densité, viscosité) pourront même être directement déterminées à l'aide d'approches informatisées, telles que la dynamique moléculaire.

Une autre voie d'application de la modélisation moléculaire concerne la substitution des substances préoccupantes. Les outils de modélisation peuvent alors aider non seulement à la caractérisation des données de sécurité mais aussi à la recherche des propriétés de service des produits et des possibles composés alternatifs. Ils permettent alors de choisir les solutions les plus probantes, en amont même de toute synthèse expérimentale.

Les travaux réalisés au cours de cette thèse s'inscrivent dans la démarche de développement de nouveaux modèles prédictifs (en l'occurrence QSPR) pour les propriétés explosives des substances chimiques, dont la caractérisation fait partie intégrante des activités de l'INERIS depuis de très nombreuses années [32].

2- Explosibilité

Une étape initiale importante de la gestion des risques industriels réside dans l'identification et la détermination du caractère potentiellement explosif des substances, quelle que soit leur utilisation finale. Aussi, la caractérisation des propriétés explosives intrinsèques est considérée de manière prioritaire en termes d'évaluation des risques physico-chimiques pour un composé chimique donné. Aussi, l'approche expérimentale est restée jusque-là la voie quasi exclusive pour l'évaluation des dangers d'explosion. Cependant, les exigences du nouveau règlement REACH rendent l'utilisation de cette seule approche très difficile voire impossible pour l'industrie dans le calendrier considéré. De plus, de part la nature même des propriétés testées, l'approche expérimentale peut présenter des problèmes de sécurité. En effet, non seulement les tests doivent être réalisés par des organismes compétents dans le domaine mais des précautions particulières sont également à prendre en termes de toxicité puisque les produits de décomposition d'une substance peuvent être eux-mêmes toxiques. Aussi, le développement de méthodes prédictives par modélisation pouvant être complémentaires aux essais expérimentaux présente un grand intérêt.

De nombreux ouvrages traitent des phénomènes d'explosion [33-39]. Le but est ici de présenter les phénomènes chimiques mis en jeu au cours d'une explosion et d'expliciter comment ces propriétés explosives sont caractérisées, en particulier expérimentalement, au niveau règlementaire basé sur les recommandations du manuel d'épreuves et de critères des Nations Unies pour le transport de marchandises dangereuses [40].

2.1- Définitions

Explosion

Une explosion résulte de la production d'une grande quantité d'énergie et de gaz dans un laps de temps très réduit. L'énergie libérée est alors dissipée par ondes de choc, par propulsion de matière ou encore par radiations thermiques. On peut distinguer différents types d'explosions en fonction de l'origine de cette énergie : explosions atomiques, physiques ou chimiques.

Les explosions atomiques, issues de réactions nucléaires, sont les plus puissantes. L'énergie qu'elles dissipent est d'un million à un milliard de fois plus importante que celle d'une explosion chimique. Cette énergie est dissipée sous forme d'onde de choc, de rayonnements thermiques et de radiations, avec des conséquences terribles à grandes échelles – à la fois spatiales et temporelles – telles que mises en évidence à Hiroshima et Nagasaki.

Les explosions physiques, quant à elles, peuvent avoir lieu sous l'effet d'une transformation physique soudaine, telle qu'une compression. L'énergie potentielle du système est alors convertie en énergie cinétique. Sa température augmente et une onde de choc se dissipe dans le milieu environnant. Ce type d'explosion peut notamment intervenir lorsqu'un liquide froid et volatil rencontre un autre liquide à une température bien supérieure à la température d'ébullition du liquide froid. Lors de l'éruption du volcan Krakatoa en 1883, par exemple, une grande quantité de lave en fusion s'est déversée dans l'océan. La vaporisation instantanée de plus d'un kilomètre cube d'eau de mer entraîna une onde de choc entendue à des milliers de kilomètres.

Enfin, les explosions chimiques résultent d'une réaction chimique exothermique extrêmement rapide générant une grande quantité de chaleur et généralement de gaz très chauds. Il s'agit d'une réaction en chaîne qui, une fois initiée, peut devenir incontrôlable et forme de nombreux produits de décomposition tels que CO_2, H_2O, NO, N_2. Ce sont plus particulièrement les explosifs chimiques qui nous intéressent ici.

Pour avoir lieu, la réaction en chaîne nécessite une certaine énergie d'activation, comme montré en Figure 2. L'énergie libérée par le processus peut avoir deux origines qui ne sont bien entendu pas exclusives :

- soit la réaction est fortement exothermique ;
- soit elle engendre des composés très réactifs engendrant eux-mêmes des réactions exothermiques avec leur environnement.

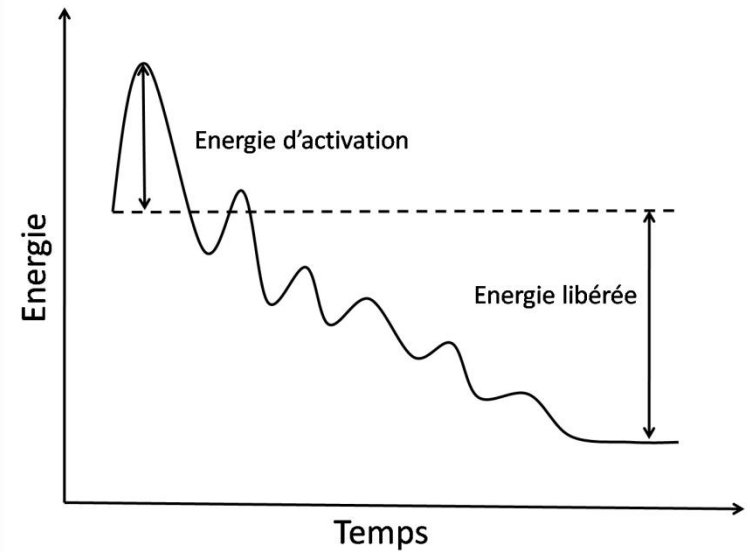


Figure 2 – Profil typique d'une décomposition explosive

De plus, pour donner lieu à une explosion, cette réaction doit être particulièrement rapide. Cela implique des conditions particulières (confinement, haute température, haute pression) et/ou la création au cours du processus de décomposition de composés hautement réactifs. En particulier, la présence au sein même du composé d'éléments comburants, tels que l'oxygène, est un facteur favorisant, puisque ces derniers constituent un comburant interne au matériau.

Finalement, on retrouve le triangle du feu, en Figure 3, qui indique que pour qu'un feu ait lieu, il faut un combustible, un comburant et une source d'inflammation. On substitue parfois ce schéma par un hexagone pour les phénomènes explosifs, en particulier pour les explosions de poussières. Dans ce cas-là, pour que l'explosion ait lieu, il ne faut pas seulement 3 mais 6 conditions réunies :

- un comburant ;
- un combustible ;
- une source d'inflammation ;
- des produits en suspensions (gaz, aérosols, poussières) ;
- l'obtention d'un domaine d'explosibilité, c'est-à-dire le domaine de concentration du combustible dans l'air à l'intérieur duquel les explosions sont possibles ;
- un confinement suffisant.

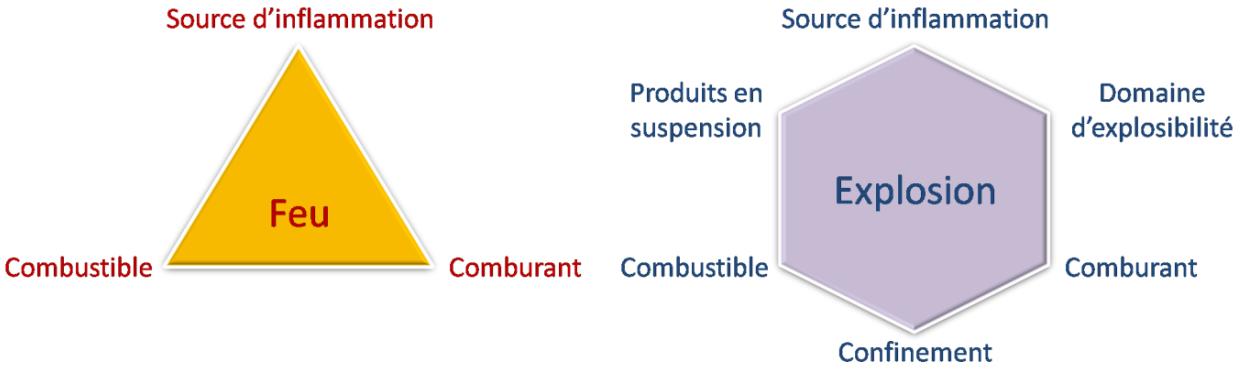


Figure 3 - Triangle du feu et hexagone de l'explosion (extrait de [41])

Détonation et déflagration

Deux types de phénomènes explosifs se distinguent : la détonation et la déflagration. Ils diffèrent notamment par leurs vitesses de propagation.

En fait, sur l'échelle des vitesses de propagation, on retrouve tout d'abord les phénomènes de combustion dans lesquels la propagation se fait de manière conductive avec une vitesse de propagation de l'ordre de quelques millimètres par seconde.

Dans une déflagration, le mélange, alors explosif, brûle à une vitesse subsonique pouvant être de plusieurs centaines de mètres par seconde. Il s'agit alors d'un phénomène convectif.

Des vitesses de propagation bien supérieures sont atteintes pour une détonation. En effet, le front de flamme se propage comme une onde de choc, à une vitesse supérieure à la vitesse du son (plus de 2000 m/s pour un mélange air/hydrocarbure). Celle-ci est entretenue par l'énergie libérée par l'onde de combustion qui la suit. Ainsi, une détonation engendre des pressions beaucoup plus importantes. Elle est donc plus destructrice qu'une déflagration. Cela dit, une déflagration peut passer en régime de détonation dans des conditions particulières, comme sous l'effet d'un confinement (dans un tube de longueur importante par exemple). Ces transitions peuvent être très puissantes puisque les vitesses atteintes peuvent excéder les vitesses de détonation en régime constant.

Matériaux énergétiques

Les matériaux énergétiques sont des matériaux, en général solides, capables de restituer de façon maîtrisée l'énergie chimique (importante) emmagasinée dans les édifices moléculaires qui les constituent. Ils trouvent des applications dans les domaines de la défense, du génie civil ou de l'automobile (airbags).

Parmi eux, les explosifs primaires sont des matériaux qui entrent en détonation lorsqu'ils sont excités par un stimulus extérieur (choc, étincelle électrique…).

Les explosifs secondaires, quant à eux, détonent sous l'effet d'un explosif primaire. Parmi eux, différentes familles existent, caractérisées par la présence de liaisons particulières, comme montré en Figure 4.

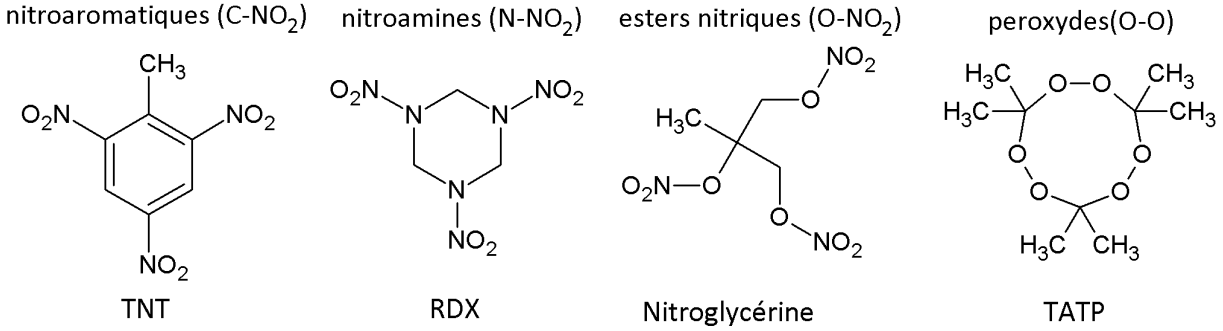


Figure 4 - Exemples de composés explosifs

Dans la famille des composés nitroaromatiques, certains composés sont très connus tels que le 2,4,6-trinitrotoluène (TNT) ou le 2,4,6-trinitrophénol (acide picrique). Ils entrent notamment dans la constitution de compositions explosives courantes telle que la pentolite constituée de 50% de TNT et 50% de PETN (tétranitrate de pentaérythrite) et utilisée dans le domaine militaire comme civil.

2.2- Evaluation expérimentale des dangers d'explosion

La caractérisation des dangers d'explosion dans les différentes règlementations en vigueur se fait à partir de dispositifs expérimentaux permettant d'appréhender les divers aspects du risque d'explosion des substances chimiques [42]. Ces tests ont été développés pour caractériser au mieux le comportement des substances explosibles, c'est-à-dire pouvant présenter des propriétés explosives, dans les conditions réelles de leur utilisation (stockage, transport, manipulation).

Des essais permettent ainsi d'évaluer l'aptitude de ces substances à la détonation, à la déflagration, le comportement dans un chauffage sous confinement et leurs niveaux de sensibilité à différents stimuli mécaniques ou thermiques. Ces essais, extraits du règlement dit méthodes d'essais de REACH [43] et du Manuel d'épreuves et de critères des Recommandations relatives au transport des marchandises dangereuses [40], sont utilisés respectivement pour la classification, l'étiquetage et l'emballage des substances explosives et leur classement au transport. Ils s'appliquent aux solides, aux liquides mais aussi aux gels ou suspensions. Des tests spécifiques sont consacrés à la caractérisation des gaz et vapeurs. La procédure de classification des substances dans la classe des explosibles est très complexe tel que représenté dans les diagrammes de décision en Annexe 1 (extraits du Manuel d'épreuves et de critères relatif à la réglementation du Transport des Marchandises Dangereuses [40]).

Aptitude à la détonation

Il s'agit de caractériser la propension d'un échantillon en environnement confiné à entrer en régime détonant. Cette propriété est caractérisée en disposant l'échantillon à tester dans un tube d'acier (afin de simuler l'effet de masse) et en le soumettant à une onde de choc intense provenant d'un détonateur ou d'un relais explosif puissant.

Aptitude à la déflagration

Il est également important de caractériser la possibilité d'amorçage et de propagation d'une déflagration. Pour cela, une épreuve pression/temps peut être réalisée. Il s'agit de placer 5 g d'échantillon dans une bombe et d'observer l'évolution de la pression en fonction du temps après amorçage. Il s'agit alors de déterminer si et à quelle vitesse la pression dans la bombe atteint 2070 kPa :

- "Oui, rapidement" : si cette pression est atteinte en moins de 30 ms ;
- "Oui, lentement" : si cette pression est atteinte en 30 ms ou plus ;
- "Non" : si cette pression n'est pas atteinte.

L'aptitude d'une matière à passer de la déflagration à la détonation est également caractérisée à l'aide d'épreuves spécifiques.

Chauffage sous confinement

Cet essai a pour but de définir la possibilité d'explosion thermique d'une petite quantité de matière afin d'évaluer si la décomposition peut devenir plus ou moins violente sous confinement comme cela peut être le cas dans un emballage ou dans un réacteur.

L'un des tests existant est l'épreuve de Koenen (en Figure 5) qui consiste à chauffer l'échantillon dans un tube d'acier fermé par des plaques percées par un trou d'un diamètre donné, afin d'évaluer la possibilité d'explosion dans des conditions de température élevée avec un confinement défini.

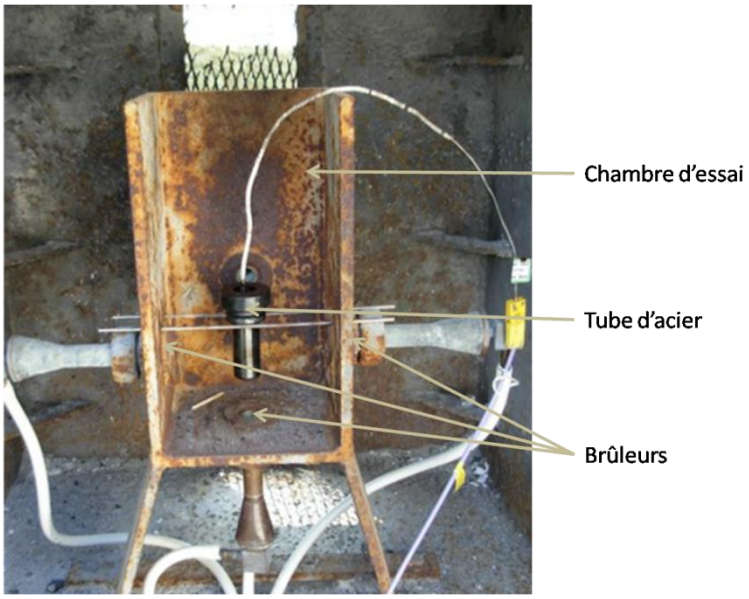


Figure 5 – Epreuve de Koenen (Appareillage INERIS)

Sensibilité

La réaction de décomposition explosive peut être initiée par divers stimuli (cf. Figure ***6***): onde de choc, choc mécanique, frottement, flamme ou encore étincelle électrique. Aussi existe-t-il différentes épreuves capables de caractériser les sensibilités des substances aux différents stimuli auxquels elle peut être confrontée.

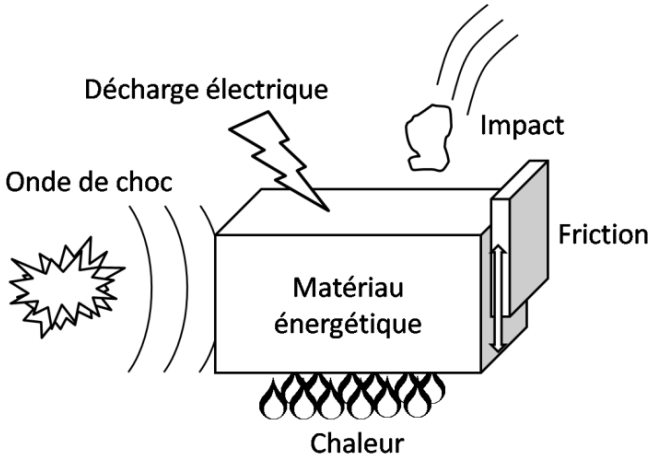


Figure 6 – Principaux modes de sensibilité des substances explosibles

D'une manière générale, il s'agit de soumettre un échantillon de matière au stimulus considéré et d'analyser son comportement afin d'observer d'éventuelles réactions (explosion, bruit, fumée, odeur). Les sensibilités à la décharge électrique et à l'impact seront étudiées plus tard. Des détails concernant les dispositifs expérimentaux seront alors proposés.

Potentiel explosif

D'autres épreuves permettent enfin de caractériser le potentiel d'une substance en cas d'explosion. Il s'agit d'évaluer la puissance et donc les dégâts éventuels que pourrait causer son explosion. Pour ce faire, l'épreuve du mortier balistique peut être employée. La Figure 7 est une photographie du dispositif employé à l'INERIS.

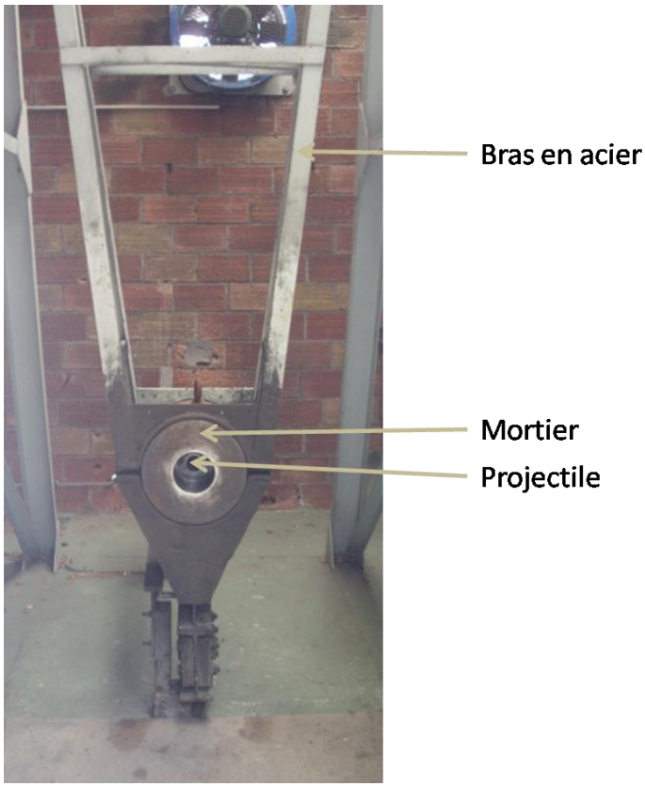


Figure 7 – Mortier balistique (appareillage INERIS)

L'échantillon est placé dans un projectile en acier, lui-même installé dans un mortier. L'échantillon est alors soumis à l'action d'un détonateur. La puissance explosive est alors déterminée par le recul du mortier, comparé à celui obtenu pour une substance de référence, l'acide picrique.

Un autre test est l'épreuve au bloc de plomb (test de Trauzl) [36] qui consiste à faire détoner un échantillon d'explosif au fond d'une cavité au sein d'un bloc de plomb et à mesurer l'augmentation du volume de la cavité, tel que montré en Figure 8.

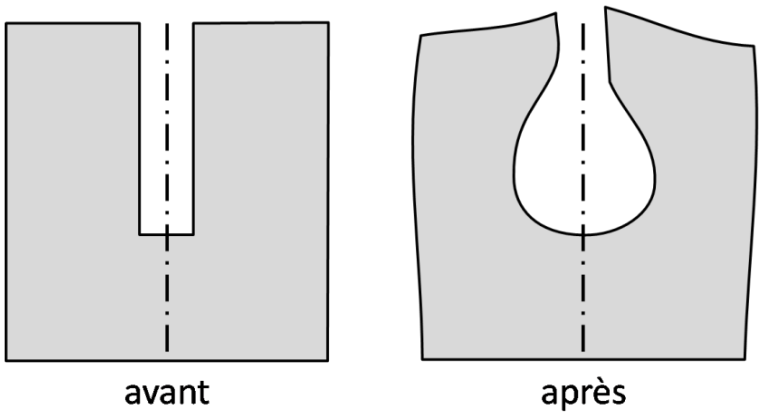


Figure 8 - Coupe du bloc de plomb avant et après le tir d'explosif au cours d'un test de Trauzl [35]

Procédures de présélection

Les tests expérimentaux précédemment présentés sont destinés à caractériser le comportement des substances dans les conditions réelles d'utilisation, de stockage et de transport. Ils nécessitent donc des quantités de matières significatives (quelques grammes à plusieurs kilogrammes). Certaines procédures, dites de présélection, consistant en une combinaison d'évaluations théoriques (d'ordre thermodynamique ou structural) et d'épreuves à petite échelle, ont donc été mises en place (cf. Appendice 6 du Manuel d'épreuves et de critères ONU [40]). Leur but est de montrer qu'une substance n'est, en aucune façon, susceptible de se décomposer rapidement en libérant des gaz ou de la chaleur (à savoir que la substance ne présente aucun risque d'explosion).

La première d'entre elle consiste tout simplement à considérer la présence de groupements chimiques associés à des propriétés explosives au sein de la molécule. Ces groupements, dits explosophores, sont :

- liaisons C-C insaturées (ex : acétylène) ;
- C-métal ou N-métal (ex : composés organo-lithium) ;
- atomes d'azote contigus (ex : azotures) ;
- atomes d'oxygène contigus (ex : peroxyde) ;
- liaisons N-O (ex : composés nitroaromatiques) ;
- liaisons N-halogène (ex : fluoroamines) ;
- liaisons O-halogène (ex : chlorates).

Un indice empirique, la balance en oxygène (BO), est également calculé. Il a été développé pour caractériser la quantité d'oxygène nécessaire pour la conversion complète du carbone et de l'hydrogène de la molécule en dioxyde de carbone (CO_2) et eau (H_2O), tel que proposé par Lothrop [44].

$$OB = \frac{-1600(2X + Y/2 - Z)}{M_w} \quad (1.1)$$

où X, Y et Z sont respectivement les nombres d'atomes de carbone, d'hydrogène et d'oxygène dans la molécule.

Si la balance en oxygène est inférieure à -200, le composé n'est pas éligible pour un classement parmi les substances explosibles.

En phase de présélection, la stabilité thermique est ensuite testée. Sa caractérisation est réalisée par calorimétrie [45]. Plus de détails concernant la détermination expérimentale seront donnés ultérieurement. Si la chaleur de décomposition est inférieure à 500 J/g et la température de décomposition inférieure à 500°C, le composé n'est pas, non plus, éligible à la classe des composés explosibles. Dès lors, il n'est pas nécessaire de réaliser la série de tests de l'Annexe 1, présentée précédemment.

2.3- Modélisation et explosibilité

Les outils de modélisation permettent de caractériser les phénomènes chimiques, biologiques ou physiques à différentes échelles (cf. Figure 9).

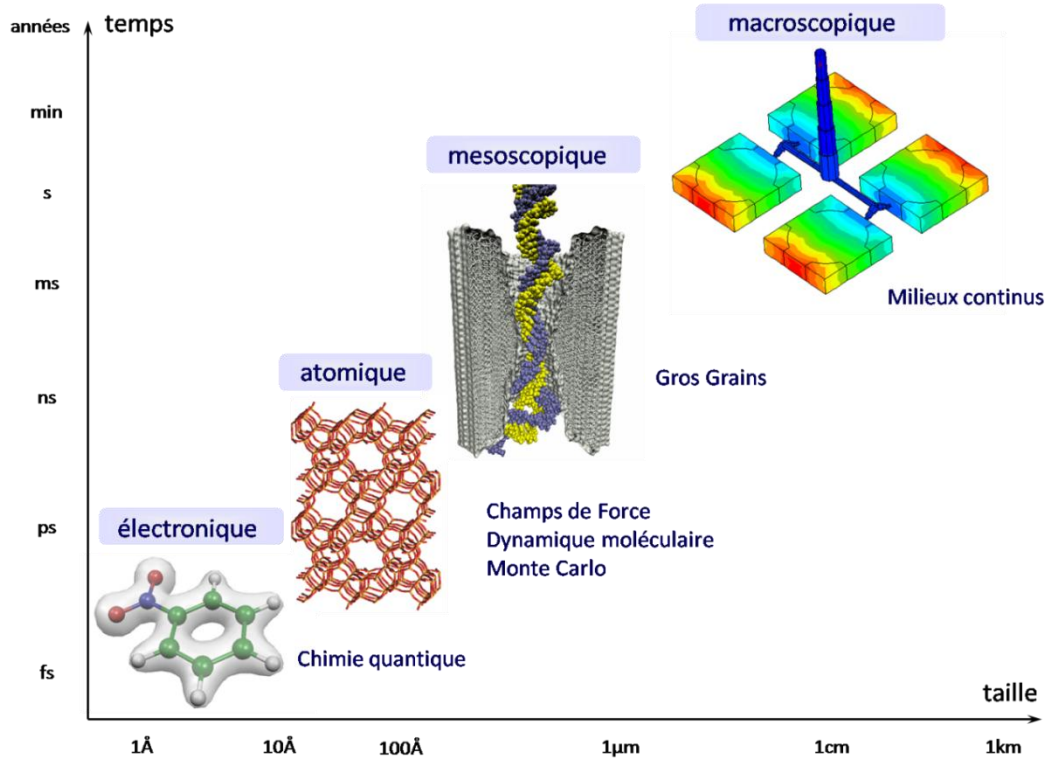


Figure 9 – Echelles de modélisation

A l'échelle macroscopique, l'approche des milieux continus permet par exemple de simuler des procédés tels que l'injection d'une matière plastique. Cette échelle permet, par exemple, de s'intéresser à des systèmes biologiques tels que l'ADN. A une échelle inférieure, les méthodes de modélisation moléculaire (chimie quantique, méthodes par champs de force) permettent de caractériser les propriétés chimiques et physiques des molécules.

Ces différentes méthodes permettent d'appréhender les phénomènes explosifs sous des angles différents. Différentes applications de modélisation à l'échelle macroscopique peuvent être trouvées en lien avec ce type de phénomènes [46-48]. Elles sont employées, par exemple, de modéliser le comportement des structures sous l'effet d'une explosion [46].

La modélisation moléculaire, quant à elle, est d'un grand intérêt pour l'analyse à une échelle inférieure [49,50]. Elle permet de caractériser, à l'échelle moléculaire, les composés explosifs ainsi que leurs mécanismes de décomposition, en identifiant les mécanismes réactionnels de décomposition [38,51]. Elle est également utilisée, comme ici, pour le développement de modèles prédictifs des propriétés macroscopiques d'explosibilité à une échelle moléculaire à l'aide d'outils statistiques et de chimie quantique [16,52-54]. Dans ce cadre, elle donne accès aux descripteurs moléculaires utilisés dans les modèles qui permettent de calculer les propriétés macroscopiques. D'autres approches à l'échelle moléculaire sont envisageables. Politzer a notamment utilisé les outils de dynamique moléculaire pour caractériser les propriétés cristallines des matériaux énergétiques solides [55].

Grâce à l'accroissement des moyens de calculs, des propriétés complexes, telles que les propriétés explosives, sont maintenant accessibles, permettant non seulement une meilleure prédiction mais aussi une meilleure compréhension des phénomènes. On notera d'ailleurs de très récents travaux de dynamique moléculaire quantique ayant permis, pour la première fois, de simuler à l'échelle moléculaire un matériau énergétique (le nitrométhane) dans des conditions proches de détonation [56]. De tels outils donnent accès à des informations non disponibles avec des dispositifs expérimentaux actuels. Cedi est dû au caractère extrême des conditions expérimentales correspondant aux régimes de déflagration et de détonation.

3- Projet REPLACE pour le développement de méthodes alternatives

Depuis de nombreuses années, l'INERIS [41], de part son savoir-faire en sécurité des procédés et dans la conduite d'essais expérimentaux à petite et grande échelle, apporte son expertise pour la prévention des risques accidentels liés à l'explosibilité des substances chimiques [32], en tant que laboratoire agréé pour diverses tâches relatives aux produits explosifs et autorité compétente technique en France en matière de classement en transport des matières et objets explosibles (classe 1) . L'Institut est également amené à prendre part à l'élaboration des règles internationales (méthodes d'évaluation, procédures de classement…) [42], telles que les recommandations de l'ONU pour le transport des marchandises dangereuses ou le nouveau règlement européen CLP [24]. A ce titre, il est tout particulièrement sensible au développement des outils nécessaires à la mise en place des nouvelles approches règlementaires. Les outils prédictifs recommandés par le nouveau contexte règlementaire en sont un exemple. Un programme de recherche, intitulé REPLACE pour Recherche sur les Propriétés et l'Activité de Composés Explosifs, a donc été initié en 2007 en collaboration avec l'équipe de Modélisation des Systèmes Complexes de Chimie Paristech. Son objectif est de développer de nouvelles méthodes prédictives (alternatives ou

complémentaires à l'expérimentation) pour l'évaluation des propriétés dangereuses à caractère physico-chimique.

Cette recherche est assez nouvelle à l'INERIS, puisque seuls quelques travaux avaient été initiés dans les années 1990 pour la prédiction de la sensibilité à l'impact [57]. Elle vise à faire évoluer les méthodes d'évaluation et proposer, dans un premier temps, un nouvel outil prédictif pour l'évaluation de propriétés reconnues comme complexes, telles que l'explosibilité, en utilisant des outils de chimie quantique et les approches de type QSPR. L'idée est d'obtenir un outil qui permettrait en partant de la seule structure moléculaire d'accéder à une première évaluation en termes de propriétés explosives à l'aide de modèles QSPR (capables de prédire les différentes propriétés nécessaires à la caractérisation de l'explosibilité des substances) et d'un outil d'aide à la décision, permettant d'analyser les propriétés obtenues dans le cadre règlementaire, pour une variété de composés la plus grande possible (cf. Figure 10).

Les présents travaux de thèse initient ce projet. Il s'agit donc de développer et d'évaluer les méthodes alternatives à l'expérimentation de type QSPR capables de prédire cette propriété dangereuse d'explosibilité, de manière suffisamment efficace en vue d'une approche de type screening dans le cadre de REACH.

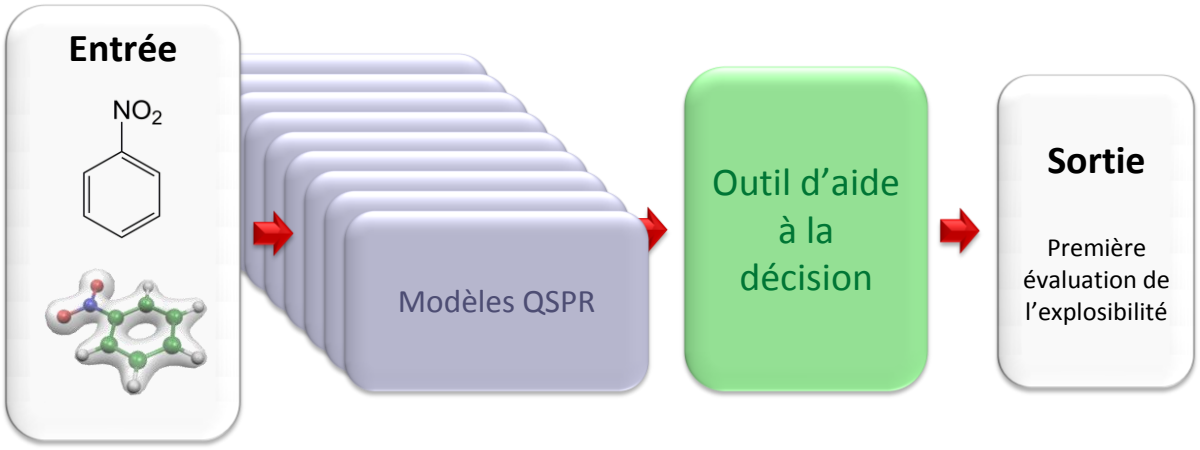


Figure 10 - Outil de prédiction des propriétés d'explosibilité

Les efforts se portent donc, dans un premier temps, sur une catégorie particulière de composés, les composés nitroaromatiques, qui représentent une classe de substances explosives [36] et dont le représentant le plus connu est le 2,4,6-trinitrotoluène (TNT), utilisé dans de nombreuses compositions pyrotechniques. Ces composés sont particulièrement complexes du point de vue de leur décomposition [58], ce qui rend leur caractérisation particulièrement intéressante.

Une problématique majeure de l'approche employée réside dans l'obtention de données expérimentales fiables et suffisamment nombreuses pour mettre en place les modèles. En fonction de la nature des données extraites des tests, quantitatives (ex : sensibilités) ou qualitatives (ex : test de Koenen), l'approche prédictive employée est différente.

Une propriété quantitative, la stabilité thermique (employée en phase de présélection) sera tout d'abord envisagée afin d'évaluer les différents outils méthodologiques disponibles et de mettre au point une démarche de développement des modèles. En effet, des données quantitatives permettent de réaliser des analyses à la fois qualitatives et quantitatives alors que des données qualitatives n'auraient pas permis d'analyses quantitatives. L'étude sera ensuite étendue à de nouvelles propriétés d'explosibilité : les sensibilités à la décharge électrique et à l'impact, deux propriétés pour lesquelles des données expérimentales ont pu être extraites de la littérature.

CHAPITRE 2

BASES THEORIQUES

Sommaire

1- Chimie quantique 46
1.1- Bases de la chimie quantique 46
1.2- Méthode Hartree-Fock 49
1.3- Méthodes semi-empiriques 52
1.4- Théorie de la Fonctionnelle de la Densité 54
2- Quantitative Structures-Property Relationships 57
2.1- Principe 58
2.2- Importance des bases de données 59
2.3- Descripteurs moléculaires 61
2.4- Méthodes d'analyse de données 68

L'utilisation de méthodes prédictives est d'un grand intérêt dans le cadre de REACH. Dans ce travail, une méthodologie combinant les approches de chimie quantique et les outils QSPR (*Quantitative Structure-Property Relationship*) a été employée. Le présent chapitre présente les principales théories et techniques utilisées au cours de ces travaux.

Dans une première partie, il s'agira d'expliciter les méthodes de chimie quantique, utilisées non seulement pour le calcul des structures moléculaires nécessaires à la mise en place de modèles prédictifs mais aussi pour les études de réactivité.

Les outils employés dans la méthodologie QSPR seront ensuite explicités. Il s'agira de traiter du principe général de cette méthode puis de donner des détails supplémentaires concernant les descripteurs de la structure moléculaire ainsi que les outils d'analyse de données nécessaires à la mise en place et à la validation des modèles prédictifs.

1- Chimie quantique

Différentes approches sont envisageables dans le cadre des outils de modélisation moléculaire. Si ceux de mécanique classique, économiques en termes de temps de calcul, permettent de traiter des systèmes moléculaires de grande taille, les méthodes quantiques (*ab initio*, semi-empiriques ou théorie de la fonctionnelle de la densité) sont, quant à elles, capables de calculer les propriétés électroniques des systèmes. C'est pour cette raison que ces approches ont été employées dans le cadre de cette étude. Des compléments sur les méthodes décrites ici peuvent être trouvés dans différents ouvrages de références [59-65].

1.1- Bases de la chimie quantique

1.1.1- Equation de Schrödinger

Les méthodes de chimie quantique, présentées par la suite, reposent toutes sur l'équation de Schrödinger, non relativiste, indépendante du temps.

$$\hat{H}\Psi = E\Psi \tag{2.1}$$

où Ψ est la fonction d'onde décrivant le système de noyaux et d'électrons, $\hat{H}$ est l'opérateur Hamiltonien relatif à ce même système et E est l'énergie correspondante, valeur propre de l'équation.

L'Hamiltonien rend compte des différentes contributions à l'énergie totale du système à partir d'opérateurs pour les énergies cinétiques des électrons et des noyaux ainsi que les interactions noyau-électron, électron-électron et noyau-noyau.

$$\hat{H} = -\frac{1}{2}\sum_i \nabla_i^2 - \sum_k \frac{1}{2}\nabla_k^2 - \sum_i \sum_k \frac{Z_k}{r_{ik}} + \sum_i \sum_{i<j} \frac{1}{r_{ij}} + \sum_k \sum_{k<l} \frac{Z_k Z_l}{r_{kl}} \tag{2.2}$$

où ∇^2 est l'opérateur d'énergie cinétique, Z_k est le numéro atomique de l'atome k, r_{ik} est la distance entre un électron i et un noyau k, r_{ij} est la distance entre deux électrons i et j et r_{kl} est la distance entre deux noyaux k et l.

Il n'est pas possible de résoudre cette équation pour des systèmes d'intérêt chimique (au-delà de H_2), de manière exacte. Il faut alors introduire différentes approximations.

1.1.2- Approximation de Born-Oppenheimer

L'approximation de Born-Oppenheimer permet une première simplification de l'Hamiltonien. Etant donnée la masse des noyaux, des milliers de fois plus grande que celle des électrons, leur mouvement est négligeable devant celui des électrons. En première approximation, il est donc possible de considérer les noyaux fixes. Il s'en suit donc que leur énergie cinétique est négligeable devant celle des électrons et que l'énergie d'interaction entre noyaux est constante.

Ce qu'il reste de $\hat{H}$ peut alors se décomposer en deux contributions distinctes : l'une électronique et l'autre nucléaire, cette dernière correspondant à la répulsion (constante) entre les noyaux. On peut donc finalement considérer de manière totalement indépendante l'Hamiltonien électronique du système.

$$\hat{H}_{elec} = -\sum_i \frac{1}{2}\nabla_i^2 - \sum_i \sum_k \frac{Z_k}{r_{ik}} + \sum_{i<j} \frac{1}{r_{ij}} \tag{2.3}$$

Il s'agit alors de résoudre l'équation de Schrödinger pour le système d'électrons en considérant une fonction d'onde électronique qui dépend des coordonnées électroniques et nucléaires (ces dernières sont alors des paramètres du problème et non plus des variables). L'énergie totale du système (pour une configuration nucléaire donnée) sera alors obtenue en ajoutant à l'énergie électronique E_{elec} du système la répulsion nucléaire.

$$E_{tot} = E_{elec} + \sum_k \sum_{k<l} \frac{Z_k Z_l}{r_{kl}} \tag{2.4}$$

Cette approximation est sous-jacente dans toutes les méthodes qui sont présentées par la suite.

1.1.3- Approximation orbitalaire et fonctions de base

La fonction d'onde multi-électronique d'un système peut être décomposée sous la forme d'un produit de fonctions d'onde mono-électroniques φ_i, appelées orbitales.

$$\Psi_{elec} = \varphi_1(\vec{r}_1)\varphi_2(\vec{r}_2)...\varphi_n(\vec{r}_n) \tag{2.5}$$

où $\vec{r}_i$ est le vecteur position de l'électron i.

En l'absence de couplage spin-orbite, les coordonnées d'espace et de spin peuvent être découplées. On parle alors de spinorbitales ϕ_i, produits d'une fonction d'espace ψ_i et d'une fonction de spin $\alpha(\omega)$ ou $\beta(\omega)$, selon le spin de la particule.

$$\phi_i = \begin{cases} \psi_i(\vec{r}_i)\alpha(\omega) \\ \psi_i(\vec{r}_i)\beta(\omega) \end{cases} \tag{2.6}$$

Telle quelle, la fonction d'onde ne satisfait pas au principe d'exclusion de Pauli qui propose que seules sont physiquement significatives les fonctions d'onde antisymétriques par échange d'une paire d'électrons. La fonction d'onde peut s'écrire sous la forme d'un déterminant dit de Slater, qui respecte ce principe d'antisymétrie.

$$\Psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\vec{r}_1) & \phi_2(\vec{r}_1) & \cdots & \phi_N(\vec{r}_1) \\ \phi_1(\vec{r}_2) & \phi_2(\vec{r}_2) & \cdots & \phi_N(\vec{r}_2) \\ \vdots & \vdots & & \vdots \\ \phi_1(\vec{r}_N) & \phi_2(\vec{r}_N) & \cdots & \phi_N(\vec{r}_N) \end{vmatrix} \tag{2.7}$$

Chaque orbitale moléculaire (OM) peut ensuite être exprimée sous la forme d'une combinaison linéaire d'orbitales atomiques (OA). Il s'agit de l'approche LCAO (*Linear*

Combinations of Atomic Orbitals). Les OA pouvant elles-mêmes être représentées par une combinaison linéaire de fonctions de bases, toute orbitale OA ou OM peut être décrite sous la forme suivante.

$$\psi = \sum_i c_i \chi_i \tag{2.8}$$

où χ_i est une fonction de base à laquelle est associé le coefficient de développement c_i.

Il s'agit d'une approximation puisque les orbitales sont représentées sur un jeu de fonctions de base fini alors qu'une formulation exacte des orbitales nécessiterait une infinité de fonctions de base. Ainsi, le nombre et la nature des fonctions de base employées est un paramètre critique des calculs de chimie quantique.

On peut distinguer deux types de fonctions de base selon la forme mathématique employée : les fonctions de Slater (STO) et les fonctions gaussiennes (GTO). Pour une fonction 1s normalisée pour l'atome d'hydrogène, par exemple, elles ont les formes suivantes.

$$\chi_{1s}^{STO}(r) = \left(\zeta^3/\pi\right)^{1/2} \exp\left(-\zeta r\right) \tag{2.9}$$

$$\chi_{1s}^{GTO}(r) = \left(2\alpha/\pi\right)^{3/4} \exp\left(-\alpha r^2\right) \tag{2.10}$$

où ζ et α sont les exposants respectifs de la fonction de Stater et de la fonction gaussienne.

L'intérêt majeur des gaussiennes réside dans le fait qu'elles facilitent le calcul des intégrales, le produit de deux gaussiennes étant une nouvelle gaussienne. Cependant, elles décrivent mal les orbitales atomiques au voisinage du noyau, au contraire des fonctions de type Slater. Aussi utilise-t-on des combinaisons linéaires de fonctions gaussiennes (dites primitives) afin de reproduire le comportement des fonctions de Slater au voisinage du noyau (cf. Figure 11). On parle alors de fonctions gaussiennes contractées.

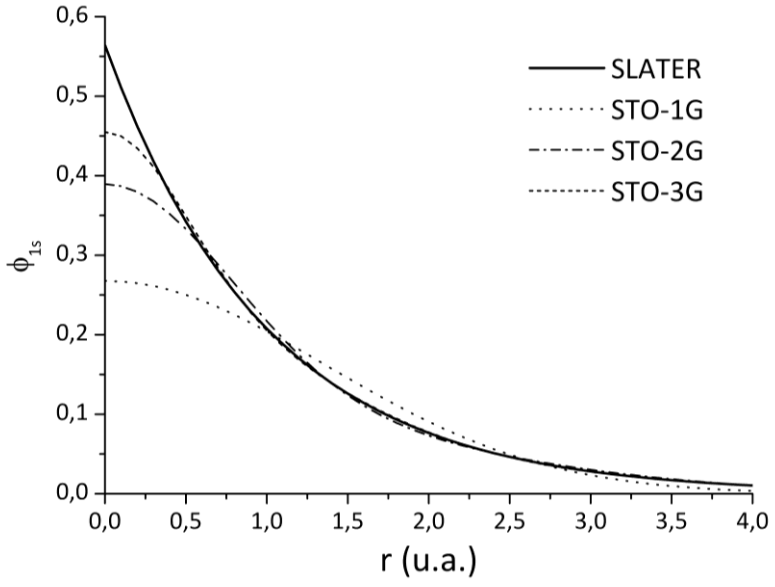


Figure 11 – Comparaison de la représentation d'une fonction de Slater 1s à partir de gaussiennes contractées STO-1G, STO-2G et STO-3G

Par exemple, les bases de fonctions STO-MG [66] ont été définies pour reproduire le comportement de fonctions de Slater, comme montré en Figure 11.

$$\chi^{STO-MG} = \sum_{i=1}^{M} c_i \chi_i^{GTO} \tag{2.11}$$

Afin d'améliorer la représentation du système électronique, le jeu des fonctions de base peut être accru. On peut, par exemple, augmenter le nombre de STO pour représenter chaque orbitale. On parle alors de jeux de fonctions de base, ou plus simplement bases, simple, double, triple zêta, etc.

Ensuite, des fonctions supplémentaires peuvent encore être ajoutées. Les fonctions de polarisation ont pour but de donner de la flexibilité angulaire aux orbitales de valence. De telles fonctions sont notamment utiles pour la description des systèmes polarisés. Les fonctions diffuses, quant à elles, sont des fonctions gaussiennes caractérisées par des exposants très faibles. Elles sont en particulier nécessaires pour la description correcte des interactions intermoléculaires.

Finalement, il s'agit de choisir le jeu de fonctions de base garantissant une représentation convenable des orbitales moléculaires pour des temps de calculs raisonnables.

1.1.4- Principe variationnel

Le principe variationnel est utilisé afin d'optimiser les fonctions d'ondes approchées. En effet, l'énergie du système dans son état fondamental E_0 est une borne inférieure pour la valeur moyenne de l'énergie E, calculée à partir de la fonction d'onde approchée Ψ.

$$E = \langle \Psi | \hat{H} | \Psi \rangle \geq E_0, \quad \langle \Psi | \Psi \rangle = 1 \tag{2.12}$$

Aussi, la qualité d'une fonction d'onde peut être caractérisée par son énergie. Plus basse est l'énergie et meilleure est la fonction d'onde.

Dans le formalisme LCAO, la fonction d'onde est caractérisée par une série de coefficients c_i qu'il s'agit alors de choisir de sorte qu'ils minimisent l'énergie. Aussi, la fonction d'onde est-elle optimisée de sorte que la dérivée de l'énergie suivant chaque paramètre soit nulle.

$$\frac{\partial E}{\partial c_i} = 0 \tag{2.13}$$

1.2- Méthode Hartree-Fock

Une approche quantique est la méthode Hartree-Fock (HF). Il s'agit de l'approche de Roothaan, une approche variationnelle. Il s'agit donc de trouver les spin-orbitales minimisant l'énergie électronique tout en conservant leur orthogonalité (cf. Eq. 2.12). Une approche itérative dite de champ auto-cohérent (*Self Consistent Fields,* SCF) est donc utilisée.

1.2.1- Equations Hartree-Fock

Pour un système moléculaire, au sein duquel les électrons sont appariés (on parle alors de système à couches fermées), l'énergie HF se décompose en une somme de termes mono et bi-électroniques.

$$E^{HF} = 2\sum_i h_{ii} + \sum_i \sum_{j>i} \left(2J_{ij} - K_{ij}\right) \tag{2.14}$$

avec $$h_{ii} = \int \psi_i^*(1)\hat{h}_1\psi_i(1)d\vec{r} \tag{2.15}$$

$$J_{ij} = \int \psi_i^*(1)\psi_j^*(2)r_{12}^{-1}\psi_i(1)\psi_j(2)d\vec{r}_1 d\vec{r}_2 \tag{2.16}$$

$$K_{ij} = \int \psi_i^*(1)\psi_j^*(2)r_{12}^{-1}\psi_j(1)\psi_i(2)d\vec{r}_1 d\vec{r}_2 \tag{2.17}$$

où $$\hat{h}_1 = -\frac{1}{2}\nabla_1^2 - \sum_A \frac{Z_A}{r_{1A}} \tag{2.18}$$

Dans cette expression, J_{ij} et K_{ij} sont respectivement des intégrales de Coulomb et d'échange, qui caractérisent les répulsions entre électrons. Les intégrales d'échange résultent de la nature antisymétrique de la fonction d'onde multiélectronique.

L'interaction de chaque électron avec les noyaux et les autres électrons environnants est introduite via l'approximation du champ moyen qui considère que chaque électron subit un potentiel V^{eff} formé par les noyaux et le champ moyen des autres électrons.

A chaque électron est alors associée une équation mono-électronique similaire à l'équation de Schrödinger. Ce sont les équations Hartree-Fock.

$$\hat{F}\psi_i = \varepsilon_i\psi_i \tag{2.19}$$

où ε_i sont les énergies des spin-orbitales.

Ces dernières sont les valeurs propres de l'opérateur Hamiltonien mono-électronique $\hat{F}$, appelé opérateur de Fock, dont l'expression pour l'électron 1 est la suivante :

$$\hat{F}(1) = \hat{h}_1(1) + V^{eff}(1) = \hat{h}_1(1) + \sum_j \left(2\hat{J}_j(1) - \hat{K}_j(1)\right) \tag{2.20}$$

dans laquelle les opérateurs d'échange $\hat{J}_j$ et de Coulomb $\hat{K}_j$ ont les expressions suivantes :

$$\hat{J}_j(1)\psi_i(1) = \int \psi_j^*(2)r_{12}^{-1}\psi_j(2)d\vec{r}_2\psi_i(1) \tag{2.21}$$

$$\hat{K}_j(1)\psi_i(1) = \int \psi_j^*(2)r_{12}^{-1}\psi_i(2)d\vec{r}_2\psi_j(1) \tag{2.22}$$

1.2.2- Equations de Roothaan-Hall

La résolution des équations Hartree-Fock, sous forme matricielle, a été proposée par Roothaan et Hall au début des années 1950 [67].

En effet, dans le cadre de l'approximation LCAO, les équations Hartree-Fock peuvent être réécrites sous la forme suivante (pour l'électron 1) :

$$\hat{F}(1)\sum_{\nu} c_{\nu i}\chi_{\nu}(1) = \varepsilon_i \sum_{\nu} c_{\nu i}\chi_{\nu}(1) \tag{2.23}$$

Multiplier à gauche chaque terme de cette équation par φ_{μ}^{*} permet de transformer l'équation précédente sous une forme matricielle.

$$\sum_{\nu} F_{\mu\nu}C_{\nu i} = \varepsilon_i \sum_{\nu} S_{\mu\nu}C_{\nu i} \tag{2.24}$$

avec $$F_{\mu\nu} = \int \chi_{\mu}^{*}(1)\hat{F}(1)\chi_{\nu}(1)d\vec{r}_1 \tag{2.25}$$

$$S_{\mu\nu} = \int \chi_{\mu}^{*}(1)\chi_{\nu}(1)d\vec{r}_1 \tag{2.26}$$

Les matrices F et S ainsi définies sont respectivement les matrices de Fock et de recouvrement. Ces équations, appelées équations de Roothaan-Hall, peuvent s'exprimer sous la forme simplifiée suivante.

$$FC = SC\varepsilon \tag{2.27}$$

Par orthogonalisation des OM, le problème peut être ramené à la résolution de l'équation $FC=CE$. Or, l'opérateur de Fock dépend des spinorbitales, et donc des solutions de l'équation. Aussi, la résolution du problème passe nécessairement par un processus itératif dit du champ auto-cohérent (SCF, pour *self consistent field*).

A partir d'une géométrie donnée (et en utilisant une base donnée), un jeu d'orbitales moléculaires initial est établi. Ce jeu initial peut être obtenu, par exemple, par calcul semi-empirique de type Hückel étendu. La valeur du potentiel HF est alors déterminée, donnant ainsi accès à l'opérateur de Fock. La résolution des équations aux valeurs propres mène ensuite aux énergies E_i ainsi qu'à un nouveau jeu d'orbitales moléculaires. A partir de ce nouveau jeu d'orbitales, un nouveau cycle peut alors débuter.

La procédure itérative prend fin lorsque que la variation des énergies devient inférieure à une certaine limite, le critère de convergence de la procédure SCF.

1.2.3- Cas des systèmes à couches ouvertes

Jusqu'à présent, le système considéré était composé d'électrons appariés (système à couches fermées). Les électrons de spins opposés occupent alors par paires les fonctions d'espace des spin-orbitales. On parle de fonction d'onde avec contrainte de spin (RHF, pour *restricted Hartree-Fock*).

Dans le cadre des systèmes à couches ouvertes, deux approches peuvent être employées : avec contrainte de spin (ROHF, pour *restricted open-shell Hartree-Fock*) ou sans contrainte de spin (UHF pour *unrestricted Hartree-Fock*).

Dans l'approche ROHF [68], une partie des fonctions d'espace correspondent à des couches fermées, toutes occupées par deux électrons, et d'autres à des couches ouvertes qui ne sont occupées que par un seul électron α. L'utilisation de cette méthode pose des problèmes notamment dans l'étude des phénomènes de dissociation.

Dans l'approche UHF, aucune contrainte de spin n'est appliquée et les électrons α et β occupent des fonctions d'espace différentes. Deux jeux d'orbitales moléculaires distincts sont alors développés et les équations HF sont exprimées de manière couplées. Les équations de Berthier-Pople-Nesbet [69,70] ainsi obtenues sont résolues de manière itérative. Le défaut de la méthode réside dans le fait que la fonction d'onde multiélectronique n'est plus fonction propre de l'opérateur de spin S^2 et l'état de spin du déterminant de Slater correspond alors à un mélange avec les états de spin supérieurs. On parle de contamination de spin. Cette dernière peut être évaluée par la différence entre la valeur moyenne de S^2 évaluée à partir du déterminant UHF et celle correspondant à un Hamiltonien non contaminé. Si cette différence est significative, alors la fonction d'onde UHF n'a pas de sens.

1.2.4- Limites de la méthode Hartree-Fock

Dans la plupart des cas, la méthode HF donne des résultats satisfaisants. Malgré tout, certaines limitations ont mené au développement de nouvelles méthodes. Le principal problème posé par cette approche découle du fait que la corrélation existant entre les mouvements des électrons n'est pas prise en compte.

L'énergie de corrélation E_{corr} est ainsi définie comme la différence entre l'énergie exacte E_{exacte} et l'énergie calculée en HF pour une base complète E_{HF}.

$$E_{corr} = E_{exacte} - E_{HF} < 0 \qquad (2.28)$$

D'autres méthodes ont été développées sur des approximations différentes et permettent ainsi la prise en compte de cette corrélation électronique. C'est le cas des méthodes post-HF (méthodes multiconfigurationnelles, perturbatives ou encore d'agrégats couplés (*coupled clusters*)) [65], qui nécessitent des temps de calculs plus importants, ou la Théorie de la Fonctionnelle de la densité (DFT), basée sur une approche différente, qui sera décrite par la suite.

1.3- Méthodes semi-empiriques

Les calculs semi-empiriques sont eux développés sur la même structure générale que les calculs HF, mais certaines parties de l'information sont sujettes à approximation ou même complètement omises, afin de les rendre moins exigeants en termes de temps de calcul.

En particulier, dans le cadre de ces approches, trois types de simplifications sont principalement réalisés :

- non-considération des électrons de cœur, puisqu'ils ne contribuent pas à la réactivité chimique du système. Ils seront alors considérés avec le noyau au sein d'une fonction paramétrée (ex : méthode de Hückel étendu [71-74]) ;
- utilisation d'un jeu réduit de fonctions de base (typiquement STO-3G) ;
- réduction du nombre d'intégrales bi-électroniques grâce à l'introduction de paramètres empiriques.

En effet, l'étape la plus exigeante en termes de temps de calcul dans la résolution des équations HF réside dans le traitement des intégrales bi-électroniques de la matrice de Fock. Pour une base de dimension n, n^4 intégrales bi-électroniques doivent être traitées. Aussi, le traitement de systèmes de taille importante nécessite des temps de calcul eux aussi importants. Afin de pouvoir traiter des systèmes moléculaires de plus grande taille, les méthodes semi-empiriques introduisent des approximations supplémentaires dans les équations HF.

Il s'agit de négliger certains recouvrements orbitalaires et d'estimer les autres à partir de considérations empiriques. L'approche CNDO (*Complete Neglect of Differential Overlap*), par exemple, néglige tous les recouvrements [75,76]. L'approche NDDO (*Neglect of Diatomic Differential Overlap*) néglige, quant à elle, uniquement les recouvrements mettant en jeu des orbitales centrées sur des noyaux différents [75,76].

L'avantage des calculs semi-empiriques est qu'ils sont donc plus rapides que les autres méthodes quantiques. Leur inconvénient réside dans le fait qu'elles sont soumises à de nombreuses approximations. En fait, peu de propriétés peuvent être prédites de manière fiable, en particulier, pour des molécules de structures trop éloignées de celles utilisées pour la paramétrisation des méthodes. En général, celle-ci est réalisée afin de reproduire différents propriétés : géométries, énergies de formation, ou encore énergies de réaction, moments dipolaires, potentiels d'ionisation voire des propriétés spécifiques telles que des spectres électroniques ou des déplacements chimiques RMN.

L'une des méthodes semi-empiriques les plus utilisées est le modèle AM1 (*Austin Model 1*) [77]. Cette approche emploie un schéma de type NDDO dans lequel les recouvrements des intégrales bi-électroniques mono-centrées sont paramétrés sur des données spectroscopiques pour des atomes isolés, les autres considérant des interactions entre multipôles. Si cette méthode est en particulier largement utilisée pour les composés organiques, elle présente quelques limitations reconnues dans l'estimation des énergies d'activation, stabilité de certains composés ou enthalpies de liaison [64].

1.4- Théorie de la Fonctionnelle de la Densité

Jusqu'à présent, les approches présentées, Hartree-Fock ou semi-empiriques, sont toutes fondées autour d'une fonction mathématique : la fonction d'onde. Même si cette dernière peut être reliée à l'énergie de la molécule, il n'en reste pas moins que cette grandeur n'a pas en soit de signification physique.

La théorie de la fonctionnelle de la densité (DFT) diffère des approches précédentes en prenant pour propriété fondamentale la densité électronique, qui est, quant à elle, une observable. Cette approche a pour origine le postulat de Thomas et Fermi selon lequel l'énergie du système, dans son état fondamental, peut être décrite sous la forme d'une fonctionnelle de la densité électronique $\rho(\vec{r})$ qui peut, elle-même, être reliée à la fonction d'onde Ψ.

$$E_0 = E[\rho] \tag{2.29}$$

avec $$\rho(\vec{r}) = \Psi^*(\vec{r})\Psi(\vec{r}) = |\Psi^2(\vec{r})| \tag{2.30}$$

Si la DFT diffère conceptuellement des précédentes approches par la considération de la densité électronique, elle n'en reste pas moins appuyée sur des approximations similaires.

1.4.1. Théorèmes de Hohenberg-Kohn

Si l'utilisation de la densité électronique comme variable de base a été envisagée par Thomas et Fermi dès 1927 puis Slater dans les années 1950 [78], le véritable formalisme de la DFT naît des théorèmes de Hohenberg et Kohn dans les années 1960 [79].

Premier théorème : théorème d'existence

Le premier théorème démontre que toute propriété électronique du système à l'état fondamental, et donc son énergie totale, est une fonctionnelle de la seule densité électronique.

$$E[\rho] = T[\rho] + V_{ee}[\rho] + V_{eN}[\rho] \tag{2.31}$$

où T est l'énergie cinétique des électrons, V_{ee} et V_{eN} sont respectivement les énergies potentielles de répulsion électron-électron et d'attraction noyau-électron.

Ce dernier terme peut être caractérisé via un potentiel externe v_{ext} qui contient les interactions avec le noyau. Une fonctionnelle $F_{HK}[\rho]$, dite universelle de ρ, est alors constituée des termes indépendants de v_{ext}.

$$E[\rho] = F_{HK}[\rho] + \int v_{ext}(\vec{r})\rho(\vec{r})d\vec{r} \tag{2.32}$$

La fonctionnelle universelle $F_{HK}[\rho]$ regroupe donc l'énergie cinétique des électrons T et la répulsion inter-électronique V_{ee}, pour tous les systèmes électroniques.

Second théorème : principe variationnel

Le second théorème applique principe variationnel à la DFT. Ainsi, l'énergie, fonctionnelle d'une densité électronique approchée, est-elle supérieure ou égale à l'énergie exacte du système dans son état fondamental.

$$F_{HK}[\rho]+V_{eN}[\rho]=E[\rho]\geq E[\rho_0]=E_0 \tag{2.33}$$

Si déterminer la densité ρ_0 pour laquelle l'énergie est minimale à partir de la fonction d'onde correspondante Ψ se fait plutôt aisément via l'équation (2.30), il existe par contre une infinité de fonctions d'onde pouvant être associées à une même densité. Dès lors, obtenir la fonction d'onde de l'état fondamental à partir de la densité ρ_0 n'est pas chose évidente. Pour ce faire, une méthode sous contrainte a été développée par Levy et Lieb [80] permettant d'isoler la fonction d'onde Ψ.

1.4.2. Approche Kohn-Sham

Malgré les théorèmes de Hohenberg-Kohn qui démontrent la faisabilité de la méthode, il n'en reste pas moins une difficulté majeure à surmonter, le traitement de l'énergie cinétique des électrons, et il faut attendre l'approche de Kohn-Sham en 1965 [81] pour voir émerger une approche DFT utilisable.

Celle-ci repose sur l'introduction d'un système fictif de référence à n particules, sans interaction, ayant la même densité électronique que le système réel. On peut alors écrire un Hamiltonien mono-électronique du système :

$$\hat{H}_{KS}=-\frac{1}{2}\nabla^2+\int\frac{\rho(\vec{r})-\rho(\vec{r}')}{|\vec{r}-\vec{r}'|}d\vec{r}'+\upsilon_{xc}[\rho]+\upsilon_{ext}(\vec{r}) \tag{2.34}$$

Dans cette équation, $\hat{H}_{KS}$ est l'hamiltonien de Kohn-Sham.

Le potentiel d'échange-corrélation du système υ_{xc} est directement lié à l'énergie d'échange-corrélation E_{xc} qui sera détaillée ultérieurement.

$$\upsilon_{xc}[\rho]=\frac{\partial E_{xc}[\rho(\vec{r})]}{\partial\rho(\vec{r})} \tag{2.35}$$

Un jeu d'orbitales dites de Kohn-Sham est alors introduit permettant la minimisation de l'énergie du système par résolution auto-cohérente des équations aux valeurs propres :

$$\hat{H}_{KS}\psi_i=\varepsilon_i\psi_i \tag{2.36}$$

Finalement, les théorèmes d'Hohenberg et Kohn imposent que l'énergie totale du système alors calculée (E_{KS}) soit égale à celle du système réel (Eq. 3.32).

$$E_{KS}[\rho]=T_s[\rho]+J[\rho]+E_{xc}[\rho]+\int V_{ext}(\vec{r})\rho(\vec{r})d\vec{r} \tag{2.37}$$

Dans ces équations, $T_s[\rho]$ est l'énergie cinétique du système sans interaction de référence, $J[\rho]$ l'énergie d'interaction coulombienne entre les électrons, $E_{xc}[\rho]$ l'énergie d'échange corrélation et $\int V_{ext}(\vec{r})\rho(\vec{r})d\vec{r}$ l'énergie d'interaction avec le potentiel externe.

Il s'ensuit que le terme d'échange-corrélation peut s'exprimer de la manière suivante :

$$E_{xc}[\rho] = (T_{réel}[\rho] - T_s[\rho]) + (V_{ee}[\rho] - J[\rho]) = \Delta T + \Delta V_{ee} \quad (2.38)$$

où ΔT et ΔV_{ee} sont respectivement les corrections à l'énergie cinétique dues aux interactions inter-électroniques et aux corrections non-classiques issues de la répulsion inter-électronique. Ce dernier terme inclut donc en son sein les différents effets quantiques d'échange et de corrélation.

1.4.3. Fonctionnelles d'échange-corrélation

Finalement, toute la difficulté de la DFT relève du terme d'échange-corrélation dont l'expression exacte reste inconnue. Aussi est-il calculé à partir de fonctionnelles d'échange-corrélation approchées qui se distinguent par les approximations qui les définissent.

Approximation de la densité locale (LDA)

L'approximation la plus simple est la LDA (*Local Density Approximation*). Elle consiste à considérer la densité électronique comme localement uniforme. L'énergie d'échange corrélation en un point donné est alors fonction de la seule densité électronique en ce point.

$$E_{xc}^{LDA}[\rho(\vec{r})] = \int \rho(\vec{r})\varepsilon_{xc}[\rho(\vec{r})]d\vec{r} \quad (2.39)$$

où ε_{xc} est l'énergie d'échange-corrélation par électron.

Si cette approximation ne prend pas en compte les fluctuations locales de densité, elle reste très utilisée pour l'état solide en particulier. En effet, elle fournit d'assez bonnes informations sur les propriétés géométriques et vibrationnelles mais est généralement défaillante pour ce qui est des données énergétiques [82].

Approximation du gradient généralisé (GGA)

L'approximation GGA (*Generalized Gradient Approximation*) franchit un échelon vis-à-vis de la LDA en considérant non seulement la densité mais aussi son gradient $\nabla\rho$.

$$E_{xc}^{GGA}[\rho(\vec{r})] = \int \rho(\vec{r})\varepsilon_{xc}[\rho(\vec{r}), \nabla\rho(\vec{r})]d\vec{r} \quad (2.40)$$

Les contributions d'échange et de corrélation sont communément traitées de manière distincte, pour être par la suite combinée de façon à donner la fonctionnelle complète. Les différents termes peuvent être obtenus de deux manières différentes en les paramétrant soit sur des données expérimentales, comme dans la fonctionnelle BP86 (échange de Becke version 1988 [83] et corrélation de Perdew version 1986 [84]), soit sur des considérations physiques, comme, pour la fonctionnelle PBE (Perdew, Burke, Ernzherhof [85,86]). Elles représentent une amélioration vis-à-vis des fonctionnelles LDA au niveau de plusieurs propriétés moléculaires telles que des longueurs et énergies de liaisons.

Fonctionnelles hybrides

Une troisième catégorie de fonctionnelles est constituée par les fonctionnelles hybrides. En fait, il s'agit d'introduire une fraction d'échange (exact) Hartree-Fock.

La fonctionnelle hybride la plus couramment utilisée dans la littérature est B3LYP. Cette fonctionnelle à 3 paramètres repose sur les termes d'échange de Becke [83] et de corrélation de Lee, Yang et Parr [87].

$$E_{xc}^{B3LYP} = E_{xc}^{LDA} + a_0\left(E_x^{HF} - E_x^{LDA}\right) + a_x\left(E_x^{GGA} - E_x^{LDA}\right) + a_c\left(E_c^{GGA} - E_c^{LDA}\right) \quad (2.41)$$

où les paramètres a_0, a_x et a_c ont été ajustés respectivement à 0,20, 0,72 et 0,81 pour une série de petites molécules [88].

Dans nos travaux, la fonctionnelle PBE0 [89] a été employée. Cette fonctionnelle peut être considérée comme non-paramétrée, le pourcentage d'échange Hartree-Fock (25%) n'ayant pas été ajusté sur des critères empiriques mais fixé sur des considérations physiques.

$$E_{xc}^{PBE0} = E_{xc}^{PBE} + \frac{1}{4}\left(E_x^{HF} - E_x^{PBE}\right) \quad (2.42)$$

Les fonctionnelles hybrides sont très couramment employées pour leur bonne précision dans une gamme importante de propriétés moléculaires.

Alors que les fonctionnelles de la densité présentées auparavant avaient tendance à entraîner des surestimations de certaines grandeurs, la théorie HF les sous-estime. C'est, par exemple, le cas pour les longueurs de liaisons [90]. Une combinaison de ces deux approches propose des résultats plus fiables.

2- Quantitative Structures-Property Relationships

Les relations quantitatives structure-activité/propriété (QSAR/QSPR) sont de plus en plus utilisées, du fait de la croissance des moyens de calculs. Très récemment, la mise en place du nouveau règlement européen REACH, qui recommande leur utilisation pour limiter le recours à l'expérience, donne un nouvel essor au développement de tels modèles prédictifs.

En fait, les premiers développements dans le sens de telles méthodologies sont plutôt anciens. Dès 1868, Crum-Brown et Fraser ont postulé l'existence de relations entre les activités physiologiques et les structures chimiques [91] en reliant les changements d'activité biologique à des modifications structurales simples, ne disposant alors pas de moyen pour caractériser les structures chimiques en termes quantitatifs.

Une avancée importante vers les modèles QSPR proprement dits a été réalisée grâce au développement des équations de Hammett [92-94] dans lesquelles les constantes du même nom σ caractérisent de manière quantitative les vitesses de réactions pour les composés organiques.

$$\log \frac{K}{K_0} = \sigma\rho \tag{2.43}$$

où K et K_0 sont les constantes respectives de la réaction étudiée et de celle d'une référence et ρ une constante de réaction dépendant du type de réaction.

Cela dit, les premiers travaux utilisant la méthodologie telle qu'employée à l'heure actuelle sont à attribuer aux contributions de Hansch [95] d'une part et de Free et Wilson [96] d'autre part.

D'un côté, Hansch a proposé des modèles reliant directement l'activité biologique des composés avec les propriétés hydrophobes, électroniques et stériques à l'échelle moléculaire. De l'autre, Free et Wilson ont développé des modèles dits de contributions de groupes pour l'activité biologique dans lesquels l'identification de groupements particuliers renvoie à une valeur tabulée (empirique). La valeur finale est donc une combinaison de ces différentes valeurs correspondant aux différents groupes identifiés au sein de la molécule.

Dans les dernières années, l'utilisation des méthodes QSAR n'a cessé de progresser. Elle est même devenue indispensable en chimie pharmaceutique et pour la conception de médicaments [8,9]. Leur développement dans une gamme plus large d'applications leur ouvre d'ailleurs de grandes perspectives (ex : points d'ébullition, températures critiques, densité) [97].

Il s'agit de présenter ici le principe des modèles QSPR ainsi que ceux des différents outils employés pour leur mise en place et leur évaluation : bases de données expérimentales, descripteurs, outils d'analyse de données.

2.1- Principe

Le principe des méthodes QSPR est, comme leur nom l'indique, de mettre en place une relation mathématique reliant de manière quantitative des propriétés moléculaires aussi bien électroniques que géométriques, appelées descripteurs, avec une observable macroscopique (activité biologique, toxicité, propriété physico-chimique, etc.), pour une série de composés chimiques similaires à l'aide de méthodes d'analyses de données. Aussi la forme générale d'un tel modèle est la suivante :

Propriété = f(Descripteurs) (2.44)

L'objectif d'une telle méthode est donc d'analyser les données structurales afin de détecter les facteurs déterminants pour la propriété mesurée. Pour ce faire, différents types d'outils peuvent être employés : régressions multi-linéaires (MLR) [98], régressions aux moindres carrés partiels (PLS) [99], arbres de décision [100], réseaux de neurones [101-103], algorithmes génétiques [104].

Une fois cette relation mise en place et validée sur un jeu validation, elle peut alors être employée pour la prédiction de la propriété de nouvelles molécules, pour lesquelles la valeur expérimentale n'est pas disponible, voire pour des molécules encore non synthétisées. De tels modèles peuvent également, dans certains cas, être utilisés pour mieux appréhender les phénomènes moléculaires mis en jeu dans la propriété d'intérêt.

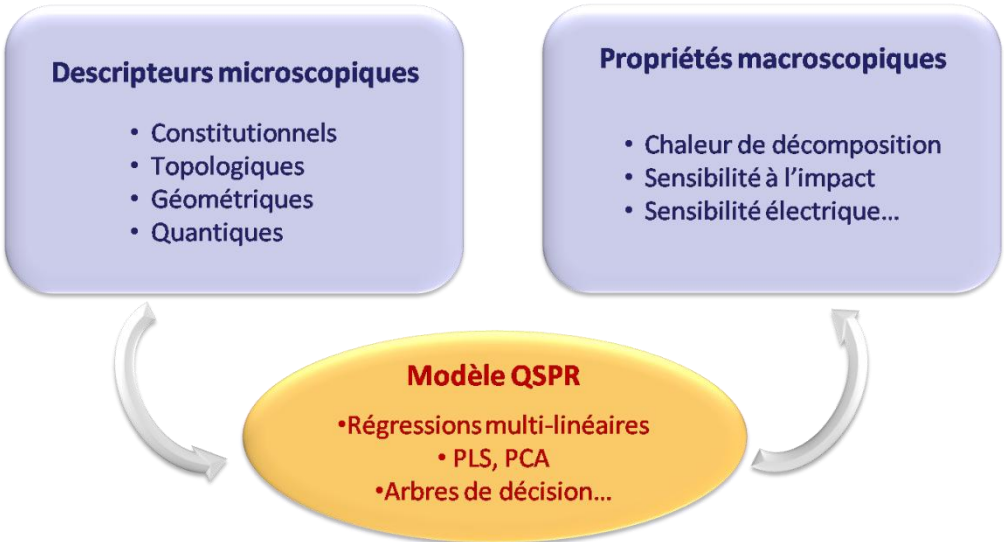


Figure 12 - Principe de la méthode QSPR

En pratique, le développement proprement dit d'un modèle débute par la collecte de données expérimentales fiables et en nombre le plus important possible. Il s'agit ensuite de développer une série de descripteurs qui caractérisent les structures moléculaires électroniques et géométriques des composés de la base de données en vue de les relier à la propriété expérimentale étudiée. Des outils d'analyse de données sont alors employés pour aider à choisir les descripteurs adéquats et mettre en place le modèle proprement dit. Une fois développé, le modèle doit alors être validé en termes de corrélation (sur le jeu de données d'entraînement). Sa robustesse, c'est-à-dire l'influence des composés du jeu d'entraînement sur le modèle, est estimée par des méthodes de validation interne. Pour estimer son pouvoir prédictif, il est ensuite nécessaire de disposer de données expérimentales supplémentaires afin de déterminer la capacité du modèle à prédire ces valeurs. Enfin, pour tout modèle, il est important de savoir pour quel type de molécules il est utilisable ou non, c'est-à-dire connaître son domaine d'applicabilité.

2.2- Importance des bases de données

Un modèle QSPR, de par sa construction, est très dépendant des données expérimentales de référence. Le choix de la base de données est donc un point critique de son développement. Dans la plupart des cas, les données expérimentales sont issues de la littérature.

Pour être de qualité, une base de données doit être composée de données expérimentales aussi fiables que possible, puisque les barres d'erreurs sur celles-ci se propageront dans le modèle final, étant donné que les paramètres de ce dernier sont ajustés par rapport à ces données. Il est donc important de choisir des données présentant des incertitudes faibles afin de limiter les barres d'erreur expérimentales. En effet, un modèle ne pourra être plus robuste statistiquement que les données théoriques à partir desquelles il a été développé.

De plus, les données doivent être obtenues suivant un protocole expérimental unique. En effet, les conditions expérimentales ont, en général, une forte influence sur les valeurs obtenues. La définition de la propriété en termes de conditions expérimentales est d'ailleurs un point important de la démarche.

Tant que faire se peut, la base de données doit également être caractéristique de l'éventail des valeurs que la propriété peut rencontrer, puisque c'est en partie sur ce critère que sera déterminé le domaine d'applicabilité du modèle final. D'une manière générale, plus ce dernier sera grand, plus des modèles prédictifs sur une large gamme de valeurs pourront être attendus.

Il faut d'ailleurs également que cette distribution soit la plus homogène (et normale) possible et ce pour deux raisons. La première vient du fait que la plupart des méthodes statistiques sont basées sur des distributions normales et sont donc plus performantes sur ce type de distributions. Ensuite, la présence de valeurs atypiques influence beaucoup les modèles et fausse les résultats. Sur la Figure 13, par exemple, la présence du composé extrême A donne l'impression d'une corrélation. De tels composés sont donc à éliminer. Sur l'illustration proposée, l'exclusion de la valeur atypique réduit la corrélation à R^2=0,04.

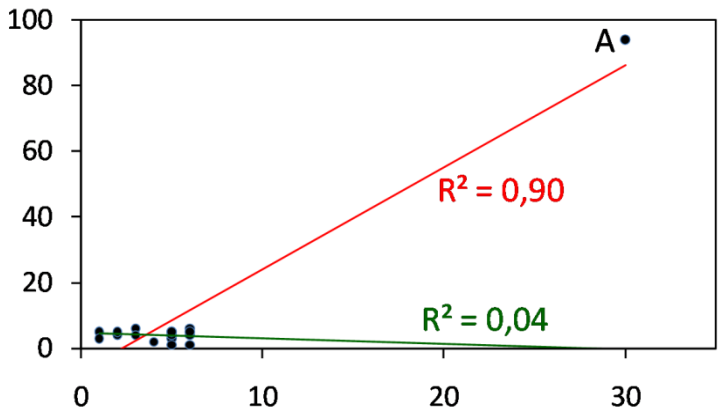


Figure 13 – Influence d'une valeur atypique (A) sur une corrélation

L'obtention d'un modèle efficace sera également dépendante du type de molécules qui y sont incluses. Un jeu convenable devrait contenir des composés pour lesquels le mécanisme mis en jeu est unique. De plus, plus le modèle présentera des molécules de structures proches, plus ce dernier aura de chances d'être performant au prix d'une réduction de son domaine d'applicabilité. Enfin, la présence de molécules au comportement déviant au sein de la base de données peut influencer de manière importante le modèle final. De même, la mise en évidence de sous-groupes de molécules dans un modèle a tendance à le dégrader. Il est préférable, dans de tels cas, de considérer des modèles multiples permettant de traiter les différents sous-groupes de manière indépendante.

2.3- Descripteurs moléculaires

Afin d'exploiter au maximum les informations contenues dans les structures moléculaires, celles-ci sont traduites en une série de grandeurs (en général scalaires) qui quantifient leurs caractéristiques physico-chimiques et structurelles. Ces grandeurs sont appelées descripteurs.

Depuis des décennies, de nombreux travaux ont été menés afin de développer des descripteurs capables de décrire de la manière la plus exhaustive possible les structures moléculaires. On en dénombre aujourd'hui des milliers [105,106]. Ils peuvent être calculés ou empiriques. Ces derniers nécessitant d'être mesurés, les descripteurs calculés seront privilégiés. Les détails les concernant sont disponibles dans l'ouvrage de Karelson [107].

2.3.1- Descripteurs constitutionnels

La manière la plus simple de représenter un système moléculaire est de considérer sa composition chimique, sans prise en compte de quelconques considérations géométriques ou électroniques. Des exemples caractéristiques de descripteurs constitutionnels sont :

- les nombres absolus et relatifs d'atomes (C, H, N…) ;
- les nombres absolus et relatifs de groupes fonctionnels (NO_2, COOH, OH…) ;
- les nombres absolus et relatifs de liaisons (simples, doubles, aromatiques…) ;
- les nombres absolus et relatifs de cycles (aromatiques ou non) ;
- la masse moléculaire.

Ces descripteurs sont très couramment utilisés du fait de leur extrême simplicité non seulement d'un point de vue conceptuel mais surtout calculatoire. Ils sont notamment à la base des analyses par groupes de contribution, comme la méthode de Benson [108], dans laquelle l'effet des substituants est introduit sous la forme d'une valeur tabulée.

Si ce type de descripteurs est employé pour obtenir les modèles QSAR/QSPR les plus simples, ces derniers peuvent poser problème pour une bonne interprétation des mécanismes d'interaction mis en jeu pour la propriété étudiée. Par exemple, ils ne permettent pas de distinguer les isomères. Or, pour la grande majorité des propriétés, la position d'un substituant modifie la valeur expérimentale d'une propriété. Les paramètres constitutionnels sont, dans de tels cas, défaillants. Il faut alors recourir à d'autres classes de descripteurs.

2.3.2- Descripteurs topologiques

Les descripteurs topologiques sont des indices obtenus à partir d'une structure 2D de la molécule, à savoir une simple table de connectivité des atomes dans la molécule. Ils contiennent en leur sein des informations sur la taille globale du système, sa forme globale et ses ramifications [105].

Ces descripteurs, ou indices topologiques, s'inspirent de la théorie des graphes appliquée à la table de connectivité qui n'est autre qu'une représentation compacte de la connectivité interatomique au sein de la molécule.

Le système moléculaire est considéré comme un graphe *G[V,R]* dans lequel les atomes sont des sommets *V* et les liaisons chimiques des arêtes *R*. La distance topologique *d* entre deux atomes est alors définie comme le nombre de liaisons minimum reliant ces atomes.

L'indice de Wiener [109], par exemple, permet de caractériser le volume moléculaire et le taux de ramification des molécules. Il est calculé comme la somme de toutes les distances topologiques entre atomes au sein de la molécule.

$$W = \sum_{i,j} d_{ij} \tag{2.45}$$

Si l'on s'intéresse au nitrobenzène, la Figure 14 représente la structure telle qu'utilisée pour calculer cet indice. Les seules informations nécessaires sont les atomes et leur connectivité d'un point de vue purement topologique. Les distances topologiques sont mesurées, comme ici pour l'atome C1 vis-à-vis de C3 et C4, puis toutes les distances au sein de la molécule sont sommées pour donner l'indice de Wiener qui vaut 88 pour le nitrobenzène. Son utilisation originale a concerné la caractérisation de la variation du point d'ébullition au sein d'une série d'alcanes Δt_b [109].

$$\Delta t_b = \frac{98}{n^2}\Delta W + 5{,}5\Delta p \tag{2.46}$$

où ΔW et Δp sont les variations respectives de l'indice de Wiener et du nombre de paires d'atomes de carbone séparés par trois liaisons carbone-carbone et *n* est le nombre d'atomes de carbone.

Parmi les autres indices topologiques courants, on retrouve l'indice de Randic [110], qui peut être interprété comme une mesure de l'aire de la molécule accessible au solvant, les indices de Kier-Hall [111] ou encore l'indice de Balaban [112]. Enfin, des indices de contenu d'information appliquent les fondements de la théorie de l'information de Shannon aux distances topologiques [113].

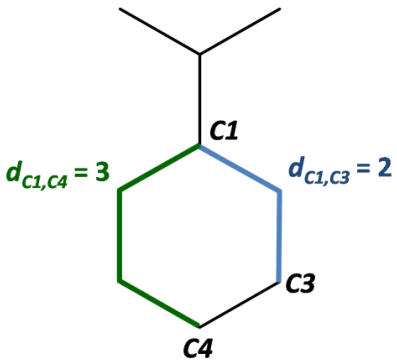


Figure 14 - Distances topologiques au sein de la molécule de nitrobenzène

En général, ce type de descripteurs simplifie grandement la représentation de la connectivité chimique au sein de la molécule puisqu'ils ne prennent pas en compte les différences de distances, d'angles et d'ordres de liaison ni même la nature des atomes dans la molécule. Si certains indices ont été développés pour intégrer de manière très approximative ce genre d'informations, ils restent souvent insuffisants pour caractériser l'intégralité des propriétés moléculaires. Finalement, les indices topologiques sont souvent considérés comme des descripteurs convenables d'un point de vue numériques. Cela dit, dans la plupart des cas, l'interprétation des équations QSAR/QSPR qui en résultent n'est pas aisée, puisqu'il est difficile de les relier aux mécanismes sous-jacents.

2.3.3- Descripteurs géométriques

Il s'agit de descripteurs évalués à partir des positions relatives des atomes d'une molécule dans l'espace, ainsi que des rayons et masses atomiques. Ils nécessitent donc de connaître la structure 3D de la molécule. Celle-ci peut être obtenue expérimentalement bien entendu mais le plus souvent par modélisation moléculaire, empirique ou *ab initio*. Par conséquent, contrairement aux descripteurs précédents, ils nécessitent un certain temps de calcul.

Parmi ces descripteurs, on retrouve le volume et la surface moléculaire, le moment d'inertie ou encore des distances, angles ou angles dièdres particuliers entre atomes dans la molécule.

Un tel paramètre géométrique a pu notamment montrer son rôle central pour caractériser l'activité des complexes bis(arylimino)pyridine du fer (II) pour l'oligomérisation de l'éthylène [114]. Il a pu être montré que l'activité de ces complexes est liée à un angle dièdre particulier φ (identifié sur la Figure 15) qui régit l'accessibilité au cœur réactif métallique du précurseur du complexe catalytique par son activateur.

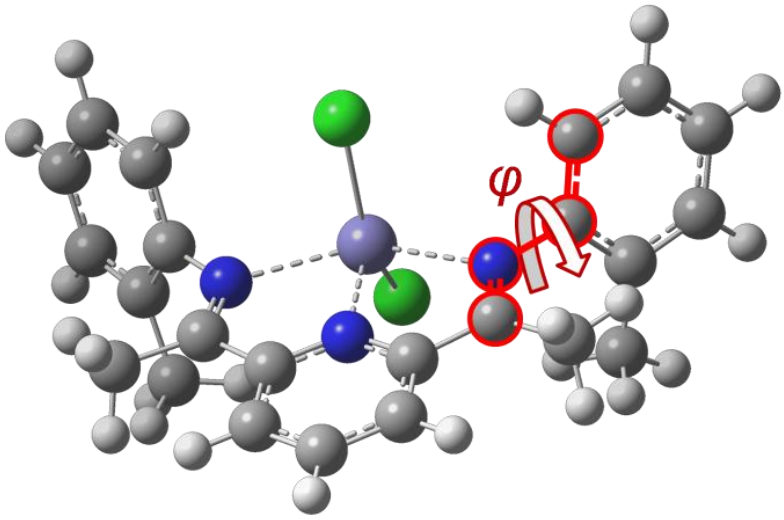


Figure 15 – Angle dièdre dirigeant l'activité catalytique des complexes bis(arylimino)pyridine du fer

2.3.4- Descripteurs quantiques

Afin d'aller plus loin dans la description des structures moléculaires, des caractéristiques supplémentaires de la structure moléculaire peuvent encore être quantifiées. Ces descripteurs qui, pour être calculés de manière fiable, nécessitent des calculs plus sophistiqués sont regroupés ici sous l'étiquette de « descripteurs quantiques ».

Les approches de chimie quantique nous donnent accès à des informations supplémentaires telles que des données énergétiques, vibrationnelles et orbitalaires du système.

Il est alors possible, par exemple, de calculer des énergies de dissociation. Pour ce faire, considérons une dissociation homolytique :

$$AB \rightarrow A^{\bullet} + B^{\bullet} \tag{2.47}$$

L'énergie de dissociation de la liaison A-B est alors la différence d'énergie entre la molécule AB et la somme des radicaux $A^{\bullet}$ et $B^{\bullet}$ issus de la dissociation de la liaison considérée.

$$E_{diss}(A\text{-}B) = E(A^{\bullet}) + E(B^{\bullet}) - E(AB) \tag{2.48}$$

A partir de tels calculs, sont également évaluées des propriétés électroniques telles que le moment dipolaire ou la polarisabilité. Des descripteurs permettent également de quantifier les différentes interactions inter- et intramoléculaires, ces dernières ayant une grande influence sur la plupart des propriétés des systèmes moléculaires. Ce sont notamment des descripteurs de réactivité pouvant être dérivés de la DFT conceptuelle ou encore des charges partielles pouvant être calculées selon différentes approches.

Descripteurs de réactivité issus de la DFT conceptuelle

La DFT conceptuelle permet de caractériser les propriétés de réactivité des composés chimiques [115,116]. Les descripteurs qui en sont issus représentent un moyen simple de rationaliser le comportement chimique des molécules, sur la base de la théorie HSAB (*Hard and Soft Acids and Bases*) de Pearson [117]. Leur fiabilité a d'ailleurs été démontrée via différentes analyses théoriques, dédiées principalement à la réactivité chimique [118-120].

La Figure 16 présente comment, en dérivant l'énergie au sens de la DFT, des descripteurs de la réactivité chimique peuvent être obtenus.

Le potentiel chimique μ mesure la tendance du nuage électronique à s'échapper de la molécule. Il s'agit d'une propriété globale du système moléculaire égal à la pente de l'énergie en fonction du nombre d'électrons N à potentiel externe $v(r)$ constant et, par la même, à l'opposé de l'électronégativité χ telle que définie par Pauling et Mulliken [121,122].

$$\mu = \left(\frac{\partial E}{\partial N}\right)_{v(r)} = -\chi \tag{2.49}$$

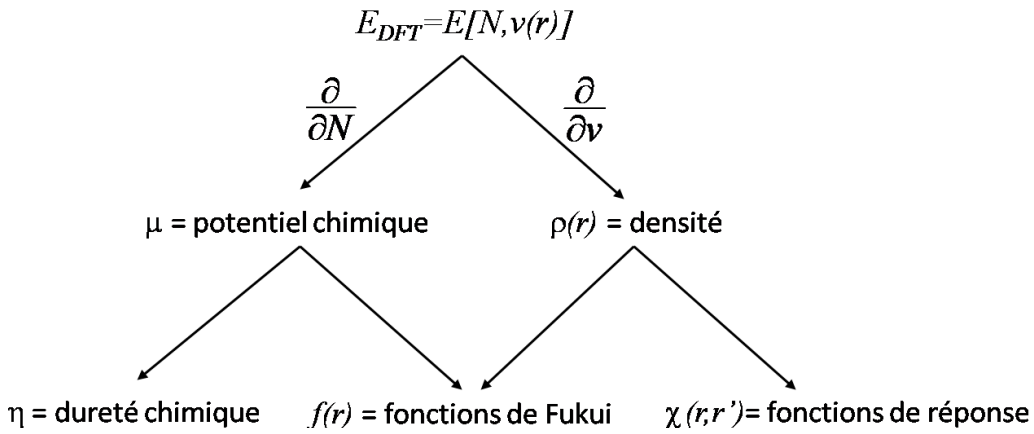


Figure 16 - Dérivées de l'énergie aux 1^er^ et 2^nd^ ordre

Via une approximation aux différences finies en trois points, μ peut alors être exprimée en fonction du potentiel d'ionisation IP et de l'affinité électronique EA.

$$\mu = -\frac{IP + EA}{2} = -\chi \tag{2.50}$$

Il s'agit d'une propriété fondamentale. En effet, quelle que soit la définition choisie pour un atome ou une molécule, quand des atomes de potentiels chimiques différents s'assemblent pour former une molécule caractérisée par son propre potentiel chimique, ils conservent leur identité mais leurs potentiels chimiques doivent s'égaliser. Il s'agit du principe d'égalisation de l'électronégativité de Sanderson [123] qui est utilisé pour le calcul des charges atomiques par les approches empiriques qui seront présentées par la suite.

En considérant la première dérivée de ce potentiel chimique, la dureté η et son inverse la mollesse S peuvent être obtenues [124,125].

$$\eta = \left(\frac{\partial \mu}{\partial N}\right)_{v(r)} = \left(\frac{\partial^2 E}{\partial N^2}\right)_{v(r)} = \frac{1}{S} \tag{2.51}$$

Là-encore, par différence finie, ces grandeurs, fondamentales dans la théorie des acides et des bases telle que développée par Pearson [126], peuvent être exprimées en fonction des IP et EA.

$$\eta = \frac{1}{S} = IP - EA \tag{2.52}$$

Afin de caractériser d'autres comportements électroniques des systèmes, d'autres descripteurs ont été développés. Pour caractériser la capacité d'une molécule à engendrer un transfert d'électron, l'indice d'électrophilicité ω est calculé selon la formule suivante [127].

$$\omega = \frac{\mu^2}{2\eta} \tag{2.53}$$

Finalement, ces descripteurs peuvent être calculés de manière très simple dans le cadre de l'approximation de Koopmans [128] selon laquelle :

$$IP=-\varepsilon_{HOMO} \tag{2.54}$$

$$EA=-\varepsilon_{LUMO} \tag{2.55}$$

où ε_{HOMO} et ε_{LUMO} sont les énergies de l'orbitale la plus haute occupée et de l'orbitale la plus basse vacante.

Les propriétés considérées jusque-là sont globales, c'est-à-dire qu'elles sont constantes sur l'ensemble du système moléculaire. Afin d'analyser les propriétés de réactivité localement, il faut s'intéresser à la dérivée du potentiel chimique, dépendante du potentiel externe.

$$f(r)=\left(\frac{\partial\mu}{\partial v(r)}\right)=\left(\frac{\partial\rho(r)}{\partial N}\right)_{v(r)} \tag{2.56}$$

Cette fonction *f(r)* est appelée fonction de Fukui [129] et mesure la sensibilité du potentiel chimique du système à une perturbation externe en un point particulier. Sur cette base, une série de descripteurs locaux peut être développée. Si l'on considère ensuite qu'un ajout ou une perte d'électrons n'entraine de modifications que dans la structure électronique, de nouvelles grandeurs sont développées sur le même schéma, pour caractériser respectivement les attaques nucléophiles, électrophiles et radicalaires.

$$f^{+}(r)=\left(\frac{\partial\rho(r)}{\partial N}\right)^{+}_{v(r)} \tag{2.57}$$

$$f^{-}(r)=\left(\frac{\partial\rho(r)}{\partial N}\right)^{-}_{v(r)} \tag{2.58}$$

$$f^{0}(r)=\left(\frac{\partial\rho(r)}{\partial N}\right)^{0}_{v(r)} \tag{2.59}$$

La DFT conceptuelle permet donc d'accéder à une série de descripteurs de la réactivité. Comme nous venons de le voir, ceux-ci sont définis, à la fois aux niveaux global et local.

Analyses de population

Les analyses de population représentent un moyen mathématique de considérer une fonction d'onde ou une densité électronique au niveau local sous la forme de charges atomiques ou encore d'ordres de liaison.

Ces populations électroniques ne sont pas des observables. Malgré tout, elles sont déterminantes dans les mécanismes de la plupart des réactions chimiques et propriétés physico-chimiques. Elles interviennent notamment dans les interactions électrostatiques au sein des systèmes moléculaires. Aussi, nombreux sont les descripteurs basés sur les charges partielles destinées à caractériser les interactions intermoléculaires dans les analyses de

type QSAR/QSPR. La distribution de charge peut être traitée à différents niveaux de théorie depuis les approches empiriques basées sur le concept d'électronégativité des atomes aux méthodes issues de la chimie quantique, telles que l'approche de Mulliken ou l'analyse de populations naturelles (NPA).

Approches empiriques

Un des schémas empiriques les plus courants est celui de Gasteiger [130,131]. Cette approche repose sur une adaptation du principe d'égalisation des électronégativités atomiques de Sanderson [123] en partant de la définition de l'électronégativité de Mulliken au niveau orbitalaire.

$$\chi_\nu = \frac{IP_\nu + EA_\nu}{2} \tag{2.60}$$

où IP_v et EA_v sont les potentiels d'ionisation et affinités électroniques des orbitales v dans les atomes. La charge atomique est alors calculée de manière itérative.

$$Q_i = \sum_\alpha q_i^{<\alpha>} \tag{2.61}$$

$$\text{avec} \quad q_i^{<\alpha>} = \left(\frac{1}{2}\right)^\alpha \sum_{\nu\in i}\left[\sum_{\mu\in j}\frac{\chi_{i\mu}^{<\alpha>} - \chi_{i\nu}^{<\alpha>}}{\chi_{i\nu}^{+}} + \sum_{\mu\in k}\frac{\chi_{k\lambda}^{<\alpha>} - \chi_{i\nu}^{<\alpha>}}{\chi_{k\lambda}^{+}}\right] \tag{2.62}$$

où $q_i^{<\alpha>}$ est la contribution à la charge atomique à l'itération α, elle-même calculée à partir des électronégativités des premiers voisins de l'atome considéré.

Zefirov [107] a proposé une autre technique basée sur le même principe d'égalisation des électronégativités atomiques de Sanderson en considérant, cette fois, l'électronégativité comme une moyenne géométrique des électronégativités atomiques :

$$\chi_i = \left(\chi_i^0 \prod_{k=1}^{n} \chi_k\right)^{1/(n+1)} \tag{2.63}$$

où χ_i est l'électronégativité de l'atome i dans la molécule, $\chi_i°$ celle de l'atome isolé et les χ_k correspondent aux électronégativités des k atomes environnants dans la molécule.

Approche de Mulliken

Les charges atomiques peuvent également être calculées à partir des orbitales moléculaires issues de la chimie quantique selon la méthode de Mulliken [132]. Il s'agit d'un des schémas les plus employés. Il repose sur l'hypothèse que le recouvrement entre deux orbitales centrées sur deux noyaux différents est partagé de manière égale.

Dans le cadre de la LCAO, la charge électronique moléculaire totale peut s'écrire :

$$\rho_{el} = 2\sum_{i,k,l} c_{ik}c_{il}S_{kl} = 2\sum_{k,l} P_{kl} \tag{2.64}$$

où c_{ik} et c_{il} sont les coefficients d'une orbitale moléculaire sur un jeu d'orbitales donné et P_{kl} les éléments de la matrice densité.

Quant à S_{kl}, il s'agit de l'intégrale de recouvrement des orbitales χ_k et χ_l.

$$S_{kl} = \int_V \chi_k \chi_l dV \tag{2.65}$$

La charge atomique q_A au sens de Mulliken est alors définie selon la relation suivante :

$$q_A = Z_A - \left(\sum_k P_{kk} + \frac{1}{2} \sum_l P_{kl} + \frac{1}{2} \sum_l P_{lk} \right) \tag{2.66}$$

où Z_A est la charge nucléaire de l'atome et la sommation des populations de Mulliken est réalisée sur toutes les orbitales atomiques (de valence) de l'atome.

Si cette représentation, très simple conceptuellement, ne respecte pas les électronégativités élémentaires, elle montre une certaine efficacité pour des petites bases mais est très dépendante de la base choisie. Elle peut donner notamment des résultats incohérents pour des bases importantes, en particulier lorsqu'elles incluent des fonctions diffuses [114].

Approche NPA

Dans le cadre de l'analyse des populations naturelles (NPA) [133], plutôt que les orbitales moléculaires, ce sont les fonctions propres de la matrice de densité réduite au premier ordre, ou orbitales naturelles, qui sont employées. Par définition, ces dernières sont localisées et orthogonales. La procédure de localisation permet aux orbitales moléculaires d'être définies de la même manière que celles centrées sur des atomes ou celles entourant une paire d'atomes.

A partir de ces orbitales, non seulement les charges atomiques peuvent être déterminées mais l'analyse des poids des fonctions d'onde et des propriétés nodales permet aussi la caractérisation des différentes orbitales (liantes, non-liantes, de cœur, Rydberg) ainsi que des liaisons chimiques via l'approche NBO (pour *Natural Bond Orbital*) [134]. En général, cette approche est moins dépendante de la base choisie que l'approche de Mulliken.

2.4- Méthodes d'analyse de données

La mise en place de modèles QSPR n'est pas une chose aisée. La première difficulté réside dans la différence d'échelles existant entre les données à corréler, la structure étant à une échelle moléculaire alors que les propriétés à prédire sont à une échelle macroscopique. De plus, il tient compte des problèmes d'incertitudes à la fois au niveau des structures moléculaires (liées niveau de calcul) et des données expérimentales (protocoles de mesures).

Un des problèmes importants réside également dans le traitement de données en grande quantité. Un grand nombre de descripteurs et de molécules peuvent être à analyser, mais aucune règle stricte n'existe quant au choix des paramètres structuraux les plus importants parmi le jeu complet de ceux disponibles.

En fait, de nombreux outils existent et il s'agit de trouver le moyen le plus adapté pour obtenir un modèle fiable à partir des données disponibles. Selon les cas, plusieurs approches sont envisageables, il faut alors choisir celle permettant de caractériser au mieux le système.

Les différentes méthodes présentées dans la suite sont celles employées au cours de l'étude, pour développer des modèles (plus ou moins de paramétrés, linéaires ou non linéaires, interprétables ou non), choisir les paramètres les plus pertinents, valider ces modèles (en interne ou en externe) et déterminer leurs domaines d'applicabilité.

2.4.1- Régression multi-linéaire

La régression multi-linéaire (MLR, pour *Multiple Linear Regression*) [135] est la méthode la plus simple et la plus communément employée pour le développement de modèles prédictifs. Elle repose sur l'hypothèse qu'il existe une relation linéaire entre une variable dépendante y (ici, la propriété) et une série de n variables indépendantes x_i (ici, les descripteurs). L'objectif est d'obtenir une équation de la forme suivante :

$$y = a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n \tag{2.67}$$

où a_i sont les coefficients de la régression.

La Figure 17 donne une représentation graphique de la méthode pour deux variables dépendantes. Il s'agit en fait de définir le plan au plus proche de tous les points de l'espace (au sens des moindres carrés). Pour n variables, il s'agira d'un hyperplan d'ordre n.

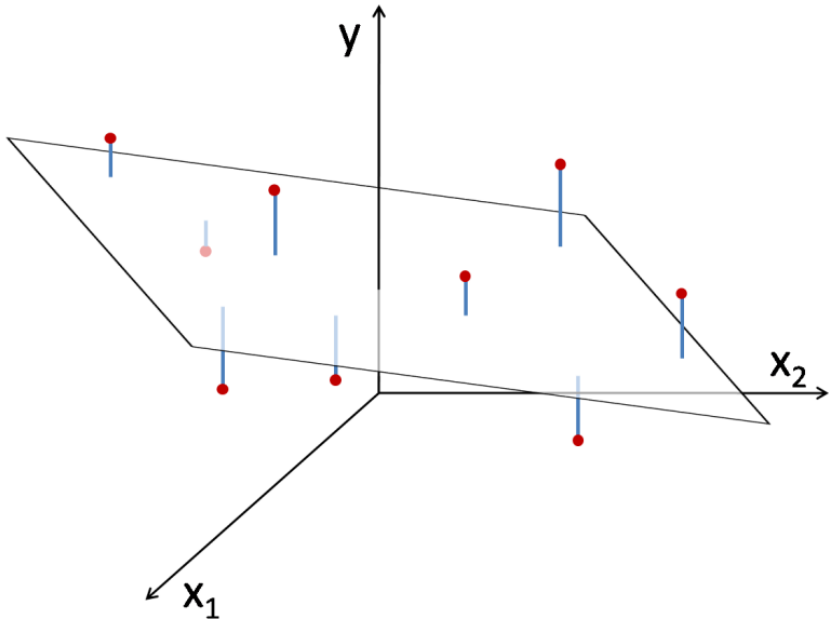


Figure 17 - Représentation graphique de la régression multi-linéaire pour deux variables indépendantes x_1 et x_2 et une variable dépendante y

La détermination de l'équation (2.67) se fait alors à partir d'une base de données de p échantillons pour laquelle à la fois les variables dépendantes et la variable indépendante sont connues. Il s'agit donc de considérer un système de p équations.

$$\begin{aligned} y_1 &= a_0 + a_1x_{1,1} + a_2x_{2,1} + \ldots + a_nx_{n,1} + \varepsilon_1 \\ y_2 &= a_0 + a_1x_{1,2} + a_2x_{2,2} + \ldots + a_nx_{n,2} + \varepsilon_2 \\ &\ldots \\ y_p &= a_0 + a_1x_{1,p} + a_2x_{2,p} + \ldots + a_nx_{n,p} + \varepsilon_p \end{aligned} \tag{2.68}$$

où les résidus ε_i représentent l'erreur du modèle, constituée par l'incertitude sur la variable dépendante y_i d'une part, sur les variables indépendantes x_i d'autre part, mais aussi par les informations contenues dans les variables indépendantes mais non exprimées via les variables dépendantes.

Ce système d'équations peut être écrit sous la forme matricielle suivante :

$$\begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{n,1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,p} & \cdots & x_{n,p} \end{pmatrix} \begin{pmatrix} a_0 \\ \vdots \\ a_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_p \end{pmatrix} \tag{2.69}$$

soit de manière condensée :

$$\boldsymbol{Y} = \boldsymbol{X}\,\boldsymbol{A} + \boldsymbol{\varepsilon} \tag{2.70}$$

La méthode consiste alors à choisir les coefficients du vecteur $\boldsymbol{A}$ en faisant en sorte de minimiser la somme des carrés des écarts entre les valeurs prédites et les valeurs réelles sur l'intégralité de la base de données et ceci sous couvert de certaines hypothèses de départ.

En premier lieu, les variables indépendantes x_i, comme leur nom l'indique, sont supposées indépendantes entre elles et leur incertitude est négligeable. Ensuite, les différents échantillons y_i sont supposés indépendants entre eux et suivent une distribution normale. L'erreur $\boldsymbol{\varepsilon}$ est elle-même supposée suivre une distribution normale, centrée en 0. Enfin, par nature, la dépendance de y vis-à-vis des x_i est supposée linéaire.

La valeur prédite de la variable dépendante est alors :

$$\hat{y}_i = \hat{a}_0 + \hat{a}_1x_{1,i} + \ldots + \hat{a}_nx_{n,i} \tag{2.71}$$

Les résidus peuvent donc être définis comme la différence entre les valeurs prédites et observées de y.

$$\varepsilon_i = y_i - \hat{y}_i \tag{2.72}$$

Il s'agit alors de trouver les coefficients $\hat{a}_i$ afin de minimiser la somme des carrés de ces résidus pour l'intégralité de la base de données.

$$\begin{aligned} min\,[\textstyle\sum(\varepsilon_i)^2] &= min\,[\textstyle\sum(y_i - \hat{y}_i)^2] = min\,[\textstyle\sum(y_i - \hat{a}_0 - \hat{a}_1x_{1,i} - \ldots - \hat{a}_nx_{n,i})^2] \\ &= min\,(\mathbf{Y}\text{-}\mathbf{X}\hat{\mathbf{A}})^{\mathrm{T}}(\mathbf{Y}\text{-}\mathbf{X}\hat{\mathbf{A}}) \end{aligned} \tag{2.73}$$

Les coefficients peuvent être obtenus à partir de l'équation matricielle suivante :

$$\hat{\mathbf{A}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \quad (2.74)$$

Bien entendu, la régression multi-linéaire souffre de certains désavantages. Le principal découle de sa linéarité. Elle est donc défaillante pour la mise en évidence de dépendances non-linéaires. Cela dit, elle n'en reste pas moins une méthode simple et efficace dans la plupart des cas. De plus, pour peu que les variables indépendantes soient choisies de manière raisonnée, les équations obtenues peuvent être interprétées d'un point de vue phénoménologique.

2.4.2- Analyse par composantes principales

L'analyse par composantes principales (PCA, pour *Principal Component Analysis*) consiste à transformer un jeu de variables corrélées entre elles en un nouveau jeu de variables, appelées composantes principales, moins nombreuses mais indépendantes. En utilisant ces nouvelles variables, la dimensionnalité du système est réduite en perdant un minimum d'information. La Figure 18 représente graphiquement le principe de la méthodologie pour 3 variables x_1, x_2 et x_3 et deux composantes principales *PC1* et *PC2.*

Considérons l'ensemble des données comme un espace à n dimensions, chaque dimension représentant une variable de la base de données. Chaque échantillon de la base de données est donc un point dans cet espace à n dimensions [x_1 … x_n].

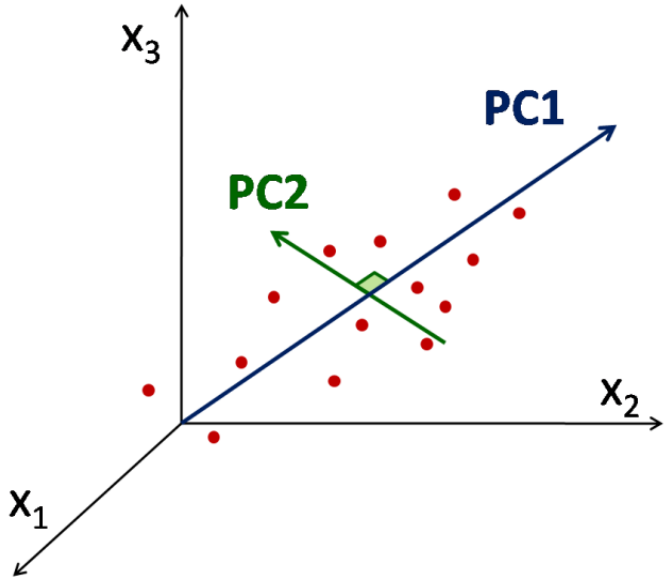


Figure 18 – Principe de mise en place de deux composantes principales PC1 et PC2 dans un espace à 3 dimensions (x_1, x_2, x_3) lors d'une analyse PCA

Dans un premier temps, il s'agit de construire une première composante principale PC1 dont la coordonnée t_1 correspond à la combinaison linéaire des n variables passant au plus proche de tous les points du système.

$$t_1 = a_{1,1}x_1 + a_{2,1}x_2 + \ldots + a_{n,1}x_n \quad (2.75)$$

où les constantes $a_{i,1}$ sont choisies pour maximiser la variance sur t_1.

Une deuxième composante principale PC2 est ensuite construite orthogonale à la première. Ainsi, PC1 et PC2 sont indépendantes.

$$t_2 = a_{1,2}x_1 + a_{2,2}x_2 + \ldots + a_{n,2}x_n \quad (2.76)$$

où les constantes $a_{i,2}$ sont choisies de sorte que la variance sur t_2 soit maximale tout en imposant l'orthogonalité des vecteurs $[a_{i,1}]$ et $[a_{i,2}]$. L'ajout de composantes principales supplémentaires se fait de la même manière.

Ainsi, la première composante principale traduit la plus grande part de la variance globale du système, les composantes successives n'ayant pour but que d'expliquer la variance résiduelle, non expliquée par les composantes qui la précèdent. Pour déterminer à quel moment le nombre de composantes principales est suffisant pour traduire la variance du système, une démarche de validation interne est réalisée et la composante PCi est considérée utile si elle contribue de manière significative à améliorer la robustesse de l'analyse.

Il est très courant d'utiliser des représentations à deux dimensions suivant les composantes PC1 et PC2, ces deux composantes étant celles qui caractérisent la plus grande part de la variance dans le système. Ce dernier peut alors être examiné sous deux points de vue différents. La matrice des coordonnées nous permet d'analyser la dispersion des individus dans le nouvel espace alors défini. Ainsi, deux échantillons proches graphiquement portent une information très similaire (ex : *A* et *B* en Figure 19).

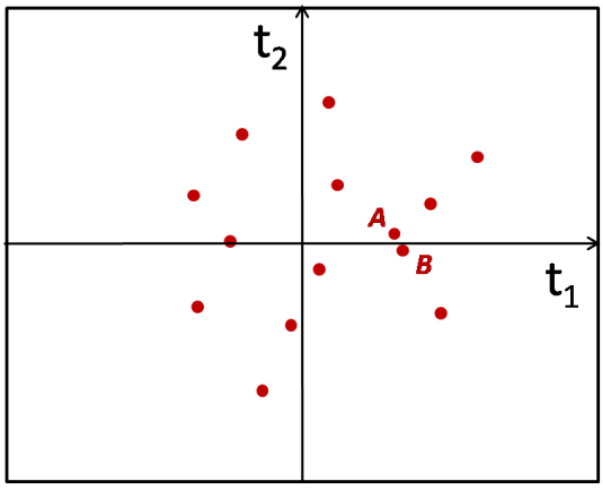


Figure 19 – Matrice des coordonnées dans le plan des deux premières composantes principales d'une analyse PCA

La matrice des poids (wc_i) des différentes variables dans les composantes principales permet quant à elle d'avoir une représentation de l'espace des variables. Celles qui contribuent le moins se situent à proximité de l'origine (ex : v_1 en Figure 20), celles contribuant le plus vers les valeurs (positives ou négatives) importantes (ex : v_2). Il s'en suit que deux variables très corrélées seront proches dans cette représentation (ex : v_3 et v_4).

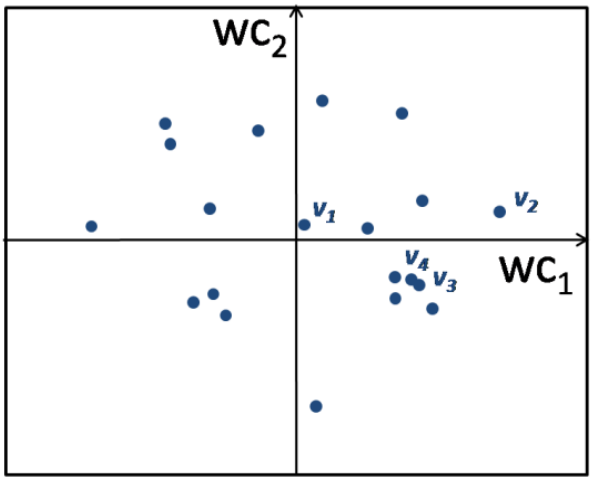


Figure 20 - Matrice des poids dans le plan des deux premières composantes principales d'une analyse PCA

Finalement, la PCA est une analyse non supervisée puisqu'elle considère toutes les variables de manière indépendante. Elle est utile dans l'identification de variables fondamentales, ainsi que de groupes de variables corrélées entre elles. Au-delà de son utilisation pour la réduction de dimensionnalité d'un problème, elle sert à faire ressortir les informations les plus caractéristiques d'un ensemble de données à la fois en termes de variables et en termes de descripteurs. Avec ce type d'analyse, il est notamment possible de trier graphiquement des échantillons en classes ou encore de mettre en évidence des variables corrélées ou non significatives d'un ensemble.

2.4.3- Régression aux moindres carrés partiels

La régression aux moindres carrés partiels (PLS, pour *Partial Least Squares* ou *Projection on Latent Structures*) [136] est en quelque sorte une version supervisée de la PCA. Il s'agit dans ce cas de considérer deux types de variables : une ou des variable(s) dépendante(s) Y_i – dans le cas d'une analyse QSPR, une ou des propriété(s) - dont la variance est expliquée par un nombre de variables indépendantes X_i – les descripteurs moléculaires.

La PLS repose (comme montré en Figure 21) sur une projection des X_i sur des composantes principales, comme dans le cadre de la PCA, à la différence près qu'ici, cette projection est guidée par leur relation avec les Y_i.

Le système peut alors être analysé, comme dans le cas d'une PCA à partir de la matrice des coordonnées et des poids. En plus des informations données par les représentations de ces matrices, l'importance des variables dans le modèle est traduite au travers d'un indice dédié, le VIP (pour *Variable Importance in the Projection*) [137].

Un des avantages de cette méthode de régression réside dans le traitement de bases de données de grande taille présentant de nombreuses variables corrélées entre elles [138]. On trouve donc des utilisations de cette méthode pour d'autres types d'applications telles que le traitement d'images [139].

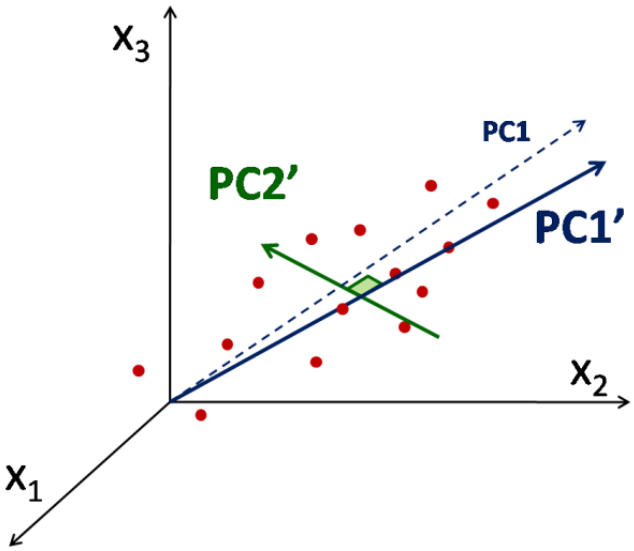


Figure 21 - Principe de mise en place de deux composantes principales PC1' et PC2' dans un espace à 3 dimensions (x_1, x_2, x_3) par analyse PLS

2.4.4- Arbre de décision

Les arbres de décision constituent une méthode de classement courante du fait de leur simplicité. Il s'agit de construire un arbre capable de réaliser un classement au sein d'un ensemble d'individus. La variable dépendante du système étant donc discrète, il s'agit d'une méthode qualitative. Un arbre de décision est représenté sur la Figure 22.

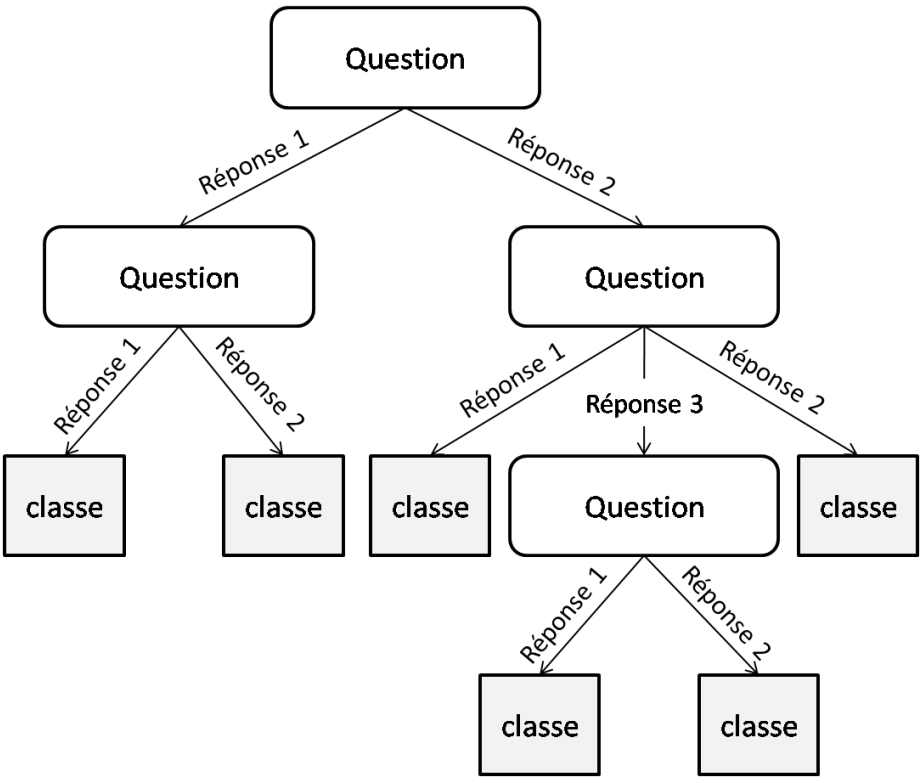


Figure 22 – Arbre de décision

Chaque nœud représente une question dont les réponses dirigent vers des branches qui mènent soit à un nouveau nœud soit à une feuille, à laquelle est associée une classe. Ainsi, chaque molécule est classée à partir de questions dont la réponse oriente le composé de proche en proche vers une classe.

Le principe de construction d'un arbre consiste à débuter d'un arbre vide et de le construire de proche en proche de manière descendante à partir du nœud racine jusqu'à l'obtention d'un échantillonnage pur, c'est-à-dire que tous les échantillons de la branche appartiennent à la même classe, ou jusqu'à ce que tous les descripteurs soient utilisés.

Les avantages de la méthode résident dans sa simplicité de compréhension et d'utilisation. Sa limite majeure réside dans sa nature. Comme tout outil de classification, un arbre de décision donne une information qualitative. Lorsque les données expérimentales d'origine sont des données quantitatives, il faut alors définir des classes de manière à obtenir une information exploitable. De même, les questions appellent des réponses discrètes. Finalement, toutes les informations quantitatives sont discrétisées, non seulement dans l'espace des propriétés, mais aussi dans l'espace des descripteurs.

2.4.5- Sélection des descripteurs

Lorsqu'une grande quantité de descripteurs est introduite, certains d'entre eux peuvent contenir des informations redondantes, entraînant un problème de colinéarité. Les paramètres employés doivent être, autant que possible, porteurs de sens et facilement interprétables d'un point de vue chimique. Les descripteurs sélectionnés seront donc d'autant plus pertinents qu'ils offrent une idée du mécanisme [34]. Au final, les modèles QSPR devraient être simples, transparents et compréhensibles d'un point de vue phénoménologique [34].

Il ne s'agit finalement que d'appliquer le principe de parcimonie qui considère que parmi un jeu d'explications toutes aussi bonnes les unes que les autres pour un phénomène donné, l'explication correcte est la plus simple. Cela dit, il ne faut pas pour autant perdre d'information. Comme a pu le dire Einstein : *"Everything should be made as simple as possible, but not simpler"* ("Tout devrait être fait aussi simple que possible, mais pas plus simple."). Dans une modélisation QSPR, ce principe signifie que le modèle doit avoir le moins de paramètres possibles tout en traduisant au mieux l'information contenue dans la propriété [140].

Aussi, le processus de simplification doit-il être une partie intégrante des tests d'hypothèse. De manière générale, pour qu'une variable soit retenue, il faut que son retrait entraîne une décroissance significative de performance du modèle. Bien entendu, en simplifiant le modèle, il faut être attentif à ne pas perdre des parts d'informations essentielles.

Différentes techniques permettent de réduire le nombre de descripteurs finalement inclus dans les modèles. L'une d'entre elles est la technique *Best Multi-Linear Regression* (BMLR) implémentée dans le logiciel CodessaPro [141]. La première étape de cette analyse consiste à exclure tous les descripteurs présentant une variance non significative. Cela permet de s'assurer que de tels descripteurs ne sont pas inclus par chance dans le modèle final alors qu'ils n'ont aucun lien avec la propriété étudiée. De même, lorsque deux descripteurs sont fortement corrélés entre eux, seul celui présentant la meilleure corrélation avec la propriété est conservé. Non seulement cette première étape évite l'introduction dans le modèle de descripteurs inappropriés mais elle rend, de surcroît, la suite de l'analyse moins coûteuse en terme de temps de calcul, puisqu'elle réduit le nombre de variables restant à traiter. Par la suite, en partant de paires de descripteurs orthogonaux (i.e. non corrélés entre eux), des modèles de rangs supérieurs sont développés en incluant de nouveaux descripteurs orthogonaux de manières successives tant qu'une augmentation de corrélation est observée. Au final, l'analyse BMLR sélectionne les meilleurs modèles à chaque rang et le modèle final doit être choisi parmi eux. Celui-ci doit être suffisamment corrélé sans être sur-paramétré, ce qui amènerait alors à une dégradation du pouvoir prédictif pour des molécules hors du jeu d'entraînement.

Finalement, le sens chimique des descripteurs utilisés doit bien entendu être pris en considération puisque, plus les descripteurs sont reliés chimiquement au phénomène, plus la probabilité de faire intervenir des descripteurs par le biais du hasard est réduite [142].

2.4.6- Validation

Coefficients et tests statistiques standard

Afin de déterminer la qualité d'un modèle, différents indicateurs statistiques sont employés. Le plus répandu d'entre eux est le coefficient de corrélation R^2 qui évalue la part de la variance expliquée par le modèle.

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \tag{2.77}$$

où $\bar{y}$ est la valeur moyenne des valeurs prédites.

Plus la valeur de R^2 sera proche de 1 (cas idéal) et plus les valeurs prédites et observées sont corrélées. Bien entendu, un autre indicateur est l'erreur absolue moyenne (*MAE*, pour *mean absolute error*), à laquelle est parfois préférée la déviation standard s.

$$MAE = \frac{\sum |\hat{y}_i - y_i|}{p} \tag{2.78}$$

$$s = \sqrt{\frac{\sum (\hat{y}_i - y_i)^2}{p - n - 1}} \tag{2.79}$$

L'indice de Fisher F est également couramment employé afin de mesurer le niveau de signifiance statistique du modèle, c'est-à-dire la qualité du choix du jeu de paramètres.

$$F = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \hat{y}_i)^2} \frac{p-n-1}{n} \tag{2.80}$$

En ce qui concerne la pertinence des descripteurs dans le modèle, elle est également évaluée par le *t*-test de Student. Il s'agit de tester l'hypothèse considérant le descripteur comme non significatif. Pour une régression multi-linéaire, cela revient à supposer le coefficient a_i qui lui est associé comme nul.

Cette hypothèse est rejetée (avec un intervalle de confiance α) si le ratio t_i entre a_i et son erreur type $s(a_i)$ atteint la valeur du fractile d'ordre $(1-\alpha/2)$ de la loi de Student à $(p-n-2)$ degrés de liberté.

$$|t_i| = \left|\frac{a_i}{s(a_i)}\right| > t^{p-n-2}_{1-\frac{\alpha}{2}} \tag{2.81}$$

Validation interne

La technique la plus employée pour déterminer la stabilité du modèle prédictif est de tester l'influence de chaque échantillon sur le modèle final. Pour ce faire, on emploie une technique de validation croisée (*cross validation* ou CV).

Ce processus consiste à extraire un certain nombre *n* de molécules du jeu initial à *k* molécules et à construire un nouveau modèle avec les *n-k* molécules restantes à l'aide des descripteurs choisis (seules les constantes de la régression changent). Ce nouveau modèle est alors utilisé pour la phase de prédiction sur les *n* molécules retirées. Ce processus est ensuite réitéré pour retirer et prédire les valeurs de toutes les molécules du jeu d'entraînement.

En fonction du nombre de molécules retirées à chaque itération, on parlera de *Leave-One-Out* (LOO) ou de *Leave-Many-Out* (LMO) selon qu'une ou plusieurs molécules est (sont) retirée(s). Dans ce dernier cas, on parle de partitionnement *N-fold* pour indiquer qu'une portion (1/N) du jeu de données est exclue à chaque itération. Ainsi le LOO correspond à une *cross validation* en *k-fold*.

D'une manière générale, ces techniques de validation interne permettent l'évaluation de la robustesse du modèle, autrement dit la stabilité des paramètres du modèle QSPR vis-à-vis des molécules du jeu d'entraînement. Cela dit, elles ne permettent en aucun cas de démontrer le pouvoir prédictif des modèles [143,144].

Validation externe

Afin de tester de manière fiable ce pouvoir prédictif, l'utilisation d'un jeu de validation externe, non employé pour le développement du modèle, est nécessaire. Pour peu que le jeu de données initial soit suffisamment large, ce dernier peut être aisément divisé en deux : un jeu d'entraînement sur lequel le modèle est développé et un jeu de validation utilisé pour caractériser son pouvoir prédictif.

Le jeu de validation doit être, tant que faire se peut, représentatif du domaine d'applicabilité du modèle donc du jeu d'entraînement. Bien entendu, il doit aussi être suffisamment important en nombre pour avoir une estimation correcte de la corrélation entre les valeurs calculées et prédites.

Plusieurs stratégies différentes sont employées pour définir ce dernier. Une procédure très courante consiste à faire une répartition aléatoire des données entre jeux d'entraînement et de validation. Afin de valider l'efficacité de la subdivision, il est recommandé, si possible, de tester l'influence de la subdivision sur les résultats en considérant une série de subdivision aléatoire. Ceci permet de déterminer les variations de performance des modèles en fonction de celle-ci. En effet, un tel partitionnement risque de donner lieu à des jeux de données dont les distributions sont incompatibles. Il faut donc vérifier que la partition choisie ne pose pas ce genre de problème.

Dans la suite de cette étude, une autre stratégie a également été employée. Celle-ci repose sur une partition basée sur la distribution des données expérimentales, en faisant en sorte d'obtenir des distributions homogènes des jeux d'entraînement et de validation avec le jeu de données complet.

Une fois le jeu de validation mis en place, il suffit alors d'appliquer le modèle aux molécules qui le composent et de déterminer la corrélation existant entre les données prédites et expérimentales. Plus cette corrélation est importante, plus le modèle est capable de prédire les propriétés pour des molécules hors du jeu d'entraînement.

2.4.7- Domaines d'applicabilité

Même les modèles les plus exhaustifs, dignes de confiance et validés, ne peuvent prédire des propriétés de manière fiable pour l'intégralité des composés chimiques existants. Aussi, les domaines d'applicabilité (AD) des modèles doivent-ils être définis et seules les prédictions concernant des molécules dans ce domaine peuvent être considérées comme recevables. En effet, un modèle QSPR n'est pas destiné à être employé en dehors de son domaine d'applicabilité, c'est-à-dire en dehors de l'espace chimique couvert par son jeu d'entraînement. La détermination des AD est donc d'une grande importance. Cette partie de l'analyse est d'ailleurs explicitement demandée dans les démarches de validation mises en place au niveau de l'OCDE [30,145].

Le domaine d'applicabilité est la région de l'espace chimique définie par les molécules du jeu d'entraînement du modèle. Il peut être caractérisé de différentes manières. Dans la suite de cette étude, une approche basée sur l'interpolation du domaine chimique des descripteurs a été employée.

L'interpolation est le processus d'estimation de valeurs en des points arbitraires d'un espace donné à partir de valeurs extrêmes connues. Aussi, dans le cas présent, la région d'interpolation pour un descripteur correspond simplement à l'intervalle des valeurs qu'il

prend au sein du jeu d'entraînement. Dans un espace à plusieurs dimensions, il existe plusieurs approches définissant des espaces géométriquement différents [146].

Une approche simple consiste à projeter directement les amplitudes des valeurs des descripteurs (à droite en Figure 23). Cette approche propose d'une manière générale des domaines d'applicabilité très restreints.

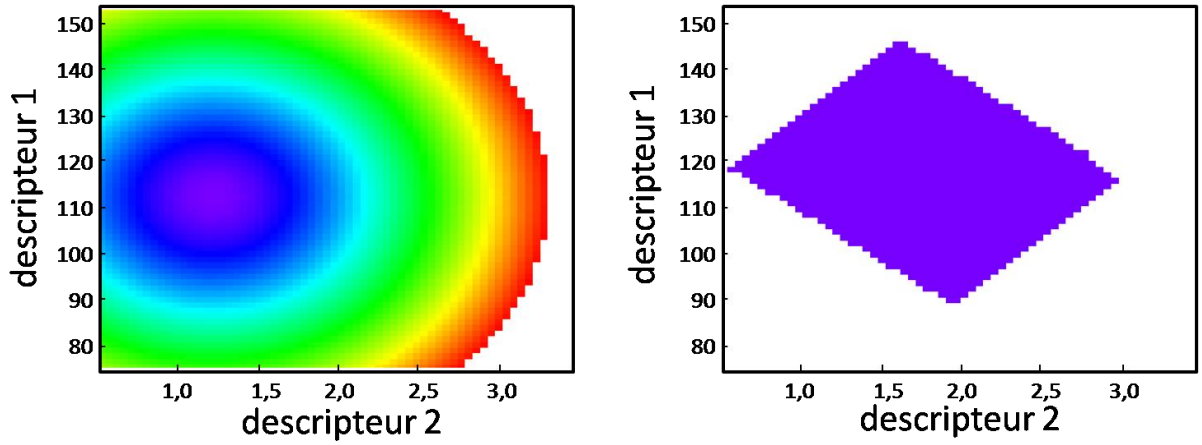


Figure 23 - Représentation géométrique du domaine d'application d'un modèle selon deux descripteurs par des approches basées sur des distances euclidiennes (à gauche) et par simple projection des éventails de valeurs (à droite)

Au cours de cette étude, l'approche envisagée est basée sur la distance des molécules du jeu d'entraînement dans l'espace des descripteurs du modèle. En particulier, des distances euclidiennes ont été considérées. Il s'agit alors de calculer la distance de chaque point par rapport au centre de la distribution des valeurs. Dans le cadre des distances euclidiennes d, cette distance est définie par la formule suivante :

$$d_i = \sqrt{(x_i - \bar{x})^T (x_i - \bar{x})} \qquad (2.82)$$

où x_i est la valeur du descripteur pour la molécule i et $\bar{x}$ est la valeur moyenne sur l'ensemble du jeu de données.

Cette approche extrapole légèrement l'espace des descripteurs. Aussi, pour limiter les problèmes liés à cette extrapolation, il peut être avantageux d'exclure une partie des molécules du jeu de données, les plus externes. Il s'agit alors de définir le domaine incluant, par exemple, 95% des molécules du jeu d'entraînement.

Ce type d'approche est particulièrement efficace pour des méthodes de régression ou de classification, en particulier lorsque le nombre de paramètres employés est limité, ce qui est le cas ici. Pour des modèles de nature différente (avec un grand nombre de paramètres notamment), d'autres approches peuvent être privilégiées (ex : similarité moléculaire [147,148]). En fait, il n'existe pas vraiment de règle quant à la mise en place des domaines d'applicabilité. L'outil doit être choisi au cas par cas, en fonction des données et du type de modèle développé.

2.4.8- Interprétation des modèles

L'interprétation des modèles en termes de mécanismes élémentaires sous-jacents est également un point important [149,150], recommandé dans la démarche de validation des modèles pour la prédiction dans un cadre règlementaire [30]. Outre l'aspect d'interprétation pure des modèles, qui peut permettre une meilleure compréhension des phénomènes chimiques mis en jeu, l'utilisation de paramètres interprétables dans les modèles prédictifs présente l'intérêt de limiter les risques d'avoir choisi ces derniers par chance, comme cela a pu être expliqué précédemment.

Bien entendu, cela n'est pas toujours simple. Les descripteurs ne sont pas forcément faciles à interpréter chimiquement. Les indices topologiques, notamment, sont des constructions mathématiques caractérisant la taille et la forme des systèmes moléculaires mais sans caractérisation explicite. Si leur efficacité en termes de prédiction n'est plus à démontrer, une interprétation physico-chimique est en général impossible.

Un autre obstacle à l'interprétation peut provenir du type de modèle choisi. Un grand nombre de descripteurs, par exemple, rend une équation difficilement interprétable du fait d'un trop grand nombre d'informations. De même, certains modèles non linéaires rendent l'interprétation de descripteurs, pourtant signifiants chimiquement, totalement impossible.

En termes de démarche, l'interprétation peut être prise en compte dès l'étape de sélection des données. En effet, au cours du processus, il peut être nécessaire de choisir entre deux descripteurs très proches statistiquement. Une technique automatisée peut, par exemple, mener au choix du moins signifiant au niveau chimique pour peu que sa corrélation avec la propriété expérimentale soit très légèrement supérieure. Inclure une phase de choix manuel des descripteurs peut alors permettre d'intégrer plus aisément des considérations physico-chimiques.

CHAPITRE 3

PREDICTION DES PROPRIETES EXPLOSIVES DES COMPOSES NITROAROMATIQUES

Sommaire

1- Méthodologie 82
1.1- Calcul des descripteurs moléculaires 82
1.2- Outils d'analyse de données 82
2- Stabilité thermique 83
2.1- Généralités 83
2.2- Travaux antérieurs 85
2.3- Chaleur de décomposition 86
2.4- Température de décomposition 123
2.5- Bilan 126
3- Sensibilité à la décharge électrique 128
3.1- Généralités 128
3.2- Travaux antérieurs 129
3.3- Analyse QSPR 131
3.4- Bilan 136
4- Sensibilité à l'impact 138
4.1- Généralités 138
4.2- Travaux antérieurs 139
4.3- Analyse QSPR 142
4.4- Bilan 149
5- Conclusion 149

La mise en place de modèles prédictifs de type QSPR a été envisagée à plusieurs reprises dans le passé pour les propriétés des substances chimiques dangereuses, mais ces études restent assez marginales en comparaison des travaux QSAR existants en toxicologie ou pharmacologie. Il s'agit ici d'avancer vers le développement de tels modèles pour les propriétés d'explosibilité d'une famille de composés explosifs, les nitroaromatiques.

Les détails relatifs aux méthodes utilisées dans ce chapitre seront tout d'abord mentionnés. Différentes approches ont ensuite été explorées, afin de prédire la chaleur de décomposition des dérivés du nitrobenzène, le but étant de mettre en évidence une méthodologie pour la mise en place de modèles prédictifs. Les approches ainsi mises en évidence seront ensuite appliquées à d'autres propriétés des substances explosibles : température de décomposition, sensibilité à la décharge électrique et sensibilité à l'impact.

1- Méthodologie

1.1- Calcul des descripteurs moléculaires

Toutes les structures moléculaires employées pour le développement des modèles QSPR ont été calculées à l'aide du logiciel Gaussian03 [151]. Deux niveaux de calculs ont été employés : la Théorie de la Fonctionnelle de la Densité (DFT), à l'aide de la fonctionnelle hybride PBE0 [89] et la base 6-31+G(d,p), et la paramétrisation semi-empirique AM1 (*Austin Method 1*) [77]. Dans les deux cas, après les optimisations de géométries, des calculs des fréquences vibrationnelles ont été réalisés au même niveau de théorie afin de s'assurer qu'aucune des structures optimisées ne présente de fréquence imaginaire. Les structures radicalaires ont été calculées dans le formalisme de Kohn-Sham sans contrainte de spin (*spin unrestricted*).

Ces structures ont ensuite été employées pour le calcul des descripteurs moléculaires. La plupart d'entre eux (plus de 300) ont été obtenus à l'aide du logiciel CodessaPro [141]. Les détails sur ces descripteurs sont disponibles dans l'ouvrage de référence de Karelson [107]. Les autres ont été calculés directement à partir des données extraites des calculs réalisés avec Gaussian03. Plus de détails sur les descripteurs sont donnés dans le chapitre 2.

1.2- Outils d'analyse de données

Développement des modèles

Les descripteurs moléculaires ont ensuite été utilisés dans des analyses statistiques par régressions linéaires et multilinéaires à l'aide du logiciel CodessaPro. En particulier, la méthode *« Best Multi Linear Regression »* (BMLR), implémentée dans CodessaPro, a été utilisée. Les détails de cette approche ont été développés dans le chapitre 2. Le modèle final est alors choisi pour représenter le meilleur compromis possible entre le nombre de descripteurs et la corrélation entre les valeurs prédites et expérimentales afin d'éviter toute sur-paramétrisation, qui mènerait à une perte de pouvoir prédictif du modèle pour des molécules hors du jeu d'entraînement.

Des analyses PCA (*Principal Component Analysis*) et PLS (*Partial Least Squares*) ont également été réalisées à l'aide du logiciel SIMCA P+ [152]. Le choix du nombre de composantes principales employées a été fait en considérant qu'une composante principale est utile au modèle si elle permet une amélioration significative de la validation interne (à savoir $\Delta R^2_{cv} \geq 0,05$).

Enfin, l'analyse par arbres de décision a été réalisée à l'aide de l'outil de classification J48 du logiciel WEKA [142,153]. Cet outil permet non seulement la mise en place des arbres de décision mais aussi le choix des descripteurs nécessaires.

Plus de détails sur ces différentes méthodes sont disponibles dans le chapitre 2.

Evaluation des modèles

La stabilité et la robustesse des modèles développés sont ensuite évaluées à l'aide de coefficients de corrélation (R^2), des déviations standard et des critères de Fisher F. De plus, le choix des descripteurs a été appuyé par un test de Student à un niveau de confiance de 95%. Tous les modèles ont été validés par *cross validation,* selon une procédure de *Leave-One-Out* (LOO). Dans certains cas, d'autres types de partition (*Leave-Many-Out*) ont également été employés. Par convention, sans indication contraire, les R^2_{cv} présentés sont issus d'une procédure LOO. Lorsque cela a été possible (nombre de molécules suffisant), le modèle a été évalué par validation externe à partir de données ne faisant pas partie du jeu d'entraînement. Le pouvoir prédictif est alors caractérisé par le coefficient de corrélation pour ce jeu de validation (R^2_{ext}).

Domaines d'applicabilité

Les domaines d'applicabilité ont enfin été définis à l'aide du logiciel Ambit Discovery [154], de sorte qu'ils incluent 95% des molécules du jeu d'entraînement. Le pouvoir prédictif du modèle au sein de son domaine d'applicabilité a ensuite été évalué par le coefficient de corrélation pour les molécules du jeu de validation qui appartiennent à ce domaine (R^2_{in}).

2- Stabilité thermique

2.1- Généralités

Les risques chimiques sont causés par la libération incontrôlée d'énergie au sein des matériaux énergétiques durant leur décomposition (potentiellement explosive). Pour évaluer expérimentalement la quantité de cette énergie libérée, des analyses calorimétriques sont réalisées sur les produits chimiques. La valeur de cette énergie est un critère de présélection pour le classement des substances explosives au sein du processus règlementaire (cf. Appendice 6 du règlement relatif au transport des marchandises dangereuses [40]).

Un essai expérimental typique est la calorimétrie différentielle à balayage (DSC, pour *Differential Scanning Calorimetry*) [33,155,156]. Cette mesure consiste à chauffer un échantillon et une référence à une même vitesse de montée en température constante, comme montré en Figure 24. Les flux de chaleur sont contrôlés afin de maintenir la même température au niveau de l'échantillon et de la référence.

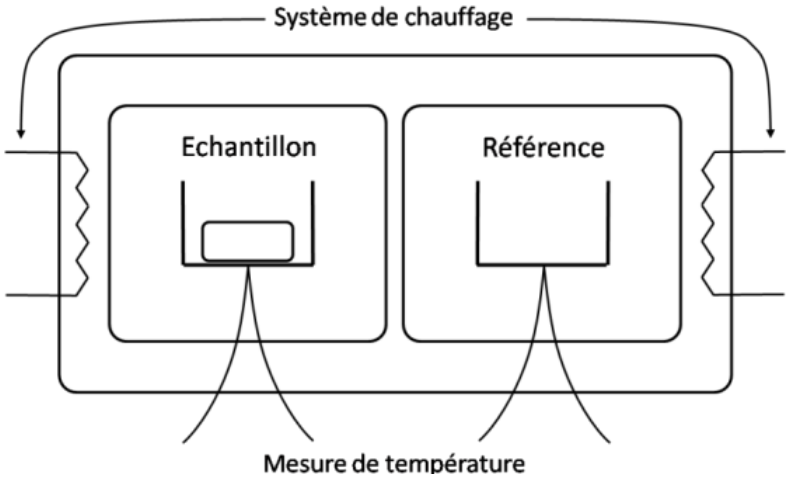


Figure 24 – Schéma de principe d'un calorimètre différentiel à balayage (DSC)

La chaleur captée ou libérée par l'échantillon est alors évaluée à partir de la différence de flux nécessaire pour que l'échantillon et la référence soient à la même température. Il est alors possible de mettre en évidence l'énergie mise en jeu par une transition de phase ou, comme dans le cas présent, par une décomposition. Ainsi, la DSC permet la détermination de la température et de la chaleur de décomposition. Sur la Figure 25, la température à laquelle la courbe s'éloigne de la ligne de base est la température de décomposition (T_0) alors que la chaleur de décomposition (-ΔH) est caractérisée par l'aire contenue entre la ligne de base (en pointillés) et la courbe.

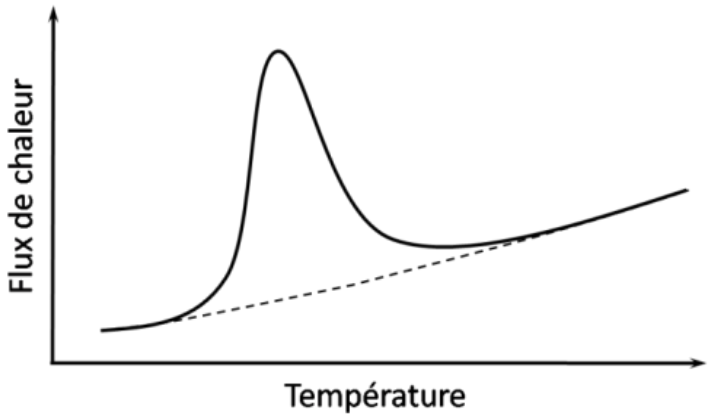


Figure 25 – Profil caractéristique d'un pic de décomposition exothermique tel qu'observé en DSC

Il est à noter que la chaleur de décomposition diffère de la chaleur de réaction thermodynamique puisqu'elle tient compte de phénomènes tels que l'évaporation ou l'absorption de chaleur par le creuset contenant l'échantillon.

Une incertitude expérimentale d'environs 5 à 10% est communément admise [157]. Elle s'explique par la sensibilité de l'appareillage. De plus, par les conditions et paramètres expérimentaux, telle que la vitesse de montée en température peuvent aussi influencer les

mesures. Parmi ces paramètres, le type de creuset utilisé dans le calorimètre est notamment un paramètre clé. En effet, pour éviter des phénomènes de réactivité pouvant intervenir entre l'échantillon et le matériau constituant le creuset, différents types de creusets existent (aluminium, acier). Les transferts de chaleur vers l'intérieur du creuset sont également différents d'un matériau à l'autre. De plus, les creusets peuvent être fermés ou ouverts et des effets d'évaporation peuvent affecter les mesures dans ce dernier cas.

Dans la suite de ces travaux, deux bases de données expérimentales de DSC seront utilisées. Dans les deux cas, il s'agit de composés nitroaromatiques, et en particulier de dérivés du nitrobenzène. Le premier jeu de données concerne 22 molécules caractérisées avec une montée en température de 4°C/min dans un creuset équipé d'un dispositif haute pression [158]. Les données du second jeu (77 molécules) ont été obtenues avec un dispositif différent, muni d'un creuset ouvert, avec une montée en température de 10°C/min [157].

2.2- Travaux antérieurs

Bien que la caractérisation expérimentale soit à présent bien définie, seuls quelques modèles prédictifs on été jusque là proposés pour les propriétés de stabilité thermique. Des travaux se sont notamment intéressés aux propriétés de stabilité thermique spécifiques de chromophores [159], polymères [160,161] ou liquides ioniques [162].

Dans le domaine des matériaux énergétiques, le logiciel CHETAH doit également être cité [13]. Cet outil, dédié à la prédiction des risques chimiques, inclut notamment un modèle de calcul de la chaleur de décomposition maximale des produits chimiques à l'aide de la méthode des groupes de contribution de Benson [163]. Néanmoins, les valeurs calculées selon cette méthode ne peuvent être directement comparées aux chaleurs de décomposition expérimentales puisqu'elle considère une décomposition "théorique" complète, c'est-à-dire jusqu'aux composés simples de type CO_2 ou NO_2.

Au début des années 1990, Grewer [15] a mis en évidence le fait que la présence de groupes fonctionnels particuliers (ex : C-NO_2, N-NO_2, C-N=O) influence la température de décomposition. Depuis, cette dernière a également été reliée aux énergies de dissociation des liaisons les plus faibles des molécules nitrées [164,165]. Dans ces études, cette dissociation a été considérée comme l'étape limitante du mécanisme de décomposition puis intégrée non pas dans une analyse statistique de type QSPR mais dans des approches cinétiques simplifiées. Mais aucune validation de ces approches n'a été réalisée.

A notre connaissance, la première étude de type QSPR concernant la stabilité thermique des composés nitroaromatiques a été réalisée par Saraf [16] sur la base de 22 données de DSC. Le modèle suivant a été proposé pour prédire la température de décomposition avec un coefficient de corrélation de 0,6.

$$T_o\ (°C) = 827{,}0 - 1036\ HPC - 4{,}43\ Sr - 5{,}07\ DM \qquad (3.1)$$

où HPC est la charge la plus positive au sein de la molécule, Sr est un indice de délocalisabilité électronique et DM est le moment dipolaire.

Le nombre de groupements nitro (n_{NO2}) a également été considéré suffisant pour estimer la chaleur de décomposition (-ΔH). Si aucune valeur de corrélation n'est proposée, une erreur moyenne d'environ 8% est calculée.

$$-\Delta H \text{ (kcal/mol)} = 75\ n_{NO2} \qquad (3.2)$$

Depuis, d'autres analyses QSPR, reposant sur des descripteurs de type constitutionnels, ont été dédiées à la prédiction de l'énergie d'activation de thermolyses. Keshavarz a ainsi proposé deux modèles, l'un pour des nitroaromatiques [166] et l'autre pour des nitramines [167], avec une corrélation significative (R^2=0,87) dans les deux cas.

L'objectif de ces travaux est donc de mettre en évidence les méthodologies nécessaires au développement d'un modèle performant et validé pour la chaleur de décomposition, dans un premier temps, puis pour la température de décomposition de composés nitroaromatiques.

2.3- Chaleur de décomposition

La première étape de la démarche concernera la mise en place d'un jeu de descripteurs adéquat pour ce type de propriétés et de molécules. La pertinence de ceux de type quantique sera tout d'abord vérifiée. Ils seront ensuite intégrés dans un jeu étendu de descripteurs (constitutionnels, topologiques, géométriques et donc quantiques). L'utilisation d'une base de données plus importante (77 molécules) permettra ensuite de réaliser une analyse complète avec validation externe et définition du domaine d'applicabilité. L'importance des mécanismes sous-jacents sera également démontrée, avec la mise en place d'un premier modèle utilisable pour des dérivés du nitrobenzène ne possédant pas de substituant en position ortho du groupement nitro. D'autres outils d'analyse de données seront ensuite utilisés (PCA, PLS, arbres de décision) pour obtenir un nouveau modèle certes qualitatif mais applicable à une plus grande variété de composés.

2.3.1- Intérêt des descripteurs quantiques

Une part du succès de tout modèle QSPR réside dans le choix des descripteurs moléculaires employés. En général, les descripteurs classiques utilisés pour de telles analyses sont des descripteurs constitutionnels, topologiques voire géométriques. Or, il est souvent difficile de relier ces paramètres aux phénomènes microscopiques mis en jeu. L'emploi de descripteurs issus de la chimie quantique est moins fréquent en QSPR alors qu'ils présentent l'avantage de pouvoir être reliés directement aux propriétés de réactivité des systèmes moléculaires. Les propriétés de décomposition (explosive) étant, par définition, des phénomènes de réactivité, leur utilisation semble donc tout indiquée, dans cette étude. La pertinence de tels descripteurs pour prédire la chaleur de décomposition (-ΔH) des 22 composés nitroaromatiques représentés sur la Figure 3 [158] a été analysée.

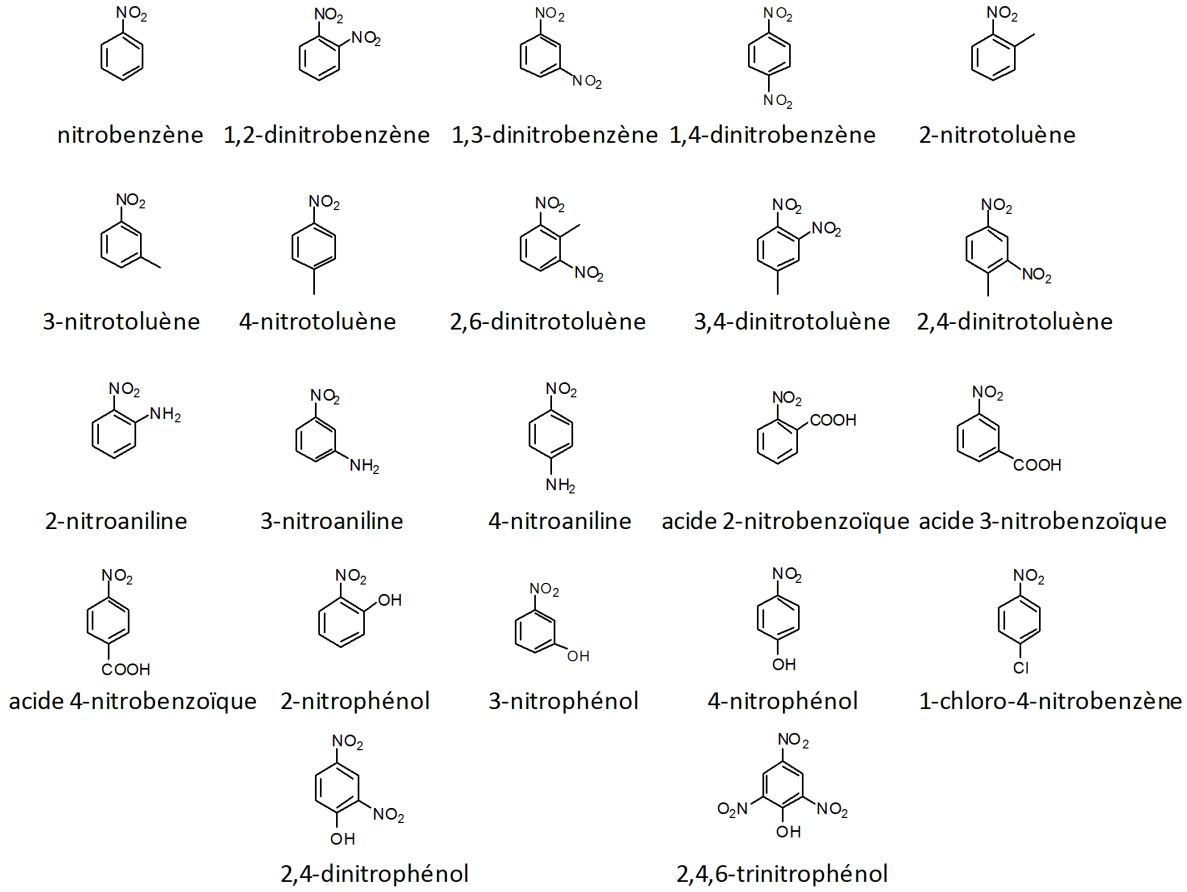


Figure 26 – 22 molécules de la base de données de Duh [158]

2.3.1.1- Energie de dissociation de la liaison C-NO_2

L'intérêt a tout d'abord été porté sur la relation entre la chaleur de décomposition expérimentale (-ΔH) et les énergies (calculées) relatives à la rupture homolytique de la liaison C-NO_2 (E_{diss}).

$$R\text{-}NO_2 \rightarrow R^{\bullet} + NO_2^{\bullet} \qquad (3.3)$$

Cette dissociation a déjà été considérée comme un paramètre déterminant pour l'étude de la sensibilité à l'impact de matériaux énergétiques à l'échelle moléculaire [168,169].

Son énergie est calculée de la manière suivante :

$$E_{diss} = E(R^{\bullet}) + E(NO_2^{\bullet}) - E(R\text{-}NO_2) \qquad (3.4)$$

où $E(R^{\bullet})$, $E(NO_2^{\bullet})$ et $E(R\text{-}NO_2)$ sont les énergies respectives des radicaux $NO_2^{\bullet}$, $R^{\bullet}$ et de la molécule initiale R-NO_2.

La liaison C-NO_2 est souvent considérée comme la liaison la plus faible chez les composés nitroaromatiques et représente donc une propriété fondamentale pour l'analyse de leur décomposition [164,165]. Une corrélation significative est donc attendue entre l'énergie de décomposition expérimentale et l'énergie de dissociation de cette liaison. Or, ce n'est pas le cas puisqu'un coefficient de corrélation linéaire R^2 de seulement 0,42 est calculé pour les 22 molécules de la base de données (cf. Tableau 4).

$$-\Delta H \text{ (kcal/mol)} = 2549{,}5 - 30{,}75\ E_{diss} \qquad (3.5)$$

Ce résultat n'est cependant pas si inattendu puisque les composés nitroaromatiques suivent des mécanismes de décomposition complexes et variés au cours du processus explosif [58]. Une discussion plus approfondie de ces mécanismes sera proposée dans le chapitre 4.

2.3.1.2- Descripteurs globaux de réactivité issus de la DFT conceptuelle

Cinq descripteurs globaux de réactivité, issus de la DFT conceptuelle, ont ensuite été calculés : potentiel d'ionisation (IP), affinité électronique (EA), électronégativité (χ), dureté (η) et indice d'électrophilicité (ω). Plus de détails sur ces descripteurs sont disponibles dans le chapitre 2. Ils ont été calculés en fonction des énergies des HOMO et LUMO (ε_{HOMO} et ε_{HOMO}), dans l'approximation de Koopmans [128].

Ces descripteurs représentent un moyen simple de rationaliser les différences de comportement chimique existant entre des espèces similaires [115,116]. Si leur fiabilité a été démontrée au travers d'analyses théoriques, en particulier relatives à la réactivité chimique [118-120], leur pertinence dans les modèles QSPR n'a été que plus sporadiquement démontrée [170,171].

Les valeurs calculées de ces cinq descripteurs, les énergies de dissociation de la liaison C-NO_2 précédemment considérées, ainsi que leurs coefficients de corrélation respectifs avec la chaleur de décomposition expérimentale, sont reportés dans le Tableau 4.

Les meilleures corrélations linéaires sont observées pour l'indice d'électrophilicité (R^2=0,75) et l'affinité électronique (R^2=0,71), celles-ci étant sensiblement meilleures que celle trouvée pour l'énergie de dissociation. Le potentiel d'ionisation et l'électronégativité ne présentent pas, pour leur part, de corrélation significative (R^2=0,35 et 0,52). La chaleur de décomposition est même indépendante de la dureté (R^2<0,01).

	E_{diss}	IP	EA	χ	η	ω
Nitrobenzène	71,2	187,5	60,7	124,1	126,8	60,7
1,2-dinitrobenzène	60,4	197,1	73,4	135,2	123,7	73,9
1,3-dinitrobenzène	68,5	206,7	76,7	141,7	129,9	77,3
1,4-dinitrobenzène	68,2	205,1	85,0	145,1	120,1	87,7
2-nitrotoluène	68,7	179,2	56,8	118,0	122,4	56,9
3-nitrotoluène	71,4	179,0	58,9	118,9	120,1	58,9
4-nitrotoluène	72,4	181,3	57,7	119,5	123,6	57,8
2,6-dinitrotoluène	63,8	195,1	69,3	132,2	125,8	69,4
3,4-dinitrotoluène	60,5	191,1	70,4	130,7	120,6	70,8
2,4-dinitrotoluène	65,9	198,6	72,2	135,4	126,4	72,5
2-nitroaniline	75,6	152,3	54,9	103,6	97,4	55,1
3-nitroaniline	71,6	153,1	56,7	104,9	96,5	57,1
4-nitroaniline	75,7	155,8	50,0	102,9	105,8	50,1
acide 2-nitrobenzoïque	62,7	189,4	63,2	126,3	126,2	63,2
acide 3-nitrobenzoïque	70,1	195,8	66,5	131,2	129,3	66,5
acide 4-nitrobenzoïque	69,5	194,3	72,2	133,3	122,1	72,7
2-nitrophénol	67,4	170,2	53,9	112,0	116,3	54,0
3-nitrophénol	70,5	169,3	60,3	114,8	109,0	60,4
4-nitrophénol	73,8	172,5	56,3	114,4	116,3	56,3
1-chloro-4-nitrobenzène	71,1	185,1	65,3	125,2	119,9	65,4
2,4-dinitrophénol	64,6	189,5	69,0	129,3	120,5	69,3
2,4,6-trinitrophénol	62,3	202,7	93,5	148,1	109,2	100,4
R^2	0,42	0,35	0,71	0,52	< 0,01	0,75

Tableau 4 – Energies de dissociation de la liaison C-NO_2 (E_{diss}) et descripteurs globaux de réactivité (en kcal/mol) calculés au niveau PBE0/6-31+G(d,p) pour 22 molécules nitroaromatiques

En Figure 27, deux descripteurs, l'électronégativité et l'affinité électronique, sont reportés en fonction des données expérimentales d'énergies de décomposition. Ce tracé permet de mettre en évidence l'importance du nombre de groupements NO_2 (ici, 1, 2 ou 3). La présence de tels groupements, dits explosophores, est d'ailleurs déjà une règle empirique utilisée pour une première présélection des produits chimiques explosibles, c'est-à-dire susceptibles de présenter des propriétés explosives (appendice 6 du manuel ONU pour le transport des marchandises dangereuses [40]). D'ailleurs, Saraf [16] avait proposé un modèle basé sur ce seul nombre de NO_2 (Eq. 3.2). Les modèles prédictifs sont donc sensés respecter cette observation. Par conséquent, le fait que les modèles obtenus ici, avec les descripteurs quantiques, présentent cette caractéristique est une preuve de la cohérence du modèle.

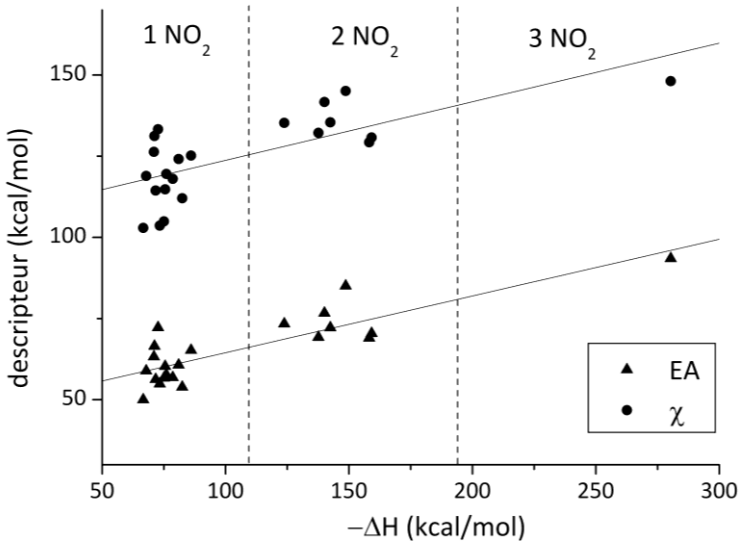


Figure 27 – Affinité électronique (EA) et électronégativité (χ) en fonction de la chaleur de décomposition expérimentale (-ΔH)

Quoi qu'il en soit, même si deux des cinq descripteurs calculés sont significativement corrélés avec les chaleurs de décomposition expérimentales, ces corrélations ne sont pas suffisantes pour des prédictions fiables. Par conséquent, les cinq descripteurs globaux de réactivité calculés ont été intégrés dans une analyse multivariée. Le meilleur modèle QSPR obtenu est une équation à deux paramètres, composée de la dureté et de l'indice d'électrophilicité.

$$-\Delta H \text{ (kcal/mol)} = -0{,}9\,\eta + 3{,}9\,\omega - 49{,}1 \qquad (3.6)$$

Ce modèle présente un R^2 de 0,77 et un R^2_{cv} de 0,63. Cette relation inclut seulement l'un des deux descripteurs les plus corrélés (EA et ω) alors que le paramètre le moins corrélé, la dureté, est présent. Ceci est dû au fait que l'affinité électronique est elle-même fortement corrélée avec l'indice d'électrophilicité. Ils sont donc redondants puisqu'ils traduisent une information similaire.

En fait, la dureté ne contribue pas de manière significative à la corrélation du modèle, la corrélation linéaire pour ω atteignant déjà 0,75. En d'autres termes, une régression ne mettant en jeu que cette classe de descripteurs ne présentera pas un pouvoir prédictif suffisamment élevé. Aussi, il faut utiliser un jeu de descripteurs plus étendu, incluant des descripteurs plus indépendants entre eux.

Si l'énergie de dissociation est alors ajoutée dans l'analyse BMLR, une régression plus corrélée est obtenue (R^2=0,82 et R^2_{cv}=0,67).

$$-\Delta H \text{ (kcal/mol)} = -1{,}4\,\eta + 3{,}2\,\omega - 3{,}4\,E_{diss} + 286{,}6 \qquad (3.7)$$

L'ajout supplémentaire du potentiel d'ionisation dans ce dernier modèle améliore encore le modèle puisque R^2 atteint 0,85 et R^2_{cv}=0,65.

$$-\Delta H \text{ (kcal/mol)} = 12{,}9\,\eta + 15{,}0\,\omega - 3{,}8\,E_{diss} - 13{,}4\,IP + 307{,}5 \qquad (3.8)$$

Bien entendu, la dureté et le potentiel d'ionisation sont, par définition, directement reliés comme montré dans le chapitre 2. L'équation (3.8) peut donc être réécrite.

$$-\Delta H \text{ (kcal/mol)} = 12{,}9\ EA + 15{,}0\ \omega - 3{,}8\ E_{diss} - 0{,}5\ IP + 307{,}5 \quad (3.9)$$

Cette équation présente alors un R^2 de 0,85 et un R^2_{cv} de 0,68. Ce n'est pas satisfaisant en termes de signifiance statistique des descripteurs, puisque la valeur du t-test pour IP n'est que de 0,60 alors que la valeur critique, pour une confiance à 90%, est déjà de 1,32 (pour une analyse à 22 échantillons). Si ce descripteur est alors ôté, l'équation suivante est obtenue :

$$-\Delta H \text{ (kcal/mol)} = -16{,}9\ EA + 18{,}1\ \omega - 3{,}6\ E_{diss} + 263{,}8 \quad (3.10)$$

Cette dernière régression a un R^2=0,85 et R^2_{cv}=0,73. Aussi, seuls deux descripteurs globaux de réactivité sont nécessaires pour améliorer significativement la corrélation de l'énergie de dissociation seule avec d'énergie de décomposition expérimentale. Dans le Tableau 5 sont reportées les chaleurs de décomposition obtenues avec les équations (3.5), (3.6) et (3.10) ainsi que les coefficients R^2 et R^2_{cv} correspondants.

	exp [158]	Eq. (3.5)	Eq. (3.6)	Eq. (3.10)
nitrobenzène	81,0	83,8	77,3	79,6
1,2-dinitrobenzène	123,8	163,6	132,1	143,1
1,3-dinitrobenzène	140,1	105,6	139,9	118,9
1,4-dinitrobenzène	148,7	105,6	189,6	167,4
2-nitrotoluène	78,6	105,6	66,3	85,0
3-nitrotoluène	67,9	83,8	76,2	77,0
4-nitrotoluène	76,0	76,6	68,8	73,0
2,6-dinitrotoluène	137,7	141,8	112,5	118,7
3,4-dinitrotoluène	159,2	163,6	122,6	136,5
2,4-dinitrotoluène	142,4	127,3	124,2	117,7
2-nitroaniline	73,4	54,9	81,4	60,6
3-nitroaniline	75,0	83,8	90,1	80,8
4-nitroaniline	66,7	54,9	54,3	52,0
acide 2-nitrobenzoïque	71,0	149,1	87,7	112,9
acide 3-nitrobenzoïque	71,2	91,1	98,0	90,3
acide 4-nitrobenzoïque	72,7	98,3	128,8	108,0
2-nitrophénol	82,5	112,8	60,3	86,9
3-nitrophénol	75,5	91,1	92,0	84,0
4-nitrophénol	71,7	69,3	69,4	65,7
1-chloro-4-nitrobenzène	86,0	83,8	102,0	87,9
2,4-dinitrophénol	158,2	134,6	116,8	118,3
2,4,6-trinitrophénol	280,4	149,1	249,3	275,4
	MAE (%)	23,5	19,4	16,0
	R^2	0,42	0,77	0,85
	R^2_{cv}	0,27	0,63	0,73

Tableau 5 - Chaleurs de décomposition (en kcal/mol) expérimentales et prédites des 22 composés nitroaromatiques étudiés

La Figure 28, représentant l'accord entre les valeurs calculées et expérimentales, illustre non seulement l'amélioration de la corrélation lorsque l'on passe du modèle (3.5) au (3.10), mais aussi la présence de cas problématiques au sein de la base de données.

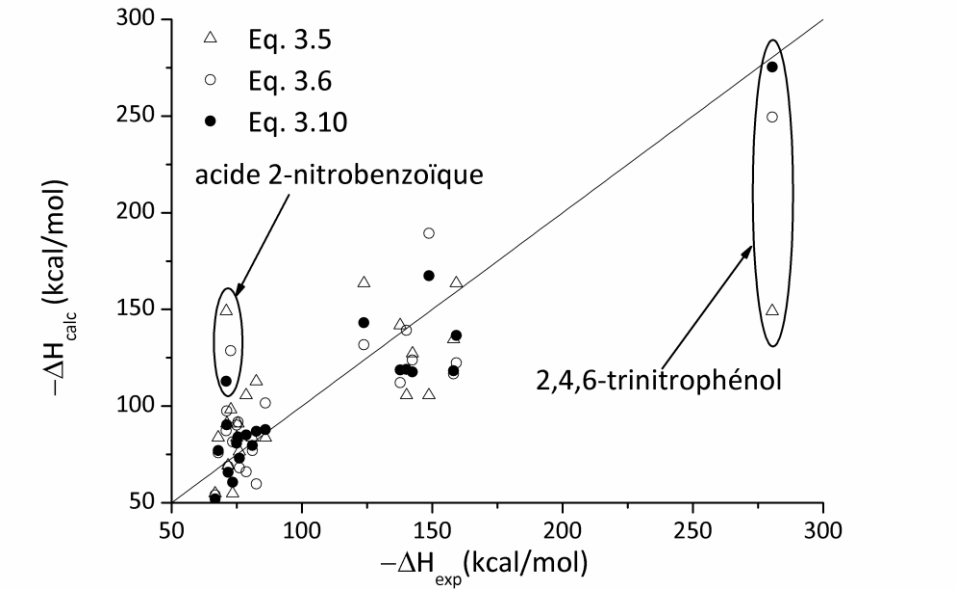


Figure 28 – Chaleurs de décomposition prédites ($-\Delta H_{calc}$) et expérimentales ($-\Delta H_{exp}$)

En effet, les chaleurs de décomposition des molécules de 2,4,6-trinitrophénol et d'acide 2-nitrobenzoïque sont particulièrement difficiles à reproduire à l'aide des modèles (3.5) et (3.6), qui montrent des déviations importantes vis-à-vis de l'expérience. Le modèle (3.10) améliore significativement la prédiction pour la première molécule, moins pour la seconde. Ces difficultés suggèrent la présence de mécanismes de décomposition différents de la simple dissociation C-NO_2 comme des études des mécanismes réactionnels de décomposition peuvent le montrer [58,172,173]. Néanmoins, ces résultats mettent en évidence à quel point les descripteurs issus de simples et "raisonnablement" rapides calculs DFT peuvent être utilisés pour la prédiction de la stabilité thermique.

2.3.1.3- Utilisation d'un jeu de descripteurs sélectionnés

Un jeu plus important de 14 descripteurs (incluant les descripteurs du paragraphe précédent) a ensuite été considéré. Ces descripteurs sont calculés de manière très simple, directement à partir des calculs DFT, voire même sans le besoin du moindre calcul quantique. Ces descripteurs ont été choisis car ils pourraient tous, a priori caractériser les propriétés de décomposition des composés nitroaromatiques.

En plus des cinq descripteurs globaux de réactivité précédemment employés, quatre descripteurs locaux décrivent le groupement NO_2 et sa liaison au cycle aromatique. Il s'agit de l'énergie de dissociation de cette liaison, E_{diss}, définie précédemment, la distance carbone-azote d_{CN}, la charge du groupement nitro Q_{NO2} (calculée à partir des charges NPA) et le potentiel à mi-distance, V_{mid}. Ce dernier est une estimation du potentiel créé par les

charges des atomes de carbone et d'azote (Q_c et Q_N) à mi-distance de la liaison C-N et a, par le passé, été corrélé à la sensibilité à l'impact de matériaux énergétiques [54,174].

$$V_{mid} = \frac{Q_C + Q_N}{0{,}5 \times d_{CN}} \tag{3.11}$$

Les calculs quantiques ont aussi donné accès à l'énergie d'atomisation qui représente l'énergie nécessaire à la rupture de toutes les liaisons dans la molécule.

$$E_{atomisation} = \sum_i E(atome\ i) - E(molécule) \tag{3.12}$$

La structure électronique de la molécule est également décrite par son moment dipolaire et sa polarisabilité, qui caractérisent l'influence d'un champ externe sur la densité électronique du système. Si le moment dipolaire (DM) est un scalaire, la polarisabilité est un tenseur à partir de laquelle est calculée la polarisabilité principale (α).

$$\alpha = \frac{\alpha_{xx} + \alpha_{yy} + \alpha_{zz}}{3} \tag{3.13}$$

où les α_{ii} sont les composantes diagonales de la matrice de polarisabilité.

Enfin, les derniers descripteurs envisagés sont obtenus sans le moindre calcul quantique. Il s'agit de la masse moléculaire (M_w) et de la balance en oxygène (OB). Cette dernière est un descripteur empirique traditionnel des matériaux énergétiques qui caractérise la quantité d'oxygène nécessaire à la conversion du carbone et de l'hydrogène de la molécule en dioxyde de carbone (CO_2) et eau (H_2O) [36,175] et est calculée selon la formule proposée par Lothrop [44] :

$$OB = \frac{-1600\left(2X + Y/2 - Z\right)}{M_w} \tag{3.14}$$

où X, Y et Z sont respectivement les nombres d'atomes de carbone, d'hydrogène et d'oxygène.

Notons que, jusqu'à présent, les chaleurs de décomposition étaient considérées en kcal/mol puisque tous les descripteurs étaient calculés dans cette unité. A présent, différents type de descripteurs, de natures et donc d'unités différentes sont employés. Les énergies en Joules (J) étant plus communément employées expérimentalement, la définition des chaleurs de décomposition en kJ/mol a été préférée pour la suite de l'étude.

Régressions linéaires

Les coefficients de corrélation entre chacun de ces 14 descripteurs et la chaleur de décomposition expérimentale des 22 composés nitroaromatiques sont reportés en Tableau ***6***. Le descripteur le plus corrélé est la masse moléculaire (R^2=0,76) et la plus faible corrélation concerne la dureté (R^2<0,01). Les descripteurs relatifs à la liaison C-NO_2 ne montrent pas de corrélation significative (R^2<0,50). Comme déjà montré précédemment, l'énergie de dissociation ne présente qu'un coefficient de corrélation R^2=0,42, en accord avec la complexité du mécanisme de décomposition des nitroaromatiques. Des corrélations

substantiellement plus importantes sont montrées pour l'indice d'électrophilicité (0,75), l'affinité électronique (0,71) et la balance en oxygène (0,64). Il est à noter que ces descripteurs sont, par définition, corrélés deux à deux, M_w avec OB et ω avec EA (cf. respectivement Eq. 3.14 et chapitre 2).

	descripteurs	R^2
d_{CN}	Distance C-N en Å	0,10
Q_{NO2}	Charge NPA sur NO_2	0,46
V_{mid}	Potentiel à mi-distance en $Å^{-1}$	0,01
E_{diss}	Energie de dissociation en kcal/mol	0,42
M_w	Masse moléculaire en g/mol	0,76
OB	Balance en oxygène en %	0,64
E_{atom}	Energie d'atomisation en kcal/mol	0,21
DM	Moment dipolaire en DM	0,10
α	Polarisabilité principale en $Å^3$	<0,01
IP	Potentiel d'ionisation en u.a.	0,35
EA	Affinité électronique en u.a.	0,71
χ	Electronégativité en u.a.	0,52
η	Dureté en u.a.	<0,01
ω	Indice d'électrophilicité en u.a.	0,75

Tableau 6 - Corrélations linéaires de 14 descripteurs sélectionnés avec la chaleur de décomposition expérimentale de 22 composés nitroaromatiques

En tous cas, même si des corrélations mono-descripteurs significatives apparaissent, elles ne sont pas suffisantes pour prédire de manière efficace les valeurs expérimentales. En particulier, les valeurs des descripteurs ne varient pas avec la position des substituants lorsque l'on considère la masse moléculaire ou la balance en oxygène. En effet, les masses moléculaires des o-, m- et p-nitrotoluènes présentent la même masse moléculaire (137 g/mol) alors que leurs chaleurs de décomposition sont différentes : respectivement 329, 284 et 318 kJ/mol. Aussi, l'utilisation d'au moins un descripteur supplémentaire est nécessaire pour caractériser l'influence de la position des substituants.

Régression multilinéaire

Une analyse multivariée a donc été employée, en utilisant le jeu complet des 14 descripteurs calculés via l'analyse *Best Mult Linear Regression* du logiciel CodessaPro. La meilleure régression, estimée en termes de R^2_{cv} (0,84), est une équation à 6 paramètres construite autour de l'indice d'électrophilicité ω, les autres descripteurs intégrés dans l'équation étant moins corrélés (cf. Tableau 6).

$$-\Delta H\ (kJ/mol) = 33854\ \eta + 40050\ \omega - 1030\ \alpha - 33785\ IP + 25\ DM - 14\ E_{diss} + 973 \qquad (3.15)$$

Evidemment, 22 molécules ne sont pas suffisantes pour estimer le pouvoir prédictif d'un modèle aussi robuste soit-il. Pour ce faire, l'obtention de bases de données expérimentales homogènes, fiables et plus nombreuses est nécessaire, bien que cela soit difficile pour cette classe de composés (potentiellement explosifs). Quoi qu'il en soit, ces résultats (avec un R^2=0,91) sont prometteurs, comme le montre la Figure 29. Le modèle a

été appliqué au 2,4,6-trinitrotoluène (TNT), une molécule non incluse dans le jeu d'entraînement. La valeur prédite est de 920 kJ/mol ce qui est proche de la valeur expérimentale (998 kJ/mol [176]).

Dans ce modèle, des descripteurs de réactivité à la fois globaux (η, ω et IP) et liés à la liaison C-NO_2 (E_{diss}) sont présents, ce qui justifie l'usage de ces types de descripteurs pour la stabilité thermique des composés nitroaromatiques. Ils distinguent tous les isomères entre eux, ce qui n'est pas le cas des descripteurs constitutionnels tels que la balance en oxygène ou encore le nombre de NO_2. Contrairement à ce qui avait été considéré par Saraf [16], ce dernier ne semble d'ailleurs pas suffisant à la caractérisation complète de la stabilité thermique des composés nitroaromatiques. Il n'en reste pas moins qu'il s'agit d'une propriété fondamentale, la présence de groupements explosophores, nitro par exemple, dans les substances chimiques étant un critère de pré-sélection dans le cadre règlementaire relatif aux substances explosibles [40]. D'ailleurs, le modèle développé reste en accord avec cette considération empirique puisqu'il permet d'identifier clairement les composés mono-, di- et trinitroaromatiques, comme le montre la Figure 29.

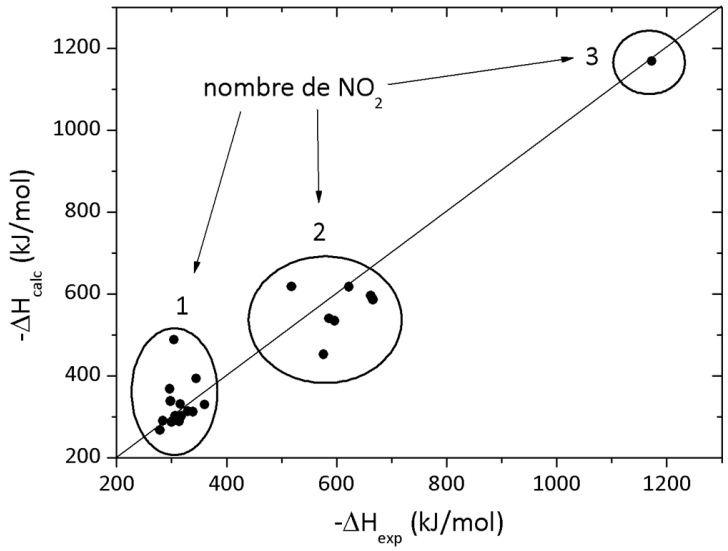


Figure 29 - Chaleurs de décomposition (en kJ/mol) prédites (Eq. 3.15) et expérimentales

Ce modèle utilisant exclusivement des descripteurs issus de calculs quantiques offre des résultats intéressants en termes de performance (R^2=0,91 et R^2_{cv}=0,84). De plus, ces descripteurs ayant un sens chimique ont l'avantage de distinguer les isomères entre eux.

2.3.2- Utilisation d'un jeu de descripteurs étendu

Les descripteurs issus de la chimie quantique se sont révélés tout à fait intéressants pour la prédiction de la chaleur de décomposition. Ils ont donc été inclus dans une analyse considérant un jeu de descripteurs plus diversifiés comportant à la fois des descripteurs quantiques plus variés, avec notamment des indices locaux de réactivité, mais aussi d'autres plus classiques des analyses structures-propriétés (constitutionnels, topologiques). Au total, plus de 300 descripteurs ont donc été obtenus à l'aide du logiciel CodessaPro [141], auxquels s'ajoutent les descripteurs précédemment considérés.

2.3.2.1- Régression multi-linéaire

Ces descripteurs, obtenus à partir des structures moléculaires optimisées au niveau PBE0/6-31+G(d,p), ont été utilisés pour le développement de modèles QSPR pour prédire la stabilité thermique des 22 composés nitroaromatiques jusqu'à présent considérés. La technique *Best Multi Linear Regression* a permis d'extraire les modèles les plus performants.

Une étape importante concerne alors la définition du nombre optimal de descripteurs à introduire dans le modèle final de sorte qu'il soit le plus robuste possible tout en évitant sa sur-paramétrisation. Cette étape est particulièrement importante, ici, du fait du grand nombre de descripteurs considérés. Pour ce faire, nous avons considéré une valeur limite pour l'augmentation du R^2 vis-à-vis du nombre de descripteurs (ici, 0,01) au-dessous de laquelle le système peut être considéré comme optimal. La progression du R^2 en fonction du nombre de descripteurs est présentée en Figure 30.

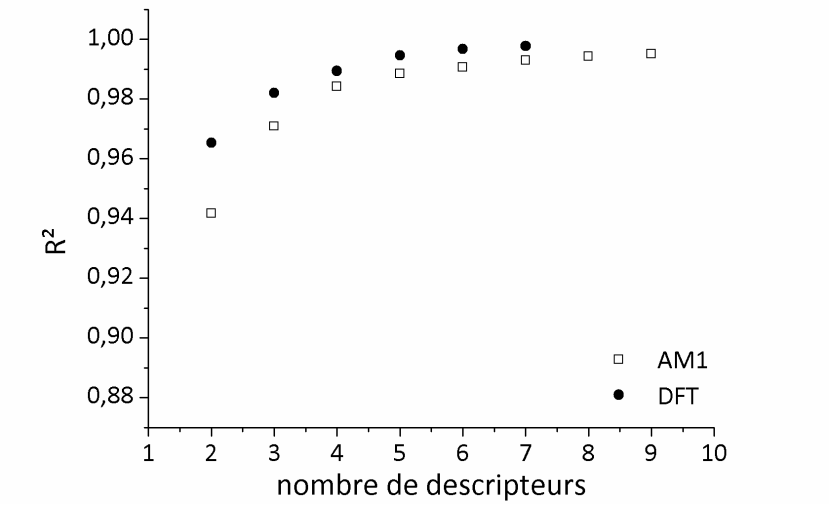


Figure 30 – Corrélation R^2 en fonction du nombre de descripteurs

L'augmentation du R^2 devient inférieure à 0,01 entre les modèles à trois et quatre descripteurs. En conséquence, l'équation à trois descripteurs est considérée comme le modèle DFT le plus prédictif évitant toute sur-paramétrisation. L'équation à trois descripteurs (également détaillée dans le Tableau 7) est la suivante :

$$-\Delta H \text{ (kJ/mol)} = 401{,}61\ n_N + 2092{,}2\ BO_{avg,N} + 13287\ E_{O,max} - 3148{,}5 \qquad (3.16)$$

Les descripteurs employés dans cette équation ont un sens chimique et peuvent d'ailleurs être reliés au processus de décomposition des composés nitroaromatiques. En effet, le nombre d'atomes d'azote n_N est lié à la présence de groupement NO_2, connus comme étant explosophores [40]. De plus, l'ordre de liaison moyen pour un atome d'azote ($BO_{avg,N}$) et l'indice de réactivité électrophile pour un atome d'oxygène ($E_{O,max}$) caractérisent la capacité des groupements nitro à se dissocier du reste de la molécule.

No.	X	±ΔX	t-test	Descripteurs
0	-3,15E+03	1,73E+02	-18,24	Intersection
1	4,02E+02	1,38E+01	29,02	Nombre de N
2	2,09E+03	1,11E+02	18,89	Ordre de liaison moyen pour un N
3	1,33E+04	1,76E+03	7,53	Indice d'électrophilicité maximum pour un O

Tableau 7 - Modèle à trois paramètres (3.16) pour la chaleur de décomposition à partir de structures calculées au niveau PBE0/6-31+G(d,p) (R^2=0,982; R^2_{cv}=0,974; F=330; s^2=963,17)

Cette équation à trois descripteurs ne donne que de faibles déviations par rapport aux valeurs expérimentales et présente une valeur de R^2 proche de la meilleure équation à quatre paramètres (0,982 contre 0,989).

$$-\Delta H\ (kJ/mol) = 3{,}98\ W - 663{,}16\ BO_{max,C} - 16{,}30\ C_{v,tr} + 6024{,}4\ FNSA3 + 1781{,}3 \qquad (3.17)$$

Dans cette dernière, l'indice de Wiener (W), qui caractérise le degré de ramification des molécules, peut être interprété comme une expression du degré de substitution du cycle aromatique. L'ordre de liaison maximum pour un atome de carbone ($BO_{max,C}$) est un descripteur de la liaison reliant le cycle aromatique au groupement nitro, qui est le premier à se dissocier dans la décomposition des nitroaromatiques. Les deux derniers descripteurs (la capacité calorifique translationnelle par atome à 300K, $C_{v,tr}$, et la portion relative de la surface moléculaire chargée négativement, (FNSA3) sont plus difficiles à interpréter d'un point de vue mécanistique. Statistiquement, ces deux derniers descripteurs sont les moins significatifs au sens du *t*-test de Student (voir Tableau 8). Ils peuvent donc être supposés superflus et représentent une sur-paramétrisation du système.

No.	X	±ΔX	t-test	Descripteurs
0	1,78E+03	1,19E+02	14,92	Intersection
1	3,98E+00	2,45E-01	16,21	Indice de Wiener
2	-6,63E+02	6,55E+01	-10,13	Ordre de liaison maximum pour un C
3	-1,63E+01	2,64E+00	-6,16	Capacité calorifique translationnelle (300K) par atome
4	6,02E+03	1,48E+03	4,08	Surface relative chargée négativement FNSA-3 [Zefirov]

Tableau 8 - Modèle à quatre paramètres (3.17) pour la chaleur de décomposition à partir de structures calculées au niveau PBE0/6-31+G(d,p) (R^2=0,989; R^2_{cv}=0,984; F=397; s^2=605,60)

L'équation à trois descripteurs est proche de celle à quatre paramètres du point de vue des corrélations. De plus, l'interprétation de ses descripteurs est plus claire, puisque deux des quatre descripteurs de celle à quatre paramètres (Eq. 3.17) sont assez difficiles à relier à la stabilité thermique. Le modèle à trois paramètres (Eq. 3.16), dont la corrélation est présentée en Figure 31, est donc recommandé puisqu'il associe des descripteurs ayant un sens chimique tout en assurant une bonne corrélation sans sur-paramétrisation.

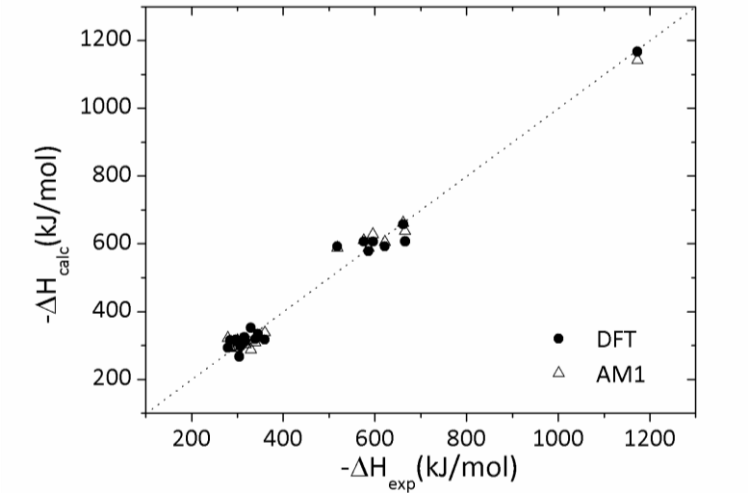


Figure 31 – Chaleurs de décompositions (en kJ/mol) prédites (AM1 et DFT) et expérimentales

2.3.2.2- Influence du niveau de calcul

Les descripteurs calculés à l'aide de Codessa ont, bien entendu, déjà été utilisés par le passé. Cela dit, les structures moléculaires sont généralement obtenues à un niveau de calcul moins important que celui considéré jusqu'à présent. En effet, des calculs semi-empiriques sont le plus souvent employés [98,177-181] puisque, étant plus approchés, ils nécessitent des temps de calcul moins importants. Nous sommes alors en droit de nous demander quel niveau de calcul est le plus adéquat dans notre cas.

Plusieurs travaux ont d'ailleurs récemment mis en garde contre l'utilisation de niveaux de calculs trop approchés [182]. Par exemple, une diminution de la corrélation comprise entre 0,02 et 0,04 a été observée pour la prédiction du coefficient de partition octanol/eau [183,184] ainsi que pour les solubilités aqueuses [185-187] lorsque la méthode semi-empirique AM1 est préférée à l'approche DFT. Cela dit, Puzyn recommande l'utilisation d'autres approches semi-empiriques (PM6, RM1) pour le même type de propriétés [188]. En tous cas, une attention toute particulière doit être portée à cette étape de détermination structurale car les conséquences sur les données statistiques peuvent être très importantes, comme cela a pu être démontré par Young [189].

Pour cette raison, le jeu complet des descripteurs utilisés précédemment ont été calculés à partir des structures moléculaires optimisées au niveau semi-empirique AM1, très couramment employés dans la littérature [190,191]. En particulier, Katritzky a utilisé des descripteurs à ce niveau de théorie pour la prédiction du point d'éclair [177,178]. Il va donc s'agir d'estimer la pertinence de ce niveau de calcul pour la prédiction de la stabilité thermique des composés nitroaromatiques.

Un modèle à la fois performant statistiquement et chimiquement sensé a été obtenu précédemment à partir des structures calculées au niveau DFT. La même analyse a été réalisée à un niveau de théorie moins coûteux en termes de temps de calcul. Les descripteurs ont été obtenus à partir de structures moléculaires optimisées à l'aide de la paramétrisation AM1. La Figure 30 représente la progression du R^2 en fonction du nombre

de descripteurs, à la fois aux niveaux de calcul AM1 et DFT. Il est tout d'abord à noter que, quelque soit le nombre de descripteurs employé, le modèle AM1 présente une corrélation légèrement plus faible qu'avec les structures DFT.

Le nombre optimal de descripteurs pour ce niveau de calcul est déterminé de manière identique à l'analyse au niveau DFT. Le modèle suivant, à 4 descripteurs, est alors obtenu:

$$-\Delta H \text{ (kJ/mol)} = 2{,}70\ W - 124{,}03\ {}^{1}IC - 4389{,}3\ Q_{C,max} + 40022\ Q_{O,max} + 3050{,}7 \qquad (3.18)$$

Ce modèle, plus détaillé dans le Tableau 9, est principalement constitué autour de l'indice de Wiener, qui caractérise, comme proposé précédemment, le nombre de substituants. Ce descripteur présente d'ailleurs, à lui seul, une corrélation de 0,81 avec la chaleur de décomposition. Les trois autres descripteurs sont moins significatifs, au regard des *t*-tests (entre 4 et 8, contre 20 pour l'indice de Wiener). Cela dit, les charges partielles maximales pour des atomes de carbone et d'oxygène, $Q_{C,max}$ et $Q_{O,max}$, calculées selon la méthode empirique de Zefirov, peuvent être attribuées au groupement nitro. L'indice topologique ${}^{1}IC$ (contenu d'information moyen d'ordre 1) est, tout comme W, supposé caractériser le taux de branchement de la molécule.

No.	X	±ΔX	t-test	Descripteurs
0	3,05E+03	5,17E+02	5,90	Intersection
1	2,70E+00	1,35E-01	20,05	Indice de Wiener
2	-1,24E+02	2,93E+01	-4,23	Indice de contenu d'information (ordre 1)
3	-4,39E+03	5,29E+02	-8,30	Charge partielle maximale pour un C [Zefirov]
4	4,00E+04	8,05E+03	4,97	Charge partielle maximale pour un O [Zefirov]

Tableau 9 – Modèle à quatre paramètres (3.18) pour la chaleur de décomposition à partir de structures calculées au niveau AM1 (R^2=0,984; R^2_{cv}=0,967; F=265; s^2=901,98)

D'un point de vue statistique, si cette équation est également significativement corrélée avec les données expérimentales que le modèle DFT (R^2 = 0,984 contre 0,982), il est très légèrement moins robuste au regard de la validation interne (R^2_{cv}=0,967 contre 0,974). Cela dit, tous les descripteurs du modèle DFT (Eq. 3.18) sont relatifs au groupement nitro alors que l'indice de Wiener, sur lequel repose principalement le modèle AM1, est un indice topologique uniquement construit à partir de la table de connectivité de la molécule et est donc moins interprétable du point de vue mécanistique.

D'ailleurs, en considérant le jeu de descripteurs employé pour le modèle AM1, mais à partir des structures DFT, une équation similaire à l'équation (3.18) est obtenue, aussi bien au niveau des coefficients de la régression qu'au niveau des corrélations (R^2=0,984).

$$-\Delta H \text{ (kJ/mol)} = 2{,}69\ W - 122{,}78\ {}^{1}IC - 4406{,}7\ Q_{C,max} + 40652\ Q_{O,max} + 3091{,}1 \qquad (3.19)$$

Cela signifie que les descripteurs de ce modèle ne sont que peu influencés par le niveau de calcul. Néanmoins, si l'analyse BMLR a favorisé l'équation (3.18), présentant des descripteurs plus sensés chimiquement, c'est que cette dernière présente un critère de Fisher F plus important (F=330 contre 266). En passant du niveau AM1 au niveau DFT, si les performances du modèle en termes de corrélation ne sont pas fondamentalement améliorées, sa compréhension chimique l'est. Le modèle semi-empirique semble donc, à

première vue, présenter une corrélation suffisamment élevée pour une utilisation en tant qu'outil de criblage. En fait, ces performances sont dues à la bonne corrélation de l'indice de Wiener, un descripteur topologique ne nécessitant pas un important niveau de théorie. Le modèle DFT assure une fiabilité similaire (comme montré en Tableau 10) avec un nombre de descripteurs moins élevé (d'où un moindre risque de sur-paramétrisation) et, en plus, inclut des descripteurs permettant une interprétation chimique plus robuste.

Composé	Exp [158]	DFT (3.18)	AM1 (3.19)
nitrobenzène	339	319	309
1,2-dinitrobenzène	518	592	588
1,3-dinitrobenzène	586	578	587
1,4-dinitrobenzène	622	592	605
2-nitrotoluène	329	352	288
3-nitrotoluène	284	315	295
4-nitrotoluène	318	313	303
2,6-dinitrotoluène	576	606	610
3,4-dinitrotoluène	666	607	637
2,4-dinitrotoluène	596	606	627
2-nitroaniline	307	298	316
3-nitroaniline	314	309	315
4-nitroaniline	279	293	321
Acide 2-nitrobenzoïque	297	317	298
Acide 3-nitrobenzoïque	298	312	292
Acide 4-nitrobenzoïque	304	267	309
2-nitrophénol	345	334	322
3-nitrophénol	316	324	311
4-nitrophénol	300	314	314
1-chloro-4-nitrobenzène	360	317	338
2,4-dinitrophénol	662	657	662
2,4,6-trinitrophénol	1173	1167	1142
	MAE (%)	5,4	4,8

Tableau 10 – Chaleurs de décomposition (en kJ/mol) expérimentales et prédites à partir des modèles DFT et AM1

2.3.3- Modèle QSPR pour une base de données étendue

Jusqu'à présent, une base de données de 22 molécules a été employée. Sa taille était suffisante pour estimer la pertinence des descripteurs employés, mais insuffisante pour réaliser la validation complète d'un modèle en incluant les validations internes, externes et la définition du domaine d'applicabilité. Pour ce faire, une base plus grande est nécessaire afin, en particulier, de pouvoir diviser le jeu de données en deux : un jeu d'entraînement et un jeu de validation. C'est ce qui a été réalisé sur une nouvelle base de donnée, plus importante (77 molécules, en Figure 32), elle aussi extraite de la littérature [157]. Les données relatives à cette dernière ont été obtenues avec un dispositif expérimental différent de la base de données précédente [158] : avec un taux de chauffage de 10°C/min à l'aide d'un appareillage de DSC pressurisé au sein duquel des échantillons de 1 à 2 mg sont placés dans des cellules en aluminium ouverte permettant l'évacuation des gaz.

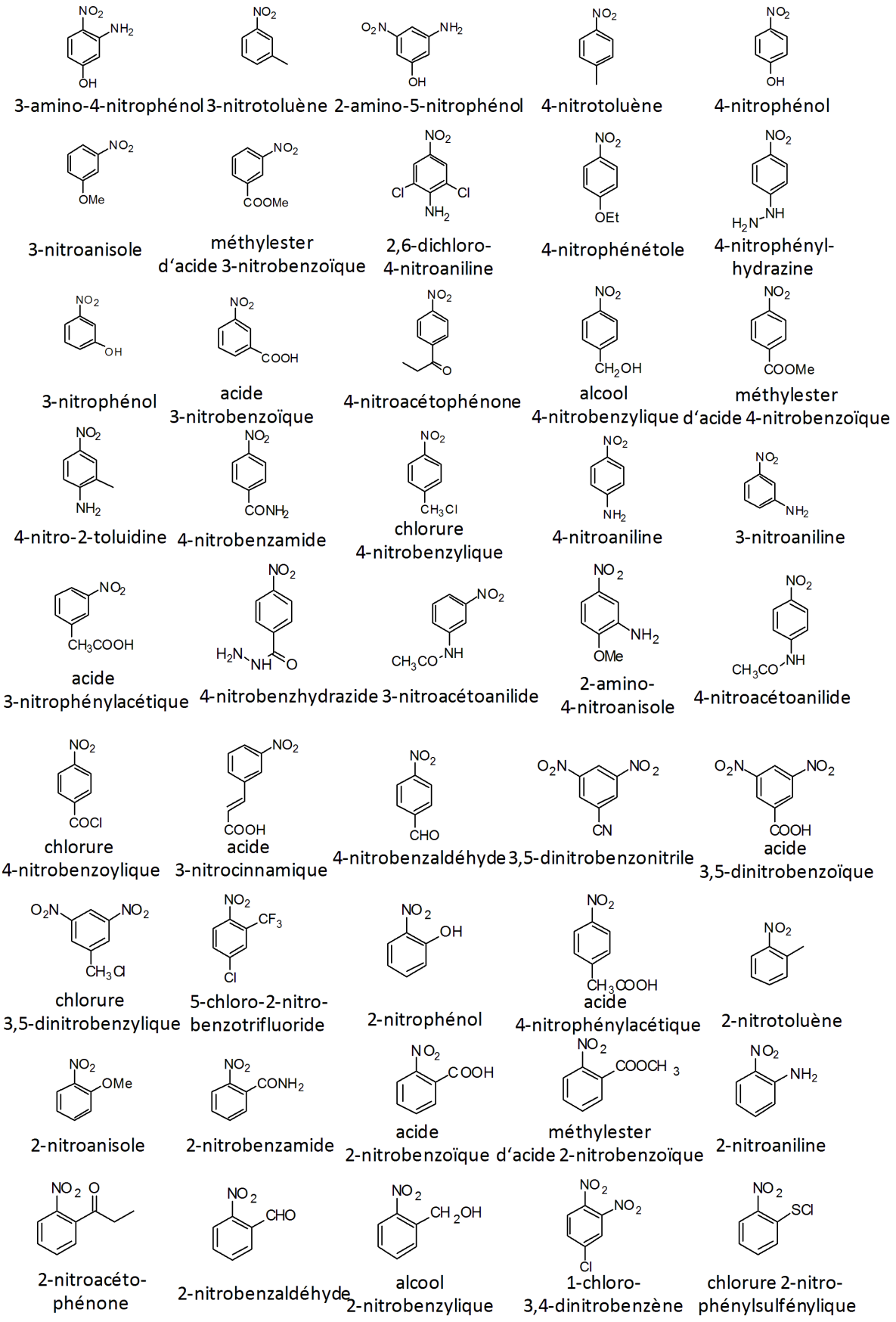


Figure 32 – 77 molécules de la base de données de Ando [157]

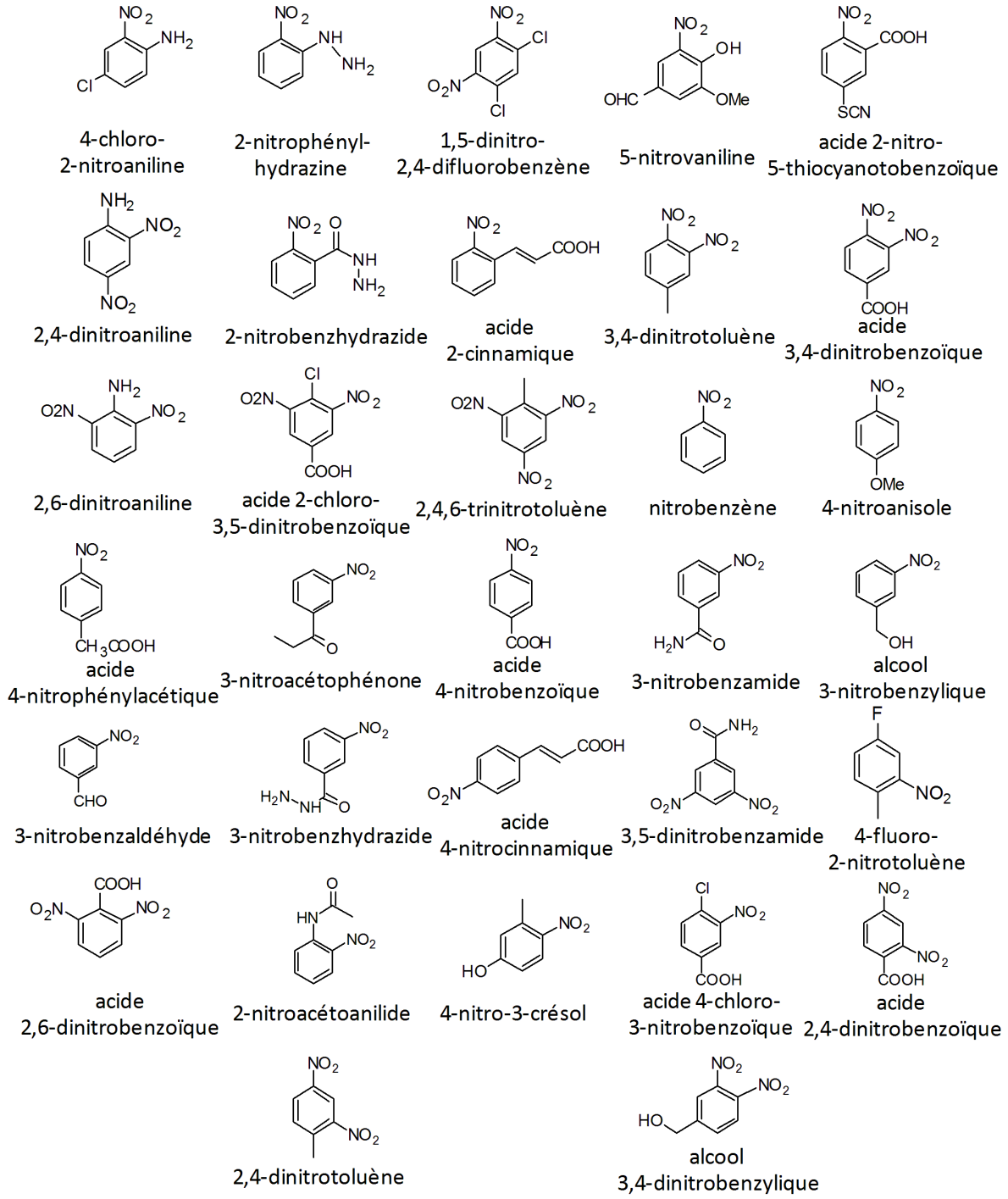


Figure 32 – 77 molécules de la base de données de Ando [157] (suite)

Par conséquent, ces deux bases ne peuvent être regroupées. D'ailleurs, pour les 16 composés communs entre les deux bases de données, la corrélation entre les deux jeux de données n'est que de 0,72, comme montré en Figure 33. C'est pour cette raison que ces nouvelles données sont considérées seules et non en addition de la base précédente.

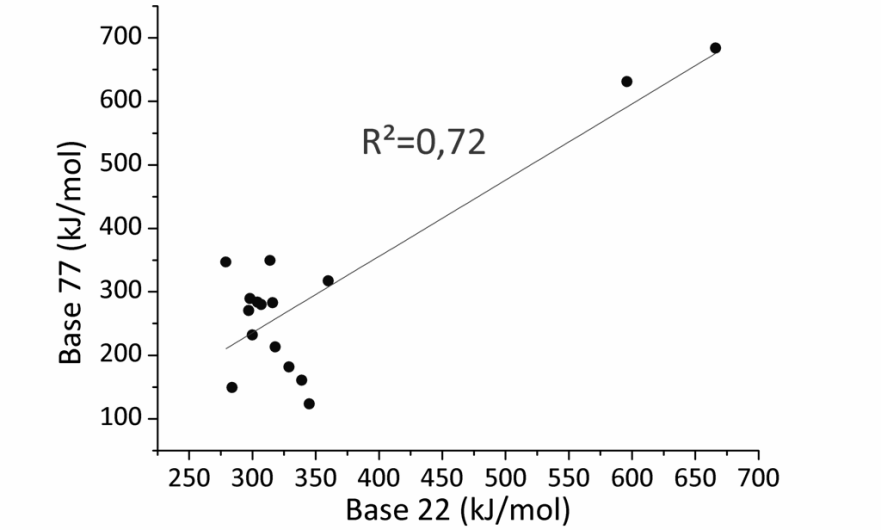


Figure 33 – Corrélation entre les données de Duh [158] (Base 22) et de Ando [157] (Base 77)

Après avoir tenté de développer des modèles prédictifs pour la base de données complète, nous verrons qu'il est intéressant de séparer les molécules sur la base d'informations mécanistiques, puisque les composés présentant un substituant en position ortho vis-à-vis de NO_2 peuvent présenter un chemin de décomposition particulier[58,172,173]. Pour cette analyse, plus de 300 descripteurs ont été calculés à partir de structures optimisées au niveau PBE0/6-31+G(d,p).

Analyse sur la base complète

Dans une première étape, l'intégralité de la base de données a été considérée. Cette dernière est suffisamment grande pour être divisée en deux : un jeu d'entraînement et un jeu de validation. Le premier sera utilisé pour développer le modèle QSPR et le second pour en évaluer le pouvoir prédictif. Afin de conserver des distributions similaires dans tous les jeux de données, les molécules de la base de données ont été classées par ordre croissant de chaleur de décomposition expérimentale. Le jeu de validation est alors composé par des molécules uniformément distribuées : une toute les 4 molécules, à partir de celle classée 3ème. Ainsi, les distributions des jeux d'entraînement et de validation (respectivement 58 et 19 molécules) restent similaires à celle de la base complète, comme montré en Figure 34.

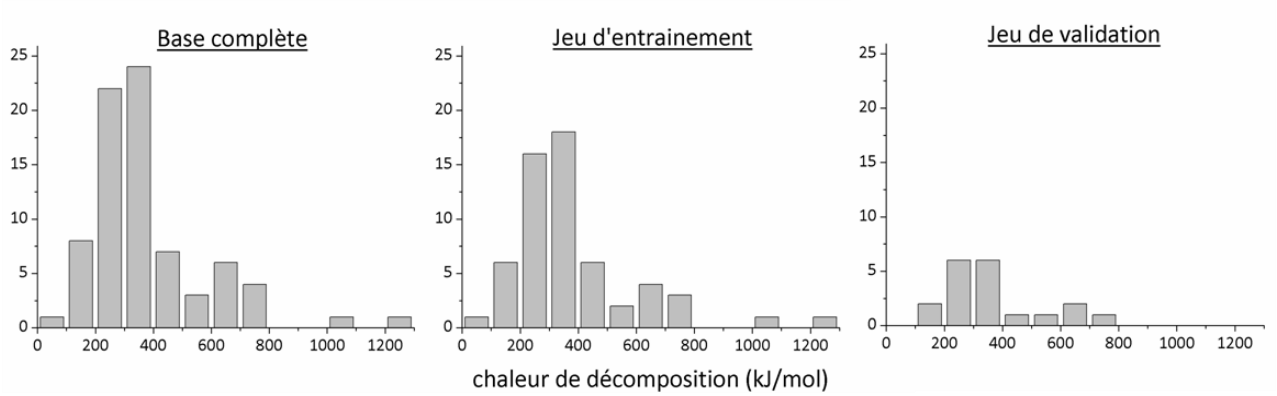


Figure 34 – Distributions des valeurs de chaleur de décomposition de la base complète et des jeux d'entraînement et de validation.

Une analyse BMLR a été réalisée sur le jeu d'entraînement (voir Tableau 11). Le meilleur compromis entre la corrélation et le degré de paramétrisation (nombre de descripteurs) est obtenu pour le modèle à quatre descripteurs suivant.

$$-\Delta H \text{ (kJ/mol)} = -282{,}3 + 333{,}5\, n_{NO2} - 1214{,}5\, E_{C,avg} + 7{,}4\, \alpha - 275{,}6\, {}^{0}IC_{avg} \qquad (3.20)$$

où n_{NO2} est le nombre de groupements nitro, $E_{C,avg}$ est l'indice de réactivité mono-électronique pour un atome de carbone et ${}^{0}IC_{avg}$ est un indice topologique de contenu d'information (à l'ordre 0).

molécules	Exp [157]	Eq. (3.20)	Eq. (3.21)	Eq. (3.22)
composés non-ortho				
2-amino-4-nitrophénol	130	238	173	-
3-nitrotoluène	149	238	212	-
2-amino-5-nitrophénol	153	201	239	-
4-nitrotoluène	213	253	192	-
4-nitrophénol	232	183	235	-
3-nitroanisole	243	223	288	-
méthylester d'acide 3-nitrobenzoïque	256	334	277	-
2,6-dichloro-4-nitroaniline	264	394	284	-
4-nitrophénétole	270	347	249	-
4-nitrophénylhydrazine	277	389	279	-
3-nitrophénol	283	165	227	-
acide 3-nitrobenzoïque	289	267	372	-
4-nitroacétophénone	291	348	343	-
alcool 4-nitrobenzylique	292	223	272	-
méthylester d'acide 4-nitrobenzoïque	302	329	264	-
4-nitro-2-toluidine	306	287	315	-
4-nitrobenzamide	319	275	321	-
chlorure 4-nitrobenzylique	337	675	333	-
4-nitroaniline	347	298	308	-
3-nitroaniline	350	302	317	-
acide 3-nitrophénylacetique	358	397	347	-
4-nitrobenzhydrazide	362	415	335	-
3-nitroacétoanilide	369	289	394	-
2-amino-4-nitroanisole	375	339	325	-
4-nitroacétoanilide	387	343	372	-
chlorure 4-nitrobenzoylique	408	463	303	-
acide 3-nitrocinnamique	414	314	417	-
4-nitrobenzaldéhyde	421	495	394	-
3,5-dinitrobenzonitrile	654	699	698	-
acide 3,5-dinitrobenzoique	674	658	679	-
chlorure 3,5-dinitrobenzylique	711	682	673	-

Tableau 11 - Chaleurs de décomposition (en kJ/mol) expérimentales et prédites pour le jeu d'entraînement

	composés ortho			
5-chloro-2-nitrobenzotrifluoride	40	96	-	7
2-nitrophénol	123	176	-	172
acide 2-nitrophenylacétique	175	288	-	266
2-nitrotoluène	182	223	-	211
2-nitroanisole	230	203	-	276
2-nitrobenzamide	256	317	-	341
acide 2-nitrobenzoïque	271	212	-	247
méthylester d'acide 2-nitrobenzoïque	274	284	-	285
2-nitroaniline	280	267	-	284
2-nitroacétophénone	308	243	-	244
2-nitrobenzaldéhyde	318	196	-	351
alcool 2-nitrobenzylique	319	350	-	239
1-chloro-3,4-dinitrobenzène	342	358	-	331
chlorure 2-nitrophenylsulfénylique	344	302	-	276
4-chloro-2-nitroaniline	349	263	-	312
2-nitrophénylhydrazine	381	448	-	298
1,5-dinitro-2,4-difluorobenzène	439	396	-	483
5-nitrovaniline	450	448	-	491
acide 2-nitro-5-thiocyanotobenzoïque	492	718	-	559
2,4-dinitroaniline	597	520	-	694
2-nitrobenzhydrazide	598	550	-	463
acide 2-nitrocinnamique	600	620	-	606
3,4-dinitrotoluène	684	658	-	626
acide 3,4-dinitrobenzoïque	701	706	-	803
2,6-dinitroaniline	719	713	-	619
acide 2-chloro-3,5-dinitrobenzoïque	1023	766	-	937
2,4,6-trinitrotoluène	1223	1149	-	1220
	R^2	0,84	0,90	0,94
	MAE (%)	23	12	17
	R^2_{cv}	0,81	0,86	0,91

Tableau 11 - Chaleurs de décomposition (en kJ/mol) expérimentales et prédites pour le jeu d'entraînement (suite)

Avant de parler de la fiabilité du modèle, il est à noter que le modèle inclut deux descripteurs qui peuvent être reliés à la stabilité thermique des nitroaromatiques. En effet, la quantité d'énergie libérée durant la décomposition est principalement reliée à la dissociation des liaisons au sein de la molécule et en particulier à la perte de groupements tels que le groupement nitro. En conséquence, la présence du nombre de ces groupements dans le modèle est très satisfaisante. De plus, $E_{C,avg}$ caractérise la réactivité des atomes de carbone dans la molécules. En particulier, le carbone lié au groupement NO_2 est supposé d'une importance toute particulière dans le processus de décomposition.

Ce modèle présente une corrélation significative avec un R^2 de 0,84 (cf. Figure 35) et est stable au regard du R^2_{cv} (0,81). Il est moins performant que celui obtenu précédemment sur la base de 22 molécules (Eq. 3.17) qui présentait un R^2 de 0,98. Néanmoins, la distribution des données expérimentales n'était pas aussi homogène, ce qui, associé au nombre de données prises en compte, augmente la confiance en ce dernier modèle.

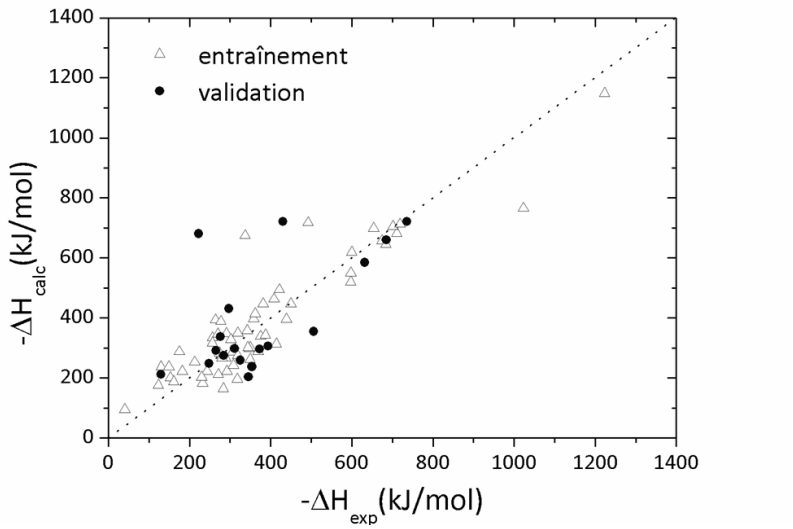


Figure 35 - Chaleurs de décomposition (en kJ/mol) expérimentales et prédites par le modèle (3.20) pour les molécules des jeux d'entraînement et de validation

Afin d'évaluer le pouvoir prédictif du modèle, les chaleurs de décomposition des composés du jeu de validation ont été calculées (cf. Tableau 12). Le modèle (Eq. 3.25) ne présente qu'une faible corrélation avec une déviation moyenne de 32% avec l'expérience (R^2_{ext}=0,43). En particulier, des déviations importantes sont obtenues pour le 4-fluoro-2-nitrotoluène et les acides 2,4- et 2,6-dinitrobenzoïques (respectivement 64, 83 et 206 %).

molécules	exp [157]	Eq.(3.20)	Eq.(3.21)	Eq.(3.22)	Eq.(3.2) [16]	Eq. (3.17)
		composés non-ortho				
Nitrobenzène	161	188	202	-	314	319
4-nitroanisole	248	248	283	-	314	312
acide 4-nitrophenylacétique	265	291	341	-	314	314
3-nitroacétophénone	276	338	364	-	314	308
acide 4-nitrobenzoïque	284	275	332	-	314	267
3-nitrobenzamide	311	298	334	-	314	39
alcool 3-nitrobenzylique	325	204	258	-	314	352
3-nitrobenzaldéhyde	373	306	389	-	314	339
3-nitrobenzhydrazide	430	355	344	-	627	631
acide 4-nitrocinnamique	506	585	414	-	314	413
3,5-dinitrobenzamide	736	721	687	-	627	560
		composés ortho				
4-fluoro-2-nitrotoluène	129	212	-	211	314	354
acide 2,6-dinitrobenzoïque	222	681	-	423	627	607
2-nitroacétoanilide	297	431	-	472	314	80
4-nitro-3-crésol	345	238	-	203	314	319
acide 4-chloro-3-nitrobenzoïque	354	297	-	463	314	325
acide 2,4-dinitrobenzoïque	394	722	-	553	314	280
2,4-dinitrotoluène	632	660	-	607	314	205
alcool 3,4-dinitrobenzylique	685	645	-	683	627	605
	R^2_{ext}	0,43	0,84	0,42	0,53	0,37
	MAE (%)	32	18	44	36	46

Tableau 12 - Chaleurs de décomposition (en kJ/mol) expérimentales et prédites pour le jeu de validation

De plus, si l'on définit le domaine d'applicabilité (AD) de ce modèle (tel qu'indiqué dans la partie méthodologique), aucune de ces molécules n'est exclue, comme le montre le Tableau 13. Le pouvoir prédictif du modèle est donc faible au sein même de son AD.

	Entraînement					Validation				
Eq.	R^2	N_{in}	N_{out}	$\%_{in}$	R^2_{in}	R^2	N_{in}	N_{out}	$\%_{in}$	R^2_{in}
3.20	0,85	55	3	95	0,82	0,43	19	0	100	0,43
3.21	0,90	29	2	94	0,88	0,84	10	1	91	0,86
3.22	0,94	26	1	96	0,94	0,42	8	0	100	0,42
3.2 [16]	0,98	19	0	100	0,98	0,53	17	2	89	0,54
3.17	0,98	21	1	95	0,99	0,37	15	4	79	0,24

N_{in}= nombre de molécules dans l'AD — $\%_{in}$= pourcentage de molécules dans l'AD
N_{out}= nombre de molécules hors de l'AD — R^2_{in}= corrélation pour les molécules dans l'AD

Tableau 13 – Analyse des domaines d'applicabilité (AD) des modèles QSPR proposés

Pour expliquer un si faible pouvoir prédictif, plusieurs causes peuvent être avancées : incertitudes des données expérimentales, présence de mécanismes sous-jacents particuliers.

En effet, pour développer un modèle QSPR fiable, les données expérimentales et leur incertitude est un point critique. Ici, les chaleurs de décomposition ont été obtenues à l'aide d'un dispositif DSC à cellule ouverte permettant l'évacuation des gaz. On peut donc supposer des effets d'évaporation pouvant dégrader les incertitudes de mesures et, en conséquence, le modèle. Cela dit, l'incertitude expérimentale n'est pas suffisante pour expliquer un pouvoir prédictif aussi faible que celui observé ici.

Un autre facteur important affectant le modèle est lié à la réactivité chimique. En effet, la décomposition des nitroaromatiques est un processus complexe. Alors que la rupture directe de la liaison carbone-nitro est le mécanisme le plus favorable pour les composés ne présentant pas de substituants en position ortho par rapport au groupement NO_2, des chemins spécifiques peuvent intervenir en présence de substituants en cette position. C'est notamment ce que nous avons pu observer pour les dérivés du 2-nitrotoluène (voir chapitre 4) [172,173]. Pour cette raison, des modèles séparés ont été mis en place selon que les composés présentent ou non un substituant en position ortho par rapport à NO_2. Ces modèles ont été développés suivant la même méthodologie que précédemment en conservant la même division entre jeux d'entraînement et de validation. D'ailleurs, Storm [192] a déjà, par le passé, considéré une distinction similaire dans la collecte des données de sensibilité à l'impact qui seront employées plus loin.

Analyse des composés "non-ortho"

L'analyse BMLR des 31 molécules du jeu d'entraînement ne présentant pas de substituants en position ortho par rapport à NO_2, donne un modèle à quatre descripteurs.

$$-\Delta H\ (kJ/mol) = 0{,}8\ G - 3{,}8\ WPSA1 - 4255{,}1\ Q_{max} + 26{,}8\ RPCS - 251{,}2 \qquad (3.21)$$

où G est l'indice gravitationnel, WPSA1 la surface pondérée chargée positivement (au sens des charges de Mulliken), Q_{max} est la charge partielle maximale au sein de la molécule (calculée selon la méthode de Gasteiger) et RPCS est la surface relative chargée positivement (au sens des charges de Zefirov).

Si Q_{max} peut être reliée aux atomes d'azote des groupements nitro, les autres descripteurs sont plus difficiles à relier au processus de décomposition. G est un descripteur topologique qui caractérise la forme moléculaire et la distribution de masse au sein de la molécule. Les descripteurs WPSA1 et RPCS sont des descripteurs caractérisant la répartition de la charge en surface (CPSA, pour *charged partial surface area*), développés pour représenter les caractéristiques responsables des interactions polaires entre molécules. Néanmoins, ces derniers sont reliés, dans un certain sens, à la distribution de la charge au sein de la molécule qui est elle-même influencée par les propriétés électroniques au niveau des groupements nitro. Il n'est malgré tout pas aisé d'en extraire directement des informations sur le mécanisme de décomposition. Comme le montre la Figure 36, le modèle est fortement corrélé (R^2=0,90), la déviation moyenne (12%) étant proche de l'incertitude expérimentale acceptée de 5-10% [157].

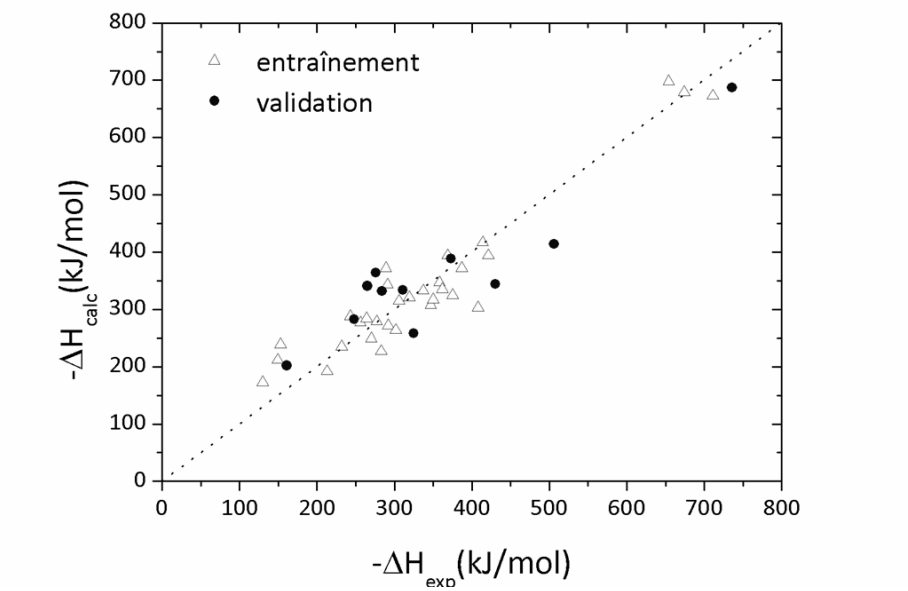


Figure 36 – Chaleurs de décomposition (en kJ/mol) prédites (Eq. 3.21) et expérimentales pour une série de composés nitroaromatiques non substitués en position ortho

La validation interne est elle aussi satisfaisante : R^2_{cv}=0,86. De plus, une bonne prédictivité du modèle est estimée via le jeu de validation (11 molécules) puisque que R^2_{ext}=0,84, correspondant à une déviation moyenne de 18%, ce qui reste satisfaisant au regard de l'incertitude expérimentale. Le pouvoir prédictif est même encore plus grand si l'on ne considère que le domaine d'applicabilité proprement dit du modèle : R^2_{in}=0,86 (voir Tableau 13).

Analyse des composés "ortho"

Pour les composés ortho, le modèle obtenu à partir des 27 molécules du jeu d'entraînement, est une autre équation à quatre paramètres.

$$-\Delta H\ (kJ/mol) = 4{,}1\ PNSA1 - 3298{,}8\ RPCG - 56228\ N_{C,min} - 1245{,}5\ S_{ZX/ZX} + 1117{,}8 \qquad (3.22)$$

où PNSA1 et RPCG sont respectivement la surface partielle chargée négativement et la charge positive relative, calculées selon la méthode de Zefirov, $N_{C,min}$ est l'indice minimal de

réactivité nucléophile pour un atome de carbone et $S_{ZX/ZX}$ est une projection de la forme moléculaire dans une direction particulière de l'espace. Dans ce modèles, le seul descripteur pouvant être relié à la stabilité thermique est $N_{C,min}$, caractérisant ainsi la réactivité des groupements nitro.

Les performances du modèle (Eq. 3.27) sont résumées dans le Tableau 11. La corrélation avec les données expérimentales est importante (R^2=0,94) avec une déviation moyenne de 17%. Bien que sa robustesse soit importante (R^2_{cv}=0,91), sa prédictivité est faible puisque les molécules "ortho" du jeu de validation (8 molécules) présentent une erreur moyenne de 44% (R^2_{ext}=0,42), comme indiqué dans le Tableau 12. Cela démontre l'importance de l'étape de validation externe du modèle.

En fait, un modèle performant était plus attendu pour les composés "non-ortho" que pour les composés "ortho" puisque, dans ce dernier cas, différents chemins de décomposition existent selon la nature du substituant en position ortho (nitro, hydroxyle, méthyle, etc.) alors qu'un mécanisme unique concerne les composés non-ortho, à savoir, la dissociation homolytique de la liaison C-NO_2 [58].

Comparaison avec les modèles antérieurs

Afin de comparer les modèles développés sur cette base de données avec les meilleurs précédemment développés, les chaleurs de décomposition des molécules du jeu de validation ont été également calculées à partir des équations (3.2) et (3.17) et apparaissent dans le Tableau 12. Aucune corrélation intéressante n'est démontrée avec ces modèles puisqu'ils présentent respectivement des déviations moyennes de 36% et 46% (R^2=0,53 et 0,37). Ceci peut être dû au manque d'homogénéité de la base de données employée pour obtenir ces modèles qui ne considère qu'une molécule trinitrée (voire aucune pour le modèle 3.2). De plus, ils ne distinguent pas les composés "ortho" des "non-ortho". Les pouvoirs prédictifs de ces modèles sont faibles même dans leurs domaines d'applicabilité respectifs : respectivement R^2_{in}=0,54 et 0,24, pour les modèles (3.2) et (3.17), comme montré dans le Tableau 13. Ces modèles présentent les mêmes limites en termes de prédiction que les équations (3.20) et (3.22).

Bilan

La connaissance des chemins de décomposition à l'échelle moléculaire est bien un point critique pour la prédiction de la stabilité thermique des composés nitroaromatiques. Sa prise en compte nous a permis de mettre en place le premier modèle QSPR (Eq. 3.21), qui s'applique à des composés dérivés du nitrobenzène non substitués en position ortho du groupement nitro, avec un prédictivité importante estimée sur un jeu de validation (R^2_{ext}=0,84) et dans un domaine d'applicabilité défini (R^2_{in}=0,86).

2.3.4- Utilisation d'autres outils de traitement de données

La méthodologie employée jusqu'à présent n'a pas permis de mettre en place un modèle prédictif pour l'intégralité des composés nitroaromatiques. D'autres outils de

traitement de données ont donc été employés afin d'étudier la possibilité de mettre en place un modèle plus global, et si possible plus prédictif, applicable à tous les dérivés du nitrobenzène, à partir de la base de données précédente de 77 molécules [157].

Sélection d'un jeu réduit de descripteurs

Dans un premier temps, la démarche employée a consisté, pour les 77 molécules de la base de données, à extraire parmi les plus de 300 descripteurs calculés précédemment un jeu réduit de descripteurs qui contient un maximum d'information en un minimum de paramètres. Non seulement cela permet de faciliter l'analyse en traitant moins de données en éliminant au maximum les informations redondantes, mais cela rend également l'interprétation phénoménologique des modèles finaux plus aisée.

Pour ce faire, l'analyse PLS a été employée sur l'intégralité de la base de données. L'objectif est de mettre en évidence les descripteurs permettant d'expliquer au mieux la variance de la propriété expérimentale, c'est-à-dire la variation de cette dernière au sein de la base de données. Dans une première étape, les descripteurs ne présentent pas de variance sur la base de données sont supprimées. Il reste alors 291 descripteurs. Ils sont représentés en Figure 37, en fonction de leurs poids wc_i sur les deux premières composantes principales. Plus un descripteur est situé proche des extrêmités des axes et plus il contribue à la composante principale correspondante. Il s'agit donc d'une représentation de l'espace des descripteurs en fonction de la variance expérimentale. Ainsi, deux descripteurs proches sur le graphe sont très corrélés entre eux : l'information est donc redondante. Il est à noter que les descripteurs ne présentant pas de variance significative sont exclus de l'analyse.

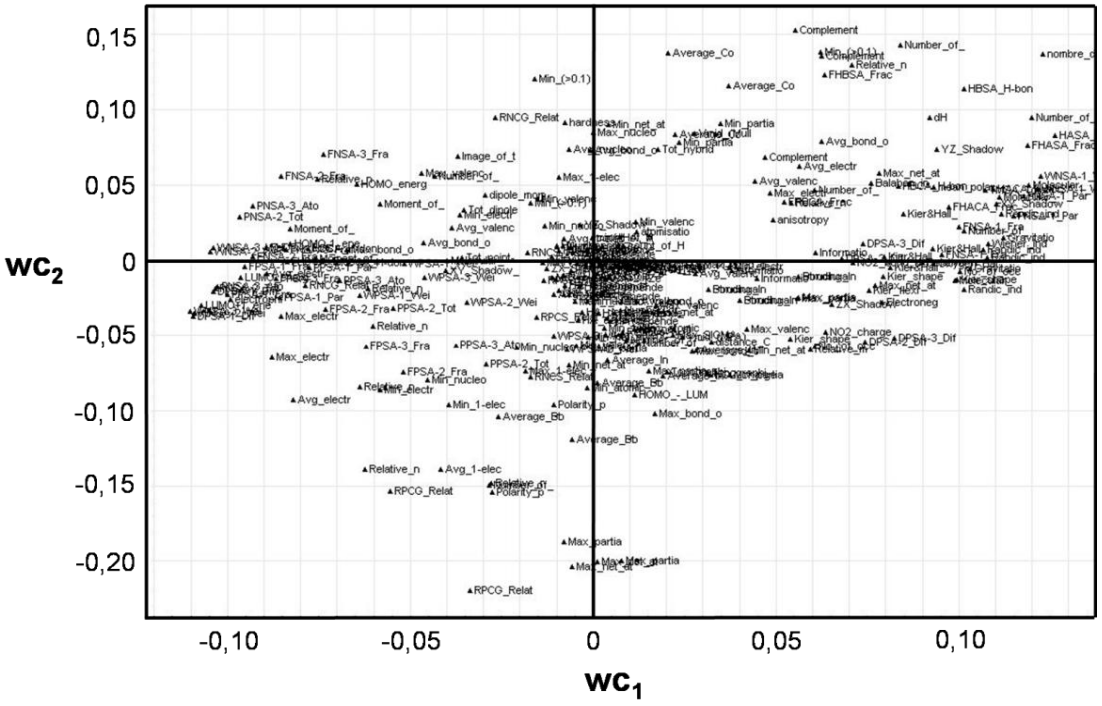


Figure 37 - Représentation du jeu complet des descripteurs dans le plan des deux premières composantes principales de l'analyse PLS

Notre objectif est donc de supprimer au maximum les descripteurs qui n'expliquent pas la variance expérimentale ainsi que de réduire la redondance d'information (descripteurs corrélés entre eux). Dans un premier temps, nous avons donc supprimé par étapes successives les descripteurs contribuant le moins aux modèles PLS en termes de VIP (*Variable Importance in the Projection*), un indice qui traduit l'importance des descripteurs sur le modèle. Au fur et à mesure de la réduction, la robustesse du modèle (évaluation par validation interne R^2_{cv}) augmente, ce qui signifie que l'on réduit ainsi l'inclusion de descripteurs redondants ou non-informatifs qui perturbent le modèle. Nous obtenons un jeu de descripteurs peut être ainsi réduit à 44 descripteurs sans perte de robustesse (cf. Tableau ***14***), même si elle reste inférieure à celle observée pour les modèles précédemment développés.

nombre de descripteurs	R^2_x	R^2_y	R^2_{cv}
291	0,336	0,733	0,597
97	0,634	0,732	0,664
44	0,789	0,724	0,684
7	0,805	0,691	0,641

Tableau 14 - Données statistiques des modèles PLS en fonction du nombre de descripteurs

Pour réduire encore le jeu des descripteurs de manière efficace, il est nécessaire de s'intéresser de plus près aux descripteurs restants. Sur le graphique de la Figure 38, représentant les poids de la PLS avec ce jeu de 44 descripteurs on peut distinguer des zones où les descripteurs s'accumulent.

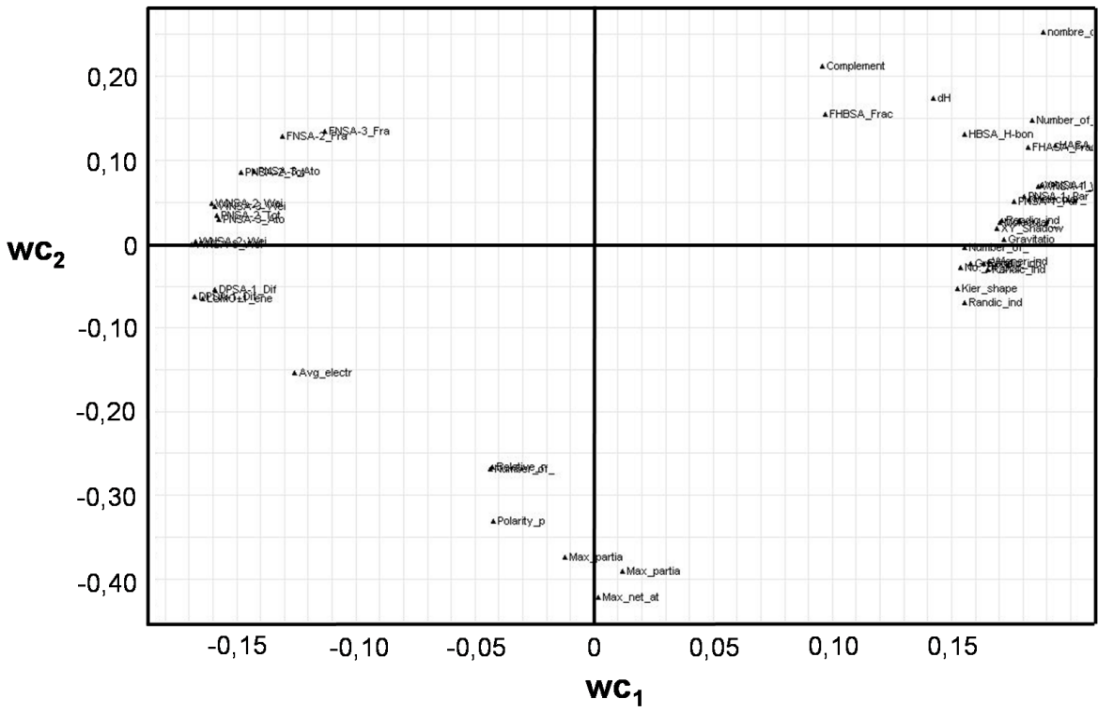


Figure 38 - Représentation du jeu réduit à 44 descripteurs dans le plan des deux premières composantes principales de l'analyse PLS

Il semble donc que ces zones soient importantes pour le modèle et donc pour la prédiction de la propriété. Mais tous les descripteurs dans ces zones caractérisent le même contenu d'information. Nous avons donc éliminé manuellement les informations redondantes en favorisant les descripteurs pouvant être reliés directement au processus chimique. Au final, nous pouvons extraire un jeu de 7 descripteurs, représenté en Figure 39, offrant une analyse PLS encore robuste comme montré dans le Tableau 14.

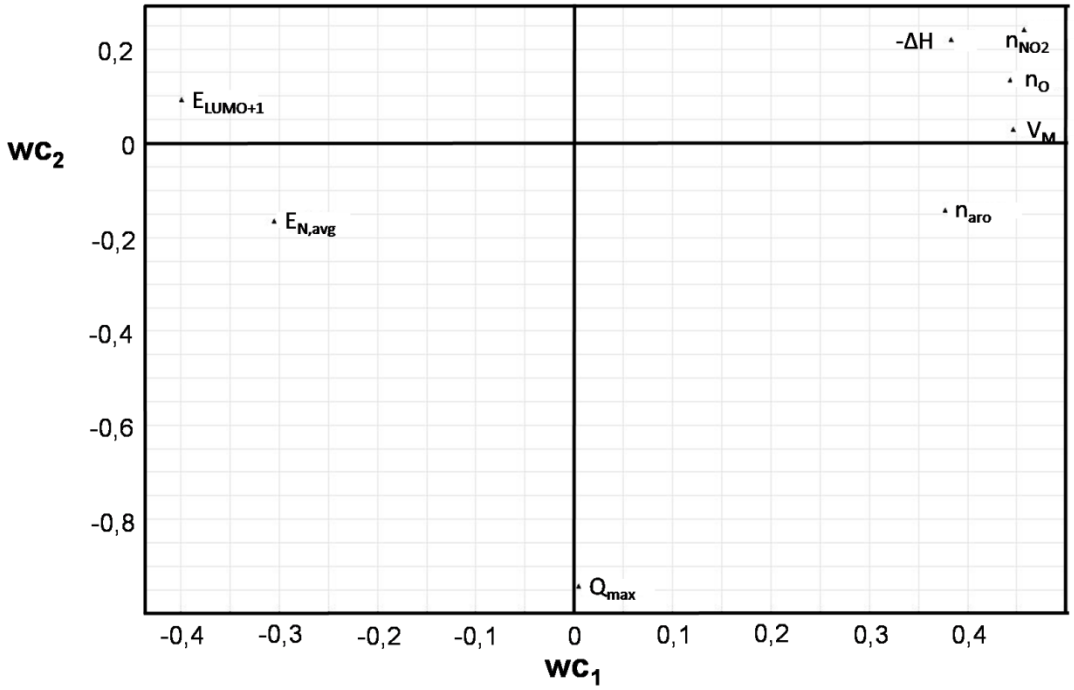


Figure 39 - Représentation du jeu réduit à 7 descripteurs et de la propriété (-ΔH) dans le plan des deux premières composantes principales de l'analyse PLS

Les 7 descripteurs sont suivants :

- → Nombre de NO_2 (n_{NO2})
- → Nombre d'atomes d'oxygène (n_O)
- → Nombre de liaisons aromatiques (n_{aro})
- → Volume moléculaire (V_M)
- → Indice d'électrophilicité moyen pour un atome d'azote ($E_{N,avg}$)
- → Charge atomique nette maximale (Q_{max})
- → Energie de la LUMO+1 (E_{LUMO+1})

Ce jeu de descripteurs est plutôt cohérent. En effet, le nombre de groupements nitro avait déjà été identifié précédemment comme un descripteur important de l'énergie libérée durant la décomposition.

Le nombre d'atomes d'oxygène est lui aussi un indicateur significatif du caractère énergétique de la molécule. En effet, l'oxygène de l'air est reconnu comme comburant. De même, la présence d'oxygène au sein même d'un matériau le rend énergétique par nature, ce dernier constitue alors un comburant interne au matériau. C'est d'ailleurs ce que caractérise également la balance en oxygène (Eq. 3.19).

Les types de liaison, telles que définies par CodessaPro, sont obtenues en fonction des ordres de liaison issus des calculs quantiques. Le nombre de liaison aromatique compte donc ici les liaisons d'ordre de liaison proche de 1,5 au sein de la molécule. De telles liaisons sont stabilisatrices. Donc, plus leur nombre est important et plus l'énergie disponible dans la molécule est importante.

Pour le volume moléculaire, l'interprétation est moins directe. Cela dit, plus le volume moléculaire est important, et plus la molécule est de taille importante. L'énergie contenue au sein de sa structure moléculaire pour la décomposer peut être supposée importante puisque le nombre de liaisons sera plus grand.

Le descripteur suivant est l'indice d'électrophilicité moyen pour un atome d'azote. Ces atomes sont au cœur des mécanismes de décomposition des nitroaromatiques [58]. D'ailleurs, la charge atomique maximale au sein de la molécule, Q_{max}, est également relative à ces atomes (comme le montre la Figure 40 pour deux molécules de la base de données).

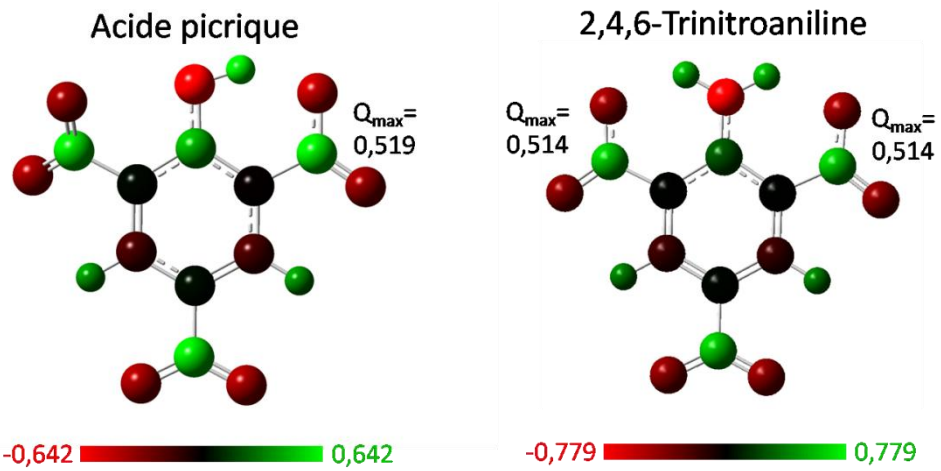


Figure 40 - Charges atomiques (calculées par analyse NPA) pour les molécules d'acide picrique ou de 2,4,6-trinitroaniline

S'il est plus difficile d'expliquer directement le rôle de la seconde orbitale moléculaire la plus basse vacante (LUMO+1), avec le processus de décomposition, le rôle des orbitales frontières en tant que descripteurs est souvent important pour les mécanismes radicalaires. Il traduit dans cette étude la stabilité globale des niveaux électroniques.

Il est à noter que les méthodes employées pour la sélection des descripteurs reposent sur des philosophies relativement différentes. Alors que la méthode BMLR précédemment employée identifiait des couples de descripteurs orthogonaux puis les incrémentait (en évitant toute sur-paramétrisation), l'utilisation des méthodes PLS part d'un nombre de descripteurs important (modèle initial sur-paramétré) et le réduit pour contenir un maximum de l'information du système tout en éliminant les redondances d'information jusqu'à avoir un modèle stable (non sur-paramétré). De plus, l'approche par projection présente l'avantage de permettre une élimination manuelle de descripteurs et ainsi de favoriser l'emploi de descripteurs interprétables d'un point de vue phénoménologique.

Mise en place des jeux d'entraînement et de validation

Afin d'évaluer le pouvoir prédictif des modèles, il est nécessaire de diviser notre base de données en deux : un jeu d'entraînement et un jeu de validation. Pour une évaluation correcte, il faut que le jeu de validation caractérise le même espace des données que le jeu d'entraînement. Précédemment, ce partage avait été réalisé au regard de la propriété, sur la base de la distribution des données expérimentales, en s'assurant que le jeu de validation soit bien représentatif du domaine défini par le jeu d'entraînement en mettant en place des distributions des données expérimentales similaires pour les différents jeux de données.

Une autre approche a été testée ici. Le partage a été réalisé à l'aide d'un tirage aléatoire. Il présente ainsi l'avantage de n'être influencé par aucune variable (descripteurs ou propriété) mais certains problèmes peuvent apparaître si le jeu de validation définit un espace trop différent de celui du jeu d'entraînement. Cela peut intervenir en particulier avec des distributions qui, comme ici, ne comprennent pas un grand nombre de molécules (plusieurs centaines).

Avec un seul tirage aléatoire, nous risquons d'être confrontés à ce type de problème. Pour s'en affranchir, une statistique sur 10 tirages aléatoires a été réalisée afin de s'assurer que la qualité du modèle n'est pas ni surestimée ni sous-estimée par ce partage. Les 10 jeux de validation ont donc été choisis de manière aléatoire avec un rapport constant (55 molécules d'entraînement et 22 de validation). Des analyses PCA et PLS ont été réalisées sur chacun des jeux d'entraînement puis la prédictivité de chaque modèle PLS a été évaluée à l'aide du jeu de validation correspondant. La différence entre les méthodes PCA et PLS réside dans la prise en compte de la propriété expérimentale. Ainsi, la PCA représente uniquement l'espace des descripteurs alors que la PLS prend en plus en compte celui de la propriété. Le Tableau 15 donne les paramètres obtenus pour les analyses PLS et PCA pour les différents partages réalisés.

partage	PCA		PLS			
	R^2_x	R^2_{cv}	R^2_x	R^2_y	R^2_{cv}	R^2_{ext}
1	0,80	0,48	0,79	0,64	0,52	0,75
2	0,80	0,47	0,78	0,56	0,43	0,91
3	0,81	0,49	0,80	0,62	0,51	0,78
4	0,83	0,55	0,83	0,69	0,62	0,78
5	0,81	0,53	0,79	0,61	0,44	0,81
6	0,80	0,46	0,80	0,74	0,65	0,64
7	0,80	0,47	0,79	0,69	0,59	0,72
8	0,81	0,49	0,80	0,69	0,61	0,67
9	0,82	0,50	0,82	0,70	0,61	0,65
10	0,84	0,59	0,83	0,69	0,61	0,78
$<R^2>$ (a)	0,81	0,50	0,80	0,66	0,56	0,75
ΔR^2 (b)	0,01	0,03	0,01	0,04	0,06	0,06

(a) Valeur moyenne (b) Déviation standard

Tableau 15 - Données statistiques pour 10 partages aléatoires entre jeux d'entraînement et de validation

Le partage entre entraînement et validation ne semble pas avoir de grande incidence sur les performances des modèles obtenus puisque les écarts types des coefficients statistiques ne sont pas particulièrement importantes, en particulier ΔR^2_{ext}=0,06.

Dans le cadre de l'analyse PLS, les corrélations respectives sur la matrice des descripteurs (R^2_x) et de la propriété (R^2_y) sont respectivement de 0,80 (±0,01) et 0,66 (±0,04), ce qui est acceptable. Bien que la corrélation avec la propriété ne soit pas particulièrement importante, la représentation de l'espace chimique que forme le jeu d'entraînement est correcte et le modèle suffisamment stable pour engendrer une bonne prédictivité. En effet, les modèles issus de l'analyse PCA présentent une corrélation moyenne $<R^2_x>$=0,81, $<R^2_{cv}>$=0,50 en validation interne.

Pour la suite de l'étude, la séparation de la base de données offrant le comportement le plus médian, c'est-à-dire présentant les paramètres sta tistiques (relatifs aux analyses PCA et PLS) les plus proches des valeurs moyennes, a été choisie. En l'occurrence, il s'agit du partage n°1 et les modèles à venir seront donc développés sur le jeu d'entraînement correspondant.

Approches quantitatives

A cette étape de l'étude, un jeu restreint de descripteurs ainsi que les jeux d'entraînement et de validation ont été mis en place. Il s'est alors agit de les employer afin de mettre en place un modèle prédictif à la fois robuste, utilisable en vue de prédictions et, autant que possible, interprétable d'un point de vue phénoménologique.

Le premier modèle quantitatif pouvant être considéré est la PLS (qui a été réalisée dans le paragraphe précédent). Ce modèle, reposant sur deux composantes principales, présente une corrélation et une robustesse assez similaires à celles de la régression multilinéaire (R^2=0,64, R^2_{cv}=0,52 et R^2_{ext}=0,75, cf. Tableau 15).

La Figure 41 représente les molécules de la base de données dans le nouveau référentiel créé, à deux composantes principales. On remarque que les molécules sont regroupées dans certaines zones du graphe. Plus particulièrement, une zone (en bleu) sur le graphique est densément peuplée.

Précédemment, les molécules ne présentant pas de substituant en position ortho par rapport au groupement nitro ont été distinguées. Si l'on considère à la fois les molécules du jeu d'entraînement et de validation, 45% des composés présentent un groupement en position ortho par rapport à NO_2. Ce ratio atteint 80% dans la zone symbolisée en rouge sur les représentations des molécules dans le plan des composantes principales (en Figure 41). A l'opposé la quantité de composés ortho dans la zone bleue n'est que de 33 %. Il semble donc que les molécules présentant des comportements particuliers sont majoritairement des composés ortho. Ceci est tout à fait en accord avec ce que nous avions pu considérer précédemment. D'ailleurs si les molécules non-ortho semblent relativement groupées, les composés ortho sont plus dispersés. Là-encore, ceci est cohérent avec les mécanismes réactionnels de décomposition puisque différents chemins peuvent intervenir selon la nature du groupement en position ortho [58].

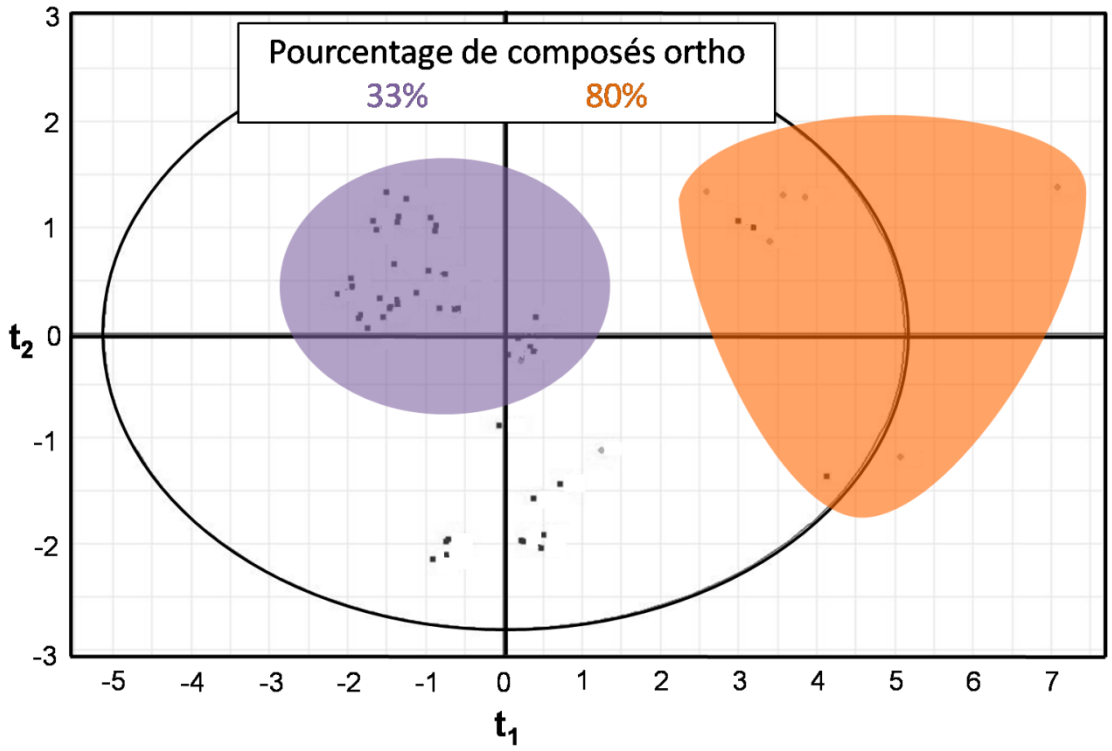


Figure 41 - Représentation des molécules dans le plan des deux premières composantes principales du modèle PLS

La méthode la plus communément utilisée en termes de régression est la régression multi-linéaire (MLR), qui avait d'ailleurs été considérée dans les études précédentes. Le modèle développé sur le jeu d'entraînement est l'équation suivante (obtenue à l'aide du logiciel WEKA) :

$$-\Delta H\ (kJ/mol) = 58{,}0\ n_{NO2} + 27{,}3\ n_O + 2{,}7\ n_{aro} + 5{,}2\ V_M - 2304{,}0\ E_{N,avg} - 567{,}7\ Q_{max} + 2{,}5\ E_{LUMO+1} - 306{,}6 \quad (3.23)$$

Ce modèle présente une corrélation R^2 de 0,69. Sa robustesse est évaluée avec une démarche de validation interne consistant non seulement en *cross-validations* en *leave-one-out* mais aussi *leave-many-out* (*10-, 5-, 3-, 2-fold*) pour s'assurer encore plus de la stabilité de la méthode comme présenté dans le Tableau 16.

R^2	Eq. (3.23)	Eq. (3.24)
Jeu d'entraînement	0,69	0,59
LOO [a]	0,47	0,38
10-fold [b]	0,48	0,37
5-fold [b]	0,54	0,46
3-fold [b]	0,47	0,40
2-fold [b]	0,53	0,41
Jeu de validation	0,80	0,67

[a] méthode de cross-validation en *Leave-One-Out*
[b] méthodes de cross-validation en *Leave-Many-Out*

Tableau 16 – Analyse de la fiabilité du modèle multi-linéaire

Si les valeurs de R^2_{cv} ne sont pas particulièrement élevées, la robustesse est relativement bonne pour le modèle (3.20), les coefficients de corrélation étant stables pour les différentes étapes de validation interne. En effet, en augmentant la taille de la partition depuis le LOO (correspondant à un *55-fold*) vers le LMO en *2-fold*, aucune perte de corrélation n'est observée.

La prédictivité du modèle est assez remarquable, puisque les chaleurs de décomposition des molécules du jeu de validation sont reproduites avec un coefficient de corrélation de 0,80.

Cela dit, ce modèle présente une corrélation et une stabilité moins élevée que celui qui avait été obtenu à l'aide de l'analyse BMLR en Eq. 3.20. Ce dernier était plus corrélé et robuste (R^2=0,84 et R^2_{cv}=0,81) mais moins prédictif (0,43 contre 0,80 ici). Ce modèle était constitué de quatre descripteurs seulement. Aussi, une réduction supplémentaire du jeu de descripteurs a été réalisée afin de s'affranchir d'une éventuelle sur-paramétrisation qui pourrait intervenir ici, ce qui pourrait expliquer la plus faible robustesse du modèle (3.23). Les descripteurs contribuant le moins au modèle PLS (n_O, V_M et Q_{max}), au sens des VIP, ont alors été exclus et une nouvelle régression multi-linéaire est réalisée :

$$-\Delta H = 205{,}8\ n_{NO2} + 28{,}5\ n_{aro} - 16{,}8\ E_{LUMO+1} + 2779{,}1\ E_{N,avg} - 75{,}6 \qquad (3.24)$$

Cette équation à quatre descripteurs est moins corrélée sur le jeu d'entraînement (R^2=0,59), moins robuste en termes de validation interne (R^2_{cv}=0,38) et moins prédictive en termes de validation externe (R^2_{ext}=0,67). Avec cette nouvelle méthodologie, le modèle à 7 descripteurs est donc le plus adéquat.

A partir des deux approches quantitatives, il est d'ores et déjà possible d'obtenir un modèle relativement prédictif. Cela dit, les modèles (PLS et Eq. 3.23) restent moins performants que le modèle obtenu en se limitant aux descripteurs non-ortho.

Approche par classification

D'autres méthodes d'analyse de données permettent le traitement de l'information en discrétisant la propriété. Autrement dit, ces méthodes consistent à bâtir des modèles de classification. Elles présentent plusieurs avantages.

En premier lieu, la discrétisation rend le modèle plus stable lorsque la propriété présente des incertitudes importantes. De plus, les modèles sont plus simples et interprétables au niveau chimique, d'une part du fait de la structure du modèle proprement dit et d'autre part parce que ce type de modèles nécessite en général un nombre plus réduit de descripteurs. Enfin, des modèles plus généraux, pouvant donc être appliqués à une plus grande variété de molécules, peuvent être obtenus. Au prix de la quantitativité, des modèles plus stables, interprétables et peut-être même plus prédictifs, pour une plus grande variété de molécules, peuvent être attendus.

Pour ce faire, la propriété quantitative de chaleur de décomposition a été discrétisée en deux classes. La classe A contient les molécules les moins énergétiques, c'est-à-dire présentant les chaleurs de décomposition les plus basses (ΔH<500 kJ/mol), la classe B les plus énergétiques (ΔH>500 kJ/mol).

L'approche employée ici consiste en la construction d'un arbre de décision, réalisée à l'aide l'outil de classification J48, implémenté dans le logiciel WEKA. Cet outil consiste à construire de proche en proche un arbre de décision permettant de classer au mieux les molécules du jeu d'entraînement dans les classes A ou B. A partir des 55 molécules du jeu d'entraînement, l'arbre représenté en Figure 42 a donc pu être construit.

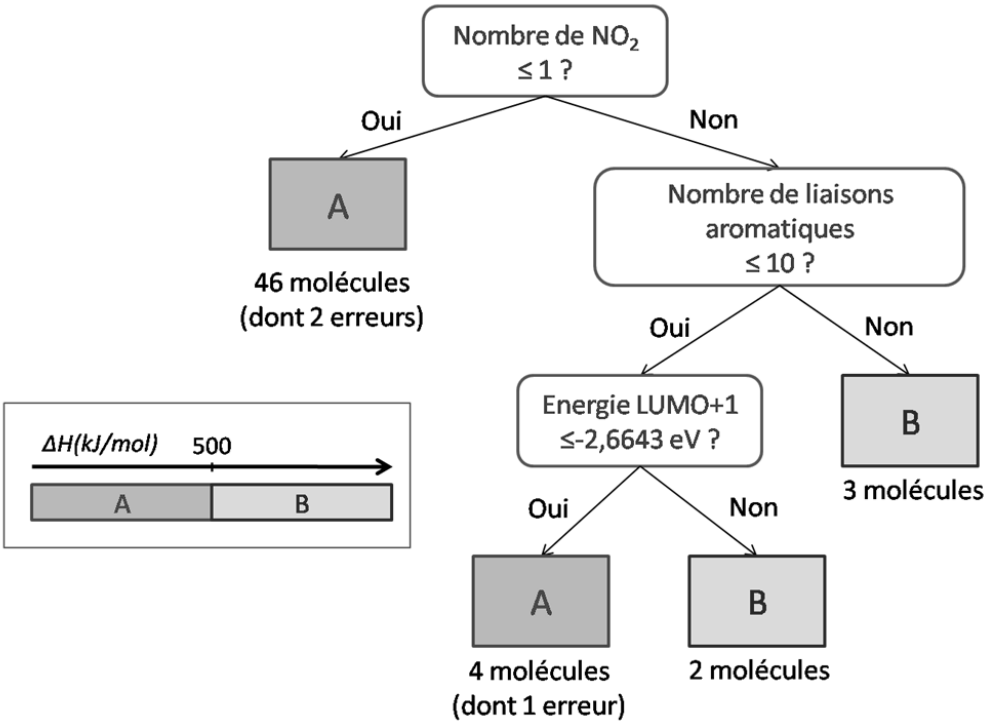


Figure 42 - Arbre de décision pour la prédiction de la chaleur de décomposition des composés nitroaromatiques du jeu d'entraînement

Le premier descripteur, le nombre de NO_2, discrimine la majeure partie de la base de données puisque 46 molécules sont déjà classées comme peu énergétiques à l'aide de ce seul descripteur. Au niveau de l'interprétation, il est tout à fait cohérent puisque plus la molécule présente de groupements nitro et plus elle est énergétique. De plus, ce descripteur a déjà été mis en évidence dans les analyses précédentes.

Parmi ces 46 molécules, 2 sont placées dans cette feuille de manière incorrecte. Ces deux molécules sont l'acide 4-nitrocinnamique et le 2-nitrobenzhydrazide qui, bien qu'elles ne possèdent qu'un groupement nitro, présentent des chaleurs de décomposition légèrement supérieures à 500 kJ/mol (respectivement 506 et 598 kJ/mol). Ces valeurs ne sont cependant pas très éloignées de la valeur limite choisie pour séparer les classes A et B. Ces molécules se trouvent donc dans la barre d'erreur expérimentale au niveau de cette limite.

Les deux autres descripteurs employés ne servent pas au classement d'autant de molécules. Cela dit, les liaisons aromatiques ayant un rôle stabilisateur, l'énergie qu'elles représentent est donc importante. Plus la molécule contient de liaisons de ce type, plus l'énergie pouvant être libérée au cours de la décomposition est importante (classe B). Ce nœud de l'arbre est donc tout à fait cohérent au niveau de son interprétation lui aussi. Le dernier nœud, bien que difficile à interpréter, permet de classer les dernières molécules de la base de données.

La seule molécule mal classée par ce dernier descripteur est le 3,5-dinitrobenzonitrile qui, malgré une chaleur de décomposition de 654 kJ/mol, est classé comme peu énergétique (A). Cette fois, l'erreur de classification provient du fait que la molécule présente 10 liaisons aromatiques, soit la valeur limite au niveau du second nœud de l'arbre. Une seule liaison supplémentaire aurait classée la molécule en classe B, ce qui aurait été correct. Il s'agit donc d'une molécule se situant dans la barre d'erreur au niveau d'un descripteur.

Pour ce qui est de la fiabilité du modèle, nous avons vu que seules 3 molécules sur les 55 employées étaient mal classées, soit une classification correcte à 94,5%. La robustesse du modèle a été vérifiée suivant la même procédure que pour la MLR (*LOO, 10-, 5-, 3-* et *2-fold*). Comme montré dans le Tableau 17, le modèle est peu sensible au retrait de certaines molécules du jeu d'entraînement, ce qui signifie que les trois seuls descripteurs utilisés au sein d'un arbre de décisions suffisent à la mise en place d'un modèle très fiable.

	Pourcentage de classifications correctes
Jeu d'entraînement	94,5
LOO [a]	83,6
10-fold [b]	83,6
5-fold [b]	85,5
3-fold [b]	80,0
2-fold [b]	83,6
Jeu de validation	81,8

[a] méthode de cross-validation en Leave-One-Out

[b] méthodes de cross-validation en Leave-Many-Out

Tableau 17 – Analyse de la fiabilité du modèle par arbre de décision

De plus, en appliquant cet arbre de décision aux molécules du jeu de validation, la prédictivité du modèle a pu être évaluée à 81,8%. 4 molécules du jeu de validation sont mal prédites par le modèle. L'acide 2-nitrocinamique est mal classé (B au lieu de A) du fait de son nombre de groupements nitro (1 seul). S'il avait passé ce nœud de l'arbre, il aurait été classé correctement. Pour les 3 autres molécules concernées (acide 2-chloro-3,5-dinitrobenzoïque, 2,4-dinitrotoluène, chlorure de 3-dinitrobenzyle), elles ont été classées parmi les moins énergétiques du fait de l'énergie de leur LUMO+1 qui est légèrement trop négative (respectivement -3,1, -2,9 et -3,4), mais la mauvaise classification doit être attribuée au nœud précédant où ces trois molécules présentent un nombre de liaisons aromatiques de 10, soit la valeur critique. Il s'agit donc d'erreurs dues à l'incertitude au niveau de ce nœud.

Il est intéressant de noter que ce type d'analyse permet d'identifier aisément les sources d'erreurs. Par exemple, ici, des composés présentant 10 liaisons aromatiques risquent d'être classés parmi les composés les moins énergétiques alors qu'ils ne le sont pas. En termes d'utilisation du modèle, une telle information est facile à prendre en compte. Le même type de traitement est beaucoup plus difficile à réaliser au niveau des analyses quantitatives puisque les descripteurs sont tous utilisés ensemble, alors qu'ils sont ici utilisés indépendamment, les uns après les autres.

Bilan

Au cours de cette analyse, plusieurs outils de traitement de données ont été testés. Il s'agit d'outils complémentaires qui permettent de traduire les informations contenues dans notre jeu de données de manières différentes.

Si les méthodes par projection (PCA et PLS) pour la sélection des descripteurs sont moins automatisées que la technique BMLR employée précédemment, elles permettent d'analyser les données sous un nouveau point de vue puisqu'elles consistent en une réduction progressive de la dimensionnalité de l'espace des descripteurs alors que la BMLR incrémente le modèle par ajouts successifs de descripteurs. Dans cette dernière, seule l'information statistique est employée alors que l'approche par projection sur des composantes principales permet un choix manuel des descripteurs sur des critères phénoménologiques.

Une nouvelle méthode a également été employée pour la mise en place des jeux d'entraînement et de validation à l'aide d'un tirage aléatoire. Pour s'assurer de l'indépendance des modèles vis-à-vis de ce partage, une statistique sur 10 partages aléatoires a été réalisée.

Ensuite, différents types de modèles ont été employés sur le jeu réduit des 7 descripteurs sélectionnés. Les résultats obtenus sont résumés dans le Tableau 18. Les modèles quantitatifs (MLR et PLS) permettent d'obtenir une certaine performance en termes de prédictivité mais une certaine faiblesse en termes de robustesse. L'analyse PLS a d'ailleurs permis d'observer des sous-groupes de molécules qui tendent à montrer que des analyses distinctes devraient être menées afin de distinguer les composés non-ortho des ortho, ce qui a été fait précédemment. L'objectif étant ici de disposer d'un modèle global, c'est-à-dire applicable à l'intégralité des composés de la base de données, une autre approche, qualitative, a été utilisée.

Un arbre de décision, à la fois corrélé, robuste et prédictif, a ainsi été développé. Au prix de prédictions quantitatives sur la propriété, un modèle de classification plus robuste, global et prédictif a pu être développé. De plus, il présente l'avantage de sa simplicité qui permet une grande transparence et une bonne interprétation au niveau phénoménologique.

Molécules	Expérience		Prédiction		
	valeur	classe	MLR	PLS	DT
Jeu d'entraînement					
2,4-dinitroaniline	597	B	597	656	B
2-nitrophenylhydrazine	381	A	306	306	A
4-chloro-2-nitroaniline	349	A	302	336	A
4-nitroanisole	248	A	299	270	A
2,6-dinitroaniline	719	B	587	590	B
4-nitrophénol	232	A	201	241	A
acide 4-nitrophénylacétique	265	A	249	236	A
acide 4-nitrocinnamique	506	B	441	404	A
3-nitrophénol	283	A	206	237	A
3-nitrotoluène	149	A	231	244	A
acide 4-chloro-3-nitrobenzoïque	354	A	297	315	A
2-amino-4-nitrophénol	130	A	295	315	A
2,6-dichloro-4-nitroaniline	264	A	374	339	A
2-nitroanisole	230	A	299	262	A
acide 3-nitrophenylacétique	358	A	249	234	A
1,5-dinitro-2,4-difluorobenzène	439	A	577	628	A
4-nitroaniline	347	A	239	284	A
méthyl-ester d'acide 4-nitrobenzoïque	302	A	338	326	A
méthyl-ester d'acide 2-nitrobenzoïque	274	A	308	301	A
4-nitrophénétole	270	A	387	297	A
acide 3,5-dinitrobenzoïque	674	B	635	665	B
acide 3-nitrobenzoïque	289	A	238	302	A
4-nitrobenzaldéhyde	421	A	301	329	A
3-nitrobenzaldéhyde	373	A	283	324	A
4-nitro-3-crésol	345	A	277	259	A
2-nitrobenzamide	256	A	336	356	A
acide 2-nitrobenzoïque	271	A	161	228	A
4-nitrobenzamide	319	A	354	370	A
2-nitrobenzaldéhyde	318	A	275	312	A
2-nitrophénol	123	A	209	240	A
chlorure de 4-nitrobenzyle	337	A	312	287	A
2,4,6-trinitrotoluène	1223	B	957	900	B
alcool de 3-nitrobenzyle	325	A	300	276	A
4-nitro-2-toluidine	306	A	325	308	A
4-nitrophénylhydrazine	277	A	318	322	A
acide 2,6-dinitrobenzoïque	222	A	548	594	A
3-nitroaniline	350	A	245	284	A
acide 2-nitrophénylacétique	175	A	223	216	A
4-nitroacétoanilide	387	A	420	378	A
1-chloro-3,4-dinitrobenzène	342	A	594	607	A
4-fluoro-2-nitrotoluène	129	A	241	242	A

Tableau 18 – Chaleurs de décomposition (kJ/mol) expérimentales et prédites par MLR (Eq. 3.23), PLS et arbres de décision (DT)

Molécules	Expérience		Prédiction		
	valeur	classe	MLR	PLS	DT
acide 2-nitro-5-thiocyanotobenzoïque	492	A	392	355	A
3,5-dinitrobenzonitrile	654	B	651	674	A
4-nitroacétophénone	291	A	349	328	A
2-nitrobenzhydrazide	598	B	384	365	A
2-nitrotoluène	182	A	212	227	A
alcool 3,4-dinitrobenzylique	685	B	675	628	B
4-nitrobenzhydrazide	362	A	423	394	A
5-nitrovanilline	450	A	427	322	A
méthyl-ester d'acide 3-nitrobenzoïque	256	A	319	322	A
alcool 4-nitrobenzylique	292	A	294	267	A
chlorure de 2-nitrophénylsulfényle	344	A	283	276	A
2-nitroaniline	280	A	231	305	A
2-amino-4-nitroanisole	375	A	397	335	A
3-nitroacétoanilide	369	A	418	379	A
Jeu de validation					
acide 2-nitrocinnamique	600	B	405	370	A
alcool 2-nitrobenzylique	319	A	262	245	A
acide 4-nitrobenzoïque	284	A	258	307	A
acide 2,4-dinitrobenzoïque	394	A	558	599	A
3,5-dinitrobenzamide	736	B	717	709	B
3-nitroanisole	243	A	300	263	A
acide 3,4-dinitrobenzoïque	701	B	628	647	B
3-nitroacétophénone	276	A	324	315	A
3-nitrobenzamide	311	A	345	376	A
acide 2-chloro-3,5-dinitrobenzoïque	1023	B	631	614	A
acide 3-nitrocinnamique	414	A	399	366	A
5-chloro-2-nitrobenzotrifluoride	40	A	40	162	A
2,4-dinitrotoluène	632	B	619	629	A
4-nitrotoluène	213	A	229	245	A
3,4-dinitrotoluène	684	B	611	605	B
3-nitrobenzhydrazide	430	A	413	402	A
2-nitroacétophénone	308	A	300	279	A
3,5-dinitrobenzylchloride	711	B	702	666	A
chlorure de 4-nitrobenzoyle	408	A	365	357	A
Nitrobenzène	161	A	143	220	A
2-amino-5-nitrophénol	153	A	295	297	A
2-nitroacétoanilide	297	A	392	361	A
		R^2	0,69	0,64	*(94,6%)*[a]
		R^2_{cv}	0,47	0,52	*(83,6%)*[a]
		R^2_{ext}	0,80	0,75	*(81,8%)*[a]

[a]pourcentage de classifications correctes

Tableau 18 (suite) - Chaleurs de décomposition (kJ/mol) expérimentales et prédites par MLR (Eq. 3.23), PLS et arbres de décision (DT)

2.4- Température de décomposition

Lorsque l'on s'intéresse à la stabilité thermique, deux caractéristiques sont à prendre en compte. La première est l'énergie libérée durant la décomposition, à savoir la chaleur de décomposition qui a été considérée jusqu'à présent. La seconde est la température à laquelle celle-ci s'initie, la température de début de décomposition, T_0, elle aussi obtenue par DSC. Dans la procédure de présélection de la règlement relative au transport des marchandises dangereuses [40], les composés dont la température de décomposition excède 500°C sont considérées comme non éligibles pour la classe des composés explosibles.

Une base de données de 22 molécules nitroaromatiques [158] a été employée pour la mise en place d'un modèle QSPR. Comme précisé précédemment, le nombre de molécules ne nous permet pas de réaliser une scission raisonnable en jeux d'entraînement et de validation. Il ne s'agit donc pas ici de réaliser le développement complet d'un modèle en vue d'usage prédictif pour lequel la validation externe est nécessaire. Il s'agit plutôt de d'analyser les aspects méthodologiques. Des modèles ont été obtenus à partir de structures optimisées aux niveaux AM1 et DFT afin d'évaluer l'influence du niveau de calcul.

Modèle AM1

Les 22 molécules de la base de données ont été optimisées à l'aide de la méthode semi-empirique AM1. A partir de ces structures, plus de 300 descripteurs ont été calculés, en partie à l'aide du logiciel CodessaPro. L'analyse BMLR a alors été employée afin de mettre en place le modèle le plus performant possible. Le meilleur compromis entre la corrélation du modèle et son degré de paramétrisation est donné pour un modèle à 5 descripteurs.

$$T_0(°C) = 2813{,}3\ n_{C,rel} - 334{,}9\ {}^2BIC - 2956{,}7\ \eta + 10{,}6\ S_{ZX} - 0{,}5\ PPSA1 - 275{,}7 \qquad (3.25)$$

où $n_{C,rel}$ est le nombre relatif d'atomes de carbone, ${}^2BIC$ est l'indicateur de contenu d'information moyen sur les liaisons (à l'ordre 2) , η est la dureté moléculaire, S_{ZX} est la projection de la molécule sur un plan ZX et PPSA1 est la portion de la surface moléculaire chargée positivement (définie à partir de la méthode de Zefirov).

Les descripteurs inclus dans ce modèle sont assez difficiles à relier directement au processus chimique de décomposition. L'indice topologique ${}^2BIC$ ou encore S_{ZX} n'ont a priori pas de relation particulière avec la décomposition des nitroaromatiques. Malgré tout, la présence de la dureté chimique η, un descripteur de la réactivité chimique, est à remarquer. La pertinence de ce type de descripteurs avait d'ailleurs pu être démontrée précédemment (cf. paragraphe 2.3.1.2). Enfin, PPSA1 est un descripteur de la charge en surface de la molécule destiné initialement à évaluer les possibilités d'interaction de la molécule avec le milieu environnant. Cela dit, la distribution de la charge en surface (et donc PPSA1) est reliée aux propriétés électroniques locales au niveau des groupements nitro, qui jouent un rôle central dans le processus de décomposition.

Du point de vue des performances statistiques, ce modèle présente un coefficient de corrélation R^2 significatif de 0,91 comme montré en Figure 43. De plus, il semble relativement robuste puisque le R^2_{cv} atteint 0,84. Bien entendu, sans validation externe, il n'est pas possible d'estimer la capacité du modèle à prédire des données en dehors du jeu d'entraînement.

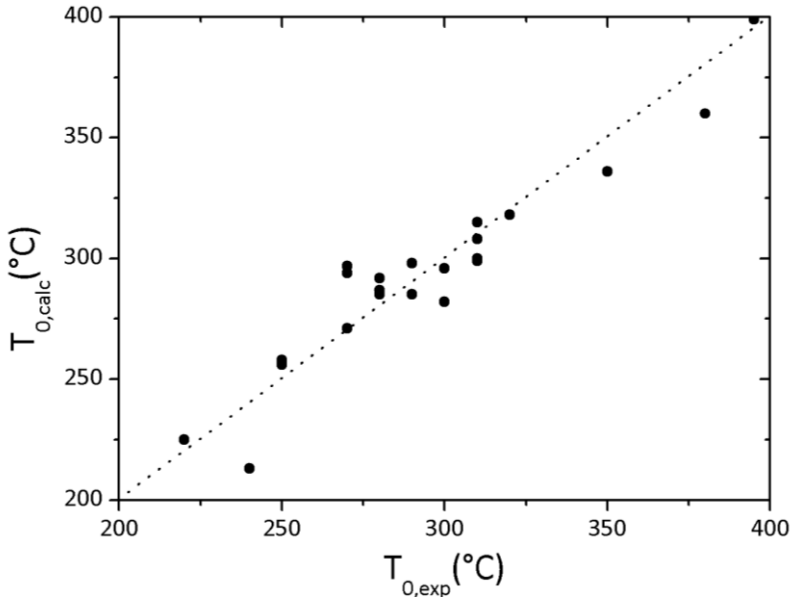


Figure 43 – Températures de décomposition (en °C) expérimentales et prédites par le modèle AM1 (Eq. 3.25)

Modèle DFT

Les structures des 22 molécules ont également été optimisées au niveau DFT, PBE0/6-31+G(d,p). Les plus de 300 descripteurs calculés ont alors été intégrés dans une analyse BMLR et le meilleur modèle obtenu est une équation à 4 paramètres.

$$T_0 = 1{,}7\ WPSA1 + 11{,}4\ RNCS + 1544{,}0\ Q_{H,min} - 31{,}2\ n_H - 420{,}5 \qquad (3.26)$$

où WPSA1, RNCS, $Q_{H,min}$ sont la surface pondérée chargée positivement, la surface relative chargée négativement et la charge minimale pour un atome d'hydrogène (calculées selon la méthode de Mulliken) et n_H est le nombre d'atomes d'hydrogène.

Les descripteurs sont là-encore relativement difficiles à relier de manière directe au processus de décomposition. WPSA1 et RNCS sont des descripteurs de la distribution de charge au niveau de la surface moléculaire qui peuvent être influencés par les propriétés électroniques au niveau du groupement NO_2. Cela dit, rien ne permet a priori de relier les propriétés et le nombre d'atomes d'hydrogène de la molécule au processus de décomposition des composés nitroaromatiques considérés.

Non seulement les valeurs obtenues avec ce modèle (en Tableau 19) reproduisent de manière efficace les valeurs expérimentales (R^2=0,90), comme montré en Figure 44, mais il est en plus relativement robuste avec un R^2_{cv} de 0,81.

molécules	exp [158]	DFT (3.25)	AM1 (3.26)
nitrobenzène	380	368	360
1,2-dinitrobenzène	280	286	292
1,3-dinitrobenzène	270	290	294
1,4-dinitrobenzène	350	341	336
2-nitrotoluène	290	293	298
3-nitrotoluène	310	313	308
4-nitrotoluène	320	328	318
2,6-dinitrotoluène	290	289	285
3,4-dinitrotoluène	280	286	287
2,4-dinitrotoluène	250	250	258
2-nitroaniline	280	297	285
3-nitroaniline	300	298	296
4-nitroaniline	310	288	300
Acide 2-nitrobenzoïque	270	277	271
Acide 3-nitrobenzoïque	300	276	282
Acide 4-nitrobenzoïque	310	328	315
2-nitrophénole	250	260	256
3-nitrophénole	310	304	299
4-nitrophénole	270	275	297
1-chloro-4-nitrobenzène	395	386	399
2,4-dinitrophénole	240	220	213
2,4,6-trinitrophénole	220	221	225
	MAE (%)	3,2	3,6
	R^2	0,90	0,91
	R^2_{cv}	0,81	0,84

Tableau 19 - Températures de décomposition (en °C) expérimentales et prédites pour une série de composés nitroaromatiques

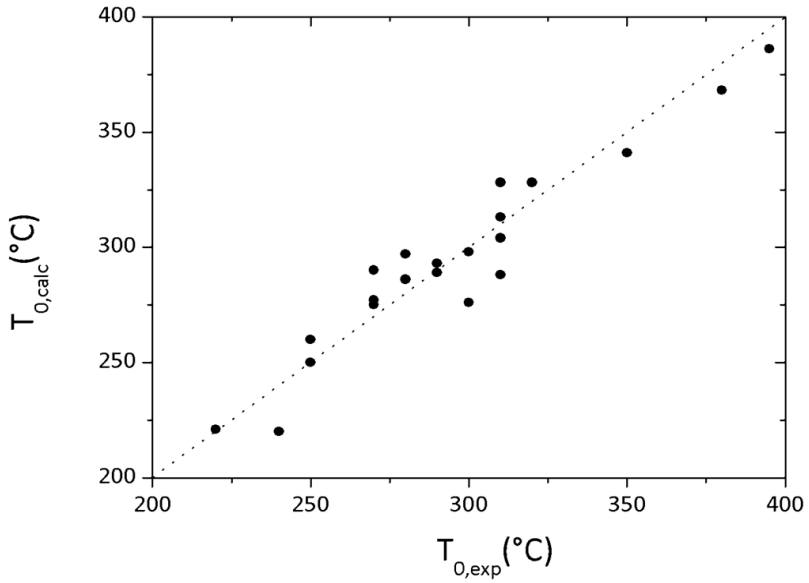


Figure 44 – Températures de décomposition (en °C) prédites par le modèle DFT (Eq. 3.26) et expérimentales

Si l'on compare les modèles DFT et AM1 obtenus, il faut noter que les deux modèles présentent des corrélations similaires et sont même caractérisés par la même erreur moyenne absolue (MAE=10 kJ/mol) vis-à-vis des données expérimentales (voir Tableau 19). Par contre, le modèle AM1 repose sur 5 descripteurs contre seulement 4 pour le modèle DFT. Si aucune validation externe n'est possible avec une base de données aussi réduite, le risque d'une sur-paramétrisation avec le modèle AM1 semble plus important.

Bilan

La taille de la base de données employée n'a pas permis la détermination des pouvoirs prédictifs des modèles développés aux niveaux DFT et AM1. Cela dit, il apparaît que l'emploi de structures au niveau DFT soit plus adéquat puisque le modèle obtenu, moins paramétré, présente moins de risque de sur-paramétrisation.

Cette étude mérite d'être approfondie. De premiers essais ont été réalisés sur une base étendue de 77 molécules [157] mais jusqu'à présent sans succès, aucun modèle n'ayant pu être mis en évidence.

2.5- Bilan

Le bilan de l'analyse QSPR de la stabilité thermique des composés nitroaromatiques est très satisfaisant, la majorité des travaux concernant leur chaleur de décomposition. Du point de vue méthodologique, les descripteurs quantiques sont très intéressants pour cette propriété, pour laquelle les mécanismes de décomposition sont très importants. Il a également été montré, à la fois sur la chaleur et la température de décomposition, qu'un niveau de théorie suffisamment précis (ici, la DFT) était nécessaire pour l'introduction de ces descripteurs dans les modèles. Dés lors, non seulement ils contribuent à l'obtention de modèles fiables mais ils permettent aussi à ces derniers d'être interprétables du point de vue phénoménologique en mettant en évidence l'importance du nombre et de la réactivité des groupements nitro dans la stabilité thermique des composés nitroaromatiques.

Une première méthode a été développée pour l'obtention de modèle prédictif sur la base de la méthode BMLR du logiciel CodessaPro. A partir de cette dernière, un modèle a été développé pour un jeu de molécules d'entraînement, en extrayant celui présentant le meilleur compromis entre corrélation et nombre de descripteurs, pour s'affranchir de toute sur-paramétrisation. Ce modèle est alors validé en interne par *cross validation*, afin d'estimer sa robustesse (R^2_{cv}), et en externe à l'aide d'un jeu de validation, afin d'estimer son pouvoir prédictif (R^2_{ext}). Enfin, son domaine d'applicabilité est défini sur la base des descripteurs du modèle et la prédictivité au sein de ce dernier est calculée (R^2_{in}).

Avec cette approche, un modèle prédictif (Eq. 3.20) a été mis en place pour la chaleur de décomposition des dérivés du nitrobenzène ne présentant pas de substituant en position ortho du groupement nitro. Ce modèle contenant 4 descripteurs est très corrélé (R^2=0,90), robuste (R^2_{cv}=0,86) et surtout prédictif (R^2_{ext}=0,84), en particulier dans son domaine d'applicabilité (R^2_{in}=0,86). Au niveau de l'interprétation, il rend compte de la charge

atomique maximale au sein de la molécule (attribuable à l'azote du groupement nitro, cf. Figure 40).

Pour obtenir un modèle avec un domaine d'applicabilité plus important (intégrant les composés ortho-substitués), d'autres outils d'analyse de données ont été employés. Les analyses par projection (PCA et PLS) forment un outil de sélection des descripteurs moins automatisé que la BMLR mais permettent de choisir manuellement les descripteurs au regard non seulement de critères statistiques mais aussi de critères phénoménologiques. Enfin, un modèle global a été développé à l'aide de la méthode de classification par arbre de décision, les méthodes quantitatives (PLS et MLR) étant moins performantes (en termes de corrélation et robustesse) avec cette méthode de sélection des descripteurs. Le modèle obtenu (utilisant seulement 3 descripteurs) est très corrélé (95% de classement correct sur le jeu d'entraînement), robuste (84% en procédure *Leave-One-Out*), prédictif (82% sur le jeu de validation) et très facilement interprétable chimiquement.

Les deux modèles ainsi mis en place sont d'ores et déjà utilisables en termes de prédiction. Pour une molécule donnée (pour laquelle les modèles sont applicables), il s'agit alors d'optimiser sa structure au niveau PBE0/6-31+G(d,p), de calculer les différents descripteurs des modèles. Pour une utilisation correcte, il est donc impératif de s'assurer d'une optimisation correcte des structures, au niveau de calcul approprié. Il suffit ensuite d'intégrer les descripteurs dans les modèles.

Les aspects méthodologiques ayant été abordés pour la stabilité thermique, la suite de l'étude est dédiée à deux autres propriétés : les sensibilités à la décharge électrique et à l'impact des composés nitroaromatiques.

3- Sensibilité à la décharge électrique

3.1- Généralités

La sensibilité est une propriété majeure des matériaux énergétiques puisqu'elle renseigne sur leur propension à réagir (de manière explosive) sous l'effet d'un stimulus externe mécanique, thermique ou, dans le cas présent, électrique. Ainsi, la sensibilité à la décharge (ou étincelle) électrique (E_{ES}) quantifie le degré de sensibilité du matériau à une telle décharge. Il est à noter que bien que non incluse dans le cadre de REACH, elle intervient dans la règlementation liée aux produits pyrotechniques [193].

Des décharges électriques ou électrostatiques peuvent se développer au cours de la manipulation de tels produits [194]. De plus, certains dispositifs de mise à feu d'explosifs reposent sur leur sensibilité électrique. Pour ces raisons, il est important de connaître la sensibilité des matériaux potentiellement explosifs à de tels stimuli.

Cette propriété est déterminée expérimentalement en appliquant une décharge électrique de haute tension, à un échantillon (à l'état solide ou liquide) à l'aide d'une capacité. Il s'agit de déterminer l'énergie de la décharge permettant d'initier une réaction au sein de l'échantillon [194]. Il peut s'agir d'une explosion, de l'apparition de fumées ou d'un simple bruit, signes du départ de la décomposition du matériau.

Différents dispositifs existent pour déterminer cette sensibilité [195,196] et les résultats peuvent varier de manière importante d'un appareillage à un autre, comme le montre le Tableau 20 pour le TNT ou le TATB [197].

	Ref [197]	Ref [198]	Ref [199]	Ref [200]	Ref [201]
TNT	4,1	0,46 / 3,75	22,38	1,26	8,57 / 5,47
TATB	17,8	4,25 / 18,14	-	-	11,88 / 13 ,52

Tableau 20 - Sensibilités à la décharge électrique (en J) obtenues à partir de différents dispositifs expérimentaux

Par exemple, la forme des électrodes et leur surface influencent la densité d'énergie déployée lors de la décharge. De même, la structure du circuit électrique et les conditions expérimentales (température, humidité, confinement) affectent les mesures [201,202].

La sensibilité à la décharge électrique dépend également des propriétés physiques de l'échantillon (granulométrie, présence de dislocations dans les cristaux solides) [202-205]. La variabilité de tous ces paramètres fait que l'incertitude sur les mesures expérimentales est relativement importante (sans être précisée dans la littérature).

Un schéma simplifié du dispositif employé pour obtenir les valeurs qui seront utilisées dans cette étude est présenté en Figure 45 [197]. Sur ce schéma, U est un générateur à haute tension, P un commutateur, V un voltmètre, C une capacité et R une résistance. L'électrode inférieure est en acier et l'électrode supérieure conique est reliée à l'échantillon proprement dit par un barreau métallique isolé de l'environnement par un tube en matière plastique.

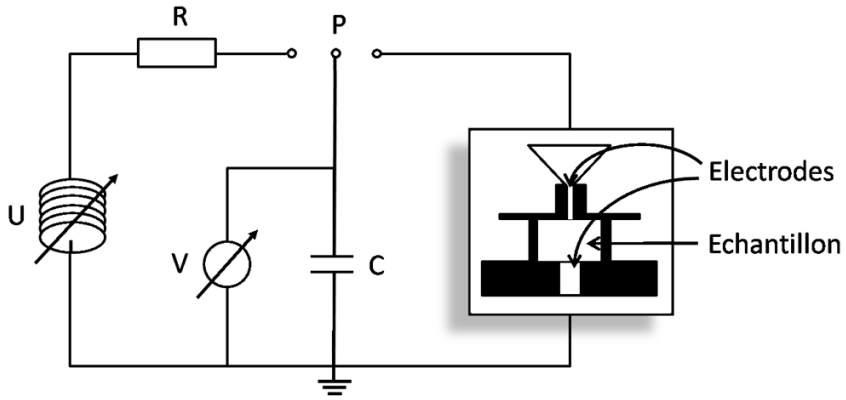


Figure 45 - Appareillage permettant la détermination de la sensibilité à la décharge électrique [197]

En fait, le phénomène d'initiation engendré par une décharge électrique ainsi que les micro-mécanismes mis en jeu n'ont pas encore été complètement élucidés [206,207]. Certains auteurs considèrent une conversion de l'énergie de la décharge électrique en un phénomène thermique [205,208,209]. Dans un premier temps, un claquage diélectrique au sein du matériau intervient et l'énergie électrique est convertie en énergie thermique par effet Joule. Des points chauds sont alors créés à partir desquels l'allumage proprement dit du matériau a lieu, en initiant des réactions de décomposition rapides. D'autres hypothèses existent selon lesquelles les processus de transferts énergétiques pourraient être plus compliqués (onde de choc, micro-arcs électriques…) [194].

Bien que les mécanismes ne soient pas encore totalement compris, certaines observations ont pu être faites d'un point de vue chimique. Les structures électroniques et les proches voisins des groupements nitro ont une influence critique sur la sensibilité de composés nitroaromatiques [210]. Cette observation est d'ailleurs en accord avec les mécanismes de décomposition qui diffèrent en fonction de la présence ou non de groupements en position ortho de NO_2 pour cette famille de composés [58,172,173].

3.2- Travaux antérieurs

Jusqu'à présent, peu d'études prédictives ont été réalisées sur la propriété de sensibilité électrique. Zeman [206,211] a ainsi proposé des corrélations avec les vitesses de détonation des composés nitrés. Par la suite, d'autres relations ont été développées pour estimer la sensibilité à la décharge électrique à partir de valeurs calculées de paramètres de détonation (tels que les vitesses et pressions de détonation) [212-214].

Par exemple, dans les travaux de Keshavarz [214], la sensibilité à la décharge électrique est prédite par la relation suivante (avec une corrélation R^2=0,94) mais pour lequel aucune validation n'a été réalisée :

$$E_{ES}(J) = 21{,}19 - 0{,}0422\ P' - 3{,}257\ SSP^- + 6{,}498\ SSP^+ \quad (3.27)$$

dans laquelle SSP^- et SSP^+ sont des paramètres qui définissent la présence de groupements particuliers dans la molécules (tels que CH_2-NNO_2 ou –(NO_2)N–C–C–N(NO_2)–) et P' est la pression de détonation maximale, elle-même calculée par la relation :

$$P' = 221{,}5 - 20{,}4\ a - 2{,}3\ b + 17{,}2\ c + 16{,}1\ d - 79{,}1\ C_{SSP} - 66{,}3\ n_N \quad (3.28)$$

où a, b, c et d sont les nombres de carbone, hydrogène, azote et oxygène dans la molécule, C_{SSP} est un autre paramètre qui définit la présence d'autres groupements particuliers (-N=N-, -O-NO_2, etc.) et n_N vaut (1,5 + 0,5 n_{NO2}) avec n_{NO2} le nombre de groupements nitro.

Des tendances, reliant la sensibilité électrique à la réactivité thermique (énergie d'activation expérimentale, obtenue par thermolyse), ont également été mises en évidence [201,207,208,210]. De même, des corrélations avec des propriétés moléculaires de nitramines (charges du groupement nitro, énergies de dissociation, orbitales moléculaires) ont été observées [203,215]. Wang et coll. [213] ont ainsi considéré la charge du groupement nitro et l'orbitale moléculaire la plus basse occupée (LUMO) comme des paramètres particulièrement importants (qualitativement) pour la sensibilité des composés nitroaromatiques, sans obtenir pour autant une relation quantitative.

A notre connaissance, peu d'études QSPR ont été menées jusque-là. On citera les travaux réalisés par Wang [212] basés sur des paramètres électroniques (charge de NO_2, énergies orbitalaires…) pour une série de nitramines et par Keshavarz [216,217] utilisant des descripteurs constitutionnels pour les nitroaromatiques et les nitramines. En particulier, un modèle à 4 paramètres a été proposé par ce dernier pour une série de composés nitroaromatiques, avec un coefficient de corrélation R^2 de 0,77 [216].

$$E_{ES}(J) = 4{,}60 - 0{,}733\ n_C + 0{,}724\ n_O + 9{,}16\ R_{nH/nO} - 5{,}14\ C_{R,OR} \quad (3.29)$$

où n_C et n_O sont les nombres d'atomes de carbone et d'oxygène, $R_{nH/nO}$ est le ratio d'atomes d'hydrogène rapporté au nombre d'atomes d'oxygène et $C_{R,OR}$ caractérise la présence de groupements alkyl (-R) ou alkoxy (-OR) le long du cycle aromatique.

Une des principales limitations de ce modèle, reposant exclusivement sur des descripteurs constitutionnels, réside dans le fait que ces derniers ne distinguent pas les isomères entre eux. Ainsi, les sensibilités calculées pour les molécules de 1,5- et 1,8-dinitronaphtalène (1,5- et 1,8-DNN) sont identiques (13,91 J) alors que les valeurs expérimentales (respectivement 11,2 et 13,9 J) présentent une différence significative (voir Figure 46).

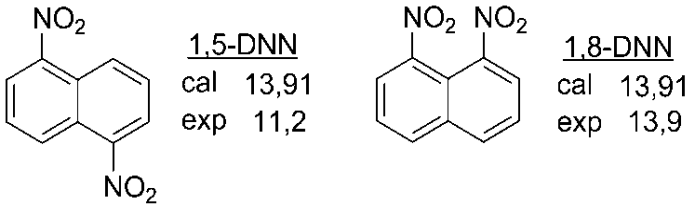


Figure 46 - Sensibilités électriques (en J) des 1,5- et 1,8-dinitronaphtalènes (DNN) calculées par Keshavarz [216] et expérimentales [211]

La méthodologie employée par la suite considère une plus grande variété de descripteurs, en incluant des propriétés issues de calculs quantiques qui permettront non seulement de dépasser cette limitation mais aussi de prendre en compte d'autres paramètres influants, relatifs notamment à la réactivité chimique.

3.3- Analyse QSPR

Comme expliqué précédemment, la propriété de sensibilité à la décharge électrique peut être mesurée à l'aide de divers appareillages expérimentaux et les résultats obtenus varient de manière importante non seulement en fonction des dispositifs mais aussi en fonction des conditions de l'expérience [216]. Toutes les données expérimentales, employées ici, ont été obtenues à l'aide d'un unique dispositif, un appareillage développé par Zeman et al. [197], décrit précédemment. Les valeurs de sensibilité des 41 molécules étudiées ont été collectées dans la littérature [207,211]. Il s'agit de composés nitroaromatiques, présentés en Figure 47.

Les structures moléculaires de ces 41 molécules ont été optimisées à l'aide de la DFT au niveau PBE0/6-31+G(d,p). Ces dernières ont été utilisées pour le calcul de plus de 300 descripteurs (pour chaque composé). Il s'agit des mêmes types de descripteurs que ceux qui ont pu être considérés précédemment pour la stabilité thermique. Le jeu de données a alors été partagé en deux afin de former un jeu d'entraînement et un jeu de validation de respectivement 26 et 15 molécules.

3.3.1- Développement du modèle (jeu d'entraînement)

Dans un premier temps, un modèle QSPR a été développé sur 26 molécules choisies de manière aléatoire dans la base de données (jeu d'entraînement). Les descripteurs correspondant à ces 26 composés ont donc été intégrés dans notre démarche d'analyse de données. Afin de mettre en place un modèle quantitatif, la technique BMLR du logiciel CodessaPro a été employée. Le choix du modèle a été réalisé pour représenter le meilleur compromis entre la corrélation et le degré de paramétrisation.

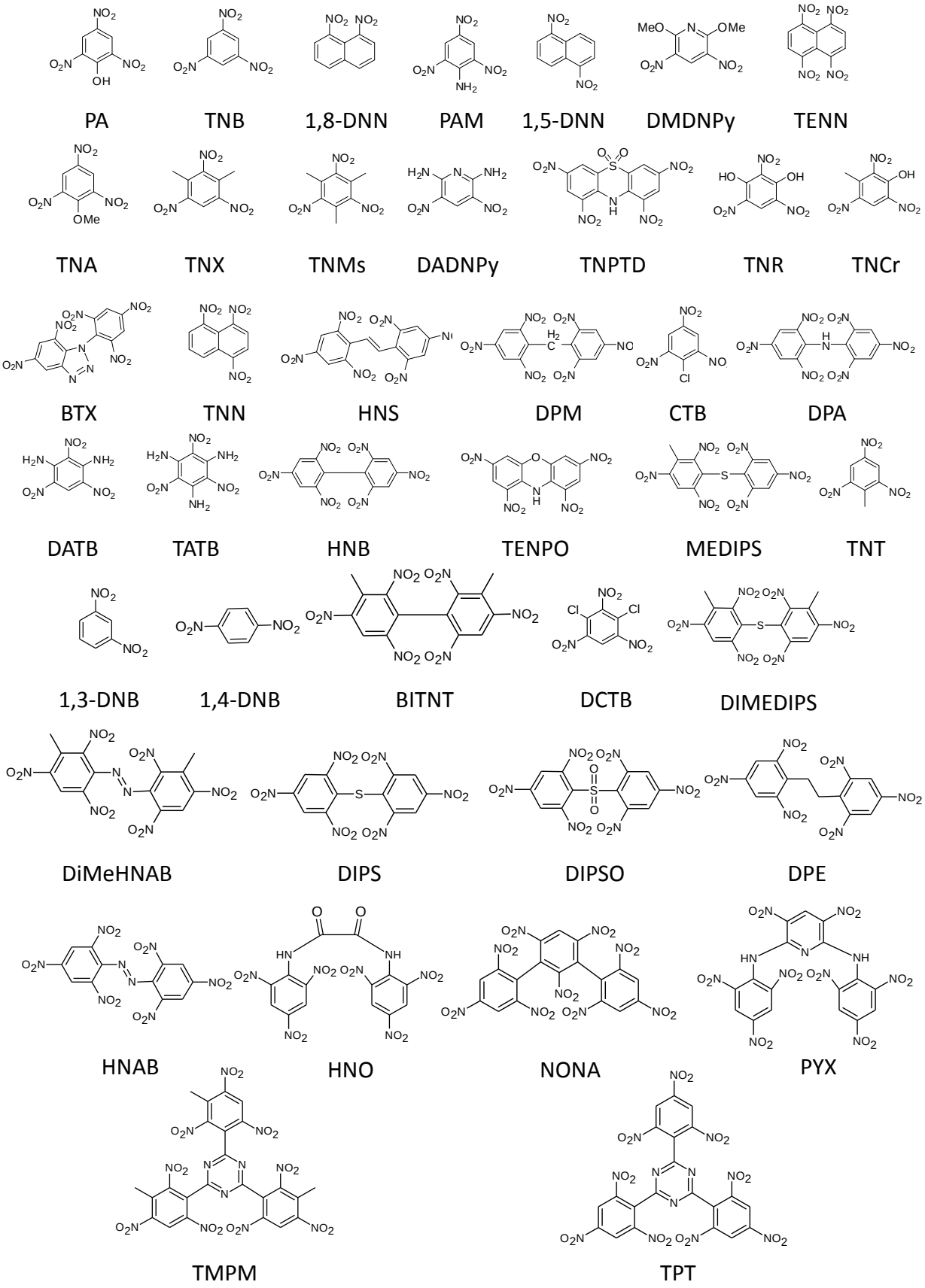


Figure 47 – 41 Molécules de la base de données

(voir Tableau 21 et Tableau 22 pour la dénomination complète des molécules)

Le meilleur modèle est une équation à quatre descripteurs, bien corrélée avec l'expérience (R^2=0,90), comme le montre le Tableau 21, et plutôt robuste (R^2_{cv}=0,78), aux vues des incertitudes expérimentales déjà importantes.

$$E_{ES}(J) = 29,6\ n_{single} + 63,3\ N_{C,max} + 168,4\ Q_{C,min} - 27,8\ V_{C,min} + 99,4 \quad (3.30)$$

où n_{single} est le nombre relatif de liaisons simples et $N_{C,max}$, $Q_{C,min}$ et $V_{C,min}$ sont respectivement l'indice maximal de réactivité nucléophile, la charge minimale et la valence minimale pour un atome de carbone.

Dans le modèle final, si n_{single} n'est a priori pas directement lié à l'initiation d'une réaction de décomposition explosive, les trois autres descripteurs le sont. En effet, le phénomène réactif initial est supposé intervenir au niveau de la liaison C-NO_2. Il est donc intéressant de voir apparaître des descripteurs liés à la réactivité et à la connectivité du carbone.

molécules	Exp [207,211]	Eq. (3.30)	Diff.
acide picrique (PA)	8,98	7,57	-1,41
1,3,5-trinitrobenzène (TNB)	6,31	9,45	3,14
1,8-dinitronaphtalène (1,8-DNN)	13,9	15,95	2,05
2,4,6-trinitroaniline (PAM)	6,85	6,32	-0,53
1,5-dinitronaphtalène (1,5-DNN)	11,2	8,21	-2,99
2,6-diméthoxy-3,5-dinitropyridine (DMDNPy)	20,57	20,36	-0,21
1,4,5,8-tetranitronaphtalène (TENN)	8,26	3,9	-4,36
2,4,6-trinitroanisole (TNA)	28,59	27,04	-1,55
trinitroxylène (TNX)	11,1	9,9	-1,2
trinitromésitylène (TNMs)	8,98	8,77	-0,21
2,6-diamino-3,5-dinitropyridine (DADNPy)	12,4	10,86	-1,54
1,3,7,9-tetranitro-10H-phénothiazine-5,5-dioxine (TNPTD)	5,78	6,91	1,13
trinitrorésorcinol (TNR)	12,3	11,39	-0,91
trinitrocrésol (TNCr)	5,21	8,3	3,09
5,7-dinitro-(2,4,6-trinitrophényl)-1H-1,2,3-benzotriazole (BTX)	6,5	7,33	0,83
1,4,5-trinitronaphtalène (TNN)	10,97	9,46	-1,51
2,2',4,4',6,6'-hexanitrostilbène (HNS)	5,32	6,62	1,3
2,2',4,4',6,6'-hexanitrodiphénylméthane (DPM)	4,1	4,65	0,55
2-chloro-1,3,5-trinitrobenzène (CTB)	6,71	5,92	-0,79
2,2',4,4',6,6'-hexanitrodiphénylamine (DPA)	5,02	6,72	1,7
1,3-diamino-2,4,6-trinitrobenzène (DATB)	10,97	12,94	1,97
1,3,5-triamino-2,4,6-trinitrobenzène (TATB)	17,75	17,73	-0,02
2,2',4,4',6,6'-hexanitrobiphényle (HNB)	5,03	4,95	-0,08
1,3,7,9-tetranitro-10H-phénoxazine (TENPO)	5,12	6,21	1,09
3-méthyle-2,2',4,4',6,6'-hexanitrodiphénylsulfure (MEDIPS)	5,71	7,4	1,69
2,4,6-trinitrotoluène (TNT)	6,85	5,63	-1,22
		MAE (%)	18,6

Tableau 21 - Sensibilités à la décharge électrique (en J) expérimentales et prédites des 26 molécules du jeu d'entraînement

D'ailleurs, en termes d'interprétation, il est également important de noter que la chaleur de décomposition était prédite à l'aide de descripteurs caractéristiques du nombre et de la réactivité des groupements nitro, alors que le nombre de NO_2 n'intervient pas dans ce modèle (Eq. 3.30). Lorsque l'on s'intéresse à la chaleur de décomposition, il s'agit de caractériser l'énergie libérée durant la décomposition. Le nombre de groupements nitro est donc important. La sensibilité, quant à elle, traduit l'énergie d'activation de la réaction de décomposition et donc seule la dissociation du premier groupement nitro avec le reste de la molécule est mise en jeu. Il s'agit de l'énergie nécessaire à la première étape de décomposition et non de l'énergie relative à l'intégralité du processus. Il y a donc bien une cohérence au niveau de l'interprétation.

3.3.2- Validation externe et domaine d'applicabilité

Si ce modèle est corrélé, robuste et intéressant d'un point de vue de l'interprétation des phénomènes, il reste à en estimer le degré de prédictivité. Le modèle a donc été appliqué aux 15 molécules de la base de données n'ayant pas encore été utilisées (jeu de validation). Les résultats en prédiction ne sont pas du tout satisfaisants (R^2<0,01), puisque les valeurs prédites (cf. Figure 48) peuvent dévier des valeurs expérimentales de manière très importante pour certains composés (avec une erreur moyenne absolue de 5 J).

La structure de certains composés, légèrement différente des composés du jeu d'entraînement, peut donner une explication (erreur de 11,72 J pour la molécule NONA) mais, pour d'autres, cette hypothèse n'est pas valable. Par exemple, l'erreur pour la molécule de 1,4-nitrobenzène (1,4-DNB) atteint 15,17 J alors que de nombreuses molécules proches en termes de structure sont présentes dans le jeu d'entraînement (PA, TNT, TNCr…).

Molécules	Exp [207,211]	Eq. (3.30)	Diff.
1,3-dinitrobenzène (1,3-DNB)	1,84	3,15	-1,31
1,4-dinitrobenzène (1,4-DNB)	3,21	18,38	-15,17
3,3'-diméthyl-2,2',4,4',6,6'-hexanitrobiphényle (BITNT)	6,73	4,28	2,45
1,3-dichloro-2,4,6-trinitrobenzène (DCTB)	8,01	2,55	5,46
3,3'-diméthyle-2,2',4,4',6,6'-hexanitrodiphénylsulfure (DIMEDIPS)	10,47	8,57	1,90
3,3'-diméthyl-2,2',4,4',6,6'-hexanitrobenzène (DiMeHNAB)	5,97	13,37	-7,40
2,2',4,4',6,6'-hexanitrodiphénylsulfure (DIPS)	5,76	2,56	3,20
2,2',4,4',6,6'-hexanitrodiphénylsulfone (DIPSO)	5,00	10,54	-5,54
2,2',4,4',6,6'-hexanitrodiphényléthane (DPE)	5,00	3,89	1,11
2,2',4,4',6,6'-hexanitroazobenzène (HNAB)	5,40	8,20	-2,80
2,2',4,4',6,6'-hexanitrooxanilide (HNO)	9,43	14,58	-5,15
2,2',2",4,4',4",6,6',6"nonanitro-m-terphényle (NONA)	4,72	16,44	-11,72
2,6-bis(picrylamino)-3,5-dinitropyridine (PYX)	12,51	8,90	3,61
2,4,6-tris(3-méthyl-2,4,6-trinitrophénylamino)-s-triazine (TMPM)	4,53	6,84	-2,31
1,4,6-tris(2,4,6-trinitrophényll-s-triazine (TPT)	4,76	10,61	-5,85
		MAE (%)	63,3

Tableau 22 - Sensibilités à la décharge électrique (en J) expérimentales et prédites des 15 molécules du jeu de validation

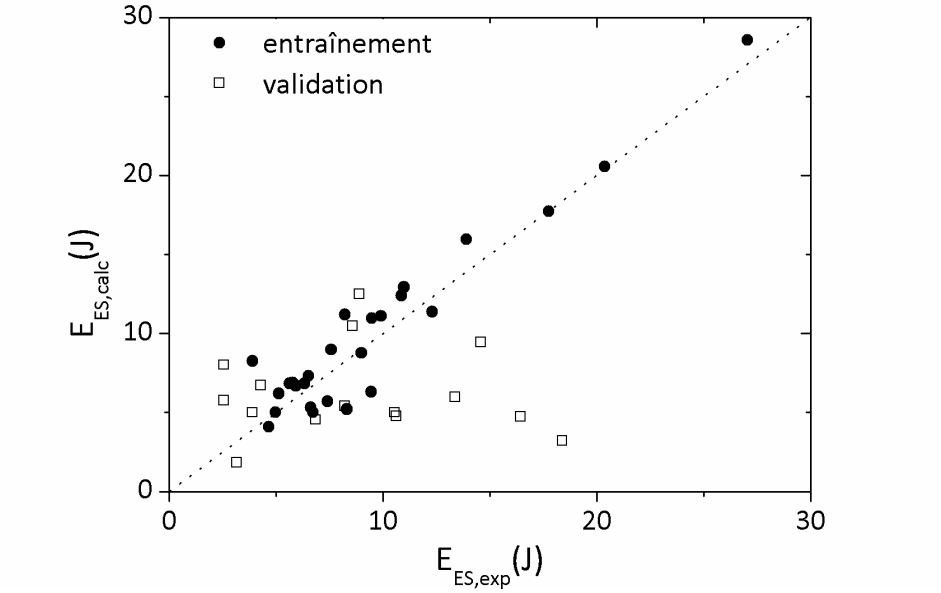


Figure 48 - Sensibilités à la décharge électrique prédites (Eq. 3.30) et expérimentales [207,211]

L'analyse du domaine d'applicabilité du modèle (réalisée sur la base des distances euclidiennes avec un critère d'intégration de 95% des molécules du jeu d'entraînement) confirme d'ailleurs cette observation. En effet, toutes les molécules du jeu de validation entrent dans ce dernier. Cela signifie donc que le manque de prédictivité ne peut être attribué à des molécules extrêmes n'entrant pas dans le domaine d'applicabilité. Le modèle n'est donc pas prédictif, au sein même de son domaine d'applicabilité (R^2_{in}<0,01).

Au final, ce modèle bien que satisfaisant en termes de corrélation, robustesse et interprétation, est inutilisable en termes de prédiction comme l'illustre la Figure 48. Afin de comprendre ce manque de prédictivité, il faut tout d'abord se souvenir que les incertitudes expérimentales sur les mesures de sensibilité à la décharge électrique sont importantes [197]. Comme ces dernières se propagent dans notre modèle, l'incertitude sur les valeurs calculées est encore plus élevée.

De plus, si les phénomènes mis en jeu au cours de l'initiation par une décharge électrique ne sont pas complètement élucidés, il est assez évident qu'il ne s'agit pas d'un processus uniquement chimique. Notamment, les phénomènes de transferts énergétiques au sein du matériau sont fortement influencés par les propriétés physiques, telles que sa granulométrie. Or, ceux-ci ne peuvent être pris en compte dans un modèle de type QSPR.

3.3.3- Analyse par composantes principales

Les considérations précédentes ne suffisent pas à expliquer complètement la faible prédictivité du modèle. Une analyse par composantes principales (PCA) a donc été réalisée. La représentation des molécules de la base de données dans le plan formé par les deux premières composantes principales de l'analyse met en évidence la présence de 3 sous-groupes de molécules. Or, leur présence a une influence très néfaste sur le modèle.

Sur la gauche du graphique (en Figure 49), sont identifiées les molécules mono-aromatiques et les naphtalènes. Plus au centre, les molécules bicycliques se distinguent. Enfin, des molécules présentant un nombre de cycles aromatiques supérieur ou égal à 3 se retrouvent sur la droite du graphique.

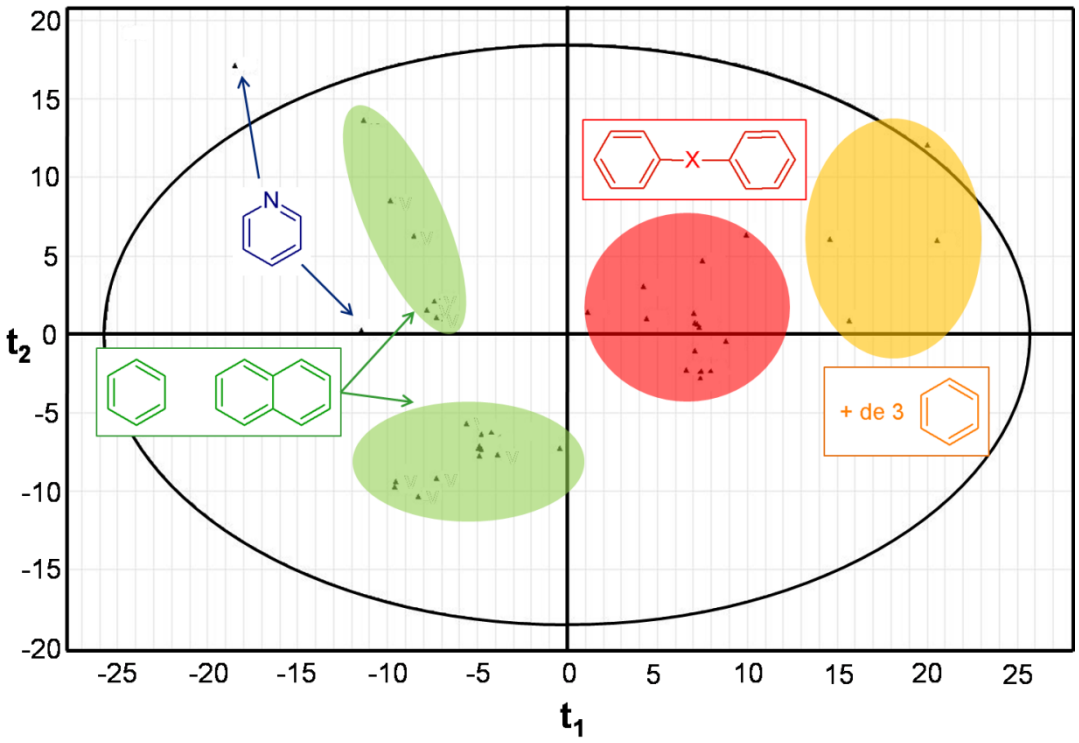


Figure 49 - Représentation des 50 molécules dans le plan des deux premières composantes principales d'une analyse PCA

Le nombre de cycles aromatiques peut être un moyen de discriminer ces molécules au sein de la base de données globale et prendre en compte ces distinctions. Or, ce descripteur est déjà inclus dans le jeu des descripteurs calculé et n'est pas suffisant pour améliorer la prédiction. Pour rappel, l'analyse PCA est une approche non supervisée, ce qui signifie qu'elle ne prend pas en compte la propriété mais définit l'espace chimique en fonction des seuls descripteurs. Il semble donc que le caractère qui discrimine le plus les molécules de la base de données dans l'espace chimique défini par les descripteurs calculés, à savoir le nombre de cycles, ne soit pas corrélé avec la propriété.

3.4- Bilan

Cette étude a permis de montrer la difficulté de mise en place de modèles QSPR prédictif avec des jeux de données aussi variées. D'autres essais de partages du jeu de données entre entraînement et validation ont été réalisés. Mais, dans tous les cas, les mêmes conclusions sont réalisées et les modèles obtenus, bien que corrélés et robustes, ne présentent aucun pouvoir prédictif. D'ailleurs, aucun modèle existant ne démontre de prédictivité significative. Pour la majeure partie des études de ce type, aucune étape de validation externe n'est même réalisée.

Seul le modèle de Keshavarz, [216] présenté en équation (3.29), présente une telle analyse. Une prédictivité de 0,77 est alors proposée. Les données expérimentales sont extraites de la même base de données que celle employée ici. Or, certaines molécules, pourtant sans particularité au niveau de leur structure chimique, n'ont pas été intégrées à l'analyse et les déviations observées pour ces valeurs sont très importantes, comme indiqué dans le Tableau 23, avec une erreur moyenne de 5,8 J (soit 91%). Ce modèle n'est donc pas prédictif. L'étude réalisée ici permet d'expliquer ce manque de prédictivité.

molécules	Exp [207,211]	Eq. (3.29) [216]	Diff.
NONA	16,44	6,98	-9,46
DIPS	2,56	7,55	4,99
PYX	8,90	7,73	-1,17
HNO	14,58	7,84	-6,74
DiMeHNAB	13,37	3,99	-9,38
1.3-DNB	3,15	12,26	9,11
1.4-DNB	18,38	12,26	-6,12
TPT	10,61	5,29	-5,32
TMPM	6,84	1,01	-5,83
DCTB	2,55	6,07	3,52
NTO	8,98	11,41	2,43
		MAE (%)	91,4

Tableau 23 - Sensibilités à la décharge électrique (en J) expérimentales et prédites des 11 molécules de la base de données n'ayant pas été employées par Keshavarz dans son analyse

Plusieurs voies d'approfondissement dans la recherche de modèles prédictifs sont envisagées :

- Une première voie peut concerner le développement de nouveaux descripteurs qui permettraient de corréler correctement, si ce n'est l'intégralité, le plus grand nombre possible de molécules de la base de données avec la sensibilité à la décharge électrique.
- Une autre approche consiste à distinguer les molécules par sous-groupes, ce qui n'a pas été possible dans notre étude du fait de la taille réduite de la base de données. En effet, le sous-groupe le plus large, constitué des composés mono-aromatiques et des naphtalènes, ne contient que 15 molécules, ce qui est insuffisant pour une scission en jeux d'entraînement et de validation. Pour aller dans cette voie, qui semble la plus viable du fait de la diversité des structures au sein de la base de données, l'obtention de nouvelles données expérimentales est indispensable. D'ailleurs, l'obtention de bases de données fiables et en nombre important est un réel problème pour ce type de propriétés.

4- Sensibilité à l'impact

4.1- Généralités

La sensibilité à l'impact (ou au choc) d'un matériau énergétique est une mesure de la tendance du matériau à engendrer une détonation explosive sous l'effet d'un impact. Si différents dispositifs expérimentaux existent [57], ils reposent tous sur le même principe. Par exemple, le dispositif utilisé à l'INERIS (appelé mouton de choc) est présenté en Figure 50. Il s'agit de laisser tomber une masse déterminée (10kg), aussi appelée "mouton", le long d'une colonne, sur un échantillon d'explosif (solide) et de déterminer la hauteur nécessaire pour le faire réagir. Leur but est de représenter au mieux le phénomène d'impact sur les matériaux énergétiques.

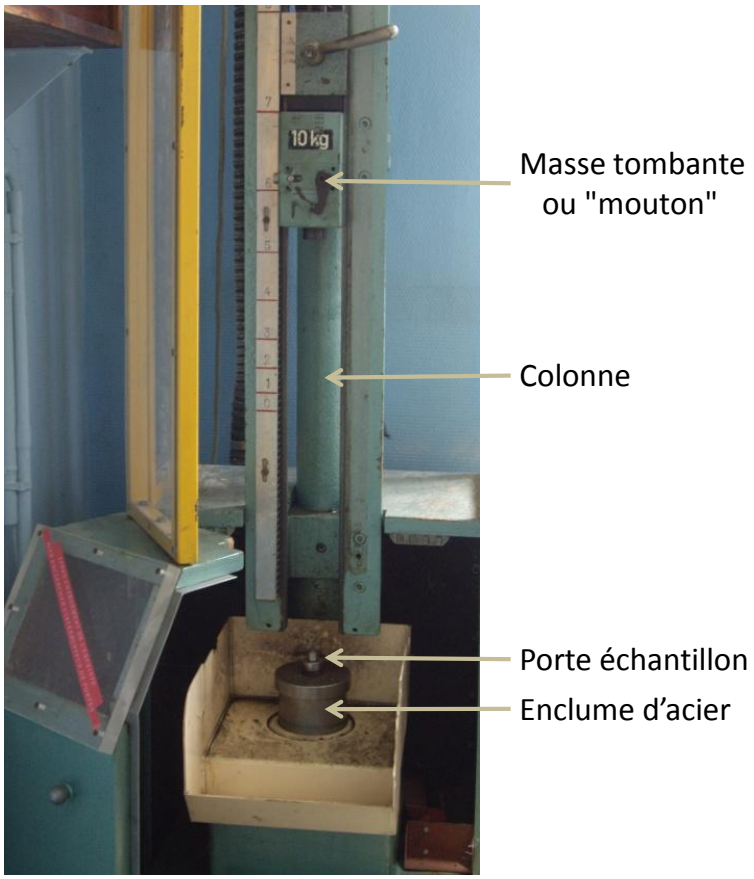


Figure 50 – Mouton de choc pour la mesure de la sensibilité à l'impact (Appareillage INERIS)

Pour un explosif donné, des essais sur une série d'échantillons sont réalisés. Le résultat d'un essai est positif si une réaction est observée au niveau de l'échantillon (explosion, apparition de fumées, bruit). Il s'agit de déterminer la hauteur $h_{50\%}$ à laquelle 50% des essais sont positifs [40].

Les matériaux possédant des $h_{50\%}$ peu élevées sont les plus sensibles à l'impact puisque l'énergie à apporter pour faire réagir le matériau est faible. Le développement ainsi que la manipulation (mise en œuvre, transport, stockage, utilisation) de tels composés est donc dangereuse. Disposer d'outils de criblage des molécules d'intérêt réel sera particulièrement intéressant afin de limiter le recours à des tests expérimentaux dangereux.

Concernant la compréhension du phénomène, au niveau macroscopique, l'initiation par impact est proposée comme principalement engendrée par la dissipation de l'énergie issue de la déformation plastique du matériau (au moment de la phase d'impact) ainsi qu'à ses propriétés en cisaillement [218]. Selon ces théories, le volume dans lequel l'énergie de l'impact a lieu est suffisamment petit pour que l'augmentation de température du matériau, à l'intérieur de ce volume, soit plusieurs fois supérieure à la température de décomposition. Dans le cas d'une initiation à l'impact de faible énergie [219], un facteur important est la vitesse à laquelle l'énergie, dissipée depuis l'évènement de l'impact au travers des modes vibrationnels du matériau énergétique, entraîne le démarrage de la chimie.

Quoi qu'il en soit, les mécanismes moléculaires conservent une grande importance dans le phénomène puisqu'il s'agit d'initier une réaction chimique de décomposition au sein du matériau. Zeman a d'ailleurs dédié une partie de ces travaux à la mise en évidence de corrélations existant entre la sensibilité à l'impact et les données de réactivité expérimentales (températures et énergies d'activation de décomposition obtenues par calorimétrie) pour les composés nitrés [220-223].

4.2- Travaux antérieurs

Le développement d'outils prédictifs de type QSPR pour la sensibilité à l'impact des matériaux énergétiques est un champ de recherche récurrent depuis un certain nombre d'années. Ceci peut s'expliquer par le fait qu'il s'agit de l'épreuve de sensibilité la plus commune. Les données expérimentales disponibles sont donc plus nombreuses que pour les autres propriétés explosives.

4.2.1- Balance en oxygène

Les premiers travaux ont considéré des relations simples avec des composantes constitutives des composés explosifs. Sur la base de cette propriété, Lothrop a développé la balance en oxygène (OB) (cf. Equation 3.14) comme indicateur des performances des composés explosifs [44].

Kamlet a alors relié cette grandeur de manière quantitative à la sensibilité à l'impact [224]. Deux équations, présentant des corrélations importantes (R^2=0,92 et 0,94), ont ainsi été développées pour les nitroaromatiques selon qu'ils présentent ou non un groupement de type CH (ex : $-CH_3$, $-CH_2-CH_3$) en position ortho de NO_2 (respectivement Eq. 3.31 et 3.32).

$$\log h_{50\%} = 1{,}73 - 0{,}32\ OB \qquad (3.31)$$

$$\log h_{50\%} = 1{,}33 - 0{,}26\ OB \qquad (3.32)$$

D'autres études plus récentes se sont également appuyées sur la balance en oxygène pour caractériser la sensibilité à l'impact de composés similaires [225,226]. En particulier, Cao [226] a proposé un modèle très corrélé (R^2=0,98) intégrant également le carré de la charge sur le groupement nitro Q_{NO2}, pour une série de 19 composés nitrés variés (nitramines, nitroaromatiques).

$$h_{50\%}\,(cm) = 0{,}1926 + 98{,}64\ Q_{NO2}^2 - 0{,}03405\ OB \qquad (3.33)$$

4.2.2- Autres descripteurs moléculaires

D'autres propriétés moléculaires, telles que l'électronégativité [227], des chaleurs de réaction ou la composition moléculaire [228], ont été corrélées linéairement avec la sensibilité à l'impact. Des efforts ont été menés pour le développement de relations avec des descripteurs dérivés des structures électroniques des composés, via la modélisation purement structurale [229-232] et des approches basées sur les mécanismes de décomposition [233-235]. Delpuech [236-238] a notamment corrélé la sensibilité à l'impact des composés explosibles avec leurs propriétés électroniques, en particulier avec leur répartition de charge à l'état excité.

Rice et coll. [54] se sont, eux, intéressés aux propriétés électroniques d'explosifs nitrés (nitroaromatiques et nitramines) dans leur état fondamental via la distribution du potentiel électrostatique en surface (ESPs) au niveau DFT. Les ESPs d'une série de composés nitroaromatiques ont été analysés afin d'identifier les comportements de la distribution électronique qui pourraient être reliés au degré de sensibilité à l'impact des matériaux étudiés. Une corrélation de $h_{50\%}$ avec l'accumulation de charge positive au niveau des liaisons covalentes de la molécule a notamment été mise en évidence.

Parmi les autres corrélations linéaires disponibles dans la littérature, Edwards et coll. [239] ont mis en relation les sensibilités à l'impact d'une série de nitramines avec différentes quantités dérivées de la mécanique quantique, telles que les énergies orbitalaires (HOMO, LUMO) et les chaleurs de détonation correspondant aux différents mécanismes réactionnels. Si aucune relation prédictive n'est observée, des corrélations significatives avec des grandeurs constitutionnelles caractéristiques simples des matériaux ont été mises en évidence : nombre d'atomes d'oxygène (R^2=0,58) et de nitro (R^2=0,61). Zhang et coll. [240] ont pour leur part calculé la charge sur NO_2 de 26 composés nitrés à l'aide de la méthode de Mulliken au niveau DFT et ont observé quelques tendances qualitatives la reliant à $h_{50\%}$. Enfin, Song et coll. [241] ont observé une relation entre $h_{50\%}$ et les énergies de dissociation des liaisons C-NO_2 mais au sein d'une très petite série de nitroalcanes (6 molécules).

4.2.3- Régressions multi-linaires (MLR)

Pour ce qui est des modèles multi-linéaires, Badders et coll. [242] ont proposé un modèle à 3 paramètres pour un jeu de 22 composés nitroaromatiques.

$$\log h_{50\%} = 7{,}096\ (\varepsilon_{LUMO}-\varepsilon_{HOMO}) - 7{,}096\ V_{mid} + 0{,}000529\ E + 0{,}711 \qquad (3.34)$$

où ε_{HOMO} et ε_{LUMO} sont les énergies respectives des HOMO et LUMO, V_{mid} est le potentiel à mi-distance, tel que défini précédemment (Eq. 3.16), et E est l'énergie totale de la molécule.

Si la corrélation est en apparence significative (R^2=0,83), la taille réduite de la base de données ne permet pas d'avoir une idée correcte de la prédictivité du modèle, puisqu'aucune validation externe n'a été réalisée.

Il en est de même pour les modèles de Keshavarz [243-245], basés sur des descripteurs constitutionnels. Les premiers d'entre eux [243] concernent une série de 104 composés explosifs variés de formules $C_aH_bN_cO_d$, considérés en 3 classes distinctes (respectivement polynitroaromatiques et benzofuroxanes, nitamines et polynitroaromatiques présentant des possibilité d'interactions C-H alpha ou N-CH alpha entre le groupement nitro et un substituant en ortho, nitroaliphatiques). Un modèle a été développé pour chaque classe de composés. Les modèles obtenus présentent des déviations moyennes de 24 cm (en moyenne pour les deux classes de nitraromatiques) et 40 cm pour les nitroaliphatiques.

$$\log h_{50\%} = (11{,}76\ a + 61{,}72\ b + 26{,}89\ c + 11{,}48\ d) / MW \qquad (3.35)$$

$$\log h_{50\%} = (47{,}33\ a + 23{,}50\ b + 2{,}357\ c - 1{,}105\ d) / MW \qquad (3.36)$$

$$\log h_{50\%} = (81{,}40\ a + 16{,}11\ b - 19{,}08\ c + 1{,}089\ d) / MW \qquad (3.37)$$

où MW est la masse moléculaire.

Plus récemment, il proposa deux nouveaux modèles : l'un pour les composés nitrohétérocycliques [244], l'autre pour les nitroaliphatiques [245]. Ils font intervenir des paramètres constitutionnels supplémentaires, caractérisant la présence de fragments particuliers, avec des déviations moyennes respectives de 59 cm et 40 cm.

$$\log h_{50\%} = (46{,}2923\ a + 35{,}6305\ b - 7{,}7005\ c + 7{,}9425\ d + 44{,}4167\ n_{\text{-CNC-}} + 102{,}2749\ n_{\text{-CNCN-}}) / MW \qquad (3.38)$$

$$\log h_{50\%} = 2{,}47 + 0{,}371\ [100(a + b/2 - d)/MW] - 0{,}485\ [100c/MW] + 0{,}185\ n_{\text{R-C(NO2)2-CH2-}} \qquad (3.39)$$

où $n_{\text{-CNC-}}$, $n_{\text{-CNCN-}}$ et $n_{\text{R-C(NO2)2-CH2-}}$ sont les nombres respectifs de fragments -CNC-, -CNCN- et R-$C(NO_2)_2$-CH_2- au sein de la molécule.

4.2.4- Autres approches multi-variables

D'autres types d'approches multi-variables ont également été envisagés. Cho et coll. [246] ont mis en place des modèles par réseaux de neurones pour une série de 258 molécules nitrées variées (nitroaromatiques, nitropyridines, nitroimidazoles, nitramines, etc.) à partir de 39 descripteurs variés (constitutionnels, géométriques, quantiques) de la structure moléculaire a priori corrélés avec la sensibilité à impact. Les modèles obtenus présentent des corrélations jusqu'à R^2=0,848. Ils ne font qu'obtenir des résultats similaires à ceux de Legendre qui, avec une méthodologie similaire mais des descripteurs différents (39 descripteurs constitutionnels et quantiques), a obtenu une corrélation R^2=0,882 pour une base de données légèrement moins nombreuses (204 molécules) [53].

Finalement, les seuls modèles, avançant des valeurs de prédictivité par validation externe, ont été développés par Wang et al. [247]. Dans cette étude, différents modèles ont été développés suivant différentes approches (MLR, PLS et réseaux de neurones) avec une série de 16 descripteurs topologiques. Sur une base de 156 nitroaromatiques, nitramines et nitroaliphatiques, le modèle par réseau de neurones est le plus performant avec une corrélation R^2=0,816 et une prédictivité R^2_{ext}=0,740. Pour les approches MLR et PLS, diviser le jeu de données en fonction du type de composés permet d'améliorer les performances des modèles pour les nitramines (respectivement R^2_{ext}=0,856 et 0,864 pour les analyses MLR et PLS) et les nitroaliphatiques (R^2_{ext}=0,932 et 0,974) mais pas pour les nitroaromatiques (R^2_{ext}=0,457 et 0,424). Cela dit, l'interprétation des modèles développés est peu aisée du fait du nombre élevé de descripteurs inclus (16).

Si des approches de type QSPR existent donc pour la sensibilité à l'impact, aucun modèle, à notre connaissance, n'a encore été à la fois validé (par validations internes et externes), défini dans un domaine d'applicabilité et interprété d'un point de vue phénoménologique.

4.3- Analyse QSPR

Les données expérimentales de sensibilité à l'impact ($h_{50\%}$) sont plus disponibles que celles caractérisant la stabilité thermique et la sensibilité à la décharge électrique qui ont été étudiées précédemment. Les données utilisées dans cette étude ont été extraites de la littérature [192]. Il s'agit de 50 composés nitroaromatiques, présentés en Figure 51. Dans la publication de référence, elles ont été considérées en deux sous-groupes afin de distinguer les molécules pouvant former, au cours du processus de décomposition, une liaison C-H alpha entre le groupement nitro et le substituant de type CH adjacent. Il s'agit en fait de distinguer les molécules pouvant avoir un mécanisme de décomposition similaire au TNT [248]. Une division similaire a d'ailleurs été réalisée pour la chaleur de décomposition où les composés substitués en position ortho avaient été distingués des non-ortho.

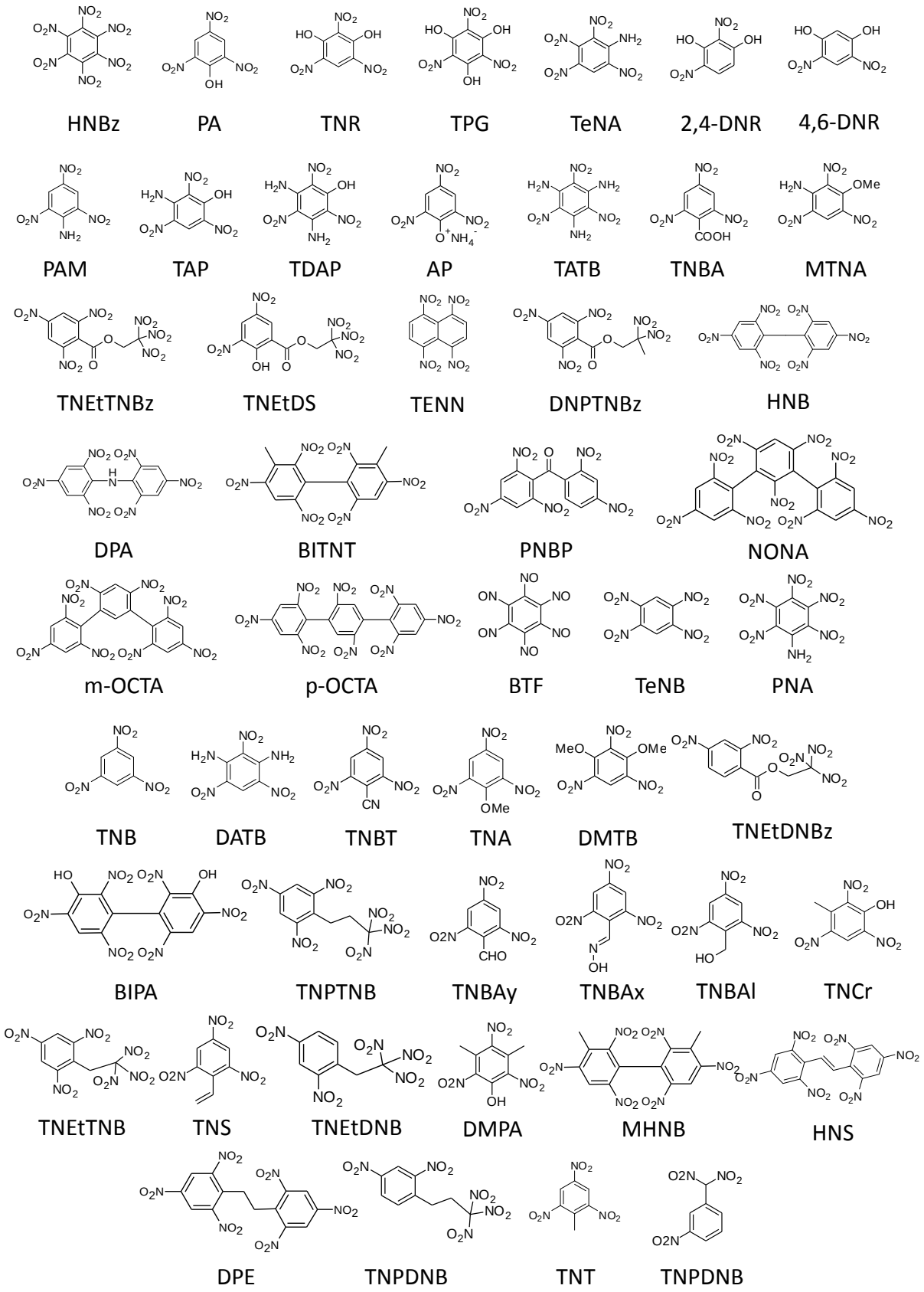


Figure 51 – 50 molécules de la base de données
(voir Tableau 24 et 25 pour la dénomination complète des molécules)

4.3.1. Base de données complète

Dans un premier temps, l'intégralité des molécules de la base de données a été considérée. Pour chacune des 50 molécules, la structure moléculaire a été optimisée au niveau PBE0/6-31+G(d,p) puis plus de 300 descripteurs moléculaires ont été calculés. La procédure BMLR a ensuite été appliquée afin d'obtenir un modèle le plus corrélé possible avec la sensibilité à l'impact expérimentale tout en ayant un nombre de descripteurs suffisamment faible pour éviter toute sur-paramétrisation et être, autant que possible, interprétable.

Avec cette méthodologie, le meilleur modèle obtenu contient deux descripteurs et ne présente ni corrélation (R^2=0,55) ni robustesse (R^2_{cv}=0,47) significative.

$$h_{50\%}\ (\text{cm}) = 618{,}3 + 450{,}4\ S_{ZX/ZX} - 279{,}1\ n_L \qquad (3.40)$$

où $S_{ZX}/_{ZX}$ est une projection relative du volume moléculaire sur un plan ZX de la molécule et n_L est le nombre de niveaux électroniques occupés par atome dans la molécule.

Les résultats en termes de corrélation et robustesse, n'étant d'ores et déjà pas satisfaisant. La prédictivité et le domaine d'applicabilité du modèle n'ont pas été envisagés.

4.3.2. Molécules sans liaison C-H alpha

La difficulté à mettre en place un modèle fiable peut s'expliquer par la complexité des mécanismes de décomposition des nitroaromatiques. Les molécules ont donc ensuite été réparties de manière similaire à la référence originale de Storm [192].

La base de données contient 35 molécules sans liaison C-H alpha possible avec le groupement nitro. Ce chiffre est suffisant pour envisager une validation externe. Un jeu de validation de 10 molécules a donc été mis à part afin d'être utilisé pour avoir une estimation du pouvoir prédictif du modèle (cf. Tableau 24). La séparation entre jeu d'entraînement (25 molécules) et de validation (10 molécules) a été réalisée de manière aléatoire.

L'analyse BMLR a été utilisée sur le jeu d'entraînement et le meilleur modèle obtenu comprend cette fois 6 descripteurs.

$$h_{50\%}\ (\text{cm}) = 1674{,}1 + 7381{,}7\ I - 1727{,}4\ q_{O,max} - 1114{,}7\ V_{C,min} + 363{,}8\ V_{C,moy} - 85{,}9\ BO_{c,min} - 11667\ q'_{O,max} \qquad (3.41)$$

où I est le moment d'inertie, $q_{O,max}$ la charge atomique nette maximale pour un atome d'oxygène (méthode de Mulliken), $V_{C,min}$ la valence minimale d'un atome de carbone, $V_{C,moy}$ la valence moyenne d'un atome de carbone, $BO_{c,min}$ l'ordre de liaison minimal d'un atome d'oxygène et $q'_{O,max}$ une autre expression de la charge partielle maximale pour un atome d'oxygène (méthode de Zefirov).

Si le modèle est cette fois bien corrélé (R^2=0,96) et, a priori, robuste (R^2=0,93), il n'a pas de pouvoir prédictif significatif puisque la corrélation pour le jeu de validation est faible (R^2_{ext}=0,57) comme illustré en Figure 52.

Molécules	Exp [192]	Eq. (3.40)	Eq. (3.41)
Entraînement			
2,2',4,4',6,6'-hexanitrobenzène (HNBz)	12	9	12
2',2'-dinitropropyl-2,4,6-trinitrobenzoate (DNPTBz)	214	39	178
2,4,6-trinitrophloroglucinol (TNG)	27	111	4
2,2',2'',4,4',4'',6,6',6''-nonanitro-m-terphényle (NONA)	39	52	32
2,3,4,6-tetranitroaniline (TeNA)	41	59	42
2,4,6-trinitroresorcinol (TNR)	43	57	74
2',2',2'-trinitroéthyl-3,5-dinitrosalicylate (TNEtDS)	45	33	46
2,2',4,4',6,6'-hexanitrodiphénylamine (DPA)	48	70	62
2,2',4,4',6-pentanitrobenzophénone (PNBP)	54	59	52
2,2',2'',4,4'',6,6',6''-octanitro-p-terphényle (p-OCTA)	59	-17	63
2,2'',4,4',4'',6,6',6''-octanitro-m-terphényle (m-OCTA)	63	62	36
2,2',4,4',6,6'-hexanitrobiphényle (HNB)	85	12	78
acide picrique (PA)	87	92	103
1,4,5,8-tetranitronaphthalène (TENN)	100	133	119
acide 2,4,6-trinitrobenzoïque (TNBA)	109	86	156
1-hydroxy-3,5-diamino-2,4,6-trinitrobenzène (TDAP)	120	151	112
3,3'-diméthyl-2,2',4,4',6,6'-hexanitrobiphényle (BITNT)	132	89	100
picrate d'ammonium (AP)	135	148	108
2,4,6-trinitro-3-aminophénole (TAP)	138	122	143
2,4,6-trinitroaniline (PAM)	177	192	174
2,4-dinitroresorcinol (2,4-DNR)	296	210	287
4,6-dinitroresorcinol (4,6-DNR)	320	212	303
3-méthoxy-2,4,6-trinitroaniline (MTNA)	320	184	337
1,3,5-Triamino-2,4,6-trinitrobenzène (TATB)	320	246	326
2',2',2'-trinitroéthyl-2,4,6-trinitrobenzoate (TNEtTNBz)	24	-8	33
	MAE (%)	49,4	20,4
Validation			
pentanitroaniline (PNA)	15	56	-86
1,2,4,5-tetranitrobenzène (TeNB)	27	69	127
benzotrifuxane (BTF)	50	26	-157
3,3'-dihydroxy-2,2',4,4',6,6'-hexanitrobiphényle (BIPA)	40	29	36
1,3,5-trinitrobenzène (TNB)	100	166	180
2',2',2'-trinitroéthyl-3,5-dinitrobenzoate (TNEtDNBz)	73	67	45
2,4,6-trinitrobenzonitrile (TNBT)	140	59	117
2,4,6-Trinitroanisole (TNA)	192	136	403
1,3-diamino-2,4,6-trinitrobenzène (DATB)	320	232	249
1,3-diméthoxy-2,4,6-trinitrobenzène (DMTB)	251	215	327
	MAE (%)	70,7	176,5

Tableau 24 - Sensibilités à l'impact (en cm) expérimentales et prédites pour les composés de la base de données ne présentant pas de possibilité de liaisons C-H alpha

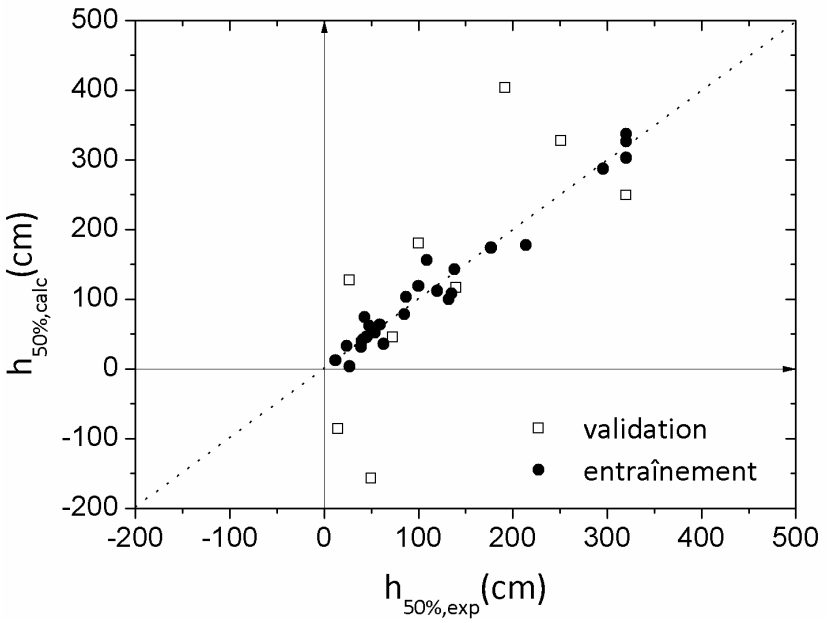


Figure 52 – Sensibilité à l'impact (en cm) expérimentales et prédites des composés nitroaromatiques de présentant pas de possibles liaisons C-H alpha

4.3.3. Molécules avec liaison C-H alpha

Pour les 15 molécules avec liaison C-H alpha, la mise en place d'un jeu de validation n'est pas possible du fait du trop petit nombre de composés (cf. Tableau 25). Un modèle a tout de même été mis en place pour mettre, éventuellement, en évidence certaines tendances ou problèmes pouvant être liés au traitement de ces molécules.

Molécules	Exp [192]	Eq. (3.40)	Eq. (3.42)
1-(3,3,3-trinitropropyl)-2,4-dinitrobenzène (TNPDNB)	31	93	23
1-(3,3,3-trinitropropyl)-2,4,6-trinitrobenzène (TNPTNB)	21	92	25
1-(2,2,2-trinitroéthyl)-2,4-dinitrobenzène (TNEtDNB)	3	92	25
2,4,6-trinitrobenzaldehyde (TNBAy)	36	97	33
2,4,6-trinitrobenzaldoxime (TNBAx)	42	100	47
2,4,6-trinitrotoluène (TNT)	160	192	160
1-dinitrométhyl-3-nitrobenzène (DNMNB)	105	137	110
2,4,6-trinitrobenzyl alcool (TNAl)	52	173	58
2,4,6-trinitro-m-crésol (TNCr)	191	143	182
1-(2,2,2-trinitroethyl)-2,4,6-trinitrobenzène (TNEtTNB)	13	17	19
2,4,6-trinitrostyrène (TNS)	32	137	25
3,5-diméthyl-2,4,6-trinitrophénole (DMPA)	77	237	78
3-methyl-2,2',4,4',6,6'-hexanitrobiphényle (MHNB)	53	55	53
hexanitrostilbène (HNS)	39	62	37
2,2',4,4',6,6'-hexanitrobibenzyle (DPE)	114	99	122
	MAE (%)	317,5	60,1

Tableau 25 - Sensibilités à l'impact (en cm) expérimentales et prédites des composés de la base de données présentant des possibilités de liaisons C-H alpha

L'analyse BMLR ne propose pas une réponse très stable, dans le sens où la robustesse des modèles (au sens du R^2_{cv}) oscille de manière importante en fonction du nombre de descripteurs comme montré en Figure 53.

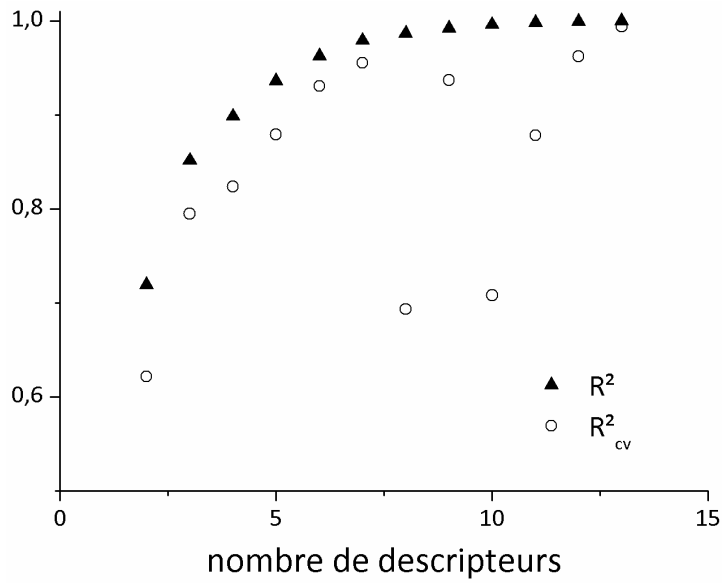


Figure 53 - Evolution des R^2 et R^2_{cv} en fonction du nombre de descripteurs

Finalement, le meilleur modèle obtenu (meilleur compromis entre corrélation et degré de paramétrisation) est une équation à 6 paramètres avec de bonnes performances entre termes de corrélation (R^2=0,99) et de robustesse (R^2_{cv}=0,85).

$$h_{50\%}\,(\text{cm}) = -52{,}0 + 10759\,R'_{O,max} - 0{,}3\,\nu_{norm,min} - 39{,}9\,\mu_{tot,mol} + 5641{,}8\,R'_{N,min} - 81{,}2\,BO_{c,min} + 1445{,}1\,IP \quad (3.42)$$

où $R'_{O,max}$ est l'indice de réactivité à 1 électron maximal pour un atome d'oxygène, $\nu_{norm,min}$ la fréquence minimale d'un mode propre de vibration, $\mu_{tot,mol}$ le moment dipolaire total de la molécule, $R'_{N,min}$ l'indice de réactivité à 1 électron minimal pour un atome d'azote, $BO_{c,min}$ l'ordre de liaison minimal d'un atome d'oxygène et IP le potentiel d'ionisation.

L'évaluation du pouvoir prédictif du modèle n'est pas possible puisque nous ne disposons pas de suffisamment de molécules externes au jeu d'entraînement. Malgré les valeurs élevées de R^2 et R^2_{cv}, quelques doutes peuvent être portés sur la prédictivité du modèle du fait du manque de stabilité observé au cours de l'analyse BMLR.

4.3.4. Analyse par composantes principales

Une analyse PCA a finalement été réalisée sur le jeu complet des 50 molécules. Plusieurs sous-classes de molécules peuvent alors être distinguées sur le plan formé par les deux premières composantes principales, ce qui signifie que des modèles séparés devraient être considérés pour chaque sous-groupe. Par contre, la distinction ne se fait pas en fonction d'une éventuelle liaison C-H alpha comme considéré précédemment.

Comme montré en Figure 54, une première subdivision peut être faite en fonction du nombre de cycles aromatiques (comme pour la sensibilité électrique). De plus, une distinction particulière doit en plus être faite parmi les composés à un seul cycle, ceux présentant des substituants trinitroéthyles ou trinitrométhyles formant un sous-groupe distinct.

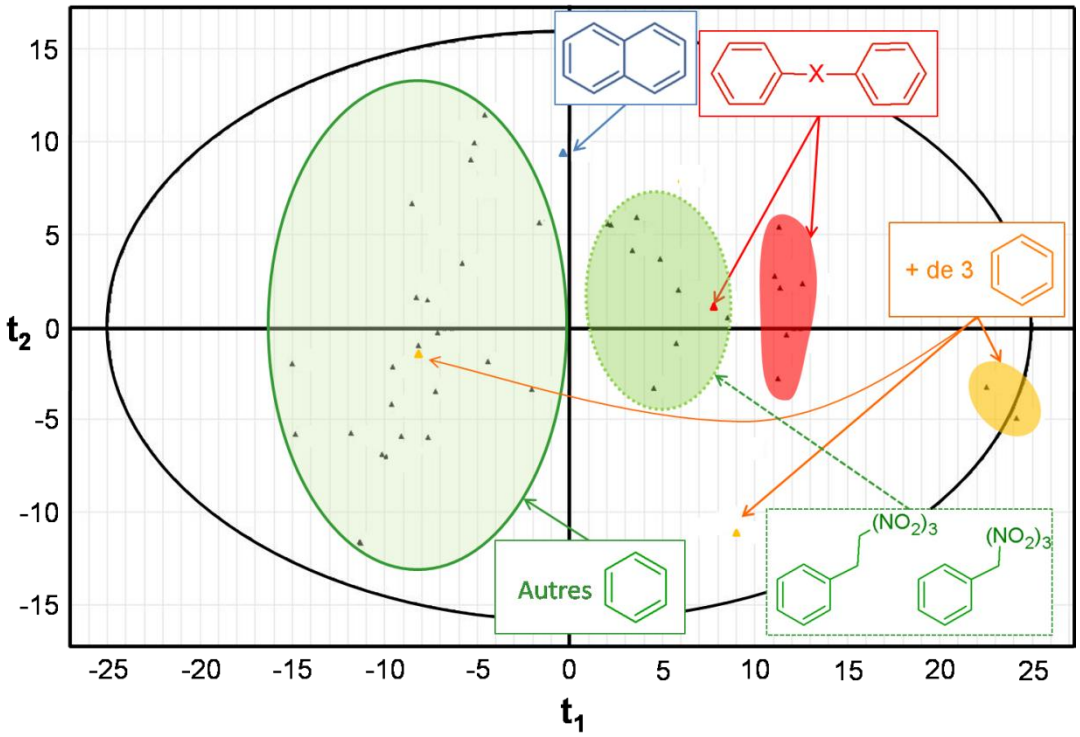


Figure 54 - Représentation des 50 molécules dans le plan des deux premières composantes principales d'une analyse PCA

Finalement, le nombre de composés dans chaque sous-groupe est insuffisant pour réaliser un partage entre jeux d'entraînement et de validation, le plus important contenant 26 molécules (molécules mono-aromatique sans groupement R-C$(NO_2)_3$). Néanmoins, ce nombre de molécules permet de développer un modèle à 5 descripteurs, à partir de l'analyse BMLR.

$$h_{50\%}\,(\text{cm}) = 2444{,}9 + 174{,}6\,\varepsilon_{HOMO-1} - 8059{,}6\,N_{C,min} - 158230\,N_{N,min} + 1354{,}2\,\text{FNSA-3} - 4{,}6\,S_{YZ} \quad (3.43)$$

où ε_{HOMO-1} est l'énergie de la LUMO-1, $N_{C,min}$ et $N_{N,min}$ les indices de réactivité nucléophile minimaux pour des atomes de carbone et d'azote, FNSA-3 est la fraction de la surface moléculaire chargée négativement (ordre 3) et S_{YZ} est une projection du volume moléculaire sur un plan YZ de la molécule.

Cette équation présente de bonnes corrélation (R^2=0,94) et robustesse (R^2_{cv}=0,86) mais sa prédictivité sur un jeu externe n'a pu être estimée du fait du manque de données disponibles, pour le type de molécules envisagé. Pour réaliser des modèles validés pour les différents sous-groupes, des données expérimentales plus nombreuses sont donc nécessaires.

4.4- Bilan

Pour cette propriété, avec la base de données employée, aucun modèle prédictif n'a pu être mis en évidence. C'est pourquoi l'analyse du domaine d'applicabilité n'a pas été réalisée. Le travail sur cette propriété mérite donc d'être consolidé.

On retrouve, dans la littérature [224,242], des corrélations réalisées avec $\log(h_{50\%})$ (cf. Eq. 3.31-3.39). Ici, une telle démarche n'a donné aucun résultat satisfaisant. La meilleure corrélation simple obtenue est de R^2=0,37 pour l'énergie de la LUMO+1 et aucun modèle multi-linéaire ne permet de l'améliorer.

Au cours de cette analyse, il a été mis en évidence l'influence néfaste d'une trop grande diversité de molécules sur le modèle, pour des bases de données de cette taille. D'ailleurs, jusque-là, malgré le nombre des analyses QSPR qui ont pu être initiées au cours du temps, les seuls modèles validés par une validation externe pour cette propriété ont été développés par Wang [247] et ceux mis en place pour les nitroaromatiques n'ont pas présenté non plus de prédictivité significative (R^2_{ext}=0,42-0,45), de meilleurs résultats étant obtenus pour d'autres classes de molécules (nitramines, nitroaliphatiques) ou en considérant les trois classes dans un seul modèle, à partir d'une base de données expérimentales importante (156 molécules). Dans ce dernier cas, le nombre de données employées permet de mieux caractériser la diversité des molécules.

Il semble donc que l'utilisation de bases de données plus importantes soit nécessaire pour mettre en place un modèle QSPR pour la sensibilité à l'impact des composés nitroaromatiques afin de développer :

- soit un modèle global concernant l'intégralité des composés nitrés mais sur la base d'un plus grand nombre de molécules représentant de manière plus étendu l'ensemble des composés de cette famille ;
- soit des modèles en fonction des sous-groupes identifiés lors de l'analyse PCA.

5- Conclusion

Des modèles QSPR ont été développés pour la prédiction des propriétés explosives des composés nitroaromatiques : en particulier leur stabilité thermique (chaleur et température de décomposition) et leurs sensibilités à la décharge électrique et à l'impact.

Dans un premier temps, différentes approches méthodologiques ont été considérées au regard de la chaleur de décomposition. Au niveau des descripteurs, l'intérêt des descripteurs issus de la chimie quantique a été démontré. Des corrélations significatives ont, en particulier, été observées pour des modèles QSPR basés sur les descripteurs globaux de la DFT conceptuels (potentiel d'ionisation, affinité électronique, électronégativité, dureté, indice d'électrophilicité). D'ailleurs, inclus dans un jeu étendu de plus de 300 descripteurs (constitutionnels, topologiques, géométriques et, donc, quantiques), ces derniers se distinguent puisqu'ils interviennent dans les meilleurs modèles QSPR, pour peu que le niveau de théorie employé pour les calculer soit adéquat. Leur insertion dans les modèles

permet de les rendre plus interprétables au niveau phénoménologique. En particulier, pour la chaleur de décomposition, si le nombre de groupements NO_2 (descripteur constitutionnel) est fondamental, il n'est pas suffisant, notamment pour distinguer les isomères entre eux. Aussi, l'inclusion de descripteurs quantiques, qui ne présentent pas la même défaillance, permet en plus de caractériser la réactivité de ces groupements au cours de la décomposition.

Au niveau de l'analyse des données, l'utilisation de l'analyse BMLR (*Best Multi Linear Regression*), telle qu'implémentée dans le logiciel CodessaPro, permet de définir, de manière automatisée, les modèles multi-linéaires réalisant les meilleurs compromis entre corrélation et nombre de descripteurs. Il s'agit ensuite de tester la validité des modèles mis en place, de manières interne (par *cross validation*) et externe (à l'aide d'un jeu de validation). Enfin, le domaine d'applicabilité est défini par les molécules du jeu d'entraînement dans l'espace des descripteurs du modèle. A l'aide de cette méthodologie, le modèle mis en place pour l'intégralité des composés nitroaromatiques de la base de données utilisée, bien que corrélé (R^2=0,85) et robuste (R^2_{cv}=0,81), ne présente pas de pouvoir prédictif (R^2_{ext}=0,43). Par contre, un premier modèle prédictif (R^2_{ext}=0,84) a été obtenu pour les nitroaromatiques ne présentant pas de substituant en position ortho du groupement nitro. En effet, un substituant dans cette position peut engendrer des mécanismes de décomposition spécifiques.

Pour obtenir un modèle plus global, d'autres méthodes d'analyse de données ont été employées. Les analyses par projection (PCA et PLS) présentent un grand intérêt dans le choix des descripteurs. Si cette méthode est moins automatisée que la BMLR, elle présente l'avantage de permettre un choix "manuel" des descripteurs, d'un grand intérêt pour des exemples aussi complexes que le traitement de l'intégralité des nitroaromatiques. Finalement, à l'aide d'une méthode de classification par arbre de décision, au prix de la quantitativité, un modèle global robuste (84% de classifications correctes en LOO), prédictif (82% de classifications correctes sur le jeu de validation) et interprétable d'un point de vue phénoménologique a été obtenu. Ces méthodes représentent des outils tout-à-fait intéressants pour les analyses QSPR aussi complexes que la stabilité thermique des composés nitroaromatiques.

Il s'agit, à notre connaissance, des deux premiers modèles QSPR pour la chaleur de décomposition dont les pouvoirs prédictifs (importants) ont été déterminés pour un jeu de validation.

Trois autres propriétés ont ensuite été envisagées. Pour la température de décomposition, la base de données employée (22 molécules) ne permet pas de réaliser une étude complète, incluant la validation externe et l'analyse du domaine d'applicabilité. Néanmoins, les modèles QSPR développés aux niveaux AM1 et DFT démontrent de fortes robustesses (respectivement R^2_{cv}=0,84 et 0,81). Une nouvelle fois, l'emploi d'un niveau de théorie suffisant (DFT) pour le calcul des descripteurs quantiques permet d'obtenir des

modèles moins paramétrés, avec donc moins de risques de sur-paramétrisation (d'où une meilleure prédictivité attendue).

Pour la sensibilité à la décharge électrique, si le modèle développé pour le jeu complet de 41 molécules n'est pas prédictif (R^2_{ext}<0,01), il présente des corrélations élevées (R^2=0,90). D'ailleurs, aucun modèle de la littérature ne présente de résultats meilleurs ou même similaires. L'intérêt de ce modèle réside dans l'inclusion de descripteurs interprétables chimiquement. En particulier, alors que les modèles QSPR pour la chaleur de décomposition incluaient des descripteurs caractérisant à la fois le nombre de groupements nitro et la réactivité de ce groupement, le présent modèle ne prend en compte que cette réactivité. Ceci est tout-à-fait cohérent puisque la sensibilité traduit l'énergie nécessaire à la perte d'un premier groupement nitro et non, comme pour la chaleur de décomposition, l'énergie libérée au cours de la décomposition. Afin de comprendre le manque de prédictivité du modèle, une analyse PCA a été réalisée. Elle a permis de mettre en évidence une diversité excessive des molécules au sein de la base de données et l'obtention de modèles prédictifs nécessite des données expérimentales supplémentaires pour pouvoir considérer les composés nitroaromatiques en sous-classes, tel que mis en évidence par l'analyse PCA.

Les modèles QSPR développés pour la sensibilité à l'impact des composés nitroaromatiques sont soumis aux mêmes difficultés. Aucun modèle robuste n'est même obtenu pour l'intégralité des composés nitroaromatiques (R^2=0,55 et R^2_{cv}=0,47). Les composés présentant des substituants de type CH (CH_3, CH_2CH_3…) suivent un mécanisme spécifique de décomposition. Aussi ont-ils été considérés séparément. Malheureusement, si des modèles corrélés sont obtenus avec une telle subdivision, aucune prédictivité importante n'est observée (R^2_{ext}=0,57). Une analyse PCA démontre d'ailleurs qu'il ne s'agit pas du facteur le plus discriminant au sein de la base de données dans l'espace chimique des descripteurs. Là-encore, des sous-groupes sont identifiés et les données inclues dans chaque sous-groupe sont là encore insuffisantes pour valider la prédictivité de modèles pour chacun d'entre eux.

Des outils méthodologiques efficaces ont donc été mis en évidence. Ils ont permis l'obtention de deux modèles prédictifs pour la chaleur de décomposition : l'un quantitatif avec un domaine d'applicabilité plutôt réduit (non applicable aux composés nitrés ortho susbstitués), l'autre qualitatif mais plus largement applicable. Pour la température de décomposition ou les sensibilités à la décharge électrique et à l'impact, les travaux méritent d'être approfondis. Il semble d'ores et déjà que des données expérimentales supplémentaires soient nécessaires pour l'obtention de modèles prédictifs.

CHAPITRE 4

MECANISMES DE DECOMPOSITION DES COMPOSES NITROAROMATIQUES

Sommaire

1- Généralités 154
2- Méthodologie 156
3- Décomposition des nitrobenzènes substitués 156
3.1- Décomposition du nitrobenzène 156
3.2- Influence des substituants en position para 158
3.3- Influence des substituants en position méta 162
3.4- Influence des substituants en position ortho 164
4- Décomposition des o-nitrotoluènes substitués 166
4.1- Décomposition de l'o-nitrotoluène 166
4.2- Influence des substituants sur le carbone 3 169
4.3- Influence des substituants sur le carbone 4 171
4.4- Influence des substituants sur le carbone 5 173
4.5- Influence des substituants sur le carbone 6 176
5- Conclusion 178

Dans le chapitre précédent, des modèles QSPR ont été développés. Ces méthodes bien que prédictives ne caractérisent pas de manière explicite les mécanismes de décomposition. Or, les phénomènes explosifs étant le résultat de réactions en chaîne, la compréhension des mécanismes chimiques mis en jeu dans ces décompositions sont particulièrement importants. L'importance de la connaissance de ces mécanismes a d'ailleurs été démontrée précédemment puisqu'ils interviennent non seulement dans la définition des descripteurs (ex : énergie de dissociation de la liaison C-NO_2) mais aussi dans celle des bases de données (ex : partage selon la présence ou non de substituants en position ortho du groupement NO_2).

Ce chapitre traite donc de l'analyse des mécanismes réactionnels mis en jeu dans la décomposition de ces nitroaromatiques, parmi lesquels on retrouve notamment le TNT ou l'acide picrique. Les méthodes de chimie quantique, et en particulier la DFT, sont tout à fait adéquates pour de telles études. Après avoir énoncé un certain nombre de généralités sur le processus de décomposition de ces substances, l'influence de la nature des substituants sera étudiée à partir de deux systèmes de référence (le nitrobenzène et le 2-nitrotoluène) pour une série de leurs dérivés.

1- Généralités

Comprendre et prédire les propriétés de décomposition des matériaux énergétiques sont très importants, bien entendu pour leur caractérisation physico-chimique mais aussi pour faire face aux problèmes de sécurité relatifs à leur manipulation et stockage. Le manque de connaissance sur les propriétés chimiques et la réactivité des matériaux énergétiques ont de tous temps été la cause d'incidents [1,249]. Il n'est donc pas surprenant que différentes approches aient été développées pour une meilleure compréhension et prédiction des risques chimiques, au moins pour des espèces chimiques simples [13,14,31,44]. Le principal objectif de telles études était de comprendre les mécanismes détaillés du processus de décomposition, non seulement pour expliquer les observations expérimentales, mais aussi afin d'ouvrir de nouvelles perspectives dans la prévention des risques chimiques [248,250-254]. En fait, la connaissance des mécanismes de décomposition, en particulier thermiques, d'une substance explosive ou potentiellement explosive, permettra d'aider à la prédiction de manière systématique de propriétés de nouvelles molécules explosives, d'évaluer la faisabilité de leur synthèse à grande échelle, de prédire leur stabilité à long terme en vue de leur stockage, ainsi que leur sensibilité à divers stimuli tels que la chaleur ou un impact mécanique [168,169,255]. Etant donné que les propriétés explosives sont intimement liées au mécanisme de décomposition, ces connaissances pourront mener à une amélioration non seulement des performances explosives mais aussi de la prévention des risques chimiques liés à ces produits. Quoi qu'il en soit, en l'absence d'information mécanistique, le comportement des explosifs ne peut être sujet qu'à des considérations purement empiriques [15,44,166,175,224]. Ceci est particulièrement vrai au regard des conditions expérimentales qui sont souvent très différentes d'un laboratoire à l'autre, ce qui complique la comparaison des données [256]. A ce titre, la chimie théorique ouvre de nouvelles perspectives [248,250-254,257,258], non seulement parce qu'elle permet d'accéder aux données thermochimiques [259] mais aussi parce qu'elle donne des informations importantes dans la recherche de modèles prédictifs pour les propriétés dangereuses des substances telles que la stabilité thermique [165,260,261].

Au sein des substances explosives, les composés nitroaromatiques sont parmi les plus couramment utilisés [36]. Pour la prédiction des propriétés explosives des composés nitrés, il est généralement considéré que la plus faible liaison des composés nitrés est la liaison C-NO_2 et que sa rupture est l'étape limitante, à partir de laquelle débute la réaction en chaîne responsable de l'explosion [164,165]. Ce mécanisme est d'ailleurs l'une des réactions prépondérantes parmi celles identifiées expérimentalement pour l'initiation de la décomposition du nitrobenzène, le composé nitroaromatique le plus simple [262-268].

$$C_6H_5NO_2 \rightarrow C_6H_5 + NO_2 \quad (4.1)$$

$$C_6H_5NO_2 \rightarrow C_6H_5O + NO \quad (4.2)$$

$$C_6H_5NO_2 \rightarrow C_6H_5NO + O \quad (4.3)$$

$$C_6H_5NO_2 \rightarrow C_6H_4 + HONO \quad (4.4)$$

Les deux premières réactions menant à la formation de radicaux NO_2 et NO sont les voies majeures de décomposition thermique et de photodissociation [264,265,268]. La suivante, elle, n'a été observée qu'en photodissociation et est négligeable en décomposition thermique [264]. Plusieurs études théoriques se sont intéressées à la décomposition thermique du nitrobenzène. Alors que la plupart des ces travaux se sont focalisés sur la dissociation de la liaison carbone-nitro [169,269], la cinétique et les mécanismes de décomposition mono-moléculaire du nitrobenzène, n'ont été envisagés que récemment [250]. En particulier, les quatre principaux chemins de décomposition ont été étudiés. Ces résultats théoriques suggèrent que les deux dernières réactions (4.3 et 4.4) sont négligeables entre 500 et 800 K, domaine dans lequel a lieu la décomposition thermique, alors que la seconde réaction (4.2) est en compétition avec la réaction de dissociation directe (4.1). Ces mécanismes peuvent être considérées comme les mécanismes principaux d'initiation de la décomposition des composés nitroaromatiques en l'absence de substituant en position ortho du groupement nitro [270]. Dans ce type de composés, un mécanisme de décomposition plus complexe peut intervenir du fait des interactions possibles entre le groupement NO_2 et ce substituant situant en ortho, par exemple méthyle pour des dérivés du toluène [58,248,251].

Or, de tels composés, dont le 2,4,6-trinitrotoluène (TNT) est un représentant, sont très largement considérés pour leurs propriétés explosives d'où un fort intérêt pour leur décomposition [264,265,270,271]. Plusieurs travaux expérimentaux [262,263,266,272,273] et théoriques [248,251,257] ont d'ailleurs été menés afin de rationaliser les chemins de décomposition des dérivés de l'*o*-nitrotoluène. Aux deux chemins principaux du nitrobenzène s'ajoute celui mettant en jeu l'interaction entre le groupement nitro et le méthyle situé en ortho comme illustré en Figure 55.

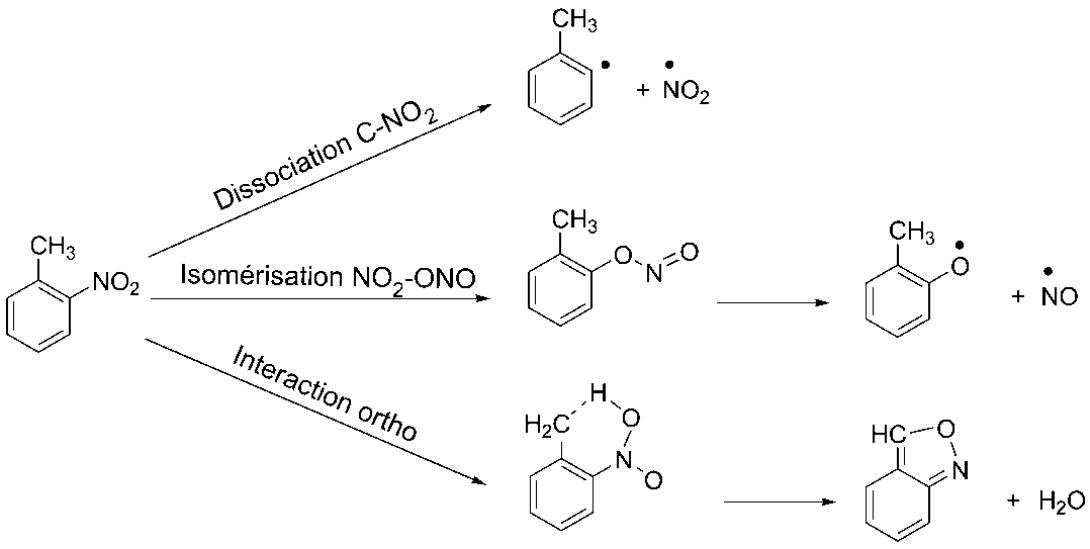


Figure 55 - Mécanismes de décomposition principaux du 2-nitrotoluène

Les différentes études expérimentales et théoriques sur de tels composés (TNT [248], 2-nitrotoluène [251] et 2,4-dinitrotoluène [257]) conduisent toutes à considérer ce dernier mécanisme comme le chemin de décomposition majoritaire.

L'importance relative de ces voies de décomposition est susceptible d'être affectée par la présence de substituants en différentes position du cycle aromatique. Une telle influence a été observée expérimentalement pour des nitrobenzènes substitués en para [270]. L'effet de déstabilisation des groupes hydroxyles sur des molécules nitroaromatiques a également été examiné [274]. Aussi, l'étude détaillée de l'influence des substituants a été réalisée sur les chemins de décomposition d'une série de molécules nitroaromatiques au niveau DFT. La dissociation homolytique de la liaison C-NO_2 (réaction 1), dans un premier temps, été envisagée en compétition avec l'isomérisation NO_2-ONO (réaction 2) pour une série de nitrobenzènes substitués [172], puis, dans un second temps, avec l'interaction nitro-méthyle (réaction 3) pour une série d'*o*-nitrotoluènes substitués [173].

2- Méthodologie

Tous les calculs réalisés dans ce chapitre ont été menés à l'aide du logiciel Gaussian03 [151] dans le cadre de la Théorie de la Fonctionnelle de la Densité (DFT). La fonctionnelle hybride PBE0 [89] a été employée dans tous les cas. Les optimisations de géométrie ont été réalisées avec une base 6-31+G(d,p) et les fréquences vibrationnelles ont été obtenues au même niveau de théorie. La nature des différents points stationnaires a alors été confirmée en ne comptant aucune fréquence imaginaire pour les espèces stables (réactifs, produits et intermédiaires réactionnels) et une seule pour les états de transition. A partir des fréquences harmoniques, les corrections thermiques aux énergie, enthalpies et énergies libres de Gibbs ont été calculées selon les approches statistiques standard, telles qu'implémentées dans Gaussian03. Les structures électroniques des différentes molécules ont été analysées à l'aide de l'analyse de population naturelle (NPA) [133]. Enfin, tous les calculs réalisés sur les espèces à couches ouvertes ont été menés dans le formalisme *spin unrestricted* en s'assurant d'une contamination de spin négligeable, les valeurs moyennes de l'opérateur S^2 étant proches de la valeur attendue, 0,75 pour les espèces radicalaires de cette étude.

3- Décomposition des nitrobenzènes substitués

3.1- Décomposition du nitrobenzène

Dans un premier temps, les deux mécanismes principaux de décomposition du nitrobenzène ont été envisagés. Les variations en énergie et énergie libre de Gibbs le long de ces deux réactions ainsi qu'un aperçu des structures moléculaires sont représentés en Figure ***56*** alors que le Tableau 26 collecte également les valeurs énergétiques, non seulement en énergie (ΔE) mais aussi en enthalpie (ΔH) et energie libre de Gibbs (ΔG).

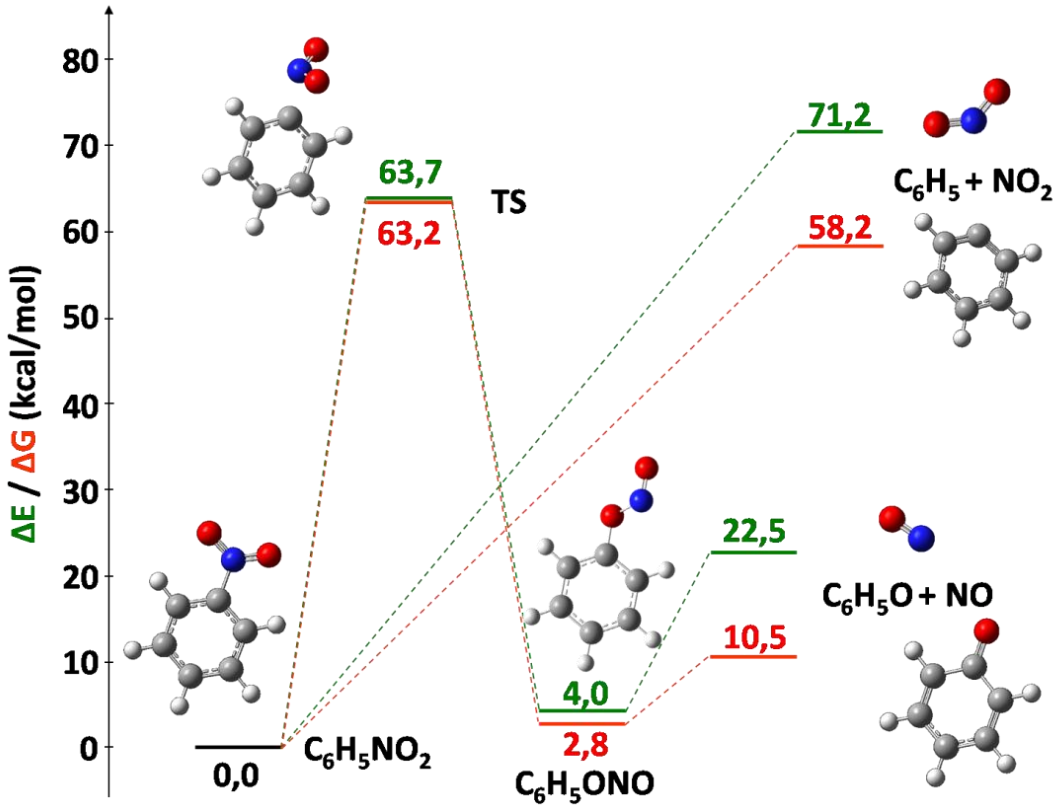


Figure 56 - Mécanismes de décomposition principaux du nitrobenzène (énergies ΔE en vert et énergies libres de Gibbs ΔG en rouge)

	ΔE		ΔH	ΔG
	PBE0	G2M[a]	PBE0	PBE0
$C_6H_5NO_2$	0,0	0,0	0,0	0,0
TS	63,7	61,1	63,7	63,2
C_6H_5ONO	4,0	1,1	4,0	2,8
C_6H_5O + NO	22,5	23,6	23,1	10,5
C_6H_5 + NO_2	71,2	74,1	71,8	58,2

[a] niveau G2M(RCC,MP2)//B3LYP/6-311G(d,p) [250]

Tableau 26 - Energies (ΔE), enthalpies (ΔH) et énergies libres de Gibbs relatives (ΔG) en kcal/mol pour les réactions de décomposition principaux du nitrobenzène au niveau PBE0/6-31+G(d,p)

Le premier mécanisme (réaction 1) est la rupture directe de la liaison carbone-azote. Il s'agit d'une dissociation homolytique menant à la formation de radicaux C_6H_5 et NO_2. La réaction 2, quant à elle, est caractérisée par un mécanisme en deux étapes. Le nitrobenzène ($C_6H_5NO_2$) s'isomérise tout d'abord en phénylnitrite (C_6H_5ONO). Cette isomérisation a pu être caractérisée théoriquement pour des composés nitro [275] et en particulier pour le nitrobenzène [276]. Puis la liaison oxygène-azote est rompue pour former les radicaux phénoxyle (C_6H_5O) et oxyde nitrique (NO).

Intéressons-nous, dans un premier temps, aux énergies relatives (ΔE) obtenues au niveau PBE0/6-31+G(d,p). Pour ce qui est du réarrangement NO_2-ONO, l'énergie d'activation, correspondant à l'état de transition (TS) de l'isomérisation est de 63,7 kcal/mol et l'isomère C_6H_5ONO est finalement 4,0 kcal/mol moins stable que le nitrobenzène initial. Ensuite, l'énergie de dissociation O-N, qui correspond à la dernière étape du mécanisme, est de 18,5 kcal/mol. Ces résultats sont globalement en bon accord avec les calculs antérieurs obtenus à l'aide d'une méthode post-HF plus sophistiquée [250]. L'énergie d'activation pour l'isomérisation vers C_6H_5ONO est plus faible que celle nécessaire à la rupture de la liaison C-NO_2 (63,7 contre 71,2 kcal/mol).

Ce résultat est en contradiction avec l'hypothèse expérimentale qui considère cette dissociation homolytique comme la voie de décomposition principale du nitrobenzène [164,165]. En se reportant aux enthalpies et énergies libres de Gibbs reportées dans le Tableau 26, il apparaît que les réactions 1 et 2 sont affectées de manière différente par les contributions thermiques et vibrationnelles. Tout d'abord, les enthalpies de dissociation calculées pour le mécanisme 1 et la dernière étape de la réaction 2, respectivement 71,8 et 19,1 kcal/mol, sont en accord avec les données expérimentales, 71,9 [277] et 20,8 kcal/mol [278]. Ceci confirme la qualité du protocole employé. En outre, si les effets entropiques sont inclus, les énergies libres de Gibbs obtenues sont cette fois en accord avec le comportement attendu et l'enthalpie libre de dissociation est plus faible que l'enthalpie libre d'activation pour la réaction d'isomérisation. Aussi, dans la suite de cette étude, seules les enthalpies libres de Gibbs seront considérées. Il s'agira alors de déterminer l'influence de la nature des substituants pour une série de dérivés du nitrobenzène (cf. Figure 57).

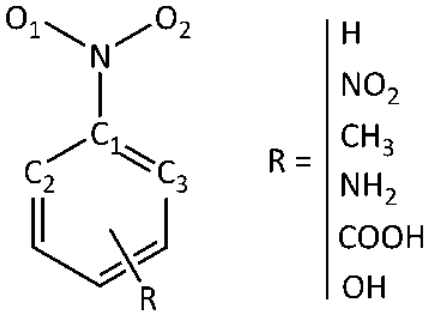


Figure 57 - Aperçu des dérivés du nitrobenzène.

3.2- Influence des substituants en position para

Afin d'étudier l'influence des groupements latéraux sur les deux chemins de décomposition principaux du nitrobenzène, 5 substituants différents ont été considérés : nitro (-NO_2), méthyle (-CH_3), amino (-NH_2), acide carboxylique (-COOH) et hydroxyle (-OH).

Le caractère électro-donneur ou accepteur des différents substituants a été étudié par le biais de la constante de Hammett (σ) [92], une approche communément employée pour l'étude et l'interprétation des mécanismes réactionnels et propriétés des composés organiques [94,279]. Il s'agit d'une des premières approches QSPR. D'autres descripteurs

moléculaires ont d'ailleurs depuis été dérivés des équations de Hammett [280]. Ces équations permettent de relier les constantes de vitesse d'une réaction donnée aux effets électroniques d'un substituant donné.

$$\log\left(\frac{K}{K_0}\right) = \rho\sigma \qquad (3.5)$$

où K et K_0 sont les constantes de vitesse pour les réactions avec les composés substitués et non-substitués, ρ est une constante qui dépend des conditions expérimentales et de la sensibilité de la réaction considérée aux effets électroniques et σ est la constante de Hammett pour le substituant étudié, caractérisant son influence sur la vitesse de réaction.

Finalement, une valeur positive de σ indique que le substituant est, par nature, accepteur d'électron. Au contraire, une valeur de σ négative caractérise son caractère donneur.

Il est à noter qu'aucune valeur de σ n'est disponible pour la position ortho pour laquelle des interactions directes (polaires ou stériques) interviennent du fait de la proximité des substituants, ce qui ne peut être représenté facilement à l'aide d'une équation aussi simple que celle de Hammett.

	σ_p	Mulliken	NPA
-H	0,00	-0,25	-0,26
-NO_2	0,78	-0,23	-0,23
-CH_3	-0,17	-0,28	-0,26
-NH_2	-0,66	-0,32	-0,30
-COOH	0,45	-0,24	-0,24
-OH	-0,37	-0,29	-0,28

Tableau 27 - Constantes de Hammett [94] pour les substituants en position para et charges de NO_2 (en $|e^-|$) calculées à l'aide des analyses de populations naturelle (NPA) et de Mulliken

Le groupement nitro est le plus accepteur (+0,78) alors que le substituant amino est le plus donneur (-0,66). Les valeurs de σ suggèrent une plus grande réactivité dans le premier cas que dans le second.

Les charges, obtenues à partir de la population NPA, sont proches de celles calculées à l'aide de la définition de Mulliken. Dans la suite de l'étude, seules les charges NPA seront discutées. Si leurs variations se situent dans un assez faible éventail de valeurs, avec des charges situées entre -0,23 pour NO_2 et -0,30 pour NH_2, les caractères donneurs ou accepteurs des différents substituants influent sur la charge du groupement nitro.

	-H	-NO_2	-CH_3	-NH_2	-COOH	-OH
Distances (Å)						
N-O_1	1,221	1,220	1,222	1,225	1,220	1,223
N-O_2	1,221	1,220	1,222	1,225	1,221	1,224
C_1-N	1,467	1,472	1,462	1,448	1,470	1,455
C_1-C_2	1,391	1,390	1,390	1,394	1,391	1,394
C_1-C_3	1,391	1,390	1,390	1,394	1,390	1,391
Angles (deg)						
C_2-C_1-N	118,790	118,622	119,086	119,516	118,714	119,312
C_3-C_1-N	118,790	118,622	119,086	119,516	118,683	119,158
C_1-N-O_1	117,623	117,343	117,672	117,872	117,474	117,704
C_1-N-O_2	117,623	117,343	117,672	117,872	117,494	117,729
Angles dièdres (deg)						
O_1-N-C_1-C_2	0,000	0,000	0,000	0,044	0,000	0,000

Tableau 28 - Paramètres structuraux principaux des nitrobenzènes substitués en position para (nomenclature en Figure 57)

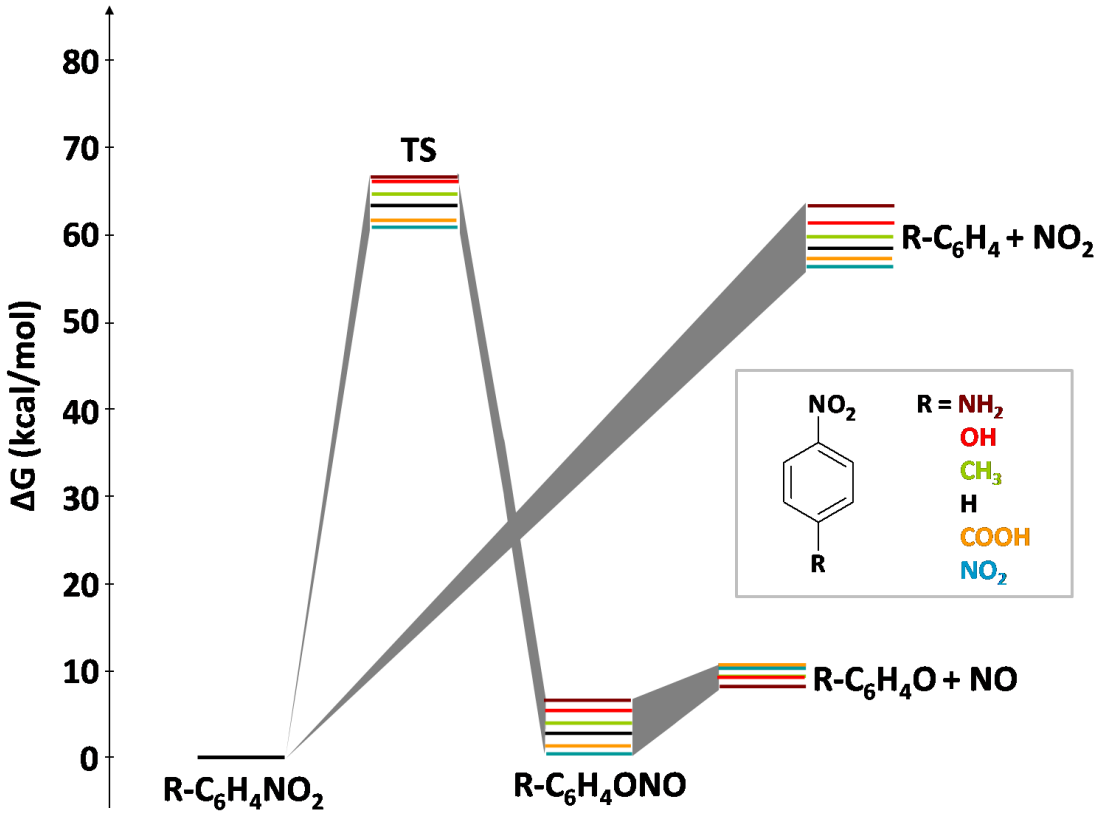


Figure 58 – Profils énergétiques pour les réactions de décomposition vers NO et NO_2 de nitrobenzènes substitués en position para

	-H	-NO_2	-CH_3	-NH_2	-COOH	-OH
R-$C_6H_4NO_2$	0,0	0,0	0,0	0,0	0,0	0,0
TS	63,2	60,7	64,3	66,9	61,4	65,9
R-C_6H_4ONO	2,8	0,4	3,9	6,6	1,4	5,1
R-C_6H_4O + NO	10,5	10,1	9,1	7,6	10,2	9,0
R-C_6H_4 + NO_2	58,2	56,2	59,4	63,0	56,9	61,0

Tableau 29 - Energies libres de Gibbs relatives (en kcal/mol) pour les réactions de décomposition vers NO et NO_2 de nitrobenzènes substitués en position para

La structure électronique est reliée aux réarrangements géométriques de sorte que les distances C-N les plus grandes sont observées pour les nitrobenzènes substitués avec les groupements les plus électro-attracteurs. L'amplitude de variation (entre 1,448 et 1,472 Å) est relativement importante (voir Tableau 28).

Les énergies libres de Gibbs pour les deux chemins de décomposition, représentées en Figure 58, sont collectées dans le Tableau 29. D'un point de vue général, la voie dissociative directe (réaction 1) est toujours plus basse en énergie que le TS pour l'isomérisation en C_6H_5ONO de la réaction 2 avec une différence de 4 kcal/mol en moyenne.

Les groupes électro-attracteurs influencent la barrière d'activation de la réaction 2, qui varie de 63 kcal/mol pour le nitrobenzène de référence à 61 kcal/mol pour les substituants NO_2 et COOH. La barrière est, par contre, plus grande pour les substituants électro-donneurs, tel que NH_2 avec 67 kcal/mol. La même tendance est à noter pour la stabilité de la structure R-C_6H_4ONO : plus le substituant est électro-donneur, moins elle sera stabilisée. Au contraire, la dissociation de la liaison O-N menant au radical R-C_6H_5O, la dernière étape de ce mécanisme, nécessite toujours autour de 10 kcal/mol.

L'influence des substituants est également importante pour la dissociation de la liaison C-NO_2 (réaction 1) : les énergies libres de Gibbs correspondantes sont comprises entre 56 et 63 kcal/mol (de NO_2 à NH_2). Finalement, la différence entre les énergies d'activation des réactions 1 et 2, et donc la compétition entre les deux chemins, n'est pas modifiée.

Comme cela a pu être observé pour la charge du groupement nitro, l'énergie est, elle aussi, directement reliée au pouvoir électro-attracteur du substituant. Il est même intéressant de noter la corrélation linéaire (R^2=0,94) observée entre les enthalpies libres de dissociation des liaisons C-NO_2 (ΔG_{diss}) et les paramètres de Hammett, en Figure 59. Une relation linéaire similaire peut être établie avec la stabilité de R-C_6H_4ONO (ΔG_{ONO}).

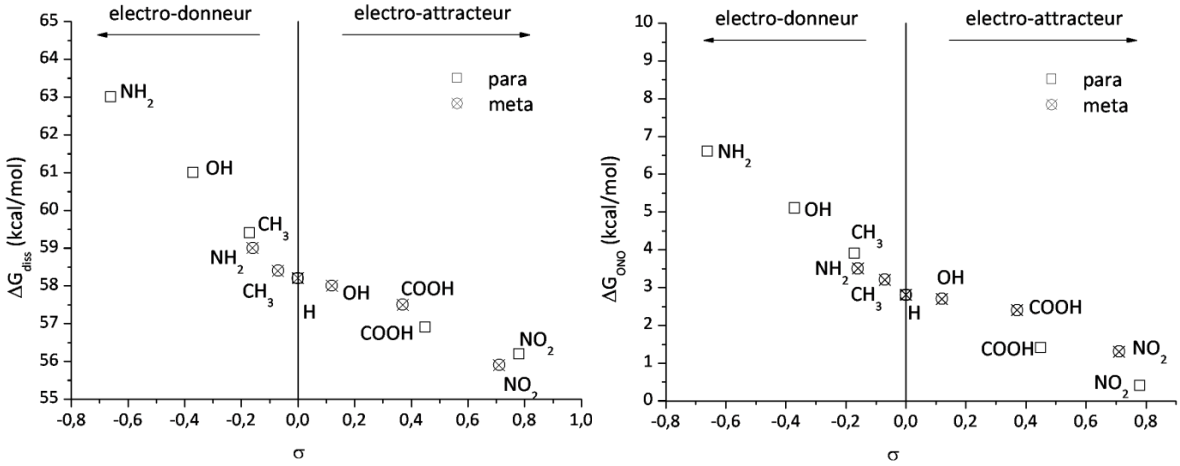


Figure 59 - Energies libres de Gibbs relatives de la dissociation C-NO_2 (ΔG_{diss}) et de la stabilité de la structure R-C_6H_4ONO (ΔG_{ONO}) en fonction des constantes de Hammett (σ)

Par contre, aucune corrélation ne peut être observée avec la barrière d'activation de cette même réaction 2. Par le passé, une telle défaillance du pouvoir prédictif des constantes de Hammett vis-à-vis des énergies d'activation avait été observée pour des équilibres conformationnels de dérivés thiobenzanilides [281].

Une comparaison directe de la différence entre les barrières des deux chemins réactionnels ne met pas en avant une variation d'un substituant à un autre. En effet, les substituants électro-attracteurs affaiblissent la liaison carbone-azote et abaissent la barrière vers l'intermédiaire R-C_6H_4ONO alors que les substituants électro-donneurs déstabilisent les produits issus de la dissociation C-N autant qu'ils augmentent la barrière d'isomérisation dans la réaction 2. En d'autres termes, le caractère électro-attracteur ou donneur des substituants en position para du groupement nitro n'influencent pas de manière significative la compétition entre les réactions 1 et 2. Ces résultats sont plutôt en contradiction avec les conclusions émises par Brill [270] qui avait observé, par analyse spectrométrique, une influence en position para. Ceci étant dit, les substituants étudiés étaient alors différents.

3.3- Influence des substituants en position méta

La substitution en position méta a ensuite été étudiée. Dans cette position, l'effet du substituant n'est que très faible, à la fois sur les paramètres géométriques (dans le Tableau ***31***) et électroniques (les charges sur le groupement nitro apparaissent dans le Tableau 30). En effet, la distance C-NO_2 n'est pas affectée par la nature du substituant (variations de moins de 0,004 Å) quand la charge sur le groupement nitro demeure proche de -0,24.

	σ_m	Mulliken	NPA
-H	0,00	-0,25	-0,26
-NO_2	0,71	-0,26	-0,23
-CH_3	-0,07	-0,26	-0,26
-NH_2	-0,16	-0,28	-0,26
-COOH	0,37	-0,25	-0,24
-OH	0,12	-0,26	-0,25

Tableau 30 - Charges de NO_2 (en $|e^-|$), calculées à l'aide des analyses de populations naturelles (NPA) et de Mulliken, et constantes de Hammett [94] pour les substituants en position méta

	-H	-NO_2	-CH_3	-NH_2	-COOH	-OH
Distances (Å)						
N-O_1	1,221	1,220	1,222	1,221	1,221	1,221
N-O_2	1,221	1,219	1,222	1,222	1,220	1,221
C_1-N	1,467	1,470	1,467	1,468	1,468	1,469
C_1-C_2	1,391	1,391	1,389	1,389	1,391	1,391
C_1-C_3	1,391	1,387	1,391	1,388	1,388	1,387
Angles (deg)						
C_2-C_1-N	118,790	118,914	118,809	118,749	118,815	118,727
C_3-C_1-N	118,790	118,494	118,602	117,985	118,700	118,103
C_1-N-O_1	117,623	117,233	117,629	117,646	117,414	117,481
C_1-N-O_2	117,623	117,382	117,700	117,748	117,504	117,674
Angles dièdres (deg)						
O_1-N-C_1-C_2	0,000	0,000	-0,040	-0,302	0,000	-0,004

Tableau 31 - Paramètres structuraux principaux des nitrobenzènes substitués en position méta. La nomenclature des atomes est indiquée en Figure 57

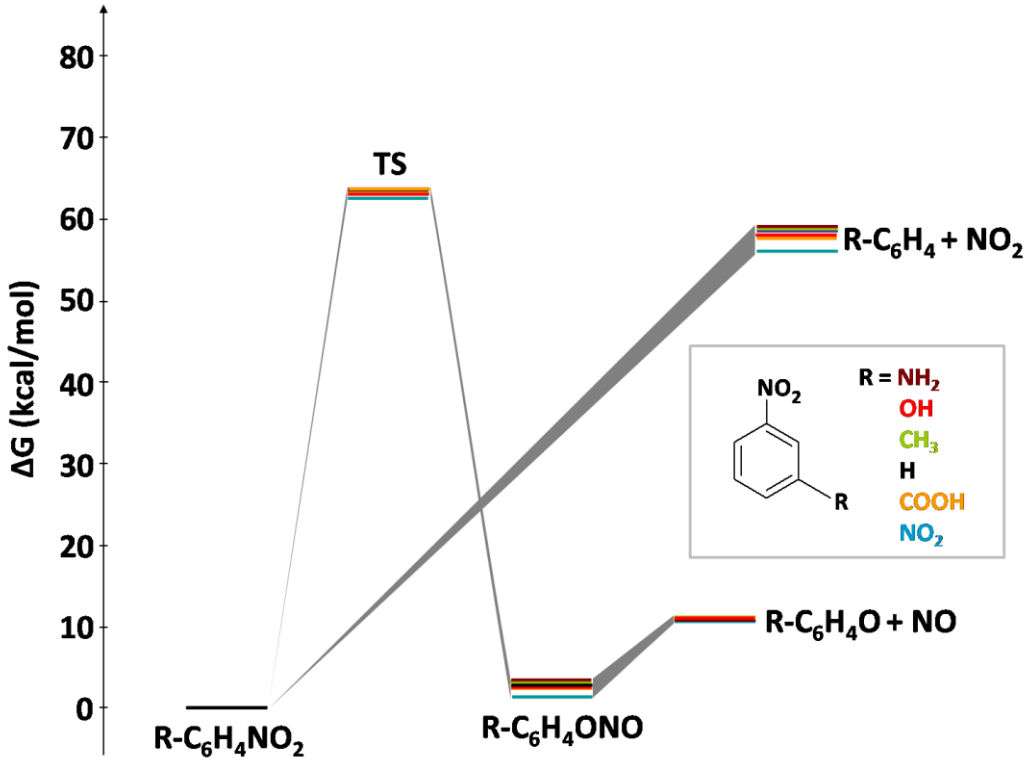


Figure 60 - Profils énergétiques pour les réactions de décomposition vers NO et NO_2 de nitrobenzènes substitués en position méta

	-H	-NO_2	-CH_3	-NH_2	-COOH	-OH
R-$C_6H_4NO_2$	0,0	0,0	0,0	0,0	0,0	0,0
TS	63,2	62,9	63,5	63,4	63,4	63,0
R-C_6H_4ONO	2,8	1,3	3,2	3,5	2,4	2,7
R-C_6H_4O + NO	10,5	10,5	11,1	10,7	10,9	10,9
R-C_6H_4 + NO_2	58,2	55,9	58,4	59,0	57,5	58,0

Tableau 32 - Energies libres de Gibbs relatives (en kcal/mol) pour les réactions de décomposition vers NO et NO_2 de nitrobenzènes substitués en position méta

Seules de faibles variations énergétiques sont observées sur les chemins de décomposition, comme le montrent la Figure 60 et le Tableau 32. Pour la réaction 2, les énergies des TS d'isomérisation sont toutes d'environ 63 kcal/mol et les produits finaux d'environ 10 kcal/mol, même si une très légère influence est observée pour la stabilité de l'intermédiaire R-C_6H_4ONO (1 à 3 kcal/mol). Une telle variation (3 kcal/mol) est également observée pour la dissociation homolytique de la liaison C-NO_2, les énergies libres de dissociation variant de 56 (NO_2) à 59 kcal/mol (NH_2), suivant les propriétés électroniques du substituant.

En fait, une dépendance linéaire (R^2=0,96) relie ces énergies et les paramètres de Hammett (Tableau 30). Ceci est particulièrement bien illustré par la Figure 59. Comme pour la substitution en position para, une telle dépendance peut être observée pour la stabilité de la structure R-C_6H_4ONO et non pour les énergies d'activation de l'isomérisation NO_2-ONO.

Il est à noter que la réactivité des nitrobenzènes substitués n'est pas influencée par la position des deux groupements les plus électro-attracteurs (NO_2, COOH) alors que des

variations significatives apparaissent pour les plus électro-donneurs (CH_3, NH_2). Entre ces deux extrêmes, la constante de Hammett du groupement hydroxyle change de signe de la position para à la position méta du fait de l'effet de résonnance de OH dans les molécules aromatiques, qui concerne uniquement les positions ortho et para.

Finalement, la différence d'énergie entre les énergies d'activation des réactions 1 et 2, indicateur de la compétition entre ces deux chemins, n'est que légèrement influencée par la nature du substituant en position méta, la rupture de la liaison C-NO_2 demeurant le mécanisme le plus favorable et ceci que le substituant présente un caractère plutôt électro-attracteur ou plutôt électro-donneur.

3.4- Influence des substituants en position ortho

Le caractère électro-donneur ou attracteur des substituants en position ortho est plus difficile à interpréter du fait des interactions directes pouvant intervenir entre le groupement nitro. Ce dernier est alors situé à proximité et peut donc directement interagir avec lui. De ce fait, les équations de Hammett et les constantes associées n'ont pu être développées pour une telle configuration.

Les charges sur le groupement nitro (Tableau 33) mais aussi les principaux paramètres géométriques (en Tableau 34 -) sont influencés par la nature du groupement en ortho mais les effets de ce dernier ne peuvent être classés aussi directement que pour les positions méta et para. Par exemple, la distance C-N est moins importante avec un substituant méthyle que pour un hydroxyle, ce qui est tout à fait contraire à ce qui a été observé en position para. Cela dit, les groupements les plus électro-attracteurs (NO_2 et COOH) ont toujours tendance à montrer les plus grandes longueurs de liaison C-NO_2 et les plus faibles charges pour le groupement nitro.

Cette observation est en adéquation avec la description énergétique des chemins de décomposition. Des variations significatives en termes de ΔG apparaissent (en Figure 61 et Tableau 35) selon la nature du substituant positionné en ortho du groupement nitro mais il est là-encore difficile de relier ceci aux propriétés électroniques de ce substituant. En particulier, pour les groupements nitro, méthyle et acide carboxylique, les barrières d'activation de la réaction 2 sont inférieures aux systèmes méta et para correspondants, alors que la tendance inverse est observée pour l'hydroxyle. Le substituant amino, quant à lui, présente un comportement intermédiaire : la barrière pour la position ortho est plus haute qu'en méta mais plus basse qu'en para. Plus intéressant, la substitution ortho mène à des dérivés R-C_6H_4ONO qui sont plus stables que les structures de type nitrobenzène dans le cas des substituants NO_2 et COOH et montrent les barrières d'isomérisation les plus basses (respectivement -55 et -56 kcal/mol).

Une nouvelle fois, de plus faibles énergies sont établies pour la dissociation directe de la liaison C-N. Ces dernières varient entre 47,8 (NO_2) et 63,0 kcal/mol (NH_2). La seule exception apparaît pour NH_2 pour lequel on note une petite différence (1 kcal/mol) entre les réactions 1 et 2.

	Mulliken	NPA
-H	-0,25	-0,26
$-NO_2$	-0,17	-0,20
$-CH_3$	-0,30	-0,26
$-NH_2$	-0,38	-0,33
-COOH	-0,17	-0,22
-OH	-0,31	-0,25

Tableau 33 - Charges de NO_2 (en $|e^-|$), calculées à l'aide des analyses de populations naturelles (NPA) et de Mulliken

	-H	$-NO_2$	$-CH_3$	$-NH_2$	-COOH	-OH
Distances (Å)						
$N-O_1$	1,221	1,218	1,223	1,223	1,220	1,225
$N-O_2$	1,221	1,217	1,223	1,237	1,220	1,219
C_1-N	1,467	1,467	1,446	1,441	1,466	1,460
C_1-C_2	1,391	1,387	1,395	1,402	1,388	1,394
C_1-C_3	1,391	1,394	1,406	1,420	1,398	1,406
Angles (deg)						
C_2-C_1-N	118,790	117,395	115,910	116,926	117,041	117,142
C_3-C_1-N	118,790	122,140	121,343	121,762	117,847	121,892
C_1-N-O_1	117,623	116,737	117,648	118,658	117,295	116,969
C_1-N-O_2	117,623	116,883	118,099	118,577	117,108	118,203
Angles dièdres (deg)						
$O_1-N-C_1-C_2$	0,000	-41,949	-18,099	0,006	35,728	-23,482

Tableau 34 - Paramètres structuraux principaux des nitrobenzènes substitués en position ortho (nomenclature en Figure 57)

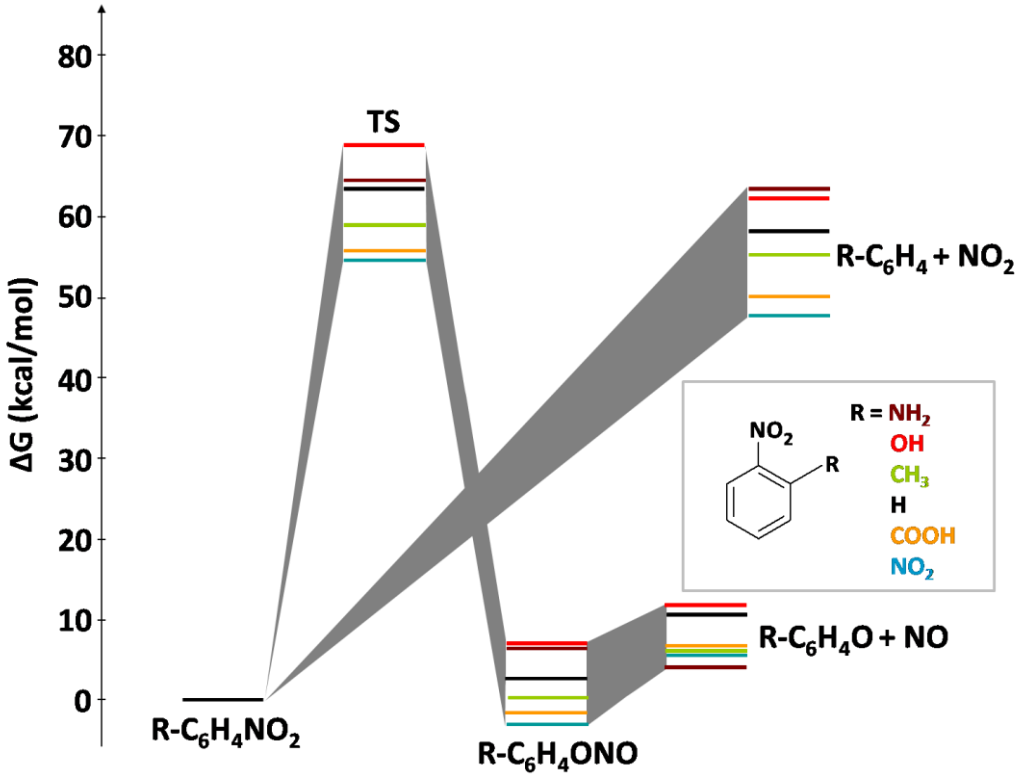


Figure 61 - Profils énergétiques pour les réactions de décomposition vers NO et NO_2 de nitrobenzènes substitués en position para

	-H	-NO_2	-CH_3	-NH_2	-COOH	-OH
R-$C_6H_4NO_2$	0,0	0,0	0,0	0,0	0,0	0,0
TS	63,2	54,6	58,9	64,4	55,9	68,8
R-C_6H_4ONO	2,8	-3,4	0,3	6,3	-1,7	7,0
R-C_6H_4O + NO	10,5	5,7	6,0	4,0	6,8	11,8
R-C_6H_4 + NO_2	58,2	47,8	55,1	63,0	50,1	62,1

Tableau 35 - Energies libres de Gibbs relatives (en kcal/mol) pour les réactions de décomposition vers NO et NO_2 de nitrobenzènes substitués en position ortho

Néanmoins, il doit être indiqué ici que d'autres mécanismes de décomposition plus complexes et plus favorables ont pu être proposés pour certains dérivés du nitrobenzène substitués en ortho [58]. De tels mécanismes interviennent notamment dans la décomposition du 2,4,6-trinitrotoluène (TNT) [248] ou de l'*o*-nitrotoluène [251] et font l'objet de la suite de cette étude.

4- Décomposition des *o*-nitrotoluènes substitués

4.1- Décomposition de l'*o*-nitrotoluène

Il a donc pu être observé, jusque-là, que la dissociation directe de la liaison (réaction 3.1) est toujours plus favorable que le mécanisme par isomérisation NO_2-ONO (réaction 3.2). Afin d'étudier, à présent, la compétition entre la réaction 3.1 et le mécanisme spécifique par interaction entre le groupement nitro et un substituant méthyle en position ortho, l'*o*-nitrotoluène est considéré comme molécule de référence. Les profils énergétiques sont présentés en Figure 62 et un aperçu des structures des états de transitions (TS) et minima locaux (LM) le long de ces deux mécanismes en Figure 63. Les données en énergies (ΔE), enthalpies (ΔH) et énergies libres de Gibbs (ΔG) sont quant à elles collectées dans le Tableau 36.

	ΔE		ΔH	ΔG
	PBE0	G2M[a]	PBE0	PBE0
$CH_3C_6H_4NO_2$	0,0	0,0	0,0	0,0
$CH_3C_6H_4$ + NO_2	68,7	76,3	69,9	55,1
TS1	40,5	46,3	41,1	42,0
LM1	39,0	37,9	39,6	39,5
TS2	40,2	38,8	40,8	40,8
LM2	33,6	31,9	34,2	33,9
TS3	53,9	52,9	53,5	53,6
LM3	30,4	28,5	31,0	30,7
TS4	47,9	46,1	48,5	49,3
LM4	11,4	11,1	12,0	12,8
TS5	50,6	48,8	51,2	52,0
C_6H_4CHON + H_2O	-17,1	-6,0	-15,9	-26,2

[a] niveau G2M(RCC,MP2)//B3LYP/6-311G(d,p) [250]

Tableau 36 - Energies (ΔE), enthalpies (ΔH) et énergies libres de Gibbs relatives (ΔG) en kcal/mol pour les réactions de décomposition principales de l'o-nitrotoluène au niveau PBE0/6-31+G(d,p)

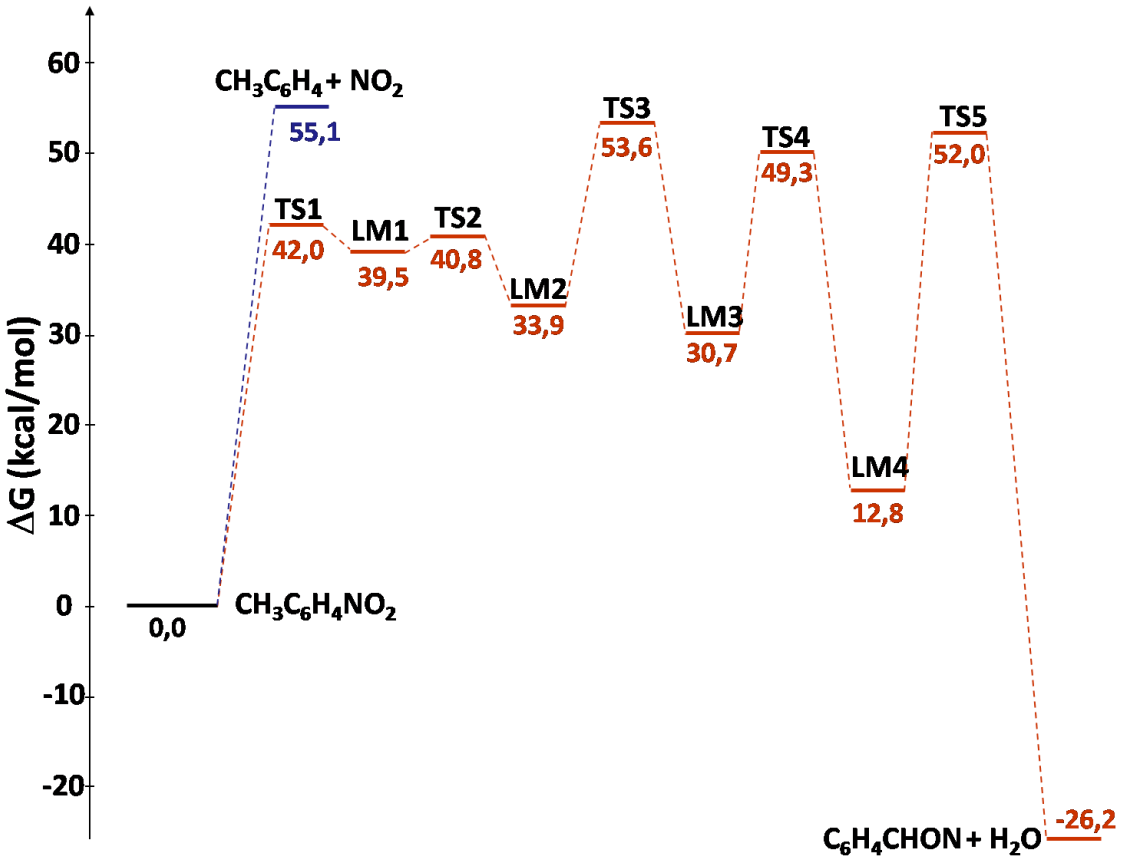


Figure 62 - Mécanismes de décomposition principaux de l'o-nitrotoluène

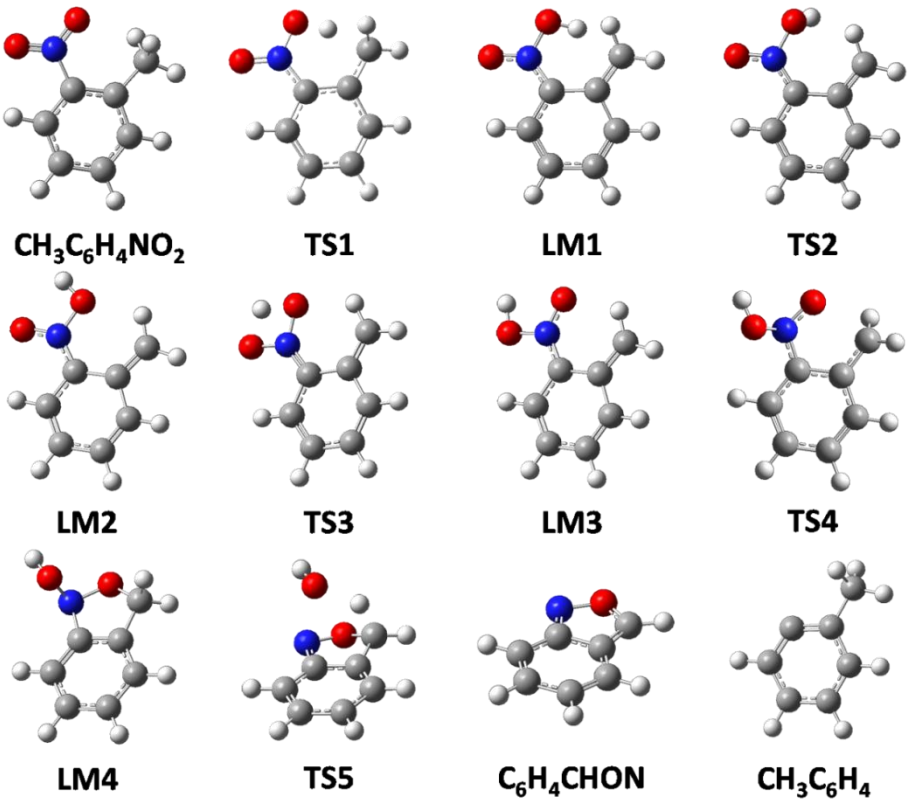


Figure 63 - Aperçu des différents points stationnaires intervenant dans les chemins de décomposition principaux de l'o-nitrotoluène

La rupture de la liaison C-NO_2 mène à la formation de radicaux tolyle ($CH_3C_6H_4$) et nitro (NO_2). Le second mécanisme (réaction 3) est par contre plus complexe. Dans un premier temps, l'étape d'initiation du mécanisme correspond à l'attaque d'une liaison C-H du méthyle par NO_2. Un atome d'hydrogène du groupement CH_3 est alors transféré vers le groupement nitro situé en ortho via un état de transition TS1. La molécule ainsi obtenue (LM1) subit ensuite une isomérisation afin de passer d'une conformation « HONO » à une conformation « ONOH » (LM3), en deux étapes : une rotation (menant à LM2) suivie d'un transfert d'hydrogène. L'atome d'oxygène le plus accessible du groupement « ONOH » ainsi formé interagit alors avec le groupement CH_2 (LM4). Finalement, deux produits sont obtenus par les dissociations concertées des liaisons C-H et N-OH (via l'état de transition TS5) : l'anthranile (C_6H_4CHON) et l'eau (H_2O).

L'étape limitante de cette réaction est le transfert initial d'hydrogène depuis le groupement méthyle vers le nitro (42 kcal/mol). L'énergie d'activation calculée est bien plus basse que l'énergie de dissociation de la liaison C-NO_2 (55,1 kcal/mol). Si ce résultat est en désaccord avec la considération usuelle globale pour les composés nitrés, qui considère la rupture de la liaison C-NO_2 comme mécanisme responsable du processus d'explosion [164,165], il est par contre en accord avec les études précédentes considèrant ce mécanisme comme majoritaire à la fois pour le 2-nitrotoluène [251] et pour le TNT [248]. Les étapes suivantes du mécanisme présentent des barrières moins importantes.

Du point de vue de la méthode employée lors de ces calculs, il est à noter que les énergies (ΔE) calculées au niveau PBE0/6-31+G(d,p) sont globalement proches de celles obtenues par Chen avec une méthode post-HF plus sophistiquée (G2M) [251]. Quelques déviations importantes ponctuelles doivent toutefois être notées. La plus importante concerne, en particulier, l'énergie de dissociation de la liaison C-NO_2 (68,7 contre 76,3 kcal/mol). Cela dit, l'enthalpie de dissociation de cette liaison, obtenues par les calculs PBE0 (69,9 kcal/mol), est en excellent accord avec les données expérimentales (70,2 ± 2,5 kcal/mol) [262], ce qui conforte la qualité du protocole de calcul qui avait d'ailleurs été utilisé avec succès pour le nitrobenzène.

Finalement, du fait de l'importance des effets entropiques (13,6 kcal/mol pour la dissociation C-NO_2), seules les énergies libres de Gibbs seront discutées dans la suite de cette étude. L'influence de la nature des substituants sur les différents sites disponibles le long du cycle aromatique (carbones 3, 4, 5 et 6 en Figure 64) sera étudiée pour une série de dérivés du 2-nitrotoluène afin de déterminer si la compétition entre les deux chemins de décomposition reste inchangée.

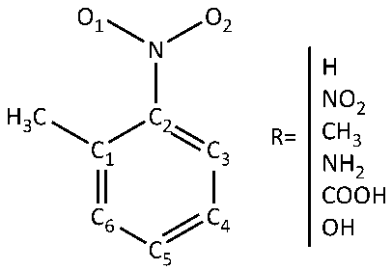


Figure 64 - Aperçu des dérivés de l'o-nitrotoluène

4.2- Influence des substituants sur le carbone 3

Les effets du substituant ont été analysés pour les cinq substituants précédemment considérés, à savoir nitro, méthyle, amino, acide carboxylique et hydroxyle. Dans cette position, les caractères électro-donneurs ou accepteurs sont en général assez difficiles à interpréter, car les effets électroniques et stériques peuvent jouer des rôles comparables, du fait de la proximité du groupement NO_2. Il s'agit d'une position de type ortho pour laquelle les équations de Hammett, présentées précédemment, ne peuvent être employées.

On peut remarquer que les charges du groupement nitro (Tableau 37) et les structures géométriques (en Tableau 38) sont influencées par la nature du substituant.

position	3	4	5	6
-H	-0,26	-0,26	-0,26	-0,26
$-NO_2$	-0,17	-0,24	-0,24	-0,23
$-CH_3$	-0,24	-0,27	-0,27	-0,25
$-NH_2$	-0,31	-0,27	-0,31	-0,25
-COOH	-0,20	-0,25	-0,25	-0,24
-OH	-0,30	-0,27	-0,29	-0,25

Tableau 37 – Charges de NO_2 (en $|e^-|$) calculées à partir de l'analyse de population naturelle (NPA)

	-H	$-NO_2$	$-CH_3$	$-NH_2$	-COOH	-OH
Distances (Å)						
$N-O_1$	1,223	1,216	1,222	1,222	1,218	1,217
$N-O_2$	1,223	1,216	1,222	1,236	1,218	1,247
C_2-N	1,466	1,472	1,466	1,444	1,472	1,439
C_1-C_2	1,406	1,397	1,401	1,418	1,397	1,423
C_2-C_3	1,394	1,393	1,401	1,423	1,398	1,425
Angles (deg)						
C_1-C_2-N	121,4	116,7	117,9	119,2	116,5	120,5
C_3-C_2-N	115,9	121,8	117,9	118,8	120,5	118,4
C_2-N-O_1	118,1	116,7	117,6	118,8	117,1	120,1
C_2-N-O_2	117,6	116,6	117,6	118,7	116,9	118,5
Angle dièdre (deg)						
$O_1-N-C_2-C_1$	-19,4	-71,2	-52,1	-25,1	-68,6	0,1

Tableau 38 – Paramètres structuraux principaux des o-nitrotoluènes substitués en position 3. La nomenclature des atomes est indiquée en Figure 64

Les variations les plus importantes concernent l'angle dièdre autour de la liaison C-N du fait à la fois de l'encombrement stérique et des possibles interactions stabilisantes entre le nitro et les substituants en ortho, comme c'est le cas entre l'hydrogène d'un groupement OH et l'oxygène du NO_2. Ces variations ne sont donc pas seulement liées aux caractéristiques intrinsèques des substituants mais aux interactions directes pouvant intervenir.

D'autres paramètres géométriques suivent, par contre, des tendances tout aussi compréhensibles au regard des propriétés électroniques des substituants. Par exemple, la liaison C-N est plus courte pour des groupements électro-donneurs (1,44 Å pour NH_2) que

pour les électro-attracteurs (1,47 Å pour NO_2 et COOH). D'ailleurs, la charge du groupement NO_2 suit la tendance associée puisqu'elle augmente de -0,31 pour NH_2 à -0,17 pour NO_2 et COOH. Il s'ensuit également que les dissociations des liaisons C-NO_2 sont énergétiquement plus aisées pour NH_2 que pour NO_2 ou COOH, tel qu'illustré par la Figure 65 et par le Tableau 39, puisque l'énergie de cette dissociation décroît clairement depuis les électro-donneurs jusqu'aux électro-attracteurs.

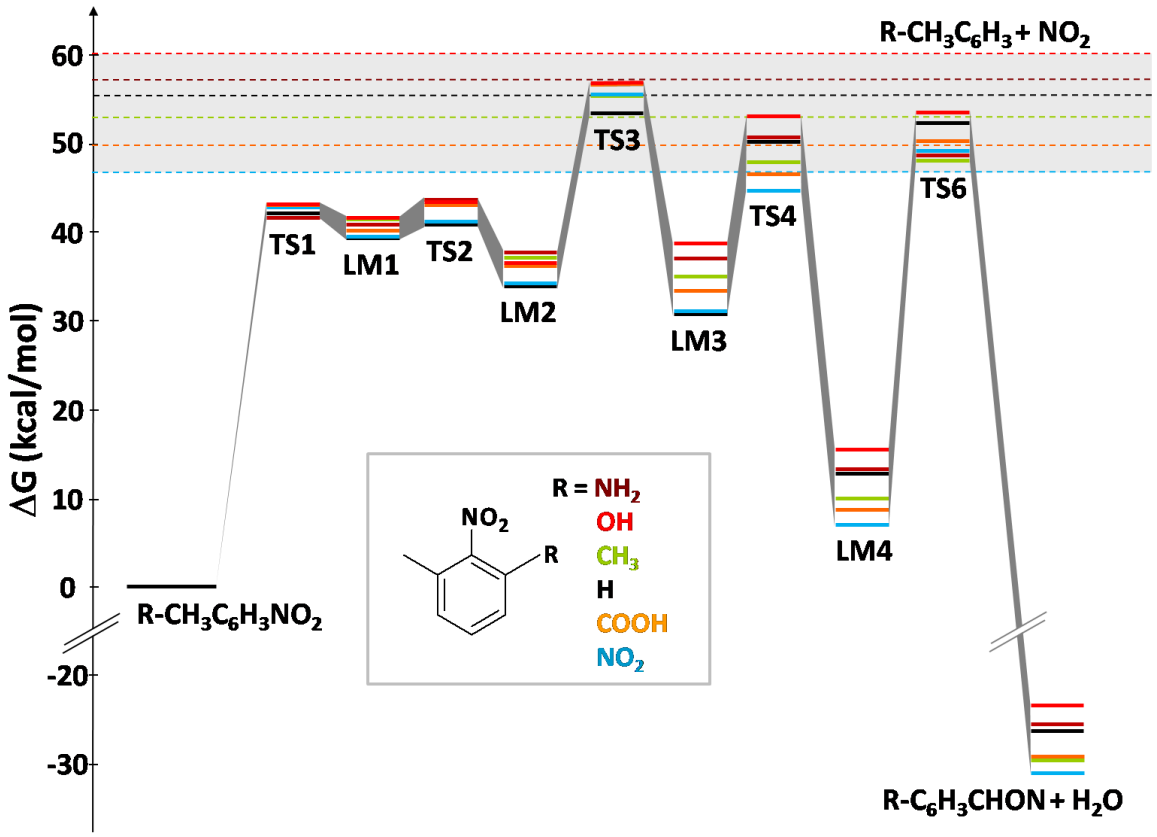


Figure 65 - Profils énergétiques pour les réactions de décomposition principales des o-nitrotoluènes substitués en position 3

	-H	-NO_2	-CH_3	-NH_2	-COOH	-OH
R-$CH_3C_6H_3NO_2$	0,0	0,0	0,0	0,0	0,0	0,0
R-$CH_3C_6H_3$ + NO_2	55,1	46,6	52,4	57,1	49,6	59,6
TS1	42,0	42,6	43,1	41,6	42,5	43,2
LM1	39,5	39,6	41,3	40,7	40,0	41,5
TS2	40,8	41,3	43,5	43,5	42,7	42,9
LM2	33,9	34,3	37,1	37,5	36,0	36,2
TS3	53,6	55,3	55,2	56,3	56,4	56,7
LM3	30,7	31,1	34,5	37,6	32,9	38,3
TS4	49,3	44,3	47,7	50,4	46,1	52,6
LM4	12,8	7,2	9,8	13,1	8,5	15,7
TS5	52,0	49,1	48,0	48,2	50,0	53,4
R-C_6H_3CHON + H_2O	-26,2	-31,1	-29,6	-25,6	-29,2	-23,6

Tableau 39 - Energies libres de Gibbs relatives (en kcal/mol) pour les réactions de décomposition principales des o-nitrotoluènes substitués en position 3

Pour le second chemin de décomposition, par contre, si quelques variations apparaissent lorsque l'on change de substituant, elles concernent uniquement les dernières étapes du mécanisme (en particulier les produits finaux) qui ne sont, en aucun cas, des étapes limitantes pour le mécanisme global. Les énergies libres relatives sont comprises entre -31,1 et -23,6 kcal/mol pour les produits finaux. Une variation très faible est également notée pour TS3 pour lequel la barrière est d'environ 2 kcal/mol pour tous les composés de la série. Cela dit, les ΔG pour TS1, qui sont les énergies d'activation de ce mécanisme, restent comprises entre 41,6 et 43,2 kcal/mol.

Finalement, même si la réaction 1 est plus fortement influencée par la nature du substituant que ce deuxième chemin de décomposition, ce dernier reste plus favorable que la dissociation C-NO_2 quel que soit le substituant étudié.

4.3- Influence des substituants sur le carbone 4

L'influence du substituant sur les paramètres géométriques (Tableau 40) et électroniques (Tableau 37) sont moins importants dans cette position. Par exemple, la charge sur le groupement nitro augmente seulement de -0,27 à -0,24 lorsque l'on passe de NH_2 à NO_2 et la longueur de la liaison C-N se situe dans un intervalle de seulement 0,004 Å. Ceci est dû au fait que l'on se situe en position méta par rapport au groupement nitro. Or, ce dernier ne présente d'effets mésomères qu'en en ortho et en para et non en méta.

Les variations concernant les énergies de dissociation de la liaison C-NO_2 sont également très faibles (de 52,3 kcal/mol pour NO_2 et 55,8 pour NH_2). De même, l'étape limitante du second mécanisme reste la même (TS1) quel que soit le substituant et le mécanisme d'interaction ortho demeure énergétiquement plus favorable que la simple dissociation homolytique. Il est d'ailleurs intéressant de remarquer que dans cette position, aucune étape du second mécanisme n'est influencée par la nature du substituant.

	-H	-NO_2	-CH_3	-NH_2	-COOH	-OH
Distances (Å)						
N-O_1	1,223	1,221	1,223	1,222	1,222	1,222
N-O_2	1,223	1,220	1,223	1,223	1,221	1,224
C_2-N	1,466	1,469	1,466	1,467	1,467	1,467
C_1-C_2	1,406	1,407	1,403	1,404	1,407	1,403
C_2-C_3	1,394	1,390	1,395	1,392	1,389	1,396
Angles (deg)						
C_1-C_2-N	121,4	121,5	121,3	121,2	121,3	121,6
C_3-C_2-N	115,9	115,6	115,8	115,3	115,9	115,3
C_2-N-O_1	118,1	117,7	118,1	118,1	117,8	118,2
C_2-N-O_2	117,6	117,4	117,7	117,8	117,5	117,7
Angle dièdre (deg)						
O_1-N-C_2-C_1	-19,4	-22,6	-18,5	-16,0	-23,6	-10,1

Tableau 40 - Paramètres structuraux principaux des o-nitrotoluènes substitués en position 4. La nomenclature des atomes est indiquée en Figure 64

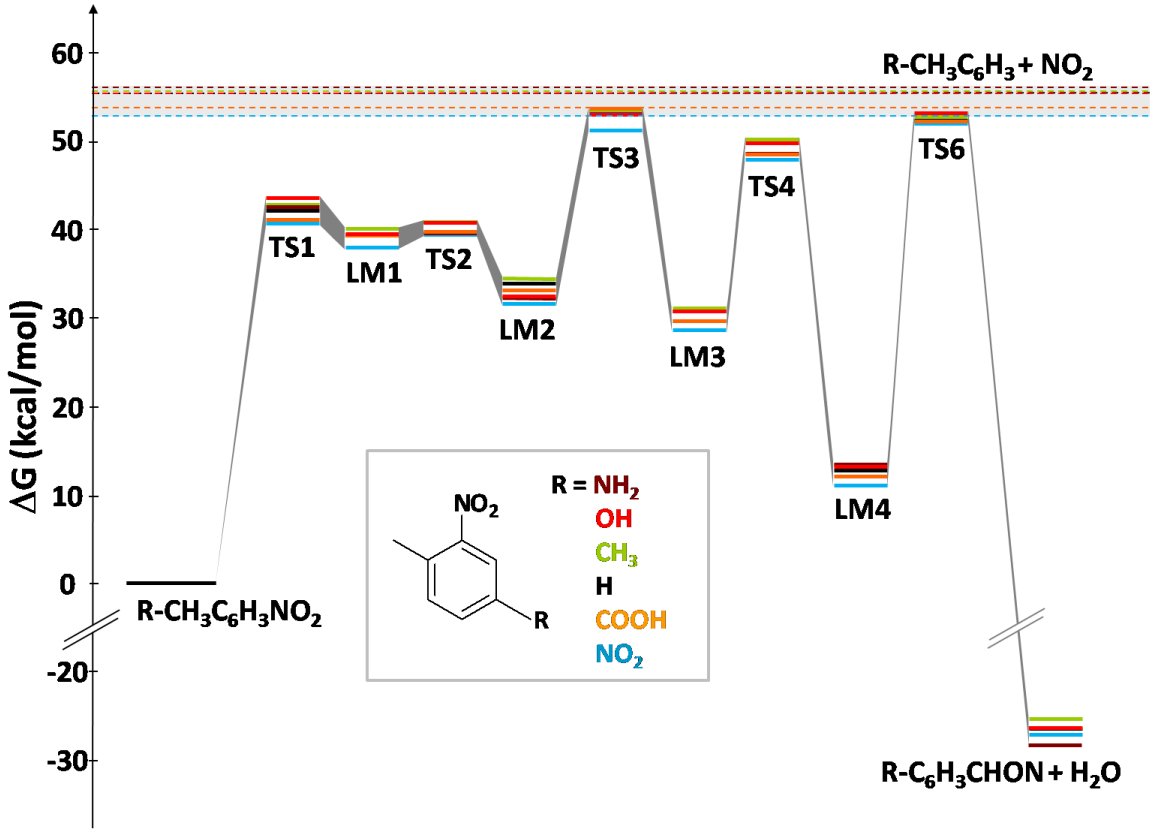


Figure 66 - Profils énergétiques pour les réactions de décomposition principales des o-nitrotoluènes substitués en position 4

	-H	-NO_2	-CH_3	-NH_2	-COOH	-OH
R-$CH_3C_6H_3NO_2$	0,0	0,0	0,0	0,0	0,0	0,0
R-$CH_3C_6H_3$ + NO_2	55,1	52,3	55,4	55,8	53,4	55,2
TS1	42,0	40,7	42,6	42,5	41,1	43,6
LM1	39,5	37,7	39,9	38,8	38,5	38,8
TS2	40,8	39,2	40,9	39,0	39,5	40,9
LM2	33,9	31,8	34,1	32,2	33,1	32,3
TS3	53,6	52,7	53,5	51,2	53,4	53,2
LM3	30,7	28,4	31,0	29,4	29,6	30,7
TS4	49,3	47,7	49,8	48,7	48,3	49,4
LM4	12,8	11,0	13,1	13,4	11,9	13,3
TS5	52,0	51,9	52,6	52,1	52,1	52,9
R-C_6H_3CHON + H_2O	-26,2	-27,0	-25,6	-28,2	-26,1	-26,2

Tableau 41 - Energies libres de Gibbs relatives (en kcal/mol) pour les réactions de décomposition principales des o-nitrotoluènes substitués en position 4

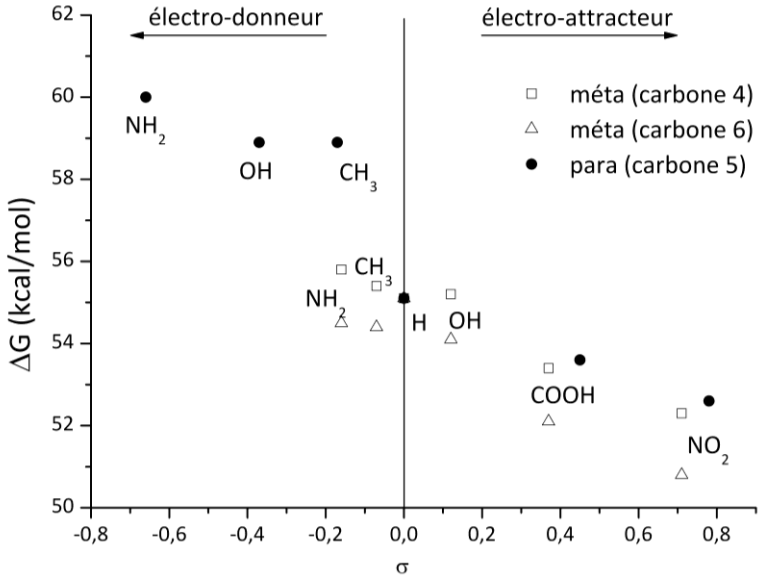


Figure 67 - Energies libres de Gibbs relatives de la dissociation C-NO_2 (ΔG) en fonction des constantes de Hammett (σ)

Malgré les faibles variations d'énergie pour la liaison C-NO_2, ces dernières peuvent être analysées en termes de constantes de Hammett (σ). Sur la Figure 67, les enthalpies libres de dissociation de la liaison C-NO_2 sont représentées en fonction du paramètre de Hammett et une corrélation linéaire (R^2=0,96) est observée alors que la dépendance sur la barrière d'activation du second chemin réactionnel n'est pas aussi significative (R^2=0,71). Ce résultat reste cohérent avec ce qui a été précédemment observé avec la réaction 2.

4.4- Influence des substituants sur le carbone 5

Les structures des différentes molécules dérivées du 2-nitrotoluène par substitution au niveau du carbone 5 sont présentées dans Tableau 42 et les charges du groupement nitro dans le Tableau 37.

	-H	-NO_2	-CH_3	-NH_2	-COOH	-OH
Distances (Å)						
N-O_1	1,223	1,221	1,224	1,227	1,221	1,224
N-O_2	1,223	1,221	1,224	1,227	1,222	1,225
C_2-N	1,466	1,469	1,462	1,450	1,468	1,457
C_1-C_2	1,406	1,405	1,405	1,412	1,405	1,412
C_2-C_3	1,394	1,393	1,396	1,398	1,394	1,395
Angles (deg)						
C_1-C_2-N	121,4	120,9	121,8	122,5	121,1	122,3
C_3-C_2-N	115,9	115,8	116,2	116,5	115,9	116,1
C_2-N-O_1	118,1	117,7	118,2	118,5	117,9	118,4
C_2-N-O_2	117,6	117,3	117,8	118,0	117,4	117,9
Angle dièdre (deg)						
O_1-N-C_2-C_1	-19,4	-26,9	-14,5	0,0	-24,9	0,2

Tableau 42 - Paramètres structuraux principaux des o-nitrotoluènes substitués en position 5. La nomenclature des atomes est indiquée en Figure 64

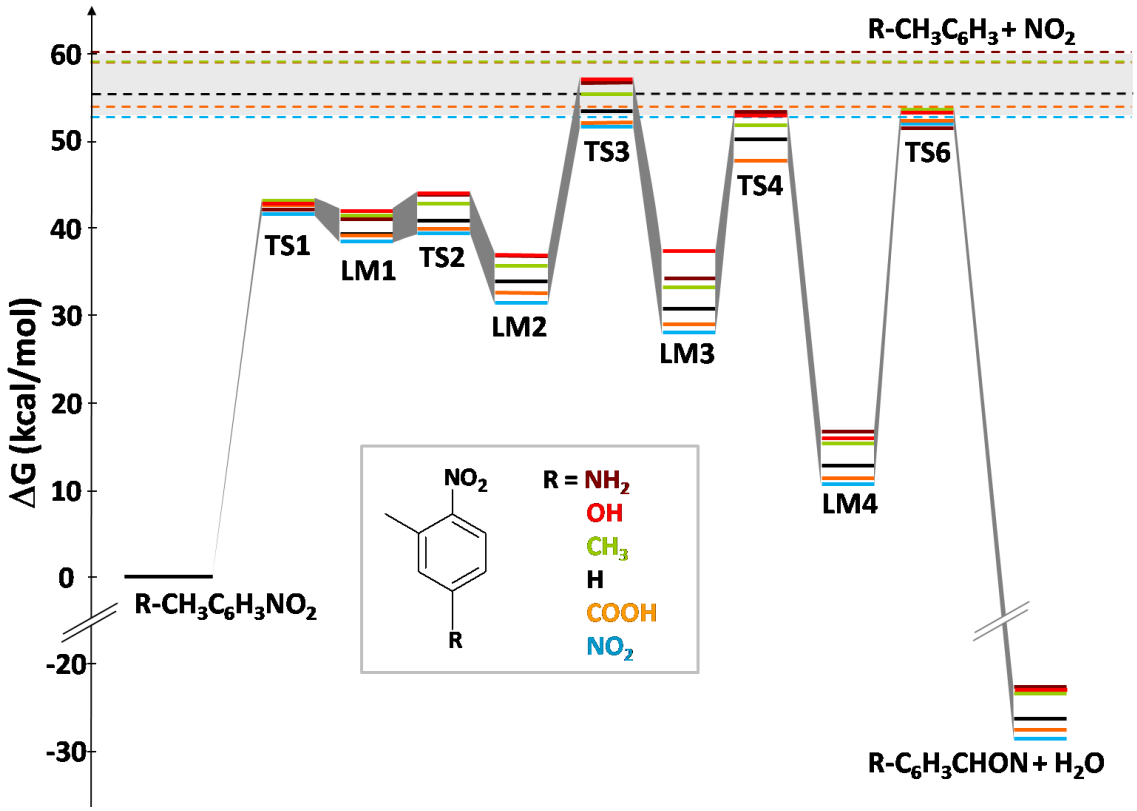


Figure 68 - Profils énergétiques pour les réactions de décomposition principales des o- nitrotoluènes substitués en position 5

	-H	-NO_2	-CH_3	-NH_2	-COOH	-OH
R-$CH_3C_6H_3NO_2$	0,0	0,0	0,0	0,0	0,0	0,0
R-$CH_3C_6H_3$ + NO_2	55,1	52,6	58,9	60,0	53,6	58,9
TS1	42,0	41,9	43,4	41,5	42,2	43,1
LM1	39,5	38,2	41,2	41,0	38,8	41,8
TS2	40,8	38,8	42,7	43,4	39,6	43,8
LM2	33,9	31,2	35,8	36,9	32,4	37,1
TS3	53,6	51,6	55,5	56,5	52,0	56,9
LM3	30,7	28,0	32,8	34,0	28,9	37,1
TS4	49,3	46,7	51,7	52,9	47,8	52,6
LM4	12,8	10,3	15,1	16,5	11,2	15,8
TS5	52,0	51,8	53,4	51,4	51,9	53,0
R-C_6H_3CHON + H_2O	-26,2	-28,3	-23,7	-22,4	-27,6	-22,6

Tableau 43 - Energies libres de Gibbs relatives (en kcal/mol) pour les réactions de décomposition principales des o-nitrotoluènes substitués en position 5

Comme prévu, le groupement le plus électro-donneur (amino) correspond à la valeur la plus élevée pour la charge du groupement nitro partant (-0,24) alors que le plus électro-attracteur (nitro) correspond à la plus faible (-0,31). En conséquent, la distance carbone-azote est maximale pour le substituant nitro (1,47 Å) et décroit jusqu'à 1,45 Å pour le groupement amino. Cela dit, ces variations géométriques ne sont pas particulièrement importantes.

Les énergies des différents points stationnaires le long des deux chemins réactionnels étudiés sont reportées dans la Figure 68 et le Tableau 43. La substitution sur le carbone 5 a plus d'influence que sur le carbone 4 du fait des propriétés mésomères de NO_2. Une fois encore, cette influence peut être corrélée au caractère électro-donneur ou accepteur des substituants, en considérant les constantes de Hammett (R^2=0,91) comme l'illustre la Figure ***67***.

En fait, aucune dépendance de ce type n'est établie pour l'énergie d'activation du second chemin réactionnel, qui correspond là encore à l'attaque initiale de la liaison C-H par un oxygène du nitro. En effet, si la plupart des points stationnaires sont influencés par la nature du substituant (les produits de réaction s'étendent, par exemple, sur un intervalle de -28,3 kcal/mol pour NO_2 à -22,4 kcal/mol pour NH_2), ce n'est pas le cas pour la première barrière (TS1) qui est l'étape limitante du mécanisme (de 41,5 kcal/mol pour NO_2 à 43,4 kcal/mol pour CH_3). Cela doit être dû à une propagation plus favorable des effets électroniques au sein de la conformation cyclique de R-C_6H_3CHON par rapport à la structure initiale de type nitrotoluène. Au contraire, des variations plus faibles sont observées pour les intermédiaires réactionnels, qui sont plus stables lorsque le groupement sur le carbone 5 est électro-donneur. Ce comportement peut être rationalisé de manière qualitative en s'intéressant de plus près aux structures de résonnance des intermédiaires LM1 et LM2 (Figure 69).

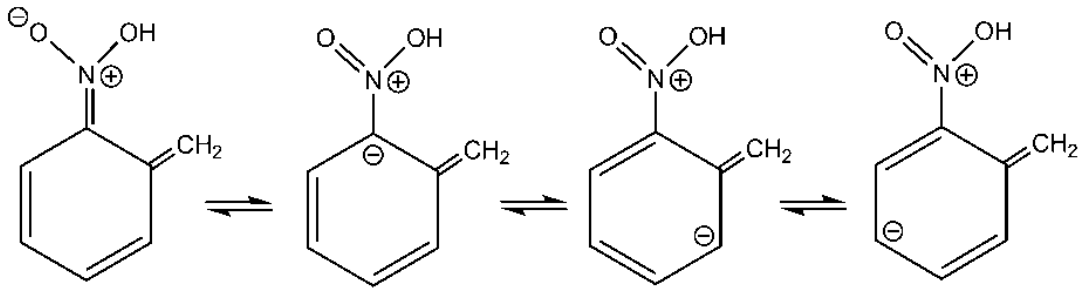


Figure 69 - Structures de résonnance pour LM1

Les groupements électro-attracteurs, en stabilisant la charge négative en position 5, accroissent la stabilité des deux intermédiaires. Ce n'est pas le cas pour les dérivés substitués en position 4 et 6. Cela intervient par contre pour ceux substitués en position 3 mais ce comportement est alors caché derrière des interactions directes fortes (stériques et électroniques) qui interviennent entre les susbtituants alors positionnés en ortho l'un par rapport à l'autre.

Finalement, quel que soit le substituant sur le carbone 5, la compétition entre les deux chemins de décomposition reste inchangée. Le processus d'interaction ortho est moins énergétique que la dissociation homolytique de la liaison C-NO_2 et reste donc le plus favorable.

4.5- Influence des substituants sur le carbone 6

La dernière position envisagée est localisée en position ortho du groupement méthyle. Les différents paramètres géométriques sont reportés dans le Tableau 44. Aucune variation significative de charge n'est observée d'un substituant à un autre : de -0,26 pour H à -0,23 pour NO_2. En parallèle, toutes les liaisons carbone-azote sont comprises dans un intervalle de 0,004 Å.

Dans cette position, le substituant se situe en position méta vis-à-vis du groupement nitro qui ne présente pas d'effet mésomère dans cette position. Néanmoins, des variations significatives apparaissent d'un substituant à un autre, pour les deux chemins réactionnels envisagés (voir Figure 70 et Tableau 45). Cela peut être expliqué par des effets ortho, c'est-à-dire des interactions avec le groupement méthyle voisin. Pour la rupture de la liaison C-NO_2, les énergies libres de Gibbs sont comprises entre 50,8 (pour NO_2) et 55,1 kcal/mol (pour H). De plus, elles sont significativement corrélées avec les constantes de Hammett (R^2=0,91, cf. Figure 67).

Pour le second mécanisme, la première partie du processus (transfert d'hydrogène depuis CH_3 vers NO_2) n'est pas affectée par la substitution alors que la seconde partie, menant à la formation de la structure R-C_6H_3CHON, l'est significativement comme cela avait été le cas pour la substitution au niveau du carbone 5. Pour autant, l'étape limitante reste le transfert initial d'hydrogène du méthyle vers le nitro.

Au final, la compétition entre les deux chemins réactionnels n'est pas plus influencée dans cette position par rapport aux trois autres positions précédemment considérées. L'interaction ortho est énergétiquement favorable par rapport à la rupture de la liaison C-NO_2.

	-H	-NO_2	-CH_3	-NH_2	-COOH	-OH
Distances (Å)						
N-O_1	1,223	1,220	1,223	1,223	1,221	1,222
N-O_2	1,223	1,221	1,223	1,222	1,222	1,222
C_2-N	1,466	1,468	1,465	1,467	1,467	1,466
C_1-C_2	1,406	1,404	1,403	1,399	1,403	1,400
C_2-C_3	1,394	1,391	1,393	1,392	1,391	1,393
Angles (deg)						
C_1-C_2-N	121,4	120,5	120,9	120,7	120,8	120,2
C_3-C_2-N	115,9	115,4	115,5	115,2	115,1	115,8
C_2-N-O_1	118,1	117,7	118,0	118,2	117,8	118,0
C_2-N-O_2	117,6	117,1	117,5	117,3	117,1	117,4
Angle dièdre (deg)						
O_1-N-C_2-C_1	-19,4	-36,4	-33,0	-39,0	-39,8	-30,9

Tableau 44 - Paramètres structuraux principaux des o-nitrotoluènes substitués en position 6. La nomenclature des atomes est indiquée en Figure 64

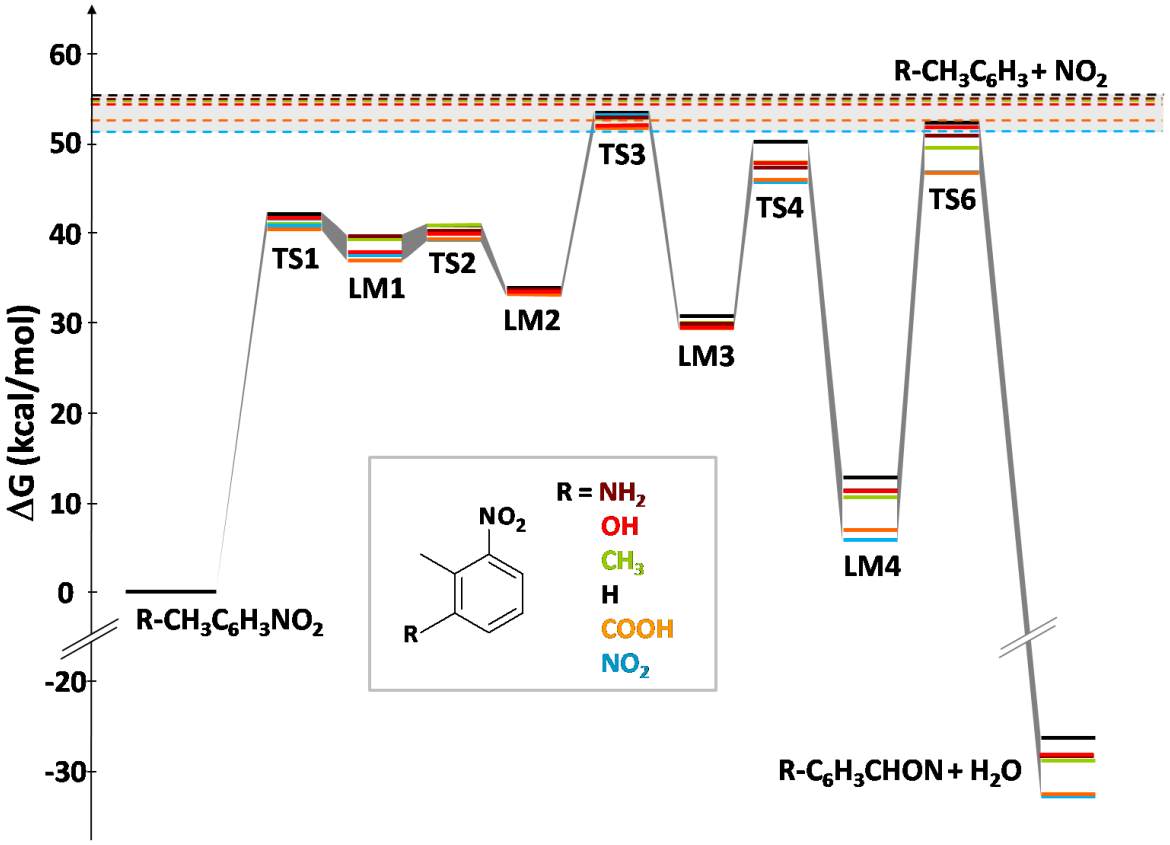


Figure 70 - Profils énergétiques pour les réactions de décomposition principales des o-nitrotoluènes substitués en position 6

	-H	-NO_2	-CH_3	-NH_2	-COOH	-OH
R-$CH_3C_6H_3NO_2$	0,0	0,0	0,0	0,0	0,0	0,0
R-$CH_3C_6H_3$ + NO_2	55,1	50,8	54,4	54,5	52,1	54,1
TS1	42,0	40,5	40,8	41,4	40,2	41,8
LM1	39,5	37,0	38,4	39,7	36,7	37,9
TS2	40,8	39,2	40,2	40,0	39,2	39,9
LM2	33,9	32,8	33,3	33,5	32,8	33,2
TS3	53,6	52,8	52,4	52,4	51,3	51,9
LM3	30,7	29,5	29,8	29,9	29,0	29,3
TS4	49,3	45,6	47,7	47,2	45,8	47,8
LM4	12,8	5,7	10,3	11,4	6,5	11,4
TS5	52,0	46,5	49,4	50,7	46,4	51,4
R-C_6H_3CHON + H_2O	-26,2	-33,0	-29,0	-28,3	-32,6	-28,0

Tableau 45 - Energies libres de Gibbs relatives (en kcal/mol) pour les réactions de décomposition principales des o-nitrotoluènes substitués en position 6

5- Conclusion

L'influence de la nature et de la position des substituants sur les chemins de décomposition d'une série de composés nitroaromatiques a été étudiée à l'aide de la théorie de la fonctionnelle de la densité au niveau PBE0/6-31+G(d,p). En particulier, nous avons étudié l'étape d'initiation de la décomposition qui est l'étape limitante du processus.

Dans un premier temps, les chemins de décomposition principaux du nitrobenzène ont été envisagés. Le mécanisme de décomposition le plus favorable est la dissociation homolytique de la liaison C-NO_2 (C_6H_5 + NO_2). Le second chemin de décomposition consiste en un processus en deux étapes, menant à la formation de deux radicaux C_6H_5O et NO. Ces réactions ont alors été étudiées pour une série de 15 composés dérivés du nitrobenzène par substitution (nitro, méthyle, amino, acide carboxylique et hydroxyle) en positions ortho, méta et para vis-à-vis du groupement nitro partant. En position méta, aucun changement significatif n'est apparu dans les profils énergétiques des réactions alors que les dérivés para et ortho montrent des variations significatives, du fait de l'effet mésomère du groupement nitro, inexistant dans cette position. D'ailleurs, des corrélations significatives ont pu être mises en évidence entre les énergies de dissociation des liaisons carbone-nitro et les constantes de Hammett, employées comme indicateurs des effets électro-donneurs et accepteurs des substituants. Néanmoins, aucune influence significative n'a été finalement observée sur la compétition entre ces deux chemins de décomposition et la rupture directe de la liaison C-NO_2 reste le processus le plus favorable.

Dans le cas de composés substitués en ortho, d'autres mécanismes peuvent intervenir mettant en jeu des interactions directes entre le groupement nitro et le substituant en ortho. C'est en particulier le cas pour le 2-nitrotoluène et ses dérivés qui ont donc été étudiés ici. La compétition entre ce nouveau chemin et la dissociation de C-NO_2 a été étudiée pour ce composé ainsi que pour 20 dérivés par substitution (nitro, méthyle, amino, acide carboxylique et hydroxyle) au niveau des quatre positions accessibles. Ce nouveau mécanisme consiste en une attaque d'une liaison C-H du méthyle menant à la formation d'anthranile (C_6H_3CHON) et d'eau (H_2O). Il est plus favorable au niveau énergétique. L'influence des substituants sur le mécanisme dissociatif C-NO_2 a, une nouvelle fois, pu être reliée aux constantes de Hammett alors que la barrière d'activation du mécanisme d'interaction ortho reste insensible à la substitution. Finalement, les résultats obtenus montrent que quelque soit le composé, l'interaction ortho constitue la voie d'initiation principale pour la décomposition.

D'un point de vue général, cette étude mécanistique a mis en évidence la complexité du processus de décomposition des composés nitroaromatiques. Différents mécanismes de décomposition interviennent, notamment sous l'effet d'interactions entre le groupement nitro partant et un substituant situé en ortho. La prise en compte de ces différents mécanismes peut être un point clés dans le développement de modèles prédictifs des propriétés en lien avec la décomposition (ex : propriétés explosives) pour de tels composés,

que ce soit pour la recherche de descripteurs ou pour la mise en place des bases de données. Cette étude est donc tout à fait complémentaire de l'analyse QSPR puisqu'elle permet de permet de mieux appréhender certains points clés du développement des modèles par une description explicite des mécanismes de décomposition.

CONCLUSION GENERALE

Si l'utilisation de modèles QSAR/QSPR est une pratique d'usage courant pour prédire les propriétés toxicologiques ou écotoxicologiques des substances chimiques, elle l'est beaucoup moins pour les propriétés physico-chimiques dangereuses, telles que l'explosibilité. Or, le nouveau cadre règlementaire européen REACH leur donne un intérêt nouveau puisque l'utilisation de telles approches y est recommandée, en particulier pour éviter le recours aux essais sur animaux. De plus, aux vues du nombre de substances concernées et du calendrier imposé, l'évaluation complète par voie expérimentale est impossible.

Dans ce contexte, l'objectif de cette thèse était de développer de nouveaux modèles QSPR efficaces pour la prédiction des propriétés des composés explosibles. Des méthodologies efficaces ont été identifiées et utilisées pour obtenir des modèles à la fois fiables, c'est-à-dire prédictifs, et interprétables du point de vue phénoménologique, et ainsi répondre aux exigences d'une utilisation règlementaire. La classification des substances explosibles nécessite la caractérisation de plusieurs propriétés. Au cours de cette thèse, la stabilité thermique (chaleur et température de décomposition) et la sensibilité (à la décharge électrique et à l'impact) d'une famille d'explosifs, les composés nitroaromatiques, ont été étudiées.

Les aspects méthodologiques du développement de modèles QSPR prédictifs ont été dans un premier temps considérés pour l'étude de la chaleur de décomposition.

Les descripteurs issus de calculs de chimique quantique (parmi lesquels les descripteurs de réactivité de la DFT conceptuelle) se sont montrés d'un grand intérêt. Non seulement ils contribuent à obtenir des modèles robustes et prédictifs mais ils rendent également ces derniers plus interprétables d'un point de vue phénoménologique. Finalement, plus de 300 descripteurs, quantiques mais aussi constitutionnels, topologiques et géométriques ont été employés. Ce jeu varié de descripteurs permet aux modèles obtenus de rendre compte des deux facteurs majeurs régissant la chaleur de décomposition des composés nitroaromatiques, à savoir le nombre de groupements nitro et leur réactivité au cours du processus de décomposition.

Au niveau de l'analyse de données, indispensable pour traiter une telle quantité d'informations, différents outils statistiques ont été employés. Une méthodologie basée sur la *Best Multi-Linear Regression* (BMLR) implémentée dans le logiciel CodessaPro a tout

d'abord été utilisée. Cette méthode permet d'extraire de manière efficace des modèles multi-linéaires corrélés et robustes en choisissant les descripteurs les plus pertinents pour le modèle de manière très automatisée. Le modèle est tout d'abord mis en place pour un jeu de molécules dit d'entraînement. Il s'agit ensuite de le valider à l'aide d'un autre jeu de molécules dit de validation. Son domaine d'applicabilité est ensuite défini à partir des valeurs prises par les descripteurs utilisés dans le modèle pour le jeu d'entraînement. Le pouvoir prédictif du modèle dans son domaine d'applicabilité est enfin déterminé à l'aide des molécules du jeu de validation inclues dans ce dernier.

Pour des propriétés et molécules aussi complexes que celles considérées ici, d'autres types d'outils, moins automatisés, peuvent être employés. Les analyses par projection (PCA et PLS) ont démontré leur efficacité pour la sélection manuelle des descripteurs sur des critères non seulement statistiques mais aussi mécanistiques. Ces analyses permettent également de mettre en évidence la présence éventuelle de sous-groupes distincts de molécules au sein de la base de données employée. Or, de tels sous-groupes sont néfastes pour l'obtention de modèles robustes et prédictifs. Enfin, une méthode de classification, par arbre de décision, a été utilisée. Cette approche, bien que qualitative, présente l'avantage de mettre en place des modèles robustes, simples et facilement interprétables au niveau phénoménologique pour peu que les descripteurs aient été correctement choisis.

Cette première étude a mené à l'obtention de deux modèles prédictifs pour la chaleur de décomposition à partir d'une base de données de 77 composés nitroaromatiques. Le premier est un modèle multilinéaire, issu de l'analyse BMLR, pour les composés nitroaromatiques ne présentant pas de substituant en position ortho du groupement nitro, de tels substituants pouvant engendrer des mécanismes de décomposition spécifiques, selon la nature du substituant. Ce modèle présente une prédictivité importante (R^2_{ext}=0,84), en particulier dans son domaine d'applicabilité (R^2_{in}=0,86). Pour une analyse plus globale, c'est-à-dire applicable à tous les dérivés du nitrobenzène, un modèle qualitatif par arbre de décision a été proposé. Ce modèle est lui aussi très prédictif puisqu'il offre 82% de classification correcte sur le jeu de validation employé. Ces deux modèles sont d'ores et déjà applicables en termes de prédiction.

Les températures de décomposition de 22 molécules dérivées du nitrobenzène ont ensuite été étudiées. La taille de la base de données ne permet pas de réaliser une étude complète incluant une validation externe et une analyse du domaine d'applicabilité. Cependant, des modèles QSPR ont été développés aux niveaux de calcul AM1 et DFT. Les deux modèles présentent des robustesses importantes (respectivement R^2_{cv}=0,84 et 0,81) mais le modèle DFT étant moins paramétré, il présente moins de risque de sur-paramétrisation et est donc préféré. Si, pour un usage prédictif, une base de données plus importante est nécessaire, afin de réaliser une validation externe, cette analyse a permis de mettre en évidence la pertinence du niveau DFT employé pour le calcul des descripteurs en particulier quantiques. La même observation avait d'ailleurs pu être faite pour la chaleur de décomposition des mêmes composés nitroaromatiques.

La troisième propriété étudiée a été la sensibilité à la décharge électrique pour 41 composés nitroaromatiques, d'une plus grande variété que les précédentes bases de données (incluant notamment des nitroaromatiques polycycliques). Si un modèle corrélé (R^2=0,90) et robuste (R^2_{cv}=0,78) a été obtenu, la validation externe a mis en évidence son absence de prédictivité (R^2_{ext}<0,01). Ce manque de pouvoir prédictif a pu être attribué, par une analyse par composantes principales (PCA), à la trop grande diversité de la base de données. En effet, des sous-groupes de molécules, qu'il faudrait considérer de manière indépendante, ont été identifiés, mais la base de données actuelle n'est pas suffisamment importante pour le faire. Des données expérimentales supplémentaires sont donc nécessaires pour l'obtention de modèles QSPR prédictifs et l'analyse PCA a permis d'identifier la manière dont devraient être choisies les molécules d'une base de données pour cette propriété. Quoi qu'il en soit, les modèles obtenus rendent compte des aspects phénoménologiques. En effet, ici, le nombre de groupements nitro n'est plus important, seule la réactivité de ce dernier intervient.

Concernant la prédiction de la sensibilité à l'impact, les mêmes difficultés ont été rencontrées. Aucun modèle QSPR prédictif n'est obtenu et la base de données, analysée par PCA, présente en son sein des sous-groupes de molécules. Cela dit, l'obtention de données supplémentaires semble plus aisée pour cette propriété, puisqu'il s'agit de la sensibilité la plus couramment testée expérimentalement.

En fait, les composés nitroaromatiques sont une classe de molécules pour laquelle la décomposition est particulièrement complexe. Ceci explique les difficultés observées. L'utilisation de la chimie quantique nous a d'ailleurs permis d'appréhender de manière explicite ces mécanismes. Pour les dérivés du nitrobenzène, différents mécanismes réactionnels ont été identifiés selon que les composés présentent ou non des substituants en position ortho du groupement nitro. Cela justifie bien le traitement spécifique des composés nitroaromatiques non-ortho-substitués pour la chaleur de décomposition puisqu'ils suivent tous un même mécanisme de décomposition simple, initié par la dissociation de la liaison C-NO_2.

Finalement, les objectifs principaux de cette thèse ont donc été remplis dans le sens où ils ont permis de mettre en évidence des méthodologies efficaces, en accord avec les règles mises en place par l'OCDE quant à la validation des modèles QSPR :

- importance de la définition et de la fiabilité de la base de données ;
- développement de modèles non ambigus à partir d'une équation ou d'un arbre de décision et de descripteurs obtenus à un niveau de calcul suffisant et défini ;
- définition des domaines d'applicabilité des modèles ;
- mesures de la corrélation et de la robustesse des modèles à partir du jeu d'entraînement et de leur pouvoir prédictif à partir d'un jeu de validation ;
- mise en évidence des mécanismes sous-jacents par l'intégration de descripteurs quantiques d'une part et à l'aide de l'analyse des mécanismes réactionnels d'autre part.

Deux modèles prédictifs ont même été obtenus pour la chaleur de décomposition des dérivés du nitrobenzène.

En vue du développement d'un outil d'évaluation global de l'explosibilité des substances chimiques, l'emploi de modèles issus de l'approche combinée QSPR-DFT utilisée dans ces travaux présente donc un fort potentiel. Les différents outils utilisés ici permettent de traiter de manière efficace les propriétés d'explosibilité pour des composés aussi complexes que les nitroaromatiques. Les problèmes rencontrés pour les sensibilités des nitroaromatiques sont attribuables à la complexité de ces mécanismes, mais pourraient ne pas être rencontrés pour d'autres familles de composés (nitramines, nitroaliphatiques, peroxydes), dont les mécanismes de décomposition sont moins complexes.

Pour en arriver à un tel outil, il est bien entendu indispensable de développer des modèles prédictifs pour toutes les propriétés d'explosibilité (puissance explosive, sensibilité à la chaleur…). Pour cela, l'obtention de bases de données expérimentales est nécessaire. Ces dernières permettront aussi la consolidation des modèles pour les propriétés de sensibilité. D'une manière générale, la collecte des données expérimentales a été l'une des difficultés majeures rencontrées au cours de cette thèse. Si elles semblent être disponibles dans la littérature pour la sensibilité à l'impact (voire pour la température de décomposition), de nouvelles bases de données devront être mises en place pour d'autres propriétés (sensibilité à la décharge électrique, à la friction ou aptitude à la déflagration). De plus, la nature du dispositif ayant une grande influence sur les résultats obtenus, les nouvelles valeurs complèteront difficilement les bases de données existantes et l'utilisation de données obtenues avec un unique dispositif sont préférables.

Enfin, des méthodes d'analyse de données performantes ont été utilisées mais d'autres types d'approches peuvent encore être utilisés selon les propriétés et molécules considérées (réseaux de neurones, régressions non-linéaires, algorithmes génétiques).

Ces travaux de thèses ouvrent donc de nombreuses perspectives en termes de méthodologies, de propriétés, de molécules et surtout d'utilisations finales des modèles développés dans le cadre de REACH.

BIBLIOGRAPHIE

[1] J.C. Etchells, *Org. Process Res. Dev.*, **1997**, *1*, 435-437.
[2] V.K. Agrawal, P.V. Khadikar, *Bioorg. Med. Chem.*, **2001**, *9*, 3035-3040.
[3] D.A. Winkler, *Brief. Bioinf.*, **2002**, *3*, 73-86.
[4] H. Gao, J.A. Katzenellenbogen, R. Garg, C. Hansch, *Chem. Rev.*, **1999**, *99*, 723-744.
[5] C.D. Selassie, R. Garg, S. Kapur, A. Kurup, R.P. Verma, S.B. Mekapati, C. Hansch, *Chem. Rev.*, **2002**, *102*, 2585-2606.
[6] S.P. Bradbury, *Toxicol. Lett.*, **1995**, *79*, 229-237.
[7] R. Garg, S.P. Gupta, H. Gao, M.S. Babu, A.K. Debnath, C. Hansch, *Chem. Rev.*, **1999**, *99*, 3525-3602.
[8] M. Grover, B. Singh, M. Bakshi, S. Singh, *Pharm. Sci. Tech. Today*, **2000**, *3*, 50-57.
[9] M. Grover, B. Singh, M. Bakshi, S. Singh, *Pharm. Sci. Tech. Today*, **2000**, *3*, 28-35.
[10] A.R. Katritzky, M. Karelson, V.S. Lobanov, *Pure Appl. Chem.*, **1997**, *69*, 245-248.
[11] A.R. Katritzky, V.S. Lobanov, M. Karelson, *Chem. Soc. Rev.*, **1995**, *24*, 279-287.
[12] J. Taskinen, J. Yliruusi, *Adv. Drug Deliv. Rev.*, **2003**, *55*, 1163-1183.
[13] E.S. Shanley, G.A. Melhem, *J. Loss. Prev. Process Ind.*, **1995**, *8*, 261-264.
[14] G.A. Melhem, E.S. Shanley, *Process Saf. Prog.*, **1996**, *15*, 168-172.
[15] T. Grewer, *Thermochim. Acta*, **1991**, *187*, 133-149.
[16] S.R. Saraf, W.J. Rogers, M.S. Mannan, *J. Hazard. Mater.*, **2003**, *98*, 15-29.
[17] Règlement (CE) n° 1907/2006 du Parlement Européen et du Conseil du 18 décembre 2006 concernant l'enregistrement, l'évaluation et l'autorisation des substances chimiques, ainsi que les restrictions applicables à ces substances (REACH), instituant une agence européenne des produits chimiques, modifiant la directive 1999/45/CE et abrogeant le règlement (CEE) n° 793/93 du Conseil et le règlement (CE) n° 1488/94 de la Commission ainsi que la directive 76/769/CEE du Conseil et les directives 91/155/CEE, 93/67/CEE, 93/105/CE et 2000/21/CE de la Commission.
[18] N. Margossian, *Le règlement REACH - La règlementation européenne sur les produits chimiques*, Dunod / L'Usine Nouvelle, Paris, **2007**.
[19] Institut National de Recherche et de Sécurité (INRS). www.inrs.fr [20 janvier 2010]
[20] European Chemicals Agency (ECHA). http://echa.europa.eu/ [28 août 2009]
[21] REACH - Service National d'Assistance Règlementaire. www.ineris.fr/reach-info [20 janvier 2010]
[22] Directive du Conseil n° 67/548/CEE du 27 juin 1967 concernant le rapprochement des dispositions législatives, réglementaires et administratives relatives à la classification, l'emballage et l'étiquetage des substances dangereuses.
[23] *Stratégie pour la future politique dans le domaine des substances chimiques, Livre Blanc, COM(2001)88 final*, Commission européenne, 2001.
[24] Règlement (CE) n° 1272/2008 du Parlement Européen et du Conseil du 16 Décembre 2008 relatif à la classification, à l'étiquetage et l'emballage des substances et des mélanges, modifiant et abrogeant les directives 67/548/CEE et 1999/45/CE et modifiant le règlement CE n° 1907/2006.
[25] Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Second Revised Edition, United Nations, 2007.

[26] F. Pedersen, J. De Bruijn, S.J. Munn, K. Van Leeuwen, *Assessment of additional testing needs under REACH. Effects of QSARs - risk based testing and voluntary industry initiatives, JRC Report EUR 20863 EN*, 2003.
[27] European Center for the Validation of Alternative Methods (ECVAM). http://ecvam.jrc.it/ [14 novembre 2009]
[28] Organisation de Coopération et de Développement Economiques (OCDE). http://www.oecd.org [20 octobre 2009]
[29] *Final Report of the OECD Workshop on Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Test Methods*, Organisation de Coopération et de Développement Economique, Paris, 2009.
[30] *Principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure-Activity Relationship Models*, Organisation de Coopération et de Développement Economique, Paris, 2009.
[31] A. Lewis, N. Kazantzis, I. Fishtik, J. Wilcox, *J. Hazard. Mater.*, **2007**, *142*, 592-602.
[32] C. Michot, M. Demissy, D. Gaston, *Propel. Explos. Pyrotechn.*, **1991**, *16*, 110-114.
[33] T. Yoshida, *Safety of Reactive Chemicals*, Elsevier, Amsterdam, **1987**.
[34] S. Mannan, *Lee's Loss Prevention in Process Industries: Hazard Identification, Assessment and Control*, Elsevier Butterworth-Heinemann, Burlington, **2005**.
[35] R. Meyer, J. Köhler, A. Homburg, *Explosives*, Wiley-VCH, Weinheim, **2002**.
[36] L. Medard, *Les explosifs occasionnels*, Tech. & Doc., Cachan, France, **1987**.
[37] J. Akhavan, *The chemistry of exposives*, Royal Society of Chemistry, Cambridge, UK, **1998**.
[38] R. Chéret, *La détonation des explosifs condensés, Tome 1*, Masson, Paris, **1988**.
[39] J.A. Conkling, *Chemistry of pyrotechnics - Basic principles and theory*, Marcel Dekker Inc., New-York, **1985**.
[40] Recommendations on the Transport of Dangerous Goods: Manual of Tests and Criteria,fifth revised edition, ST/SG/AC.10/11/Rev.5, United Nations, 2009.
[41] Institut National de l'Environnement Industriel et des Risques (INERIS). www.ineris.fr]
[42] C. Michot, *Revue générale de sécurité*, **1992**, *113*, 41-47.
[43] Règlement (CE) n° 761/2009 de la Commission du 23 juillet 2009 modifiant, aux fins de son adaptation au progrès technique, le règlement (CE) n° 440/2008 établissant des méthodes d'essai conformément au règlement (CE) n° 1907/2006 du Parlement européen et du Conseil concernant l'enregistrement, l'évaluation et l'autorisation des substances chimiques, ainsi que les restrictions applicables à ces substances (REACH).
[44] W.C. Lothrop, G.R. Handrick, *Chem. Rev.*, **1949**, *44*, 419-445.
[45] B.T. Federoff, O.E. Sheffield, G.D. Clift, E.F. Reese, *Encyclopedia of Explosives and related items, Volume 2* Picatinny Arsenal, Dover, **1962**.
[46] M.-F. Robbe, *Energetika*, **2007**, *53*, 62-71.
[47] M. Belk, V. Volpert, *Chaos*, **2004**, *14*, 263-273.
[48] M.S. Miller, Burning-Rate Models and their successors, a personal perspective, dans: Overviews of recent research on energetic materials, R. Shaw, T.B. Brill, D.L. Thompson, World Scientific Publishing, 2005, p. 419-472.
[49] B.M. Rice, Applications of theoretical chemistry in assessing energetic materials for performance or sensitivity, dans: Overviews of recent research on energetic materials, R. Shaw, T.B. Brill, D.L. Thompson, World Scientific Publishing, 2005, p. 335-368.
[50] L.E. Fried, M.R. Manaa, J.P. Lewis, Modeling the reactions of energetic materials in the condensed phase, dans: Overviews of recent research on energetic materials, R. Shaw, T.B. Brill, D.L. Thompson, World Scientific Publishing, 2005, p. 275-302.
[51] D.L. Thompson, Gas Phase Decomposition of Energetic Molecules, dans: Overviews of recent research on energetic materials, R. Shaw, T.B. Brill, D.L. Thompson, World Scientific Publishing, 2005, p. 241-274.
[52] M.H. Keshavarz, M. Jaafari, *Propel. Explos. Pyrotechn.*, **2006**, *31*, 216-225.
[53] H. Nefati, J.-M. Cense, J.-J. Legendre, *J. Chem. Inf. Comput. Sci.*, **1996**, *36*, 804-810.

[54] B.M. Rice, J.J. Hare, *J. Phys. Chem. A*, **2002**, *106*, 1770-1783.
[55] P. Politzer, S. Boyd, *Struct. Chem.*, **2002**, *13*, 105-113.
[56] E.J. Reed, M.R. Manaa, L.E. Fried, K.R. Glaesemann, J.D. Joannopoulos, *Nature Physics*, **2008**, *4*, 72-76.
[57] H. Nefati, Prédiction de la sensibilité au choc des substances explosives ou non : Approches statistiques et neuronales, Université Paris VI, Paris, 1994.
[58] T.B. Brill, K.J. James, *Chem. Rev.*, **1993**, *93*, 2667-2692.
[59] L. Pauling, E.B.J. Wilson, *Introduction to Quantum Mechanics with Applications to Chemistry*, McGraw-Hill Book Company Inc., New York, **1935**.
[60] K.S. Pitzer, *Quantum Chemistry*, Prentice-Hall Inc., Englewood Cliffs, New Jersey, **1953**.
[61] A. Szabo, N.S. Ostlund, *Modern Quantum Chemistry - Introduction to Advanced Electronic Structure Theory*, Macmillan Publishing Co. Inc., New York, **1982**.
[62] R.G. Parr, W. Yang, *Density Functional Theory of Atoms and Molecules*, Oxford University Press, New-York, **1989**.
[63] J.-L. Rivail, *Elements de chimie quantique à l'usage des chimistes*, CNRS Editions, Paris, **1999**.
[64] D.C. Young, *Computational Chemistry: A Practical Guide for Applying Techniques to Real-World Problems*, John Wiley & Sons Inc., New York, **2001**.
[65] C.J. Cramer, *Essentials of Computational Chemistry - Theories and Models*, Wiley, Chicherster, U.K., **2004**.
[66] W.J. Hehre, R.F. Stewart, J.A. Pople, **1969**, *51*, 2657-2664.
[67] C.C.J. Roothaan, *Rev. Mod. Phys.*, **1951**, *23*, 69-89.
[68] C.C.J. Roothaan, *Rev. Mod. Phys.*, **1960**, *32*, 179-185.
[69] J.A. Pople, R.K. Nesbet, *J. Chem. Phys.*, **1954**, *22*, 571-572.
[70] G. Berthier, *C.R. Acad. Sci. Paris*, **1954**, *238*, 91-93.
[71] R. Hoffmann, *J. Chem. Phys.*, **1963**, *39*, 1397-1412.
[72] R. Hoffmann, *J. Chem. Phys.*, **1964**, *40*, 2745-2745.
[73] R. Hoffmann, *J. Chem. Phys.*, **1964**, *40*, 2474-2480.
[74] R. Hoffmann, *J. Chem. Phys.*, **1964**, *40*, 2480-2488.
[75] J.A. Pople, D.P. Santry, G.A. Segal, *J. Chem. Phys.*, **1965**, *43*, S129-S135.
[76] J.A. Pople, G.A. Segal, *J. Chem. Phys.*, **1965**, *43*, S136-S151.
[77] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, *J. Am. Chem. Soc.*, **1985**, *107*, 3902-3909.
[78] J.C. Slater, *Phys. Rev.*, **1951**, *81*, 385-390.
[79] P. Hohenberg, W. Kohn, *Phys. Rev.*, **1964**, *136*, B864-B871.
[80] M. Levy, *Proc. Natl. Acad. Sci. USA*, **1979**, *76*, 6062-6065.
[81] W. Kohn, L.J. Sham, *Phys. Rev.* , **1965**, *140*, A1133-A1138.
[82] W. Koch, M.C. Holthausen, *A Chemist's Guide to Density Functional Theory*, Wiley, Weinheim, **2001**.
[83] A.D. Becke, *Phys. Rev. A*, **1988**, *38*, 3098-3100.
[84] J.P. Perdew, *Phys. Rev. B*, **1986**, *33*, 8822-8824.
[85] J.P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.*, **1996**, *77*, 3865-3868.
[86] J.P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.*, **1997**, *78*, 1396.
[87] C. Lee, W. Yang, R.G. Parr, *Phys. Rev. B*, **1988**, *37*, 785-789.
[88] A.D. Becke, *J. Chem. Phys.*, **1993**, *98*, 5648-5652.
[89] C. Adamo, V. Barone, *J. Chem. Phys.*, **1999**, *110*, 6158-6170.
[90] B.G. Johnson, P.M.W. Gill, J.A. Pople, *J. Chem. Phys.*, **1993**, *98*, 5612-5626.
[91] A. Crum Brown, T.R. Fraser, *Trans. Roy.Soc. Edinburgh*, **1868**, *25*, 151-203.
[92] L.P. Hammett, *J. Am. Chem. Soc.*, **1937**, *59*, 96-103.
[93] C. Hansch, *Acc. Chem. Res.*, **1969**, *2*, 232-239.
[94] C. Hansch, A. Leo, R.W. Taft, *Chem. Rev.*, **1991**, *91*, 165-195.
[95] C. Hansch, T. Fujita, *J. Am. Chem. Soc.*, **1964**, *86*, 1616-1626.
[96] S.M. Free, J.W. Wilson, *J. Med. Chem.*, **1964**, *7*, 395-399.

[97] A.R. Katritzky, D.C. Fara, R.O. Petrukhin, D.B. Tatham, U. Maran, A. Lomaka, M. Karelson, *Curr. Top. Med. Chem.*, **2002**, *2*, 1333-1356.
[98] J. Ghasemi, S. Saaidpour, S.D. Brown, *J. Mol. Struct. (Theochem)*, **2007**, *805*, 27-32.
[99] P. Geladi, B.R. Kowalski, *Anal. Chim. Acta*, **1986**, *185*, 1-17.
[100] A.J. Myles, R.N. Feudale, Y. Liu, N.A. Woody, S.D. Brown, *J. Chemom.*, **2004**, *18*, 275-285.
[101] A.F. Duprat, T. Huynh, G. Dreyfus, *J. Chem. Inf. Comput. Sci.*, **1998**, *38*, 586-594.
[102] I.V. Tetko, A.E.P. Villa, D.J. Livingstone, *J. Chem. Inf. Comput. Sci.*, **1996**, *36*, 794-803.
[103] J. Gasteiger, J. Zupan, *Angew. Chem. Int. Ed. Engl.*, **1993**, *32*, 503-527.
[104] R. Leardi, *J. Chemometr.*, **2001**, *15*, 559-569.
[105] M. Karelson, *Molecular Descriptors in QSAR/QSPR*, Wiley, New York, **2000**.
[106] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley, Weinheim, **2000**.
[107] M. Karelson, V.S. Lobanov, A.R. Katritzky, *Chem. Rev.*, **1996**, *96*, 1027-1044.
[108] S.W. Benson, J.H. Buss, *J. Chem. Phys.*, **1958**, *29*, 546-572.
[109] H. Wiener, *J. Am. Chem. Soc.*, **2002**, *69*, 17-20.
[110] M. Randic, *J. Am. Chem. Soc.*, **2002**, *97*, 6609-6615.
[111] L.B. Kier, L.H. Hall, *J. Pharm. Sci.*, **1981**, *70*, 583-589.
[112] A.T. Balaban, *Chem. Phys. Lett.*, **1982**, *89*, 399-404.
[113] S.C. Basak, D.K. Harriss, V.R. Magnuson, *J. Pharm. Sci.*, **1984**, *73*, 429-437.
[114] G. Fayet, P. Raybaud, H. Toulhoat, T. de Bruin, *J. Mol. Struct. (Theochem)*, **2009**, *903*, 100-107.
[115] P. Geerlings, F. De Proft, W. Langenaeker, *Chem. Rev.*, **2003**, *103*, 1793-1874.
[116] H. Chermette, *J. Comput. Chem.*, **1999**, *20*, 129-154.
[117] R.G. Pearson, *J. Am. Chem. Soc.*, **2002**, *85*, 3533-3539.
[118] C.A. Caro, J.H. Zagal, F. Bedioui, C. Adamo, G.I. Cardenas-Jiron, *J. Phys. Chem. A*, **2004**, *108*, 6045-6051.
[119] G.I. Cardenas-Jiron, S. Gutierrez-Oliva, J. Melin, A. Toro-Labbe, *J. Phys. Chem. A*, **1997**, *101*, 4621-4627.
[120] P.K. Chattaraj, P. Perez, J. Zevallos, A. Toro-Labbe, *J. Phys. Chem. A*, **2001**, *105*, 4272-4283.
[121] R.G. Parr, R.A. Donnelly, M. Levy, W.E. Palke, *J. Chem. Phys.*, **1978**, *68*, 3801-3807.
[122] R.S. Mulliken, *J. Chem. Phys.*, **1934**, *2*, 782-793.
[123] R.T. Sanderson, *Science*, **1951**, *114*, 670-672.
[124] R.G. Parr, R.G. Pearson, *J. Am. Chem. Soc.*, **1983**, *105*, 7512-7516.
[125] W. Yang, R.G. Parr, *Proc. Natl. Acad. Sci. U.S.A.*, **1985**, *82*, 6723-6726.
[126] R.G. Pearson, *J. Am. Chem. Soc.*, **1963**, *85*, 3533-3539.
[127] R.G. Parr, L.v. Szentpaly, S. Liu, *J. Am. Chem. Soc.*, **1999**, *121*, 1922-1924.
[128] T. Koopmans, *Physica*, **1934**, *1*, 104-113.
[129] R.G. Parr, W. Yang, *J. Am. Chem. Soc.*, **1984**, *106*, 4049-4050.
[130] J. Gasteiger, *Tetrahedron Lett.*, **1978**, *34*, 3181-3184.
[131] J. Gasteiger, M. Marsili, *Tetrahedron*, **1980**, *36*, 3219-3228.
[132] R.S. Mulliken, *J. Chem. Phys.*, **1955**, *23*, 1833-1840.
[133] A.E. Reed, F. Weinhold, *J. Chem. Phys.*, **1985**, *83*, 1736-1740.
[134] A.E. Reed, L.A. Curtiss, F. Weinhold, *Chem. Rev.*, **1988**, *88*, 899-926.
[135] M. Lejeune, *Statistiques : la théorie et ses applications*, Springer-Verlag, Paris, **2004**.
[136] A.-L. Boulesteix, K. Strimmer, *Brief. Bioinform.*, **2006**, *8*, 32-44.
[137] *User's Guide to SIMCA-P, SIMCA-P+, Version 11.0*, Umetrics AB, 2005.
[138] J.P. Gauchi, *Rev. Stat. Appl.*, **1995**, *43*, 65-89.
[139] D.P. Mesquita, O. Dias, A.M.A. Dias, A.L. Amaral, E.C. Ferreira, *Anal. Chim. Acta*, **2009**, *642*, 94-101.
[140] M.J. Crawley, *Statistics : an introduction using R*, Wiley, Chichester, UK, **2005**.
[141] *CodessaPro*, University of Florida, 2002.
[142] I.H. Witten, E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Publishers, San Francisco, **2005**.

[143] A. Golbraikh, A. Tropsha, *J. Mol. Graph. Model.*, **2002**, *20*, 269-276.
[144] A. Tropsha, P. Gramatica, K.V. Gombar, *QSAR Comb. Sci.*, **2003**, *22*, 69-77.
[145] J. Tunkel, K. Mayo, C. Austin, A. Hickerson, P. Howard, *Environ. Sci. Techn.*, **2005**, *39*, 2188-2199.
[146] J. Jaworska, N. Nikolova-Jeliazkova, T. Aldenberg, *Altern. Lab. Anim.*, **2005**, *33*, 445-459.
[147] A. Bender, R.C. Glen, *Org. Biomol. Chem.*, **2004**, *2*, 3204-3218.
[148] N. Nikolova, J. Jaworska, *QSAR Comb. Sci.*, **2003**, *22*, 1006-1026.
[149] D.T. Stanton, *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, 1423-1433.
[150] A.R. Katritzky, R. Petrukhin, D. Tatham, S. Basak, E. Benfenati, M. Karelson, U. Maran, *J. Chem. Inf. Model.*, **2001**, *41*, 679-685.
[151] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery, Jr., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, 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, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople, *Gaussian03*, Gaussian Inc., Wallington CT, 2004.
[152] *SIMCA P+, Version 12.0*, Umetrics AB, Umeå, Sweden, 2009.
[153] M. Hall, E. Franck, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Written, *SIGKDD Explor.*, **2009**, *11*, 10-18.
[154] J. Jaworska, N. Jeliazkova, *Ambit Discovery* 2007.
[155] D.E.G. Jones, R.A. Augsten, *Thermochim. Acta*, **1996**, *286*, 355-373.
[156] S. Chervin, G.I. Bodman, *Process Saf. Prog.*, **2003**, *22*, 241-243.
[157] T. Ando, Y. Fujimoto, S. Morisaki, *J. Hazard. Mater.*, **1991**, *28*, 251-280.
[158] Y.S. Duh, C. Lee, C.C. Hsu, D.R. Hwang, C.S. Kao, *J. Hazard. Mater.*, **1997**, *53*, 183-194.
[159] L.J.O. Figueiredo, F.M.S. Garrido, *J. Mol. Struct. (THEOCHEM)*, **2001**, *539*, 75-81.
[160] X. Yu, Z. Xie, B. Yi, X. Wang, F. Liu, *Eur. Polym. J.*, **2007**, *43*, 818-823.
[161] D. Ajloo, A. Sharifian, H. Behniafar, *Bull. Korean Chem. Soc.*, **2008**, *29*, 2009-2016.
[162] M.C. Kroon, W. Buijs, C.J. Peters, G.-J. Witkamp, *Thermochim. Acta*, **2007**, *465*, 40-47.
[163] S.W. Benson, *Thermochemical kinetics*, Wiley, New York, **1976**.
[164] E. Theerlynck, D. Mathieu, P. Simonetti, *Thermochim. Acta*, **2005**, *426*, 123-129.
[165] S.R. Saraf, W.J. Rogers, M.S. Mannan, *Ind. Eng. Chem. Res.*, **2003**, *42*, 1341-1346.
[166] M.H. Keshavarz, H.R. Pouretedal, A. Shokrolahi, A. Zali, A. Semnani, *J. Hazard. Mater.*, **2008**, *160*, 142-147.
[167] M.H. Keshavarz, *J. Hazard. Mater.*, **2009**, *162*, 1557-1562.
[168] P. Politzer, J.S. Murray, *J. Molec. Struct.*, **1996**, *376*, 419-424.
[169] B.M. Rice, S. Sahu, F.J. Owens, *J. Mol. Struct. (THEOCHEM)*, **2002**, *583*, 69-72.
[170] R. Vijayaraj, V. Subramanian, P.K. Chattaraj, *J. Chem. Theory Comput.*, **2009**, *5*, 2744-2753.
[171] P. Sarmah, R. Deka, *J. Comput. Aided Mol. Des.*, **2009**, *23*, 343-354.
[172] G. Fayet, L. Joubert, P. Rotureau, C. Adamo, *J. Phys. Chem. A*, **2008**, *112*, 4054-4059.
[173] G. Fayet, L. Joubert, P. Rotureau, C. Adamo, *J. Phys. Chem. A*, **2009**, *113*, 13621-13627.
[174] J.S. Murray, P. Lane, P. Politzer, P.R. Bolduc, *Chem. Phys. Lett.*, **1990**, *168*, 135-139.
[175] E.S. Shanley, G.A. Melhem, *Process Saf. Prog.*, **1995**, *14*, 29-31.
[176] T. Grewer, *Thermal Hazards of Chemical Reactions*, Elsevier, Amsterdam, **1994**.
[177] A.R. Katritzky, R. Petrukhin, R. Jain, M. Karelson, *J. Chem. Inf. Comput. Sci.*, **2001**, *41*, 1521-1530.

[178] A.R. Katritzky, I.B. Stoyanova-Slavova, D.A. Dobchev, M. Karelson, *J. Mol. Graph. Model.*, **2007**, *26*, 529-536.
[179] A.R. Katritzky, S.H. Slavov, D.A. Dobchev, M. Karelson, *Comput. Chem. Eng.*, **2007**, *31*, 1123-1130.
[180] A.R. Katritzky, L. Pacureanu, D. Dobchev, M. Karelson, *J. Chem. Inf. Model.*, **2007**, *47*, 782-793.
[181] M.H. Fatemi, M. Haghdadi, *J. Mol. Struct.*, **2008**, *886*, 43-50.
[182] T. Puzyn, J. Falandysz, *J. Phys. Chem. Ref. Data*, **2007**, *36*, 203-214.
[183] X.-Y. Han, Z.-Y.W.Z.-C.Z.L.-S. Wang, *QSAR Comb. Sci.*, **2006**, *25*, 333-341.
[184] W. Zhou, Z. Zhai, Z. Wang, L. Wang, *J. Mol. Struct. (Theochem)*, **2005**, *755*, 137-145.
[185] G.Y. Yang, J. Yu, Z.Y. Wang, X.L. Zeng, X.H. Ju, *QSAR Comb. Sci.*, **2007**, *26*, 352-357.
[186] G. Yang, X. Zhang, Z. Wang, H. Liu, X. Ju, *J. Mol. Struct. (Theochem)*, **2006**, *766*, 25-33.
[187] X.-Y. Wei, Z.-G. Ge, Z.-Y. Wang, J. Xu, *Chinese J. Struct. Chem.*, **2007**, *26*, 519-528.
[188] T. Puzyn, N. Suzuki, M. Haranczyk, J. Rak, *J. Chem. Inf. Model.*, **2008**, *48*, 1174-1180.
[189] D. Young, T. Martin, R. Venkatapathy, P. Harten, *QSAR Comb. Sci.*, **2008**, *27*, 1337-1345.
[190] S. Sild, M. Karelson, *J. Chem. Inf. Model.*, **2002**, *42*, 360-367.
[191] S. Yin, Z. Shuai, Y. Wang, *J. Chem. Inf. Model.*, **2003**, *43*, 970-977.
[192] C.P. Storm, J.R. Stine, Sensitivity Relationships in Energetic Materials, dans: Chemistry and Physics of Energetic Materials, S.N. Bulusu, Kluwer Academic Publishers, Netherlands, 1990, p. 605-630.
[193] Directive 93/15/CEE du Conseil du 5/04/93 relative à l'harmonisation des dispositions concernant la mise sur le marché et le contrôle des explosifs à usage civil.
[194] *Electricité statique en pyrotechnie, Synthèse de l'état des connaissances*, Groupe de travail de pyrotechnie, 2000.
[195] *Matériaux énergétiques de défense - Sécurité, vulnérabilité - Sensibilité à l'étincelle électrique - Appareil GEMO*, Norme NF T70-539, Décembre 2009.
[196] *Matériaux énergétiques de défense - Sécurité, vulnérabilité - Sensibilité à l'étincelle électrique - Appareil SNPE*, Norme NF T70-540, Mars 2009.
[197] V. Zeman, J. Koci, S. Zeman, *HanNeng CaiLiao*, **1998**, *6*, 118-122.
[198] T.E. Larson, P. Dimas, C.E. Hannaford, *Electrostatic Sensitivity Testing of Explosives at Los Alamos*, Inst. Phys. Conf. Ser. (Electrostatics '91), IOP Publishing Ltd., 1991, p. 107-117.
[199] S. Amari, F. Hosoya, Y. Mizushima, T. Yoshida, *Electrostatic Spark Ignitability of Energetic Materials*, Proc. of the 21st Int. Pyrotech. Seminar, Moscou, 1995.
[200] M. Roux, A. Trevino, M. Auzanneau, C. Brassy, *Sensibilité des substances explosives : Etude de la sensibillité électro-statique d'explosifs polynitrés aromatiques*, Proc. of the 16th Annual Conf. ICT, Karlsruhe, 1985.
[201] F. Hosoya, K. Shiino, K. Itabashi, *Propel. Explos. Pyrotechn.*, **1991**, *16*, 119-122.
[202] M. Roux, M. Auzanneau, C. Brassy, *Propel. Explos. Pyrotechn.*, **1993**, *18*, 317-324.
[203] S. Zeman, V. Pelikan, J. Majzlik, *Cent. Eur. J. Energ. Mater.*, **2006**, *3*, 27-44.
[204] S. Zeman, V. Pelikan, J. Majzlik, *Cent. Eur. J. Energ. Mater.*, **2006**, *3*, 45-51.
[205] M. Auzanneau, M. Roux, *Propel. Explos. Pyrotechn.*, **1995**, *20*, 96-101.
[206] V. Zeman, J. Koci, S. Zeman, *HanNeng CaiLiao*, **1999**, *7*, 127-132.
[207] S. Zeman, J. Koci, *HanNeng CaiLiao*, **2000**, *8*, 18-26.
[208] D. Skinner, D.O.A. Block-Bolten, *Propel. Explos. Pyrotechn.*, **1998**, *23*, 34-42.
[209] M.M. Chaudry, F.A. Al-Ramadhan, I.U. Haq, *Dielectric breakdown and its influence on ignition*, Defense Technical Information Center, 1993.
[210] S. Zeman, A study of chemical micro-mechanisms of initiation of organic polynitro compounds, dans: Energetic Materials. Part 2: Detonation, Combustion, P. Politzer, J.S. Murray, Elsevier, Amsterdam, 2003, p. 25-52.
[211] V. Zeman, J. Koci, S. Zeman, *HanNeng CaiLiao*, **1999**, *7*, 172-175.
[212] G.-X. Wang, H.-M. Xiao, X.-J. Xu, X.-H. Ju, *Propel. Explos. Pyrotechn.*, **2006**, *31*, 102-109.
[213] G. Wang, H. Xiao, X. Ju, X. Gong, *Propel. Explos. Pyrotechn.*, **2006**, *31*, 361-368.

[214] M.H. Keshavarz, H.R. Pouretedal, A. Semnani, *J. Hazard. Mater.*, **2009**, *167*, 461-466.
[215] L. Türker, *J. Hazard. Mater.*, **2009**, *169*, 454-459.
[216] M.H. Keshavarz, *J. Hazard. Mater.*, **2008**, *153*, 201-206.
[217] M.H. Keshavarz, M.H. Moghadas, M.K. Terani, *Propel. Explos. Pyrotechn.*, **2009**, *34*, 136-141.
[218] J.J. Dick, *Appl. Phys. Lett.*, **1984**, *44*, 859-861.
[219] K.L. McNesby, C.S. Coffey, *J. Phys. Chem. B*, **1997**, *101*, 3097-3104.
[220] S. Zeman, *Propel. Explos. Pyrotechn.*, **2000**, *25*, 66-74.
[221] S. Zeman, M. Krupka, *Propel. Explos. Pyrotechn.*, **2003**, *28*, 249-255.
[222] S. Zeman, M. Krupka, *Propel. Explos. Pyrotechn.*, **2003**, *28*, 301-307.
[223] S. Zeman, *Propel. Explos. Pyrotechn.*, **2003**, *28*, 308-313.
[224] M.J. Kamlet, H.G. Adolph, *Propel. Explos.*, **1979**, *4*, 30-34.
[225] C. Brassy, M. Roux, M. Auzanneau, *Propel. Explos. Pyrotechn.*, **1987**, *12*, 53-59.
[226] C. Cao, S. Gao, *J. Phys. Chem. B*, **2007**, *111*, 12399-12402.
[227] J. Mullay, *Propel. Explos. Pyrotechn.*, **1987**, *12*, 60-63.
[228] M. Vaullerin, A. Espagnacq, L. Morin-Allory, *Propel. Explos. Pyrotechn.*, **1998**, *23*, 237-239.
[229] J. Sharma, B.C. Beard, M. Chaykovsky, *J. Phys. Chem.*, **1991**, *95*, 1209-1213.
[230] Y. Kohno, K. Maekawa, T. Tsuchioka, T. Hashizume, A. Imamura, *Chem. Phys. Lett.*, **1993**, *214*, 603-608.
[231] X.-J. Xu, W.-H. Zhu, H.-M. Xiao, *J. Phys. Chem. B*, **2007**, *111*, 2090-2097.
[232] F.J. Owens, *J. Mol. Struct. (THEOCHEM)*, **1985**, *121*, 213-220.
[233] F. Jianfen, X. Heming, *J. Mol. Struct. (THEOCHEM)*, **1996**, *365*, 225-229.
[234] C. Meredith, T.P. Russell, R.C. Mowrey, J.R. McDonald, *J. Phys. Chem. A*, **1998**, *102*, 471-477.
[235] X.J. Xu, H.M. Xiao, X.D. Gong, X.H. Ju, Z.X. Chen, *J. Phys. Chem. A*, **2005**, *109*, 11268-11274.
[236] A. Delpuech, J. Cherville, *Propel. Explos.*, **1979**, *4*, 61-65.
[237] A. Delpuech, J. Cherville, *Propel. Explos.*, **1979**, *4*, 121-128.
[238] A. Delpuech, *J. Phys. Col. C4*, **1987**, *48*, 353-365.
[239] J. Edwards, C. Eybl, B. Johnson, *Int. J. Quantum Chem.*, **2004**, *100*, 713-719.
[240] C. Zhang, Y. Shu, Y. Huang, X. Zhao, H. Dong, *J. Phys. Chem. B*, **2005**, *109*, 8978-8982.
[241] X.-S. Song, X.-L. Cheng, X.-D. Yang, B. He, *Propel. Explos. Pyrotechn.*, **2006**, *31*, 306-310.
[242] N.R. Badders, C. Wei, A.A. Aldeeb, W.J. Rogers, M.S. Mannan, *J. Energ. Mater.*, **2006**, *24*, 17 - 33.
[243] M.H. Keshavarz, H.R. Pouretedal, *J. Hazard. Mater.*, **2005**, *124*, 27-33.
[244] M.H. Keshavarz, H.R. Pouretedal, A. Semnani, *J. Hazard. Mater.*, **2007**, *141*, 803-807.
[245] M.H. Keshavarz, *J. Hazard. Mater.*, **2007**, *148*, 648-652.
[246] S.G. Cho, K.T. No, E.M. Goh, J.K. Kim, J.H. Shin, Y.D. Joo, S. Seong, *Bull. Korean Chem. Soc.*, **2005**, *26*, 399-408.
[247] R. Wang, J. Jiang, Y. Pan, H. Cao, Y. Cui, *J. Hazard. Mater.*, **2009**, *166*, 155-186.
[248] R. Cohen, Y. Zeiri, E. Wurzberg, R. Kosloff, *J. Phys. Chem. A*, **2007**, *111*, 11074-11083.
[249] J.L. Gustin, *Org. Process Res. Dev.*, **1998**, *2*, 27-33.
[250] S. Xu, M.C. Lin, *J. Phys. Chem. B*, **2005**, *109*, 8367-8373.
[251] S.C. Chen, S.C. Xu, E. Diau, M.C. Lin, *J. Phys. Chem. A*, **2006**, *110*, 10130-10134.
[252] A. Gindulyte, L. Massa, L. Huang, J. Karle, *J. Phys. Chem. A*, **1999**, *103*, 11040-11044.
[253] A. Gindulyte, L. Massa, L. Huang, J. Karle, *J. Phys. Chem. A*, **1999**, *103*, 11045-11051.
[254] M.-H. Liu, C.C.Y.-S. Hong, *Int. J. Quantum Chem.*, **2005**, *102*, 398-408.
[255] F.J. Owens, *J. Mol. Struct. (THEOCHEM)*, **1996**, *370*, 11-16.
[256] R.J.A. Kersten, M.N. Boers, M.M. Stork, C. Visser, *J. Loss Prev. Process Ind.*, **2005**, *18*, 145-151.
[257] G. Tanaka, C. Weatherford, *Int. J. Quantum Chem.*, **2008**, *108*, 2924-2934.
[258] V. Korolev, T. Petukhova, T. Pivina, A. Porollo, A. Sheremetev, K. Suponitskii, V. Ivshin, *Russ. Chem. Bull. Int. Ed.*, **2006**, *55*, 1388-1410.
[259] S.R. Saraf, W.J. Rogers, D.M. Ford, M.S. Mannan, *Fluid Phase Equil.*, **2004**, *222-223*, 205-211.
[260] G. Fayet, L. Joubert, P. Rotureau, C. Adamo, *Chem. Phys. Lett.*, **2009**, *467*, 407-411.
[261] G. Fayet, P. Rotureau, L. Joubert, C. Adamo, *J. Hazard. Mater.*, **2009**, *171*, 845-850.

[262] A.C. Gonzalez, C.W. Larson, D.F. McMillen, D.M. Golden, *J. Phys. Chem.*, **1985**, *89*, 4809-4814.
[263] W. Tsang, D. Robaugh, W.G. Mallard, *J. Phys. Chem.*, **1986**, *90*, 5968-5973.
[264] D.B. Galloway, J.A. Bartz, L.G. Huey, F.F. Crim, *J. Chem. Phys.*, **1993**, *98*, 2107-2114.
[265] D.B. Galloway, T. Glenewinkelmeyer, J.A. Bartz, L.G. Huey, F.F. Crim, *J. Chem. Phys.*, **1994**, *100*, 1946-1952.
[266] C. Kosmidis, K.W.D. Ledingham, H.S. Kilic, T. McCanny, R.P. Singhal, A.J. Langley, W. Shaikh, *J. Phys. Chem. A*, **1997**, *101*, 2264-2270.
[267] Y.-M. Li, J.-L. Sun, H.-M. Yin, K.-L. Han, G.-Z. He, *J. Chem. Phys.*, **2003**, *118*, 6244-6249.
[268] E. McCarthy, K. O'Brien, *J. Org. Chem.*, **1980**, *45*, 2086-2088.
[269] J. Shao, X. Cheng, X. Yang, *Struct. Chem.*, **2006**, *17*, 547-550.
[270] T.B. Brill, K.J. James, R. Chawla, G. Nicol, A. Shukla, J.H. Futrell, *J. Phys. Org. Chem.*, **1999**, *12*, 819-826.
[271] E. Fields, S. Meyerson, *J. Am. Chem. Soc.*, **1967**, *89*, 3224-3228.
[272] Y.Z. He, J.P. Cui, W.G. Mallard, W. Tsang, *J. Am. Chem. Soc.*, **1988**, *110*, 3754-3759.
[273] J.-D. Shao, T. Baer, *Int. J. Mass Spectrom. Ion Proc.* , **1988**, *86*, 357-367.
[274] P. Politzer, J.M. Seminario, P.R. Bolduc, *Chem. Phys. Lett.*, **1989**, *158*, 463-469.
[275] R. Saxon, M. Yoshimine, *J. Phys. Chem.*, **1989**, *93*, 3130-3135.
[276] T. Glenewinkel-Meyer, F.F. Crim, *J. Mol. Struct. (THEOCHEM)*, **1995**, *337*, 209-224.
[277] J.X. Luo, *Handbook of bond dissociation energies in organic compounds*, CRC Press, Boca Raton, FL, **2003**.
[278] F. Berho, F. Caralp, M.-T. Rayez, R. Lesclaux, *J. Phys. Chem. A*, **1998**, *102*, 1-8.
[279] D. Jacquemin, B. Champagne, J.-M. André, *Synth. Met.*, **1996**, *80*, 205-210.
[280] P. Perez, Y. Simon-Manso, A. Aizman, P. Fuentealba, R. Contreras, *J. Am. Chem. Soc.*, **2000**, *122*, 4756-4762.
[281] K.J. Palat, S. Böhm, G. Braunerova, K. Waisser, O. Exner, *New J. Chem.*, **2002**, *26*, 861-866.

ANNEXES

Sommaire

ANNEXE I : Procédure de classification des substances, mélanges et objets explosibles dans le Manuel d'épreuves et de critères relatif à la réglementation du Transport des Marchandises Dangereuses (5ème édition révisée, 2009)..... 195
ANNEXE II : Liste des publications personnelles..... 199

ANNEXE I : Procédure de classification des substances, mélanges et objets explosibles dans le Manuel d'épreuves et de critères relatif à la réglementation du Transport des Marchandises Dangereuses (5ème édition révisée, 2009)

Figure 2.1.1

Diagramme d'ensemble de la procédure de classification d'une substance, d'un mélange ou d'un objet dans la classe des explosibles (Classe 1 pour le transport)

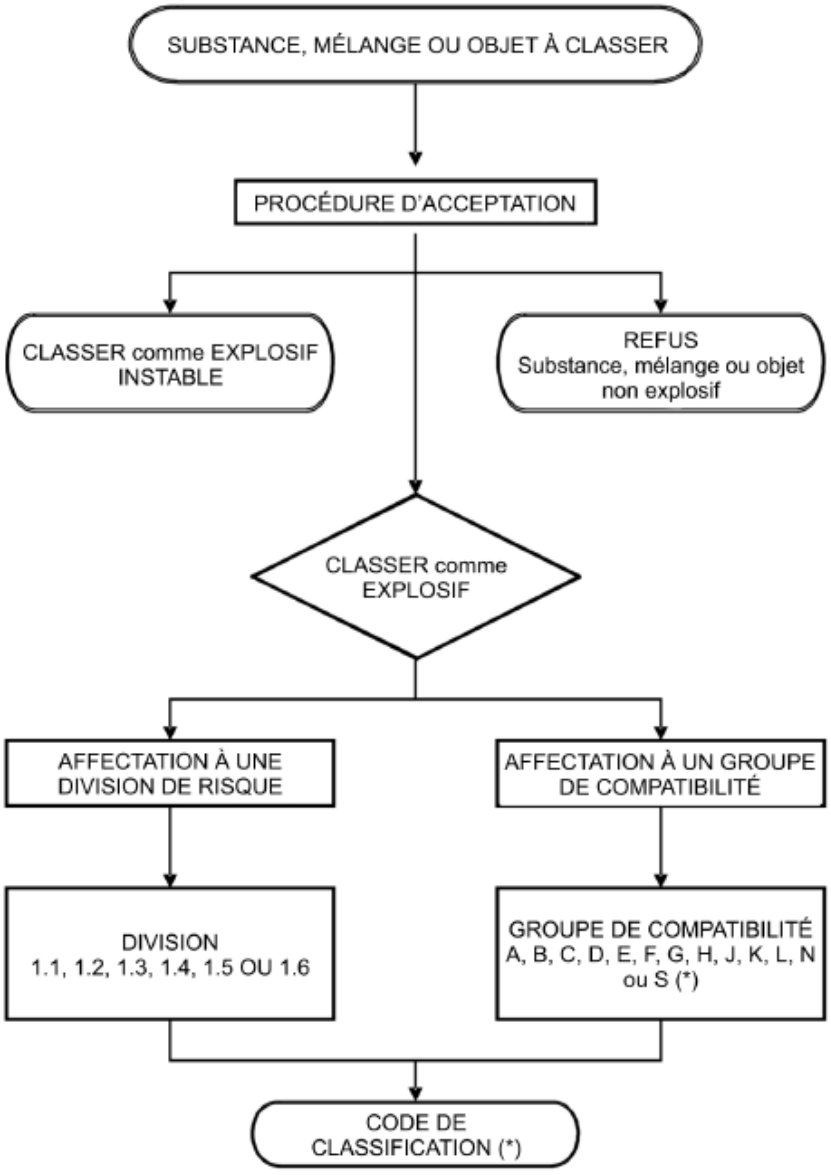
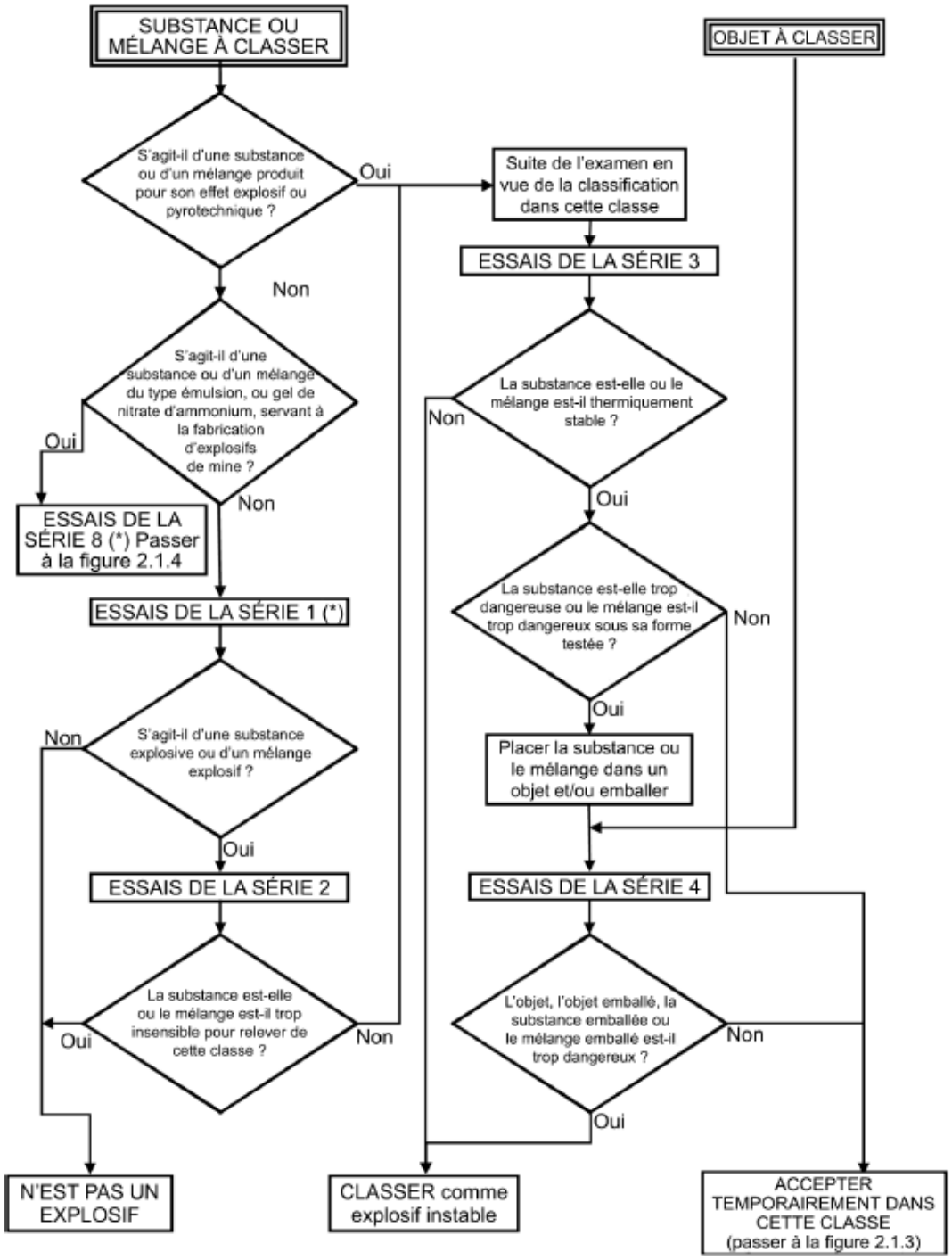


Figure 2.1.2

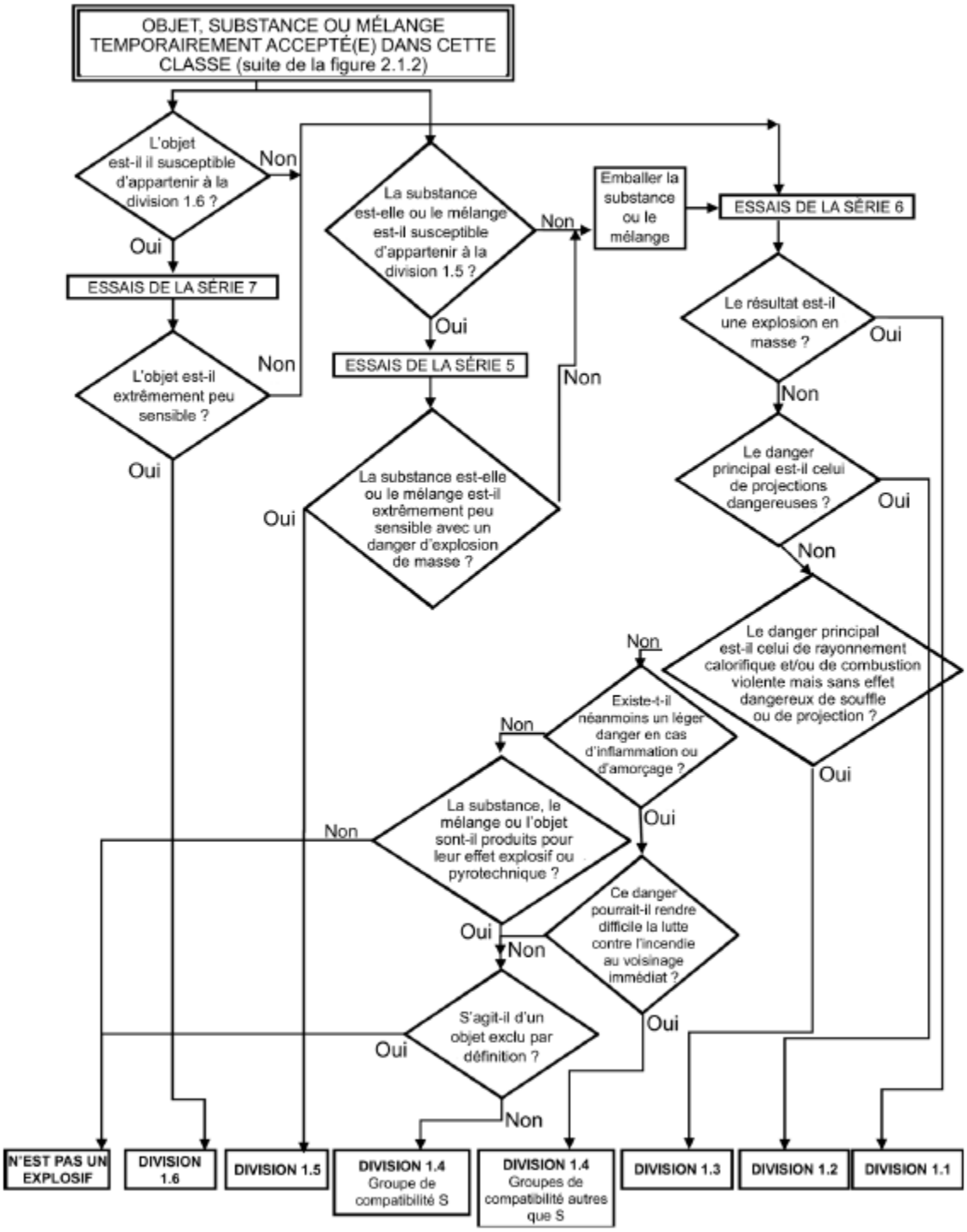
Procédure d'acceptation temporaire d'une substance, d'un mélange ou d'un objet dans la classe des explosibles (Classe 1 pour le transport)



(*) Aux fins de la classification, il convient de commencer par les essais de la série 2.

Figure 2.1.3

Procédure d'affectation à une division de la classe des explosibles (Classe 1 pour le transport)



ANNEXE II : Liste des publications issues de ces travaux de thèse

1. **G. Fayet**, L. Joubert, P. Rotureau, C. Adamo
   Theoretical study of the decomposition reactions in substituted nitrobenzenes
   Journal of Physical Chemistry A*,* **2008***, 112*, 4054-4059
2. **G. Fayet**, L. Joubert, P. Rotureau, C. Adamo
   On the use of descriptors arising from the conceptual density functional theory for the prediction of chemicals explosibility
   Chemical Physics Letter, **2009***, 467*, 407-411
3. **G. Fayet**, P. Rotureau, L. Joubert, C. Adamo
   On the prediction of thermal stability of nitroaromatic compounds using quantum chemical calculations
   Journal of Hazardous Materials, **2009**, *171*, 845-850
4. **G. Fayet**, P. Rotureau, L. Joubert, C. Adamo
   A theoretical study of the decomposition mechanisms in substituted o-nitrotoluenes
   Journal of Physical Chemistry A, **2009**, *113*, 13621-13627
5. P. Rotureau, **G. Fayet**, G. Marlair, C. Michot, L. Joubert, C. Adamo
   Evaluer les risques d'explosion des substances chimiques – Des approches expérimentales classiques à la prédiction par la chimie quantique et les méthodes statistiques QSPR
   Actualité Chimique, **2010**, *337*, 51-58
6. **G. Fayet**, P. Rotureau, L. Joubert, C. Adamo
   QSPR modeling of thermal stability of nitroaromatic compounds: DFT vs. AM1 calculated descriptors
   Journal of Molecular Modelling, sous presse
7. **G. Fayet**, P. Rotureau, L. Joubert, C. Adamo
   Predicting explosibility properties of chemicals from quantitative structure-property relationships
   Process Safety Progress, sous presse
8. **G. Fayet**, P. Rotureau, L. Joubert, C. Adamo
   Development of QSPR models on thermal stability of nitroaromatic compounds considering their decomposition mechanisms
   Manuscrit en préparation

RESUME

L'objectif de ces travaux était de développer et d'évaluer des modèles quantitatifs structure-propriété (QSPR) pour la prédiction des propriétés explosives des composés nitroaromatiques, en vue d'une utilisation dans un cadre règlementaire, en particulier celui du nouveau règlement européen REACH.

Différentes approches méthodologiques (régressions multi-linéaires, PCA, PLS, arbres de décision) ont été utilisées pour mettre en place des modèles pour la prédiction de la chaleur de décomposition. Les descripteurs des modèles ont été sélectionnés dans un jeu étendu de plus de 300 descripteurs (constitutionnels, topologiques, géométriques et quantiques). Deux premiers modèles avec des domaines d'applicabilité définis et des pouvoirs prédictifs importants ont été obtenus.

Des modèles pour trois autres propriétés explosives (la température de décomposition, les sensibilités à la décharge électrique et à l'impact) ont ensuite été développés, avec des performances similaires voire supérieures aux modèles existants.

Enfin, l'analyse des mécanismes réactionnels sous-jacents, menée à l'aide de la DFT, a permis de mettre en évidence la présence de chemins de décomposition spécifiques au sein des composés nitroaromatiques et a ainsi complété l'approche QSPR en termes d'interprétation phénoménologique.

Cette étude a donc pris en compte l'intégralité des principes mis en place par l'OCDE pour la validation des modèles QSAR/QSPR dans un usage règlementaire (cible expérimentale, structure du modèle, validation, domaine d'applicabilité et interprétation des mécanismes sous-jacents). Deux modèles prédictifs ont même été développés pour la chaleur de décomposition des composés nitroaromatiques.

ABSTRACT

The aim of these works consisted in the development and evaluation of quantitative structure property models (QSPR) for the prediction of explosive properties of nitroaromatic compounds, to be used within a regulatory context, in particular the new European regulation called REACH.

Different methodological approaches (multilinear regressions, PCA, PLS, decision trees) were used to develop models for the prediction of the heat of decomposition. The descriptors of the model are selected among an extended set of more than 300 (constitutional, topological, geometric and quantum chemical) descriptors. Applicability domains were defined and the predictive power was determined using a set of validation.

Three other properties (decomposition temperature, electric spark and impact sensitivities) were also investigated, leading to similar or better performances than existing models.

Then, subjacent mechanisms, analyzed from DFT calculations, brought to light specific decomposition paths for any nitroaromatic compounds. This study completed the QSPR analyses in terms of phenomenological interpretation.

Finally, this study took into account all OECD principles for the validation of QSAR/QSPR models for regulatory uses (experimental endpoint, model structure, validation, applicability domain and interpretation of subjacent mechanisms). Moreover, two predictive models were developed for the heat of decomposition of nitroaromatic compounds.