



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE ABOU-BEKR BELKAID - TLEMCCEN

MEMOIRE

Présenté à

FACULTE DES SCIENCES – DEPARTEMENT DE CHIMIE

Pour l'obtention du diplôme de :

MASTER

Filière: **Chimie**

Option: Chimie Théorique et Modélisation

Par :

M^r MAMMERI Ibrahim

Sur le Thème

PREDICTION DES CONSTANTES D'ACIDITE (pK_a) DES DERIVES DE LA COUMARINE PAR LES METHODES QUANTIQUES. ETUDE QSPR

Soutenu publiquement le 19 Juin 2017 à Tlemcen devant le jury composé de :

M ^{me} CEMOURI Hafida	Maître de Conférences B	ESSA-Tlemcen	Président
M ^{me} ERRAHOUI née BELLIFA Khadidja	Maître de Conférences B	Université de Tlemcen	Examineur
M ^r CHARIF Imad Eddine	Maître de Conférences B	Université de Tlemcen	Encadreur

*Laboratoire de Thermodynamique Appliquée et Modélisation Moléculaire (LATA2M), N° 53
BP 119, 13000 Tlemcen - Algérie*

Dédicace

A mes très chers parents

A tous les membres de ma famille

A tous mes chers ami(e)s

A tous ceux qui me sont chers

...Je dédie ce modeste travail

Remerciement

Avant toute chose je tiens à remercier Dieu pour m'avoir donné la volonté, la force et le courage d'achever le travail de ce mémoire.

Le travail présenté dans ce mémoire a été dirigé par Monsieur I.E. CHARIF Maitre de Conférences B au Département de Chimie Faculté des Sciences, Université A. Belkaïd de Tlemcen et il a été réalisé au Laboratoire de Recherche « Thermodynamique Appliquée et Modélisation Moléculaire LATA2M » domicilié à l'Université A. Belkaïd de Tlemcen

Je tiens prioritairement à remercier très sincèrement monsieur I.E. CHARIF pour l'encadrement de ce mémoire et de m'avoir guidé, ainsi pour ces précieux conseils ses orientations et son aide à la réalisation de ce modeste travail.

Je tiens également à remercier tous les membres du jury pour leur participation à la présentation de ce travail.

J'adresse mes sincères remerciements à madame H. CHEMOURI Maitre de conférences B à l'école ESSA de Tlemcen qui m'a fait l'honneur de présider le jury de ce mémoire.

Je remercie madame K. ERRAHOUÏ née BELLIFA Maitre de Conférences B à l'université A. Belkaïd de Tlemcen pour ses conseils et remarques, son soutien, sa généreuse collaboration et d'avoir accepté d'examiner et évaluer ce travail.

Enfin, un vif remerciement à tous mes enseignants qui ont contribué à cette formation et à mes collègues au laboratoire de recherche « Thermodynamique Appliquée et Modélisation Moléculaire », à tous mes chers amis pour leur soutien moral et motivations.

A tous les personnes qui ont beaucoup contribués à la réalisation de ce modeste travail.

SOMMAIRE

INTRODUCTION GENERALE.....	01
Bibliographie.....	03
CHAPITRE I : METHODES DE CALCULS QUANTO-CHIMIQUE	
I. Les méthodes de la chimie quantique.....	05
I.1. Equation de Schrödinger.....	05
I.2. Méthode de Hartree-Fock-Roothan.....	06
I.2.1. Approximation du champ moyen de Hartree.....	06
I.2.2. Méthode de Hartree-Fock.....	06
I.2.3. Méthode de Hartree-Fock-Roothaan.....	07
I.3. Théorie de la Fonctionnelle de Densité (DFT).....	08
I.3.1. Fondement de la théorie DFT.....	08
I.3.2. Méthode de Kohn et Sham.....	10
I.3.3. Approximation de la densité locale LDA.....	12
I.3.4. Approximation du Gradient Généralisé (GGA).....	13
I.3.5. Fonctionnelle hybride B3LYP.....	14
I.4. Bases d'orbitales atomiques.....	14
I.5. Les Méthodes de solvation.....	15
I.5.1. Modèle SCRF.....	16
I.5.2. Modèle PCM.....	17
I.5.3. Modèle SMD	17
Bibliographie du chapitre I.....	19
CHAPITRE II : MODELISATION QSPR ET ANALYSE DE DONNEES.....	
I. Principe des Méthodes QSPR	20
I.1. Méthodes d'analyse de données.....	22
I.2.1. La régression linéaire simple	22
I.2.2. . La régression linéaire multiple.....	29
I.2.3. Interprétation et validation d'un modèle QSPR.....	30

I.2.3.1. Analyse des paramètres statistiques d'un modèle QSPR.....	30
1. Test de la signification globale de la régression.....	30
I.2.3.2. Validation interne : <i>Y-Randomization</i>	33
I.2. 3. 3. Validation externe.....	34
I.2.3.4. Domaine d'applicabilité.....	35
Bibliographie du chapitre II.....	36

CHAPITRE III : RESULTATS ET DISCUSSIONS

I. Introduction.....	38
II. Méthodologie de calculs.....	40
II.1 Calculs des descripteurs théoriques.....	40
II.2. Résultats et discussions.....	42
II.2.1. La Phase gazeuse.....	42
II.2 .2. La Phase aqueuse	42
II.2.3. Validation externe des modèles.....	48
II.2.4. Domaine d'applicabilité.....	49
Bibliographie du chapitre III.....	51
CONCLUSION GENERALE	53

LISTE DES ABREVIATIONS

B3LYP	Becke 3-Parameter Lee-Yang-Parr
CLOA	Combinaison Linéaire d'Orbitales Atomiques
CGTO	Contracted Gaussian Type Orbital Configuration
DFT	Density Functional Theory
GGA	Generalized Gradient Approximation
GTO	Gaussian Type Orbital
HF	Hartree-Fock
HFR	Hartree-Fock-Roothaan
HOMO	Highest Occupied Molecular Orbital
KS	Kohn et Sham
LDA	Local Density Approximation
LUMO	Lowest Unoccupied Molecular Orbital
MP2	Möller –Plesset d'ordre 2
OA	Orbitale Atomique
OM	Orbitale Moléculaire
PCM	Polarizable Continuum Model
CPCM	Conductor Polarizable Continuum Model
SMD	Solvation Model based on Density
SCF	Self Consistent Field
STO	Slater Type Orbital
SCRf	Self-Consistent Reaction Field
UAHF	United Atom Topological Model applied on radii optimized for the HF/6-31G(d)
UAKS	United Atom Topological Model applied on radii optimized for the PBE1PBE/6-31G(d)
QSPR	Quantitative Structure-Property Relationships
QSAR	Quantitative Structure-Activity Relationships
SLR	Simple Linear Regression
MLR	Multiple Linear Regression
SSR	Sum of Squares Regression
SSE	Sum of Squares Error
SST	Sum of Squares Total
NPA	Natural Population Analysis

INTRODUCTION GENERALE

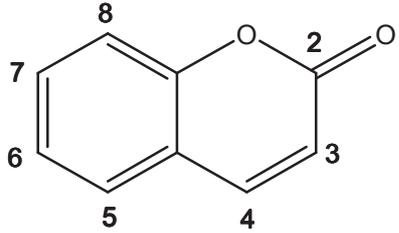
L'acidité ou la constante d'acidité, pK_a , est l'une des propriétés moléculaires physico-chimiques les plus importantes en chimie organique et inorganique. L'acidité gère, par exemple, la solubilité et la modification de cette propriété constitue un facteur crucial pour l'activité de nombreuses molécules biologiquement importantes. La connaissance des valeurs de pK_a nous permet, dans une réaction acide-base, de prédire facilement le sens privilégié de l'équilibre acide-base ainsi que les concentrations des différentes espèces. Par exemple dans les réactions de substitution nucléophile, Les valeurs pK_a permettent de prédire le pouvoir nucléophile et les meilleurs groupes sortants [1].

Il existe plusieurs méthodes théoriques pour calculer ou prédire les propriété/activité biologique des molécules en solution. [2-7]. Les méthodes QSPR/QSAR sont largement utilisées dans la prédiction de la propriété/activité biologique [8-11]. Le principe est de déterminer des équations mathématiques appelés modèles reliant propriété/activité à des paramètres appelées descripteurs liés à la structure de la molécule. Ces modèles obtenus peuvent être utilisés dans la prédiction de la propriété/activité de nouvelles molécules où les données expérimentales sont indisponibles.

Les méthodes QSPR représentent parmi les méthodes les plus efficaces dans la prédiction des constantes d'acidité. Plusieurs travaux ont été publiés sur l'élaboration des modèles pour prédire les valeurs de pK_a des composés organiques [12-17] tels que l'imidazole [18] et les benzimidazoles [19] en utilisant les méthode de régression linéaire et les méthodes de la chimie quantiques PM3, PM6 , DFT...

Les hydroxy-coumarines et leurs dérivés (schéma1) sont des molécules phénoliques connues pour leurs propriétés fluorescentes et leurs activités anticoagulantes [20-22]. Elles existent dans plusieurs plantes à savoir la lavande et le mélilot. Les coumarines existent également dans plusieurs produits naturels tels que le thé vert et le miel. La détermination de l'acidité des coumarines et ses dérivés est une étape importante puis qu'elles exercent une forte activité biologique dans l'action toxique de la vitamine K [23].

Schéma 1

	Composé	R ₆	R ₇	R ₈
	Daphnétine	H	OH	OH
	Esculétine	OH	H	OH
	Scopolétine	OCH ₃	H	OH
	Umbélliférone	H	H	OH

L'objectif de ce travail est l'élaboration des modèles QSPR reliant l'acidité expérimentale en phase aqueuse (pK_a) d'une série de 20 hydroxy-coumarines à un ensemble de descripteurs théoriques. Ces derniers ont été calculés en solution au moyen de la méthode quantique (DFT/ B3LYP/6-311++G**) et en utilisant les méthodes d'analyse de données : la régression linéaire simple SLR (*Simple Linear Regression*) et la régression linéaire multiple MLR (*Multiple Linear Regression*).

Ce manuscrit est présenté en trois chapitres :

- Dans le chapitre 1, nous présentons les méthodes de la chimie quantique (méthode HFR, méthodes DFT) et les modèles de solvation PCM et SMD.
- Dans le chapitre 2, nous présentons les méthodes statistiques, l'analyse de données et la modélisation QSPR.
- Le dernier chapitre est consacré pour l'interprétation des résultats obtenus de la prédiction des constantes d'acidité (pK_a) des dérivés des hydroxy-coumarines par la modélisation QSPR en utilisant les méthodes quato-chimiques.

Bibliographie

- [1] R.D. Rossi, *J. Chem. Educ.* 90 (2013), 183
- [2] Alongi, K. S.; Shields, G. C., Theoretical calculations of acid dissociation constants: a review article. In *Annual Reports in Computational Chemistry*, Wheeler, R. A., Ed. Elsevier: 2010; Vol. 6, pp 113-138.
- [3] Tomasi, J.; Mennucci, B.; Cammi, R., *Chem. Rev.* 105, (2005) 2999
- [4] Cramer, C. J.; Truhlar, D. G., *Acc. Chem. Res.* 41 (2008) 760.
- [5] Lee, A. C.; Crippen, G. M., *J. Chem. Inf. Model.* 49 (2009) 2013.
- [6] Ho, J.; Coote, M. L., *Theor. Chem. Acc.*, 125 (2010) 3.
- [7] Ho, J.; Coote, M. L., *WIREs Comp. Molec. Sci.*, 1 (2011) 649.
- [8] D.A. Winkler, *Brief. Bioinf.*, 3 (2002) 73.
- [9] C.D. Selassie, R. Garg, S. Kapur, A. Kurup, R.P. Verma, S.B. Mekapati, C. Hansch, *Chem. Rev.*, 102 (2002) 2585.
- [10] M. Grover, B. Singh, M. Bakshi, S. Singh, *Pharm. Sci. Tech. Today*, 3 (2000) 50.
- [11] J. Taskinen, J. Yliruusi, *Adv. Drug Deliv. Rev.*, 55 (2003) 1163.
- [12] V. Consonni, R. Todeschini, M. Pavan, P. Gramatica, *J. Chem. Inf. Comput. Sci.* 42 (2002) 693.
- [13] G. Krenkel, E. A. Castro, A. A. Toropov, *J. Mol. Struct. (THEOCHEM)* 542 (2001) 107.
- [14] L. B. Kier, L. H. Hall, "Molecular Connectivity in Structure–Activity Analysis", RSP-Wiley, Chetster, UK, 1986.
- [15] J. Ghasemi, S. Shahmirani, E. V. Farahani, *Ann. Chim.* 96 (2006) 327.
- [16] J. Ghasemi, S. Shahmirani, E. V. Farahani, *Ann. Chim.* 96 (2006) 327.
- [17] J. Ghasemi, S. Saaidpour, S.D. Brown, *J. Mol. Struct. (THEOCHEM)* 805 (2007) 27.
- [18] E. Soriano, S. Cerdán, P. Ballesteros, *J. Mol. Struct. (THEOCHEM)* 684 (2004) 121.
- [19] T. N. Brown, N. Mora-Diez, *J. Phys. Chem. B* 110 (2006) 20546.
- [20] Adamowicz, P.; Kała, M., *Probl. Forensic Sci.*, 64 (2005) 373.
- [21] Jain, P. K.; Joshi, H., *J. Appl. Pharm. Sci.* 2 (2012) 236.

- [22] Abdou, M. M.; El-Saeed, R. A.; Bondock, S., *Arabian J. Chem.* 2015, 1.
- [23] J.W. Suttie, G.L. Nelsestuen, *Crit. Rev. Biochem.* 8 (1980) 191–223.

CHAPITRE I :
METHODES DE LA CHIMIE QUANTIQUE

I. Les méthodes de la chimie quantique:

I.1. Equation de Schrödinger:

Les méthodes de la chimie quantique sont toutes basées sur l'équation connue sous le nom d'équation de Schrödinger indépendante du temps [1]. Cette équation permet de décrire un système moléculaire[1]:

$$H \Psi = E \Psi \quad (1)$$

H est l'opérateur hamiltonien du système, Ψ la fonction d'onde qui est obtenue en résolvant l'équation différentielle et E est l'énergie du système.

En tenant compte de l'approximation Born et Oppenheimer l'hamiltonien H peut être simplifié [2]:

$$H = -\frac{\hbar^2}{2m_e} \sum_{i=1}^n \Delta_i - \sum_{K=1}^N \sum_{i=1}^n \frac{Z_K e^2}{R_{Ki}} + \sum_{i>j}^n \frac{e^2}{r_{ij}} \quad (2)$$

Avec i et j représentant les électrons, K et L représentant les noyaux, Δ_i l'opérateur laplacien, m_e est la masse d'électron, M_k et Z_k représentent respectivement la masse et le numéro atomique du noyau k et r la distance noyau-électron.

La résolution exacte de l'équation (1) n'est exacte que pour les systèmes mono-électroniques. Pour les systèmes à plusieurs électrons, il est nécessaire de faire appel aux méthodes approchées pour résoudre l'équation de Schrödinger.

La résolution de l'équation de Schrödinger permet de calculer les propriétés telle que : géométries moléculaires, énergies des états de transition, fréquences de vibration, grandeurs thermochimiques, mécanisme réactionnels, les populations électroniques etc.

I-2.Méthode de Hartree-Fock-Roothaan:

I.2.1. Approximation du champ moyen de Hartree:

En 1927, Hartree [3] a proposé l'approximation de remplacer l'interaction d'un électron avec le reste des électrons du système $\sum_j e^2 / r_{ij}$ par un champ moyen engendré par ces électrons définit le potentiel mono-électronique $U(i)$. La fonction d'onde totale s'écrit sous forme d'un produit de fonctions d'onde mono-électroniques:

$$\Psi = \Psi_1(1) \cdot \Psi_2(2) \cdot \Psi_3(3) \dots \Psi_n(n) \quad (3)$$

I.2.2. Méthode de Hartree-Fock:

La méthode de Hartree-Fock résout l'équation de Schrödinger d'une manière approchée pour des systèmes poly-électroniques en utilisant le principe variationnel. La fonction d'onde totale est exprimée sous la forme d'un déterminant de Slater d'un système à couches fermées ($n=2m$) [4]:

$$\Psi(1,2, \dots, n) = \frac{1}{(n!)^{1/2}} |\Phi_1(1)\bar{\Phi}_1(2) \dots \Phi_m(2m-1)\bar{\Phi}_m(2m)| \quad (4)$$

Avec:

$$\Phi_1(1) \equiv \Phi_1(1)\alpha(1) \quad (5)$$

$$\bar{\Phi}_1(2) \equiv \Phi_1(2)\beta(2) \quad (6)$$

Φ Est une orbitale moléculaire mono électronique. α et β sont les fonctions de spin.

L'énergie HF s'écrit sous la forme suivante:

$$E^{HF} = 2 \sum_i h_{ii} + \sum_{j>i} \sum (2J_{ij} - K_{ij}) \quad (7)$$

Avec :

$$h_{ii} = \int \Phi_i^*(1) \hat{h}_1 \Phi_i(1) d\tau_1 \quad (8)$$

Et :

$$\hat{h}_1 = -\frac{1}{2}\Delta_1^2 - \sum_K \frac{Z_K e^2}{r_{1K}} \quad (9)$$

$$J_{ij} = \int \Phi_i^*(1)\Phi_i(1)r_{12}^{-1}\Phi_j^*(2)\Phi_j(2)d\tau_1d\tau_2 \quad (10)$$

$$K_{ij} = \int \Phi_i^*(1)\Phi_j(1)r_{12}^{-1}\Phi_i^*(2)\Phi_j(2)d\tau_1d\tau_2 \quad (11)$$

Les intégrales J_{ij} de Coulomb et K_{ij} d'échange définissent les répulsions électroniques. Les équations Hartree-Fock sont:

$$F\Phi_i = E_i\Phi_i \quad (12)$$

Où: F est l'opérateur de Fock

$$F(1) = \hat{h}_1(1) + V^{eff}(1) = \hat{h}_1(1) + \sum_i \left(2\hat{J}_i(1) - \hat{K}_i(1) \right) \quad (13)$$

E_i sont les énergies des spin-orbitales.

1.2.3. Méthode de Hartree-Fock-Roothaan:

Dans la méthode de Hartree-Fock, les expressions analytiques des orbitales moléculaires (OMs) Φ_i ne sont pas connues. Afin de construire les OmsRoothaan[5] a proposé l'approximation de décrire les OM Φ_i comme une combinaison linéaire d'orbitales atomiques ϕ_μ (OM-CLOA):

$$\Phi_i = \sum_{\mu=1}^N C_{i\mu} \phi_\mu \quad (14)$$

$C_{i\mu}$ sont les coefficients à faire varier. N étant le nombre d'OAs.

En procédant par la méthode des variations et la condition de minimisation, les meilleurs coefficients sont ceux qui minimisent l'énergie. Les équations de Roothaan sont les suivantes suivant (système séculaire) [5]:

$$\sum_{r=1}^N C_{kr} (F_{rs} - \epsilon_k S_{rs}) = 0 \quad s = 1, 2, \dots, N \quad (15)$$

Avec:

$$\left\{ \begin{array}{l} F_{rs} = h_{rs}^c + \sum_{p=1}^n \sum_{q=1}^n P_{pq} \{2\langle rs|pq\rangle - \langle rq|ps\rangle\} \\ S_{rs} = \langle \phi_r | \phi_s \rangle \\ h_{rc}^c = \int \phi_r^*(i) h^c \phi_s(i) d\tau_i \end{array} \right. \quad (16)$$

Où r, s, p et q désignent les OA, P_{pq} est l'élément de la matrice densité. Les termes $\langle rs|pq\rangle$ et $\langle rq|ps\rangle$ représentent les intégrales bi-électroniques coulombiennes et d'échange respectivement. S_{rs} est une intégrale de recouvrement. L'expression de l'énergie électronique s'écrit:

$$E = 2 \sum_i^{N/2} h_{ii} + \sum_{ij}^{N/2} (2J_{ij} - K_{ij}) \quad (17)$$

I.3. Théorie de la fonctionnelle de densité (DFT):

I.3.1. Fondement de la théorie DFT:

La Théorie de la fonctionnelle de la densité est devenue, ces dernières années, un outil théorique important, elle est utilisée pour la description et l'analyse des propriétés physiques et chimiques pour les systèmes complexes, particulièrement à ce contenant un grand nombre d'électrons [6,7].

- **1^{er} théorème de Hohenberg et Kohn:**

«L'énergie moléculaire, la fonction d'onde et toutes les autres propriétés électroniques de l'état fondamental sont déterminées à partir de la densité électronique de l'état fondamental $\rho_0(x,y,z)$ ». [8]

Pour un système pol-yélectronique l'expression de l'Hamiltonien total :

$$H = -\frac{1}{2} \sum_i^n \Delta_i + \sum_{i>j}^n \frac{1}{r_{ij}} + \sum_i^n v(r_i) \quad (18)$$

Avec:

$$v(r_i) = - \sum_{\alpha} \frac{Z_{\alpha}}{r_{i\alpha}} \quad (19)$$

$v(r_i)$: Le potentiel externe de l'électron. Il définit l'attraction de l' e^- (i) avec tous les noyaux (système considéré comme externe vis-à-vis les électrons).

$\rho_0(r)$: exprime la densité électronique au point r. Le nombre total d'électrons obtenu après intégration sur cette densité ponctuelle sur tout l'espace :

$$\int \rho_0(r) dr = n \quad (20)$$

L'énergie totale peut s'écrire comme la somme de trois fonctionnelles:

$$E_0[\rho_0] = V_{ne}[\rho_0] + T[\rho_0] + V_{ee}[\rho_0] \quad (21)$$

Avec:

$$V_{ne}[\rho_0] = \int \rho_0(r)v(r)dr \quad (22)$$

Par conséquent, la fonctionnelle de l'énergie peut s'écrire:

$$E_0[\rho] = \int \rho_0(r)v(r)dr + F[\rho_0] \quad (23)$$

Avec:

$$F[\rho_0] = T[\rho_0] + V_{ee}[\rho_0] \quad (24)$$

La fonctionnelle $F[\rho_0]$ est inconnue

• **2^{ème} théorème de Hohenberg et Kohn:**

Enoncé : « Pour une densité d'essai $\tilde{\rho}(r)$, telle que $\tilde{\rho}(r) \geq 0$ et $\int \tilde{\rho}(r)dr = n$ l'inégalité suivante est vérifiée:

$$E_0 \leq E[\tilde{\rho}] \quad (25)$$

I.3.2 Méthode de Kohn et Sham:

La procédure pour calculer l'énergie E_0 à partir de ρ_0 , n'a pas été élaborée par les théorèmes de Hohenberg et Kohn. En 1965, Kohn et Sham(KS) ont proposé une méthode pratique pour trouver E_0 à partir de ρ_0 [9]. Ils ont suggéré un système fictif de référence, noté s , constitué par les n électrons non interagissant.

Le système de référence est choisi de telle façon à avoir:

$$\rho_s(r) = \rho_0(r) \quad (26)$$

Etant donné que les électrons n'interagissent pas entre eux dans le système de référence, l'hamiltonien de système de références'écrit:

$$\hat{H}_s = \sum_{i=1}^n [-1/2\nabla_i^2 + v_s(r_i)] = \sum_{i=1}^n h_i^{ks} \quad (27)$$

Avec:

$$h_i^{ks} = -1/2\nabla_i^2 + v_s(r_i) \quad (28)$$

Par conséquent, les équations de Kohn et Sham, pour l'électron i , sont les suivantes :

$$h_i^{ks} \theta_i^{ks} = \varepsilon_i^{ks} \theta_i^{ks} \quad (29)$$

θ_i^{KS} : Orbitale de Kohn et Sham de l'électron i .

- **Terme d'échange-corrélation**

Soit ΔT la différence de l'énergie cinétique entre le système réel (électrons interagissants) et le système de référence.

$$\Delta T = T[\rho] - T_s[\rho] \quad (30)$$

Et

$$\Delta V_{ee} = V_{ee}[\rho] - 1/2 \int \int \frac{\rho(r_1)\rho(r_2)}{r_{12}} dr_1 \quad (31)$$

ΔV_{ee} : Est la différence entre la répulsion électronique réelle et la répulsion coulombienne entre deux densités ponctuelles. L'expression de l'énergie est:

$$E_v[\rho] = \int \rho(r)v(r)dr + T_s[\rho] + 1/2 \int \int \frac{\rho(r_1)\rho(r_2)}{r_{12}} dr_1 dr_2 + \Delta T + \Delta V_{ee}[\rho] \quad (32)$$

La fonctionnelle d'énergie d'échange-corrélation est exprimée par :

$$E_{xc}[\rho] = \Delta T[\rho] + \Delta V_{ee}[\rho] \quad (33)$$

$$E_v[\rho] = \int \rho(r)v(r)dr + T_s[\rho] + 1/2 \int \int \frac{\rho(r_1)\rho(r_2)}{r_{12}} dr_1 dr_2 + E_{xc}[\rho] \quad (34)$$

Dans les calculs DFT, selon le schéma de Kohn et Sham, la difficulté réside dans la qualité (meilleure) de l'approximation pour l'énergie échange- corrélation E_{xc} .

La densité électronique ρ_0 est calculée par les orbitales de Kohn et Sham selon la formule suivante:

$$\rho_0(r) = \rho_s(r) = \sum_{i=1}^n |\theta_i^{KS}|^2 \quad (35)$$

Ces orbitales KS servent à calculer l'énergie cinétique du système de référence T_s . L'énergie E_0 peut s'écrire:

$$E_0 = - \sum_{\alpha} Z_{\alpha} \int \frac{\rho(r_1)}{r_{1\alpha}} dr_1 - 1/2 \sum_{i=1}^n \langle \theta_i^{KS}(1) | \nabla_1^2 | \theta_i^{KS}(1) \rangle + 1/2 \int \int \frac{\rho(r_1)\rho(r_2)}{r_{12}} dr_1 dr_2 + E_{xc}[\rho] \quad (36)$$

L'équation aux valeurs propres correspondante est de la forme:

$$\left[-1/2 \nabla_1^2 - \sum_{\alpha} \frac{Z_{\alpha}}{r_{1\alpha}} + \int \frac{\rho(r_2)}{r_{12}} dr_2 + v_{xc}(1) \right] \theta_i^{KS}(1) = \varepsilon_i^{KS} \theta_i^{KS}(1) \quad (37)$$

Le potentiel d'échange-corrélation v_{xc} est la dérivée de l'énergie échange-corrélation E_{xc} par rapport à la densité électronique:

$$v_{xc}(r) = \frac{\partial E_{xc}[\rho(r)]}{\partial \rho(r)} \quad (38)$$

Il existe plusieurs approximations de ce potentiel d'échange-corrélation.

1.3.3 Approximation de la densité locale LDA:

Afin trouver une expression pour le terme d'échange-corrélation plusieurs approximations ont été proposées pour l'approximation de la densité locale (LDA), l'une de ces propositions, qui considère que la densité varie lentement en fonction de la position.

L'énergie d'échange-corrélation $E_{xc}[\rho_s]$ peut s'écrire:

$$E_{xc}^{LDA}[\rho] = \int \rho(r) \varepsilon_{xc}(\rho) dr \quad (39)$$

ε_{xc} : étant l'énergie d'échange-corrélation par électron. Cette quantité est exprimée comme la somme des deux contributions:

$$\varepsilon_{xc}(\rho) = \varepsilon_x(\rho) + \varepsilon_c(\rho) \quad (40)$$

Avec:

$$\varepsilon_x(\rho) = -\frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} (\rho(r))^{1/3} \quad (41)$$

Donc:

$$E_x^{LDA} = \int \rho \varepsilon_x dr = -\frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} \int [\rho(r)]^{4/3} dr \quad (42)$$

Le terme de corrélation $\varepsilon_c(\rho)$ est exprimé par la formule de Vosko, Wilk, et Nusair (VWN) [10,11].

Fonctionnelles E_x et E_c :

L'énergie E_{xc} d'échange-corrélation est la somme de deux fonctionnelles d'échange E_x et de corrélation E_c :

$$E_{xc} = E_x + E_c \quad (43)$$

E_x est l'énergie d'échange et E_c est l'énergie de corrélation.

Une expression similaire à celle de HF est obtenue et exprimée par les orbitales de Kohn et Sham.

$$E_x = -\frac{1}{4} \sum_{i=1}^n \sum_{j=1}^n \langle \theta_j^{KS}(1) \theta_j^{KS}(2) | 1/r_{12} | \theta_i^{KS}(1) \theta_i^{KS}(2) \rangle \quad (44)$$

L'énergie de corrélation est calculée comme la différence entre E_{xc} et E_x .

$$E_c = E_{xc} - E_x \quad (45)$$

1.3.4 Approximation du Gradient Généralisé (GGA):

L'approximation LDA est basée sur le fait que la densité électronique ρ est uniforme et varie très peu avec la position. L'amélioration de cette approximation nécessite l'introduction des gradients des densités. Cette technique est connue sous les noms appelées l'approximation du gradient généralisé GGA (Generalized gradient approximation). L'énergie d'échange-corrélation s'écrit alors:

$$E_{xc}^{GGA} = \int \varepsilon_{xc}^{GGA}(\rho, \Delta\rho) dr \quad (46)$$

Où ε_{xc}^{GGA} est la densité d'énergie d'échange-corrélation. La difficulté réside dès lors dans la recherche d'expressions analytiques de ε_{xc}^{GGA} .

E_{xc}^{GGA} est divisé en deux contributions: échange et corrélation:

$$E_{xc}^{GGA} = E_x^{GGA} + E_c^{GGA} \quad (47)$$

Plusieurs fonctionnelles ont été élaborées pour les termes d'échange et corrélation. Parmi les fonctionnelles d'échange on cite Becke (B88) [12] et de Perdew et Wang (PW91) [13]. Pour la corrélation, on trouve les fonctionnelles de Perdew (P86) [13], de Lee, Yang et Parr (LYP) [18] et de Perdew et Wang (PW91) [13].

I.3.5 Fonctionnelle hybride B3LYP:

La fonctionnelle hybride B3LYP (Becke 3-parameters Lee-Yang-Parr) consiste à mélanger de plusieurs fonctionnelles de différentes méthodes selon l'expressionsuivante:

$$E_{xc}^{B3LYP} = (1 - a_0 - a_x)E_x^{LSDA} + a_0E_x^{exact} + a_xE_x^{B88} + (1 - a_c)E_c^{VWN} + a_cE_c^{LYP} \quad (48)$$

E_x^{LDA} , E_x^{B88} , E_x^{HF} sont les fonctionnelles d'échange local de Becke et l'énergie d'échange HF respectivement.

E_c^{VWN} est la fonctionnelles de corrélation locale et corrigée du gradient de Lee, Yang et Parr

Les valeurs des 3 paramètres d'ajustement sont: $a_0 = 0.20$; $a_x = 0.72$; $a_c = 0.81$

I.4 Bases d'orbitales atomiques:

Il y a deux sortes de fonctions de base qui sont d'un usage courant. Le premier type de bases sont les orbitales de type Slater STO [12] qui sont les meilleures OA analytiques définies par:

$$\Psi_{nlm} = N_n r^{n^*-1} \exp(-\zeta r) Y_{lm}(\theta, \phi) \quad (49)$$

Où N_n est le constante de normalisation et ζ est l'exposant de Slater, $Y_{lm}(\theta, \phi)$ sont les parties angulaires.

A cause de la complexité du calcul d'intégrales moléculaires les OAs type Slater(STOs) ne sont pas utilisées dans les programmes moléculaires *abinitio*. Ces derniers (tel que Gaussian) utilisent les bases de fonctions gaussiennes (GTOs) proposées par Boys [13,14].

$$g(\alpha, \vec{r}) = c X^n y^l Z^m \exp(-\alpha r^2) \quad (50)$$

Dans cette équation, α est une constante déterminant la taille de la fonction. La somme $(n+l+m)$ définit le type de l'orbitale atomique.

L'outil des fonctions gaussiennes est justifié par le fait que « Le produit de deux gaussiennes centrées en deux points A et B est équivalent à une gaussienne centrée en un point situé sur le segment AB. Cette caractéristique facilite énormément le calcul d'intégrales bi-électroniques.

Les bases gaussiennes présentent une mauvaise représentation des orbitales atomiques à l'origine et aux grandes distances. Pour améliorer cet inconvénient des combinaisons linéaires de gaussiennes sont utilisées comme fonctions de base. Ces fonctions sont appelées « fonctions gaussiennes contractées ». Une fonction gaussienne contractée (CGTO) est une combinaison linéaire de gaussiennes primitives (PGTO_s):

$$G^{CGTO} = \sum_{\lambda=1}^k d_{\lambda} g_{\lambda}^{PGTO} \quad (51)$$

d_{λ} étant le coefficient de contraction de la gaussienne primitive g_{λ} . k est le degré de contraction.

6-311++G** est une base SV-TZ 6-311G, les orbitales de couches internes sont exprimées par une contraction de 6 gaussiennes primitives. Les orbitales de la split couche de valence sont exprimées par des contractions de 3, 1 et 1 gaussiennes primitives. C'est une base avec des orbitales diffuses sur tous les atomes, et des orbitales de polarisation uniquement sur les atomes lourds et aux hydrogènes.

I.5. Les méthodes de la solvation:

Il existe deux façons de modéliser un solvant explicite et implicite. La différence fondamentale entre ces deux modèles est que le modèle explicite décrit les molécules de solvant de manière atomique (par dynamique moléculaire classique ou quantique) tandis que le modèle implicite représente le solvant comme un milieu homogène dit continuum de constante diélectrique « ϵ », à la place de molécules, le soluté est placé dans une cavité de forme définie date des travaux de Kirkwood [15] et Onsager concernant les effets de solvation sur les molécules polaires [16-18].

1.5.1. Modèle SCRF:

Dans la figure 2 nous avons représenté trois manières de construire les cavités.

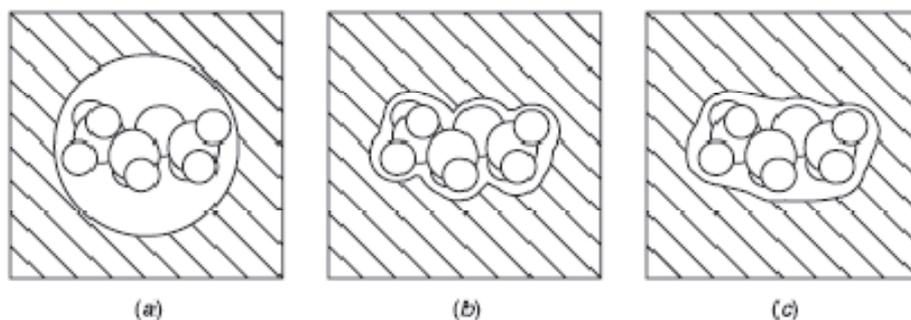


Figure2: (a) un soluté intégré dans un milieu diélectrique avec une cavité sphérique, (b) une cavité formée en unissant des sphères atomiques de rayon égal au rayon de van der Waals, (c) une surface accessible au solvant.

La construction de la cavité coûte une quantité d'énergie qui est positive nommée l'énergie libre de cavitation (ΔG_{cav}). Cette quantité dépend de la nature du solvant ainsi que de la forme de la cavité. Le soluté est mis dans la cavité et il va interagir avec le continuum. Il existe trois types d'interactions soluté-continuum: interactions électrostatiques ΔG_{elec} , répulsives et dispersives.

L'interaction d'une molécule dans sa cavité est résolue d'une manière auto-cohérente (SCRF): la surface de la cavité du continuum est polarisée par la distribution de charge du soluté et cela conduit à l'apparition des charges à l'interface soluté-continuum. Ces charges engendrent un potentiel de réaction qui à son tour, modifie la densité électronique du soluté et une nouvelle modification de la distribution de charges à l'interface soluté-continuum apparaît et ainsi de suite jusqu'à l'obtention de la convergence électrostatique entre la distribution de charges propre au soluté et celle de la surface de la cavité. Ce terme énergétique est appelé la contribution électrostatique (ΔG_{elec}) et il est toujours négatif.

Les deux termes de dispersion/répulsion sont souvent associés, ces deux derniers (ΔG_{disp} et $\Delta G_{rép}$) définissent l'interaction du soluté avec le solvant à l'interface de la cavité, et donne une contribution négative/positive à la variation d'énergie.

Finalement, l'énergie totale d'interactions c'est la somme d'un terme électrostatique et de trois termes non-électrostatiques:

$$\Delta G_{\text{solv}} = \Delta G_{\text{cav}} + \Delta G_{\text{elec}} + \Delta G_{\text{disp}} + \Delta G_{\text{rép}} \quad (52)$$

1.5.2. Modèle PCM:

Le modèle PCM [19-21] permet de définir la cavité du soluté comme un assemblage de sphères atomiques de Van der Waals. Chaque sphère est calculée en fonction du rayon de van der Waals de chaque atome et la densité de charge est répartie sur la surface de la cavité. De plus cette cavité a une forme plus réaliste dont la surface est découpée en aires élémentaires constituée de petits polygones sphériques (appelées tesserae) (Figure 3)



Figure 3: Découpage de la surface d'une cavité en un ensemble de tesserae.

Le modèle C-PCM est une extension du modèle PCM, ce modèle permet de simplifier le calcul des densités de charge à la surface de la cavité.

1.5.3. Modèle SMD:

Le modèle SMD [22] est un modèle continuum de solvation basé sur la densité de charge d'un soluté interagissant avec le milieu diélectrique «continuum» qui est le solvant, d'où «D» Signifie "densité" désigne l'utilisation de la densité électronique complète des solutés, calculées par les méthode quantiques, sans définir les charges partielles des atomes. Le modèle SMD est un modèle universel de solvation, qui peut être appliqué à tout soluté chargé ou non chargé dans n'importe quel solvant ou liquide pour lequel les

informations clés telles que: la constante diélectrique, l'indice de réfraction, la tension superficielle et l'acidité doivent être disponibles.

Selon le modèle SMD, l'énergie libre de solvation se décompose en trois composantes suivant la relation:

$$\Delta G_s = \Delta G_{ENP} + \Delta G_{CDS} + \Delta G_{CONC} \quad (53)$$

ΔG_{ENP} Désigne les termes de l'énergie libre (l'électronique (E), nucléaire (N), et La polarisation du continuum (P). Si la géométrie est la même dans la phase gazeuse et en solution l'énergie ΔG_{ENP} devienne ΔG_{EP} . De même comme le modèle PCM, le calcul de cette énergie est basé sur champ de réaction auto-cohérent donnée par l'expression suivante:

$$\Delta G_{EP} = \left\langle \Psi \left| H^{(0)} - \frac{e}{2} \varphi \right| \Psi \right\rangle + \frac{e}{2} \sum_k Z_k \varphi_k - \langle \Psi^{(0)} | H^{(0)} | \Psi^{(0)} \rangle \quad (54)$$

Avec:

e est l'unité atomique de la charge:

φ_k Est le champ de réaction évaluée à l'atome k et Z_k est le numéro atomique de l'atome k,

$H^{(0)}$ Est l'Hamiltonien électronique de soluté.

$\Psi^{(0)}$ Est la fonction d'onde électronique dans la phase gazeuse.

Ψ Est la fonction d'onde électronique polarisée de soluté en solution.

ΔG_{CDS} Désigne les termes de l'énergie libre de cavitation (C), les variations de l'énergie de dispersion (D), et les changements structuraux du solvant (S).

ΔG_{CONC} désigne le terme de correction de la variation de l'énergie libre ΔG entre l'état standard en phase gazeuse (1 atm) et l'état standard en solution (1 M).

Bibliographie

- [1] E. Schrödinger, *Ann. Phys. Leipzig*, 76 (1926) 361.
- [2] M. Born and J. R. Oppenheimer, *Ann. Phys.*, 84 (1927) 457.
- [3] V. Minkine, B. Simkine and R. Minaev, « Théorie de la structure moléculaire », Edition Mir, Moscou, (1982).
- [4] J. C. Slater, *Phys. Rev.*34 (1929) 1293etJ. C. Slater, *Phys. Rev*, 38 (1931) 38.
- [5] C. C. Roothaan, *Rev. Mod. Phys.* 23 (1951) 69.
- [6] R. M. Dreizler, E. K. U. Gross, *Density Functional Theory: “An Approach to the QuantumMany-Body Problem”*, Springer-Verlag, Berlin, (1990).
- [7] R. G. Parr, W. Yang, *Density-Functional“Theory of Atoms and Molecules”*, Oxford UniversityPress, Oxford, (1989).
- [8] P. Hohenberg and W. Kohn, *Phys. Rev.*, 136 (1964) B846.
- [9] W. Khon and L. J. Sham, *Phys. Rev.*, 140 (1965) A1133.
- [10] S. J. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.*, 58 (1980) 1200.
- [11] F. Jensen « Introduction to Computational Chemistry », John Wiley & Sons, (1999).
- [12] J. C. Slater, *J. Chem. Phys.*, 36 (1930) 57 .
- [13] S. F. Boys, *Proc. Roy. Soc.*, A200 (1950) 542.
- [14] E. Clementi, Ed., «Modern Techniques in Computationnal Chemistry», OTECC 89, (ESCOM, Leiden), (1989).
- [15] J. G. Kirkwood, *J. Chem. Phys.*, 2 (1934) 351.
- [16] K. Baldrige and A. Klamt, *J. Chem. Phys.*, 106 (1997) 6622.
- [17] L. Onsager, *J. Am. Chem. Soc.*, 58 (1936) 1486.
- [18] M. Cossi, *J. Comp. Chem.*, 19 (1998) 833.
- [19] S. Miertuš, E. Scrocco, and J. Tomasi, *J. Chem. Phys.*, 55 (1981) 117.
- [20] J. Tomasi and M. Persico, *Chem. Rev.*, 94 (1994) 2027.
- [21] R. Cammi and J. Tomasi, *J. Comp. Chem.*, 16 (1995) 1449.
- [22] A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B.*, 113 (2009) 6378.

CHAPITRE II :

LA MODELISATION QSPR ET ANALYSE DE DONNEES

I.1. Principe des méthodes QSPR

Les méthodes QSPR (*Quantitative Structure-Property Relationships*) sont largement utilisées en chimie, biologie ainsi d'autres domaines de recherche. Ces méthodes sont basées sur le principe d'élaborer une relation quantitative des **descripteurs** qui peuvent être des propriétés moléculaires, électroniques ou géométriques ou des propriétés physico-chimiques mesurées expérimentalement d'une famille de composés chimiques.

La forme générale d'un tel modèle est: $Propriété = f(\text{descripteurs})$

Les méthodes QSPR sont basées sur des méthodes d'analyses de données telle que : régressions linéaires simples (SLR) et multiples (MLR) [6], régressions aux moindres carrés partiels (PLS) [7]. L'objectif est de chercher les paramètres déterminants pour la propriété.

Après avoir validé le modèle QSPR obtenu, il peut être utilisé pour la prédiction de la propriété non mesurée de nouvelles molécules synthétisées ou non.

En pratique, l'élaboration d'un modèle suit les étapes suivantes (figure1):

- Collection des données expérimentales (propriété) pour une famille de composés : c'est une étape cruciale dans le développement des modèles QSPR. Les données expérimentales doivent être aussi fiables que possible afin de minimiser les erreurs dans le modèle final.
- Calcul des descripteurs moléculaires, qui peuvent décrire la propriété : ces descripteurs peuvent être constitutionnels [8], topologiques [9-13], géométriques [14] et quantiques [15-18].
- Analyse des données afin de choisir les meilleurs descripteurs et élaborer le modèle.
- Validation du modèle obtenu.
- Tester le pouvoir prédictif du modèle et déterminer le domaine d'applicabilité.

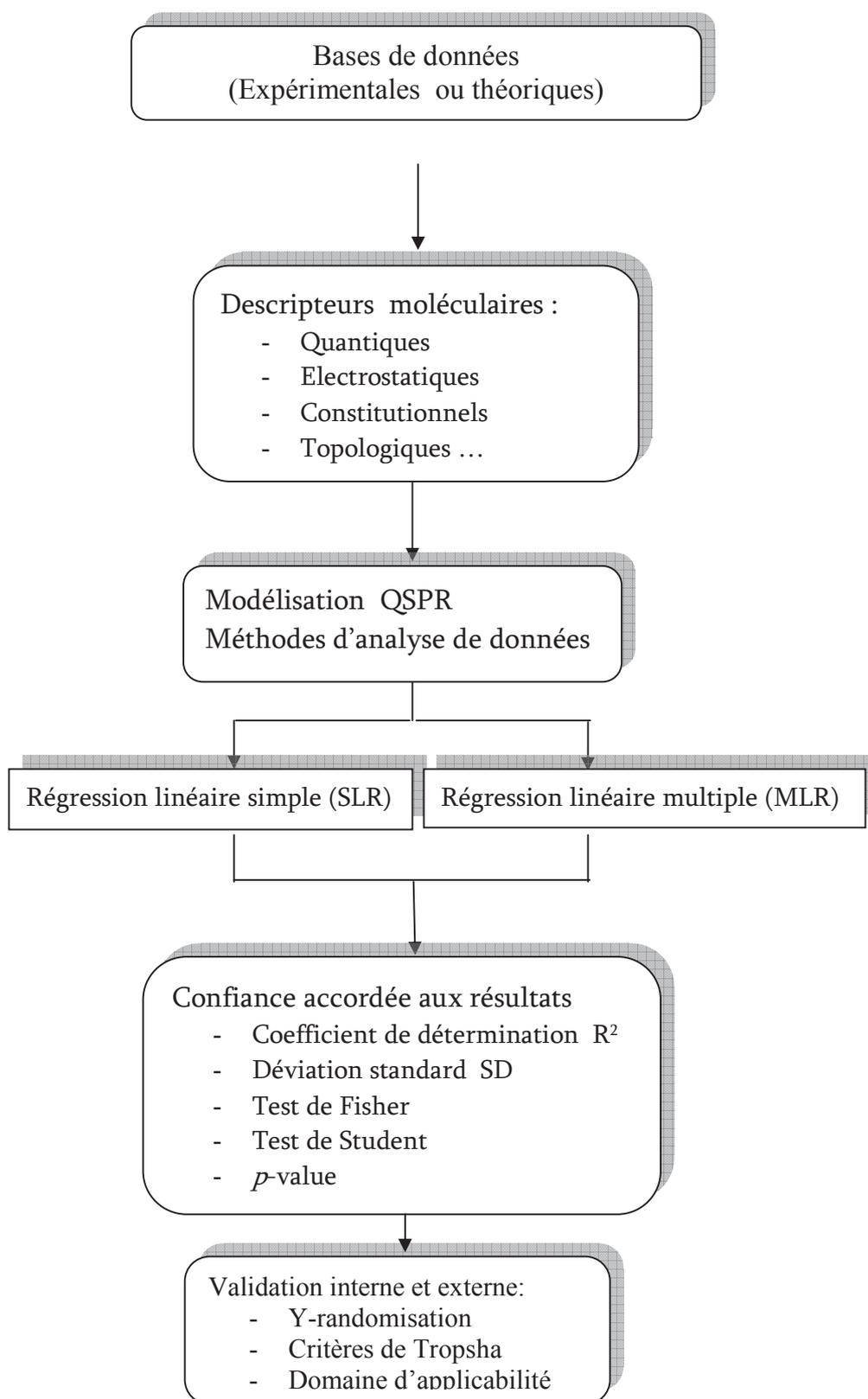


Figure 1: Principe de la méthode QSPR

I.2. Méthodes d'analyse de données

Il est à noter que parmi les difficultés rencontrées dans l'élaboration des modèles QSPR est la différence d'échelles existant entre les descripteurs théoriques (l'échelle moléculaire) et les propriétés physico-chimiques (l'échelle macroscopique). Dans cette partie nous présenteront deux types de méthodes d'analyse : la régressions linéaires simples (SLR) et multiples (MLR).

I.2.1. Régression linéaire simple

La régression linéaire simple (SLR) est un outil statistique permettant de modéliser l'équation linéaire entre une variable explicative « x » (descripteur) et une variable à expliquer « y » (la propriété). Donc il y a une dépendance linéaire entre le descripteur et la propriété mais cette relation est perturbée par un bruit « ϵ » ou une erreur :

$$y = \beta_0 + \beta_1 x + \epsilon. \quad (1)$$

Cette équation 1 est appelée modèle de régression linéaire. Les β_i , à déterminer, sont appelés les coefficients de régression. Comment trouver la droite qui passe « au plus près » de tous les points :

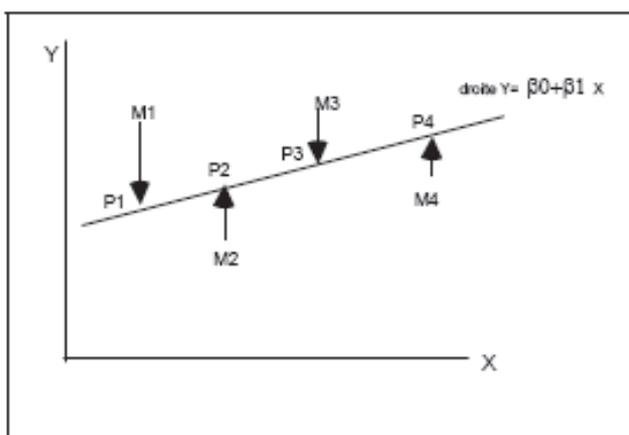


Figure 2.

L'obtention de la meilleure droite de régression (passant par un maximum de points) est basée sur le *critère d'ajustement*. Le principe repose la projection des points de M_i parallèlement à l'axe y d'où les points P_i (**Figure 2**) et l'objectif est d'avoir la somme des carrés des écarts (SCE) des points observés M_i à la droite la plus minimum possible. Le critère d'ajustement est appelé « critère des Moindres Carrés Ordinaires » (*Ordinary Least Squares*).

1. Méthode d'estimation des paramètres β_0 et β_1

La Somme des Carrés des Ecarts (SCE) est donnée par :

$$S(\beta_0, \beta_1) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \quad (2)$$

Si les dérivées de la somme S par rapport aux coefficients β_0, β_1 s'annulent. Alors la valeur la fonction $S(\beta_0, \beta_1)$ est minimum. On obtient les équations suivantes :

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_0} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (3)$$

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_1} = -2 \sum_{i=1}^n x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (4)$$

$\hat{\beta}_0$ et $\hat{\beta}_1$ sont les estimateurs des moindres carrés des coefficients β_0 et β_1 . La première équation donne :

$$\hat{\beta}_0 n + \hat{\beta}_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \quad (5)$$

et nous avons un estimateur de l'ordonnée à l'origine (en divisant par n)

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (6)$$

où $\bar{x} = \sum_{i=1}^n x_i / n$.

En remplaçant $\hat{\beta}_0$ par son expression, la seconde équation nous donne une première écriture de $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{\sum x_i y_i - \sum x_i \bar{y}}{\sum x_i^2 - \sum x_i \bar{x}} \quad (7)$$

Cet estimateur peut être écrit sous une autre forme :

$$\hat{\beta}_1 = \frac{\sum x_i (y_i - \bar{y})}{\sum x_i (x_i - \bar{x})} = \frac{\sum (x_i - \bar{x}) (y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\text{cov}(x, y)}{\text{var}(x)} \quad (8)$$

cov(x,y) est la covariance et var(x) est la variance.

Les points qui sont sur la droite de régression ont pour ordonnée:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \quad (9)$$

Où \hat{y} est l'estimation de y obtenue à partir de l'équation de régression.

2. Analyse de la variance

A partir de l'équation de la droite de régression (modèle retenu), on peut pour tout point i d'abscisse x_i calculer son estimation (ordonnée) y_i . on obtient :

$$\hat{y}_i = \bar{y} + \hat{\beta}_1 (x_i - \bar{x}) \quad (10)$$

Ou encore :

$$\hat{y}_i - \bar{y} = \hat{\beta}_1 (x_i - \bar{x}) \quad (11)$$

En un point i l'écart ou le résidu est :

$$y_i - \hat{y}_i = (y_i - \bar{y}) - (\hat{y}_i - \bar{y}) \quad (12)$$

On élève les deux membres au carré et on somme sur les observations i :

$$\sum (y_i - \hat{y}_i)^2 = \sum (y_i - \bar{y})^2 + \sum (\hat{y}_i - \bar{y})^2 - 2 \sum (y_i - \bar{y})(\hat{y}_i - \bar{y}) \quad (13)$$

En utilisant la l'équation 11 on obtient :

$$\sum (y_i - \hat{y}_i)^2 = \sum (y_i - \bar{y})^2 + \sum (\hat{y}_i - \bar{y})^2 - 2 \sum (y_i - \bar{y}) \hat{\beta}_1 (x_i - \bar{x}) \quad (14)$$

En utilisant la transformation de l'équation 8 : $\hat{\beta}_1 \sum (x_i - \bar{x})^2 = \sum (x_i - \bar{x}) (y_i - \bar{y})$

L'équation 14 devient:

$$\sum (y_i - \hat{y}_i)^2 = \sum (y_i - \bar{y})^2 + \sum (\hat{y}_i - \bar{y})^2 - 2 \hat{\beta}_1^2 \sum (x_i - \bar{x})^2 \quad (15)$$

En utilisant l'équation 11 :

$$\sum (y_i - \hat{y}_i)^2 = \sum (y_i - \bar{y})^2 + \sum (\hat{y}_i - \bar{y})^2 - 2 \sum (\hat{y}_i - \bar{y})^2 \quad (16)$$

On aboutit enfin à l'égalité fondamentale :

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2 \quad (17)$$

Ou bien : $SS\ Total = SSR\ regression + SS\ Error$, $SCE\ totale = SCE\ modèle + SCE\ erreur$

* SST (Sum of Squares Total): la somme des carrés des écarts à la moyenne \bar{y} (SCE):

$$\sum_{i=1}^n (y_i - \bar{y})^2 = SCE \text{ totale (SST)} \quad (18)$$

* SSR (Sum of Squares Regression or model) : la somme des carrés des écarts à la moyenne :

$$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = SCE \text{ mod èle (SSR)} \quad (19)$$

* SSE (Sum of Squares Error) représente la somme des carrés des écarts, ou erreurs:

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n \varepsilon_i^2 = SCE \text{ erreur (SSE)} \quad (20)$$

Représentation de l'analyse de la variance:

Tous les programmes de régression donnent un tableau de l'analyse de la variance dit table d'ANOVA (ANalysis Of Variance). Ce dernier est formé de 3 lignes. La 3^{ème} ligne « la totale » est décomposée en deux parties : la partie régression et la partie erreur.

Table d'analyse de variance (ANOVA)

Source de variation	Degrés de liberté (Degrees of Freedom)	Somme des carrés (Sum of Squares)	Moyenne de la somme des carrés (Mean Square)
Source	DF	SS	MS
Regression	p	SSR	SSR/p
Error	n-p-1	SSE	SSE/(n-p-1)
Total	n-1	SST	

DF : le degrés de liberté respectivement égal à n-1, p, n-p-1

n : le nombre d'observations,

p : le nombre de variables régresseurs (la variable x₀, constante égale à 1, correspondant au paramètre β₀, n'est pas comprise).

Toutes les quantités SS ET MS sont représentées dans la **Figure 3** :

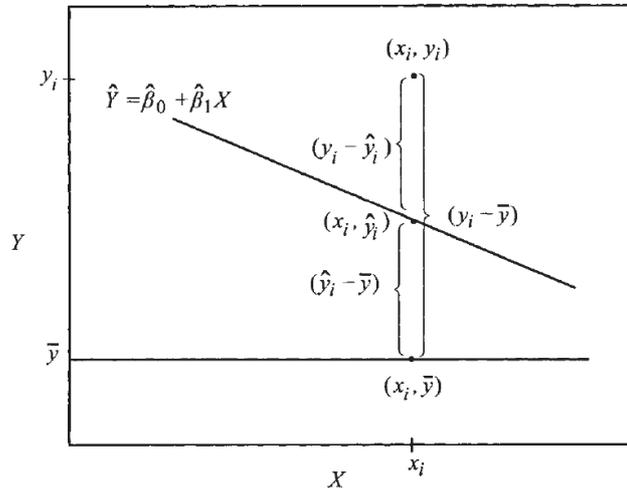


Figure 3: représentation graphique des différentes quantités calculées après la régression.

Par suite, on obtient la très importante formule d'analyse de variance :

$$SS \text{ Total} = SSR \text{ regression} + SS \text{ Error} \quad \text{ou bien :} \quad SCE \text{ totale} = SCE \text{ modèle} + SCE \text{ erreur}$$

3. Validation du modèle

- **Le coefficient de corrélation R^2**

Le coefficient de corrélation est le rapport des quantités SCE totale (SST) et SCE modèle (SSR).

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{SSR}{SST} \quad (21)$$

$$\text{ou encore} \quad R^2 = 1 - \frac{SSE}{SST} \quad (22)$$

avec : $SSR \leq SST$ et $0 < R^2 < 1$

Dans un bon modèle, les valeurs estimées de \hat{y}_i sont proches des vraies valeurs de y_i . Les deux quantités sont des sommes de carrés toujours positives ou nulles (avec $SSR \leq SST$). Si les valeurs prédites \hat{y}_i sont très proches des valeurs expérimentales y_i tout les points seront alignés sur la droite de régression et cette dernière sera parfaite (SCE Erreur = 0, $R^2 = 1$).

Un R^2 proche de zéro signifie les descripteurs n'ont aucun effet sur la variable à expliquer y_i .

- **Coefficient de corrélation ajusté**

Le coefficient de corrélation ajusté R^2 -adj est défini par :

$$R_{ajusté}^2 = 1 - \frac{SSE / (n - p - 1)}{SST / (n - Intercept)} \quad (23)$$

$$R_{ajusté}^2 = 1 - \frac{(n - Intercept)}{n - p - 1} (1 - R^2) \quad (24)$$

Avec $Intercept = 0$, si il n'y a pas de constante β_0 à l'origine si non $Intercept = 1$.

Cette formule indique notamment que R^2 -adj est toujours inférieur à R^2 , et ceci d'autant plus que le modèle contient un grand nombre de prédicteurs.

- **Estimation de la variance σ^2 (ou s^2).**

La variance σ^2 (moyenne de la somme des écarts MSE) est estimée par la formule suivante:

$$\sigma^2 = s^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - p - 1} = \frac{SSE}{n - p - 1} = MSE \quad (25)$$

L'écart moyen résiduel connu sous le nom la déviation standard est calculé par :

$$s = SD = \sqrt{s^2} \quad (26)$$

I.2.2. Régression linéaire multiple

Le modèle de régression multiple MLR (*Multiple Linear Regression*) [19-23] est une généralisation de la méthode SLR. Il est basé sur le principe d'élaborer une relation linéaire entre une variable dépendante y (la propriété) et une série de p descripteurs x_i :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_p x_p + \varepsilon \quad (27)$$

où β_i sont les coefficients de la régression.

La détermination de cette se fait alors à partir d'une base de données de n échantillons de variables y_i et x_i connues. Il s'agit donc de considérer un système de p équations.

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{1,1} + \beta_2 x_{2,1} + \dots + \beta_p x_{n,1} + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_{1,2} + \beta_2 x_{2,2} + \dots + \beta_p x_{n,2} + \varepsilon_2 \\ &\dots \\ y_n &= \beta_0 + \beta_1 x_{1,p} + \beta_2 x_{2,p} + \dots + \beta_p x_{n,p} + \varepsilon_n \end{aligned} \quad (28)$$

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

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & & & & \\ 1 & & & & \\ \dots & & & & \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \dots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \dots \\ \varepsilon_n \end{pmatrix} \quad (29)$$

On ajoute à la matrice x (n,p) une première colonne un vecteur constitué uniquement de 1.

Ce vecteur correspond à la constante x_0 . la matrice x est alors de dimension $(n, p+1)$.

Le modèle s'écrit alors sous forme $Y = \beta X + \varepsilon$

Le principe est, le même à celui de la méthode SLR, de déterminer les coefficients du vecteur β afin de minimiser la somme des carrés des écarts entre les valeurs estimées les valeurs expérimentales.

La valeur estimée de la variable y peut être écrite:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \dots + \hat{\beta}_p x_{p,i} \quad (30)$$

Les résidus peuvent être définis :

$$\varepsilon_i = y_i - \hat{y}_i \quad (31)$$

I.2.3. Interprétation et validation d'un modèle QSPR

Une fois développé, le modèle doit être validé en analysant tous les paramètres statistiques obtenus.

I.2.3.1. Analyse des paramètres statistiques d'un modèle QSPR

En plus des paramètres cités précédemment (paragraphe II.2.1) il est nécessaire dans le cas de la régression linéaire multiple d'analyser les paramètres suivants :

1. Test de la signification globale de la régression

- **Test Fisher-Snedecor**

Ce test à surtout un intérêt dans le cadre de la régression multiple, c'est à dire avec p régresseurs.

$$\mu_i = E (y_i) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (32)$$

Ce test permet de connaître l'apport global de l'ensemble des variables x_1, x_2, \dots, x_p à la détermination de y .

On veut tester l'hypothèse nulle:

$H_0: \beta_1 = \dots = \beta_p = 0$ contre $H_a: \text{il existe au moins un } \beta_j \text{ parmi } \beta_1, \dots, \beta_p \text{ non égal à } 0.$

On calcule la statistique de test (test de Fisher) $F = \frac{MS\ Model}{MS\ Error}$ (33)

Avec : $MS\ model = \frac{SS\ Model}{p}$ (34)

Et $MS\ Error = \frac{SS\ Error}{n - p - 1}$ (35)

MSmodel et MSerror représentent respectivement la somme des carrés des écarts moyens pour la régression et pour l'erreur.

Le test de Fisher mesure le rapport entre la variance de la variable y expliquée et non expliquée par le modèle de régression. Le test Fisher-Snedecor permet de tester l'hypothèse nulle (chaque β est significativement différent de zéro).

Raisonnement sur la p-value

Le logiciel de statistique fournit la un paramètre dit p-value associée au F observé. La p-value est comparée à une valeur α choisie (par exemple $\alpha=0.05$).

Si $p\text{-value} \leq \alpha$, alors on rejette l'hypothèse nulle $\beta_1 = \dots = \beta_p = 0$

Les tables d'analyse de variance complètes est du type suivant :

Table d'analyse de variance

Source	DF	SS	MS	F	p
Regression	p	SSR	SSR/p	MSR/MSE	p value
Error	n-p-1	SSE	SSE/(n-p-1)		
Total	n-1	SST			
R^2		SSR/SST			

- **Test de Student (t-statistic)**

- ***Statistiques liées au paramètre β_1***

Le *t*-test de Student permet de tester l'hypothèse de significativité du descripteur ou non. Pour une régression MLR, cela revient à supposer le coefficient β_i qui lui est associé comme nul.

On s'intéresse au test de l'hypothèse nulle:

H_0 : paramètre $\beta_1 = 0$ contre H_a : paramètre $\beta_1 \neq 0$

On calcule la statistique de test : $t\text{-observé} = \frac{\hat{\beta}_1}{s(\hat{\beta}_1)}$ (36)

Avec: $s(\hat{\beta}_1) = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}}$ (37)

$s(\beta_i)$ est l'erreur type du paramètre β_i

Raisonnement

On compare la p-value associée à *t*-observé, au risque α choisi (par ex: $\alpha=0.05$).

Si p-value $\leq \alpha$

Alors on rejette l'hypothèse $\hat{\beta}_1 = 0$

Conclusion : $\hat{\beta}_1$ est significativement différent de zéro au niveau α .

- ***Statistiques liées au paramètre $\hat{\beta}_0$*** :

L'estimateur de la variance du coefficient $\hat{\beta}_0$ est :

$$s^2(\hat{\beta}_0) = \frac{s^2 \sum x_i^2}{n \sum (x_i - \bar{x})^2} \quad (38)$$

On s'intéresse au test de l'hypothèse nulle

H_0 : paramètre $\hat{\beta}_0 = 0$ contre H_a : paramètre $\hat{\beta}_0 \neq 0$

On calcule la statistique de test : $t - \text{observé} = \frac{\hat{\beta}_0}{s(\hat{\beta}_0)}$ (39)

Raisonnement

On compare la *p-value* associée à t-observé, c'est-à-dire la probabilité de dépasser le t-observé en valeur absolue, au risque α choisi (par exemple $\alpha=0.05$).

Si $p\text{-value} \leq \alpha$

Alors on rejette l'hypothèse $\hat{\beta}_0 = 0$

Et $\hat{\beta}_0$ est significativement différent de zéro au niveau α .

La valeur que doit atteindre le test de *Student* pour que l'on puisse rejeter l'hypothèse nulle dépend du nombre d'observations et du niveau de confiance recherchée (de 90% à 99% en général). En pratique la valeur critique oscille le plus souvent autour de 2.

I.2.3.2. Validation interne : *Y-Randomization*

La qualité du modèle et l'influence des composés de la série d'apprentissage sur le modèle est estimée par des techniques appelées validation interne.

La technique de *Y-randomization* est une méthode de vérification de la crédibilité du modèle QSPR obtenu [24]. Le principe de *Y-randomization* est de mélanger la propriété expérimentale d'une manière aléatoirement pour la série d'apprentissage en utilisant les mêmes descripteurs et obtenir des modèles QSPR de mauvaise qualité.

I.2. 3. 3. Validation externe

La validation externe consiste à estimer la propriété d'une série de composés dite série de test qui ne sont pas inclus dans la série d'étude. Le principe de cette technique est basé sur le calcul des coefficients de corrélation R^2 (test). Ces paramètres sont appelés « critères de validation externe » ou « critères de Trophsa » (*Trophsa criteria*) [25].

Critères de validation Externe (série de test)

- $R^2 > 0.7$ (critère 1)
- $R^2_{cv} > 0.6$ (critère 2)
- $\frac{R^2 - R_0^2}{R^2} < 0.1$ et $0.85 \leq k \leq 1.15$ (critère 3)
- $\frac{R^2 - R_0'^2}{R^2} < 0.1$ et $0.85 \leq k' \leq 1.15$ (critère 4)
- $|R^2 - R_0^2| \leq 0.3$ (critère 5)

Avec

R^2 Coefficient de corrélation pour la série de test.

R_0^2 coefficient de corrélation entre les valeurs prédites et expérimentales pour la série de test.

$R_0'^2$ coefficient de corrélation entre les valeurs expérimentales et prédites pour la série de test.

k : est la constante de la droite (à l'origine) de corrélation (valeurs prédites en fonction des valeurs expérimentales)

k' : est la constante de la droite (à l'origine) de corrélation (valeurs expérimentales en fonction des valeurs prédites)

I.2.3.4. Domaine d'applicabilité

Méthode de "leverage" : cette méthode de détecter le domaine d'applicabilité [26] ou l'intervalle dans lequel nos composés pourront être prédits avec confiance. Cette technique est basée sur la variation des résiduels standardisés de la variable dépendante avec "leverage" (la distance entre les valeurs des descripteurs et leurs moyennes). Le principe est que le leverage d'un composé doit être inférieur à la valeur de seuil $h^* = 3p/n$ (ou p est le nombre de descripteurs plus 1 et n le nombre de points). Dans le cas contraire le composé sera considéré en dehors du domaine d'applicabilité du modèle obtenu.

Bibliographie

- [1] L.P. Hammett, *J. Am. Chem. Soc.* 59 (1937) 96.
- [2] C. Hansch, *Acc. Chem. Res.* 2 (1969) 232.
- [3] C. Hansch, A. Leo, R.W. Taft, *Chem. Rev.* 91 (1991) 165.
- [4] C. Hansch, T. Fujita, *J. Am. Chem. Soc.* 86 (1964) 1616.
- [5] S.M. Free, J.W. Wilson, *J. Med. Chem.* 7 (1964) 7 395.
- [6] J. Ghasemi, S. Saaidpour, S.D. Brown, *J. Mol. Struct. (Theochem)* 807 (2007) 27.
- [7] P. Geladi, B.R. Kowalski, *Anal. Chim. Acta* 185 (1986) 1.
- [8] M. Karelson, "Molecular Descriptors in QSAR/QSPR", Wiley, New York, 2000.
- [9] H. Wiener, *J. Am. Chem. Soc.* 69 (2002) 17.
- [10] M. Randic, *J. Am. Chem. Soc.* 97 (2002) 6609.
- [11] L.B. Kier, L.H. Hall, *J. Pharm. Sci.* 70 (1981) 583.
- [12] A.T. Balaban, *Chem. Phys. Lett.* 89 (1982) 399.
- [13] S.C. Basak, D.K. Harriss, V.R. Magnuson, *J. Pharm. Sci.* 73 (1984) 429.
- [14] G. Fayet, P. Raybaud, H. Toulhoat, de Bruin, *J. Mol. Struct. (Theochem)* 903 (2009) 100.
- [15] R.S. Mulliken, *J. Chem. Phys.* 23 (1955) 1833.
- [16] A.E. Reed, F. Weinhold, *J. Chem. Phys.* 83 (1985) 1736.
- [17] A.E. Reed, L.A. Curtiss, F. Weinhold, *Chem. Rev.* 88 (1988) 899.
- [18] (a) R. G. Parr, W. Yang, "Density Functional Theory of Atoms and Molecules", Oxford University Press, Oxford, UK, 1989.
(b) P. Geerlings, F. De Proft, W. Langenaeker, *Chem. Rev.* 103 (2003) 1793.
(b) H. Chermette, *J. Comput. Chem.* 20 (1999) 129.
- [19] S. Chatterjee and A. S. Hadi, "Regression Analysis by example", fourth Edition, a John Wiley & Sons, Inc., Hoboken, New Jersey, 2006.
- [20] S. Weisberg, "Applied Linear Regression", thirth Edition, John Wiley & Sons, Inc., Hoboken, New Jersey 2005.

- [21] A. C. Rencher, G. B. Schaalje, “Linear Models in Statistics”, Second Edition, John Wiley & Sons, Inc., Publication, 2008.
- [22] P. A. Cornillon, E.M.atzner-Løber, “Régression théorie et Applications ”, Springer-Verlag France, Paris, 2007.
- [23] M. Lejeune, “Statistiques : la théorie et ses applications”, Springer-Verlag, Paris, 2004
- [24] L. He, P.C. Jurs, *J. Mol. Graph. Model.*, 23 (2005) 503.
- [25] A. Golbraikh, A. Tropsha, Beware of q^2 ! *J. Mol. Graph. Model.* 20 (2002) 269.
- [26] T. Netzeva, et *al. Altern. Lab. Anim.* 33 (2005) 1.

CHAPITRE III :
RESULTATS ET DISCUSSIONS

I. Introduction

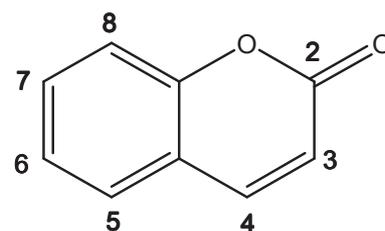
Les constantes d'acidité (pK_a) jouent un rôle essentiel dans la compréhension de nombreuses réactions fondamentales en chimie et en biochimie [1-5]. Lors de la comparaison des composés, les valeurs de pK_a permettent aux chercheurs de comparer le caractère des acides et des bases, les variations de l'énergie libre de Gibbs et les constantes d'équilibre des réactions d'ionisation. La connaissance des valeurs de pK_a nous permet, dans une réaction acide-base, de prédire facilement le sens privilégié de l'équilibre acide-base ainsi que les concentrations des différentes espèces. Par exemple dans les réactions de substitution nucléophile, Les valeurs pK_a permettent de prédire le pouvoir nucléophile et les meilleurs groupes sortants [5].

Les méthodes QSPR représentent parmi les méthodes les plus efficaces dans la prédiction des propriétés physico-chimiques. Plusieurs travaux ont été publiés sur l'élaboration des modèles pour prédire les valeurs aqueuses de pK_a des composés organiques [6-8,9,10] tels que l'imidazole [11] et les benzimidazoles [12] avec la méthode DFT/B3LYP/6-31+G(d,p).

En utilisant la série d'apprentissage de 20 hydroxy-coumarines (voir le **Tableau 1**), des descripteurs ont été calculés et leurs corrélations avec des valeurs aqueuses expérimentales de pK_a ont été effectuées. Des régressions linéaires simples (SLR) et multiples (MLR) ont été élaborés afin de d'obtenir les meilleures modèles QSPR pour prédire les pK_a d'un nombre de d'hydroxy-coumarines non incluses dans la série d'étude.

Tableau 1: Les constantes d'acidité aqueuse expérimentales (pK_a) des dérivés de la coumarine..

Substituant	pK_a
1 3- hydroxy	7.60 ^a
2 4- hydroxy	5.10 ^a
3 6- hydroxy	9.40 ^a
4 7- hydroxy	8.30 ^a
5 8- hydroxy	8.50 ^a
6 4-Methyl-5-hydroxy	8.26 ^a
7 4-Methyl-6-hydroxy	9.14 ^a
8 4-Methyl-7-hydroxy	7.80 ^a
9 4-Methyl-6,8-difluoro-7-hydroxy	4.90 ^a
10 6,7-dihydroxy	7.13 ^b
11 3-(2-furyl)-7-hydroxy	7.41 ^a
12 3-(2-Thienyl)-7-hydroxy	7.71 ^a
13 3-Phenyl-7-hydroxy	7.80 ^a
14 3-(2-Benzoxazolyl)-7-hydroxy	6.84 ^a
15 3-(2-Benzothiazolyl)-7-hydroxy	7.02 ^a
16 4-Phenyl-7-hydroxy	7.74 ^b
17 3-(2-Benzothiazolyl)- 4-cyano-7-hydroxy	6.38 ^a
18 6-Methoxy-7-hydroxy	7.60 ^c
19 6-Methoxy-7,8-dihydroxy	7.11 ^b
20 4-Methyl-7,8-dihydroxy	7.73 ^b
21 4-Methyl-5,7-dihydroxy	7.55 ^b
22 4-Propyl-5,7-dihydroxy	7.74 ^b
23 4-Methyl-6,7-dihydroxy	6.80 ^a
24 7,8-dihydroxy	7.41 ^b



^a Ref. [13]

^b Ref. [14]

^c Ref. [15]

II. Méthodologie

II.1. Calcul des descripteurs théoriques.

Les géométries d'équilibre des coumarines neutres ainsi que leurs formes anioniques ont été obtenues avec la méthode B3LYP/6-311++G(d,p) en utilisant le programme G09W [16].

L'énergie libre de Gibbs de déprotonation des hydroxy-coumarines étudiées en phase gazeuse ($\Delta G^\circ_{\text{gas}}$) en utilisant l'équation 1.



Avec $G^\circ_{\text{gas}}(\text{H}^+) [\text{exp}] = -6.28 \text{ kcal/mol}$ [17].

L'énergie libre $\Delta G^\circ_{\text{aq}}$ de déprotonation en phase aqueuse a été, également, calculée à partir d'un calcul single point d'énergie sur les géométries précédentes en moyen des modèles de solvation PCM/UAHF, CPCM/UAKS et SMD. Ces modèles nous permettent d'obtenir les énergies libres de solvation $\Delta G^\circ_{\text{solv}}$ pour chaque espèce. Ces dernières seront utilisées dans le calcul des énergies libre en phase gazeuse ($G^\circ_{\text{aq}}(\text{X})$) suivant l'équation 4.



$$\Delta G^\circ_{\text{aq}} = G^\circ_{\text{aq}}(\text{A}^-) + G^\circ_{\text{aq}}(\text{H}^+) - G^\circ_{\text{aq}}(\text{AH}) \quad (3)$$

$$G^\circ_{\text{aq}}(\text{X}) = G^\circ_{\text{gas}}(\text{X}) + \Delta G^\circ_{\text{solv}}(\text{X}) \quad (4)$$

$\Delta G^\circ_{\text{solv}}(\text{H}^+) = -265.9 \text{ kcal/mol}$ [18].

Chapitre III: Résultats et Discussion

Afin de convertir l'état standard en phase gazeuse (1 atm) en solution (1 mole. dm⁻³) la formule suivante peut être utilisée:

$$G^{\circ}_{\text{gas}} (1 \text{ mole. dm}^{-3}) = G^{\circ}_{\text{gas}} (1 \text{ atm}) + RT \ln (V_m) \quad (5)$$

à T=298.15 K, V_m = 24.44 L, R=1.987207 cal/mol.K, et RTln (V_m)=1.8929329 kcal/mol.

Dans notre étude QSPR, nous avons étudié 8 descripteurs de l'acide et 7 descripteurs de sa base conjuguée. Nous avons établi leurs corrélation avec les valeurs expérimentales de pK_a en phase aqueuses. Ces descripteurs sont

- les orbitales frontières : ε_{HOMO}, ε_{LUMO}

- La charge NPA [19] de l'hydrogène acide Q_H^{acid} et celle de l'atome d'oxygène portant la charge négative Q_O^{base}.

- l'énergie libre aqueuse de Gibbs (G_{aq})

-

- l'électronégativité (χ)

$$\chi = (\varepsilon_{\text{HOMO}} + \varepsilon_{\text{LUMO}})/2 \quad (6)$$

- la dureté (η)

$$\eta = (\varepsilon_{\text{LUMO}} - \varepsilon_{\text{HOMO}}) \quad (7)$$

- la mollesse (S)

$$S = \eta^{-1} \quad (8)$$

Nous avons engendré d'autres descripteurs calculés en utilisant les descripteurs précédemment cités. Ces descripteurs sont:

- ΔQ_O: la variation de la charge NPA de l'oxygène de la base conjugué,

- $\Delta\chi$, $\Delta\eta$ et ΔS : la variation de l'électronégativité, de la dureté et de la mollesse lors de déprotonation des hydroxy-coumarines.
- ΔG_{aq} : L'énergie libre aqueuse de la dissociation de l'acide

Nous avons établi les corrélations de tout les descripteurs avec les valeurs expérimentales de pK_a en phase aqueuses. Les résultats (R^2 , R^2_{adj} , et SD) sont donnés dans les **Tables 1-12** (*voir annexe n°1*).

Dans ce travail les régressions linéaires simples et multiples ont été effectuées en utilisant le programme d'analyse statistique de données MINITAB [20].

II.2. Résultats et discussions :

II.2.1. La phase gazeuse :

L'analyse des résultats obtenus en phase gazeuse, montre la corrélation (SLR) des différents descripteurs avec les pK_a expérimentaux présente de mauvais coefficients de corrélation. La meilleure corrélation obtenue dans cette phase est celle de l'énergie libre de dissociation ΔG_{gas} ($R^2 = 0.42$) et de la variation de la mollesse ($R^2 = 0.41$). A l'issue de ces résultats, l'étude QSPR dans la phase gazeuse ne présente aucun intérêt d'où la nécessité de faire des calculs et une étude QSPR en présence de deux phases.

II.2.2. La phase aqueuse :

Les résultats obtenus de la régression linéaire simple en utilisant le modèle PCM/UAHF montrent que les descripteurs possédant les corrélation les plus importantes ($R^2 \geq 0.60$) avec les pK_a expérimentaux sont : ΔG_{aq} (0.91), $Q_{\text{O}}^{\text{base}}$ (0.63) et $Q_{\text{H}}^{\text{acid}}$ (0.65). Dans le cas du modèle CPCM/UAKS les mêmes descripteurs, avec des corrélation légèrement

Chapitre III: Résultats et Discussion

meilleures, sont retenues: ΔG_{aq} (0.92), Q_H^{acid} (0.75) et Q_O^{base} (0.64). Pour le modèle SMD les meilleurs coefficients de corrélation sont obtenus pour les descripteurs suivants : ΔG_{aq} (0.89) et Q_O^{base} (0.63).

A l'aide du programme MINTAB et en utilisant la technique "Best Subsets Analysis" nous avons élaborés pour chaque modèle de solvation des ensembles réduits de descripteurs. Les meilleures modèles QSPR obtenues sont présentés dans le tableau suivant :

Tableau 2: Les meilleures modèles QSPR obtenus par la régression linéaire multiple.

	R²	R²_{cv}	SD	F-test
B3LYP/6-311++G(d,p)-PCM/UAHF				
(9) $pK_a = 2.18 + 0.385 \Delta G_{aq-pcm}$	0.910	0.880	0.34	182.30
(10) $pK_a = 76.2 - 156 Q_H^{acid} - 17.2 Q_O^{base}$	0.825	0.710	0.49	40.02
B3LYP/6-311++G(d,p)-CPCM/UAKS				
(11) $pK_a = 2.21 + 0.386 \Delta G_{aq-cpcm}$	0.916	0.890	0.33	196.39
(12) $pK_a = 98.3 - 194 Q_H^{acid} - 15.2 Q_O^{base}$	0.880	0.770	0.41	62.60
B3LYP/6-311++G(d,p)-SMD				
(13) $pK_a = 3.90 + 0.338 \Delta G_{aq-smd}$	0.890	0.850	0.38	145.26

Ces modèles ont été validés après vérification des paramètres statistiques de la régression tels que les coefficients R^2 , les testes de Fischer (F-test), les déviations standards SD, les t-student ainsi que la probabilité $P \leq 0.05$.

Dans le cas de la méthode B3LYP/6-311++G (d,p)-PCM/UAHF, seulement deux modèles ont été retenus. Le modèle à un seul descripteur inclut l'énergie de déprotonation ΔG_{aq-pcm} des coumarines avec $R^2 = 0.91$, $SD = 0.34$, et $F\text{-test} = 182.30$ et le deuxième modèle, moins satisfaisant, a été trouvé est celui à 2 variables comporte les descripteurs Q_O^{base} et Q_H^{acid} ($R^2 = 0.825$, $SD = 0.49$, $F\text{-test} = 40.02$).

Dans le cas des calculs B3LYP/6-311++G (d,p)-CPCM/UAKS des modèles identiques ont été obtenus à ceux trouvés avec la méthode B3LYP/6-311++G (d,p)-PCM/UAHF. Le modèle à une variable inclut l'énergie de déprotonation $\Delta G_{\text{aq-cpcm}}$ des coumarines avec $R^2 = 0.916$, $SD=0.33$, et $F\text{-test}=196.39$. Le deuxième modèle est celui à 2 variables comporte les descripteurs $Q_{\text{O}}^{\text{base}}$ et $Q_{\text{H}}^{\text{acid}}$ ($R^2 = 0.880$, $SD=0.41$, $F\text{-test}=62.60$). Les modèles QSPR obtenus par cette méthode (B3LYP/6-311++G (d,p)-CPCM/UAKS) sont légèrement meilleurs que celui obtenu avec le modèle de solvation PCM/UAHF. Contrairement aux méthodes précédentes, un seul modèle QSPR est validé avec le niveau de calcul B3LYP/6-311++G (d,p)-SMD. Ce Modèle inclut un seul descripteur c'est $\Delta G_{\text{aq-smd}}$ avec les paramètres $R^2 = 0.890$, $SD=0.38$, $F\text{-test}=145.26$.

Le pouvoir prédictif et la stabilité interne des modèles 9, 11 et 13 sont justifiés par les valeurs considérables des coefficients de régression R^2_{cv} de *leave-one-out* et par les faibles valeurs des déviations standards SD (voir Tableau 2)

Dans les **Figures 1-3** nous avons présenté la corrélation linéaire entre les valeurs de pK_a prédites par les modèles eq. 9, 11 et 13 avec les valeurs de pK_a expérimentales aqueuses.

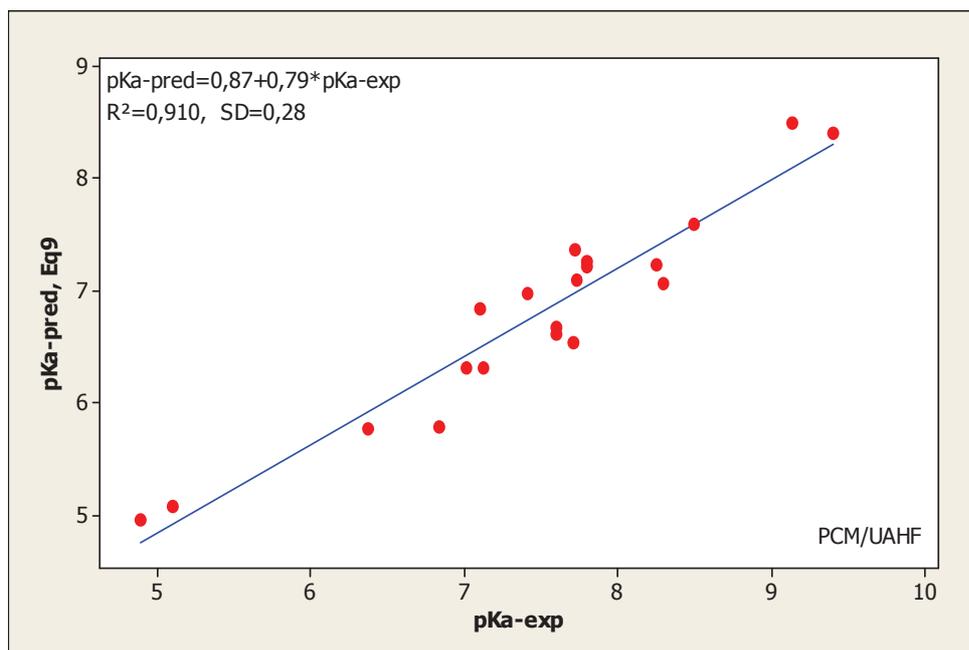


Figure 1. Corrélation des valeurs de pK_a prédites par Eq.9 (B3LYP/6-311++G(d.p)/PCM/UAHF) avec les valeurs pK_a expérimentales aqueuses.

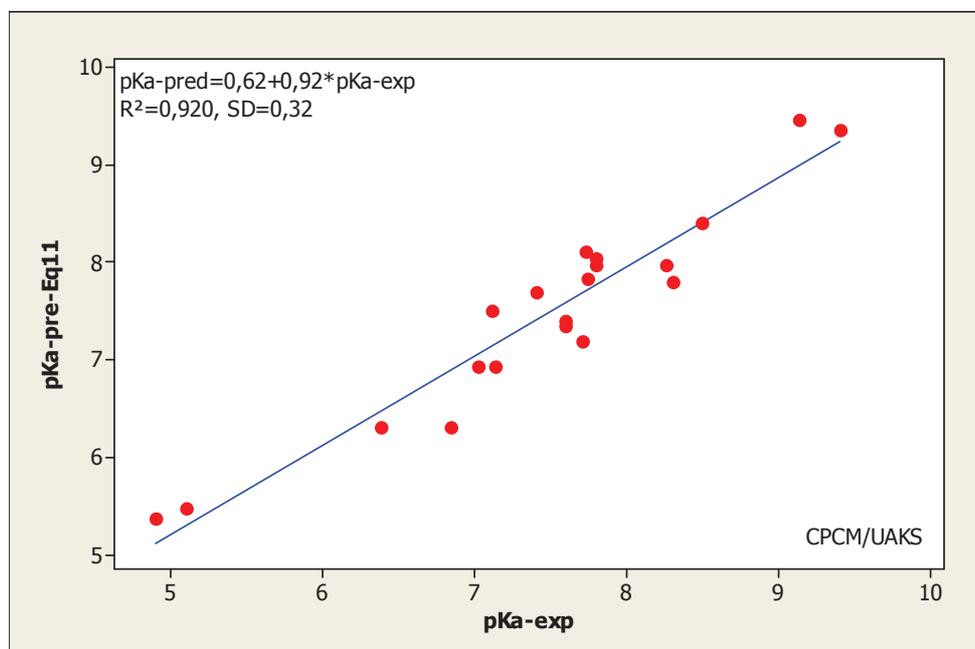


Figure 2. Corrélation des valeurs de pK_a prédites par Eq.11 (B3LYP/6-311++G(d.p)/CPCM/UAKS) avec les valeurs pK_a expérimentales aqueuses

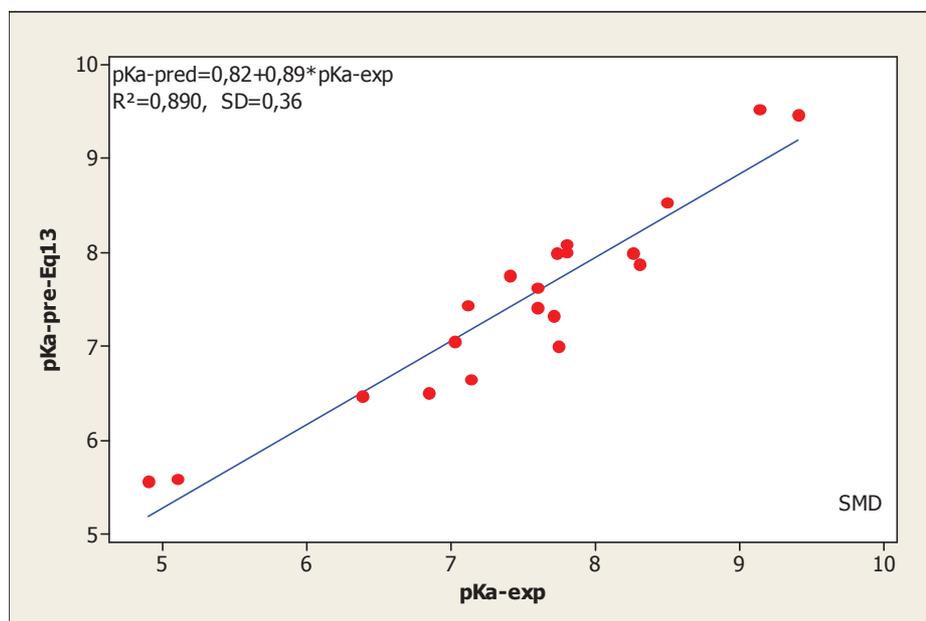


Figure 3. Corrélation des valeurs de pK_a prédites par Eq.13 (B3LYP/6-311++G(d,p)/SMD) avec les valeurs pK_a expérimentales aqueuses

Afin de finaliser cette validation interne nous avons effectué la procédure Y-randomisation pour 7 itérations. Les résultats, montrent que toutes les valeurs de coefficients de corrélation R^2 des modèles aléatoires sont plus faibles que le R^2 des modèles QSPR trouvés. Ce résultat exclu toute chance dans l'obtention des modèles QSPR. Dans le **Tableau 3** et **4** nous avons donné les résultats des Y-randomisation des modèles présentes par l'équation 9-13.

Tableau. 3. Valeurs de R^2 et R^2_{cv} des les différentes Y-randomisations des modèles (Eqs9,11,13)

<i>Itération</i>	R^2 (Eq.9)	R^2_{cv} (Eq.9)	R^2 (Eq.11)	R^2_{cv} (Eq.11)	R^2 (Eq.13)	R^2_{cv} (Eq.13)
1	0.190	0.00	0.031	0.00	0.076	0.00
2	0.001	0.00	0.007	0.00	0.009	0.00
3	0.043	0.00	0.009	0.00	0.155	0.00
4	0.174	0.00	0.004	0.00	0.064	0.00
5	0.011	0.00	0.022	0.00	0.054	0.00
6	0.032	0.00	0.080	0.00	0.001	0.00
7	0.022	0.00	0.067	0.00	0.086	0.00

Tableau. 4. Valeurs de R^2 et R^2_{cv} des les différentes Y-randomisations des modèles Eq.10 et 12

<i>Itération</i>	R^2 (Eq.10)	R^2_{cv} (Eq.10)	R^2 (Eq.12)	R^2_{cv} (Eq.12)
1	0.133	0.00	0.104	0.00
2	0.035	0.00	0.005	0.00
3	0.002	0.00	0.001	0.00
4	0.348	0.00	0.018	0.00
5	0.025	0.00	0.069	0.00
6	0.092	0.00	0.135	0.00
7	0.016	0.00	0.048	0.00

II.2.3. Validation externe des modèles.

Après avoir validé les modèles QSPR obtenus par l'analyse des paramètres statistiques ainsi que Y- randomisation, l'étape suivante consiste à utiliser ces modèles pour prédire la constante d'acidité de quelques hydroxy-coumarines (série de test) non incluses dans la série d'apprentissage de départ. Cette étape est appelée la validation externe. L'analyse des coefficients R^2 montrent que les modèles calculés avec B3LYP/6-311++G (d,p)-CPCM/UAKS (Eq.9, $R^2=0.87$) et B3LYP/6-311++G(d,p)-PCM/UAHF (Eq.11, $R^2=0.84$) présentent un pouvoir prédictif considérable. Par contre les autres modèles QSPR Eqs 10, 12 et 13 la prédiction des constantes d'acidité de la série de test est faible avec des $R^2 < 0.45$.

Tableau. 5. Validation interne et externe des modèles QSPR retenus

	Série de test			Critères Tropsha			
	n	R^2	SD	R^2_0	k	$R^2 - R^2_0 / R^2$	$ R^2 - R^2_0 $
PCM/UAHF							
Eq.(9)	4	0.838	0.10	0.838	0.46	0.00	0.00
				(0.838)	(1.83)	0.00	0.00
Eq.(10)	4	0.415	0.20	0.415	0.35	0.00	0.00
				(0.415)	(1.19)	(0.00)	(0.00)
CPCM/UAKS							
Eq.(11)	4	0.871	0.10	0.871	0.52	0.00	0.00
				0.871	(1.68)	(0.00)	(0.00)
Eq.(12)	4	0.330	0.22	0.330	0.31	0.00	0.00
				(0.330)	(1.06)	(0.00)	0.00
SMD							
Eq.(13)	4	0.144	0.180	0.144	0.15	0.00	0.00
				0.144	0.96	0.00	0.00

Les valeurs entre par anthèses sont les critères de Tropsha avec R^2_0

Nous avons également examiné la rigidité de nos modèles obtenus en vérifiant les critères de Tropsha. Pour les modèles obtenus avec la méthode PCM-Eq.9, les cinq critères de Tropsha ne sont pas tous vérifiés, car la valeur de k et k' présente une valeur de l'ordre 0.46 et 1.83 légèrement en dehors de $0.85 \leq k(k') \leq 1.15$. Les mêmes constatations sont faites avec le modèle CCPM/UAKS-Eq.11 avec une légère amélioration des constantes k et k' (0.52 et 1.63 respectivement). Par contre, les modèles PCM-Eq.10, CPCM-Eq.12 et SMD-Eq.13 présentent des mauvais critères. Ces constantes peuvent être améliorées en augmentant le nombre de composés de la série d'apprentissage et surtout celui de la série de test.

Finalement, nous pouvons dire que les modèles QSPR à un descripteur avec les modèles de solvation CPCM/UAKS et PCM/UAHF sont rigides et présentent un pouvoir prédictif des constantes d'acidité pK_a des hydroxy-coumarines. Le descripteur, ΔG_{aq} , est considéré comme pertinent et ce résultat est en très bon accord avec les résultats trouvés très récemment [21] où il a été démontré que la grande acidité de quelques hydroxy-coumarines est principalement liée aux grandeurs thermodynamiques ΔG , ΔH , et ΔS .

II.2.4. Domaine d'applicabilité

Dans les figures 4 et 5 nous avons présenté le domaine d'applicabilité des meilleurs modèles QSPR validés, à savoir Eq9 et 11. Cette procédure nous permet de localiser les valeurs aberrantes de la série d'études. On remarque à partir de ces figures qu'il n'y a aucun composé de la série possédant un h supérieur à la valeur critique ($h^* = 0.30$) et ceci est valable pour les trois modèles de solvation.

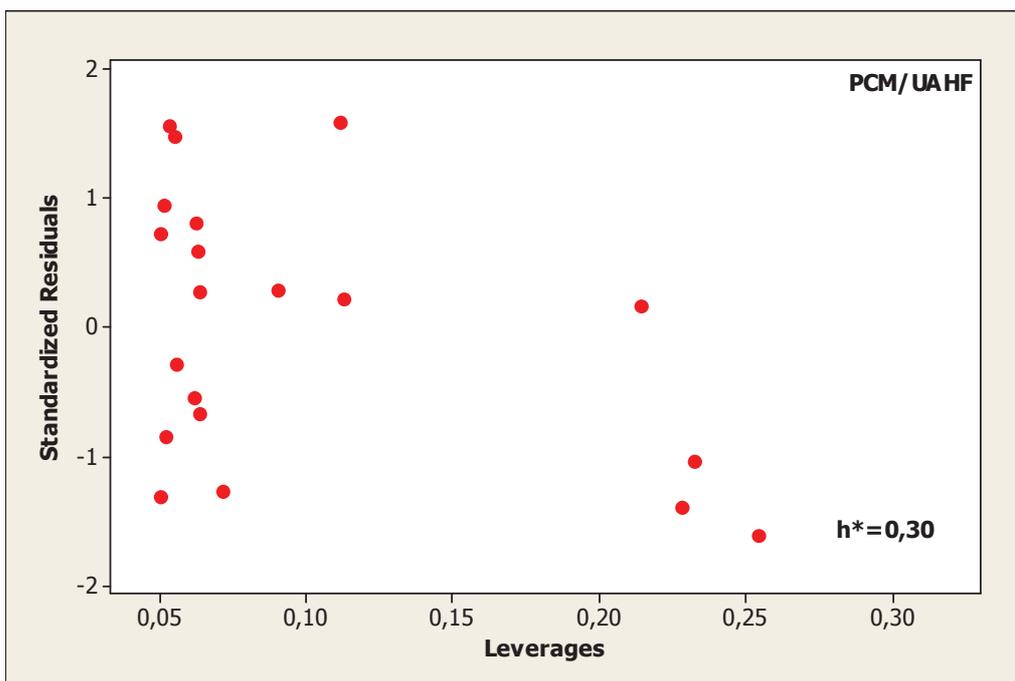


Figure 4: Domaine d'applicabilité du modèle QSPR (Eq.9)

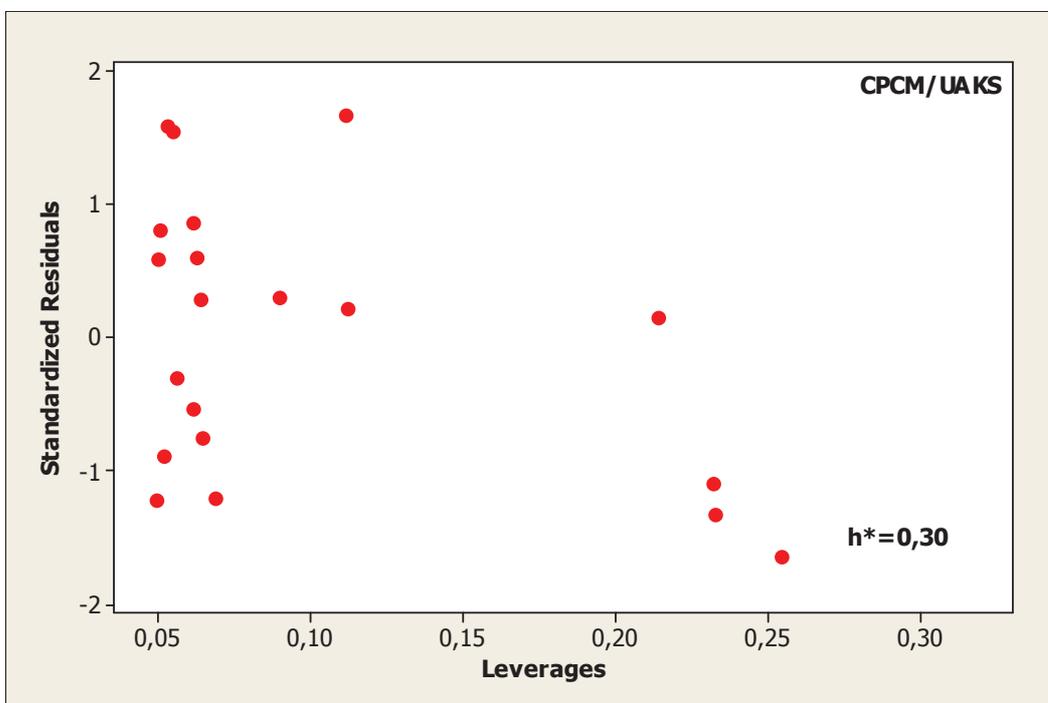


Figure 5: Domaine d'applicabilité du modèle QSPR (Eq.11)

Bibliographie

- [1] K. Alongi and G. Shields, *Annu. Rep. Comput. Chem.* 6 (2010), 113.
- [2] R.D. Rossi, *J. Chem. Educ.* 90 (2013), pp. 183–190.
- [3] W.C. Kreye and P.G. Seybold, *Int. J. Quantum Chem.* 109 (2009), 3679.
- [4] P.G. Seybold and W.C. Kreye, *Int. J. Quantum Chem.* 112 (2012), 3769.
- [5] R.D. Rossi, *J. Chem. Educ.* 90 (2013), 183.
- [6] V. Consonni, R. Todeschini, M. Pavan, P. Gramatica, *J. Chem. Inf. Comput. Sci.* 42 (2002) 693.
- [7] G. Krenkel, E. A. Castro, A. A. Toropov, *J. Mol. Struct. (THEOCHEM)* 542 (2001) 107.
- [8] (a) L. B. Kier, L. H. Hall, “Molecular Connectivity in Structure–Activity Analysis”, RSP-Wiley, Chetster, UK, 1986.
(b) J. Ghasemi, S. Shahmirani, E. V. Farahani, *Ann. Chim.* 96 (2006) 327.
- [9] J. Ghasemi, S. Shahmirani, E. V. Farahani, *Ann. Chim.* 96 (2006) 327.
- [10] J. Ghasemi, S. Saaïdpour, S.D. Brown, *J. Mol. Struct.:(THEOCHEM)* 805 (2007) 27.
- [11] E. Soriano, S. Cerdán, P. Ballesteros, *J. Mol. Struct. (THEOCHEM)* 684 (2004) 121.
- [12] T. .N. Brown, N. Mora-Diez, *J. Phys. Chem. B* 110 (2006) 20546.
- [13] A.M. Ferrari et al. / *European Journal of Medicinal Chemistry* 42 (2007) 1028.
- [14] P.M. Nowak et al., *J. Chromatogr. A* 1446 (2016) 149.
- [15] Gerald J. Smith et al. *Photochemistry and Photobiology*, 87 (2011) 45.
- [16] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Jr. Montgomery, 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. and Pople, Gaussian 03, revision E.01; Gaussian, Inc.: Wallingford CT, 2009.
- [17] M. D. Liptak, K. C. Grass, P. G. Seybold, *J. Am. Chem. Soc.* 124 (2002) 6421.
- [18] M.D. Tissandier, K.A. Cowen, W.Y. Feng, E. Gundlach, M.H. Cohen, A.D. Earhart, J.V. Coe, and T.R. Tuttle, *J. Phys. Chem. A* 102 (1998), pp. 7787–7794.
- [19] (a) A. E. Reed, F. Weinhold, *J. Chem. Phys.* 78 (1983) 4066.
(b) A.E. Reed, F. Weinhold, *Isr. J. Chem.* 31 (1991) 277.
- [20] MINITAB State College, Minitab, Inc, PA 2006.
- [21] P. M. Nowak, F. Sagan and M. P. Mitoraj, *J. Phys. Chem. B* (accepted) (2017), DOI: 10.1021/acs.jpcc.7b01849

CONCLUSION GENERALE

Dans ce travail, nous avons établi des modèles QSPR prédictifs reliant les constantes d'acidité (pK_a) expérimentales aqueuses des hydroxy-coumarines à des descripteurs théoriques. Ce travail est basé sur la régression linéaire multiple (MLR), en utilisant le logiciel MINITAB.

Dans la phase gazeuse, les calculs B3LYP/6-311++G(d,p) ont montré qu'il existe des mauvaises corrélations entre l'énergie libre de déprotonation ainsi que les autres descripteurs quantiques tels que la charge de l'hydrogène acide, la charge de l'oxygène de la base conjugué.

L'effet de solvation a été considéré avec l'utilisation de trois méthodes PCM/UAHF, CPCM/UAKS et SMD. Dans le cas de la méthode B3LYP/6-311++G(d,p)-PCM/UAHF un modèle QSPR satisfaisant a été obtenu comportant l'énergie libre aqueuse de déprotonation ($\Delta G_{\text{aq-pcm}}$) des hydro-coumarines ($R^2 = 0.91$, $SD = 0.34$, et $F\text{-test} = 182.30$). Un deuxième modèle QSPR, moins acceptable que le modèle précédent, comporte les charges NPA $Q_{\text{O}}^{\text{base}}$ et $Q_{\text{H}}^{\text{acid}}$ ($R^2 = 0.825$, $SD = 0.49$, $F\text{-test} = 40.02$).

Dans le cas des calculs B3LYP/6-311++G(d,p)-CPCM/UAKS les mêmes modèles ont été trouvés avec des meilleurs paramètres statistiques que dans le cas B3LYP/6-311++G(d,p)-PCM/UAHF. Le modèle QSPR à un seul descripteur qui comporte l'énergie libre aqueuse de déprotonation ($\Delta G_{\text{aq-cpcm}}$) présente les paramètres $R^2 = 0.916$, $SD = 0.33$, et $F\text{-test} = 196.39$. Des paramètres statistiques moins satisfaisant ont été obtenus pour le deuxième modèle avec les descripteurs $Q_{\text{O}}^{\text{base}}$ et $Q_{\text{H}}^{\text{acid}}$ ($R^2 = 0.880$, $SD = 0.41$, $F\text{-test} = 62.60$).

Contrairement aux méthodes précédentes, un seul modèle QSPR est validé avec le niveau de calcul B3LYP/6-311++G(d,p)-SMD comportant le descripteur, $\Delta G_{\text{aq-smd}}$, avec les paramètres $R^2 = 0.890$, $SD = 0.38$, $F\text{-test} = 145.26$. Tous ces modèles QSPR peuvent décrire environ 90% de la variance de la constante d'acidité aqueuse expérimentale des hydroxy-coumarines.

Conclusion Générale

Les modèles QSPR obtenus ont été validés au moyen de la prédiction d'une série de test d'hydroxy-coumarines qui ne sont pas inclus dans la série d'apprentissage de départ. Finalement, seulement les modèles à un seul descripteur comportant l'énergie de déprotonation ont été validés et ils sont considérés comme fiables et prédictifs de l'acidité des dérivés des hydroxy-coumarines pour lesquels les données expérimentales sont inconnues.

Comme perspectives à ce travail, nous envisageons :

- L'élaboration des modèles QSPR plus fiables avec d'autres descripteurs quantiques.
- L'élaboration des modèles QSPR plus prédictifs en utilisant des méthodes quantiques plus précises CBS-QB3....
- Etablir des modèles QSPR plus fiables en utilisant d'autres méthodes de solvation.

ANNEXE N°1

Table 1: Descripteurs B3LYP/6-311++G(d,p) et leurs corrélation avec les valeurs aqueuses pK_a

Acid	ϵ_{HOMO}	ϵ_{LUMO}	S	η	χ	Q _O	Q _H	G _{gas}
1	-0,24066	-0,07911	6,19003	0,16155	-0,15989	-0,65974	0,49513	-572,312465
2	-0,25069	-0,07634	5,73559	0,17435	-0,16352	-0,65387	0,47901	-572,313025
3	-0,23892	-0,08559	6,52188	0,15333	-0,16226	-0,67156	0,46892	-572,306264
4	-0,24190	-0,07857	6,12257	0,16333	-0,16024	-0,65861	0,4703	-572,309139
5	-0,24492	-0,08260	6,16067	0,16232	-0,16376	-0,64955	0,46939	-572,302161
6	-0,24669	-0,07394	5,78871	0,17275	-0,16032	-0,6719	0,4735	-611,608097
7	-0,23573	-0,08050	6,44205	0,15523	-0,15812	-0,67345	0,46873	-611,610381
8	-0,23860	-0,07411	6,07940	0,16449	-0,15636	-0,66127	0,47155	-611,613559
9	-0,25108	-0,08818	6,13874	0,16290	-0,16963	-0,63974	0,48701	-810,151803
10	-0,23145	-0,07905	6,56168	0,15240	-0,15525	-0,69401	0,48479	-647,553286
11	-0,21512	-0,08394	7,62311	0,13118	-0,14953	-0,65942	0,47012	-801,161474
12	-0,21822	-0,08773	7,66342	0,13049	-0,15298	-0,65876	0,4703	-1124,142815
13	-0,22868	-0,08073	6,75904	0,14795	-0,15471	-0,65934	0,47011	-803,345433
14	-0,22827	-0,09769	7,65814	0,13058	-0,16298	-0,6542	0,47149	-970,862989
15	-0,22772	-0,09838	7,73156	0,12934	-0,16305	-0,65478	0,47154	-1293,852106
16	-0,23855	-0,07905	6,26959	0,15950	-0,15880	-0,65963	0,47002	-803,34515
17	-0,23841	-0,12537	8,84643	0,11304	-0,18189	-0,65875	0,47186	-1386,109525
18	-0,22536	-0,07580	6,68628	0,14956	-0,15058	-0,64572	0,47092	-686,83059
19	-0,22714	-0,08033	6,81153	0,14681	-0,15374	-0,64638	0,48844	-762,077282
20	-0,23692	-0,07781	6,28496	0,15911	-0,15737	-0,66093	0,48884	-686,860414
R²	0.036	0.089	0.022	0.010	0.140	0.188	0.175	0,118
SD	1.12	1.09	1.13	1.13	1.06	0.03	1.03	1,07
R ² _{adj}	0.000	0.039	0.000	0.000	0.092	0.142	0.129	0,069

Table 2 : Descripteurs B3LYP/6-311++G(d,p) et leurs corrélations avec les valeurs aqueuses pK_a

Base	ϵ_{HOMO}	ϵ_{LUMO}	S	η	χ	Q _O	G _{gas}
1	-0,04144	0,07929	8,28294542	0,12073	0,018925	-0,69604	-571,782675
2	-0,06352	0,07841	7,04572677	0,14193	0,007445	-0,71773	-571,802413
3	-0,0394	0,05455	10,6439596	0,09395	0,007575	-0,7732	-571,775702
4	-0,05567	0,07166	7,85360873	0,12733	0,007995	-0,71478	-571,791003
5	-0,04024	0,06213	9,76848686	0,10237	0,010945	-0,73501	-571,77285
6	-0,05098	0,07705	7,81066937	0,12803	0,013035	-0,73715	-611,090973
7	-0,03847	0,05741	10,4297038	0,09588	0,00947	-0,77294	-611,078876
8	-0,05446	0,07093	7,97511763	0,12539	0,008235	-0,71895	-611,092583
9	-0,06588	0,05808	8,06711843	0,12396	-0,0039	-0,68176	-809,643358
10	-0,05715	0,06864	7,94975753	0,12579	0,005745	-0,75348	-647,048567
11	-0,0613	0,05288	8,75810124	0,11418	-0,00421	-0,69854	-800,64771
12	-0,06479	0,04439	9,15918666	0,10918	-0,0102	-0,69422	-1123,633009
13	-0,06484	0,05037	8,6798021	0,11521	-0,007235	-0,69867	-802,831603
14	-0,07986	0,03421	8,76654686	0,11407	-0,022825	-0,67693	-970,362156
15	-0,08093	0,03055	8,97021887	0,11148	-0,02519	-0,67648	-1293,349798
16	-0,06045	0,04824	9,20047842	0,10869	-0,006105	-0,70996	-802,827357
17	-0,09408	0,0006	10,5618927	0,09468	-0,04674	-0,65677	-1385,615427
18	-0,05092	0,07069	8,22300798	0,12161	0,009885	-0,69056	-686,312382
19	-0,04679	0,06845	8,67754252	0,11524	0,01083	-0,66765	-761,552318
20	-0,04886	0,06622	8,68960723	0,11508	0,00868	-0,69741	-686,333138
R²	0.365	0.021	0.228	0.245	0.135	0.362	0,118
SD	0.91	1.13	1.00	0.99	1.06	0.91	1,07
R ² _{adj}	0.329	0.000	0.185	0.203	0.086	0.327	0,069

Table 3 : Descripteurs B3LYP/6-311++G(d,p) et leurs corrélations avec les valeurs aqueuses pK_a

Acid/ Base	ΔS	$\Delta \eta$	$\Delta \chi$	ΔQ_O	ΔG_{gaz}^c
1	2,09291	-0,04082	0,17881	-0,03630	326,17
2	1,31014	-0,03242	0,17096	-0,06386	314,13
3	4,12208	-0,05938	0,16983	-0,10164	326,65
4	1,73103	-0,03600	0,16823	-0,05617	318,86
5	3,60782	-0,05995	0,17471	-0,08546	325,87
6	2,02196	-0,04472	0,17335	-0,06525	318,22
7	3,98765	-0,05935	0,16759	-0,09949	327,24
8	1,89572	-0,03910	0,16459	-0,05768	320,64
9	1,92838	-0,03894	0,16573	-0,04202	312,77
10	1,38808	-0,02661	0,16100	-0,05947	310,44
11	1,13499	-0,01700	0,14532	-0,03912	316,11
12	1,49576	-0,02131	0,14278	-0,03546	313,63
13	1,92076	-0,03274	0,14747	-0,03933	316,15
14	1,10841	-0,01651	0,14016	-0,02273	308,00
15	1,23866	-0,01786	0,13786	-0,02170	308,92
16	2,93089	-0,05081	0,15270	-0,05033	318,64
17	1,71547	-0,01836	0,13515	0,00198	303,77
18	1,53673	-0,02795	0,16047	-0,04484	318,90
19	1,86602	-0,03157	0,16457	-0,02127	323,14
20	2,40465	-0,03157	0,16605	-0,03648	324,59
R²	0.413	0.289	0.056	0.273	0.416
SD	0.87	0.96	1.11	0.97	0.87
R ² _{adj}	0.380	0.250	0.004	0.233	0,384

Table 4: Descripteurs B3LYP/6-311++G(d,p)/PCM/UAHF et leurs corrélations avec les valeurs aqueuses pK_a

Acid	ϵ_{HOMO}	ϵ_{LUMO}	S	η	χ	Q_O	Q_H
1	-0,23142	-0,07088	6,2289772	0,16054	-0,15115	-0,69802	0,53316
2	-0,24616	-0,06814	5,61734637	0,17802	-0,15715	-0,68328	0,54065
3	-0,22948	-0,0773	6,57116572	0,15218	-0,15339	-0,721	0,52851
4	-0,23237	-0,07186	6,23014142	0,16051	-0,152115	-0,70553	0,53207
5	-0,23571	-0,07638	6,27628193	0,15933	-0,156045	-0,70893	0,5329
6	-0,23406	-0,06936	6,07164542	0,1647	-0,15171	-0,7053	0,53208
7	-0,22802	-0,07334	6,46495992	0,15468	-0,15068	-0,72241	0,52854
8	-0,23123	-0,06804	6,12782646	0,16319	-0,149635	-0,70697	0,5314
9	-0,23886	-0,07541	6,11807892	0,16345	-0,157135	-0,68204	0,54073
10	-0,22143	-0,07139	6,66488936	0,15004	-0,14641	-0,71447	0,52937
11	-0,21093	-0,08002	7,63883584	0,13091	-0,145475	-0,70489	0,53195
12	-0,21352	-0,08274	7,64642912	0,13078	-0,14813	-0,70408	0,53203
13	-0,22306	-0,07568	6,78518116	0,14738	-0,14937	-0,70544	0,53192
14	-0,22469	-0,09208	7,54090943	0,13261	-0,158385	-0,69687	0,53405
15	-0,2245	-0,09347	7,63184004	0,13103	-0,158985	-0,69759	0,53406
16	-0,23227	-0,07542	6,37551801	0,15685	-0,153845	-0,70564	0,53186
17	-0,23142	-0,12134	9,08430233	0,11008	-0,17638	-0,69273	0,53558
18	-0,21985	-0,07102	6,71907546	0,14883	-0,145435	-0,69792	0,53378
19	-0,22034	-0,07283	6,77920141	0,14751	-0,146585	-0,69517	0,53305
20	-0,22525	-0,06918	6,4073813	0,15607	-0,147215	-0,70348	0,5319
R²	0.079	0.051	0.010	0.001	0.134	0.805	0.651
SD	1.09	1.11	1.13	1.14	1.06	0.795	0.67
R ² _{adj}	0.028	0.000	0.000	0.000	0.086	0.795	0.632

Table 5: Descripteurs B3LYP/6-311++G(d,p)/PCM/UAHF et leurs corrélations avec les valeurs aqueuses pK_a

Base	ϵ_{HOMO}	ϵ_{LUMO}	S	η	χ	Q _O
1	-0,18715	-0,04691	7,1306332	0,14024	-0,11703	-0,8549
2	-0,21885	-0,04405	5,7208238	0,1748	-0,13145	-0,81327
3	-0,18825	-0,06758	8,28706389	0,12067	-0,127915	-0,91602
4	-0,19431	-0,05737	7,30246823	0,13694	-0,12584	-0,85182
5	-0,19102	-0,0644	7,8976465	0,12662	-0,12771	-0,88959
6	-0,19047	-0,05428	7,34268302	0,13619	-0,122375	-0,84913
7	-0,18696	-0,06347	8,09782169	0,12349	-0,125215	-0,91691
8	-0,19339	-0,05442	7,19579765	0,13897	-0,123905	-0,85892
9	-0,2004	-0,06182	7,21604849	0,13858	-0,13111	-0,82344
10	-0,18938	-0,057	7,55401118	0,13238	-0,12319	-0,85575
11	-0,18478	-0,06576	8,40194925	0,11902	-0,12527	-0,84053
12	-0,18616	-0,06881	8,52151683	0,11735	-0,127485	-0,83729
13	-0,18951	-0,06378	7,95355126	0,12573	-0,126645	-0,84223
14	-0,19672	-0,07688	8,34445928	0,11984	-0,1368	-0,81142
15	-0,19615	-0,07895	8,53242321	0,1172	-0,13755	-0,81168
16	-0,19483	-0,06274	7,57059581	0,13209	-0,128785	-0,85115
17	-0,20278	-0,10749	10,4942806	0,09529	-0,155135	-0,78794
18	-0,18783	-0,05635	7,6057195	0,13148	-0,12209	-0,83487
19	-0,18305	-0,05907	8,06581707	0,12398	-0,12106	-0,82934
20	-0,18664	-0,05687	7,70594128	0,12977	-0,121755	-0,8545
R²	0.462	0.005	0.031	0.093	0.156	0.634
SD	0.83	1.14	1.12	1.09	1.05	0.69
R ² _{adj}	0.433	0.000	0.000	0.043	0.109	0.614

Table 6: Descripteurs B3LYP/6-311++G(d,p)/PCM/UAHF et leurs corrélationsavec les valeurs aqueuses pK_a

Acid/ Base	ΔS	$\Delta\eta$	$\Delta\chi$	ΔQ_O	ΔG_{aq}
1	0,90166	-0,0203	0,03412	-0,15688	13,25
2	0,10348	-0,00322	0,0257	-0,12999	8,66
3	1,71590	-0,03151	0,025475	-0,19502	18,61
4	1,07233	-0,02357	0,026275	-0,14629	14,61
5	1,62136	-0,03271	0,028335	-0,18066	16,16
6	1,27104	-0,02851	0,029335	-0,14383	15,08
7	1,63286	-0,03119	0,025465	-0,1945	18,87
8	1,06797	-0,02422	0,02573	-0,15195	15,06
9	1,09797	-0,02487	0,026025	-0,1414	8,30
10	0,88912	-0,01766	0,02322	-0,13661	12,35
11	0,76311	-0,01189	0,020205	-0,12139	14,31
12	0,87509	-0,01343	0,020645	-0,13321	13,00
13	1,16837	-0,02165	0,022725	-0,13679	15,16
14	0,80355	-0,01277	0,021585	-0,11455	10,76
15	0,90058	-0,01383	0,021435	-0,11409	12,32
16	1,19508	-0,02476	0,02506	-0,14551	14,68
17	1,40998	-0,01479	0,021245	-0,09521	10,72
18	0,88664	-0,01735	0,023345	-0,13695	13,44
19	1,28662	-0,02353	0,025525	-0,13417	13,94
20	1,29856	-0,0263	0,02546	-0,15102	15,50
R²	0.420	0.429	0.037	0.458	0.910
SD	0.87	0.86	0.12	0.84	0.34
R ² _{adj}	0.388	0.397	0.000	0.428	0,905

Table 7: Descripteurs B3LYP/6-311++G(d,p)/CPCM/UAKS et leurs corrélations avec les valeurs aqueuses pK_a

Acid	ϵ_{HOMO}	ϵ_{LUMO}	S	η	χ	Q _O	Q _H
1	-0,23153	-0,07116	6,23558	0,16037	-0,15135	-0,69953	0,53587
2	-0,24685	-0,06862	5,61073	0,17823	-0,15774	-0,68314	0,54142
3	-0,22987	-0,07782	6,57678	0,15205	-0,15385	-0,72135	0,52918
4	-0,23278	-0,07243	6,23636	0,16035	-0,15261	-0,70584	0,53278
5	-0,23592	-0,07691	6,28891	0,15901	-0,15642	-0,70946	0,53367
6	-0,23425	-0,06993	6,08569	0,16432	-0,15209	-0,70555	0,53272
7	-0,22843	-0,07386	6,46956	0,15457	-0,15115	-0,72275	0,52922
8	-0,23166	-0,06860	6,13271	0,16306	-0,15013	-0,70729	0,53209
9	-0,23920	-0,07586	6,12220	0,16334	-0,15753	-0,68237	0,54139
10	-0,22184	-0,07197	6,67245	0,14987	-0,14691	-0,71469	0,52986
11	-0,21120	-0,08041	7,64584	0,13079	-0,14581	-0,70522	0,53264
12	-0,21379	-0,08307	7,64994	0,13072	-0,14843	-0,70445	0,53276
13	-0,22341	-0,07612	6,78933	0,14729	-0,14977	-0,70576	0,53262
14	-0,22515	-0,09255	7,54148	0,13260	-0,15885	-0,69713	0,5348
15	-0,22490	-0,09389	7,63301	0,13101	-0,15940	-0,69786	0,53475
16	-0,23267	-0,07595	6,38081	0,15672	-0,15431	-0,706	0,53261
17	-0,23173	-0,12201	9,11411	0,10972	-0,17687	-0,693	0,53632
18	-0,22029	-0,07156	6,72359	0,14873	-0,14593	-0,69833	0,53451
19	-0,22070	-0,07335	6,78656	0,14735	-0,14703	-0,69556	0,53369
20	-0,22545	-0,06973	6,42178	0,15572	-0,14759	-0,70373	0,53245
R²	0.081	0.051	0.010	0.001	0.135	0,814	0.748
SD	1.09	1.11	1.13	1.14	1.06	0,49	0.57
R ² _{adj}	0.030	0.000	0.000	0.000	0.087	0,804	0.734

Table 8 : Descripteurs B3LYP/6-311++G(d,p)/CPCM/UAKS et leurs corrélations avec les valeurs aqueuses pK_a

Base	ϵ_{HOMO}	ϵ_{LUMO}	S	η	χ	Q_{O}
1	-0,18757	-0,04767	7,14796	0,13990	-0,11762	-0,85531
2	-0,21997	-0,04454	5,70028	0,17543	-0,13226	-0,81216
3	-0,18864	-0,06813	8,29807	0,12051	-0,12839	-0,91626
4	-0,19478	-0,05800	7,31101	0,13678	-0,12639	-0,85180
5	-0,19134	-0,06498	7,91390	0,12636	-0,12816	-0,89009
6	-0,19081	-0,05492	7,35889	0,13589	-0,12287	-0,84932
7	-0,18737	-0,06403	8,10767	0,12334	-0,12570	-0,91710
8	-0,19384	-0,05504	7,20461	0,13880	-0,12444	-0,85897
9	-0,20078	-0,06234	7,22335	0,13844	-0,13156	-0,82367
10	-0,18984	-0,05762	7,56315	0,13222	-0,12373	-0,85534
11	-0,18515	-0,06617	8,40477	0,11898	-0,12566	-0,84061
12	-0,18652	-0,06140	7,99233	0,12512	-0,12396	-0,83740
13	-0,18993	-0,06420	7,95355	0,12573	-0,12707	-0,84230
14	-0,19728	-0,07735	8,33820	0,11993	-0,13732	-0,81123
15	-0,19664	-0,07937	8,52733	0,11727	-0,13801	-0,81161
16	-0,19528	-0,06330	7,57691	0,13198	-0,12929	-0,85115
17	-0,20323	-0,10824	10,52742	0,09499	-0,15574	-0,78743
18	-0,18835	-0,05698	7,61209	0,13137	-0,12267	-0,83496
19	-0,18350	-0,05966	8,07494	0,12384	-0,12158	-0,82967
20	-0,18700	-0,05747	7,72022	0,12953	-0,12224	-0,85472
R²	0.463	0.006	0.031	0.096	0.161	0.637
SD	0.83	1.14	1.12	1.08	1.04	0.69
R ² _{adj}	0.433	0.000	0.000	0.045	0.114	0.616

Table 9: Descripteurs B3LYP/6-311++G(d,p)/CPCM/UAKS et leurs corrélations avec les valeurs aqueuses pK_a

Acid/ Base	ΔS	$\Delta \eta$	$\Delta \chi$	ΔQ_O	ΔG_{aq}^c
1	0,91238	-0,02047	0,03373	-0,15578	13,47
2	0,08955	-0,00280	0,02548	-0,12902	8,48
3	1,72128	-0,03154	0,02546	-0,19491	18,51
4	1,07465	-0,02357	0,02622	-0,14596	14,49
5	1,62498	-0,03265	0,02826	-0,18063	16,05
6	1,27321	-0,02843	0,02923	-0,14377	14,95
7	1,63811	-0,03123	0,02545	-0,19435	18,77
8	1,07190	-0,02426	0,02569	-0,15168	14,93
9	1,10115	-0,02490	0,02597	-0,14130	8,18
10	0,89070	-0,01765	0,02318	-0,13617	12,25
11	0,75893	-0,01181	0,02015	-0,13539	14,21
12	0,34239	-0,00560	0,02447	-0,13295	12,92
13	1,16422	-0,02156	0,02270	-0,13654	15,10
14	0,79672	-0,01267	0,02154	-0,11410	10,64
15	0,89432	-0,01374	0,02139	-0,11375	12,21
16	1,19610	-0,02474	0,02502	-0,14515	14,58
17	1,41332	-0,01473	0,02114	-0,09443	10,62
18	0,88849	-0,01736	0,02326	-0,13663	13,29
19	1,28837	-0,02351	0,02545	-0,13411	13,71
20	1,29844	-0,02619	0,02536	-0,15099	15,30
R²	0.335	0.361	0.051	0.463	0.916
SD	0.93	0.91	1.11	0.83	0.33
R ² _{adj}	0.298	0.326	0.000	0.434	0.911

Table 10: Descripteurs B3LYP/6-311++G(d,p)/SMD et leurs corrélations avec les valeurs aqueuses pK_a

Acid	ϵ_{HOMO}	ϵ_{LUMO}	S	η	χ	Q _O	Q _H
1	-0,23368	-0,07307	6,22626237	0,16061	-0,153375	-0,69073	0,51547
2	-0,24794	-0,06966	5,60915414	0,17828	-0,1588	-0,67155	0,51998
3	-0,23203	-0,0781	6,49645943	0,15393	-0,155065	-0,71223	0,50943
4	-0,23423	-0,07279	6,19425173	0,16144	-0,15351	-0,6967	0,51234
5	-0,23817	-0,07725	6,21426796	0,16092	-0,15771	-0,70023	0,51369
6	-0,23643	-0,07075	6,03573153	0,16568	-0,15359	-0,69584	0,51216
7	-0,23077	-0,07436	6,3934531	0,15641	-0,152565	-0,71337	0,50912
8	-0,23319	-0,06905	6,09236018	0,16414	-0,15112	-0,69788	0,51134
9	-0,2414	-0,07687	6,07791892	0,16453	-0,159135	-0,67078	0,51928
10	-0,22484	-0,07305	6,58804928	0,15179	-0,148945	-0,70626	0,51267
11	-0,21222	-0,0808	7,6091919	0,13142	-0,14651	-0,6962	0,51229
12	-0,21499	-0,08358	7,60977095	0,13141	-0,149285	-0,69531	0,51233
13	-0,2243	-0,07619	6,75173857	0,14811	-0,150245	-0,69681	0,51209
14	-0,22558	-0,09222	7,4985003	0,13336	-0,1589	-0,6882	0,51397
15	-0,22524	-0,09347	7,5889808	0,13177	-0,159355	-0,68889	0,51386
16	-0,2397	-0,08138	6,31632137	0,15832	-0,16054	-0,67567	0,49183
17	-0,23216	-0,12142	9,03016074	0,11074	-0,17679	-0,68397	0,5153
18	-0,22168	-0,07219	6,68941066	0,14949	-0,146935	-0,68998	0,51421
19	-0,22299	-0,07454	6,73627484	0,14845	-0,148765	-0,68774	0,51651
20	-0,22849	-0,07075	6,33954609	0,15774	-0,14962	-0,69435	0,51386
R²	0.057	0.054	0.120	0.002	0.078	0,691	0.229
SD	1.11	1.11	1.14	1.14	1.07	0,63	1.00
R ² _{adj}	0.005	0.010	0.000	0.000	0.127	0,674	0.186

Table 11 : Descripteurs B3LYP/6-311++G(d,p)/SMD et leurs corrélation avec les valeurs aqueuses pK_a

Base	ϵ_{HOMO}	ϵ_{LUMO}	S	η	χ	Q _o
1	-0,18683	-0,04849	7,2285673	0,13834	-0,11766	-0,83427
2	-0,22057	-0,04461	5,6831098	0,17596	-0,13259	-0,79266
3	-0,18723	-0,0687	8,4366827	0,11853	-0,127965	-0,89555
4	-0,19426	-0,05829	7,35456351	0,13597	-0,126275	-0,83146
5	-0,18977	-0,06539	8,03987779	0,12438	-0,12758	-0,86905
6	-0,19044	-0,05555	7,41344799	0,13489	-0,122995	-0,82965
7	-0,18607	-0,06469	8,23858955	0,12138	-0,12538	-0,89633
8	-0,19333	-0,05535	7,24742716	0,13798	-0,12434	-0,83905
9	-0,19945	-0,06217	7,28438228	0,13728	-0,13081	-0,80518
10	-0,19107	-0,05861	7,54944889	0,13246	-0,12484	-0,83707
11	-0,18512	-0,06616	8,40618695	0,11896	-0,12564	-0,82036
12	-0,18647	-0,06924	8,5302397	0,11723	-0,127855	-0,81753
13	-0,18945	-0,06399	7,97066794	0,12546	-0,12672	-0,82262
14	-0,19657	-0,07646	8,32570144	0,12011	-0,136515	-0,79362
15	-0,19564	-0,0782	8,51498638	0,11744	-0,13692	-0,79474
16	-0,19476	-0,06364	7,62660159	0,13112	-0,1292	-0,83072
17	-0,20212	-0,10656	10,4646296	0,09556	-0,15434	-0,771
18	-0,18777	-0,05077	7,29927007	0,137	-0,11927	-0,81489
19	-0,18362	-0,06017	8,10044552	0,12345	-0,121895	-0,80774
20	-0,18697	-0,05786	7,74533344	0,12911	-0,122415	-0,83213
R²	0.489	0.002	0.046	0.115	0.166	0.635
SD	0.81	1.14	1.11	1.07	1.04	0.70
R ² _{adj}	0.460	0.000	0.000	0.066	0.119	0.615

Table 12 Descripteurs B3LYP/6-311++G(d,p)/SMD et leurs corrélation avec les valeurs aqueuses pK_a

Acid/ Base	ΔS	$\Delta\eta$	$\Delta\chi$	ΔQ_O	ΔG_{aq}^c
1	1,00230	-0,02227	0,03572	-0,14354	11,03
2	0,07396	-0,00232	0,02621	-0,12111	4,97
3	1,94022	-0,03540	0,02710	-0,18332	16,41
4	1,16031	-0,02547	0,02724	-0,13476	11,70
5	1,82561	-0,03654	0,03013	-0,16882	13,67
6	1,37772	-0,03079	0,03060	-0,13381	12,05
7	1,84514	-0,03503	0,02719	-0,18296	16,62
8	1,15507	-0,02616	0,02678	-0,14117	12,14
9	1,20646	-0,02725	0,02833	-0,13440	4,86
10	0,96140	-0,01933	0,02411	-0,12730	8,13
11	0,79700	-0,01246	0,02087	-0,12416	11,38
12	0,92047	-0,01418	0,02143	-0,12222	10,07
13	1,21893	-0,02265	0,02353	-0,12581	12,35
14	0,82720	-0,01325	0,02239	-0,10542	7,64
15	0,92601	-0,01433	0,02244	-0,10585	9,29
16	1,31028	-0,02720	0,03134	-0,15505	9,15
17	1,43447	-0,01518	0,02245	-0,08703	7,56
18	0,60986	-0,01249	0,02767	-0,12491	10,34
19	1,36417	-0,02500	0,02687	-0,12000	10,43
20	1,40579	-0,02863	0,02721	-0,13778	12,07
R²	0.411	0.409	0.043	0.426	0.890
SD	0.88	0.88	1.12	0.86	0.38
R ² _{adj}	0.378	0.376	0.000	0.394	0.884

ملخص

تهدف الدراسة المقدمة في هذه المذكرة إلى:
1 - إيجاد علاقات كمية بين ثوابت الحموضة المائية لمركبات هيدروكسيد الكومارين مع المؤشرات النظرية مثل الطاقات الحرة للتشرد.
2 - تنبؤ ثوابت الحموضة المائية لهذه المركبات عن طريق بلورة علاقات كمية بنية-حموضة (دراسة QSPR).

قمنا بهذه الدراسة النظرية باستعمال مقاربات نظرية إحصائية وكمية
قمنا بالحسابات باستعمال برنامج Gaussian و بالطرق DFT/B3LYP/6-311++G**

الكلمات المفتاحية: هيدروكسيد الكومارين , دراسة QSPR ، ثابت الحموضة المائية ، مؤشرات نظرية نموذج SMD نموذج PCM ,

Résumé

L'objectif de cette étude présentée de ce mémoire est :

- 1- Elaboration des corrélations quantitatives entre les constantes d'acidité aqueuse expérimentales des hydroxy-coumarines avec des descripteurs théoriques, notamment les énergies libres de déprotonation.
- 2- Elaboration des modèles de corrélation quantitatives structure-acidité (QSPR) pour la prédiction des constantes d'acidité aqueuse (pK_a) des hydroxy-coumarines en utilisant les méthodes de régression linéaires.

Les approches statistiques, en l'occurrence la régression linéaire simple et multiple et des approches quantiques ont été utilisées. Les calculs ont été réalisés avec le programme *Gaussian* en utilisant les méthodes DFT/B3LYP/6-311++G**.

Mots-Clés : *hydroxy-coumarines ; QSPR ; pK_a aqueux ; Descripteurs théoriques ; Modèle PCM, Modèle SMD*

Abstract

The aim of this study:

- 1- Elaboration of quantitative correlations between the aqueous acidity constants of hydroxy-coumarines, namely, deprotonation free energies.
- 2 - Prediction of aqueous acidity constants (pK_a) hydroxyl-coumarines by developing QSPR models using linear regression methods.

This theoretical study was performed using statistic approaches, namely, simple and multiple linear regression and quantum chemistry approaches. The calculations were carried out with Gaussian program using the DFT/B3LYP/6-311++G**

Keywords: *hydroxy-coumarines; QSPR; aqueous pK_a ; Theoretical descriptors ; PCM Model; SMD Model*