

# REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE



MINISTRE DE L'ENSEIGNEMENT SUPERIEUR  
ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE ABOU BEKR BELKAÏD DE TLEMEN  
FACULTE DES SCIENCES  
DEPARTEMENT DE CHIMIE

**THÈSE**  
pour l'obtention du diplôme de Doctorat  
en Chimie Physique  
(Option: *Chimie Théorique et Modélisation Moléculaire*)

Thème :

***ELABORATION DES CORRELATIONS QUANTITATIVES STRUCTURE-ACIDITE DES  
ACIDES CARBONES. ETUDE THEORIQUE DES EFFETS DE SOLVANTS SUR LES  
EQUILIBRES CETO-ENOLIQUES DES COMPOSES  $\beta$ -DICARBONYLES CYCLIQUES  
ET DE LEURS ANALOGUES A CHAINES OUVERTES***

Présentée par :

**Mr CHARIF Imad Eddine**

Soutenu publiquement le: .. / /2011 devant le jury d'examen composé de :

Melle Latifa NEGADI	Professeur	U. A. Belkaïd de Tlemcen	Présidente
Mr. Hacène MEGHEZZI	Professeur	U.S.T.H.B Alger	Examineur
Mr. Meziane BRAHIMI	Professeur	U.S.T.H.B Alger	Examineur
Mr. Bachir MOSTEFA-KARA	Professeur	U. A. Belkaïd de Tlemcen	Examineur
Mr. Sidi Mohamed MEKELLECHE	Professeur	U. A. Belkaïd de Tlemcen	Directeur de thèse
Mr. Didier VILLEMIN	Professeur	U. de Caen, France	Co-Directeur de thèse

*A la mémoire de ma nièce CHARIF Khalida*

*A mes très chers parents*

*A ma femme et à mon fils Naoufel Rachid*

*A mon frère et à mes sœurs*

*A ma grande famille*

*A tous ceux qui me sont chers*

## ◦◦◦ REMERCIEMENTS ◦◦◦

*Le travail présenté dans cette thèse a été dirigé par Monsieur Sidi Mohamed MEKELLECHE Professeur au Département de Chimie Faculté des Sciences, Université Abou-Bekr 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 à exprimer toute ma reconnaissance à Monsieur S.M. MEKELLECHE pour m'avoir guidé durant mon travail avec une disponibilité permanente et m'avoir fait bénéficier de ses connaissances en chimie théorique.*

*J'exprime ma profonde et respectueuse gratitude à Mademoiselle L. NEGADI, Professeur à l'Université A. Belkaïd de Tlemcen, qui m'a fait l'honneur d'accepter de présider le jury de cette thèse.*

*Je tiens à adresser mes vifs remerciements et l'expression de mon profond respect à Monsieur H. MEGHEZZI et à Monsieur M. BRAHIMI, Professeurs à l'USTHB d'Alger, pour l'honneur qu'ils nous ont fait en acceptant d'examiner notre travail et de faire le déplacement à Tlemcen pour participer au jury. Mes vifs et sincères remerciements vont également à Monsieur B. MOSTEFA-KARA, Professeur à l'Université A. Belkaïd de Tlemcen, pour l'honneur qu'il nous a fait d'accepter de juger notre travail.*

*Mes vifs et sincères remerciements vont également au professeur Didier VILLEMIN pour son accueil à l'Université de Caen-France pour sa collaboration et ses conseils fructueux. Mes vifs et sincères remerciements vont également à Madame Nelaine MORADIEZ, Professeur associé à l'Université de Kamloops (Thompson Rivers University, Canada) pour son intérêt au travail et pour sa contribution fructueuse à certaines publications liées à ce travail.*

*Enfin, j'adresse mes remerciements à tous les collègues de l'équipe « chimie théorique et modélisation moléculaire » du laboratoire de recherche « Thermodynamique Appliquée et Modélisation Moléculaire » et à tous ceux qui ont contribué de près ou de loin à la réalisation de ce travail.*

# SOMMAIRE

INTRODUCTION GENERALE.....	1
Bibliographie.....	7
<b>CHAPITRE I : METHODES DE CALCULS QUANTO-CHIMIQUES</b>	
Introduction .....	9
I.1. Méthode de Hartree-Fock-Roothaan .....	11
I.1.1. Approximation du champ moyen de Hartree .....	11
I.1.2. Méthode de Hartree-Fock .....	11
I.1.3. Méthode de Hartree-Fock-Roothaan .....	13
I.2. Méthodes Post-SCF .....	14
I.2.1. Méthode de Möller-Plesset MP2 .....	15
I.3. Théorie de la fonctionnelle de densité (DFT) .....	16
I.3.1. Fondements de la théorie DFT .....	16
I.3.2. Méthode de Kohn et Sham .....	19
I.3.3. Approximation de la densité locale LDA.....	22
I.3.4. Approximation du Gradient Généralisé (GGA) .....	23
I.3.5. Fonctionnelle hybride B3LYP.....	24
I.4. Bases d'orbitales atomiques .....	24
I.5. Les effets de solvation .....	27
I.5.1. Modèle SRCF .....	28
I.5.2. Modèle PCM .....	30
• <i>Algorithme PCM</i> .....	31
I.5.3. Les termes non-électrostatiques .....	34
• Le terme de cavitation.....	34
• Le terme de dispersion .....	34
• Le terme de répulsion .....	35
I.6. L'analyse NBO.....	36
I.6.1. La théorie des perturbations pour les orbitales dans le cadre Hartree-Fock: Interaction donneur-accepteur .....	36
I.6.2. Les orbitales naturelles NAO et NBO .....	39
I.7. Thermodynamique statistique.....	43
I.7.1. La fonction de partition.....	44
I.7.2. Expression de l'énergie.....	45
I.7.3. Expression de la capacité calorifique.....	46

I.7.4. Expression de l'entropie.....	46
I.7.5. Contribution des différents mouvements à la fonction de partition .....	47
a. Contribution du mouvement de translation.....	47
b. Contribution du mouvement électronique.....	50
c. Contribution du mouvement de rotation .....	51
d. Contribution du mouvement de vibration .....	55
Bibliographie du chapitre I.....	59

## CHAPITRE II : MODELISATION QSPR ET ANALYSE DE DONNEES

Introduction .....	62
II.1. Principe des méthodes QSPR.....	62
II.2. Importance des bases de données.....	65
II.3. Descripteurs moléculaires.....	65
II.4. Méthodes d'analyse de données .....	66
II.4.1. Régression linéaire simple .....	67
II.4.2. Régression linéaire multiple .....	75
II.4.3. Confiance accordée aux résultats.....	77
Bibliographie du chapitre II.....	82

## CHAPITRE III : RESULTATS ET DISCUSSIONS

<b>PARTIE 1: Corrélation des constantes d'acidité aqueuse des acides carbonés         avec les descripteurs théoriques.</b>	84
1. Introduction .....	84
2. Méthodologie des calculs .....	87
2.1. Calcul d'acidité en phase gazeuse.....	87
2.2. Calcul d'acidité en phase aqueuse.....	87
3. Résultats et discussions .....	88
3.1. L'acidité en phase gazeuse .....	88
3.2. L'acidités en phase aqueuse .....	92
4. Conclusion.....	96
Bibliographie de la partie 1.....	97
Annexe n°1.....	

<b>PARTIE 2: Prédiction des constantes d'acidité aqueuse (pK<sub>a</sub>) des acides carbonés.</b>	<b>100</b>
Etude QSPR (Quantitative Structure-Property Relationships)	
1. Introduction.....	100
2. Méthodologie des calculs.....	105
2.1. Calcul de l'énergie libre de déprotonation $\Delta G^{\circ}_{aq}$ .....	105
3. Résultats et discussions.....	106
3.1. Prédiction des valeurs aqueuses de pK <sub>a</sub> des acides carbonés d'une série de test.....	113
4. Conclusion.....	115
Bibliographie de la partie 2.....	116
Annexe n°2.....	
<b>PARTIE 3: Etude théorique des effets de solvants sur l'équilibre tautomère céto-énolique des composés <math>\beta</math>-dicarbonylés cycliques et de leurs analogues à chaînes ouvertes.</b>	<b>119</b>
1. Introduction.....	120
2. Méthodologie des calculs.....	122
3. Premier cas: Etude des équilibres tautomères céto-énoliques de l'acétylacétate d'éthyle (EAA) et de l'acide tétronique (TA).....	124
3.1. Etude structurale de l'acétylacétate d'éthyle (EAA) .....	124
3.2. Etude théorique des effets de solvants sur le déplacement de l'équilibre céto-énol de l'acétylacétate d'éthyle (EAA) .....	126
3.3. Etude structurale de l'acide tétronique (TA).....	128
3.4. Etude théorique des effets de solvants sur le déplacement de l'équilibre céto-énol de l'acide tétronique (TA).....	129
4. Second Cas: Etude des équilibres tautomères céto-énoliques de l'acétylacétone (AcAc) et de la dimédone (DA).....	130
4.1. Etude structurale de l'Acétylacétone (AcAc).....	130
4.2. Etude théorique des effets de solvants sur le déplacement de l'équilibre céto-énol de l'acétylacétone (AcAc).....	132
4.3. Etude structurale de la dimédone.....	134
4.4. Etude théorique effets de solvants sur le déplacement de l'équilibre céto-énol de la dimédone.....	136
5. Conclusions.....	138

Bibliographie de la partie 3.....	139
<b>PARTIE 4 : Etude théorique des effets stéréoelectroniques sur l'acidité carbonée des composés <math>\beta</math>-dicarboxylés cycliques et de leurs analogues à chaînes ouvertes.</b>	145
1. Introduction.....	145
2. Résultats et discussions.....	147
2.1. Calculs des énergies libres de déprotonation $\Delta G^\circ$ .....	147
2.2. L'analyse NBO .....	149
2.3. Etude des effets stéréoelectroniques sur l'acidité carbonée des composés $\beta$ -dicarboxylés	150
• Cas n° 1 : Acide de Meldrum / Malonate diéthyle .....	150
• Cas n° 2 : Acide barbiturique /Malonamide .....	154
• Cas n° 3 : Acide tétronique /Acétylacétate d'éthyle .....	156
3. Conclusion .....	159
Bibliographie de la partie 4.....	160
<b>CONCLUSION GENERALE.....</b>	161

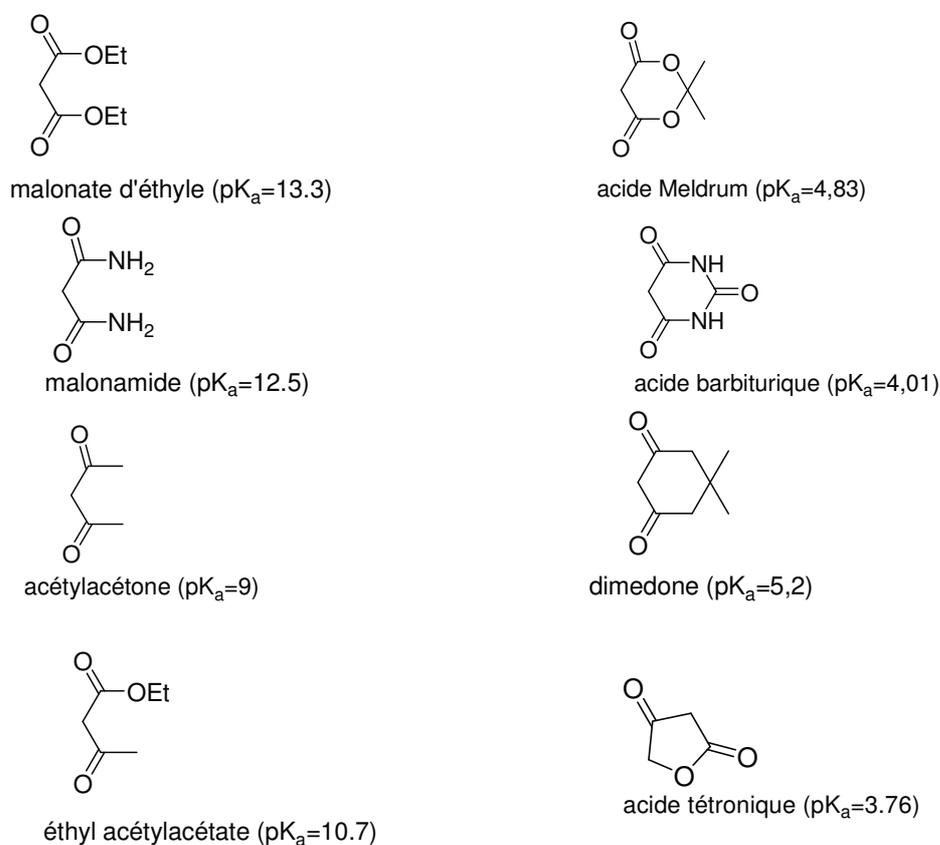
## LISTE DES ABREVIATIONS

AcAc	Acetylacetone
B3LYP	Becke 3-Parameter Lee-Yang-Parr
CLOA	Combinaison Linéaire d'Orbitales Atomiques
CGTO	Contracted Gaussian Type Orbital Configuration
CT	Charge transfer
DFT	Density Functional Theory
DIM	Dimedone
DMSO	Dimethyl sulfoxide
EAA	Ethyl Acetoacetate
FMO	Frontier Molecular Orbital
GGA	Generalized Gradient Approximation
GTO	Gaussian Type Orbital
HF	Hartree-Fock
HOMO	Highest Occupied Molecular Orbital
KS	Kohn et Sham
LDA	Local Density Approximation
LUMO	Lowest Unoccupied Molecular Orbital
MLR	Multiple Linear Regression
MP2	Möller –Plesset d'ordre 2
NAO	Natural Atomic Orbitals
NHO	Natural Hybrid Orbitals
NBO	Natural Bond Orbitals
NPA	Natural Population Analysis
OA	Orbitale Atomique
OM	Orbitale Moléculaire
PCM	Polarizable Continuum Model
QSPR	Quantitative Structure-Property Relationships
QSAR	Quantitative Structure-Activity Relationships
SLR	Simple Linear Regression
SSR	Sum of Squares Regression
SSE	Sum of Squares Error
SST	Sum of Squares Total
SCF	Self Consistent Field
STO	Slater Type Orbital
SEP	Surface d'énergie potentielle
SCRF	Self-Consistent Reaction Field
TA	Tetronic acid

# ***INTRODUCTION GENERALE***

Les acides carbonés ont un rôle fondamental en synthèse organique puisqu'ils sont les précurseurs d'organométalliques c'est-à-dire des carbanions potentiels. Les acides carbonés les plus acides (malonates, dicarboxylés) sont des synthons extrêmement importants pour la synthèse de liaison C-C par alkylation, aldolisation, réactions de Claisen, Knoevenagel, de Michael,...

Un certain nombre de composés  $\beta$ -dicarboxylés cycliques sont connus depuis le 19<sup>ème</sup> siècle par leurs propriétés acides (carbonées) surprenantes, c'est le cas de l'acide tétronique ( $pK_a=3.76$ ), l'acide barbiturique ( $pK_a=4.01$ ), l'acide de Meldrum ( $pK_a=4.83$ ) et la dimédone ( $pK_a=5.25$ ) (**Figure 1**). D'une part, cette acidité carbonée est comparable à l'acidité oxygénée en phase aqueuse de l'acide acétique ( $CH_3COOH$ ,  $pK_a=4.75$ ). D'autre part, cette acidité des ces  $\beta$ -dicarboxylés est d'autant plus remarquable que les composés analogues à chaînes ouvertes qui ne sont pas aussi acide tel que l'acétylacétate d'éthyle ( $pK_a=10.7$ ), le malonamide ( $pK_a=12.5$ ), le malonate diéthyle ( $pK_a=13.5$ ) et l'acétylacétone ( $pK_a=9$ ) (**Figure 1**). On note que ces acidités exceptionnelles ont été constatées expérimentalement mais n'ont pas été expliquées de façon rationnelle.



**Figure 1:** Acidité aqueuse des  $\beta$ -dicarboxylés cycliques et de leurs analogues acycliques.

Généralement, ce phénomène est décrit de façon discrète produit par produit (c'est-à-dire cas par cas et non pas dans le cas général). Dans l'étude expérimentale des condensations sans solvant de Knoevenagel catalysées par des catalyseurs inorganiques, il a été considéré que ce phénomène est général et il a été exploité au point de vue synthétique avec l'acide de Meldrum [2], l'acide barbiturique [3], l'acide tétronique [4] et d'autres composés [5-7].

La prédiction et la détermination de l'acidité ou de la constante d'acidité ( $K_a$ ,  $pK_a$ ) de manière aussi fiable que possible des composés organiques joue un rôle fondamental dans beaucoup de procédures analytiques telles que les titrages acido-basiques, les extractions de solvants, formation des complexes et le transport d'ions. Il a été montré que la propriété acido-basique des acides et des bases organiques affecte la toxicité, le comportement chromatographique et les propriétés pharmaceutiques de ces composés [8]. Le chimiste s'oriente, dans plusieurs cas, vers des méthodes de recherche, qui consistent à prédire les propriétés physico-chimiques des molécules avant même que celles-ci ne soient synthétisées. Une discipline faisant partie de la chimie « *computationnelle* » a été alors développée. Il s'agit de la modélisation QSPR (*Quantitative Structure-Property Relationships*). L'élaboration des modèles mathématiques QSPR reliant les propriétés physico-chimiques à la structure moléculaire permet, d'une part, d'expliquer l'origine de ces propriétés et, d'autre part, de les prédire pour des molécules pour lesquelles les données expérimentales ne sont pas disponibles. Les relations entre les structures des molécules et leurs propriétés sont généralement établies à l'aide de méthodes de modélisation basées sur des analyses statistiques. Les méthodes QSPR reposent sur la caractérisation des molécules par un ensemble de descripteurs, nombres réels mesurés ou calculés à partir des structures moléculaires. Il est alors possible d'établir une relation quantitative entre ces descripteurs et la propriété physico-chimique ou biologique. Les premières applications des méthodes QSPR sont biologiques [9,10], toxicologiques [11,12] et pharmaceutiques [13,14]. Néanmoins, ces dernières années ces méthodes ont été étendues pour la prédiction d'autres propriétés, en particulier, physico-chimiques [15-17].

Dans la première partie de notre travail, on se propose d'élaborer des modèles QSPR fiables reliant l'acidité expérimentale en phase aqueuse ( $pK_a$ ) d'une série constituée de 21 acides carbonés (**Schéma 2**) aux descripteurs quantiques en utilisant la régression linéaire simple SLR (*Simple Linear Regression*) et la régression linéaire multiple MLR (*Multiple Linear Regression*).

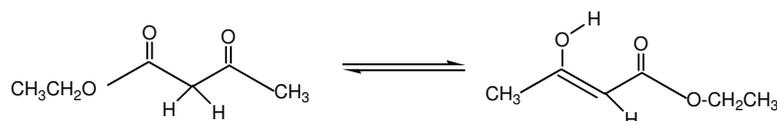
- Dans le modèle QSPR-SLR, nous avons corrélié les constantes d'acidité  $pK_a$  expérimentales aux valeurs d'énergies libres de déprotonation des acides étudiés calculées en phase gazeuse ( $\Delta G^\circ_{\text{gas}}$ ) et en phase aqueuse ( $\Delta G^\circ_{\text{aq}}$ ).
- Dans le modèle QSPR-MLR nous avons corrélié l'acidité expérimentale en phase aqueuse aux descripteurs théoriques (énergie libre de déprotonation, les charges nettes, les énergies des orbitales frontières HOMO et LUMO, la dureté chimique, la mollesse chimique, la variation du volume de cavité du solvant,...).

Tous les calculs quantiques ont été effectués avec le programme Gaussian en utilisant la méthode DFT/ B3LYP/6-311++G\*\*. La pertinence des différents descripteurs et sera testée statistiquement.

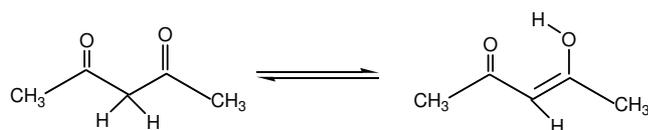
Dans la seconde partie de notre travail nous avons étudié les effets des substituants et les effets de solvant sur les équilibres céto-énoliques des composés  $\beta$ - ou 1,3-dicarbonylés. L'étude théorique des équilibres tautomères ont suscité un grand intérêt au cours de ces dernières décennies [18-23]. Des études expérimentales et théoriques en phase gazeuse sur les équilibres céto-énoliques des composés  $\beta$ -dicarbonylés ont montré que les tautomères énol sont prédominants lorsque la fonction énol est adjacente à des substituants de nature donneurs d'électrons ( $R=H, C(CH_3)_3$ ) [21,22]. Contrairement, les composés 1,3-dicarbonylés avec substituants de nature accepteurs d'électrons ( $R=F, Cl, OCH_3, NH_2$ ) existent préférentiellement sous la forme dicétonique [21,22].

Les équilibres céto-énolique dépendent également de la nature de solvant. Expérimentalement les deux tautomères sont présents en solution dans différentes

proportions [24]. Dans le cas des composés  $\beta$ -dicarboxylés à chaînes ouvertes, le pourcentage du tautomère énol augmente dans les solvants non polaires, c.-à-d., quand les liaisons d'hydrogène intermoléculaires avec le solvant sont absentes. En revanche, dans les solvants protiques polaires, les liaisons hydrogène intermoléculaires avec les molécules du solvant stabilisent la forme dicétonique. Par exemple, pour l'acétylacétate d'éthyle (**Schéma 1**) l'énol est présent en plus grande quantité (15-30%) dans les solvants non polaires, comme de tétrachlorure de carbone ( $\text{CCl}_4$ ,  $\epsilon=2.23$ ) et le benzène ( $\epsilon=2.25$ ) que dans les solvants plus polaires comme l'eau ( $\epsilon=78.39$ ) et l'acétone ( $\epsilon=20.7$ ) (5% d'énol dans l'acétone, 1% d'énol dans l'eau). Les constantes d'équilibre  $K_T=\text{énol}/\text{céto}$  sont de l'ordre de 0.3 dans  $\text{CCl}_4$  et 0.07 dans l'eau [24].

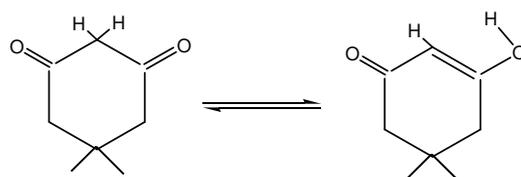
**Schéma 1**

Pour l'acétylacétone (**Schéma 2**) les constantes d'équilibre  $K_T=\text{énol}/\text{céto}$  dans les mêmes solvants sont de l'ordre de 29 et 0.23 respectivement [23].

**Schéma 2**

Dans le cas des  $\beta$ -dicarboxylés cycliques l'effet de solvant sur l'équilibre céto-énolique est opposé à celui de leurs analogues à chaînes ouvertes [24, 25]. Par exemple, pour la dimédone (**Schéma 3**), un composé analogue à l'acétylacétone, le pourcentage tautomère énol est supérieur à 95% dans l'eau ( $\epsilon=78.39$ ) et inférieur à 2% dans le cyclohexane ( $\epsilon=2.02$ )

Schéma 3



Dans la dernière partie de ce travail, on se propose de donner une explication théorique de la différence d'acidité aqueuse entre les composés  $\beta$ -dicarbonylés cycliques et de leurs analogues acycliques (**Figure 1**) en se basant sur la quantification des effets stéréoélectroniques et les interactions donneurs-accepteurs à l'aide de l'analyse NBO (Natural Bond Orbital).

Le manuscrit de cette thèse est présenté en trois chapitres

Dans le chapitre 1, nous présentons les méthodes de la chimie quantique (méthode de Hartree-Fock-Roothaan, méthode post-SCF, MP2, méthodes DFT). Le modèle de solvation PCM (Polarizable Continuum Model) [26] ainsi que l'analyse de populations électroniques NBO [27] sont également présentés.

Le chapitre 2 est consacré à la description de la modélisation QSPR. Le principe de la méthodologie QSPR, les outils statistiques et d'analyse de données ainsi qu'une brève description des méthodes de régression SLR et MLR sont également présentés.

Dans le troisième chapitre, nous présentons les résultats obtenus des différentes parties effectuées :

**Partie 1:** Corrélation des constantes d'acidité aqueuse des acides carbonés avec les descripteurs théoriques. [I. E. Charif, S. M. Mekelleche, D. Villemin, N. Mora-Diez, *J. Mol. Struct. (THEOCHEM)* 818 (2007) 1]

**Partie 2 :** Prédiction des constantes d'acidité aqueuse ( $pK_a$ ) des acides carbonés. Etude QSPR (Quantitative Structure-Property Relationships). [*I.E. Charif, S.M. Mekelleche, D. Villemin, N. Mora-Diez, article soumis*]

**Parti 3:** Etude théorique des effets de solvant sur l'équilibre tautomère céto-énolique des acides carbonés cycliques et de leurs analogues à chaînes ouvertes.

1<sup>er</sup> cas : Acide tétronique / Acétylacétate d'éthyle. [*I. E. Charif, S. M. Mekelleche, D. Villemin J. Theor. Comp. Chem. 9(6) (2010) 1021*].

2<sup>ème</sup> cas: Acide Dimédone /Acétylacétone.

**Partie 4 :** Etude des effets stéréoélectroniques sur l'acidité carbonée des acides  $\beta$ -dicarbonylés cycliques et de leurs analogues à chaînes ouvertes.

Le manuscrit de cette thèse est bouclé par les principales conclusions ainsi que les perspectives plausibles de ce travail.

**Bibliographie**

- [1] (a) E.P. Serjeant, B. Dempsey, Constantes de l'Ionisation d'Acides Organiques dans Solution aqueuse, IUPAC Série des Données Chimique, Pergamon Press, N.Y, 1979.  
(b) R.G. Pearson, R.L. Dillon, J., Est. Chem. Soc. 75, (1953) 2439. (d) F.G. Bordwell, Acc. Chem. Loi. 21 (1988) 456, 463.
- [2] (a) Didier Villemin, Chem. Ind. (London) (1983) 478-479.  
(b) Didier Villemin, Tetrahedron Lett. 24 (1983) 2855.
- [3] Didier Villemin and Bouchta Labiad, *Synth. Commun.* 20 (1990) 3333.
- [4] Didier Villemin and Bouchta Labiad, *Synth. Commun.*, 20 (1990) 3207.
- [5] (a) Didier Villemin and Bouchta Labiad, *Synth. Commun.* 20 (1990) 3213.  
(b) Didier Villemin and Benoît Martin, *Synth. Commun.* 25 (1995) 3135.
- [6] (a) Didier Villemin, Benoît Martin and Mohamed Khalid, *Synth. Commun.* 28, 17 (1998) 3201.  
(b) Didier Villemin, Benoît Martin and Nathalie Bar, *Molecules* 3 (1998) 88.
- [7] Didier Villemin, Nathalie Bar, Mohamed Khalid, Paul-Alain Jaffrès and J. Sopkova-de Oliveira Santos, *J. Chem. Res.* (2003) 433-435.
- [8] (a) C. Hansh, P. G. Sammes, J. B. Taylor, Eds. "Comprehensive Medicinal Chemistry. Quantitative Drug Design"; Pergamon Press; Oxford, UK., Vol. 4, 1990.  
(b) D.D. Perrin, B. Dempsey, E. P. Serjeant, "pK<sub>a</sub> Prediction for Organic Acids and Bases", Chapman and Hall; New York, 1981.
- [9] D.A. Winkler, *Brief. Bioinf.*, 3(2002) 73.
- [10] H. Gao, J.A. Katzenellenbogen, R. Garg, C. Hansch, *Chem. Rev.*, 99 (1999) 723.
- [11] C.D. Selassie, R. Garg, S. Kapur, A. Kurup, R.P. Verma, S.B. Mekapati, C. Hansch, *Chem. Rev.*, 102 (2002) 2585.
- [12] S.P. Bradbury, *Toxicol. Lett.*, 79 (1995) 229.
- [13] M. Grover, B. Singh, M. Bakshi, S. Singh, *Pharm. Sci. Tech. Today*, 3 (2000) 50.
- [14] M. Grover, B. Singh, M. Bakshi, S. Singh, *Pharm. Sci. Tech. Today*, 3 (2000) 28.
- [15] A.R. Katritzky, M. Karelson, V.S. Lobanov, *Pure Appl. Chem.*, 69 (1997) 245.
- [16] A.R. Katritzky, V.S. Lobanov, M. Karelson, *Chem. Soc. Rev.*, 24 (1995) 279.

- [17] J. Taskinen, J. Yliruusi, *Adv. Drug Deliv. Rev.*, 55 (2003) 1163.
- [18] C. Reichard, "solvents and solvents effects in organic chemistry ", New York, 1988.
- [19] J. Emsley, "Structure and Bonding", Springer, Berlin, Vol. 57, p. 147, 1984.
- [20] M. Huang, W. Graham Richards, Guy H. Grant, *J. Phys. Chem. A* 109 (2005) 5198.
- [21] (a) N.V. Belova,, G.V. Girichev, S.A. Shlykov, and H. Oberhammer, *J. Org. Chem.* 71 (2006) 5298.  
(b) N.V. Belova, H. Oberhammer, G.V. Girichev, S.A. Shlykov, *J. Phys. Chem. A* 111 (2007) 2248.
- [22] (a) N.V. Belova, H. Oberhammer, G.V. Girichev, *J. Mol. Struct.* 689 (2004) 255  
(b) N.V. Belova, V.V. Sliznev, H. Oberhammer, G.V. Girichev, *J. Mol. Struct.* 978 (2010) 282.
- [23] (a)H. Meghezzi, A. Boucekkine, *J. Mol. Struct. (THEOCHEM)* 254 (1992) 359.  
(b) H. Meghezzi, A. Boucekkine, *J. Mol. Struct. (THEOCHEM)* 257 (1992) 175.  
(c) A. Ladjarafi, H. Meghezzi, A. Boucekkine, *J. Mol. Struct. (THEOCHEM)* 709(2004) 129.
- [24] (a) K. D. Grande, S. M. Rosenfeld, *J. Org. Chem.* 45 (1980) 1626.  
(b) S. G. Mills, P. Beak, *J. Org. Chem.* 50 (1980) 1216.
- [25] P. Vogel, "Chimie Organique, Méthodes et Modèles ", université DeBoeck, 1998.
- [26] (a) S. Miertus, E. Scrocco, J. Tomasi, *J. Chem. Phys.* 55 (1981) 117.  
(b) S. Miertus, J. Tomasi, *J. Chem. Phys.* 65 (1982) 239.  
(c) M. Cossi, V. Barone, R. Cammi, J. Tomasi, *J. Chem. Phys. Lett.* 255 (1996) 327.
- [27] (a) A. E. Reed, , F. Weinhold, *J. Chem. Phys.* 78 (1983) 4066.  
(b) A. E. Reed, R. B. Weinstock, F. Weinhold, *J. Chem. Phys.* 83 (1985) 735.  
(c) A. E. Reed, , F. Weinhold, *J. Chem. Phys.* 83 (1985) 1736.

# **CHAPITRE I**

## ***METHODES DE CALCULS QUANTO-CHIMIQUES***

## Introduction

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

Les méthodes de chimie quantique, présentées par la suite, reposent toutes sur l'équation de Schrödinger [1], non relativiste, indépendante du temps.

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

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

L'hamiltonien  $H$  total est défini par la somme de cinq termes (terme cinétique des électrons, terme cinétique des noyaux, terme de répulsions électrons-électrons, terme de répulsions noyaux-noyaux et terme d'attractions électrons-noyaux).

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

Born et Oppenheimer [2] ont proposé l'approximation des noyaux fixes qui consiste à séparer l'hamiltonien électronique de l'hamiltonien nucléaire. Dans le cadre de cette approximation l'hamiltonien  $H$  peut se réduire à la forme suivante :

$$H = -\frac{\hbar^2}{2m_e} \sum_{i=1}^n \Delta_i - \sum_{K=1}^N \sum_{i=1}^n \frac{Z_K e^2}{R_{Ki}} + \sum_{i>j}^n \frac{e^2}{r_{ij}} \quad (3)$$

La résolution exacte de l'équation 1 n'est possible que pour l'atome d'hydrogène et les systèmes hydrogénoïdes. Pour les systèmes poly-électroniques, il est nécessaire de faire appel aux méthodes d'approximation pour résoudre l'équation de Schrödinger d'une manière approchée.

Les propriétés moléculaires qui peuvent être calculées par la résolution de l'équation de Schrödinger sont multiples. On peut citer, entre autres,

- Structures et énergies moléculaires
- Energies et structures des états de transition
- Fréquences de vibration
- Spectres IR et Raman
- Propriétés thermochimiques
- Energies de liaison
- Chemins réactionnels
- Orbitales moléculaires
- Charges atomiques
- Moments multipolaires
- Déplacements chimiques RMN et susceptibilités magnétiques
- Affinités électroniques et potentiels d'ionisation
- Polarisabilités et hyperpolarisabilités
- Potentiels électrostatiques et densités électroniques
- etc...

## I.1. Méthode de Hartree-Fock-Roothaan:

### I.1.1. Approximation du champ moyen de Hartree:

L'approximation du champ moyen, proposée par Hartree [3] en 1927, consiste à remplacer l'interaction d'un électron avec les autres électrons par l'interaction de celui-ci avec un champ moyen créé par la totalité des autres électrons ; ce qui permet de remplacer le potentiel biélectronique  $\sum_j e^2/r_{ij}$  qui exprime la répulsion entre l'électron  $i$  et les autres électrons  $j \neq i$  par un potentiel monoélectronique moyen de l'électron  $i$  de la forme  $U(i)$ . Par conséquent et en se basant sur le théorème des électrons indépendants, nous pouvons écrire la fonction d'onde totale comme le produit de fonctions d'onde monoélectroniques:

$$\Psi = \phi_1(1) \cdot \phi_2(2) \cdot \phi_3(3) \dots \phi_n(n) \quad (4)$$

### I.1.2. Méthode de Hartree-Fock :

La fonction d'onde polyélectronique de Hartree (Eq. 4) ne vérifie ni le principe d'indiscernabilité des électrons ni le principe d'exclusion de Pauli. Pour tenir compte de ces deux principes, Fock [4] a proposé d'écrire la fonction d'onde totale  $\Psi$  sous forme d'un déterminant, appelée déterminant de Slater [5], dont la forme abrégée pour un système à couches fermées est  $n=2m$ :

$$\Psi(1,2,\dots,n) = \frac{1}{(n!)^{1/2}} \begin{vmatrix} \phi_1(1) & \bar{\phi}_1(2) & \dots & \phi_m(2m-1) & \bar{\phi}_m(2m) \end{vmatrix} \quad (5)$$

avec :

$$\phi_1(1) = \psi_1(1)\alpha(1) \quad (6)$$

$$\bar{\phi}_1(2) = \psi_1(2)\beta(2) \quad (7)$$

$\phi$  est la fonction d'espace monoélectronique et  $\alpha$  et  $\beta$  sont les fonctions de spin.

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

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

$$E^{HF} = 2 \sum_i h_{ii} + \sum_i \sum_{j>i} (2J_{ij} - K_{ij}) \quad (8)$$

Avec :

$$\begin{aligned} h_{ii} &= \int \phi_i^*(1) \hat{h}_1 \phi_i(1) d\vec{r}_1 \\ J_{ij} &= \int \phi_i^*(1) \phi_j^*(2) r_{12}^{-1} \phi_i(1) \phi_j(2) d\vec{r}_1 d\vec{r}_2 \\ K_{ij} &= \int \phi_i^*(1) \phi_j^*(2) r_{12}^{-1} \phi_j(1) \phi_i(2) d\vec{r}_1 d\vec{r}_2 \end{aligned} \quad (9)$$

$$\text{où : } \hat{h}_1 = -\frac{1}{2} \Delta_1^2 - \sum_K \frac{Z_K e^2}{r_{1K}} \quad (10)$$

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

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

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

$$\hat{F} \phi_i = E_i \phi_i \quad (11)$$

où  $E_i$  sont les énergies des spin-orbitales.

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

$$\hat{F}(1) = \hat{h}_1(1) + V^{eff}(1) = \hat{h}_1(1) + \sum_i (2\hat{J}_i(1) - \hat{K}_i(1)) \quad (12)$$

### I.1.3. Méthode de Hartree-Fock-Roothaan:

Les expressions analytiques des orbitales moléculaires  $\phi_i$  n'ont pas été définies dans le cadre de la méthode de Hartree-Fock. C'est Roothaan [6] qui a utilisé la technique OM-CLOA pour construire les OM. Cette méthode consiste à exprimer l'orbitale moléculaire  $\Psi_i$  par une combinaison linéaire d'orbitales atomiques  $\varphi_\mu$ :

$$\phi_i = \sum_{r=1}^N C_{ir} \varphi_r \quad (13)$$

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

Les meilleurs coefficients sont ceux qui minimisent l'énergie. En procédant par la méthode des variations et après certaines manipulations algébriques, on aboutit aux équations de Roothaan définies par le système séculaire suivant [6]:

$$\sum_{r=1}^N C_{ir} (F_{rs} - \varepsilon_i S_{rs}) = 0 \quad s = 1, 2, \dots, N \quad (14)$$

Avec:

$$\left\{ \begin{array}{l} F_{rs} = h_{rs}^c + \sum_{p=1}^n \sum_{q=1}^n P_{pq} \{ 2 \langle rs | pq \rangle - \langle rq | ps \rangle \} \\ S_{rs} = \langle \varphi_r | \varphi_s \rangle \\ h_{rs}^c = \int \varphi_r^*(i) h^c \varphi_s(i) d\tau_i \end{array} \right. \quad (15)$$

Où r, s, p et q symbolisent 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 biélectroniques coulombienne et d'échange respectivement.  $S_{rs}$  est une intégrale de recouvrement.

## I.2. Méthodes Post-SCF :

La méthode Hartree-Fock-Roothaan présente l'inconvénient majeur de ne pas tenir compte de la **corrélacion électronique** qui existe entre le mouvement des électrons. Ceci rend cette méthode relativement restreinte dans le calcul quantitatif des propriétés thermodynamiques telles que l'enthalpie d'activation, l'énergie de Gibbs de réactions, énergies de dissociation.

Ces propriétés peuvent être calculées d'une manière efficace par les méthodes Post-SCF en tenant compte de la corrélacion électronique. Les deux familles importantes de méthodes qui ont été développées sont celles d'interaction de configurations (CI) [7,8] et la théorie des perturbations Moller-Plesset d'ordre n (MPn) et les méthodes DFT.

L'énergie de corrélacion d'un système correspond à la différence entre l'énergie Hartree-Fock et l'énergie exacte non-relativiste du système :  $E_{\text{cor r}} = E_{\text{HF}} - E$

**I.2.1. Méthode de Möller-Plesset d'ordre 2 (MP2) :**

Cette approche, proposée par Moller-Plesset [9], tient compte de la corrélation électronique en utilisant la théorie des perturbations. L'hamiltonien total est séparé en deux: une partie qui a les fonctions propres et les valeurs propres du déterminant Hartree-Fock, et une partie perturbée V. L'énergie est alors exprimée comme une somme de ces deux contributions :

$$H|\phi_i\rangle = (F + V)|\phi_i\rangle = E_i|\phi_i\rangle \quad (16)$$

F étant l'opérateur de Fock, et V étant le potentiel de corrélation défini par :

$$V = \frac{1}{2} \sum_i \sum_{j \neq i} \frac{1}{r_{ij}} - \sum_i (2J_i - K_i) \quad (17)$$

On connaît déjà les solutions de l'équation :

$$F|\phi_i\rangle = E_i^{(0)}|\phi_i\rangle \quad (18)$$

La théorie des perturbations stipule que si V est petit par rapport à F, on peut alors développer l'opérateur  $H = F + \lambda V$  en série de Taylor selon  $\lambda$ , d'où :

$$E_i = E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \dots \quad (19)$$

$$\phi_i = \phi_i^{(0)} + \lambda \phi_i^{(1)} + \lambda^2 \phi_i^{(2)} + \dots \quad (20)$$

Et on peut ainsi montrer que :

$$E_i^{(1)} = \langle \phi_i^{(0)} | V | \phi_i^{(0)} \rangle \quad (21)$$

$$\begin{aligned} E_n^2 &= \frac{1}{4} \sum_r \sum_s \sum_t \sum_u \frac{\langle rs / tu \rangle^2}{e_r + e_s + e_t + e_u} \\ E_{MP2} &= E_n^{HF} + E_n^2 \\ \Rightarrow E_{MP2} &= \sum_k e_k - \frac{1}{2} \sum_i \sum_j (J_{ij} - K_{ij}) + \frac{1}{4} \sum_r \sum_s \sum_t \sum_u \frac{\langle rs / tu \rangle^2}{e_r + e_s + e_t + e_u} \end{aligned} \quad (22a)$$

$$\begin{aligned} \langle rs / tu \rangle = & \iint \phi_r(1) \phi_s(1) \frac{1}{r_{12}} \phi_t(1) \phi_u(2) d\tau_1 d\tau_2 \\ \text{Où :} & \\ & - \iint \phi_r(1) \phi_s(2) \frac{1}{r_{12}} \phi_u(1) \phi_t(2) d\tau_1 d\tau_2 \end{aligned} \quad (22b)$$

De la même manière, on obtient les autres ordres de perturbation  $E_i^{(1)}$ ,  $E_i^{(2)}$  et  $E_i^{(3)}$  ...

La perturbation la plus couramment utilisée est la perturbation au deuxième ordre. Elle est connue sous le nom de « MP2 » [10]. Cette méthode permet de récupérer une grande partie de l'énergie de corrélation. Elle n'est cependant pas variationnelle et est basée sur une référence unique (la fonction d'onde de Hartree-Fock). Cette méthode est très efficace et requière dans la pratique des temps de calculs acceptables, proportionnels à  $N^5$ , où N est le nombre d'électrons du système étudié.

Les techniques post-HF sont en général très efficaces pour retrouver l'énergie de corrélation, mais cependant à l'heure actuelle elles sont, pour la majeure partie d'entre-elles, trop lourdes pour être applicables à des systèmes dont le nombre d'atomes est grand. Il s'est ainsi parallèlement développé à ces techniques un modèle alternatif qui a atteint le statut de théorie à la fin des années 60. La théorie de la fonctionnelle de la densité (DFT) est actuellement la seule permettant l'étude de systèmes chimiques de grande taille avec la prise en compte des effets de la corrélation électronique de manière satisfaisante.

### **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 basée sur le postulat proposé par Thomas et Fermi qui dit que les propriétés électroniques peuvent être décrites en terme de fonctionnelles de la densité électronique, en appliquant localement des relations appropriées à un système électronique homogène [11].

Hohenberg et Kohn, en 1964 [12], ont repris la théorie de Thomas-Fermi et ont montré qu'il existe une fonctionnelle de l'énergie  $E[\rho(r)]$  associée à un principe variationnel, ce qui a permis de jeter les bases de la théorie de la fonctionnelle de la densité

**1<sup>er</sup> théorème de Hohenberg et Kohn:**

Enoncé : « 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  $\rho_0(x, y, z)$  » [12].

On peut alors utiliser la densité électronique comme variable de base pour la résolution de l'équation de Schrödinger électronique. Etant donné que  $\rho(r)$  est liée au nombre d'électrons du système, elle peut en effet également déterminer les fonctions propres de l'état fondamental ainsi que toutes les autres propriétés électroniques du système ; si  $n$  est le nombre d'électrons du système, on a que :

$$\int \rho(\mathbf{r}) \, d\mathbf{r} = n \quad (23)$$

Rappelons l'expression de l'Hamiltonien électronique d'un système polyélectronique :

$$H = -\frac{1}{2} \sum_i^n \Delta_i + \sum_{i>j}^n \frac{1}{r_{ij}} + \sum_i^n v(r_i) \quad (24)$$

avec : 
$$v(\mathbf{r}_i) = -\sum_{\alpha} \frac{Z_{\alpha}}{r_{i\alpha}} \quad (25)$$

$v(\mathbf{r}_i)$  : potentiel externe de l'électron  $i$ .

Ce potentiel correspond à l'attraction de l' $e^-$  ( $i$ ) avec tous les noyaux qui sont externes par rapport au système d'électrons.

Connaissant la densité électronique  $\rho(r)$  d'un système, on a donc accès au nombre d'électrons, au potentiel externe, ainsi qu'à l'énergie totale  $E[\rho(r)]$ . Celle-ci peut s'écrire comme une somme de trois fonctionnelles :

$$E[\rho] = V_{ne}[\rho] + T[\rho] + V_{ee}[\rho] \quad (26)$$

avec :

$$V_{ne}[\rho] = \int \rho(r)v(r)dr \quad (27)$$

$$T[\rho] = \int \left[ -\frac{1}{2} \nabla^2 \rho(r) \right] dr \quad (28)$$

$$J[\rho] = \frac{1}{2} \iint \frac{1}{r_{12}} \rho(r_1)\rho(r_2) dr_1 dr_2$$

$$V_{ee}[\rho] = \quad (29)$$

$$K[\rho] = \frac{1}{4} \iint \frac{1}{r_{12}} \rho(r_1, r_2)\rho(r_1 r_2) dr_1 dr_2$$

On constate que le terme  $V_{ee}[\rho]$  est composé de deux parties; la première correspond à l'interaction coulombienne classique  $J[\rho]$ , et la seconde partie dite non-classique est appelée « énergie d'échange et de corrélation  $K[\rho]$  ».

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

$$E_0[\rho] = \int \rho_0(r)v(r)dr + F[\rho_0] \quad (30)$$

où  $F[\rho_0] = T[\rho_0] + V_{ee}[\rho_0]$  est la fonctionnelle universelle de Hohenberg et Kohn.

$F[\rho_0]$  est une fonctionnelle prenant en compte tous les effets interélectroniques ; elle est indépendante du potentiel externe, et elle est donc valable quelque soit le système étudié. La connaissance de  $F[\rho]$  permet l'étude de tous les systèmes moléculaires, malheureusement la forme exacte de cette fonctionnelle est à l'heure actuelle loin d'être connue, et il faut avoir recours à des approximations La fonctionnelle  $F[\rho]$  est inconnue.

- 2<sup>ème</sup> 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 :

$$E_0 \leq E[\tilde{\rho}] \quad (31)$$

La condition pour qu'une fonctionnelle telle que  $E_0$  admette un extremum est que sa dérivée fonctionnelle s'annule. D'après la définition :

$$\delta E = \int \frac{\delta E}{\delta \rho} d\rho dr = 0 \quad (32)$$

La relation  $\delta E=0$  est donc vérifiée si :  $\frac{\delta E}{\delta \rho} = 0$

La résolution du problème consiste dès lors à chercher à minimiser  $E[\rho]$  avec la contrainte  $\int \tilde{\rho}(r) dr = n$ .

Finalement on obtient l'équation fondamentale de la DFT :

$$\mu = \frac{\delta E[\rho]}{\delta \rho} = v(r) + \frac{\delta F_{HK}[\rho]}{\delta \rho} \quad (33)$$

où la quantité  $\mu$  est appelée « potentiel chimique » du système.

### 1.3.2. Méthode de Kohn et Sham :

Les théorèmes de Hohenberg et Kohn ne donnent pas une procédure pour calculer l'énergie  $E_0$  à partir de  $\rho$ , ni comment déterminer  $\rho$  sans déterminer, au préalable, la fonction d'onde. C'est Kohn et Sham, en 1965, qui ont élaboré une méthode pratique pour trouver  $E_0$  à partir de  $\rho_0$  [13]. Ils ont considéré un système fictif de référence, noté  $s$ , constitué par les  $n$  électrons non interagissant. Kohn et Sham ont reformulé le problème de manière à ce que le système de référence ait la même densité électronique que l'état fondamental du système étudié.

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

$$\hat{H}_s = \sum_{i=1}^n \left[ -1/2 \nabla_i^2 + v_s(r_i) \right] = \sum_{i=1}^n h_i^{KS} \quad (34)$$

avec :

$$h_i^{KS} = -1/2 \nabla_i^2 + v_s(r_i) \quad (35)$$

Par conséquent, les équations de Kohn et Sham, pour l'électron i, peuvent s'écrire comme suit :

$$h_i^{KS} \theta_i^{KS} = \epsilon_i^{KS} \theta_i^{KS} \quad (36)$$

$\theta_i^{KS}$  : Orbitale de Kohn et Sham de l'électron i.

- **Terme d'échange-corrélation:**

Soit  $\Delta T$  la différence de l'énergie cinétique entre le système réel (électrons interagissants) et le système fictif (électrons non interagissant). La quantité  $\Delta T$  étant cependant faible:

$$\Delta T = T[\rho] - T_s[\rho] \quad (37)$$

donc :

$$\Delta V_{ee} = V_{ee}[\rho] - 1/2 \iint \frac{\rho(r_1)\rho(r_2)}{r_{12}} dr_1 dr_2 \quad (38)$$

$\Delta V_{ee}$  est la différence entre la vraie répulsion électron-électron et la répulsion coulombienne entre deux distributions de charge ponctuelle. L'énergie s'écrit alors :

$$E[\rho] = \int \rho(r)v(r)dr + T_s[\rho] + 1/2 \iint \frac{\rho(r_1)\rho(r_2)}{r_{12}} dr_1 dr_2 + \Delta T[\rho] + \Delta V_{ee}[\rho] \quad (39)$$

La fonctionnelle d'énergie d'échange- corrélation est définie comme suit :

$$E_{xc}[\rho] = \Delta T[\rho] + \Delta V_{ee}[\rho] \quad (40)$$

$$E[\rho] = \int \rho(r)v(r)dr + T_s[\rho] + 1/2 \iint \frac{\rho(r_1)\rho(r_2)}{r_{12}} dr_1 dr_2 + E_{xc}[\rho] \quad (41)$$

Le problème majeur pour les calculs DFT, selon le schéma de Kohn et Sham, est de trouver une bonne approximation pour l'énergie échange- corrélation  $E_{xc}$ .

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

$$\left[ -1/2 \nabla_1^2 - \sum_{\alpha} \frac{Z_{\alpha}}{r_{1\alpha}} + \int \frac{\rho(r_2)}{r_{12}} dr_2 + v_{xc}(1) \right] \theta_i^{KS}(1) = \epsilon_i^{KS} \theta_i^{KS}(1) \quad (42)$$

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

$$v_{xc}(r) = \frac{\partial E_{xc}[\rho(r)]}{\partial \rho(r)} \quad (43)$$

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

L'équation aux valeurs propres peut également s'écrire sous la forme:

$$\left[ -\frac{1}{2} \nabla_1^2 + V_{eff} \right] \theta_i^{KS}(1) = \epsilon_i^{KS} \theta_i^{KS}(1) \quad (44)$$

$V_{eff}$  est appelé potentiel effectif.

Les orbitales de KS permettent de calculer la densité électronique  $\rho_0$  à l'aide de la formule suivante :

$$\rho_0(r) = \rho_s(r) = \sum_{i=1}^n |\theta_i^{KS}|^2 \quad (45)$$

Pratiquement, on choisit une densité d'essai à partir de laquelle on calcule un potentiel effectif  $v_{\text{eff}}(r)$ . En injectant  $v_{\text{eff}}(r)$  dans l'expression 44 on obtient une nouvelle densité électronique. La convergence est alors atteinte lorsque le potentiel effectif ne varie plus.

### 1.3.3. Approximation de la densité locale LDA:

Hohenberg et Kohn ont montré que si  $\rho$  varie extrêmement lentement avec la position, l'énergie d'échange-corrélation  $E_{xc}[\rho_s]$  peut s'écrire comme suit :

$$E_{xc}^{LDA}[\rho] = \int \rho(r) \mathcal{E}_{xc}(\rho) dr \quad (46)$$

$\mathcal{E}_{xc}$  : étant l'énergie d'échange-corrélation par électron. Cette quantité est exprimée comme la somme des deux contributions (énergie d'échange  $\epsilon_x$  et énergie de corrélation  $\epsilon_c$ ) telle que :

$$\mathcal{E}_{xc}(\rho) = \epsilon_x(\rho) + \epsilon_c(\rho) \quad (47)$$

avec :

$$\epsilon_x(\rho) = -\frac{3}{4} \left( \frac{3}{\pi} \right)^{1/3} (\rho(r))^{1/3} \quad (48)$$

donc :

$$E_x^{LDA} = \int \rho \epsilon_x dr = -\frac{3}{4} \left( \frac{3}{\pi} \right)^{1/3} \int [\rho(r)]^{4/3} dr \quad (49)$$

Le terme de corrélation  $\epsilon_c(\rho)$  est exprimé par la formule de Vosko, Wilk, et Nusair (VWN) [14]. Cette formule assez compliquée est donnée dans la référence [15].

### I.3.4. Approximation du Gradient Généralisé (GGA) :

La densité électronique d'un système est non seulement pas uniforme, mais peut même varier très rapidement dans l'espace (lorsqu'on passe d'une couche électronique à l'autre dans un atome, ou lorsqu'on passe d'un atome à l'autre dans une molécule). La première amélioration que l'on puisse apporter à la méthode LDA consiste donc à exprimer la fonctionnelle d'énergie d'échange-corrélation en fonction de la densité électronique et de son gradient. Cette technique est appelée « approximation de l'expansion du gradient » (GEA).

$$E_{xc}^{GGA} = \int \varepsilon_{xc}^{GGA}(\rho, \nabla\rho) dr \quad (50)$$

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

De nombreuses fonctionnelles ont été développées depuis, tant pour l'échange que pour la corrélation. Parmi les plus connues et les plus utilisées on peut citer les fonctionnelles d'échange de Becke (B88) [16] et de Perdew et Wang (PW91) [17]. Pour la corrélation, on dispose, entre autres, des fonctionnelles de Perdew (P86) [17], de Lee, Yang et Parr (LYP) [18] et de Perdew et Wang (PW91) [17]. Toutes ces fonctionnelles permettent une amélioration de l'estimation des énergies de liaison dans les molécules, ainsi que des barrières d'énergie par rapport à l'approximation locale LDA.

#### **Remarque :**

Il est fort intéressant de noter que n'importe quelle fonctionnelle d'échange (ou de corrélation) d'une méthode peut être combinée avec n'importe quelle fonctionnelle d'échange (ou de corrélation) d'une autre méthode. Par exemple, la notation BLYP/6-31G\* indique qu'il s'agit d'un calcul DFT avec la fonctionnelle d'échange de Becke 1988 et la fonctionnelle de corrélation de Lee-Yang-Parr, avec les orbitales de Kohn et Sham (KS) développées sur base gaussienne de type 6-31G\*.

### I.3.5. Fonctionnelle hybride B3LYP :

La fonctionnelle hybride B3LYP (Becke 3-parameters Lee-Yang-Parr) est une fonctionnelle à trois paramètres combinant les fonctionnelles d'échange local, d'échange de Becke et d'échange HF, avec les fonctionnelles de corrélation locale (VWN) et corrigée du gradient de Lee, Yang et Parr :

$$E_{xc}^{B3LYP} = (1 - a_0 - a_x) E_x^{LDA} + a_0 E_x^{HF} + a_x E_x^{B88} + a_c E_c^{LYP} (1 - a_c) + E_c^{VWN} \quad (51)$$

où les paramètres  $a_0$ ,  $a_x$  et  $a_c$  ont été ajustés respectivement à 0,20, 0,72 et 0,81 [19].

### I.4. Bases d'orbitales atomiques :

Comme il l'a été dit, le choix de la base de fonctions représentant les orbitales atomiques est important car il peut influencer tant la précision des résultats obtenus que les temps de calculs. 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* [20] qui sont Les meilleures OA analytiques définies par:

$$\Psi_{nlm} = N_n r^{n^*-1} \exp(-\zeta r) Y_{lm}(\theta, \phi) \quad (52)$$

Où  $N_n$  est le facteur de normalisation et  $\zeta$  est l'exponentielle orbitale (exposant de Slater, déterminant la taille de l'orbitale.),  $Y_{lm}(\theta, \phi)$  sont les harmoniques sphériques.

Les fonctions de types Slater (STOs) présentent une forme analytique simple mais elles ne sont pas utilisées à grande échelle dans les programmes moléculaires *ab initio*. Cela est dû à la complexité du calcul d'intégrales moléculaires sur la base STO.

Les programmes *ab initio* de chimie quantique (Gaussian par exemple), utilisent le second type de bases, fonctions gaussiennes (GTOs) proposées par Boys [21].

$$g(\alpha, \vec{r}) = c x^n y^l z^m \exp(-\alpha r^2) \quad (53)$$

Dans cette équation,  $\alpha$  est une constante déterminant la taille de la fonction,  $N$  constante de normalisation et  $n, l, m$  sont des nombres simulant les nombres quantiques.

Par exemple :

$$\text{fonction } s : \begin{cases} i = 0 \\ j = 0 \\ k = 0 \end{cases} \quad \text{fonction } p_x : \begin{cases} i = 1 \\ j = 0 \\ k = 0 \end{cases}$$

Les gaussiennes sont des fonctions très populaires en chimie quantique, spécialement pour les méthodes *ab initio* [22], car le produit de deux gaussiennes centrées sur deux atomes A et B différents peut s'écrire à l'aide d'une seule gaussienne centrée en un point situé sur le segment AB. Le calcul des intégrales biélectroniques en ressort ainsi considérablement simplifié.

Les bases gaussiennes ont, par contre, une assez mauvaise représentation des orbitales atomiques car elles n'ont pas un comportement exact à l'origine (dérivée devant être nulle), ni aux grandes distances (décroissance trop rapide avec  $r$ ). Pour compenser la représentation incomplète des orbitales atomiques des fonctions gaussiennes, on utilise donc des combinaisons linéaires de gaussiennes comme fonctions de base. Ces fonctions sont appelées « fonctions gaussiennes contractées ». Il faut en général utiliser trois fonctions gaussiennes pour que l'ajustement des parties radiales soit satisfaisant.

$$G^{CGTO} = \sum_{\lambda=1}^k d_{\lambda} g_{\lambda}^{PGTO} \quad (54)$$

$d_{\lambda}$  étant le coefficient de contraction de la gaussienne primitive  $g_{\lambda}$ .

La plus simple des bases gaussiennes est la base **STO-3G**, aussi appelée « base minimale ». Le sigle « 3G » signifie que les orbitales de type Slater (STO) sont représentées par trois fonctions gaussiennes. La base **3-21G** est une *Split Valence-Double Zeta* (SV-DZ),

où chaque orbitale atomique des couches internes est décrite par une contraction de 3 gaussiennes primitives. Les orbitales de la couche de valence sont réparties en deux groupes : les orbitales proches du noyau sont décrites par une contraction de 2 primitives, et les orbitales éloignées par une seule gaussienne primitive.

La base **6-311G** est une *Split Valence-Triple Zeta* (SV-TZ) dans laquelle les orbitales de cœur (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 primitives respectivement.

Pour une plus grande flexibilité on peut encore rajouter des fonctions de *polarisation* (déformation du nuage électronique, et des distorsions dues à l'environnement). La dénomination la plus ancienne est l'ajout d'un astérisque sur la base en question (par exemple 6-31G\*), et dans une désignation plus récente, le caractère de la fonction ajoutée est explicitement donné : 6-31G(d). La base 6-31G\* ou 6-31G(d) signifie ainsi qu'un jeu de fonctions « d » a été ajouté à tous les atomes (sauf H) dans la molécule, alors que 6-31G\*\* ou 6-31G(p,d) signifie qu'un jeu de fonctions « p » a été ajouté aux hydrogènes et que des fonctions « d » ont été ajoutées aux autres atomes.

Un autre type de fonctions est indispensable à inclure dans la base d'orbitale atomique chaque fois que le phénomène physique décrivant la propriété étudiée nécessite une bonne description de l'espace situé au-delà des orbitales de valence (espace diffus). Ce sont les fonctions *diffuses*, qui augmentent la taille du nuage électronique. Pour les espèces ayant des doublets libres et les espèces chargées (anions), la présence d'orbitales diffuses est indispensable. On note par le signe +, signifiant la présence d'orbitales diffuses, celle des orbitales de polarisation est notée par un astérisque (\*).

Par exemple la base **6-31+G\*** désigne une base SV-DZ 6-31G avec des orbitales diffuses, et de polarisation sur les atomes lourds ; **6-311++G\*** est une base SV-TZ 6-311G avec des orbitales diffuses sur tous les atomes, et des orbitales de polarisation uniquement

sur les atomes lourds. D'autres bases gaussiennes ont été proposées par Dunning et Huzinaga [23,24]. Malgré les divers perfectionnements apportés à la base gaussienne, l'utilisation de ces bases présente plusieurs inconvénients [25].

### 1.5. Les effets de solvation:

Il est intéressant de pouvoir comparer des propriétés moléculaires obtenues à l'aide d'outils quantiques à des données expérimentales. Très souvent ces dernières sont obtenues en phase condensée. L'effet de solvant joue un rôle important notamment avec la formation de liaisons hydrogènes, les interactions hydrophobes ou hydrophiles, . . . Il est donc nécessaire d'inclure ces effets de manière précise dans les calculs quantiques. Les méthodes de chimie théorique sont de très puissants outils quant à la compréhension des interactions soluté-solvant, la prédiction des structures, des spectres, ainsi que pour la description de la réactivité en phase liquide [26]. Il existe deux grandes manières de représenter cet environnement : implicite et explicite. La différence fondamentale entre les deux schémas est que le modèle explicite décrit les molécules de solvant de manière atomique tandis que le modèle implicite représente le solvant comme un milieu homogène dit continuum diélectrique. L'idée de modéliser les interactions électrostatiques dues au solvant en plaçant le soluté dans une cavité de taille définie date des travaux de Kirkwood [27] et Onsager concernant les effets de solvation sur les molécules polaires [28-30]. Dans ce dernier, le soluté, traité de manière quantique, est placé dans une cavité entourée de molécules de solvant considérées comme un continuum.

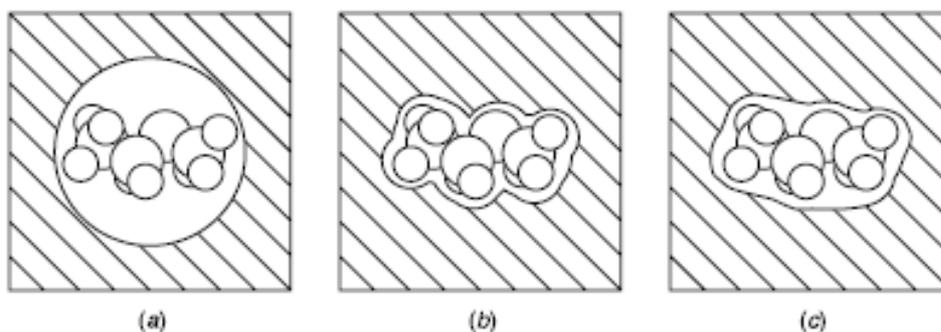
Le modèle implicite le plus simple est celui de Born (1912) qui considère une charge ponctuelle placée dans une cavité de solvant de forme sphérique (**Figure 1**). Onsager (1936) l'étend à un dipôle placé dans une cavité sphérique (**Figure 1**) [29]. Alors que le modèle de Born n'est approprié que pour les espèces chimiques possédant une charge formelle, celui d'Onsager est applicable aux nombreuses molécules.



*Figure 1: Modèle de Born et d'Onsager*

### I.5.1. Modèle SCRF

Dans un premier temps, nous allons décrire les différents termes énergétiques et le bilan énergétique du modèle de continuum tenant compte de différents types d'interactions. La première étape implique la création d'une cavité dans le continuum de solvant (Figure 2).



*Figure 2. (a) Schématisation d'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, et (c) une surface accessible au solvant.*

En pratique, la cavité est construite à partir d'un ensemble de sphères centrées sur les noyaux et ayant un rayon de type van der Waals. La formation de cette cavité coûte une certaine quantité d'énergie (positive):  $\Delta G_{\text{cav}}$ , l'énergie libre de cavitation. Cette quantité dépend de la nature du solvant ainsi que de la topologie de la cavité. En second temps, le soluté est placé dans la cavité et celui-ci interagit avec le continuum. On distingue trois types d'interactions soluté-continuum: électrostatiques  $\Delta G_{\text{elec}}$ , répulsives et dispersives. Les deux dernières sont calculées grâce à des relations empiriques.

Le processus d'interaction SCRF ( Self-Consistent Reaction Field) d'une molécule dans sa cavité est autocohérent résolu: la distribution de charge de ce soluté polarise la surface de la cavité du continuum. Des charges apparaissent donc à l'interface soluté-continuum. Ces dernières induisent un potentiel de réaction qui, à son tour, modifie la densité électronique du soluté. Ensuite, le continuum doit s'adapter et il y a une nouvelle modification de la distribution de charges à l'interface soluté-continuum . . . et ainsi de suite jusqu'à l'obtention d'une convergence électrostatique entre la distribution de charges propre au soluté et celle de la surface de la cavité. Ce terme énergétique, toujours négatif, est la contribution électrostatique ( $\Delta G_{\text{elec}}$ ).

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

Au final, l'énergie totale d'interactions s'exprime en un terme électrostatique et en trois termes non-électrostatiques :

$$\Delta G_{\text{solv}} = \Delta G_{\text{cav}} + \Delta G_{\text{elec}} + \Delta G_{\text{disp}} + \Delta G_{\text{rép}} \quad (55)$$

Pour résoudre le problème les modèles de continuum utilisent l'équation de Poisson de l'électrostatique classique :

$$-\nabla[\epsilon_r(r) \nabla V(r)] = 4\pi\rho_m(r) \quad (56)$$

$\epsilon_r(r)$  est une fonction diélectrique dans le milieu et le potentiel électrostatique total  $V(r)$  est la somme du potentiel électrostatique  $V_\rho(r)$  généré par la distribution de charges du soluté  $\rho$  et le potentiel de réaction  $V_\sigma(r)$  créé par la polarisation du milieu diélectrique :

$$V(r) = V_\rho(r) + V_\sigma(r) \quad (57)$$

La constante diélectrique ( $\epsilon_r(r)=\epsilon_s/\epsilon_0$ ) peut prendre deux valeurs:

$$\epsilon_r(r)= 1, \quad \text{si } r \in v_{\text{int}}$$

$$\epsilon_r(r)=\epsilon, \quad \text{si } r \in v_{\text{ext}}$$

où  $v_{\text{int}}$  ( $v_{\text{ext}}$ ) est le volume à l'intérieur (à l'extérieur) de la cavité. Pour les deux régions, l'équation 2 devient :

$$-\nabla^2 V(r) = 4\pi\rho_m(r), \quad r \in v_{\text{int}} \quad (58)$$

$$-\epsilon \nabla^2 V(r) = 0, \quad r \in v_{\text{ext}} \quad (59)$$

Les éqs (4) et (5) doivent être accompagnées de conditions frontières, à l'infini mais aussi et surtout à la surface de la cavité,  $|v_{\text{int}} - v_{\text{ext}}| = 0$  afin de garantir la continuité du potentiel.

La distribution de charge surfacique  $\sigma(r_s)$  peut être exprimée ainsi en termes de quantités aisément calculables avec des processus actuels :

$$\sigma(r_s) = \frac{(\epsilon - 1)}{4\pi\epsilon} E(r_s) \quad (60)$$

où  $E(r) = \left( \frac{\partial V(r)}{\partial n} \right)_k$  est le champ électrique perpendiculaire à la cavité.

### I.5.2. Modèle PCM ( Polarizable continuum model):

Le modèle PCM [31-33] permet de travailler avec des cavités de forme plus réaliste, avec une surface découpée en aires élémentaires constituée de petits polygones sphériques (appelées tesserae) assez petites pour considérer constant le  $\sigma(s)$  de chacun de ces tesserae. Dans ce modèle, la molécule se trouve dans une cavité formée par des sphères emboîtées et centrées sur les atomes (**Figure 3**). 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. Les triangles à l'intersection de deux ou plusieurs sphères sont modifiées avec un algorithme qui conserve les caractéristiques de la surface de la cavité et la distribution de charge  $\sigma(s)$  [31].



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

- **Algorithme PCM :**

Dans cette méthode, la fonction d'onde électronique du système est déterminée à partir d'un hamiltonien électronique modifié par addition d'un terme électrostatique dû à la présence de charges surfaciques ( $\sigma_s(r_s)$ ) à l'interface entre le soluté et le solvant. L'hamiltonien de solvation s'écrit alors :

$$H = H_0 + V_\sigma \quad (61)$$

$$\text{avec } V_\sigma(r_s) = \int \frac{\sigma_s(r_s)}{|r - r_s|} dr_s$$

où  $H_0$  est l'hamiltonien électronique dans le vide et  $V_\sigma$  est le potentiel électrostatique dû à la solvation (c'est une perturbation). L'hamiltonien dépend donc de ses fonctions propres, d'où une résolution par itération.

Le but du processus itératif est l'évaluation de  $\sigma(r_k)$  de chaque tesserae, avec une surface  $\Delta S_k$  et une charge  $q_k$  au point  $r_k$  suivant la relation :

$$q_k = \sigma(r_k) \Delta S_k \quad (62)$$

Ce processus itératif se fait selon les étapes suivantes :

*Etape 1:*

- On part d'une valeur d'essai de  $\sigma(r_k)$  correspondant à un potentiel dû à la distribution de charges du soluté uniquement  $V_\sigma(r)=0$

d'où :

$$V(r) = V_\rho(r) \tag{63}$$

On appelle  $\sigma_k^{00}$  les charges surfaciques correspondant à cette approximation. Le premier indice 0 correspond aux ponctuelles de soluté seul. Le second indice 0 correspond au départ, donc au fait que l'on suppose :  $V_\sigma(r) = 0$

On a alors :

$$\sigma_k^{00} = - \left[ \frac{\epsilon - 1}{4\pi} \right] \left( \frac{\partial V(r)}{\partial n} \right)_k \tag{64}$$

- On obtient alors:

$$q_k^{00} = \sigma_k^{00}(s) \Delta S_k \tag{65}$$

Ces charges produisent au centre des éléments de surface une contribution supplémentaire au potentiel électrostatique et au champ électrique d'où :

$$V^{00}(r) = V_\rho^{00}(r) + V_\sigma^{00}(r) \tag{66}$$

On calcul alors une nouvelle distribution de charge surfacique  $\sigma_k^{01}$

$$\sigma_k^{01} = - \left[ \frac{\epsilon - 1}{4\pi} \right] \left( \frac{\partial V(r_s)}{\partial n} \right)_k \tag{67}$$

alors :

$$q_k^{01} = \sigma_k^{01}(s) \Delta S_k \tag{68}$$

L'itération est effectuée jusqu'à atteindre la convergence pour laquelle  $V_{\sigma}^{0f}(r)$  est obtenu à partir des  $q_k^{0f}$

On ajoute alors le potentiel  $V_{\sigma}(r)$  à l'hamiltonien du soluté  $H_0 : H = H_0 + V_{\sigma}$

On résout alors les équations HF ou Kohn-Sham (dans le cas de la DFT) avec cet hamiltonien.

**Etape 2 :**

On obtient alors une nouvelle distribution de charge pour le soluté à partir de laquelle on déduit un nouveau jeu de départ de charges surfaciques  $q_k^{10}$  et on itère jusqu'à obtenir  $q_k^{1f}$  et donc  $V_{\sigma}^{1f}(r)$ . On répète la même procédure jusqu'à atteindre la convergence globale et donc l'obtention de  $\psi^f$ .

On peut alors définir une énergie libre électrostatique  $G_{elec}$  comme :

$$H\psi = [H^0 + V^R]\psi = E\psi$$

En résolvant cette équation on obtient:

$$\begin{aligned} G_{el} &= E - \frac{1}{2} \langle \Psi^f | V^R | \Psi^f \rangle = \langle \Psi^f | H^0 | \Psi^f \rangle + \langle \Psi^f | V^R | \Psi^f \rangle - \frac{1}{2} \langle \Psi^f | V^R | \Psi^f \rangle \\ &= \langle \Psi^f | H^0 | \Psi^f \rangle + \frac{1}{2} \langle \Psi^f | V^R | \Psi^f \rangle \end{aligned} \tag{69}$$

L'énergie libre de solvation s'écrit alors :

$$\Delta G_{el} = G_{el} - \langle \Psi^0 | H^0 | \Psi^0 \rangle$$

Soit :

$$\Delta G_{el} = \langle \Psi^f | H^0 | \Psi^f \rangle + \frac{1}{2} \langle \Psi^f | V^f | \Psi^f \rangle - \langle \Psi^0 | H^0 | \Psi^0 \rangle \tag{70a}$$

Avec :

$$\langle \Psi^f | V^R | \Psi^f \rangle = \int V(r) \rho(r) dr \tag{70b}$$

### I.5.3. Les termes non-électrostatiques

Les termes non-électrostatiques sont de trois types: cavitation, dispersion et répulsion [34].

- **Le terme de cavitation**

Le transfert d'une molécule d'une phase gazeuse à une phase liquide se produit en deux étapes. La première est la formation de la cavité dans la solution et la seconde est l'introduction de la molécule de soluté à l'intérieur de la cavité. Ces deux étapes résultent en un excès d'énergie libre. Le terme de cavitation  $\Delta G_{cav}$  correspond à la première étape. Selon Uhlig [35], le travail pour former une cavité macroscopique est fonction de la tension de surface  $\gamma$  du liquide et de la surface de la cavité  $S_m$  :

$$\Delta G_{cav} = \gamma S_m \quad (71)$$

- **Le terme de dispersion**

Les interactions de Van der Waals sont la combinaison de deux termes : **un terme répulsif** traduisant le recouvrement des nuages électroniques de même polarité à courtes distances et un **terme attractif** dû aux forces dispersives qui apparaissent instantanément durant les fluctuations des nuages électroniques.

Les forces de dispersion [36] sont des forces faibles intermoléculaires créées par des dipôles instantanés : le dipôle instantané d'une molécule A induit un dipôle instantané sur une molécule B et interagit avec lui. Ces forces représentent en général la plus importante composante des forces de Van der Waals (entre 0.5 et 40 kcal/mol). Elles apparaissent lorsque la densité électronique d'une molécule n'est pas équitablement répartie autour de celle-ci, ce qui crée ainsi un léger moment dipolaire. Ces moments dipolaires instantanés varient très rapidement au cours du temps. Ainsi à chaque distribution inhomogène se crée un moment dipolaire instantané qui peut interagir avec les moments dipolaires qu'il induit sur les molécules voisines. . Ce qui permet d'exprimer l'énergie ou potentiel de dispersion par relation:

$$E_{disp}^{VDW}(r) = -Ctes \cdot \left(\frac{\sigma}{r}\right)^6 \quad (72)$$

- **Le terme de répulsion**

Les forces répulsives sont les plus locales et les plus intenses. Le potentiel ou l'énergie de répulsion est une fonction rapidement croissante lorsque la distance  $r$ , séparant les deux atomes, diminue.

$$E_{rép}^{VDW}(r) = \left( \frac{\sigma}{r} \right)^n \quad (73)$$

avec une valeur pouvant  $n$  allant de 9 à 12.

Dans le PCM, le terme de répulsion est évalué par l'approche d'Amovilli et Mennucci [37].

$$G_{rép} = \rho_s \int dr U_{ms}^{rép}(r) g_{ms}(r) \quad (74)$$

avec  $m$  le soluté,  $s$  le solvant,  $r$  est un ensemble de coordonnées qui définissent la géométrie du complexe  $ms$ ,  $\rho_s$  est la densité du solvant et  $g_{ms}$  est une fonction de corrélation qui vaut 0 à l'intérieur de la cavité et 1 à l'extérieur. En pratique, on estime le potentiel de répulsion ( $U_{rep}$ ) suivant une approche de type Lennard-Jones.

La plus connue des fonctions de potentiel de type Van der Waals est la fonction de Lennard-Jones, dans laquelle les énergies de dispersion et de répulsion sont définie par une seule et même expression qui s'écrit pour les deux molécules, neutres et non polaires,  $i$  et  $j$  sous la forme:

$$V^{VDW-LJ}(r) = E_{rép} + E_{disp} = \sum_i \sum_j 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (75)$$

Cette équation contient deux paramètres ajustables : le diamètre de collision ( $\sigma_{ij}$ ) qui est la distance minimale d'approche entre les deux atomes  $i$  et  $j$  pour laquelle l'énergie entre deux atomes est nulle et la profondeur du puits ( $\varepsilon_{ij}$ ) qui représente le minimum de l'énergie potentielle, ce qui correspond à l'interaction la plus stable.

Enfin, notons que bien que le potentiel de Lennard-Jones soit le plus utilisé pour décrire les interactions de Van der Waals, vu sa simplicité et le peu de paramètres qu'il fait intervenir.

## I.6. L'analyse NBO (Natural Bond Orbital)

La résolution de l'équation de Schrödinger conduit à la fonction d'onde qui contient toute l'information sur la structure électronique de la molécule. La structure électronique de la molécule est souvent envisagée comme une perturbation de la structure électronique des atomes constitutifs suite à la création de liaisons entre certains atomes. Le chimiste essaie de décrire la structure électronique délocalisée de la molécule à l'aide des blocs constitutifs atomiques pour mettre en évidence les schémas de liaison dans la molécule. Il se pose alors le problème de la partition de l'information moléculaire contenue dans  $\Psi_i$  sur chacun des atomes constitutifs de la molécule.

L'objectif de la méthode NBO est de proposer un schéma de partition de l'espace fonctionnel associé à la fonction d'onde permettant de décrire la structure électronique de la molécule à l'aide du formalisme simple proposé par Lewis :

- doublets d'électrons de cœur
- doublets non liant dans la couche de valence
- doublets de liaison résultant de la mise en commun par deux atomes de deux électrons situés dans des orbitales hybrides.

L'information sur laquelle se fait la partition est contenue dans la fonction d'onde moléculaire et par conséquent des écarts à l'idéalité de la structure de Lewis [38] sont attendus et devront pouvoir être décrits par la méthode NBO.

### I.6.1. La théorie des perturbations pour les orbitales dans le cadre Hartree-Fock:

#### Interaction donneur-accepteur

En utilisant le cadre Hartree-Fock, les équations de la théorie des perturbations [39,40] peuvent être écrites en termes d'orbitales  $\phi_i$ , les énergies d'orbitales  $\epsilon_i$  et l'occupation des orbitales ( $n_i$ ). Ces équations de perturbation nous permettent d'étudier les interactions électroniques d'un système à plusieurs électrons (décrit par l'opérateur de Fock  $F$ ) et en considérant le système simple non perturbé (décrit par l'opérateur non perturbé de Fock  $F^{(0)}$ ).

- *Théorie des perturbations-Hartree-Fock*

Dans le modèle de Hartree-Fock la fonction  $\Psi^{(0)}$  est un produit antisymétrique des spin-orbitales  $\phi_i$ .

$$\Psi^{(0)}(1,2,\dots,N) = \hat{A} \left| \phi_1^{(0)}(\vec{r}_1) \phi_2^{(0)}(\vec{r}_2) \dots \phi_N^{(0)}(\vec{r}_N) \right| \quad (76)$$

$E^{(0)}$  est la somme de leurs énergies correspondantes :

$$E^{(0)} = \varepsilon_1^{(0)} + \varepsilon_2^{(0)} + \dots + \varepsilon_N^{(0)} \quad (77)$$

Et  $H^{(0)}$  est la somme des opérateurs non perturbés de Fock mono électroniques :

$$\hat{H}^{(0)} = \hat{F}_1 + \hat{F}_2 + \dots + \hat{F}_N \quad (78)$$

L'opérateur polyélectronique de perturbation  $H^{(pert)}$  peut être écrit :

$$\hat{H}^{(pert)} = \hat{F}_1^{(pert)} + \hat{F}_2^{(pert)} + \dots + \hat{F}_N^{(pert)} \quad (79)$$

Pour un électron  $\mu$  :

$$\hat{F}_\mu^{(pert)} = \hat{F}_\mu - \hat{F}_\mu^{(0)} \quad (80)$$

On note que le suffixe  $\mu$  de  $F^{(pert)}$  est le nom de l'électron décrit par l'équation de Schrödinger non perturbée.

$$\hat{F}_\mu^{(0)} \phi_i^{(0)}(\vec{r}_\mu) = \varepsilon_i^{(0)} \phi_i^{(0)}(\vec{r}_\mu) \quad (81)$$

Cependant le principe de Pauli annonce que tous les électrons sont identiques (indiscernables), alors on peut, en général, supprimer le suffixe  $\mu$  dans les équations citées au dessous. Les solutions de l'équation précédente sont mutuellement orthogonales

$\langle \phi_i^{(0)} | \phi_j^{(0)} \rangle = \delta_{ij}$  et leurs énergies associées  $\varepsilon_i^{(0)}$  sont réelles pour  $F^{(0)}$ .

En substituant les équations 76-80 dans les équations de la théorie des perturbations nous obtenons les corrections du premier et second d'ordre [39]:

$$E^{(1)} = \sum_i \varepsilon_i^{(1)} \quad (82)$$

$$E^{(2)} = \sum_i \varepsilon_i^{(2)} \quad (83)$$

$$\text{où } E^{(1)} = n_i^{(0)} \langle \phi_i^{(0)} / F^{\wedge(\text{pert})} / \phi_i^{(0)} \rangle \quad (84)$$

$$E^{(2)} \leq -n_i^{(0)} \frac{\langle \phi_i^{(0)} / F^{\wedge(\text{pert})} / \tilde{\phi} \rangle^2}{\langle \tilde{\phi} / F^{\wedge(0)} - \epsilon_i^{(0)} / \tilde{\phi} \rangle} \quad (85)$$

Dans ces expressions  $n_i^{(0)}$  est le nombre d'électrons occupant l'orbitale spatiale  $\phi_i^{(0)}$  dans  $\Psi^{(0)}$  et  $\tilde{\phi}$  est une fonction d'essai variationnelle quelconque (orthogonale à  $\phi_i^{(0)}$ ) pour la correction d'orbitale de premier ordre  $\phi_i^{(1)}$ .

Nous pouvons simplifier l'équation 85 en remplaçant  $F^{\wedge(\text{pert})}$  dans le numérateur par  $F$ , parce que  $\langle \phi_i^{(0)} / F^{\wedge(0)} / \tilde{\phi} \rangle = 0$  pour n'importe quelle  $\tilde{\phi}$  orthogonale à  $\phi_i^{(0)}$ .

Nous pouvons aussi choisir  $\tilde{\phi}$  comme une fonction d'essai normalisée et définie :

$$\tilde{\epsilon} = \langle \tilde{\phi} / F^{\wedge(0)} / \tilde{\phi} \rangle \quad (86)$$

Donc l'équation 10 est exprimée par :

$$E^{(2)} \leq -n_i^{(0)} \frac{\langle \phi_i^{(0)} / F / \tilde{\phi} \rangle^2}{\tilde{\epsilon} - \epsilon_i^{(0)}} \quad (87)$$

La meilleure fonction d'essai perturbatrice  $\tilde{\phi}$  est dominée par la seule orbitale non perturbée  $\phi_{j^*}^{(0)}$  (avec  $n_{j^*}=0$ ) alors on peut écrire :

$$\tilde{\phi} = \phi_{j^*}^{(0)} \quad (88a)$$

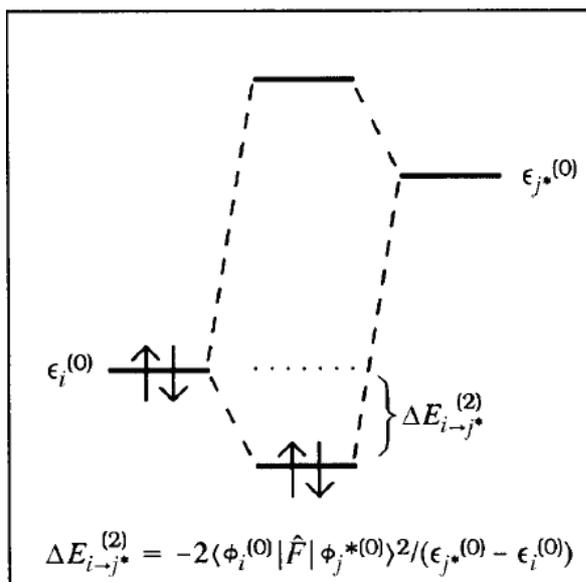
$$\tilde{\epsilon} = \epsilon_{j^*}^{(0)} \quad (88b)$$

Dans ce cas, l'expression de la correction à seconde ordre  $i \rightarrow j^*$  (eq 85) devienne :

$$E_{i \rightarrow j^*}^{(2)} = -n_i^{(0)} \frac{\langle \phi_i^{(0)} / F / \phi_{j^*}^{(0)} \rangle^2}{\epsilon_{j^*}^{(0)} - \epsilon_i^{(0)}} \quad (89)$$

- **La stabilisation donneur-accepteur:**

L'équation 14 présente un concept physique simple et fort comme il est représenté dans la figure suivante ( $n_i^{(0)}=2$ ) :



**Figure 1.** L'interaction stabilisante biélectronique entre une orbitale occupée (donneur) et une orbitale vacante (accepteur).

Cette figure montre que les niveaux non perturbés pour les orbitales doublement occupées  $\phi_i^{(0)}$  (donneur) et les orbitales vacantes (accepteur)  $\phi_j^{*(0)}$  interagissent pour donner une énergie plus faible  $E^{(2)} i \rightarrow j^*$  exprimée par l'équation 89. A partir de la **Figure 1** et l'équation on remarque que la grandeur  $E^{(2)}$  est nécessairement négative ( $E_j^* \geq E_i$ ) et l'interaction est, par conséquent, stabilisante. L'interaction donneur-accepteur de la **Figure 1** est une interaction stabilisante à 2 électrons.

### 1.6.2. Les orbitales naturelles NAO et NBO

L'analyse NBO a été proposée par Weinhold et al. [41] dont la procédure se déroule en trois étapes. La première étape consiste à diagonaliser la matrice densité d'ordre un en blocs mono-centriques atomiques sans diagonaliser la matrice dans son entier (orbitales atomiques naturelles NAO) [42,43]. La seconde étape du processus NBO (Natural Bond Orbitals) [44-46] consiste à orthogonaliser les orbitales NAO. La dernière étape est une recherche d'orbitales naturelles hybrides, processus qui diagonalise des blocs bicentriques d'orbitales naturelles orthogonalisées. Ce concept d'orbitales naturelles utilise la matrice densité pour distribuer les électrons dans des orbitales dans le but de calculer des charges et de caractériser des liaisons. La matrice densité  $\rho$  d'un système à plusieurs centres A, B, C..., peut s'écrire:

$$\rho = \begin{pmatrix} \rho^{AA} & \rho^{AB} & \rho^{AC} & \dots \\ \rho^{AB} & \rho^{BB} & \rho^{BC} & \dots \\ \rho^{AC} & \rho^{BC} & \rho^{CC} & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix} \quad (90)$$

Les orbitales naturelles atomiques NAO (Natural Atomic Orbitals) pour l'atome A dans le système moléculaire sont celles qui diagonalisent le bloc  $\rho^{AA}$ , les NAO pour l'atome B sont celles qui diagonalisent le bloc  $\rho^{BB}$  et ainsi de suite. Par définition, les orbitales naturelles  $\phi_i$  sont les fonctions propres de l'opérateur de densité  $\hat{\Gamma}$  pour une fonction à N électrons quelconque  $\Psi$ :

$$\hat{\Gamma} \phi_i = q_i \phi_i \quad (91a)$$

avec:

$$\hat{\Gamma} = 2 \sum_{i=1}^{occ} |\phi_i\rangle \langle \phi_i| \quad (91b)$$

Cette expression peut être utilisée dans l'évolution des éléments de la matrice  $\Gamma_{jk}$  de l'opérateur dans un ensemble de base d'orbitales atomiques  $\{\chi_j\}$ .

$$\Gamma_{jk} = \langle \chi_j | \hat{\Gamma} | \chi_k \rangle = 2 \sum_{i=1}^{occ} \langle \chi_j | \phi_i \rangle \langle \phi_i | \chi_k \rangle \quad (92)$$

En utilisant l'approximation CLOA-OM pour  $\phi_i$

$$\phi_i = 2 \sum_{j=1} c_{ij} \chi_j \quad (93)$$

Pour un ensemble de base orthonormé on obtient:

$$\langle \chi_j | \phi_i \rangle = c_{ij} \quad , \quad \langle \phi_i | \chi_k \rangle = c_{ik}^* \quad (94)$$

Les éléments de la matrice  $\Gamma_{jk}$  de l'opérateur de densité

$$\Gamma_{jk} = (D)_{jk} = 2 \sum_{i=1}^{occ} c_{ij} c_{ik}^* \quad (95)$$

Dans la méthode de Hückel simple l'élément diagonale de la matrice densité  $(D)_{jj}$  représente la charges  $\pi$  de l'atome  $j$  et l'élément non diagonale  $(D)_{jk}$  représente l'ordre de liaison entre les atomes  $j$  et  $k$ .

Les NAO  $\phi_i^{(A)}$  de l'atome  $A$  avec une occupation  $n_i^{(A)}$  se réduisent aux orbitales naturelles ordinaires des atomes isolés dans la limite de la dissociation, mais ils retiennent la propriété d'occupation maximale dans l'environnement moléculaire et donc elle continue à servir comme les meilleurs orbitales atomiques possibles.

$$\hat{\Gamma}^{NAO} = \begin{vmatrix} \hat{\Gamma}^A & \hat{\Gamma}^{AB} \\ \hat{\Gamma}^{AB} & \hat{\Gamma}^B \end{vmatrix} \rightarrow \hat{\Gamma}^{(A)} = \begin{vmatrix} CR^{(A)} & 0 & 0 & 0 \\ 0 & LP^{(A)} & & \\ 0 & 0 & Val^{(A)} & 0 \\ 0 & 0 & 0 & RY^{(A)} \end{vmatrix} \quad (96)$$

Dans la région de la liaison diatomique A-B, on peut de même obtenir les orbitales naturelles de liaison NBO  $\Omega_{AB}$  avec une occupation  $n_{AB} \approx 2$ . En fait les NAO de valence servent à construire les orbitales hybrides NHO (Natural Hybrid Orbitals) sur chaque atome dont le recouvrement définit les orbitales NBO (NBO type Lewis  $\Omega_{AB}$  et NBO type non-Lewis) :

$$\hat{\Gamma}^{(AB)} = \begin{vmatrix} Val^{(A)} & \text{couplage} \\ \text{couplage} & Val^{(B)} \end{vmatrix} \rightarrow \begin{aligned} \Omega_{AB} &= C_A h_A + C_B h_B \\ \Omega_{AB}^* &= C_A h_A - C_B h_B \end{aligned} \quad (97)$$

$h_A$  et  $h_B$  sont les Orbitales Naturelles Hybrides (NHO), combinaisons linéaires des orbitales atomiques naturelles (NAO).  $C_A$  et  $C_B$  sont des coefficients de polarisation tel que  $|C_A|^2 + |C_B|^2 = 1$ .

L'ensemble des NBO type Lewis contient les orbitales de cœur mono-centre (CR), les orbitales des paires électroniques de valence (LP) et les orbitales de liaison

bi-centriques (BD). Tandis que L'ensemble type non-Lewis contient les orbitales de couches de valence lointaine de Rydberg (RY\*) les orbitales de valence anti liantes (BD\*).

En résumé, l'analyse NBO est une série ou une succession de transformations d'un ensemble de base  $\{\chi_i\}$  en plusieurs ensembles localisés :

Input bases  $\rightarrow$  NAOs  $\rightarrow$  NHOs  $\rightarrow$  NBOs

Ces différentes étapes sont automatiquement réalisées par le programme NBO

- **L'analyse de population naturelle (NPA)**

L'ensemble orthonormé des NAO  $\{\theta_i^{(A)}\}$  est considéré comme la base pour l'analyse de population naturelle qui corrige beaucoup de défauts l'analyse de population de Mulliken [47]. La population naturelle  $q_i^{(A)}$  de l'orbitale  $\theta_i^{(A)}$  sur un atome A est simplement l'élément diagonal de la matrice de densité dans la base NAO.

$$q_i^{(A)} = \langle \theta_i^{(A)} | \hat{\Gamma}^{(A)} | \theta_i^{(A)} \rangle \quad (98)$$

Ces populations satisfaits, automatiquement, le principe de Pauli ( $0 \leq q_i^{(A)} \leq 2$ ) et donnent les populations atomiques  $q^{(A)}$  :

$$q^{(A)} = \sum_i q_i^{(A)}, \quad N_{electrons} = \sum_A^{atomes} q^{(A)} \quad (99)$$

L'analyse de la population naturelle est appliquée a une grande variété de composés chimiques et elle présente une bonne stabilité numérique et en très bon accord avec d'autres mesures théoriques et expérimentales des distributions de charge.

Finalement, l'analyse NBO a été développée afin de décrire la structure électronique des systèmes chimiques. Les orbitales naturelles de liaison NBO occupées décrivent différentes liaisons ou paires libres de la molécule et réfèrent ainsi à une

structure de Lewis (dans laquelle les électrons sont strictement localisés). Le système réel dévie de cette description « idéalisée » par des interactions donneur-accepteur entre NBOs occupées et NBOs vacantes. Cette analyse fournit une évaluation des énergies de ces effets de délocalisation et d'hyperconjugaison.

Les interactions donneurs-accepteurs ou interactions de transfert de charge représentent une composante de l'expression de l'énergie d'interaction entre les orbitales moléculaires. Cette dernière peut être décomposée en utilisant l'analyse de décomposition de l'énergie naturelle dans le programme NBO:

$$\Delta E = CT + ES + DEF \quad (100)$$

où

CT : énergie de transfert de charge

ES : énergie d'interaction électrostatique

DEF : énergie de déformation

Cette analyse de décomposition de l'énergie du programme NBO (version 3.1) est partiellement implémentée dans le programme GAUSSIAN. Seulement la composante CT est évaluée mais pas les composantes ES et DEF. Pour cette raison, il faut faire appel à des versions plus récentes (version 5.x) du programme NBO.

## I.7. Thermodynamique statistique

La mécanique statistique (ou thermodynamique statistique) a pour objet l'étude des propriétés macroscopiques (fonction d'état, chaleur spécifique...) à partir de la description d'un ensemble constitué de systèmes microscopiques [48]. Cette discipline permet de relier les grandeurs macroscopiques (énergie, entropie) aux grandeurs microscopiques calculées par la mécanique quantique et la spectroscopie [49].

L'énergie totale d'une molécule isolée est la somme des énergies de translation, de rotation, de vibration, électronique et de répulsions nucléaires :

$$E_{Tot} = E_{tr} + E_{rot} + E_{vib} + E_{elec} + E_{nuc} \quad (101)$$

De même, les contributions à l'enthalpie et à l'entropie peuvent s'écrire [50] :

$$H_{Tot} = H_{tr} + H_{rot} + H_{vib} + H_{elec} \quad (102)$$

$$S_{Tot} = S_{tr} + S_{rot} + S_{vib} + S_{elec} \quad (103)$$

### 1.7.1. La fonction de partition :

La fonction de partition d'une molécule peut être définie comme le produit de fonctions de partition relatives aux quatre types de mouvement (translation, rotation, vibration, et électronique) [51] :

$$Q = Q_{tr} \cdot Q_{rot} \cdot Q_{vib} \cdot Q_{elec} \quad (104)$$

Pour définir la fonction de partition, il est nécessaire de faire appel à la statistique de distributions donnée par l'équation de Boltzmann [52] :

$$N_i / N_0 = e^{-\left(\varepsilon_i - \varepsilon_0\right) / k_B T} \quad (105)$$

$N_i$  et  $N_0$  étant les nombres de molécules réparties sur les niveaux  $i$  et  $0$ , respectivement.

$\varepsilon_i$  = énergie du niveau  $i$ .

$\varepsilon_0 = 0$ , énergie du niveau  $0$ .

$k_B$  = constante de *Boltzmann*.

S'il existe plusieurs niveaux de même énergie, on doit écrire:

$$N_i / N_0 = \left(g_i / g_0\right) e^{-\varepsilon_i / k_B T} \quad (106)$$

Où  $g_i$ ,  $g_0$  sont les facteurs statistiques (degrés de dégénérescence) pour les niveaux  $i$  et  $0$ , respectivement. On obtient ainsi ;

$$N_i = \left(N_0 / g_0\right) g_i e^{-\varepsilon_i / k_B T} \quad (107)$$

Soit  $N$  le nombre total de molécules ( $N_0/g_0$  est une constante):

$$N = \sum N_i = \frac{N_0}{g_0} \sum g_i e^{-\varepsilon_i/k_B T} \quad (108)$$

Finalement, la fonction de partition peut être écrite comme suit :

$$q = \sum g_i e^{-\varepsilon_i/k_B T} \quad (109)$$

La somme  $q$  suppose que les énergies des molécules sont indépendantes les unes des autres ; ce qui veut dire que les molécules considérées constituent un gaz parfait.

### I.7.2. Expression de l'énergie :

L'énergie totale pour  $N$  molécules d'un gaz parfait vaudra :

$$E - E_0 = \Delta E = \sum N_i \varepsilon_i \quad \text{à n'importe quelle température.}$$

En pratique, les  $\varepsilon_i$  se réfèrent aux énergies de translation ( $E_{trans}$ ), de rotation ( $E_{rot}$ ), et de vibration ( $E_{vib}$ ). Dans ce cas  $E_0$  correspond à la somme des énergies électronique ( $E_{elec}$ ) et nucléaire ( $E_{nucl}$ ), à  $T=0 K$ . En utilisant la relation 108, on peut écrire :

$$\Delta E = \left( \frac{N_0}{g_0} \right) \sum g_i \varepsilon_i e^{-\varepsilon_i/k_B T} \quad (110)$$

En dérivant 109 par rapport à  $T$  et à volume constant, on tire :

$$k_B T^2 \left( \frac{\partial q}{\partial T} \right)_V = \sum g_i \varepsilon_i e^{-\varepsilon_i/k_B T} \quad (111)$$

En combinant 8, 110 et 111 on obtient :

$$\Delta E = E - E_0 = \frac{N k_B T^2}{q} \left( \frac{\partial q}{\partial T} \right)_V = N k_B T^2 \left( \frac{\partial \ln q}{\partial T} \right)_V \quad (112)$$

Si  $N=N_A$  (une mole de gaz parfait), on aura  $Nk_B=R$ , d'où :

$$E - E_0 = RT^2 \left( \frac{\partial \ln q}{\partial T} \right)_V \quad (113)$$

**I.7.3. Expression de la capacité calorifique :**

En dérivant 113 par rapport à  $T$  et à volume constant, la capacité calorifique à volume constant se ramène à :

$$C_v = \left( \frac{\partial E}{\partial T} \right)_v = \frac{\partial}{\partial T} \left[ RT^2 \left( \frac{\partial \ln q}{\partial T} \right) \right]_v \quad (114)$$

**I.7. 4. Expression de l'entropie :**

L'entropie varie avec la température selon l'expression 115 :

$$S - S_0 = \Delta S = \int_0^t \frac{C_v}{T} dT \quad (115)$$

L'équation précédente peut être développée comme suit :

$$\begin{aligned} S - S_0 = \Delta S &= \int_0^t \frac{C_v}{T} dT \\ &= \int_0^t \frac{1}{T} \frac{\partial}{\partial T} \left[ RT^2 \left( \frac{\partial \ln q}{\partial T} \right) \right]_v dT \\ &= \int_0^t \frac{1}{T} \left[ 2RT \left( \frac{\partial \ln q}{\partial T} \right) + RT^2 \frac{\partial}{\partial T} \left( \frac{\partial \ln q}{\partial T} \right) \right]_v dT \\ &= \int_0^t 2R \left( \frac{\partial \ln q}{\partial T} \right)_v dT + \int_0^t RT \frac{\partial}{\partial T} \left( \frac{\partial \ln q}{\partial T} \right)_v dT \\ &= 2R [\ln q]_0^t + RT \left( \frac{\partial \ln q}{\partial T} \right)_v - \int_0^t R \left( \frac{\partial \ln q}{\partial T} \right)_v dT \\ &= 2R \ln q_t - 2R \ln q_0 + RT \left( \frac{\partial \ln q}{\partial T} \right)_v - R [\ln q]_0^t \\ &= 2R \ln q_t - 2R \ln q_0 + RT \left( \frac{\partial \ln q}{\partial T} \right)_v - R \ln q_t + R \ln q_0 \\ &= RT \left( \frac{\partial \ln q}{\partial T} \right)_v + R \ln q_t - R \ln q_0. \end{aligned}$$

Finalement, on obtient :

$$S - S_0 = RT \left( \frac{\partial \ln q}{\partial T} \right)_V + R \ln q - R \ln q_0 \quad (116)$$

A  $T = 0$  K, les  $N$  molécules sont au niveau  $0$ , et comme les  $N$  molécules sont indiscernables, on aura :

$$S_0 = k_B \ln \frac{g_0^N}{N!} = N k_B \ln g_0 - k_B \ln N! = R \ln g_0 - k_B \ln N! \quad (117)$$

et  $q_0 = g_0 e^{-0/k_B T} = g_0$

Ce qui nous donne donc :

$$S_0 = R \ln q_0 - k_B \ln N! \quad (118)$$

En remplaçant  $S_0$  dans l'équation 116, on obtient finalement :

$$S = RT \left( \frac{\partial \ln q}{\partial T} \right)_V + R \ln q - k_B \ln N! \quad (119)$$

En connaissant les fonctions de partition  $q$  des molécules à plusieurs températures on peut donc évaluer l'entropie. L'énergie peut être estimée également à partir des relations 112 et 113.  $C_v$  est déterminable également grâce à l'équation 114. On note que la mécanique quantique donne les fonctions de partition pour les molécules isolées.

Les autres fonctions d'état thermodynamiques  $H$ ,  $G$  et  $F$  et la pression  $P$  s'expriment en fonction de la fonction de partition comme suit :

$$F = -kT \ln q \quad (120)$$

$$P = - \left( \frac{\partial A}{\partial V} \right)_T = kT \left( \frac{\partial \ln q}{\partial V} \right)_T \quad (121)$$

$$H = U + PV = kT^2 \left( \frac{\partial \ln q}{\partial T} \right)_V + kTV \left( \frac{\partial \ln q}{\partial V} \right)_T \quad (122)$$

$$G = H - TS = kTV \left( \frac{\partial \ln q}{\partial V} \right)_T - kT \ln q \quad (123)$$

### I.7.5. Contribution des différents mouvements à la fonction de partition :

#### a. Contribution du mouvement de translation :

Pour une molécule non linéaire à  $N$  atomes, il y a  $3N$  degrés de liberté au total.

-3 degrés de liberté pour la translation.

-3 degrés de liberté pour la rotation.

-3N-6 degrés de liberté pour la vibration.

Dans le cadre du modèle du puits de potentiel infini PPI, l'énergie d'une particule de masse  $M$  placée dans un PPI de longueur  $l$  :

$$\mathcal{E}_n = \frac{n^2 h^2}{8\pi^2 m} \quad n=1, 2, 3, \dots \quad (124)$$

Bien que les niveaux d'énergie soient quantifiés, la différence d'énergie entre les niveaux est si petite que la distribution puisse être traitée comme continue.

Par conséquent, la sommation discontinue dans la fonction de partition peut être remplacée par une intégrale.

$$q_{trans} = \sum_0^{\infty} e^{-\mathcal{E}_n/kT} \approx \int_0^{\infty} e^{-\mathcal{E}_n/kT} dn \quad (125)$$

Par intégration, on obtient :

$$q_t = \left( \frac{2\pi m k_B T}{h^2} \right)^{3/2} V. \quad (126)$$

Où :  $m$  = masse de la molécule,  $h$  = constante de *Planck*.

La dérivée partielle de  $\ln q_t$  par rapport à  $T$  se ramène à :

$$\left( \frac{\partial \ln q_t}{\partial T} \right)_V = \frac{3}{2T} \quad (127)$$

Cette équation servira pour le calcul de l'énergie interne  $E_t$  et le troisième terme dans l'équation 101. Dans le second terme de l'équation 126, le volume  $V$  est inconnu.

Cependant, l'expression classique de l'équation d'état d'un gaz parfait :  $PV = nRT$  ou

$$V = \frac{k_B T}{P}.$$

Par conséquent :

$$q_t = \left( \frac{2\pi m k_B T}{h^2} \right)^{3/2} \frac{k_B T}{P}. \quad (128)$$

Et c'est la formule utilisée pour le calcul de  $q_t$  dans le programme *Gaussian*.

La contribution de la translation à l'enthalpie s'écrit comme suit :

$$H_t = \frac{5}{2} RT. \quad (129)$$

La fonction de partition de translation est utilisée pour le calcul de l'entropie de translation (qui inclut le facteur  $e$  qui vient de « *l'approximation de Stirling* »).

$$\begin{aligned} S_t &= R \left( \ln(q_t e) + T \left( \frac{3}{2T} \right) \right) \\ &= R (\ln q_t + 1 + 3/2) \\ &= R \ln q_t + \frac{5}{2} R. \end{aligned} \quad (130)$$

La contribution de la translation à l'énergie interne peut être déduite comme suit :

$$\begin{aligned} E_t &= N_A k_B T^2 \left( \frac{\partial \ln q}{\partial T} \right)_V \\ &= RT^2 \left( \frac{3}{2T} \right) \\ &= \frac{3}{2} RT \end{aligned} \quad (131)$$

De même, la capacité calorifique à volume constant est donnée par :

$$\begin{aligned} C_{v,t} &= \frac{\partial E_t}{\partial T} \\ &= \frac{3}{2} R. \end{aligned} \quad (132)$$

### b. Contribution du mouvement électronique :

La fonction de partition électronique implique une somme sur les états quantiques électroniques. Ce sont les solutions de l'équation électronique de Schrödinger, c-à-d l'état le plus bas (fondamental) et tous les états excités possibles. Pour la plupart des molécules, la différence d'énergie entre l'état fondamental et les états excités est considérable par rapport à  $kT$ , ce qui signifie que seulement le premier terme (l'énergie de l'état fondamental) dans la sommation des fonctions de partition est important.

$$q_{elect} = \sum_{i=0}^{\infty} g_i e^{-\varepsilon_i/kT} \approx g_0 e^{-\varepsilon_0/kT} \quad (133)$$

Définissant le point zéro pour l'énergie comme l'énergie électronique du réactif. Les fonctions de partition électroniques pour les réactifs et les états de transition sont données comme suit :

$$\begin{aligned} q_{elect}^{reactant} &= g_0 \\ q_{elect}^{TS} &= g_0 e^{-\Delta E^\ddagger/kT} \end{aligned} \quad (134)$$

Le terme  $\Delta E^\ddagger$  est la différence de l'énergie électronique entre le réactif et le TS, et  $g_0$  est la dégénérescence électronique de la fonction d'onde (état fondamental). La dégénérescence peut être dans la partie du spin ( $g_0 = 1$  pour un singulet, 2 pour un doublet, 3 pour un triplet, etc.) ou dans la partie spatiale ( $g_0 = 1$  pour les fonctions d'onde appartenant à une représentation A, B ou  $\Sigma$  du groupe ponctuel, 2 pour une représentation E,  $\Delta$  ou  $\Phi$ , 3 pour une représentation T, etc.). La grande majorité de molécules stables ont des fonctions d'onde non dégénérées de l'état fondamental, et par conséquent  $g_0 = 1$ .

L'entropie due au mouvement électronique pour un réactif est :

$$\begin{aligned}
 S_e &= R \left( \ln q_e + T \left( \frac{\partial \ln q_e}{\partial T} \right)_V \right) \\
 &= R(\ln q_e + 0).
 \end{aligned}
 \tag{135}$$

Puisqu'il n'y a pas de termes dépendant de la température dans la fonction de partition, l'énergie interne et la capacité calorifique dues au mouvement électronique sont nulles.

$$E_e = 0 \quad \text{et} \quad C_{v,e} = 0. \tag{136}$$

### c. Contribution du mouvement de rotation :

L'énergie de rotation d'une molécule, de moment d'inertie  $I$ , dans le cadre du modèle du rotateur rigide, est donnée par :

$$\varepsilon_j = J(J+1) \frac{h^2}{8\pi^2 I} \quad j=0,1,2,\dots \tag{137}$$

Les écarts entre les niveaux rotationnels sont négligeables devant  $k_B T$ . Par conséquent, la sommation peut être remplacée par une intégration :

$$q_{rot} = \sum_{J=0}^{\infty} e^{-\varepsilon_J / k_B T} \approx \int_0^{\infty} e^{-\varepsilon_i / k_B T} dJ \tag{138}$$

Par intégration, on obtient :

$$q_{rot} = \frac{8\pi^2 I k T}{h^2 \sigma} \tag{139}$$

L'indice de symétrie  $\sigma$  est 2 pour un système homonucléaire et 1 pour une molécule diatomique hétéronucléaire. Pour une molécule polyatomique, le moment d'inertie est donné par une matrice  $3 \times 3$ .

$$I = \begin{pmatrix} \sum_i m_i (y_i^2 + z_i^2) & -\sum_i m_i x_i y_i & -\sum_i m_i x_i z_i \\ -\sum_i m_i x_i y_i & \sum_i m_i (x_i^2 + z_i^2) & -\sum_i m_i y_i z_i \\ -\sum_i m_i x_i z_i & -\sum_i m_i y_i z_i & \sum_i m_i (x_i^2 + y_i^2) \end{pmatrix} \quad (140)$$

Ici les coordonnées sont encore relativement au centre de la masse. En choisissant une transformation appropriée, cette matrice peut être diagonalisée ([3] section 16. 2), avec les valeurs propres étant les moments d'inertie et les vecteurs propres appelés les axes principaux d'inertie. Pour une molécule polyatomique quelconque, les niveaux d'énergie de rotation ne peuvent pas être écrits sous une forme simple. Une bonne approximation, cependant, peut être obtenue à partir de la mécanique classique, ayant pour résultat la fonction de partition suivante :

$$q_{rot} = \frac{\sqrt{\pi}}{\sigma} \left( \frac{8\pi^2 kT}{h^2} \right)^{3/2} \sqrt{I_1 I_2 I_3} \quad (141)$$

Ici les  $I_i$  ( $i=1, 2, 3$ ) sont les trois moments d'inertie correspondants aux trois axes. L'indice de symétrie  $\sigma$  est l'ordre du sous-groupe de rotation dans le groupe ponctuel moléculaire, c-à-d le nombre des opérations de symétrie appropriées); pour H<sub>2</sub>O il est 2, pour NH<sub>3</sub> c'est 3, pour le benzène c'est 12, etc. La fonction de partition de rotation exige seulement des informations sur les masses atomiques et les positions (eq.141)), c-à-d la géométrie moléculaire.

La discussion de la rotation moléculaire peut être traitée en 3 cas : atomes individuels, molécules polyatomiques linéaires et les molécules polyatomiques non linéaires.

- Pour un atome individuel,  $q_r=1$ . Sachant que  $q_r$  ne dépend pas de la température, les contributions de la rotation à l'énergie interne, à la capacité calorifique et à l'entropie sont toutes nulles.

- Pour une molécule linéaire, la fonction de partition de rotation est :

$$q_r = \frac{1}{\sigma_r} \left( \frac{T}{\theta_r} \right) \quad (142)$$

Où :  $\theta_r = h^2/8\pi^2 I k_B$ ,  $I$  est le moment d'inertie.

La contribution à partir de l'enthalpie est :

$$H_{rot} = RT \quad (143)$$

La contribution de la rotation à l'entropie est :

$$\begin{aligned} S_r &= R \left( \ln q_r + T \left( \frac{\partial \ln q_r}{\partial T} \right)_V \right) \\ &= R (\ln q_r + 1). \end{aligned} \quad (144)$$

La contribution de la rotation à l'énergie interne est :

$$\begin{aligned} E_{rot} &= RT^2 \left( \frac{\partial \ln q_r}{\partial T} \right)_V \\ &= RT^2 \left( \frac{1}{T} \right) \\ &= RT \end{aligned} \quad (145)$$

et la contribution à la capacité calorifique est :

$$\begin{aligned} C_{rot} &= \left( \frac{\partial E_r}{\partial T} \right)_V \\ &= R \end{aligned} \quad (146)$$

-Dans le cas général des molécules polyatomiques non linéaires, la fonction de partition est :

$$q_r = \frac{\pi^{1/2}}{\sigma_r} \left( \frac{T^{3/2}}{(\theta_{r,x}\theta_{r,y}\theta_{r,z})^{1/2}} \right) \quad (147)$$

Etant donné que  $\left(\frac{\partial \ln q}{\partial T}\right) = \frac{3}{2T}$ , l'entropie pour cette fonction de partition peut s'écrire :

$$\begin{aligned} S_{rot} &= R \left( \ln q_r + T \left( \frac{\partial \ln q_r}{\partial T} \right)_V \right) \\ &= R \left( \ln q_r + \frac{3}{2} \right) \end{aligned} \quad (148)$$

et la contribution de l'enthalpie est :

$$H_{rot} = \frac{3}{2} RT \quad (149)$$

Finalement, la contribution à l'énergie interne est :

$$\begin{aligned} E_{rot} &= RT^2 \left( \frac{\partial \ln q_r}{\partial T} \right)_V \\ &= RT^2 \left( \frac{3}{2T} \right) \\ &= \frac{3}{2} RT \end{aligned} \quad (150)$$

et la contribution à la capacité calorifique est :

$$\begin{aligned} C_r &= \left( \frac{\partial E_r}{\partial T} \right)_V \\ &= \frac{3}{2} R. \end{aligned} \quad (151)$$

**d. Contribution du mouvement de vibration :**

Les vibrations moléculaires peuvent être décrites sur la base du modèle d'un oscillateur harmonique. Le développement de l'énergie, en série de Taylor, en fonction de coordonnées nucléaires autour de la géométrie d'équilibre s'écrit, pour une molécule diatomique, comme suit :

$$E(R) = E(R_0) + \frac{dE}{dR}(R - R_0) + \frac{1}{2} \frac{d^2E}{dR^2}(R - R_0)^2 + \frac{1}{6} \frac{d^3E}{dR^3}(R - R_0)^3 + \dots \quad (152)$$

R : est la distance internucléaire R

Le premier terme peut être pris comme zéro, puisque c'est juste le point zéro pour l'énergie. Le deuxième terme (le gradient) disparaît puisque le développement est autour de la géométrie d'équilibre. En gardant juste le terme le plus bas non nul résultant de l'approximation harmonique, on obtient :

$$E(\Delta R) \cong \frac{1}{2} \frac{d^2E}{dR^2} \Delta R^2 = \frac{1}{2} k \Delta R^2 \quad (153)$$

Où k est la constante de force.

Les niveaux d'énergie obtenus à partir de la résolution de l'équation de Schrödinger pour un oscillateur harmonique unidimensionnel (système diatomique) sont donnés par :

$$\begin{aligned} \varepsilon_n &= \left( n + \frac{1}{2} \right) h\nu & n = 0, 1, 2, \dots \\ \nu &= \frac{1}{2\pi} \sqrt{\frac{k}{\mu}} \\ \mu &= \frac{m_1 m_2}{m_1 + m_2} \end{aligned} \quad (154)$$

$\mu$  étant la masse réduite.

Contrairement aux niveaux d'énergie de translation et de rotation, l'espacement entre les niveaux d'énergie de vibration est comparable à  $kT$  pour les températures autour de 300 K, et la sommation pour le  $q_{vib}$  ne peut pas être remplacée par une intégration. En raison de l'espacement régulier, cependant, l'addition infinie peut être écrite dans une forme fermée (close) :

$$\begin{aligned}
 q_{vib} &= \sum_{n=0}^{\infty} e^{-\epsilon_n/kT} = e^{-h\nu/2kT} + e^{-3h\nu/2kT} + e^{-5h\nu/2kT} + \dots \\
 q_{vib} &= e^{-h\nu/2kT} (1 + e^{-h\nu/kT} + e^{-2h\nu/kT} + \dots) \\
 q_{vib} &= \frac{e^{-h\nu/2kT}}{1 - e^{-h\nu/kT}}
 \end{aligned}
 \tag{155}$$

Etant donné que l'énergie des  $3N-6$  vibrations peut être écrite comme une somme, la fonction de partition peut être écrite comme un produit sur toutes les  $3N-6$  fonctions de partition de vibration.

$$\begin{aligned}
 E_{vib} &= \sum_{i=1}^{3N_{atome}-6(5)} \left( n_i + \frac{1}{2} \right) h\nu_i \\
 q_{vib} &= \prod_{i=1}^{3N_{atome}-6(5)} \frac{e^{-h\nu_i/2kT}}{1 - e^{-h\nu_i/kT}}
 \end{aligned}
 \tag{156}$$

Les fréquences de vibration sont nécessaires pour le calcul de  $q_{vib}$ , et peuvent être obtenues à partir de la matrice de la constante de force et des masses atomiques.

Les contributions de la fonction de partition de vibration à l'entropie, à l'énergie interne et à la capacité calorifique à volume constant sont composées d'une somme (ou produit) des contributions de chaque mode de vibration,  $K$ . Seuls les modes réels sont considérés (les modes avec des fréquences imaginaires sont ignorés). Chacun des  $3n_{atomes}-6$  (ou  $3n_{atomes}-5$  pour les molécules linéaires) modes a une température de vibration caractéristique,  $\theta_{v,k} = h\nu_k / k_B$ .

Il y a deux méthodes pour calculer la fonction de partition, dépendant du choix du zéro de l'énergie :

- Soit le bas du puits de l'énergie potentielle internucléaire.
- Soit le premier niveau vibrationnel.

Si on choisit le bas du puits de potentiel (*BOT*) comme point de référence zéro, la contribution de la fonction de partition à partir d'un mode vibrationnel donné est :

$$q_{v,k} = \frac{e^{-\theta_{v,k}/2T}}{1 - e^{-\theta_{v,k}/T}} \quad (157)$$

et la fonction de partition globale est :

$$q_v = \prod_k \frac{e^{-\theta_{v,k}/2T}}{1 - e^{-\theta_{v,k}/T}} \quad (158)$$

Si on choisit le premier niveau de l'énergie vibrationnelle comme l'énergie zéro ( $V=0$ ), la fonction de partition pour chaque niveau vibrationnel est :

$$q_{v,k} = \frac{1}{1 - e^{-\theta_{v,k}/T}} \quad (159)$$

et la fonction de partition de vibration globale est :

$$q_v = \prod_k \frac{1}{1 - e^{-\theta_{v,k}/T}} \quad (160)$$

Le programme *Gaussian* utilise le bas du puits comme énergie zéro (*BOT*) pour déterminer les autres quantités thermodynamiques, mais il affiche également la fonction de partition  $V=0$ .

On note que la seule différence entre les deux méthodes est le facteur supplémentaire  $\theta_{v,k} / 2$ , (qui est l'énergie vibrationnelle au point zéro) dans l'équation de l'énergie interne  $E_v$ .

Dans les expressions de l'entropie et de la capacité calorifique, ce facteur disparaît quand on dérive par rapport à la température (T).

La contribution de la vibration à l'enthalpie est :

$$H_{vib} = R \sum_{i=1}^{3N-6(5)} \left( \frac{h\nu_i}{2k} + \frac{h\nu_i}{k} \cdot \frac{1}{e^{-h\nu_i/kT} - 1} \right) \quad (161)$$

De même, la contribution de la vibration à l'entropie est :

$$\begin{aligned} S_{vib} &= R \left( \ln(q_v) + T \left( \frac{\partial \ln q}{\partial T} \right)_V \right) \\ &= R \left( \ln \left( \prod_k \frac{e^{-\theta_{v,k}/2T}}{1 - e^{-\theta_{v,k}/T}} \right) + T \left( \frac{\partial}{\partial T} \ln \left( \prod_k \frac{e^{-\theta_{v,k}/2T}}{1 - e^{-\theta_{v,k}/T}} \right) \right) \right) \\ &= R \left( - \sum_k \left( \frac{\theta_{v,k}}{2T} + \ln(1 - e^{-\theta_{v,k}/T}) \right) + T \left( \sum_k \frac{\theta_{v,k}}{2T^2} + \sum_k \frac{(\theta_{v,k}/T^2) e^{-\theta_{v,k}/T}}{1 - e^{-\theta_{v,k}/T}} \right) \right) \\ &= R \left( - \sum_k \ln(1 - e^{-\theta_{v,k}/T}) + \left( \sum_k \frac{(\theta_{v,k}/T) e^{-\theta_{v,k}/T}}{1 - e^{-\theta_{v,k}/T}} \right) * \frac{e^{\theta_{v,k}/T}}{e^{\theta_{v,k}/T}} \right) \\ &= R \sum_k \left( \frac{\theta_{v,k}/T}{e^{\theta_{v,k}/T} - 1} - \ln(1 - e^{-\theta_{v,k}/T}) \right) \end{aligned} \quad (162)$$

La contribution de la vibration moléculaire à l'énergie interne est :

$$E_{vib} = R \sum_k \theta_{v,k} \left( \frac{1}{2} + \frac{1}{e^{\theta_{v,k}/T} - 1} \right) \quad (163)$$

Finalement, la contribution de la capacité calorifique à volume constant est :

$$C_{vib} = R \sum_k e^{\theta_{v,k}/T} \left( \frac{\theta_{v,k}/T}{e^{-\theta_{v,k}/T} - 1} \right)^2 \quad (164)$$

**Bibliographie**

- [1] E. Schrödinger, Ann. Phys. Leipzig. 76 (1926) 361.
- [2] M. Born, J. R. Oppenheimer, Ann. Phys. 84 (1927) 457.
- [3] V. Minkine, B. Simkine, R. Minaev, "Théorie de la structure moléculaire " Edition Mir, Moscou, 1982.
- [4] V. Fock, Z. Physik. 61 (1930) 126.
- [5] (a) J.C. Slater, Phys. Rev. 48 (1930), 48, 35.  
(b) J.C. Slater, Adv. Quantum Chem. 6 (1972) 1.
- [6] C. C. Roothaan, Rev. Mod. Phys. 23 (1951) 69.
- [7] I. Shavitt, "Methods of Electronic Structure Theory" H. F. Shaefer, Ed., Plenum Press, New York, p189 1977.
- [8] A. Jugl, "Chimie Quantique Structurale et Eléments de Spectroscopie Théorique ", 1978.
- [9] C. Moller, M. S. Plesset, Phys. Rev. 46 (1934) 618.
- [10] J. L. Rivail, "Eléments de chimie quantique" InterEditions/CNRS Editions, Paris, 1994.
- [11] (a) R. G. Parr, W. Yang "Density Functional Theory" Oxford University Press, 1989.  
(b) L. J. Bartolotti, K. Flurchick, Rev. Comput. Chem. 7 (1996) 187.  
(c) St-Amant. Rev. Comput. Chem. 7 (1996) 217.  
(d) T. Ziegler. Chem. Rev. 91 (1991) 651.  
(e) E. J. Baerends et O. V. Gritsenko. J. Phys. Chem. 101 (1997) 5383.
- [12] P. Hohenberg, W. Kohn, Phys. Rev. 136 (1964) B846.
- [13] W. Kohn, L. J. Sham, Phys. Rev. 140 (1965) A1133.
- [14] S. J. Vosko, L. Wilk, M. Nusair, Can. J. Phys. 58 (1980) 1200.
- [15] F. Jensen "Introduction to Computational Chemistry" John Wiley & Sons, 1999.
- [16] A. D. Becke, Phys. Rev. B 38 (1988) 3098.
- [17] (a) Perdew, J.P. "Electronic Structure of Solids" P. Ziesche and H. Eschrig, Akademie Verlag, Berlin, 1991.(b) Perdew, J.P. Phys. Rev. B 33 (1986) 8822

- [18] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* 37 (1988) 785.
- [19] A. D. Becke, *J. Chem. Phys.* 98 (1993) 5648.
- [20] J. C. Slater, *J. Chem. Phys.* 36 (1930) 57.
- [21] S. F. Boys, *Proc. Roy. Soc. A*200 (1950) 542.
- [22] E. Clementi, Ed., "Modern Techniques in Computational Chemistry", MOTTECC™ 89, (ESCOM, Leiden), 1989.
- [23] S. Huzinaga, *J. Chem. Phys.* 42 (1965) 1293.
- [24] T. H. Dunning, *J. Chem. Phys.* 55 (1971) 716.
- [25] S. M. Mekelleche, "Thèse de doctorat d'état ", Université de Tlemcen, 2000.
- [26] C. J. Cramer et D. G. Truhlar, *Chem. Rev.* 99 (1999) 2161.
- [27] J.G. Kirkwood, *J. Chem. Phys.* 2 (1934) 351.
- [28] K. Baldrige, A Klamt, *J. Chem. Phys.* 106 (1997) 6622.
- [29] L. Onsager, *J. Am. Chem. Soc.* 58 (1936) 1486.
- [30] M. Cossi, *J. Comp. Chem.* 19 (1998) 833.
- [31] S. Miertuš, E. Scrocco, J. Tomasi, *J. Chem. Phys.* 55 (1981) 117.
- [32] J. Tomasi, M. Persico, *Chem. Rev.* 94 (1994) 2027.
- [33] R. Cammi, J. Tomasi, *J. Comp. Chem.* 16 (1995) 1449.
- [34] A. D. Laurant, Thèse de doctorat, Université de Henri Poincaré, Nancy 2010.
- [35] (a) H. H. J. Uhlig, *Phys. Chem.* 41 (1937) 1215.  
(b) I. Tunon, E. Silla, J. L. Pascual-Ahuir, *Chem. Phys. Lett.* 203 (1993) 289.
- [36] F. London, *Z. Phys.* 60 (1930) 245.
- [37] C. Amovilli et B. Mennucci, *J. Phys. Chem.* 101 (1997) 1051.
- [38] (a) G. N. Lewis, *J. Am. Chem. Soc.* 38 (1916) 762.  
(b) G. N. Lewis, *Valence and the Structure of Atoms and Molecules*, New York, The Chemical Catalog Co., 1923. As observed by Pauling (in note 51, p. 5), this remarkable work "forms the basis of the modern electronic theory of valence."
- [39] (a) I. N. Levine, "Quantum Chemistry", Englewood Cliffs, NJ, Prentice Hall, 2000.  
(b) J. R. Barrante, "Applied Mathematics for Physical Chemistry", Upper Saddle River, NJ, Prentice Hall, 1998.

- [40] E. A. Hylleraas, *Z. Phys.* 65 (1930), 209, note 6, 279.
- [41] (a) F. Weinhold, "Natural bond orbital methods", in P. v. R. Schleyer, N. L. Allinger, T. Clark et al.(eds.), *Encyclopedia of Computational Chemistry*, Vol. 3, Chichester, John Wiley & Sons, 1998, pp. 1792–1811.  
(b) F. Weinhold, C. R. Landis, "Valency and Bonding, A Natural Bond Orbital Donor–Acceptor Perspective", Cambridge university Press, New York, 2005
- [42] P.-O. Löwdin, *Phys. Rev.* 97 (1955), 1474.
- [43] A. E. Reed, R. B. Weinstock, and F. Weinhold, *J. Chem. Phys.* 83 (1985), 735.
- [44] J. P. Foster and F. Weinhold, *J. Am. Chem. Soc.* 102 (1980), 7211.
- [45] The program NBO 5.0 was developed by E. D. Glendening, J. K. Badenhoop, A. E. Reed, J. E. Carpenter, J. A. Bohmann, C. M. Morales, and F. Weinhold, Theoretical Chemistry Institute, University of Wisconsin, Madison 2001; F. Weinhold, NBO 5.0 Program Manual: Natural Bond Orbital Analysis Programs, Madison, WI, Theoretical Chemistry Institute, 2001.
- [46] The NBO Website URL is <http://www.chem.wisc.edu/~nbo5>.
- [47] (a) R. S. Mulliken, *Phys. Rev.* 41 (1932) 49.  
(b) M. Wolfsberg and L. Helmholz, *J. Chem. Phys.* 20 (1952) 837.  
(c) R. Hoffmann, *J. Chem. Phys.* 39 (1963) 1397.
- [48] C. Lhuillier, *Cours de Physique Statistique*, Chap. 1, p 7, 2004.
- [49] P. Vogel, "Chimie organique : méthodes et modèles", De Book Université, 1997.
- [50] (a) F. Jensen, "Introduction to Computational Chemistry", John Wiley and Sons, England, 2007, p 302.  
(b) F. Jensen, "Introduction to Computational Chemistry", John Wiley and Sons, England, 2007, p 428.
- [51] C. Lhuillier, "Cours de Physique Statistique", Chap. 4, p 37, 2004
- [52] Joseph W. Ochterski, Ph. D, "Thermochemistry in Gaussian". [help@gaussian.com](mailto:help@gaussian.com), 2000.

# **CHAPITRE II**

## ***MODELISATION QSPR ET ANALYSE DE DONNEES***

## Introduction

Les méthodes QSPR (*Quantitative Structure-Property Relationships*) sont de plus en plus utilisées en chimie du fait de la croissance spectaculaire des moyens de calculs. Les premiers travaux utilisant ces méthodes sont dus aux équations de Hammett [1-5] dans lesquelles les constantes  $\sigma$  caractérisent de manière quantitative les vitesses de réactions pour les composés organiques.

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

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

Dans ce que suit, on présente le principe de ces méthodes QSPR ainsi que les différents outils employés dans ces techniques :

- Choix des bases de données expérimentales
- Choix des descripteurs
- Choix des méthodes de régressions et d'analyse de données.

### II.1. Principe des méthodes QSPR

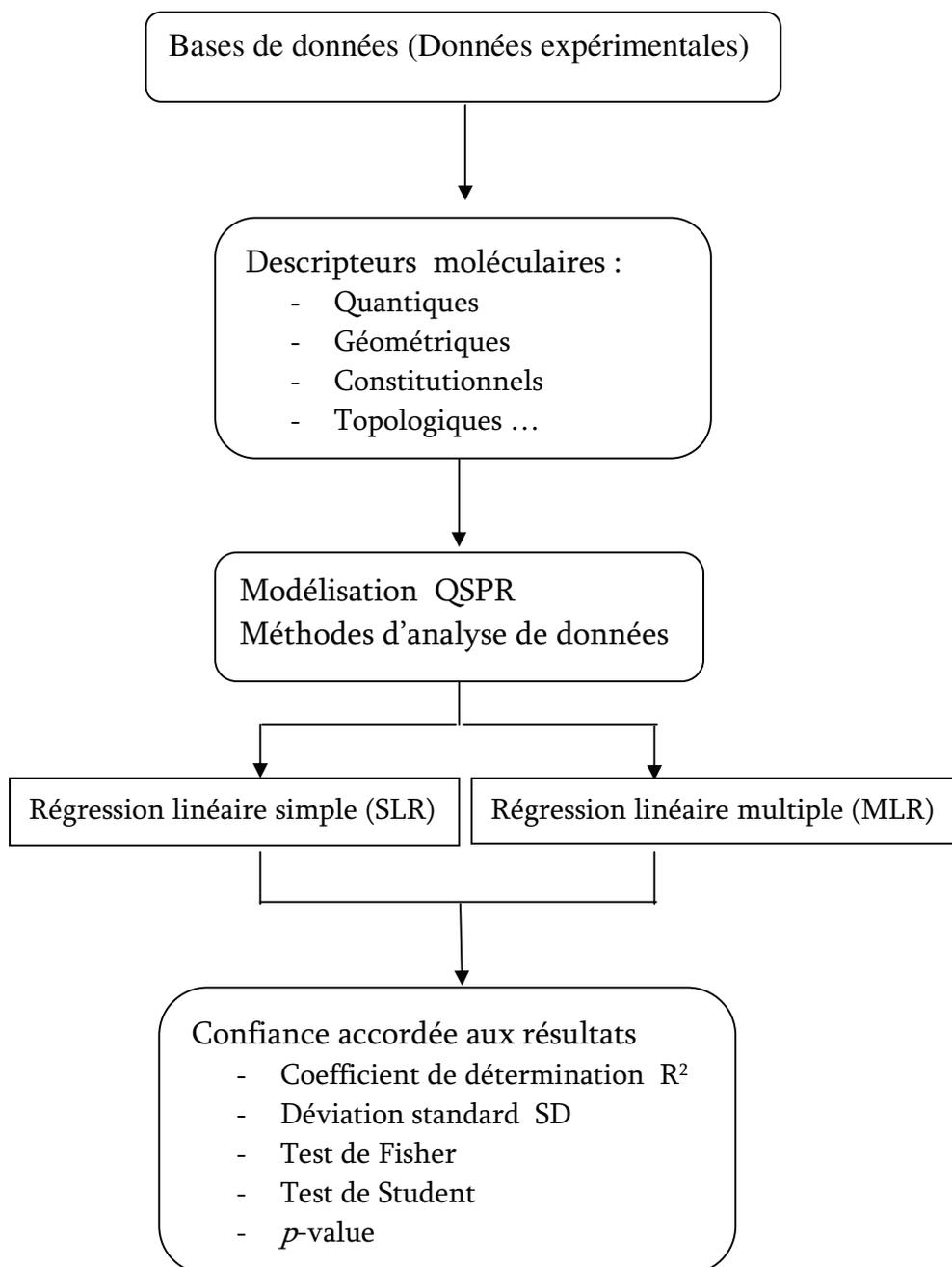
Le principe des méthodes QSPR est de mettre en œuvre une relation mathématique reliant de manière quantitative des propriétés moléculaires aussi bien électroniques que géométriques, appelées **descripteurs**, avec une observable macroscopique (propriété physico-chimique,...), pour une série de composés chimiques similaires à l'aide de méthodes d'analyses de données.

La forme générale d'un tel modèle est:  $Propriété/activité = f(descripteurs)$

L'objectif de ces méthodes est donc d'analyser les données structurales afin de détecter les facteurs déterminants pour la propriété (ou l'activité) mesurée. Pour ce faire, différents types d'outils peuvent être employés : régressions linéaires simples (SLR) et multiples (MLR) [6], régressions aux moindres carrés partiels (PLS) [7], arbres de décision [8], réseaux de neurones [9-11], algorithmes génétiques [12].

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

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



*Figure 1: Principe de la méthode QSPR*

## II.2. Importance des bases de données

Un modèle QSPR, est très dépendant des données expérimentales de référence. Le choix de la base de données est donc un point critique de son développement. Dans la plupart des cas, les données expérimentales sont issues de la littérature. Pour être de qualité, une base de données doit être composée de données expérimentales aussi fiables que possible, puisque les barres d'erreurs sur celles-ci se propageront dans le modèle final. L'obtention d'un modèle efficace sera également dépendante du type de molécules qui y sont incluses. Un jeu convenable devrait contenir des composés pour lesquels le mécanisme mis en jeu est unique. Plus le modèle présentera des molécules de structures proches, plus ce dernier aura de chances d'être performant au prix d'une réduction de son domaine d'applicabilité.

## II.3. Descripteurs moléculaires

Afin d'exploiter au maximum les informations contenues dans les structures moléculaires, celles-ci sont traduites en une série de grandeurs qui quantifient leurs caractéristiques physico-chimiques et structurelles. Ces grandeurs sont appelées descripteurs. Depuis des décennies, de nombreux travaux ont été menés afin de développer des descripteurs capables de décrire de la manière la plus exhaustive possible les structures moléculaires. On en dénombre aujourd'hui des milliers [13-15]. Ils peuvent être calculés ou empiriques. Ces derniers nécessitant d'être mesurés, les descripteurs calculés seront privilégiés. Parmi les différents types de descripteurs on peut citer : les descripteurs constitutionnels [16], les descripteurs topologiques [13,17-21], les descripteurs géométriques [22] et les descripteurs quantiques.

### II.3.1. Descripteurs quantiques

Afin d'aller plus loin dans la description des structures moléculaires, des caractéristiques supplémentaires de la structure moléculaire peuvent encore être quantifiées. Ces descripteurs qui, pour être calculés de manière fiable, nécessitent des calculs plus

sophistiqués sont regroupés ici sous l'étiquette de « descripteurs quantiques ». Les approches de chimie quantique nous donnent accès à des informations supplémentaires telles que des données énergétiques, vibrationnelles et orbitales (les orbitales frontières HOMO/LUMO) du système. Il est alors possible, par exemple, de calculer des énergies de dissociation. Pour ce faire, considérons une dissociation hétérolytique:



L'énergie de dissociation de la liaison A-B est alors la différence d'énergie entre la molécule AB et la somme des espèces ioniques  $A^+$  et  $B^-$  issus de la dissociation de la liaison considérée.

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

A partir de tels calculs, sont également évaluées des propriétés électroniques telles que le moment dipolaire, la polarisabilité, les charges électroniques calculées avec l'analyse de Mulliken [22,23] ou l'analyse de populations naturelles (NPA) [24,25]. Des descripteurs permettent également de quantifier les différentes interactions inter- et intramoléculaires, ces dernières ayant une grande influence sur la plupart des propriétés des systèmes moléculaires. Ce sont notamment des descripteurs de réactivité pouvant être dérivés de la DFT conceptuelle [26]

#### II.4. Méthodes d'analyse de données

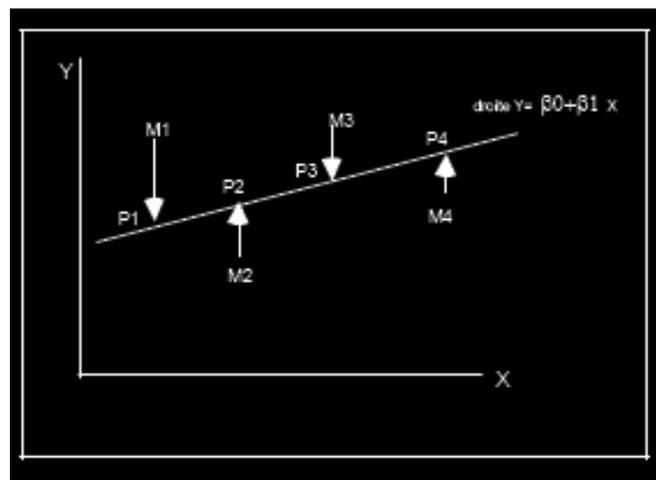
La mise en place de modèles QSPR n'est pas une chose aisée. La première difficulté réside dans la différence d'échelles existant entre les données à corrélérer, la structure étant à une échelle moléculaire alors que les propriétés à prédire sont à une échelle macroscopique. De plus, il tient compte des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

### II.4.1. Régression linéaire simple

Dans cette partie, nous allons présenter succinctement la régression linéaire simple [27-30]: nous pouvons la voir comme une technique statistique permettant de modéliser la relation linéaire entre une variable explicative (notée  $x$ ) et une variable à expliquer (notée  $y$ ). Nous supposons alors que la variable  $x$  dépend linéairement de la variable  $y$  mais cette liaison est perturbée par un « bruit ». Nous supposons en fait que les données suivent le modèle suivant :

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

L'équation 1 est appelée modèle de régression linéaire et dans ce cas précis modèle de régression linéaire simple. Les  $\beta_j$  appelés les paramètres du modèle, ou coefficients de régression, sont fixes mais inconnus. La quantité notée  $\varepsilon$  est appelée bruit, ou erreur, et elle est aléatoire et inconnue. Afin d'estimer les paramètres inconnus du modèle, nous mesurons dans le cadre de la régression simple une seule variable  $x$  appelée « régresseur » ou variable indépendante et une variable  $y$  appelée variable « réponse » ou variable dépendante. Comment trouver la droite qui passe « au plus près » de tous les points :



*Figure 2.*

Pour trouver la droite qui passe « au plus près » de tous les points il faut se donner un *critère d'ajustement*. On projette les points de  $M_1$  à  $M_4$  parallèlement à l'axe  $y$ . Sur la droite on obtient les points  $P_1$  à  $P_4$ , comme le montre la **Figure 2**. Le critère retenu pour déterminer la droite passant au plus près de tous les points sera tel que : « La somme des carrés des écarts (SCE) des points observés  $M_i$  à la droite solution soit minimum ».

La droite solution sera appelée droite de régression de  $y$  sur  $x$ . Le critère d'ajustement est le « critère des Moindres Carrés Ordinaires » (MCO, *Ordinary Least Squares*). Les écarts sont calculés en projetant les points  $M_i$  parallèlement à l'axe  $y$ .

### 1. Méthode d'estimation des paramètres $\beta_0$ et $\beta_1$

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

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

La valeur de cette fonction  $S(\beta_0, \beta_1)$  est minimum lorsque les dérivées de  $S$  par rapport à  $\beta_0, \beta_1$  s'annulent. Nous obtenons un système d'équations appelées « équations normales » :

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

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

$\hat{\beta}_0$  et  $\hat{\beta}_1$  sont les estimateurs des moindres carrés des paramètres inconnus  $\beta_0$  et  $\beta_1$ . La première équation donne :

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

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

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{6}$$

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  $\hat{\beta}_1$  :

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

Cet estimateur peut être écrit sous une autre forme :

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

cov(x,y) et var(x) sont la covariance et la variance. Le coefficient  $\hat{\beta}_1$  dépend au numérateur de la covariance entre x et y et de la variance de x pour le dénominateur.

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

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \tag{9}$$

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)  $y_i$ . A partir des équations 6 et 9 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 :

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

Ces diverses sommes des carrés sont définies comme suit :

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

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

\* La quantité SSR (Sum of Squares Regression or model) représente la somme des carrés des écarts à la moyenne lorsque l'on remplace les observations  $y_i$  par les prévisions obtenues à l'aide de la régression :

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

\* La quantité 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:

On a l'habitude de représenter l'analyse de la variance sous forme d'un tableau, appelé table d'analyse de variance ou table d'ANOVA (ANalysis Of Variance) faisant apparaître les sources de variation : le total (en 3<sup>ème</sup> ligne de la table) qui se décompose deux parties : la partie modèle et la partie erreur. A chaque source de variation correspond un nombre de degrés de liberté (DF) respectivement égal à  $n-1$ ,  $p$ ,  $n-p-1$  (ou  $n$  est le nombre d'observations,  $p$  le nombre de variables régresseurs (la variable  $x_0$ , constante égale à 1, correspondant au paramètre  $\beta_0$ , n'est pas comprise).

Nous présentons le tableau général de l'analyse de variance pour p régresseurs. Tous les logiciels effectuant des calculs de régression, donnent comme sortie ce tableau.

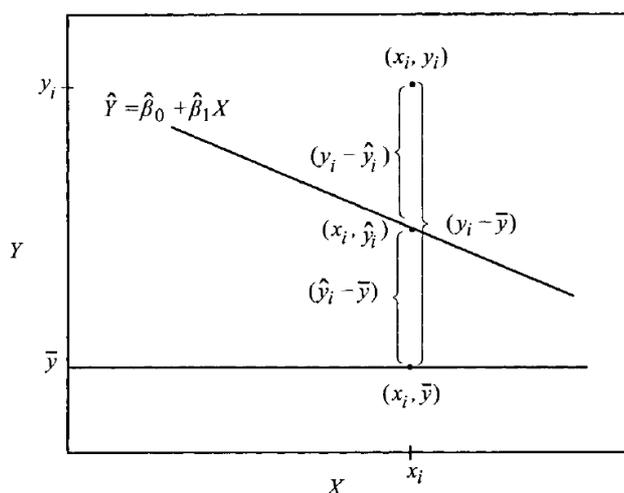
**Table d'analyse de variance (ANOVA)**

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	

Tous ces indicateurs SS, MS, MSE, vont jouer un rôle important dans l'appréciation du modèle calculé à partir des observations.

Toutes ces quantités s'interprètent de manière géométrique comme le montre la **Figure 3**

où l'on a représenté les vecteurs  $y$ ,  $\hat{y}$  et  $\bar{y}$  :



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

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

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

Cette formule montre que les variations de  $y$  de sa moyenne, c'est-à-dire SCE Totale (SS Total) peuvent être expliquées par le modèle grâce SCE Modèle (SS Model ou SS regression) et ce que ne peut être expliqué par le modèle est dans SCE Erreurs (SS Error).

### 3. Validation du modèle

- **Le coefficient de détermination  $R^2$**

Un modèle, que l'on qualifiera de bon, possédera des estimations  $\hat{y}_i$  proches des vraies valeurs de  $y_i$ . Les deux quantités SCE totale (SST) et SCE modèle (SSR) sont des sommes de carrés donc toujours positives ou nulles et telles que :  $SSR \leq SST$ . Le rapport des ces deux quantités est compris entre 0 et 1 que l'on appelle le coefficient de détermination :

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

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

Si tous les points  $y_i$  observés sont alignés sur la droite de régression, le modèle est parfaitement adapté et SCE Erreur = 0 (cas particulier).

#### Interprétation de $R^2$

$R^2$  qui varie entre 0 et 1, mesure la proportion de variation totale de  $y$  autour de la moyenne expliquée par la régression. Plus la valeur de  $R^2$  sera proche de 1 (cas idéal) et plus les

valeurs prédites et observées sont corrélées. Un  $R^2$  faible signifie que le modèle à un faible pouvoir explicatif et les descripteurs sont sans effet sur la réponse.

Il n'est pas recommandé d'utiliser  $R^2$  pour comparer des modèles. Par exemple, si on hésite entre  $y \approx \beta_0 + \beta_1 x_1$  et  $y \approx \beta_0 + \beta_1 x_1 + \beta_2 x_2$ , le coefficient  $R^2$  nous dira toujours de choisir le second modèle car son  $R^2$  sera plus important (on projette sur un espace plus grand), même si la variable  $x_2$  est sans rapport avec la réponse  $y$ . Pour comparer des modèles, en pénalisant les modèles les plus complexes, il existe de nombreux indicateurs. Parmi ceux-ci, le coefficient de détermination ajusté découle simplement de notre table d'analyse de variance.

- **Coefficient de détermination ajusté**

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

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

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

Avec  $Intercept=0$ , si il n'y a pas de constante  $\beta_0$  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  $\sigma^2$ (ou  $s^2$ ).**

La variance  $\sigma^2$  est estimée par la formule suivante:

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

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

$$s = SD = \sqrt{s^2} \tag{26}$$

### II.4.2. Régression linéaire multiple

Le modèle de régression multiple MLR (*Multiple Linear Regression*) est une généralisation du modèle de régression simple. La régression multi-linéaire [28,29-31] est la méthode la plus simple et la plus employée pour le développement de modèles prédictifs. Elle repose sur l'hypothèse qu'il existe une relation linéaire entre une variable dépendante  $y$  (ici, la propriété) et une série de  $p$  variables indépendantes  $x_i$  (ici, les descripteurs). L'objectif est d'obtenir une équation de la forme suivante :

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

où  $\beta_i$  sont les coefficients de la régression.

La détermination de l'équation 27 se fait alors à partir d'une base de données de  $n$  échantillons pour laquelle à la fois les variables dépendantes et la variable indépendante sont connues. Il s'agit donc de considérer un système de  $p$  équations.

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

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

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

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

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  $Y=\beta X + \varepsilon$

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

En premier lieu, les variables indépendantes  $x_i$ , comme leur nom l'indique, sont supposées indépendantes entre elles et leur incertitude est négligeable. Ensuite, les différents échantillons  $y_i$  sont supposés indépendants entre eux.. Enfin, par nature, la dépendance de  $y$  vis-à-vis des  $x_i$  est supposée linéaire.

La valeur prédite de la variable dépendante est alors :

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

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

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

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

### II.4.3. Confiance accordée aux résultats.

La logique statistique nous conduit à effectuer des tests d'hypothèses, et à déterminer des intervalles de confiance autour des valeurs estimées. Successivement on va chercher à:

- tester la signification globale de la régression,
- tester l'hypothèse nulle  $\beta_1=0$  et à calculer l'intervalle de confiance de  $\beta_1$ ,
- tester l'hypothèse nulle  $\beta_0=0$  et à calculer l'intervalle de confiance de  $\beta_0$ ,
- calculer la précision de l'estimation de  $y$  pour la moyenne et pour une observation individuelle.

#### 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  $H_a$ : il existe au moins un  $\beta_j$  parmi  $\beta_1, \dots, \beta_p$  non égal à 0.

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

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

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

Les expressions 34 et 35 représentent respectivement la somme des carrés des écarts moyens pour le modèle et pour l'erreur.

Le test de Fisher mesure le rapport entre la variance de la variable dépendante expliquée et non expliquée par le modèle de régression. En d'autres termes le test Fisher-Snedecor permet de tester l'hypothèse nulle selon laquelle chaque  $\beta$  est significativement différent de zéro, ce qui est signe d'une relation évidente entre la variable expliquée et les variables explicatives.

Intuitivement, nous rejetterons l'hypothèse nulle lorsque la somme des carrés expliquée par la régression est grande. En d'autres termes, la région critique de ce test est de la forme  $\{F > \text{seuil}\}$ . Si la quantité F observée dépasse le seuil, on rejette l'hypothèse  $H_0$  dans le cas contraire, on conserve  $H_0$ .

**Raisonnement sur la p-value**

Pour éviter de raisonner sur F, le programme fournit la p-value associée au F observé. La p-value est le niveau de significativité du test de Fisher-Snedecor, c'est-à-dire la probabilité de dépasser le F observé si l'hypothèse nulle est vraie. On compare la p-value au risque  $\alpha$  choisi (par exemple  $\alpha=0.05$ ).

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

Ces résultats permettent d'interpréter les tables d'analyse de variance complètes fournies par tout logiciel mettant en oeuvre la régression linéaire. La table complète 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  $\beta_1$**

En ce qui concerne la pertinence des descripteurs dans le modèle, elle est également évaluée par le  $t$ -test de Student. Il s'agit de tester l'hypothèse considérant le descripteur comme non significatif. Pour une régression multi-linéaire, cela revient à supposer le coefficient  $\beta_i$  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 : 
$$t - \text{observé} = \frac{\hat{\beta}_1}{s(\hat{\beta}_1)} \tag{36}$$

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

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

**Raisonnement**

On compare la p-value associée à  $t$ -observé, au risque  $\alpha$  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  $\alpha$ .

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

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

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

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  $\alpha$  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  $\alpha$ .

Finalement, on peut dire que la statistique t permet de tester l'hypothèse (nulle) selon laquelle la valeur des coefficients de régression ne sont pas significativement différents de zéro (en d'autres termes, qu'il existe bien une relation entre la variable dépendante et la variable indépendante en question). 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.

Tableau récapitulatif des statistiques relatives à l'analyse de la variance

Statistique	Formule	Signification
Mean square	$MS = \frac{SS}{DF}$	Le rapport d'une somme des carrés des écarts (SS) divisée par le rapport de degrés de liberté (DF)
F-value	$F = \frac{MS Model}{MS Error}$	Statistique d Fisher-Snedecor pour tester si tous les paramètres $\beta$ sont nuls
s=SD	$\sqrt{MS Error}$	Standard déviation, l'écart moyen résiduel C'est l'estimation de « s », l'écart-type des erreurs
R-square	$R^2 = \frac{SS Model}{SS Total}$ $R^2 = 1 - \frac{SS Error}{SS Total}$	F et R-square ( $R^2$ ) ont liés par la relation suivante $F = \frac{(n - p - 1)}{p} * \frac{R^2}{(1 - R^2)}$
Adj. R square	$R_{ajusté}^2 = 1 - \frac{(n - Intercept)}{n - p - 1} (1 -$	$R^2$ ajusté en fonction du nombre des régresseurs du modèle.
Mean Value	$\bar{y} = \sum_{i=1}^n \frac{y_i}{n}$	Moyenne de la variable réponse y.

## 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] A.J. Myles, R.N. Feudale, Y. Liu, N.A. Woody, S.D. Brown, *J. Chemom.* 18 (2004) 275.
- [9] A.F. Duprat, T. Huynh, G. Dreyfus, *J. Chem. Inf. Comput. Sci.* 38 (1998) 586.
- [10] I.V. Tetko, A.E.P. Villa, D.J. Livingstone, *J. Chem. Inf. Comput. Sci.*, 36 (1996) 794.
- [11] J. Gasteiger, J. Zupan, *Angew. Chem. Int. Ed. Engl.* 32 (1993) 503.
- [12] R. Leardi, *J. Chemometr.* 15 (2001) 559.
- [13] M. Karelson, "Molecular Descriptors in QSAR/QSPR", Wiley, New York, 2000.
- [14] R. Todeschini, V. Consonni, "Handbook of Molecular Descriptors", Wiley, 2000.
- [15] M. Karelson, V.S. Lobanov, A.R. Katritzky, *Chem. Rev.* 96 (1996) 1027.
- [16] S.W. Benson, J.H. Buss, *J. Chem. Phys.* 29 (1958) 546.
- [17] H. Wiener, *J. Am. Chem. Soc.* 69 (2002) 17.
- [18] M. Randic, *J. Am. Chem. Soc.* 97 (2002) 6609.
- [19] L.B. Kier, L.H. Hall, *J. Pharm. Sci.* 70 (1981) 583.
- [20] A.T. Balaban, *Chem. Phys. Lett.* 89 (1982) 399.
- [21] S.C. Basak, D.K. Harriss, V.R. Magnuson, *J. Pharm. Sci.* 73 (1984) 429.
- [22] G. Fayet, P. Raybaud, H. Toulhoat, de Bruin, *J. Mol. Struct. (Theochem)* 903 (2009) 100.
- [23] R.S. Mulliken, *J. Chem. Phys.* 23 (1955) 1833.
- [24] A.E. Reed, F. Weinhold, *J. Chem. Phys.* 83 (1985) 1736.
- [25] A.E. Reed, L.A. Curtiss, F. Weinhold, *Chem. Rev.* 88 (1988) 899.

- [26] (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.
- [27] S. Chatterjee and A. S. Hadi, "Regression Analysis by example", fourth Edition, a John Wiley & Sons, Inc., Publication, 2006.
- [28] S. Weisberg, "Applied Linear Regression", third Edition, John Wiley & Sons, Inc., Publication, 2005.
- [29] A. C. Rencher, G. B. Schaalje, "Linear Models in Statistics", Second Edition, John Wiley & Sons, Inc., Publication, 2008.
- [30] P. A. Cornillon, E.M.atzner-Løber, "Régression théorie et Applications ", Springer-Verlag France, Paris, 2007.
- [31] M. Lejeune, "Statistiques : la théorie et ses applications", Springer-Verlag, Paris, 2004

# **CHAPITRE III**

## ***RESULTATS ET DISCUSSIONS***

***PARTIE 1 :***

*Corrélation des constantes d'acidité aqueuse des acides carbonés avec les descripteurs théoriques.*

**Résumé**

*Des calculs théoriques ont été effectués pour prédire les acidités en phase gazeuse et en phase aqueuse d'une série de 21 acides carbonés avec des valeurs de  $pK_a$  variant de -6.20 à 50 unités. Des corrélations linéaires satisfaisantes ( $R^2 > 0.93$ ,  $SD < 4$ ) ont été obtenues entre les variations d'énergies libres de déprotonation calculées et les valeurs expérimentales de  $pK_a$  mesurées dans l'eau. Les effets de solvant ont été considérés au moyen du modèle PCM (Polarizable Continuum Model). Nos calculs montrent également que l'acidité élevée des composés carbonyles et des méthanes substitués par des groupements électro-attracteurs, peut être justifiée par les fortes interactions de transfert de charge vicinal  $n_c \rightarrow \pi^*$  dans les espèces anioniques (bases conjuguées). Les calculs ont été effectués au niveau B3LYP/6-311++G (d, p).*

**1. Introduction**

Les énergies de déprotonation des acides organiques et les affinités protoniques des bases conjuguées correspondantes sont largement utilisées pour la prédiction de l'acidité de Brønsted en phases gazeuse et aqueuse [1-8]. Les acides forts ont de petites valeurs d'énergie de déprotonation (i.e., une libération facile du proton) alors que les bases fortes ont des grandes valeurs d'affinité protonique (i.e., une forte fixation du proton). Plusieurs travaux sur la prédiction théorique de l'acidité des acides organiques et inorganiques peuvent être trouvés dans la littérature. Par exemple, Smith et Radom [9,10] ont montré que les méthodes G2 et

G2(MP2) fournissent d'excellents résultats pour les enthalpies de déprotonation et les affinités protoniques de petites molécules. Catalan et Palomar [11] ont étudié les acidités en phase gazeuse de plusieurs espèces et ont montré que les calculs B3LYP avec les bases 6-311+G(d) et 6-311+G(3df,3pd) concordent bien avec les données expérimentales. De bonnes corrélations ont été également obtenues entre les valeurs expérimentales de  $pK_a$  d'une large gamme d'acides organiques de Brønsted et leurs enthalpies déprotonation calculées en phase gazeuse [12]. Des corrélations entre les descripteurs théoriques et les constantes d'acidité en phase gazeuse ont été élaborées pour les amines, les alcools et les thiols [13-15]. Une excellente corrélation a été également obtenue, avec la méthode HF/3-21G(d), entre l'acidité en phase aqueuse et les valeurs expérimentales de  $pK_a$  pour une série de bases azotées [16].

Récemment, divers descripteurs théoriques ont été utilisés pour étudier leur corrélation avec l'acidité carbonée des composés ayant la structure  $CHR_1R_2R_3$ . Une corrélation satisfaisante a été obtenue aux niveaux HF/3-21G et B3LYP/6-31G(d) [17,18] entre les énergies de déprotonation et les valeurs de  $pK_a$  mesurées en phase aqueuse. L'aptitude de prédire l'acidité en utilisant des approches théoriques sans faire appel aux données expérimentales serait très utile aux chimistes surtout avec le développement spectaculaire de la chimie computationnelle. Cependant, les prédictions théoriques en solution ne sont pas toujours satisfaisantes. La difficulté est due principalement à la bonne estimation de l'énergie de solvatation.

Dans cette application, on se propose d'élaborer un modèle théorique pour corréler les valeurs expérimentales de  $pK_a$  aqueuses avec les enthalpies libres de déprotonation calculées au niveau B3LYP/6-311++G(d,p). Les corrélations ont été élaborées en phase gazeuse et en phase aqueuse pour une série constituée de 21 acides carbonés ayant des valeurs de  $pK_a$  comprises entre -6.20 et 50 [18-21] (**Figure 1**). En outre, les énergies de transfert de charge  $E^{(2)}$ , calculées par la méthode NBO (*Natural Bond Orbital*) [22-26],

permettent d'expliquer l'acidité élevée des composés dicarboxylés et alcanes substitués par des groupements attracteurs d'électrons.

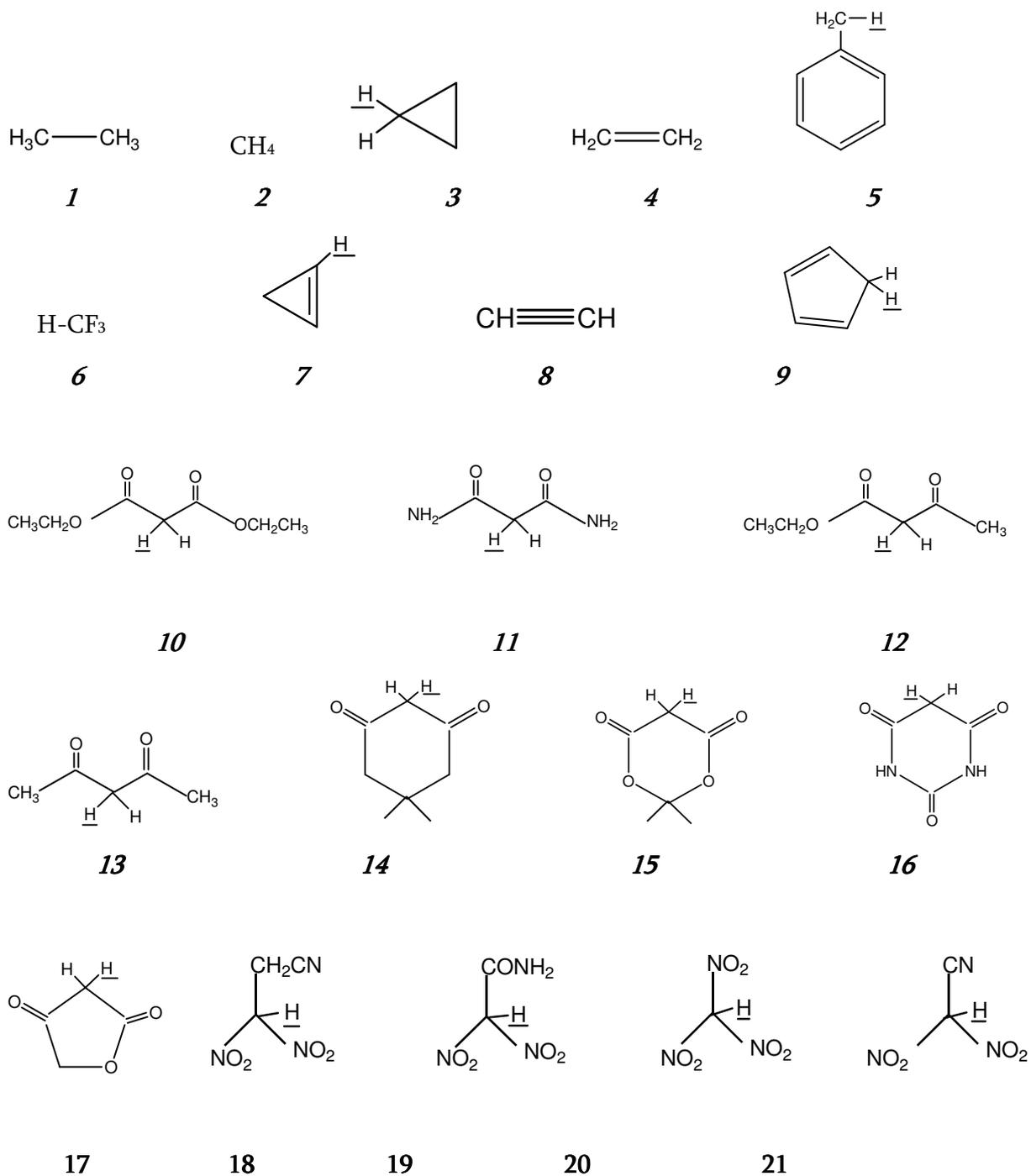


Figure 1: Structures des acides carbonés

## 2. Méthodologie des calculs

### 2.1. Calcul d'acidité en phase gazeuse

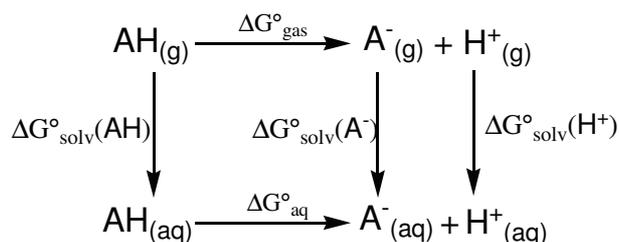
Les calculs ont été effectués à l'aide du programme G98W [27]. Les géométries d'équilibre des espèces neutres et anioniques ont été optimisées avec la méthode B3LYP/6-311++G(d,p). Afin de caractériser les points stationnaires obtenus, les calculs de fréquence ont été réalisés au même niveau de théorie. L'énergie libre de Gibbs de déprotonation de l'acide en phase gazeuse ( $\Delta G^\circ_{\text{gas}}$ ) est calculée en utilisant l'équation 1.



La valeur expérimentale  $G^\circ_{\text{gas}}(\text{H}^+) = -6.28$  kcal/mol [28] a été utilisée dans les calculs.

### 2.2. Calcul d'acidité en phase aqueuse

Les effets de solvant ont été considérés au moyen du modèle PCM [29-31] en effectuant des calculs *single point* d'énergie sur les géométries optimisées en phase gazeuse. Les calculs PCM utilisent le rayon atomique UAHF (*United Atom Topological Model applied on radii optimized for the HF/6-31G(d) level of theory*) pour calculer l'énergie libre de Gibbs de solvation  $\Delta G^\circ_{\text{solv}}$ . Le cycle thermodynamique (**schéma 1**) et les équations 2 et 3 représentent un outil fiable pour calculer l'énergie de déprotonation en phase aqueuse  $\Delta G^\circ_{\text{aq}}$  en combinant  $\Delta G^\circ_{\text{gas}}$  avec la variation de l'énergie libre de solvation ( $\Delta \Delta G^\circ_{\text{solv}}$ ) entre l'acide AH et sa base conjuguée A<sup>-</sup>. La valeur expérimentale  $G^\circ_{\text{solv}}(\text{H}^+)$  utilisée dans le calcul est de l'ordre de -264.61 kcal/mol [28]. Le pK<sub>a</sub> aqueux est calculé suivant l'équation 4.



*Schéma 1 : Cycle thermodynamique utilisé pour le calcul de  $\Delta G^{\circ}_{\text{aq}}$*

$$\Delta G^{\circ}_{\text{aq}} = G^{\circ}_{\text{aq}}(\text{A}^{-}) + G^{\circ}_{\text{aq}}(\text{H}^{+}) - G^{\circ}_{\text{aq}}(\text{AH}) \quad (2)$$

$$\Delta G^{\circ}_{\text{aq}} = \Delta G^{\circ}_{\text{gas}} + \Delta \Delta G^{\circ}_{\text{solv}} \quad (3)$$

$$\text{avec: } \Delta G^{\circ}_{\text{gas}} = G^{\circ}_{\text{gas}}(\text{A}^{-}) + G^{\circ}_{\text{gas}}(\text{H}^{+}) - G^{\circ}_{\text{gas}}(\text{AH})$$

$$\Delta \Delta G^{\circ}_{\text{solv}} = \Delta G^{\circ}_{\text{solv}}(\text{A}^{-}) + (G^{\circ}_{\text{solv}}(\text{H}^{+}) - (G^{\circ}_{\text{solv}}(\text{AH}))$$

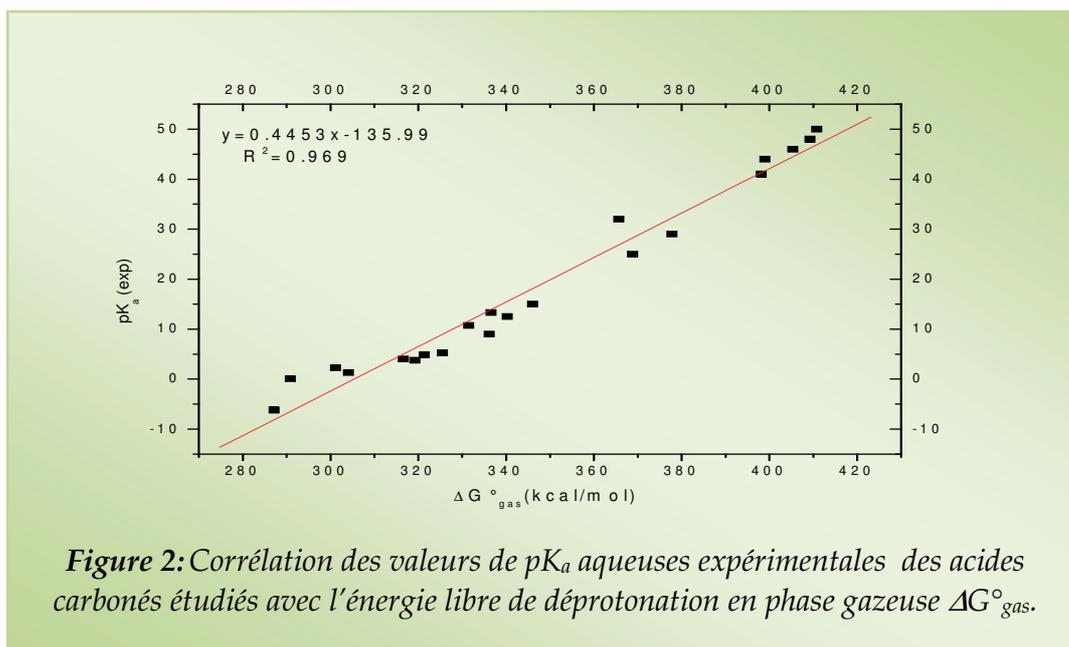
$$\text{pK}_{\text{a}} = \frac{\Delta G^{\circ}_{\text{aq}}}{RT \ln 10} \quad (4)$$

### 3. Résultats et discussions

#### 3.1. L'acidité en phase gazeuse

Les énergies  $\Delta G^{\circ}_{\text{gas}}$  en phase gazeuse et les valeurs expérimentales aqueuses de  $\text{pK}_{\text{a}}$  à 298.15 K des acides carbonés étudiés [18-21] sont regroupées dans le **Tableau 1**.

La **Figure 2** représente la corrélation des valeurs  $\Delta G^{\circ}_{\text{gas}}$  avec les  $\text{pK}_{\text{a}}$  aqueuses expérimentales.



La corrélation linéaire obtenue est satisfaisante. Les valeurs  $\Delta G^\circ_{gas}$  de déprotonation de l'acide sont inversement liées à l'acidité carbonée. L'équation de régression avec les valeurs correspondantes de la déviation standard (SD) et le coefficient de corrélation ( $R^2$ ) sont donnés dans l'équation 5.

$$pK_a = 0.4453 \Delta G^\circ_{gas} - 135.99 \quad (R^2=0.969, SD=3.27) \quad (5)$$

Les valeurs de  $pK_a$  calculées à partir  $\Delta G^\circ_{gas}$  en utilisant l'équation 5 pour les acides carbonés étudiés sont résumées dans le **Tableau 1**. Les enthalpies libres calculées et les enthalpies de solvation des différents acides et bases conjuguées sont regroupées dans le **Tableau S0** (voir annexe I)

**Tableau 1 :** Les valeurs calculées de l'énergie de dissociation de l'acide  $\Delta G^\circ$  (en kcal/mol) et les constantes d'acidité  $pK_a$  calculées en phase gazeuse, aqueuse et expérimentales des composés étudiés.

Composé	Phase gazeuse		Phase aqueuse			
	$\Delta G^\circ_{\text{gas}}(\text{calc})$	$pK_a(\text{calc})^a$	$\Delta G^\circ_{\text{aq}}(\text{calc})$	$pK_a(\text{calc})^b$	$pK_a(\text{calc})^c$	$pK_a(\text{exp})$
1	410.77	46.91	78.78	57.75	48.80	50 <sup>d</sup>
2	409.25	46.24	53.56	39.26	32.62	48 <sup>d</sup>
3	405.33	44.49	69.86	51.21	43.08	46 <sup>d</sup>
4	398.98	41.66	58.28	42.72	35.65	44 <sup>d</sup>
5	398.12	41.28	73.87	54.15	45.65	41 <sup>e</sup>
6	365.67	26.83	36.84	27.01	21.89	32 <sup>e</sup>
7	377.80	32.23	46.66	34.20	28.19	29 <sup>d</sup>
8	368.83	28.24	28.95	21.22	16.82	25 <sup>d</sup>
9	346.06	18.10	16.99	12.45	9.15	15 <sup>e</sup>
10	336.62	13.90	16.38	12.01	8.76	13.30 <sup>f</sup>
11	340.31	15.54	21.85	16.02	12.27	12.50 <sup>f</sup>
12	331.50	11.62	13.06	9.57	6.63	10.70 <sup>f</sup>
13	336.15	13.69	21.17	15.52	11.83	9.00 <sup>f</sup>
14	325.46	8.93	12.14	8.90	6.03	5.25 <sup>e</sup>
15	321.34	7.09	9.87	7.24	4.58	4.83 <sup>e</sup>
16	316.57	4.97	10.12	7.42	4.74	4.01 <sup>e</sup>
17	319.21	6.14	7.88	5.78	3.30	3.76 <sup>e</sup>
18	301.18	-1.88	-2.47	-1.81	-3.34	2.27 <sup>f</sup>
19	304.05	-0.60	-0.45	-0.33	-2.05	1.30 <sup>f</sup>
20	290.84	-6.49	-6.74	-4.94	-6.08	0.06 <sup>f</sup>
21	287.15	-8.13	-15.17	-11.12	-11.49	-6.20 <sup>f</sup>

<sup>a</sup> Valeurs calculées avec l'équation de corrélation entre  $pK_a(\text{exp})$  et  $\Delta G^\circ_{\text{gas}}$  (eq 6, voir figure 2).

<sup>b</sup> Valeurs calculées avec l'équation 5. <sup>c</sup> Valeurs calculées en utilisant l'équation de corrélation entre  $pK_a(\text{exp})$  et  $pK_a(\text{cal})$  (eq 7, voir figure 4).<sup>d</sup> Ref. [18a], <sup>e</sup> Ref.[21] <sup>f</sup> Ref. [18b, 19, 20]

La déviation moyenne MD (mean deviation, éq. 6) et la déviation moyenne absolue AD (absolute deviation, éq. 7) entre les valeurs des constantes d'acidité expérimentales et prédites par l'équation de régression (éq. 5) sont -0.009 et 2.71 unités de pK<sub>a</sub>, respectivement. Les plus grandes valeurs de déviation absolue AD (supérieur à 5 unités de pK<sub>a</sub>) correspondent aux composés **6** et **20**. La prédiction des pK<sub>a</sub> aqueuses, en utilisant l'équation de régression 6, des composés **6** et **20** sont 26.83 (au lieu de 32) et -6.49 (au lieu de 0.06), respectivement. Il devrait être noté que les constantes pK<sub>a</sub> aqueuses exactes (relatif à l'eau) peuvent être obtenus seulement pour les acides plus faibles que l'ion hydronium H<sub>3</sub>O<sup>+</sup> (pK<sub>a</sub> = -1.74) et plus forts que l'eau H<sub>2</sub>O (pK<sub>a</sub>=15.74) [1]. Les valeurs expérimentales de pK<sub>a</sub> aqueuses des acides carbonés forts (< -1.74) et faibles (> 15.74) donnés dans le **Tableau 1** doivent être considérées comme approximatives.

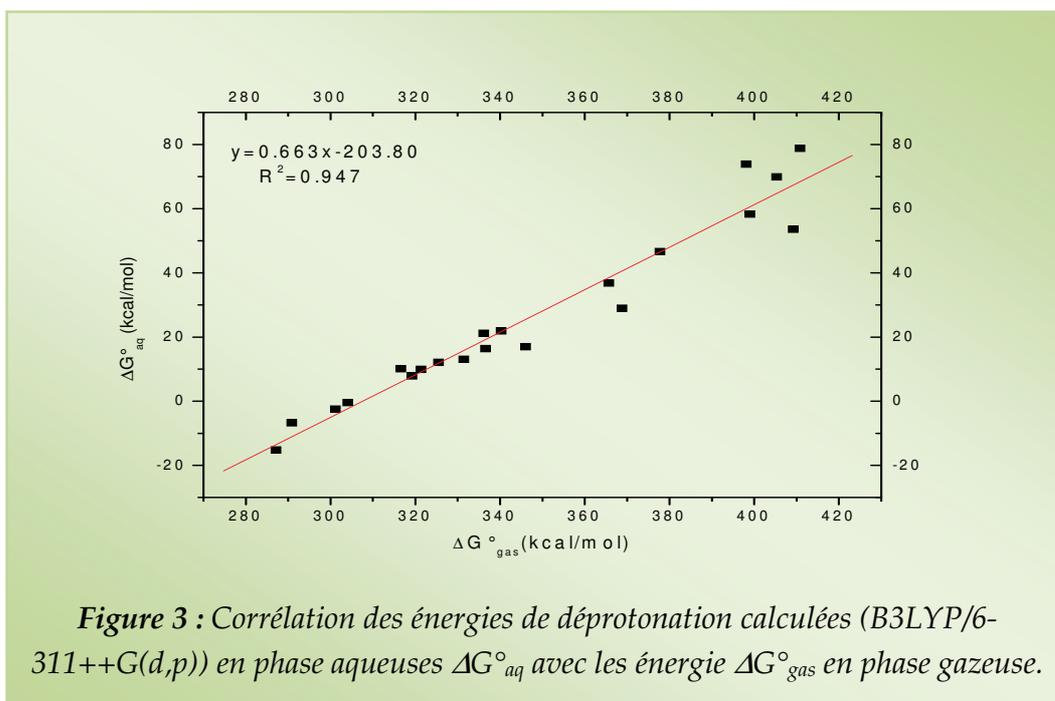
$$MD = \left[ \sum_{i=1}^{21} (pKa_i(\text{exp}) - pKa_i(\text{cal})) \right] / 21 \quad (6)$$

$$AD = \left[ \sum_{i=1}^{21} |pKa_i(\text{exp}) - pKa_i(\text{cal})| \right] / 21 \quad (7)$$

A partir des résultats obtenus, nous pouvons déduire que l'ordre d'acidité en phase aqueuse pour la famille des acides carbonés étudiés est similaire à celui obtenu en phase gazeuse. Par conséquent, il est judicieux d'utiliser les calculs en phase gazeuse pour prédire les constantes d'acidité en phase aqueuse. On note que pour certaines familles, l'ordre d'acidité en phase gazeuse (où les effets de solvatation sont complètement absents) peut être différent de celui de la phase aqueuse. Les alcools simples constituent un exemple de ce phénomène. Pendant que les alcools tertiaires sont plus acides que les alcools secondaires et primaires en phase gazeuse, l'ordre d'acidité est complètement renversé en solution aqueuse car l'anion tertiaire assez volumineux est faiblement solvato. Le même phénomène est observé dans le cas des acides carboxyliques aliphatiques simples et dans le cas de la basicité des amines [1].

### 3.2. L'acidités en phase aqueuse

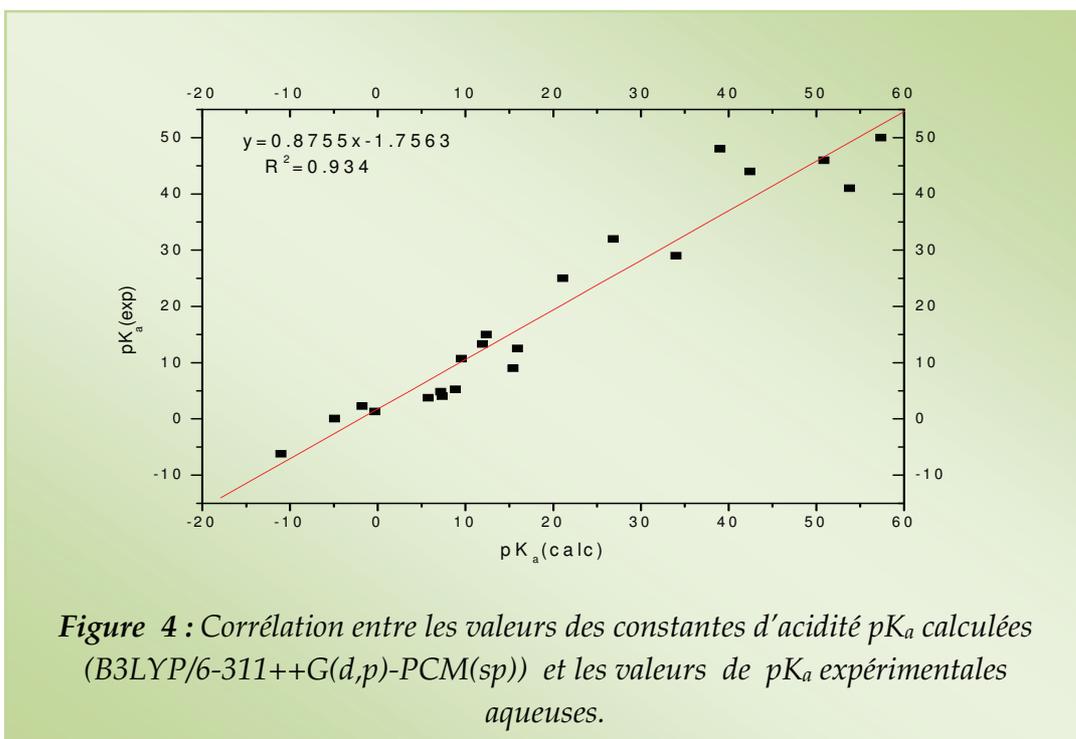
Les valeurs calculées de  $\Delta G^{\circ}_{aq}$  en phase aqueuse et les valeurs expérimentales des  $pK_a$  des acides carbonés étudiés (en utilisant l'équation 4 à 298.15 K) sont reportées dans le **Tableau 1**. La droite de la corrélation  $\Delta G^{\circ}_{aq}$  avec  $\Delta G^{\circ}_{gas}$  est montrée dans la **Figure 3**. La corrélation satisfaisante obtenue entre ces deux variations d'énergie libre de Gibbs ( $R^2=0.947$ ) pourrait aider à expliquer la bonne corrélation précédemment trouvée entre  $\Delta G^{\circ}_{gas}$  et les valeurs expérimentales aqueuses de  $pK_a$ .



La **Figure 4** représente la corrélation entre les valeurs expérimentales aqueuses et les valeurs calculées de  $pK_a$ . L'équation de régression avec les valeurs de la déviation standard correspondante (SD) et du coefficient de corrélation ( $R^2$ ) sont données dans l'équation 8.

$$pK_a(\text{exp}) = 0.8755 pK_a(\text{calc}) - 1.7563 \quad (8)$$

( $R^2=0.934$ , SD=4.80)



Les déviations moyennes MD et les déviations moyennes absolues AD entre les valeurs l'expérimentales et les valeurs calculées sont -0.64 et 4.39 unités de  $pK_a$ , respectivement. Les plus grandes valeurs de AD (supérieur à 5 unités du  $pK_a$ ) correspondent aux les composés **1**, **2**, **3**, **5**, **7** et **13**.

Les valeurs de  $pK_a$  aqueuses des composés **6** et **20** prédites avec équation 8 sont 21.89 et -6.08 respectivement. Les valeurs calculées de  $pK_a$  en phase aqueuse (obtenues en utilisant l'équation 8) sont donnés dans **Tableau 1**.

En comparant les valeurs de AD obtenues en phase gazeuse et en phase aqueuse on peut conclure que le calcul de  $pK_a$  au moyen de l'équation 5 (phase gazeuse) semble être plus fiable que l'équation 8 (phase aqueuse). Ceci peut être expliqué par les insuffisances du modèle PCM dans le calcul précis *single point* des énergies de solvation. L'inclusion des effets du solvant dans les optimisations de la géométrie et dans le calcul des fréquences de vibration s'imposent [32].

L'acidité en phase aqueuse des acides carbonés de la série étudiée peut être expliquée par la stabilité relative de leurs bases conjuguées. En effet, les anions sont stabilisés par une combinaison des effets de résonance et des effets inductifs [33]. En outre, il a été montré que la délocalisation est un facteur important responsable de l'augmentation de l'acidité [34-36]. Cependant, les effets inductifs produits par des substituants attracteurs d'électrons, ont le même ordre d'importance que les effets de délocalisation. L'acidité s'élève quand la délocalisation de la charge négative de la base conjuguée augmente [34]. Les valeurs d'acidité  $\Delta G^\circ_{\text{gas}}$  des composés **1-21** diminuent avec l'accroissement des effets de délocalisation. La délocalisation de la charge (épuisement de la charge négative) peut être expliquée par l'interaction donneur-accepteur entre l'orbitale non-liante  $n_C$  localisée sur l'atome de carbone (portant la charge négative) de la base conjuguée et l'orbitale vacante anti-liante vicinale  $\pi^*$  liée aux groupements C=C, C=O ou N=O. Cette interaction orbitale est appelée transfert de charge vicinale  $n_C \rightarrow \pi^*$ . L'énergie de stabilisation de cette interaction peut être calculée par la méthode NBO et la théorie des perturbations de second ordre selon l'équation 9 [22-26], où  $\epsilon_n$  et  $\epsilon_{\pi^*}$  sont les énergies des orbitales et F sont les éléments de la matrice NBO de Fock.

$$\Delta E_{n_C \rightarrow \pi^*}^{(2)} = -2(F_{n \rightarrow \pi^*})^2 / (\epsilon_n - \epsilon_{\pi^*}) \quad (9)$$

Par exemple, l'énergie de transfert de charge dans la base conjuguée du cyclopentadiène **9** est  $\Delta E_{n_C \rightarrow \pi_{C=C}^*}^{(2)} = 75.72$  kcal/mol. Cette énergie est remarquablement importante à cause de la conjugaison de la charge négative avec les doubles liaisons C=C de la molécule de cyclopentadiène. Ceci explique la faible valeur de  $pK_a = 15$  en la comparant avec les alcanes **1** et **2** ou les autres composés insaturés **4** et **5**.

Dans les composés dicarbonylés **10-17** (voir le **Tableau 2**), la charge négative des carbanions est conjuguée avec deux liaisons adjacentes  $\pi_{C=C}$  et par conséquent, ces anions sont plus stabilisés par les effets de délocalisation. Nous notons que ces effets stabilisateurs sont absents dans les composés **1-6**. Pour les nitro- et les cyano-alcanes **19-21**, la contribution des effets de résonance et des effets inductifs est très importante et la synergie de ces deux effets augmente l'acidité de ces composés. L'énergie de la stabilisation  $\Delta E_{n_C \rightarrow \pi_{N=O}^*}^{(2)}$  des formes anioniques des acides **20** et **21** est très importante en les comparant aux valeurs précédentes : 317.35 kcal/mol (tri-nitro méthane **20**) et 345.02 kcal/mol (cyano, di-nitro méthane **21**). L'acidité élevée de ces composés peut être expliquée par la forte délocalisation de la charge négative portée par le carbone acide sur tous les fragments de la molécule.

**Tableau 2:** Les énergies de transfert de charge  $E^{(2)}$  des bases conjuguées 9-16, 20,21

Composé	$\epsilon_n - \epsilon_{\pi^*}$ (au)	$F_{n \rightarrow \pi^*}$ (au)	$\Delta E^{(2)} n_c \rightarrow \pi^*$ kcal/mol
9	0.13	0.095	75.72
10	0.33	0.208	147.34
11	0.16	0.118	98.43
12	0.14	0.122	110.98
13	0.13	0.120	120.19
14	0.13	0.121	115.18
15	0.39	0.208	124.43
16	0.32	0.215	161.84
20	0.04	0.108	317.35
21	0.04	0.101	345.01

#### 4. Conclusion

Dans cette application, nous avons élaboré des corrélations linéaires satisfaisantes entre les constantes d'acidité ( $pK_a$ ) expérimentales aqueuses d'une variété d'acides carbonés et les énergies libres de déprotonation calculées en phase gazeuse et en phase aqueuse. La corrélation linéaire obtenue en phase gazeuse est légèrement meilleure ( $R^2=0.969$ ,  $SD=3.27$ ) en la comparant avec celle obtenue en phase aqueuse ( $R^2=0.934$ ,  $SD=4.80$ ).

La différence d'acidité des composés étudiés peut être interprétée par les effets inductifs et les effets de résonance. L'acidité élevée de certains acides carbonés a été rationalisée en calculant l'énergie de transfert de charge  $n_c \rightarrow \pi^*$  entre l'orbitale non-liante localisée sur l'atome de carbone de la base conjuguée (la charge négative  $n_c$ ) et l'orbitale antiliante vacante adjacente  $\pi^*$  des groupements C=C, C=O ou N=O.

**Bibliographie**

- [1] M. B. Smith, J. March, "March's Advanced Organic Chemistry", Sixth ed., John Wiley, New York, 2007, pp. 359–364.
- [2] M. J. S. Dewar, K.M. Dieter, *J. Am. Chem. Soc.* 108 (1986) 8075.
- [3] M. R. Siggel, T. D. Thomas, L. J. Saethre, *J. Am. Chem. Soc.* 110 (1988) 91.
- [4] S. R. Kass, *J. Comp. Chem.* 11 (1990) 94.
- [5] P. Burk, I. A. Koppel, *Theor. Chim. Acta* 86 (1993) 417.
- [6] P. Burk, K. Herodes, I. Koppel, I. A. Koppel, *Int. J. Quantum. Chem.* 27 (1993) 633.
- [7] I. A. Koppel, R. W. Taft, F. Anvia, S. Z. Zhu, L. Q. Hu, K.S. Sun, D.D. DesMarteau, L. M. Yagupolskii, Yu. L. Yagupolskii, N. V. Ignat'ev, N. V. Kondratenko, A. Yu. Volkonskii, V. M. Vlassov, R. Notario, P. C. Maria, *J. Am. Chem. Soc.* 116 (1994) 3047.
- [8] P. Burk, I. A. Koppel, L. M. Yagupolskii, R. W. Taft, *J. Comp. Chem.* 17 (1996) 30.
- [9] B. J. Smith, L. Radom, *J. Phys. Chem.* 99 (1995) 6468.
- [10] B. J. Smith, L. Radom, *J. Phys. Chem. Lett.* 245 (1995) 123.
- [11] J. Catalan, J. Palomar, *Chem. Phys. Lett.* 293 (1998) 511.
- [12] K. Choho, G. Van Lier, G., Van de Woude, P. Geerlings, *J. Chem. Soc. Perkin 2* (1996) 1723.
- [13] W. Yang, W. J. Mortier, *J. Am. Chem. Soc.* 108 (1986) 5708
- [14] R. Contreras, P. Fuentealba, M. Galvan, P. Perez, *Chem. Phys. Lett.* 304 (1999) 405
- [15] P. Pérez, Y. Simon-Manso, A. Aizman, P. Fuentealba, R. Contreras, *J. Am. Chem. Soc.* 122 (2000) 4756.
- [16] F. D'Souza, M. E. Zandler, G. R. Deviprasad, W. Kutner, *J. Phys. Chem.* 104 (2000) 6887.
- [17] M. C. Rezende, *J. Br. Chem. Soc.* 12 (2001) 73.
- [18] (a) I. Alkorta, J. Elguero, *Tetrahedron* 53 (1997) 9741.  
(b) M. C. Rezende, *Tetrahedron* 57 (2001) 5923.

- [19] E. P. Serjeant, B. Dempsey, "Ionization Constants of Organic Acids in Aqueous Solution" IUPAC Chemical Data Series, Pergamon Press, New York, 1979.
- [20] R. G. Pearson, R. L. Dillon, *J. Am. Chem. Soc.* 75 (1953) 2439
- [21] F. G. Bordwell, *Acc. Chem. Res.* 21 (1988) 456.
- [22] A. E. Reed, L. A. Curtiss, F. Weinhold, *Chem. Rev.* 88 (1988) 899.
- [23] A. E. Reed, R. B. Weinstock, F. Weinhold, *J. Chem. Phys.* 83 (1985) 735
- [24] T. K. Brunck,; F. Weinhold, *J., Am. Chem. Soc.* 101 (1979) 1700.
- [25] A. E. Reed, F. Weinhold, *J. Chem. Phys.* 83 (1985) 1736.
- [26] (a) A. E. Reed, F. Weinhold, *J. Chem. Phys.* 78 (1983) 4066.  
(b) A.E. Reed, F. Weinhold, *Isr. J. Chem.* 31 (1991) 277.
- [27] M. J. Frisch, G. W. Camions, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr, R. E. Stratmann, J. C. Burant, S. B Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D.Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, L. Komaromi, R. Gomperts, R. L. Martin, D. J. Renard, T. Keith, M. A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, J. A. Pople, GAUSSIAN 98W, Revision A.9, Gaussian, Inc., Pittsburgh, PA, 1998.
- [28] M. D. Liptak, K. C. Grass, P. G. Seybold, *J. Am. Chem. Soc.* 124 (2002) 6421.
- [29] S. Miertus, E. Scrocco, J. Tomasi, *J. Chem. Phys.* 55 (1981) 117.
- [30] S. Miertus, J. Tomasi, *J. Chem. Phys.* 65 (1982) 239.
- [31] M. Cossi, V. Barone, R. Cammi, J. Tomasi, *J. Chem. Phys. Lett.* 255 (1996) 327.
- [32] T. N. Brown, N. Mora-Diez, *J. Phys. Chem. B* 110 (2006) 9270.

- [33] F. G. Bordwell, J. E. Bares, J. E. Bartness, G.J. McCollum, M. V. D. Puy, N. R. Vanier, W. S. Matthews, *J. Org. Chem.* 42 (1977) 321.
- [34] F. G. Bordwell, G. J. McCollum, *J. Org. Chem.* 41 (1976) 2391.
- [35] P. C Hiberty, C. P. Byrman, *J. Am. Chem. Soc.* 117 (1995) 9875.
- [36] K. B. Wiberg, J. Ochterski, A. Streitwieser, *J. Am. Chem. Soc.* 118 (1996) 8291.

## ANNEXE N°1

**Tableau S0:** Les valeurs calculées des enthalpies  $G^{\circ}_{\text{gas}}$  (in au) et  $\Delta G^{\circ}_{\text{solv}}$  (en kcal/mol) des acides carbonés étudiés et de leurs bases conjuguées.

Composé	Acide <sup>a</sup>		Base Conjuguée <sup>a</sup>	
	$G^{\circ}_{\text{gas}}$	$\Delta G^{\circ}_{\text{solv}}$	$G^{\circ}_{\text{gas}}$	$\Delta G^{\circ}_{\text{solv}}$
1	-79.804731	1.89	-79.140122	-65.46
2	-40.509087	1.82	-39.846900	-89.79
3	-117.874367	0.95	-117.218424	-69.91
4	-78.585627	1.21	-77.939959	-74.87
5	-271.540629	-4.32	-270.896168	-63.85
6	-338.350391	-1.38	-337.757650	-65.59
7	-116.622586	-1.41	-116.010507	-67.93
8	-77.348553	0.28	-76.750782	-74.99
9	-194.090793	-4.24	-193.529296	-68.70
10	-574.922676	-7.83	-574.376218	-63.41
11	-377.979146	-15.89	-377.426834	-69.74
12	-460.231960	-5.96	-459.693678	-59.79
13	-345.815257	-5.11	-345.269580	-55.47
14	-462.514793	-5.93	-461.986138	-54.63
15	-534.428074	-9.54	-533.905966	-56.40
16	-490.158613	-13.29	-489.644121	-55.13
17	-380.567396	-11.16	-380.048676	-58.32
18	-581.173389	-9.68	-580.683418	-48.73
19	-618.340241	-11.28	-617.845693	-51.17
20	-654.137920	-11.30	-653.664425	-44.28
21	-541.855993	-7.27	-541.388378	-44.98

<sup>a</sup> Les optimisations de géométrie et les calculs de fréquence en phase gazeuse sont réalisés en utilisant la méthode B3LYP/6-311++G(d,p). Les valeurs de  $\Delta G^{\circ}_{\text{solv}}$  sont obtenues en effectuant un calcul single-point PCM sur les géométries optimisées en phase gazeuse.

## Correlation of aqueous $pK_a$ values of carbon acids with theoretical descriptors: A DFT study

I.E. Charif <sup>a</sup>, S.M. Mekelleche <sup>a,\*</sup>, D. Villemin <sup>b</sup>, N. Mora-Diez <sup>c</sup>

<sup>a</sup> *Département de Chimie, Faculté des Sciences, Université A. Belkaid, B.P.119, Tlemcen 13000, Algeria*

<sup>b</sup> *ENSICAEN, Université de Caen, LCMT, UMR 6507 CNRS, 6, Bd Maréchal Juin, Caen 14050, France*

<sup>c</sup> *Department of Chemistry, Thompson Rivers University, Kamloops, BC, Canada V2C 5N3*

Received 17 January 2007; received in revised form 15 April 2007; accepted 27 April 2007

Available online 10 May 2007

### Abstract

Theoretical calculations are carried out to predict gas- and aqueous-phase acidities of a series of 21 carbon acids with  $pK_a$  values varying from  $-6.20$  to  $50$ . Acceptable linear correlations ( $R^2 > 0.93$ ,  $SD < 4$ ) are obtained between calculated deprotonation Gibbs free energy changes and experimental  $pK_a$  values (measured in water). Solvent effects are taken into account by means of the polarizable continuum model (PCM). Our calculations also show that the high acidity of  $\alpha$ -proton carbonyl compounds and electron-withdrawing substituted methanes can be related to the strong increase in the vicinal charge interactions  $n_c \rightarrow \pi^*$  in anion species. Calculations are performed at the B3LYP/6-311++G(d,p) level of theory.

© 2007 Elsevier B.V. All rights reserved.

**Keywords:** Carbon acidity; Aqueous  $pK_a$ ; Deprotonation energy; DFT calculations; NBO analysis

### 1. Introduction

The deprotonation energies of organic acids and the proton affinities of the corresponding conjugate bases are widely used for the prediction of gas-phase and aqueous-phase Brønsted acidities [1–8]. Strong acids have small values of deprotonation energy (*i.e.*, the release of the proton is easier) while strong bases have large values of proton affinity (*i.e.*, the binding to the proton is stronger). Several works on the prediction of the acidity of organic and inorganic acids can be found in the literature. For instance, Smith and Radom [9,10] have shown that the G2 and G2(MP2) methods provide excellent results for both deprotonation enthalpies and proton affinities of small molecules. Catalan and Palomar [11] have investigated gas-phase acidities of a number of species and have shown that calculations at the B3LYP method with 6-311+G(d) and 6-

311+G(3df,3pd) basis sets correlate well with the experimental data. Good correlations have been obtained between experimental  $pK_a$  values of a wide range of organic Brønsted acids and their calculated gas-phase deprotonation enthalpies [12]. Correlations between theoretical descriptors and gas-phase acid-base equilibrium constants of organic compounds have been reported for amines, alcohols and thiols [13–15]. An excellent correlation was obtained between the aqueous-phase acidity calculated with the HF/3-21G(d) method and experimental  $pK_a$  values for a series of nitrogen bases [16].

Recently, various theoretical descriptors were used to investigate their correlation with the carbon acidity of compounds having the  $CHR_1R_2R_3$  structure. A reasonable correlation was obtained between the deprotonation energies of the compounds calculated at the HF/3-21G and B3LYP/6-31G(d) levels of theory and their aqueous  $pK_a$  values [17,18]. The ability to predict acidity using a coherent, well-defined theoretical approach, without external approximation or fitting to experimental data, would be very useful to chemists. Thanks to the rapid development

\* Corresponding author.

E-mail address: sm\_mekelleche@mail.univ-tlemcen.dz (S.M. Mekelleche).

of computational chemistry, the acidity of small molecules in the gas phase can be calculated with equivalent or greater accuracy than can be obtained experimentally. However, the current situation is less satisfactory in solution, mostly due to the difficulty of calculating solvation energies with adequate accuracy.

The aim of this work is to establish a correlation model for determining the aqueous  $pK_a$  values of carbon acids at 298.15 K by working with a series of 21 compounds (Fig. 1) that exhibit  $pK_a$  values from  $-6.2$  to  $50.0$  (see Table 1) [18–21]. In addition, the second-order charge-transfer energies, calculated using the natural bond orbital (NBO) method [22–26], are used to explain the high acidity of the dicarbonyl compounds and the electron-withdrawing substituted methanes included in this study.

## 2. Computational details

### 2.1. Gas-phase acidity calculations

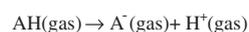
Calculations are performed with the Gaussian 98W software package [27]. The geometries of the neutral and deprotonated species are fully optimized at the B3LYP/6-

311++G(d,p) level of theory. Frequency calculations are performed at the same level of theory to characterize the stationary points obtained. The gas-phase Gibbs free-energy change ( $\Delta G_{\text{gas}}^{\circ}$ ) of Scheme 1 is calculated using Eq. (1). For  $G_{\text{gas}}^{\circ}(\text{H}^+)$  the experimental value of  $-6.28$  kcal/mol is used [28].

$$\Delta G_{\text{gas}}^{\circ} = G_{\text{gas}}^{\circ}(\text{A}^-) + G_{\text{gas}}^{\circ}(\text{H}^+) - G_{\text{gas}}^{\circ}(\text{AH}) \quad (1)$$

### 2.2. Aqueous-phase acidity calculations

Solvent effects are taken into account by means of the polarizable continuum model (PCM) [29–31] through single-point energy calculations at the B3LYP/6-311++G(d,p) level of theory (using the gas-phase optimized geometries). The PCM calculations, using Gaussian 98W, employ the UAHF atomic radii when constructing the solvent cavity for the calculation of the Gibbs free-energy of solvation. A common practice to calculate the



Scheme 1.

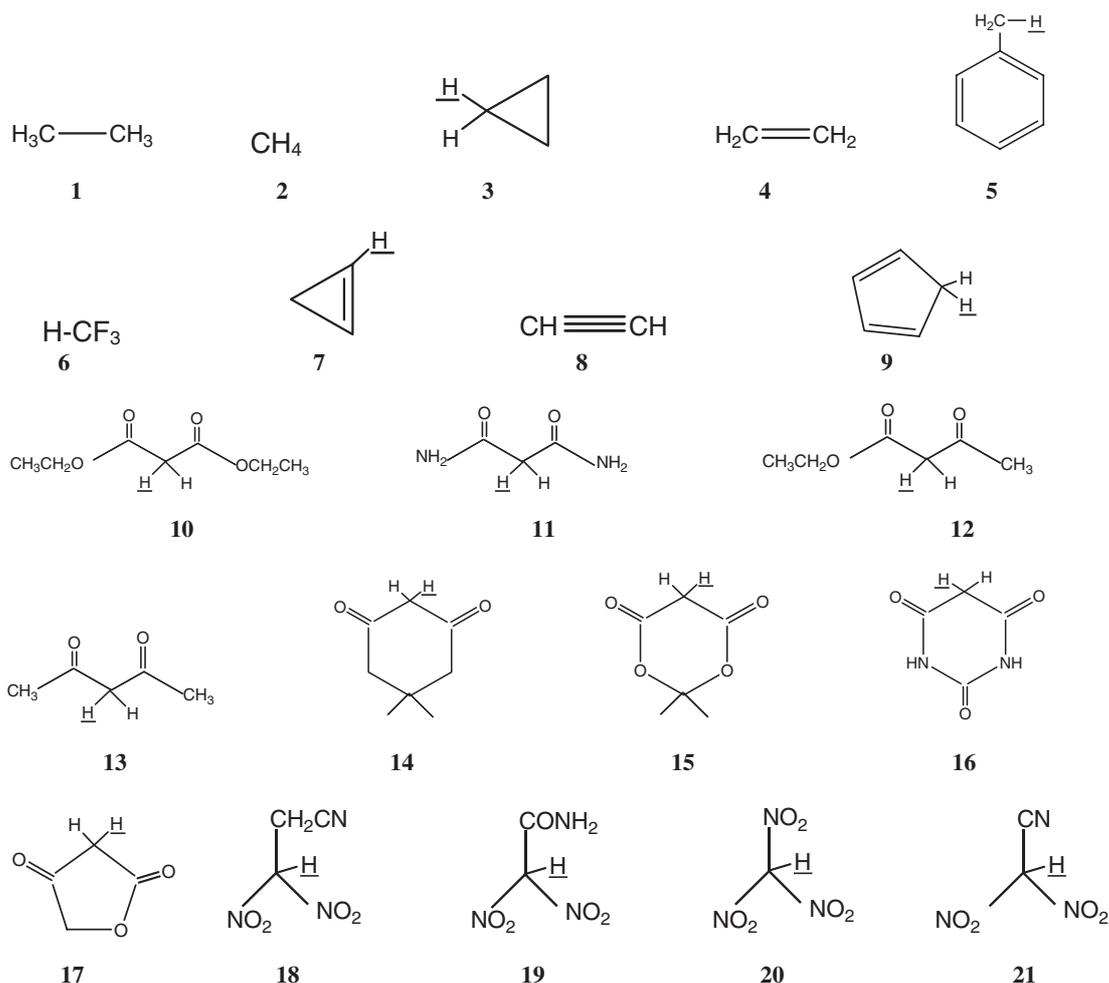


Fig. 1. Carbon acids considered in this study.

Table 1

Calculated gas- and aqueous-phase  $\Delta G^\circ$  values (in kcal/mol) of the acid dissociation and aqueous  $pK_a$  values of the compounds under study, together with the experimental aqueous  $pK_a$  values at 298.15 K

Compound number	Gas-phase		Aqueous-phase			
	$\Delta G^\circ_{\text{gas}}$	$pK_a(\text{calc})^a$	$\Delta G^\circ_{\text{aq}}$	$pK_a(\text{calc})^b$	$pK_a(\text{calc})^c$	$pK_a(\text{exp})$
1	410.77	46.91	78.78	57.75	48.80	50 <sup>d</sup>
2	409.25	46.24	53.56	39.26	32.62	48 <sup>d</sup>
3	405.33	44.49	69.86	51.21	43.08	46 <sup>d</sup>
4	398.98	41.66	58.28	42.72	35.65	44 <sup>d</sup>
5	398.12	41.28	73.87	54.15	45.65	41 <sup>e</sup>
6	365.67	26.83	36.84	27.01	21.89	32 <sup>e</sup>
7	377.80	32.23	46.66	34.20	28.19	29 <sup>d</sup>
8	368.83	28.24	28.95	21.22	16.82	25 <sup>d</sup>
9	346.06	18.10	16.99	12.45	9.15	15 <sup>e</sup>
10	336.62	13.90	16.38	12.01	8.76	13.30 <sup>f</sup>
11	340.31	15.54	21.85	16.02	12.27	12.50 <sup>f</sup>
12	331.50	11.62	13.06	9.57	6.63	10.70 <sup>f</sup>
13	336.15	13.69	21.17	15.52	11.83	9.00 <sup>f</sup>
14	325.46	8.93	12.14	8.90	6.03	5.25 <sup>e</sup>
15	321.34	7.09	9.87	7.24	4.58	4.83 <sup>e</sup>
16	316.57	4.97	10.12	7.42	4.74	4.01 <sup>e</sup>
17	319.21	6.14	7.88	5.78	3.30	3.76 <sup>e</sup>
18	301.18	-1.88	-2.47	-1.81	-3.34	2.27 <sup>f</sup>
19	304.05	-0.60	-0.45	-0.33	-2.05	1.30 <sup>f</sup>
20	290.84	-6.49	-6.74	-4.94	-6.08	0.06 <sup>f</sup>
21	287.15	-8.13	-15.17	-11.12	-11.49	-6.20 <sup>f</sup>

<sup>a</sup> Values calculated using the correlation equation between  $pK_a(\text{exp})$  and  $\Delta G^\circ_{\text{gas}}$  (Eq. (6), see Fig. 2).

<sup>b</sup> Values calculated using Eq. (5).

<sup>c</sup> Values calculated using the correlation equation between  $pK_a(\text{exp})$  and  $pK_a(\text{cal})$  (Eq. (7), see Fig. 4).

<sup>d</sup> Ref. [18a].

<sup>e</sup> Ref. [21].

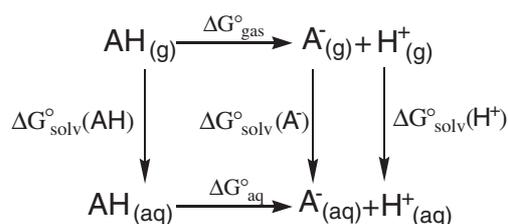
<sup>f</sup> Ref. [18b–20].

aqueous Gibbs free-energy change of an acid dissociation ( $\Delta G^\circ_{\text{aq}}$ , see Eq. (2)) is by combining the  $\Delta G^\circ_{\text{gas}}$  of the deprotonation process with the change in Gibbs free-energy of solvation ( $\Delta \Delta G^\circ_{\text{solv}}$ ), using the thermodynamic cycle of Scheme 2 and Eq. (3). For  $G^\circ_{\text{solv}}(\text{H}^+)$  the experimental value of  $-264.61$  kcal/mol is used [28]. The aqueous  $pK_a$  is calculated according to Eq. (5). All the calculations and experimental data reported in this paper are at 298.15 K.

$$\Delta G^\circ_{\text{aq}} = G^\circ_{\text{aq}}(\text{A}^-) + G^\circ_{\text{aq}}(\text{H}^+) - G^\circ_{\text{aq}}(\text{AH}) \quad (2)$$

$$\Delta G^\circ_{\text{aq}} = \Delta G^\circ_{\text{gas}} + \Delta \Delta G^\circ_{\text{solv}} \quad (3)$$

$$\begin{aligned} \Delta G^\circ_{\text{aq}} = & G^\circ_{\text{gas}}(\text{A}^-) + G^\circ_{\text{gas}}(\text{H}^+) - G^\circ_{\text{gas}}(\text{AH}) \\ & + \Delta G^\circ_{\text{solv}}(\text{A}^-) + \Delta G^\circ_{\text{solv}}(\text{H}^+) \\ & - \Delta G^\circ_{\text{solv}}(\text{AH}) \end{aligned} \quad (4)$$



Scheme 2.  $\Delta G^\circ_{\text{solv}}(\text{X})$  represents the Gibbs free-energy of solvation of species X.

$$pK_a = \frac{\Delta G^\circ_{\text{aq}}}{RT \ln 10} \quad (5)$$

### 3. Results and discussion

#### 3.1. Gas-phase acidities

The calculated gas-phase  $\Delta G^\circ$  of the acid dissociation and the experimental aqueous  $pK_a$  values (relative to water) at 298.15 K [18–21] are reported in Table 1. The raw data for these calculations are shown in Table S1 of the supplementary material. Fig. 2 shows the correlation of the  $\Delta G^\circ_{\text{gas}}$  values with the experimental aqueous  $pK_a$  values. A satisfactory linear correlation is obtained showing that the gas-phase Gibbs free-energy change of deprotonation generally increases when the acidity of these compounds decreases. For four of the 21 compounds studied (7, 11, 13 and 19) the calculated  $\Delta G^\circ_{\text{gas}}$  seems a bit higher than expected. The regression equation with the corresponding standard deviation (SD) and correlation coefficient ( $R^2$ ) values are given in Eq. (6).

$$\begin{aligned} pK_a = & 0.4453 \Delta G^\circ_{\text{gas}} - 135.99 \quad (R^2 = 0.969, \text{SD} \\ & = 3.27) \end{aligned} \quad (6)$$

If aqueous  $pK_a$  values are obtained from the calculated  $\Delta G^\circ_{\text{gas}}$  values using Eq. (6), the mean (MD) and mean absolute (AD) deviations between the experimental and

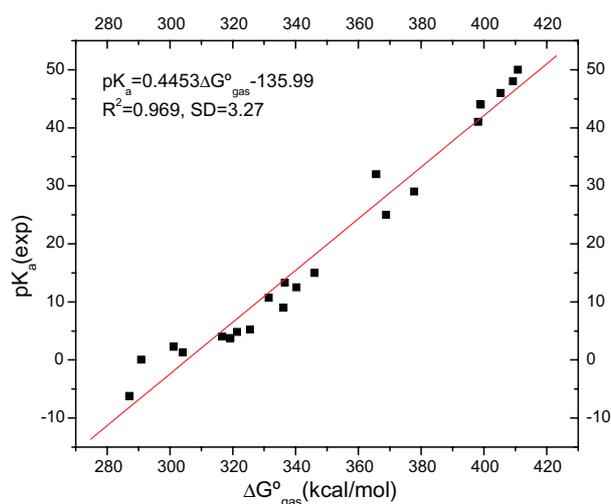


Fig. 2. Correlation of the experimental aqueous  $pK_a$  values of the carbon acids studied with the calculated (B3LYP/6-311++G(d,p)) gas-phase deprotonation Gibbs free-energy changes ( $\Delta G_{\text{gas}}^{\circ}$ ) at 298.15 K.

the calculated  $pK_a$  values are of  $-0.009$  and  $2.71$   $pK_a$  units, respectively. The largest AD values (greater than  $\pm 5$   $pK_a$  units) are detected for compounds **6** and **20**. Using the correlation Eq. (6) the predicted aqueous  $pK_a$  values for compounds **6** and **20** would be  $26.83$  (instead of  $32$ ) and  $-6.49$  (instead of  $0.06$ ), respectively. It should be noted that very accurate aqueous  $pK_a$  values (relative to water) can only be obtained for acids weaker than the hydronium ion ( $pK_a = -1.74$ ) and stronger than water ( $pK_a = 15.74$ ) [1]. The experimental aqueous  $pK_a$  values of the very strong and very weak carbon acids listed in Table 1 must be regarded as approximate.

It is well known that in the gas phase, where solvation effects are completely or almost completely absent, acidity orders may be much different than in aqueous solution for certain families of compounds. Simple alcohols are an example of this: while tertiary alcohols are more acidic than secondary and primary alcohols in the gas phase, in aqueous solution the acidity order is completely reversed because the much bulkier tertiary anion is poorly solvated. A similar pattern is found in simple aliphatic carboxylic acids and when analyzing the basicity order of amines [1]. From the results we have obtained, we can infer that the aqueous-phase acidity order of the family of carbon acids studied is very similar to that in the gas phase. Hence, it might be possible to use gas-phase calculations to predict the aqueous  $pK_a$  values of carbon acids. The predicted aqueous  $pK_a$  values for the carbon acids under study using Eq. (6) are listed in Table 1.

### 3.2. Aqueous-phase acidities

The calculated aqueous-phase  $\Delta G^{\circ}$  of the acid dissociation and  $pK_a$  values of the compounds under study (using Eq. (5)) at 298.15 K are reported in Table 1. The plot of  $\Delta G_{\text{aq}}^{\circ}$  versus  $\Delta G_{\text{gas}}^{\circ}$  is displayed in Fig. 3. A satisfactory

correlation is obtained between these two Gibbs free-energy changes ( $R^2 = 0.947$ ) that might help explain the good correlation previously found between  $\Delta G_{\text{gas}}^{\circ}$  and the experimental aqueous  $pK_a$  values. Fig. 4 shows the correlation between the experimental and calculated aqueous  $pK_a$  values. The regression equation with the corresponding standard deviation (SD) and correlation coefficient ( $R^2$ ) values are given in Eq. (7).

$$pK_a = 0.8755 pK_a(\text{calc}) - 1.7563 \quad (R^2 = 0.934, \text{SD} = 4.80) \quad (7)$$

The MD and AD between the experimental and the calculated values are of  $-0.64$  and  $4.39$   $pK_a$  units, respectively. The largest AD values (greater than  $\pm 5$   $pK_a$  units) are detected in compounds **1**, **2**, **3**, **5**, **7** and **13**. If Eq. (7) is used to predict what the aqueous  $pK_a$  of compounds **6** and **20** should be, values of  $21.89$  and  $-6.08$  are obtained. The predicted aqueous  $pK_a$  values for the carbon acids under study using Eq. (7) are listed in Table 1. The predicted values are smaller than those calculated with Eq. (5) which indicates that the calculated  $\Delta G_{\text{aq}}^{\circ}$  values might be slightly overestimated.

The aqueous phase  $pK_a$  calculation using gas-phase calculated data (by means of Eq. (6)) seems to be a more reliable approach for the family of organic acids considered since the AD obtained using the directly calculated aqueous  $pK_a$  is almost twice as large as in the first approach described. The reasons for this should lie in the deficiencies of continuum solvation models to accurately calculate solvation energies since explicit solute–solvent interactions are ignored. The inclusion of continuum solvent effects in the geometry optimizations and frequency calculations might improve these results slightly [32].

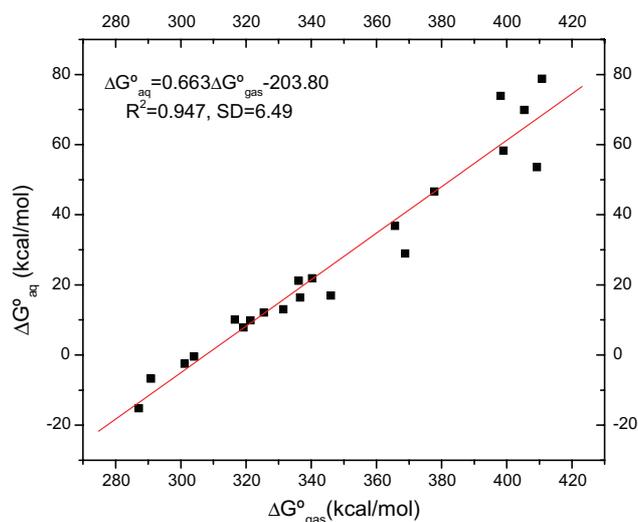


Fig. 3. Correlation between the calculated (B3LYP/6-311++G(d,p)) aqueous ( $\Delta G_{\text{aq}}^{\circ}$ ) and gas-phase ( $\Delta G_{\text{gas}}^{\circ}$ ) deprotonation Gibbs free-energy changes of the carbon acids studied at 298.15 K.

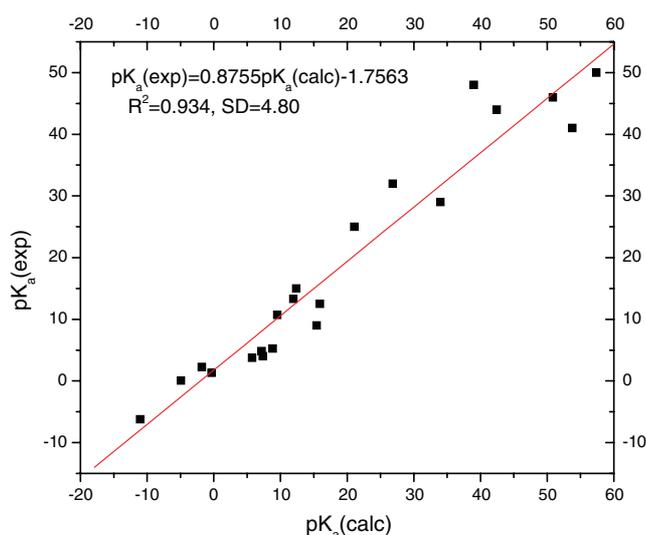


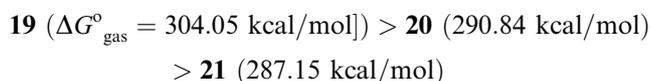
Fig. 4. Correlation between the experimental and calculated (B3LYP/6-311++G(d,p)-PCM(sp)) aqueous  $pK_a$  values of the carbon acids studied at 298.15 K.

The aqueous-phase acidity of the carbon acid series under investigation can be explained by the relative stabilities of their conjugate bases. The anions are stabilized by a combination of resonance and inductive effects [33]. Furthermore, it has been showed that delocalization is an important factor responsible for enhanced acidity [34–36]. However, inductive effects, produced by electron-withdrawing substituents, have the same order of magnitude as delocalization effects. Acidity increases when the delocalization of the negative charge (in the carbanion) increases [34]. The values of intrinsic acidity ( $\Delta G_{\text{gas}}^{\circ}$ ) of the compounds **1–21** decrease with increasing delocalization effects. The charge delocalization (depletion of the negative charge) can be explained by the donor–acceptor interaction between the non-bonding orbital localized on the carbon atom of the base with the negative charge ( $n_C$ ) and the vicinal unoccupied antibonding orbital  $\pi^*$  associated with the C=C, C=O or N=O groups. This molecular orbital interaction is called  $n_C \rightarrow \pi^*$  vicinal charge transfer. The stabilization energy of this interaction can be calculated by the NBO method using the second order perturbation theory as indicated by Eq. (8) [22–26], where  $\epsilon_n$ ,  $\epsilon_{\pi^*}$  are orbital energies and  $F_{n \rightarrow \pi^*}$  is the off-diagonal NBO Fock matrix element.

$$\Delta E_{n_C \rightarrow \pi^*}^{(2)} = -2(F_{n \rightarrow \pi^*})^2 / (\epsilon_n - \epsilon_{\pi^*}) \quad (8)$$

For instance, the charge transfer energies ( $\Delta E_{n_C \rightarrow \pi_{C=C}^*}^{(2)}$ ) in the conjugate bases of cyclopentadiene (**9**), ( $\Delta E_{n_C \rightarrow \pi_{C=O}^*}^{(2)}$ ) acetyl acetone (**13**) and barbituric acid (**16**) are 75.72, 120.19, and 161.84 kcal/mol, respectively (see Table S2 of the supplementary material). Therefore, the charge-transfer energy in these systems is remarkably important in explaining their acidity order. In the dicar-

bonyl compounds (**10–17**) the negative charge of the anions is conjugated with the two vicinal  $\pi$  bonds. Therefore, these anions are more stabilized by charge transfer delocalization. We note that these stabilizing effects are absent in compounds **1–6**. For nitro and cyano alkanes the following sequence of  $\Delta G_{\text{gas}}^{\circ}$  values is obtained.



In these compounds, the contribution of resonance and inductive effects are important and the synergy of the two effects originates a remarkably high acidity. The stabilization energy of the anionic forms ( $\Delta E_{n_C \rightarrow \pi_{N=O}^*}^{(2)}$ ), caused by the  $n_C \rightarrow \pi_{N=O}^*$  interaction, is very significant: 317.35 kcal/mol, for compound **20**, and 345.02 kcal/mol for compound **21**. Consequently, the negative charge on the acidic carbon of the anion is strongly delocalized over the whole molecule.

#### 4. Conclusions

In the present work, we have shown satisfactory linear correlations between experimental aqueous  $pK_a$  values of a variety of carbon acids and theoretical gas- and aqueous-phase deprotonation Gibbs free-energy changes. The linear correlation using the gas-phase descriptor is slightly better ( $R^2 = 0.969$ ,  $SD = 3.27$ ) than using the aqueous-phase one ( $R^2 = 0.934$ ,  $SD = 4.80$ ). The correlation equation between the experimental aqueous  $pK_a$  values and the calculated gas-phase deprotonation Gibbs free-energy changes was used to predict better  $pK_a$  values for two of the compounds studied ( $\text{HCF}_3$  and  $\text{HC}(\text{NO}_2)_3$ ). To assess the validity of the predictions in Table 1, further experimental work is required.

The difference in acidity of the compounds studied can be interpreted through inductive and resonance effects. The strong acidity of some carbon acids has been rationalized by calculating the  $n_C \rightarrow \pi^*$  charge transfer energy between the non-bonding orbital localized on the carbon atom of the base with the negative charge ( $n_C$ ) and the vicinal unoccupied antibonding orbital  $\pi^*$  associated with the C=C, C=O or N=O groups. These stabilization interaction energies are calculated by the NBO method using the second-order perturbation theory.

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.theochem.2007.04.037.

#### References

- [1] M.B. Smith, J. March, March's Advanced Organic Chemistry, Sixth ed., John Wiley, New York, 2007, pp. 359–364.
- [2] M.J.S. Dewar, K.M. Dieter, J. Am. Chem. Soc. 108 (1986) 8075.

- [3] M.R. Siggel, T.D. Thomas, L.J. Saethre, *J. Am. Chem. Soc.* 110 (1988) 91.
- [4] S.R. Kass, *J. Comp. Chem.* 11 (1990) 94.
- [5] P. Burk, I.A. Koppel, *Theor. Chim. Acta.* 86 (1993) 417.
- [6] P. Burk, K. Herodes, I. Koppel, I.A. Koppel, *Int. J. Quantum. Chem.* 27 (1993) 633.
- [7] I.A. Koppel, R.W. Taft, F. Anvia, S.Z. Zhu, L.Q. Hu, K.S. Sun, D.D. DesMarteau, L.M. Yagupolskii, Yu.L. Yagupolskii, N.V. Ignat'ev, N.V. Kondratenko, A.Yu. Volkonskii, V.M. Vlassov, R. Notario, P.C. Maria, *J. Am. Chem. Soc.* 116 (1994) 3047.
- [8] P. Burk, I.A. Koppel, I. Koppel, L.M. Yagupolskii, R.W. Taft, *J. Comp. Chem.* 17 (1996) 30.
- [9] B.J. Smith, L. Radom, *J. Phys. Chem.* 99 (1995) 6468.
- [10] B.J. Smith, L. Radom, *J. Phys. Chem. Lett.* 245 (1995) 123.
- [11] J. Catalan, J. Palomar, *Chem. Phys. Lett.* 293 (1998) 511.
- [12] K. Choho, G. Van Lier, G. Van de Woude, P. Geerlings, *J. Chem. Soc. Perkin 2* (1996) 1723.
- [13] W. Yang, W.J. Mortier, *J. Am. Chem. Soc.* 108 (1986) 5708.
- [14] R. Contreras, P. Fuentealba, M. Galvan, P. Perez, *Chem. Phys. Lett.* 304 (1999) 405.
- [15] P. Pérez, Y. Simon-Manso, A. Aizman, P. Fuentealba, R. Contreras, *J. Am. Chem. Soc.* 122 (2000) 4756.
- [16] F. D'Souza, M.E. Zandler, G.R. Deviprasad, W. Kutner, *J. Phys. Chem. A* 104 (2000) 6887.
- [17] M.C. Rezende, *J. Br. Chem. Soc.* 12 (2001) 73.
- [18] (a) I. Alkorta, J. Elguero, *Tetrahedron* 53 (1997) 9741;  
(b) M.C. Rezende, *Tetrahedron* 57 (2001) 5923.
- [19] E.P. Serjeant, B. Dempsey, *Ionization Constants of Organic Acids in Aqueous Solution* IUPAC Chemical Data Series, Pergamon Press, New York, 1979.
- [20] R.G. Pearson, R.L. Dillon, *J. Am. Chem. Soc.* 75 (1953) 2439.
- [21] F.G. Bordwell, *Acc. Chem. Res.* 21 (1988) 456–463.
- [22] A.E. Reed, L.A. Curtiss, F. Weinhold, *Chem. Rev.* 88 (1988) 899.
- [23] A.E. Reed, R.B. Weinstock, F. Weinhold, *J. Chem. Phys.* 83 (1985) 735.
- [24] T.K. Brunck, F. Weinhold, *J. Am. Chem. Soc.* 101 (1979) 1700.
- [25] A.E. Reed, F. Weinhold, *J. Chem. Phys.* 83 (1985) 1736.
- [26] (a) A.E. Reed, F. Weinhold, *J. Chem. Phys.* 78 (1983) 4066;  
(b) A.E. Reed, F. Weinhold, *Isr. J. Chem.* 31 (1991) 277.
- [27] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A. Montgomery, Jr, R.E. Stratmann, J.C. Burant, S.B. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, A.G. Baboul, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, L. Komaromi, R. Gomperts, 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, J.L. Andres, C. Gonzalez, M. Head-Gordon, E.S. Replogle, J.A. Pople, *GAUSSIAN 98W*, Revision A.9, Gaussian, Inc., Pittsburgh, PA, 1998.
- [28] M.D. Liptak, K.C. Grass, P.G. Seybold, S. Feldgus, G.C. Shields, *J. Am. Chem. Soc.* 124 (2002) 6421.
- [29] S. Miertus, E. Scrocco, J. Tomasi, *J. Chem. Phys.* 55 (1981) 117.
- [30] S. Miertus, J. Tomasi, *J. Chem. Phys.* 65 (1982) 239.
- [31] M. Cossi, V. Barone, R. Cammi, J. Tomasi, *J. Chem. Phys. Lett.* 255 (1996) 327.
- [32] T.N. Brown, N. Mora-Diez, *J. Phys. Chem. B* 110 (2006) 9270.
- [33] F.G. Bordwell, J.E. Bares, J.E. Bartness, G.J. McCollum, M.V.D. Puy, N.R. Vanier, W.S. Matthews, *J. Org. Chem.* 42 (1977) 321.
- [34] F.G. Bordwell, G.J. McCollum, *J. Org. Chem.* 41 (1976) 2391.
- [35] P.C. Hiberty, C.P. Byrman, *J. Am. Chem. Soc.* 117 (1995) 9875.
- [36] K.B. Wiberg, J. Ochterski, A. Streitwieser, *J. Am. Chem. Soc.* 118 (1996) 8291.

***PARTIE 2:***

*Prédiction des constantes d'acidité aqueuse ( $pK_a$ ) des acides carbonés.*

*Etude QSPR (Quantitative Structure-Property Relationships)*

**Résumé**

*Des modèles QSPR (Quantitative Structure-Property Relationships) entre les valeurs expérimentales aqueuses de  $pK_a$  de 21 acides carbonés et des descripteurs théoriques calculés par la méthode B3LYP/6-311++G(d,p) ont été élaborés. Les effets de solvant ont été considérés au moyen du modèle de solvation PCM sur des calculs d'énergie single point et dans les optimisations de la géométrie et les calculs de fréquence de vibration. Des régressions linéaires multiples ont été élaborées en utilisant plusieurs descripteurs théoriques: les charges NPA (Natural Population Analysis) de l'hydrogène acide et de l'atome de carbone portant la charge négative de la base conjuguée, les énergies des orbitales frontières, le volume de la cavité du solvant et l'énergie libre de Gibbs de déprotonation de l'acide. Des modèles QSPR satisfaisants ont été obtenus avec différentes combinaisons de descripteurs théoriques. Le pouvoir prédictif de ces modèles a été testé avec succès sur une série constituée de quatre acides carbonés non inclus dans la série d'apprentissage.*

**1. Introduction**

La constante d'acidité ( $K_a$ ,  $pK_a = -\log K_a$ ) des composés organiques joue un rôle fondamental dans beaucoup de procédures analytiques telles que les titrages acido-basiques, les extractions de solvants, formation des complexes et le transport d'ions. Il a été montré que la propriété acido-basique des acides et des bases organiques affecte la toxicité, le comportement chromatographique et les propriétés pharmaceutiques de ces composés [1,2].

L'une des approches les plus réussies pour prédire les propriétés chimiques à partir de la structure moléculaire est le développement des modèles QSPR (*Quantitative Structure-Property Relationships*). L'étape importante dans l'élaboration d'un modèle QSPR est de trouver le minimum de descripteurs moléculaires dans une famille de composés qui explique la propriété physique, chimique ou biologique [3-11]. A cet égard, différentes corrélations linéaires entre le pK<sub>a</sub> expérimental et un ou plusieurs descripteurs théoriques simples ont été précédemment étudiées avec succès dans la littérature [12-25].

La détermination théorique des valeurs de pK<sub>a</sub> **en solution** peut rendre, dans certains cas, les calculs théoriques plus compliqués, plus fastidieux et d'une précision limitée. Dans le cas des acides carboxyliques, des valeurs de pK<sub>a</sub> calculées en utilisant le modèle CPCM, ont été estimées avec une précision de moins de 0.5 unité par rapport aux valeurs de pK<sub>a</sub> expérimentales [22]. Pour les imidazoles substitués [23], l'utilisation du modèle PCM au niveau B3LYP/6-311+G(d,p) a donné des valeurs à moins d'une unité de pK<sub>a</sub>. Pour les amines aliphatiques, acycliques et aromatiques [24], les valeurs calculées en utilisant les méthodes B3LYP/6-31G\*-SCI-PCM et B3LYP/aug-cc-pVDZ//B3LYP/6-31G\*-PCM étaient estimées à moins de 0.7 unité. On note qu'il est nécessaire de considérer différentes combinaisons de descripteurs théoriques dans l'établissement des modèles QSPR pour mieux décrire les facteurs électrostatiques et électroniques qui déterminent la constante d'acidité d'un système chimique et améliorer la précision de la prédiction de cette propriété [10].

Plusieurs études ont été concentrées sur le développement des modèles QSPR avec des descripteurs simples pour prédire les valeurs aqueuses de pK<sub>a</sub> des composés organiques [7-9,26,27]. Soriano et col. [25] ont développé un modèle QSPR pour calculer la constante d'acidité aqueuse d'une série de quinze dérivés de l'imidazole en utilisant deux descripteurs moléculaires en phase gazeuse: la charge atomique sur l'hydrogène acide, l'énergie de l'orbitale moléculaire LUMO. Ils ont conclu que ces deux descripteurs sont fortement liés à l'acidité des espèces chimiques. Brown et Mora-Diez [10] ont développé un modèle QSPR

fiable pour estimer les  $pK_a$  aqueux des benzimidazoles protonés avec trois descripteurs moléculaires calculés au niveau B3LYP/6-31+G(d,p)-PCM (opt): la variation du volume de la cavité en allant de l'acide à la base conjuguée, la charge atomique NPA de l'hydrogène acide et l'énergie de l'orbitale LUMO+2 de l'acide.

Nous avons étudié dans l'application 1 [21b], les acidités en phase gazeuse et aqueuse de 21 acides carbonés avec les valeurs expérimentales aqueuses de  $pK_a$  variant de -6.20 à 50 [28-30]. Une corrélation linéaire acceptable a été obtenue ( $R^2=0.934$ ,  $SD=4.80$ ) entre la variation d'énergie libre aqueuse de déprotonation ( $\Delta G_{aq}$ ) calculée au niveau B3LYP/6-31++G(d, p)-PCM (*single point*) et les valeurs expérimentales aqueuses de  $pK_a$ . Cependant, la corrélation avec  $\Delta G_{gas}$  de déprotonation en phase gazeuse était meilleure ( $R^2=0.969$ ,  $SD=3.27$ ).

En utilisant la série d'acides carbonés (voir le **Tableau 1** et la **Figure 1**), un ensemble de descripteurs ont été calculés et leurs corrélations avec des valeurs aqueuses expérimentales de  $pK_a$  ont été élaborées et comparées dans la présente étude. Les charges atomiques NPA, les énergies des orbitales frontières, les volumes de cavité des acides et de leurs bases conjuguées, et l'énergie libre  $\Delta G_{aq}$  de déprotonation, ont été considérés pour élaborer les modèles QSPR. Des régressions linéaires multiples (MLR) sont effectuées pour corréler les différentes combinaisons de ces descripteurs avec les valeurs expérimentales de  $pK_a$ . Quatre acides carbonés supplémentaires (hors-série) ont été utilisés pour tester les pouvoirs prédictifs des modèles QSPR obtenus. Il est à fort intéressent de noter que les valeurs aqueuses expérimentales de  $pK_a$  (relatif à l'eau) ne peuvent être obtenues avec une grande précision que pour les acides plus faibles que l'ion de hydronium  $H_3O^+$  ( $pK_a = -1.74$ ) et plus forts que l'eau  $H_2O$  ( $pK_a = 15.74$ ) [28-30]. Par conséquent, les valeurs aqueuses expérimentales de  $pK_a$  des acides carbonés très forts ou très faibles, donnés dans le **Tableau 1**, doivent être considérées comme des valeurs approximatives.

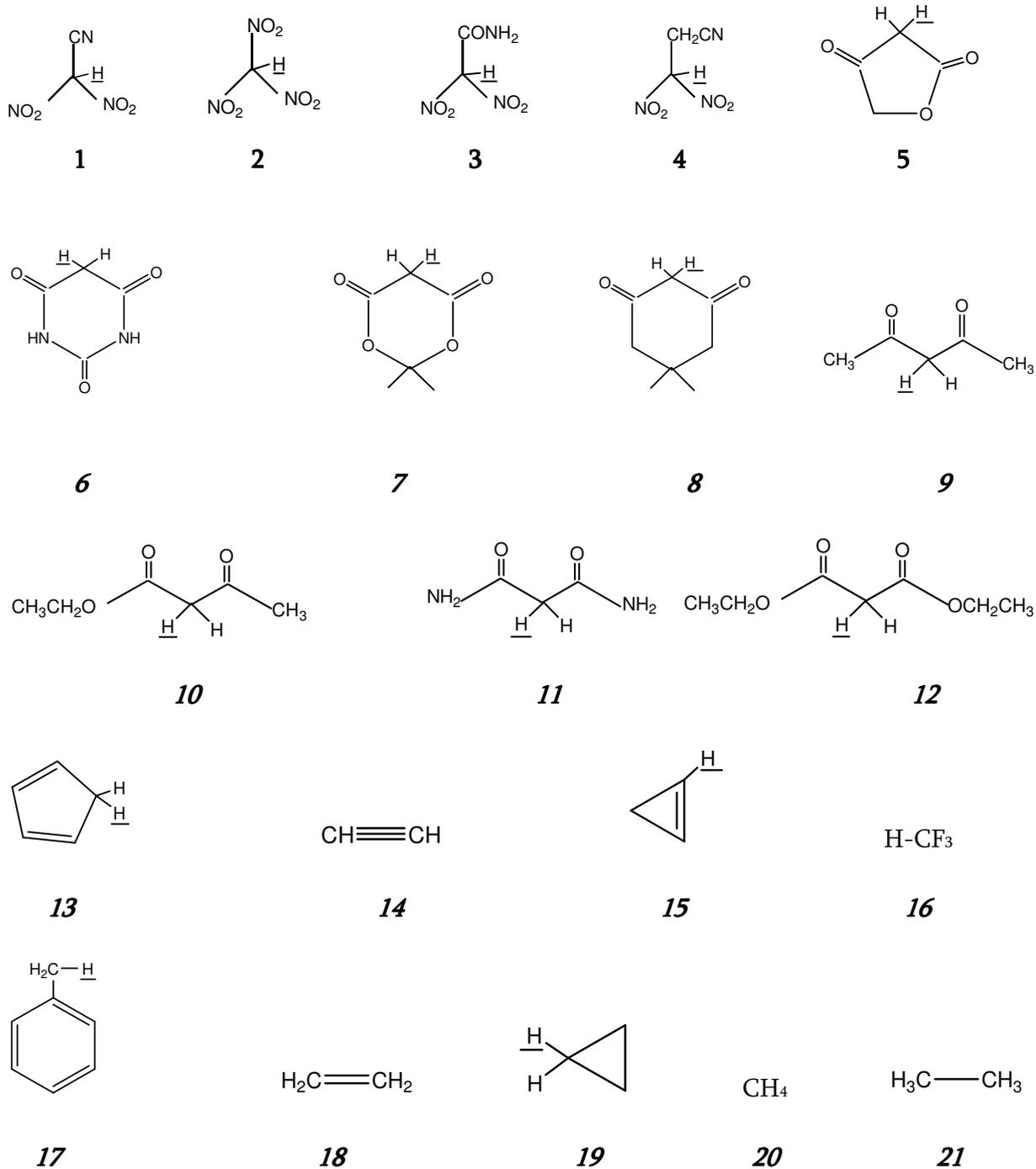


Figure 1: Structures des acides carbonés étudiés

**Tableau 1:** Les constantes d'acidité aqueuse expérimentales ( $pK_a$ ) à 298.15 K des acides carbonés.

	Carbon acid	$pK_a$
1	Dinitro-cyanomethane	-6.20 <sup>c</sup>
2	Trinitromethane	0.06 <sup>c</sup>
3	Dinitro-formamidylmethane	1.30 <sup>c</sup>
4	1,1-dinitro-2-cyanoethane	2.27 <sup>c</sup>
5	Tetronic acid	3.76 <sup>b</sup>
6	Babituric acid	4.01 <sup>b</sup>
7	Meldrum's acid	4.83 <sup>b</sup>
8	Dimedone	5.25 <sup>b</sup>
9	Acetylacetone	9.00 <sup>c</sup>
10	Ethylacetoacetate	10.70 <sup>c</sup>
11	Malonamide	12.50 <sup>c</sup>
12	Diethylmalonate	13.30 <sup>c</sup>
13	Cyclopentadiene	15 <sup>b</sup>
14	Acetylene	25 <sup>a</sup>
15	Cyclopropene	29 <sup>a</sup>
16	Trifluoromethane	32 <sup>b</sup>
17	Toluene	41 <sup>b</sup>
18	Ethylene	44 <sup>a</sup>
19	Cyclopropane	46 <sup>a</sup>
20	Methane	48 <sup>a</sup>
21	Ethane	50 <sup>a</sup>
22	Dimethylsulfoxide	32 <sup>d</sup>
23	Tetramic acid	6.4 <sup>e</sup>
24	Malonaldehyde	5 <sup>f</sup>
25	Dinitro-methane	3.57 <sup>f</sup>

<sup>a</sup> Ref. [28]

<sup>b</sup> Ref. [29]

<sup>c</sup> Ref. [28,29,30]

<sup>d</sup> Ref. [37a]

<sup>e</sup> Ref. [38]

<sup>f</sup> Ref. [37b]

## 2. Méthodologie des calculs

### 2.1. Calcul de l'énergie libre de déprotonation $\Delta G^{\circ}_{aq}$

L'énergie libre  $\Delta G^{\circ}_{aq}$  aqueuse de déprotonation d'un acide AH a été calculée à partir de l'énergie libre aqueuse des espèces impliquées ( $G^{\circ}_{aq}(X)$ ) en utilisant l'équation 2. Les valeurs de  $G^{\circ}_{aq}$  ont été obtenues à partir des énergies libres de Gibbs en phase gazeuse ( $G^{\circ}_{gas}$ ) et les énergies libres de solvation ( $\Delta G^{\circ}_{solv}$  au même état de référence que  $G^{\circ}_{gas}$ ) en utilisant l'équation 3. Les valeurs de  $G^{\circ}_{aq}$  ont été également calculées à partir de l'énergie totale aqueuse ( $E_{aq}$ ) et de la correction thermique en phase gazeuse de l'énergie libre de Gibbs ( $TCG_{gas}$ ) en utilisant l'équation 4.



$$pK_a = \frac{\Delta G^{\circ}_{aq}}{RT \ln 10} \quad (1)$$

$$\Delta G^{\circ}_{aq} = G^{\circ}_{aq}(A^{-}) + G^{\circ}_{aq}(H^{+}) - G^{\circ}_{aq}(AH) \quad (2)$$

$$G^{\circ}_{aq}(X) = G^{\circ}_{gas}(X) + \Delta G^{\circ}_{solv}(X) \quad (3)$$

$$G^{\circ}_{aq}(X) = E_{aq}(X) + TCG_{gas}(X) \quad (4)$$

Pour  $H^{+}$ , nous avons utilisé les valeurs expérimentales suivantes:  $G^{\circ}_{gas}(H^{+}) = -6.28$  kcal/mol et  $\Delta G^{\circ}_{solv}(H^{+}) = -264.61$  kcal/mol [31]. L'état standard en phase gazeuse (1 atm) doit être converti en solution (1 mole.  $dm^{-3}$ ), et ceci par l'utilisation la formule suivante:

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

à  $T=298.15$  K,  $Vm = 24.44$  L,  $R=1.987207$  cal/mol.K, et  $RT \ln(Vm)=1.8929329$  kcal/mol.

Les calculs des structures électroniques ont été réalisés à l'aide du programme G03 [32]. La méthode B3LYP/6-311++G(d,p) combinée au modèle de solvation PCM a été employée dans tous les calculs [23-35]. Le modèle PCM a été utilisé dans les calculs *single point* des énergies (PCM(sp)) et dans les optimisations des géométries et les calculs de fréquence (PCM(opt)). Les calculs PCM utilisent les rayons atomiques UAHF (*United Atom Topological Model applied on radii optimized for the HF/6-31G(d) level of theory*) dans la construction de la cavité du solvant. L'option « SCFVAC » a été employée pour obtenir l'énergie libre de solvation,  $\Delta G_{\text{solv}}$  des différentes espèces. Les descripteurs théoriques, en phase gazeuse, des acides et des bases conjuguées ont été également considérés dans la présente étude. Diverses régressions linéaires multiples ont été établies et analysées en utilisant le programme d'analyse statistique de données MINITAB [36].

Dans le cas des calculs PCM (sp), l'énergie libre aqueuse de Gibbs des espèces chimiques a été calculée en utilisant les équations 3 et 4. L'énergie de solvation  $\Delta G^{\circ}_{\text{solv}}$  dans l'équation 3 est calculée suivant deux manières: la première à partir d'un calcul *single point* d'énergie B3LYP/6-311++G (d, p)-PCM (eq. 3a) et la seconde à partir d'un calcul *single point* HF/6-31G (d)-PCM (eq. 3b). Les résultats des calculs de  $G_{\text{aq}}$  sont reportés dans le **Tableau S1** (voir *annexe n°2*).

### 3. Résultats et discussions

Dans les études QSPR, les descripteurs moléculaires ( $x_i$ ) sont corrélés avec une ou plusieurs variables de réponse ( $y$ ) et cette corrélation est bien représentée par un modèle linéaire qui peut être exprimé par:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots \quad (6)$$

Dans l'équation 6, les constantes  $b_i$  sont des inconnues appelées les coefficients de régression et l'objectif de l'analyse de la régression est de les estimer.

Pour déterminer quels sont les descripteurs moléculaires qui peuvent être utilisés dans l'étude QSPR, 13 descripteurs de l'acide et 12 descripteurs de sa base conjuguée ont été au départ considérés et leur corrélation avec les valeurs aqueuses expérimentales de  $pK_a$  ont été élaborées. Ces descripteurs sont

- 6 énergies des orbitales frontières ( $\epsilon_{\text{HOMO}-2}$  jusqu'à  $\epsilon_{\text{LUMO}+2}$ )
- 2 charges atomiques NBO (charge sur l'hydrogène acide,  $Q_H$ , et l'atome de carbone portant la charge négative  $Q_C$ )
- l'énergie libre aqueuse de Gibbs ( $G_{\text{aq}}$ )
- le volume de la cavité du solvant ( $V_{\text{sc}}$ ).
- l'électronégativité ( $\chi$ )

$$\chi = (\epsilon_{\text{HOMO}} + \epsilon_{\text{LUMO}})/2 \quad (7)$$

- la dureté ( $\eta$ )

$$\eta = (\epsilon_{\text{LUMO}} - \epsilon_{\text{HOMO}})/2 \quad (8)$$

- la mollesse ( $S$ )

$$S = \eta^{-1} \quad (9)$$

Des descripteurs supplémentaires ont été obtenus en combinant les descripteurs précédemment mentionnés. Ces descripteurs combinés sont les suivants:

- la somme et la différence entre l'énergie de LUMO de l'acide et la HOMO de la base ( $\epsilon_{\text{LUMO}}^{\text{acid}} + \epsilon_{\text{HOMO}}^{\text{base}}$  et  $\epsilon_{\text{LUMO}}^{\text{acid}} - \epsilon_{\text{HOMO}}^{\text{base}}$ , respectivement), le changement de la charge NBO sur l'atome de carbone de la base  $\Delta Q_C$ , la variation de l'électronégativité, de la dureté et de la mollesse ( $\Delta\chi$ ,  $\Delta\eta$  et  $\Delta S$  respectivement) entre l'acide et sa base conjuguée. L'énergie libre aqueuse de la dissociation de l'acide ( $\Delta G_{\text{aq}}$  calculée suivant l'équation 2) et la variation du volume de la cavité du solvant ( $\Delta V_{\text{sc}}$ ) ont été également considérées.

Les valeurs de tous les descripteurs considérés ainsi que les valeurs de  $R^2$  et les coefficients de leur corrélation linéaire avec les valeurs aqueuses expérimentales de  $pK_a$  sont regroupées dans les **Tableaux S2-S4** (*voir annexe n°2*).

Après la visualisation des orbitales moléculaires calculées au niveau B3LYP/6-311++G (d, p)-PCM (opt) (ou PCM (sp)), il s'avère que la paire électronique de la charge négative est localisée sur l'orbitale HOMO des bases conjuguées (voir **Schéma 1**). Pour cette raison, la corrélation des énergies de cette orbitale avec les valeurs expérimentales de  $pK_a$  est la plus remarquable ( $R^2 = 0.60$ ) en la comparant avec celle de la HOMO-2 ( $R^2=0.04$ ) et la HOMO-1 ( $R^2 =0.01$ ). Ceci est évident du fait que c'est l'orbitale la plus haute occupée de la base qui serait impliquée dans le processus de protonation.

Dans le cas des acides, l'énergie de l'orbitale LUMO possède la meilleure corrélation ( $R^2 = 0.59$ ) Comme il a été mentionné dans des travaux antérieurs [10,25], l'énergie LUMO de l'acide est reliée à la formation des liaisons hydrogène avec les molécules de solvant. Une valeur faible de  $\epsilon_{LUMO}^{acid}$  conduit à une formation facile des liaisons hydrogène avec les molécules d'eau permettant une déprotonation plus facile et par conséquent des valeurs faibles de  $pK_a$ .

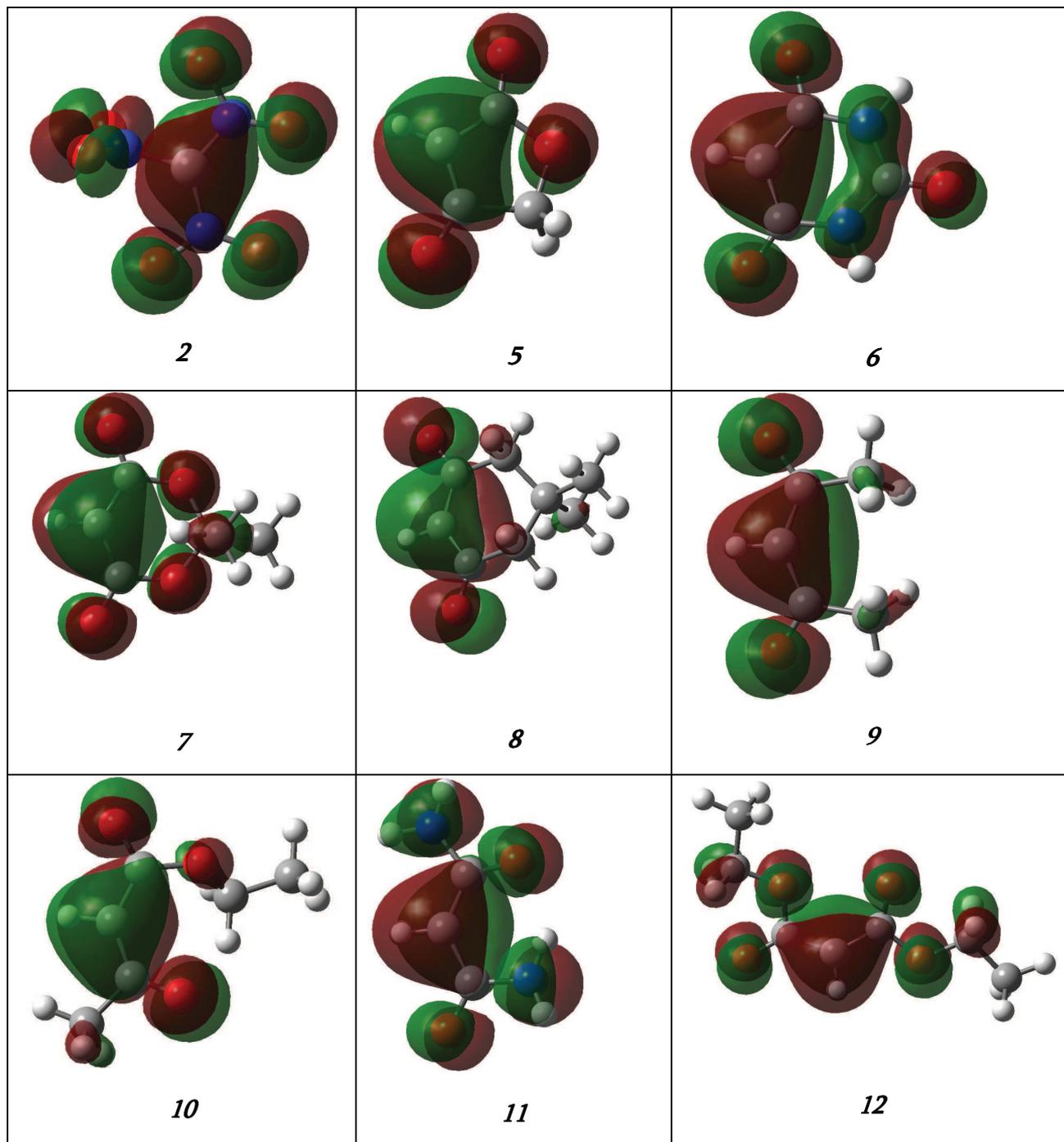


Schéma 1: Visualisation des orbitales moléculaires HOMO des bases conjuguées 2, 5-12.

Dans le cas où le modèle PCM est utilisé dans l'optimisation des géométries (PCM(opt)), les descripteurs ayant les corrélations les plus élevées ( $R_2 \geq 0.60$ ) avec les  $pK_a$  expérimentaux sont :  $\epsilon_{\text{HOMO}}^{\text{base}}$  (0.60),  $\Delta V_{\text{sc}}$  (0.63),  $Q_{\text{H}}^{\text{acid}}$  (0.81),  $\Delta\eta$  (0.66),  $\epsilon_{\text{LUMO}}^{\text{acid}} + \epsilon_{\text{HOMO}}^{\text{base}}$  (0.69),  $\Delta Q_{\text{C}}$  (0.85) et  $\Delta G_{\text{aq}}$  (0.97). Les descripteurs possédant les corrélations les plus élevées dans les calculs PCM(sp) sont:  $\Delta V_{\text{sc}}$  (0.54),  $\Delta\eta$  (0.67),  $Q_{\text{H}}^{\text{acid}}$  (0.81),  $\epsilon_{\text{LUMO}}^{\text{acid}} + \epsilon_{\text{HOMO}}^{\text{base}}$  (0.69),  $\Delta Q_{\text{C}}$  (0.83) et  $\Delta G_{\text{aq}}$  (0.96, eq 4). Il s'avère que les modèles de corrélation obtenus avec les deux niveaux de calculs (PCM(sp) et PCM(opt)) sont similaires. Les corrélations plus élevées ( $R_2 \geq 0.60$ ) des descripteurs en phase gazeuse sont:  $\epsilon_{\text{HOMO}}^{\text{base}}$  (0.88),  $Q_{\text{H}}^{\text{acid}}$  (0.63),  $\Delta\eta$  (0.83),  $\epsilon_{\text{LUMO}}^{\text{acid}} + \epsilon_{\text{HOMO}}^{\text{base}}$  (0.79),  $\Delta S$  (0.83),  $\Delta Q_{\text{C}}$  (0.77) et  $\Delta G_{\text{aq}}$  (0.95).

En raison des limitations dans les régressions linéaires multiples, le nombre de descripteurs considérés ne peut pas être plus grand que le nombre de points, ainsi il n'était pas possible de considérer simultanément tous les descripteurs pour la régression linéaire multiple des 21 acides. Pour éviter ce problème, des ensembles réduits de descripteurs sont considérés en utilisant la procédure "*Best Subsets Analysis*" du programme MINTAB. Le **Tableau S5** (voir annexe n°2) regroupe les meilleurs résultats de cette analyse pour des équations de corrélation à deux, trois et quatre variables en considérant trois méthodes de calculs. Une série de combinaisons a été effectuée et les meilleures équations QSPR obtenues sont présentées ci-dessous (**Tableau 2**).

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

<b>B3LYP/6-311++G(d,p)-PCM(opt)</b>	<b>R<sup>2</sup></b>	<b>SD</b>	<b>F-test</b>
(10) $pK_a = 1.0 + 0.525 \Delta G_{aq} - 17.4 \Delta Q_C$	0.982	2.5	499.5
(11) $pK_a = -0.722 + 0.480 \Delta G_{aq} + 0.985 \Delta S - 16.5 \Delta Q_C$	0.990	2.0	563.6
<b>B3LYP/6-311++G(d,p)-PCM(sp) (eq 3a)</b>			
(12) $pK_a = 1.18 + 0.513 \Delta G_{aq} - 18.1 \Delta Q_C$	0.980	2.7	438.8
(13) $pK_a = 0.126 + 0.464 \Delta G_{aq} + 0.809 \Delta S - 18.0 \Delta Q_C$	0.987	2.2	443.4
<b>B3LYP/6-311++G(d,p)-PCM(sp) (eq 3b, HF/6-31G(d))</b>			
(14) $pK_a = 1.74 + 0.519 \Delta G_{aq} - 21.8 \Delta Q_C$	0.982	2.6	493.0
(15) $pK_a = 0.83 + 0.473 \Delta G_{aq} - 0.701 \Delta S - 21.7 \Delta Q_C$	0.988	2.2	449.3
<b>B3LYP/6-311++G(d,p)-PCM(sp) (eq 4)</b>			
(16) $pK_a = 0.79 + 0.521 \Delta G_{aq} - 17.5 \Delta Q_C$	0.980	2.7	450.5
(17) $pK_a = -0.186 + 0.472 \Delta G_{aq} + 0.785 \Delta S - 17.5 \Delta Q_C$	0.987	2.2	445.0
<b>B3LYP/6-311++G(d,p)</b>			
(18) $pK_a = -120 + 0.396 \Delta G_{gas} - 13.5 \Delta Q_C$	0.970	3.5	261.4
(19) $pK_a = -157 + 0.495 \Delta G_{gas} - 68.5 \epsilon_{LUMO}^{acid} - 8.91 \Delta Q_C$	0.980	2.9	250.9

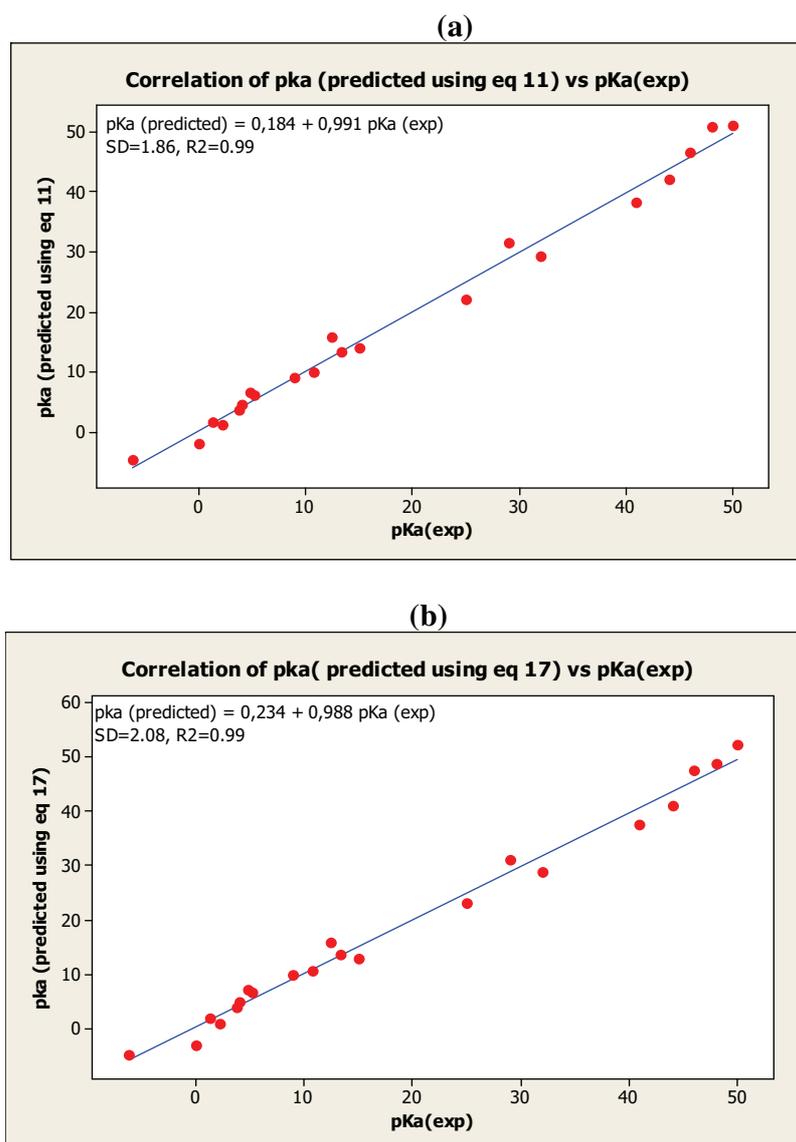
Ces modèles MLR présentent

- les valeurs des coefficients R<sup>2</sup> les plus élevées
- les F-test les plus élevées
- les déviations standards SD les plus faibles

En phase aqueuse, avec les calculs B3LYP/6-311++G (d,p)-PCM(opt), la meilleure corrélation à 2 variables inclut les descripteurs  $\Delta Q_C$  et  $\Delta G_{aq}$  avec la plus grande valeur de F-test ( $R^2=0.982$ ,  $SD=2.5$ ,  $F=499.5$ , équation 10) et le meilleur modèle à 3 variables comporte les descripteurs  $\Delta G_{aq}$ ,  $\Delta S$  et  $\Delta Q_C$  ( $R^2 = 0.99$ ,  $SD=2.0$ ,  $F= 563.6$ , équation 11). Ce dernier modèle QSPR présente les meilleurs paramètres statistiques de notre étude MLR.

En phase aqueuse, avec les calculs B3LYP/6-311++G (d,p)-PCM(sp, Eq. 4), les descripteurs  $\Delta G_{aq}$  et  $\Delta Q_C$  forme la meilleure combinaison à 2 variables ( $R^2=0.980$ ,  $SD=2.7$ ,  $F= 450.5$ , équation 16). L'ajout d'un troisième descripteur ( $\Delta S$ ) améliore légèrement la déviation standard ( $R^2 = 0.987$ ,  $SD=2.2$ ,  $F=445.0$ , équation 17).

En phase gazeuse, avec les calculs B3LYP/6-311++G (d,p), la meilleure corrélation à 2 descripteurs fait intervenir les descripteurs  $\Delta G_{aq}$  et  $\Delta Q_C$  ( $R^2=0.970$ ,  $SD=3.5$ ,  $F=261.4$ , équation 18) et la meilleure corrélation à 3 variables comporte les descripteurs  $G_{aq}$ ,  $\epsilon_{LUMO}^{acid}$  et  $\Delta Q_C$  ( $R^2=0.980$ ,  $SD=2.9$ ,  $F=250.9$ , équation 19). La **Figure 2** illustre de la corrélation linéaire entre les  $pK_a$  expérimentales aqueuses et les valeurs de  $pK_a$  calculées en utilisant les équations 11 et 17.



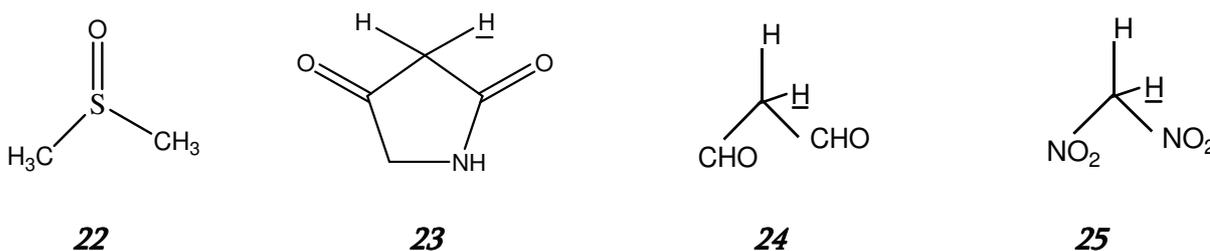
**Figure 2.** Corrélation des valeurs de  $pK_a$  calculées avec les valeurs expérimentales aqueuses. (a) B3LYP/6-311++G(d,p) PCM(opt); (b) B3LYP/6-311++G(d,p) PCM(sp).

Les valeurs de  $R^2$  des calculs PCM(opt) sont légèrement meilleurs que celles des calculs PCM(sp) avec le même nombre (2 ou 3) de descripteurs. L'augmentation du nombre de descripteurs (plus que 3) n'améliore pas de manière significative la valeur de  $R^2$ . Les corrélations des calculs PCM(sp) utilisant les valeurs de  $\Delta G_{\text{aq}}$  calculées à partir des équations 3 et 4 sont très semblables.

On note que l'équation 4 (utilisant  $E_{\text{aq}}$  à la place de  $\Delta G_{\text{solv}}$ ) semble être une meilleure approche pour estimer les valeurs de  $G_{\text{aq}}$  en effectuant un calcul PCM *single point* sur des géométries optimisées en phase gazeuse.

### 3.1 Prédiction des valeurs aqueuses de $\text{pK}_a$ des acides carbonés d'une série de test.

Les valeurs aqueuses de  $\text{pK}_a$  de 4 acides carbonés [37-38] non inclus dans la série d'apprentissage (21 acides) ont été estimées en utilisant les modèles QSPR élaborés précédemment (Tableau 2). Ces composés sont numérotés de 22 à 25 (voir le Tableau 1) et leurs structures sont schématisées dans la Figure 3. Les valeurs de  $\text{pK}_a$  estimées des ces acides sont regroupées dans le Tableau 3.



**Figure 3.** Structures des acides carbonés de la série test.

**Tableau 3:** Prédiction des  $pK_a$  des acides carbonés **22-25** en utilisant les équations MLR 10-19.

	Carbon Acid	<b>22</b>	<b>23</b>	<b>24</b>	<b>25</b>	AD <sup>a</sup> (MD) <sup>b</sup>
Level of theory	$pK_a$ (exp) [37,38]	33	6.4	5	3.57	
B3LYP/6-311++G(d,p)-PCM(opt)	$\Delta G_{aq}, \Delta Q_C$ $\Delta G_{aq}, \Delta S, \Delta Q_C$	34.43 33.76	10.0 8.61	4.60 4.00	4.62 5.69	1.61 (1.42) 1.52 (1.02)
B3LYP/6-311++G(d,p)-PCM(sp) (eq 3a)	$\Delta G_{aq}, \Delta Q_C$ $\Delta G_{aq}, \Delta S, \Delta Q_C$	34.43 33.70	10.41 9.00	4.61 4.15	3.19 4.40	1.55 (1.17) 1.24 (0.82)
B3LYP/6-311++G(d,p)-PCM(sp) (eq 3b, HF/6-31G(d))	$\Delta G_{aq}, \Delta Q_C$ $\Delta G_{aq}, \Delta S, \Delta Q_C$	36.12 35.35	10.08 8.88	4.03 3.70	4.06 5.03	2.06 (1.58) 1.89 (1.24)
B3LYP/6-311++G(d,p)-PCM(sp) (eq 4)	$\Delta G_{aq}, \Delta Q_C$ $\Delta G_{aq}, \Delta S, \Delta Q_C$	34.88 34.13	9.99 8.68	4.45 4.03	3.59 4.73	1.50 (1.24) 1.37 (0.90)
B3LYP/6-311++G(d,p)	$\Delta G_{gas}, \Delta Q_C$ $\Delta G_{gas}, \epsilon_{LUMO}^{acid}, \Delta Q_C$	30.66 29.67	10.53 10.48	7.15 7.48	1.42 3.87	2.70 (0.43) 2.50 (0.88)

<sup>a</sup> AD: Mean absolute deviation<sup>b</sup> MD: Mean deviation

Il s'avère que les meilleures prédictions sont obtenues avec les calculs PCM(opt) PCM(sp, eq4) pour les modèles MLR à 2 ( $\Delta G_{aq}, \Delta Q_C$ ) et à 3 ( $\Delta G_{aq}, \Delta Q_C, \Delta S$ ) descripteurs. Les modèles MLR-PCM(opt), équations 10 et 11, conduisent à des meilleures estimations de  $pK_a$  aqueuses avec des déviations absolues AD acceptables (1.61 et 1.52 respectivement) et des déviations moyennes MD (1.42 et 1.02, respectivement). Les modèles MLR-PCM(sp, eq4), équations 16 et 17, conduisent à des meilleures estimations de  $pK_a$  aqueuses avec les faibles déviations absolues AD (1.50 et 1.37 respectivement) et des déviations moyennes MD (1.24 et 0.90, respectivement). Les modèles MLR établis, à partir des descripteurs en phase gazeuse, possèdent les valeurs de  $R^2$  les plus élevées et des faibles déviations moyennes mais des déviations absolues moyennes plus élevées.

Les valeurs de  $pK_a$  calculées avec les modèles QSPR de type PCM(opt) et PCM(sp) ne sont pas trop différentes des valeurs expérimentales pour les composés 22, 24 et 25. Cependant, toutes les prédictions considérées indiquent une valeur excessive du  $pK_a$  aqueux du composé 23. A partir de ces résultats on suggère que cette valeur expérimentale soit remesurée.

#### 4. Conclusion

Dans cette application, plusieurs modèles QSPR faisant intervenir des descripteurs quantiques des acides et de leurs bases conjuguées ont été élaborés pour prédire les valeurs expérimentales de  $pK_a$  aqueuses des acides carbonés. Quatre composés non inclus dans la série d'étude ont été utilisés pour tester le pouvoir prédictif des meilleurs modèles QSPR obtenus. Dans le cas des calculs B3LYP/6-311++G(d,p)-PCM(opt), le modèle QSPR le plus simple comporte l'énergie libre aqueuse de déprotonation ( $\Delta G_{aq}$ ) et la variation de la charge NBO sur le carbone basique de la base conjuguée ( $\Delta Q_C$ ) avec une estimation de l'ordre de 1.61 unité de  $pK_a$ . En ajoutant un troisième descripteur  $\Delta S$  (la variation de la mollesse en allant de l'acide à sa base conjuguée) au modèle précédent, on remarque une légère amélioration de la prédiction estimée à 1.52 unité  $pK_a$  par rapport aux valeurs expérimentales. Dans le cas des calculs B3LYP/6-311++G(d,p)-PCM(sp) en utilisant l'énergie  $E_{aq}$  dans le calcul de  $\Delta G_{aq}$  (l'équation 4), le modèle QSPR le plus simple comporte l'énergie libre aqueuse de déprotonation ( $\Delta G_{aq}$ ) et la variation de la charge NBO ( $\Delta Q_C$ ) avec une estimation de l'ordre de 1.50 unité de  $pK_a$ . En ajoutant un troisième descripteur  $\Delta S$  au modèle précédent, on remarque une amélioration de la prédiction estimée à 1.37 unité  $pK_a$  par rapport aux valeurs expérimentales. Les meilleures équations QSPR trouvées pourraient être utiles pour prédire les valeurs aqueuses de  $pK_a$  des acides carbonés très forts ( $pK_a < pK_a(H_3O^+) = -1.74$ ) et très faibles ( $pK_a > pK_a(H_2O) = 15.74$ ) pour lesquels il est difficile d'obtenir des données expérimentales suffisamment précises.

## Bibliographie

- [1] C. Hansh, P. G. Sammes, J. B. Taylor, Eds. "Comprehensive Medicinal Chemistry. Quantitative Drug Design"; Pergamon Press; Oxford, UK., Vol. 4, 1990.
- [2] D.D. Perrin, B. Dempsey, E. P. Serjeant, "pK<sub>a</sub> Prediction for Organic Acids and Bases", Chapman and Hall; New York, 1981.
- [3] M. Ravi, A. J. Hopfinger, R. E. Hormann, L. Dinan, *J. Chem. Inf. Comput. Sci.* 41 (2001) 1587.
- [4] B. T. Luke, *J. Mol. Struct. (THEOCHEM)* 13 (1999) 468.
- [5] P. Bruneau, *J. Chem. Inf. Comput. Sci.* 41 (2001) 1605.
- [6] A. R. Katritzky, R. Petrukhin, D. Tatham, *J. Chem. Inf. Comput. Sci.* 41 (2001) 679.
- [7] V. Consonni, R. Todeschini, M. Pavan, P. Gramatica, *J. Chem. Inf. Comput. Sci.* 42 (2002) 693.
- [8] G. Krenkel, E. A. Castro, A. A. Toropov, *J. Mol. Struct. (THEOCHEM)* 542 (2001) 107.
- [9] (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.
- [10] T. N. Brown, N. Mora-Diez, *J. Phys. Chem. B* 110 (2006) 20546.
- [11] B. Kier, H. Hall, "Molecular Structure Description: The Electrotopological State" 1999, Academic Press, San Diego; J. Devillers, Balaban (Eds.), "Topological Indices and Related Descriptors in QSAR and QSPR", Gordon and Breach Science Publishers, Amsterdam 1999.
- [12] K. C. Gross, P. G. Seybold, Z. Peralta-Inga, J. S. Murray, P. Politzer, *J. Org. Chem.* 66 (2001) 6919.
- [13] P. Nagy, K. Novak, G. Szasz, *J. Mol. Struct. (THEOCHEM)* 30 (1989) 257.
- [14] M. Haeberlein, J. S. Murray, T. Brinck, P. Politzer, *Can. J. Chem.* 70 (1992) 2209.
- [15] H. B. Broughton, S. M. Green, H. S. Rzepa, *J. Chem. Soc., Perkin Trans. 2* (1995) 431.

- [16] J. Graton, M. Berthelot, C. Laurence, *J. Chem. Soc., Perkin Trans. 2* (2001) 2130.
- [17] K. G. Gross, P. G. Seybold, *Int. J. Quantum Chem.* 80 (2000) 1107.
- [18] G. R. Famini, C. A. Penski, L. Y. Wilson, *J. Phys. Org. Chem.* 5 (1992) 395.
- [19] P. Perez, Y. Simon-Manso, A. Aizman, P. Fuentealba, R. Contreras *J. Am. Chem. Soc.* 122 (2000) 4756.
- [20] F. D'Souza, M. E. Zandler, G. R. Deviprasad, W. Kutner, *J. Phys. Chem. A* 104 (2000) 6887.
- [21] (a) T. N. Brown, N. Mora-Diez, *J. Phys. Chem. B* 110 (2006) 9270.  
(b) I. E. Charif, S. M. Mekelleche, D. Villemin, N. Mora-Diez, *J. Mol. Struct. (THEOCHEM)* 818 (2007) 1.  
(c) G. Da Silva, E. M. Kennedy, Z. Bogdan, Dlugogorski, *J. Phys. Chem. A* 110 (2006) 11371
- [22] M. D. Liptak, G. C. Shields, *J. Am. Chem. Soc.* 123 (2001) 7314.
- [23] (a) I. A. Topol, G. J. Tawa, S. K. Burt, A. A. Rashin, *J. Phys. Chem. A* 101 (1997) 10075.  
(b) I. A. Topol, G. J. Tawa, R. A. Caldwell, M. A. Eissenstat, S. K. Burt, *J. Phys. Chem. A* 104 (2000) 9619.
- [24] B. Kallies, R. Mitzner, *J. Phys. Chem. B* 101 (1997) 2959.
- [25] E. Soriano, S. Cerdán, P. Ballesteros, *J. Mol. Struct. (THEOCHEM)* 684 (2004) 121.
- [26] J. Ghasemi, S. Shahmirani, E. V. Farahani, *Ann. Chim.* 96 (2006) 327.
- [27] J. Ghasemi, S. Saaidpour, S.D. Brown, *J. Mol. Struct.:(THEOCHEM)* 805 (2007) 27.
- [28] M. C. Rezende, *Tetrahedron* 57 (2001) 5923.
- [29] E. P. Serjeant, B. Dempsey "Ionization Constants of Organic Acids in Aqueous Solution", Pergamon Press, New York 1979.
- [30] (a) R. G. Pearson, R. L. Dillon, *J. Am. Chem. Soc.* 75 (1953) 2439.  
(b) M. B. Smith, J. March, "March's Advanced Organic Chemistry", Sixth ed., John Wiley, New York, 2007, pp. 359–364.
- [31] M. D. Liptak, K. C. Gross, P. G. Seybold, S. Feldgus, G. C. Shields, *J. Am. Chem.*

- Soc. 124 (2002) 6421.
- [32] 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, 2004.
- [33] S. Miertus, E. Scrocco, J. Tomasi, *J. Chem. Phys.* 55 (1981) 117.
- [34] S. Miertus, J. Tomasi, *J. Chem. Phys.* 65 (1982) 239.
- [35] M. Cossi, V. Barone, R. Cammi, J. Tomasi, *J. Chem. Phys. Lett.* 255 (1996) 327.
- [36] MINITAB State College, Minitab, Inc, PA 2006.
- [37] (a) see the website <http://www.chem.wisc.edu/areas/reich/pkatable/>  
(b) R. P. Bell, "The Proton in Chemistry", Ithica, Cornell UP, 1959.
- [38] G. V. Perez, A. L. Perez, *J. Chem. Ed.* 77 (2000) 910.

## ANNEXE N°2

**Table S1:**

Raw data (in atomic units, unless otherwise indicated) for the aqueous  $pK_a$  calculations (at 298.15 K) of the carbon acids under study at the B3LYP/6-311++G(d,p)-PCM//B3LYP/6-311++G(d,p) level of theory. The data provided corresponds to the neutral (Acid) and charged species (Base) according to the labels used in Table 1 and Figure S1.

Species	$G_{\text{gas}}$	$\Delta G_{\text{solv}}^a$ (kcal/mol)	$\Delta G_{\text{solv}}^b$ (kcal/mol)	$E_{\text{aq}}$	$TCC_{\text{gas}}$
Acid1	-541.85599	-10.75	-12.63	-541,89658	0,01511
Base1	-541.38838	-45.93	-46.97	-541,47178	0,00408
Acid2	-654.13792	-8.87	-10.5	-654,18001	0,01899
Base2	-653.66443	-43.53	-44.34	-653,74734	0,00549
Acid3	-618.34024	-12.83	-13.86	-618,41157	0,04277
Base3	-617.84569	-50.78	-52.16	-617,96439	0,03133
Acid4	-581.17339	-12.22	-13.71	-581,24378	0,04228
Base4	-580.68342	-48.69	-49.26	-580,79773	0,02995
Acid5	-380.56740	-13.55	-14.15	-380,63841	0,04743
Base5	-380.04868	-57.73	-59.38	-380,17920	0,03628
Acid6	-490.15861	-16.25	-17.14	-490,24397	0,05652
Base6	-489.64412	-56.63	-58.34	-489,78300	0,04567
Acid7	-534.42807	-10.87	-11.74	-534,55343	0,10415
Base7	-533.90597	-54.79	-56.53	-534,08908	0,09176
Acid8	-462.51479	-7.66	-7.84	-462,68338	0,15119
Base8	-461.98614	-53.69	-56.19	-462,21560	0,13902
Acid9	-345.81526	-5.52	-5.69	-345,91621	0,08740
Base9	-345.26958	-56.35	-59.19	-345,43918	0,07656
Acid10	-460.35714	-7.4	-7.92	-460,49234	0,118225
Base10	-459.81773	-54.29	-56.43	-460,01339	0,104976
Acid11	-377.97915	-16.37	-16.66	-378,07675	0,06920
Base11	-377.42683	-64.23	-66.68	-377,58921	0,05785
Acid12	-574.92268	-6.28	-6.44	-575,08623	0,14740
Base12	-574.37622	-54.77	-56.54	-574,60127	0,13335
Acid13	-194.09079	-3.31	-2.77	-194,16218	0,06551
Base13	-193.52930	-54.97	-57.37	-193,66992	0,05153
Acid14	-77.34855	0.2	0.08	-77,360604	0,00810
Base14	-76.75078	-72.4	-78.54	-76,863645	-0,00490
Acid15	-116.62259	-1.25	-1.1	-116,65793	0,03176
Base15	-116.01051	-64.28	-66.57	-116,13242	0,01763
Acid16	-338.35039	-1.37	-1.4	-338,35328	-0,00118
Base16	-337.75765	-62.72	-65.97	-337,84238	-0,01745
Acid17	-271.54063	-1.06	-1.16	-271,64293	0,09823
Base17	-270.936151	-48.47	-50.53	-271,098263	0,081802
Acid18	-78.58563	1.39	1.45	-78,616738	0,02991
Base18	-77.93996	-66.6	-69.61	-78,061211	0,01343
Acid19	-117.87437	0.97	1.06	-117,93197	0,05656
Base19	-117.21842	-68.89	-72.07	-117,36863	0,04000
Acid20	-40.50909	1.07	1.12	-40,534294	0,02487
Base20	-39.84690	-74.58	-85.63	-39,977400	0,00931
Acid21	-79.80473	1.91	1.95	-79,856794	0,05184
Base21	-79.14012	-71.1	-76.95	-79,289973	0,03421

<sup>a</sup> B3LYP/6-311++G(d,p)-PCM; <sup>b</sup> HF/6-31G(d)-PCM.

**Table S2:** Calculated descriptors at the B3LYP/6-311++G(d,p)-PCM(opt) level of theory and their correlation to the experimental aqueous pK<sub>a</sub> values.<sup>a</sup>

Acid	$\epsilon_{\text{HOMO-2}}$	$\epsilon_{\text{HOMO-1}}$	$\epsilon_{\text{HOMO}}$	$\epsilon_{\text{LUMO}}$	$\epsilon_{\text{LUMO+1}}$	$\epsilon_{\text{LUMO+2}}$	Q <sub>C</sub>	Q <sub>H</sub>	G <sub>aq</sub>	V <sub>sc</sub>	$\chi$	$\eta$	S
1	-0.35924	-0.35154	-0.34800	-0.12917	-0.12096	-0.05207	-0.07155	0.38653	-541.890652	146.490	-0.23858	0.10941	9.13951
2	-0.35461	-0.35216	-0.35208	-0.13640	-0.13625	-0.12502	0.14815	0.36680	-654.180454	155.113	-0.24424	0.10784	9.27299
3	-0.33891	-0.30673	-0.30317	-0.12356	-0.11475	-0.03637	-0.03246	0.34547	-618.405594	161.436	-0.21336	0.08980	11.13585
4	-0.34964	-0.34511	-0.33879	-0.12224	-0.11582	-0.01270	0.07021	0.34333	-581.237389	169.472	-0.230515	0.10827	9.235742
5	-0.32152	-0.31484	-0.27503	-0.05208	-0.01490	-0.00291	-0.62555	0.30501	-380.592708	113.297	-0.16356	0.11148	8.97062
6	-0.31306	-0.31139	-0.28097	-0.05552	-0.01829	-0.00692	-0.59836	0.30379	-490.191359	141.585	-0.16825	0.11273	8.87115
7	-0.33213	-0.32216	-0.30078	-0.04985	-0.02729	-0.01071	-0.63549	0.29990	-534.450278	183.846	-0.17532	0.12546	7.97038
8	-0.33693	-0.27577	-0.26111	-0.04636	-0.03096	-0.00680	-0.59311	0.29887	-462.534598	220.060	-0.15374	0.10738	9.31315
9	-0.37030	-0.27992	-0.26597	-0.05753	-0.01780	-0.00632	-0.58085	0.25943	-345.831493	158.552	-0.16175	0.10422	9.59509
10	-0.31424	-0.29872	-0.27450	-0.05192	-0.00896	-0.00696	-0.59713	0.27480	-460.376806	193.181	-0.16321	0.11129	8.98553
11	-0.28313	-0.27933	-0.27015	-0.01128	-0.00194	0.00883	-0.57489	0.28158	-378.009680	126.806	-0.14072	0.12944	7.72589
12	-0.31677	-0.31238	-0.28554	-0.01971	-0.00879	-0.00628	-0.58138	0.27606	-574.940306	231.584	-0.15263	0.13292	7.52361
13	-0.35984	-0.30541	-0.22481	-0.02600	-0.00199	0.01762	-0.47845	0.25965	-194.098241	109.202	-0.12541	0.09941	10.05986
14	-0.49801	-0.30269	-0.30327	0.00559	0.01396	0.01396	-0.25066	0.25066	-77.3527910	73.903	-0.14884	0.15443	6.47543
15	-0.35166	-0.29644	-0.25838	-0.00846	0.00598	0.01866	-0.18931	0.23180	-1116.627187	76.929	-0.13342	0.12496	8.00256
16	-0.45278	-0.43184	-0.41412	0.00755	0.03756	0.04032	0.92418	0.15866	-338.355030	72.103	-0.20329	0.21084	4.74305
17	-0.34151	-0.25886	-0.24924	-0.01873	-0.01416	-0.00522	-0.59795	0.21569	-271.545275	164.412	-0.13399	0.11526	8.67641
18	-0.42662	-0.35870	-0.28170	-0.01111	0.00285	0.02236	-0.37668	0.20451	-78.5871410	85.900	-0.14641	0.13530	7.39126
19	-0.36426	-0.29733	-0.29731	0.00154	0.01685	0.01687	-0.41141	0.20570	-117.875712	107.558	-0.14789	0.14943	6.69232
20	-0.39338	-0.39338	-0.39338	-0.00414	0.04154	0.04154	-0.81342	0.20336	-40.5096130	61.450	-0.19876	0.19462	5.13822
21	-0.36941	-0.34461	-0.34460	-0.00462	0.01642	0.02915	-0.57688	0.19243	-79.8051650	100.459	-0.17461	0.16999	5.88270
<b>R<sup>2</sup></b>	<b>0.21</b>	<b>0.04</b>	<b>0.03</b>	<b>0.59</b>	<b>0.51</b>	<b>0.44</b>	<b>0.01</b>	<b>0.81</b>	<b>0.73</b>	<b>0.36</b>	<b>0.15</b>	<b>0.49</b>	<b>0.54</b>
<b>A<sup>b</sup></b>	-41.74	-11.30	-0.80	31.60	24.50	19.72	16.82	91.11	45.14	48.86	53.15	-34.03	85.00
<b>B<sup>b</sup></b>	-168.0	-93.10	-64.30	298.0	248.0	337.37	-5.03	-269.0	0.08	-0.22	200.5	409.0	-8.16

**Table S2: (Cont.)**

Base	$\epsilon_{\text{HOMO-2}}$	$\epsilon_{\text{HOMO-1}}$	$\epsilon_{\text{HOMO}}$	$\epsilon_{\text{LUMO}}$	$\epsilon_{\text{LUMO+1}}$	$\epsilon_{\text{LUMO+2}}$	$Q_C$	$G_{\text{aq}}$	$V_{\text{sc}}$	$\chi$	$\eta$	S
1	-0.31269	-0.28331	-0.25989	-0.09641	-0.05186	-0.00645	0.03210	-541.472027	137.112	-0.17815	0.08174	12.23391
2	-0.31219	-0.28960	-0.26280	-0.11158	-0.10178	-0.01586	0.26742	-653.738624	147.934	-0.18719	0.07561	13.22576
3	-0.27486	-0.26924	-0.23942	-0.09684	-0.05428	0.00463	0.05647	-617.959537	152.427	-0.16813	0.07129	14.02721
4	-0.29732	-0.26477	-0.24481	-0.08795	-0.00801	0.00021	0.19171	-580.774040	157.557	-0.16638	0.07843	12.75022
5	-0.27442	-0.23160	-0.20994	0.00224	0.00277	0.02285	-0.56609	-380.143646	109.676	-0.10385	0.10609	9.42596
6	-0.27337	-0.24556	-0.20823	0.00016	0.00638	0.01553	-0.54233	-489.739841	137.876	-0.10404	0.10420	9.59739
7	-0.27604	-0.25952	-0.20849	-0.00327	0.00033	0.01569	-0.61981	-534.000889	179.985	-0.10588	0.10261	9.74564
8	-0.23377	-0.21998	-0.19702	-0.00728	0.00494	0.01430	-0.51651	-462.080474	214.059	-0.10215	0.09487	10.54074
9	-0.24175	-0.21960	-0.19410	-0.00804	0.00358	0.01788	-0.52150	-345.364980	147.834	-0.10107	0.09303	10.74922
10	-0.26156	-0.22395	-0.19928	-0.00852	-0.00266	0.00878	-0.58750	-459.910580	185.541	-0.10390	0.09538	10.48438
11	-0.24428	-0.23746	-0.18732	-0.00096	0.01647	0.02207	-0.60227	-377.54012	120.467	-0.09414	0.09318	10.73192
12	-0.25678	-0.24905	-0.19378	-0.00416	0.00406	0.01559	-0.60774	-574.47260	222.756	-0.09897	0.09481	10.54741
13	-0.31333	-0.16293	-0.16290	0.00343	0.01892	0.01893	-0.38292	-193.61796	113.023	-0.07975	0.08318	12.02212
14	-0.25035	-0.24166	-0.24082	0.01456	0.04382	0.04382	-0.63488	-76.868322	44.950	-0.11313	0.12769	7.83147
15	-0.25450	-0.20624	-0.17468	0.01151	0.02129	0.02728	-0.51783	-116.11453	66.957	-0.08159	0.09310	10.74172
16	-0.37707	-0.36408	-0.22106	0.03447	0.04589	0.04590	0.44155	-337.86003	58.450	-0.09330	0.12777	7.82687
17	-0.24855	-0.22925	-0.12036	-0.00155	0.01311	0.01330	-0.76994	-271.098268	151.71	-0.06095	0.0600	16.6666
18	-0.32641	-0.22450	-0.16442	0.01086	0.02590	0.02833	-0.79766	-78.047210	53.500	-0.07678	0.08764	11.41031
19	-0.25244	-0.23334	-0.14947	0.01144	0.01958	0.02019	-0.88697	-117.32851	62.588	-0.06902	0.08046	12.42931
20	-0.32125	-0.32073	-0.15265	0.01280	0.05736	0.05741	-1.52034	-39.971285	24.429	-0.06993	0.08273	12.08824
21	-0.30446	-0.28468	-0.13742	0.01246	0.03086	0.04467	-1.16415	-79.259040	41.503	-0.06248	0.07494	13.34401
<b>R<sup>2</sup></b>	<b>0.04</b>	<b>0.01</b>	<b>0.60</b>	<b>0.40</b>	<b>0.47</b>	<b>0.53</b>	<b>0.39</b>	<b>0.73</b>	<b>0.52</b>	<b>0.59</b>	<b>0.01</b>	<b>0.03</b>
<b>A<sup>b</sup></b>	-8.2	5.0	88.2	22.63	17	3.74	6.83	45.10	45.54	57.7	29.8	0.20
<b>B<sup>b</sup></b>	-95	-54.4	354	269	346	752	-24.1	0.08	-0.22	370	-123	1.62

**Table S2 (Cont.)**

Acid/ Base	$\epsilon_{\text{LUMO}}^{\text{acid}} + \epsilon_{\text{HOMO}}^{\text{base}}$	$\epsilon_{\text{LUMO}}^{\text{acid}} - \epsilon_{\text{HOMO}}^{\text{base}}$	$\Delta Q_{\text{C}}$	$\Delta \chi$	$\Delta \eta$	$\Delta S$	$\Delta G_{\text{aq}}^{\text{c}}$	$\Delta V_{\text{sc}}$
1	-0.38906	0.13072	0.10365	0.06043	0.02767	3.09440	-10.9	-9.38
2	-0.3992	0.1264	0.11927	0.05705	0.03223	3.95277	-6.41	-7.18
3	-0.36298	0.11586	0.08893	0.04523	0.01851	2.89136	2.16	-9.01
4	-0.36705	0.12257	0.1215	0.06413	0.02984	3.51448	1.07	-11.91
5	-0.26202	0.15786	0.05946	0.05971	0.00539	0.45534	10.44	-3.62
6	-0.26375	0.15271	0.05603	0.06421	0.00853	0.72624	11.31	-3.71
7	-0.25834	0.15864	0.01568	0.06944	0.02285	1.77526	12.1	-3.86
8	-0.24338	0.15066	0.07660	0.05159	0.01251	1.22759	14.56	-6.00
9	-0.25163	0.13657	0.05935	0.06068	0.01119	1.15413	20.18	-10.72
10	-0.25120	0.14736	0.00963	0.05931	0.01591	1.49884	19.31	-7.64
11	-0.19860	0.17604	-0.02738	0.04658	0.03626	3.00603	27.22	-6.34
12	-0.21349	0.17407	-0.02636	0.05366	0.03811	3.02381	22.31	-8.83
13	-0.18893	0.13693	0.09553	0.04566	0.01623	1.96226	29.79	3.82
14	-0.23523	0.24641	-0.38422	0.03571	0.02674	1.35604	31.52	-28.95
15	-0.18314	0.16622	-0.32852	0.05184	0.03187	2.73915	50.08	-9.97
16	-0.21351	0.22861	-0.48263	0.10999	0.08307	3.08382	39.63	-13.65
17	-0.13904	0.10163	-0.1720	0.07304	0.05526	7.99000	59.03	-12.72
18	-0.17553	0.15331	-0.42098	0.06963	0.04766	4.01906	66.23	-32.40
19	-0.14793	0.15101	-0.47556	0.07887	0.06897	5.73699	70.32	-44.97
20	-0.15679	0.14851	-0.70692	0.12884	0.11190	6.95003	68.68	-37.02
21	-0.14204	0.13280	-0.58727	0.11213	0.09505	7.46131	72.31	-58.96
<b>R<sup>2</sup></b>	<b>0.69</b>	<b>0.02</b>	<b>0.85</b>	<b>0.40</b>	<b>0.66</b>	<b>0.52</b>	<b>0.97</b>	<b>0.63</b>
<b>A<sup>b</sup></b>	64.0	6.1	10.3	-13.9	-0.46	-0.89	-1.28	4.54
<b>B<sup>b</sup></b>	189	82.2	-62.1	489.0	503.0	6.06	0.68	-0.91

<sup>a</sup> Orbital energies ( $\epsilon_{\text{HOMO}-2}$  to  $\epsilon_{\text{LUMO}+2}$ ). NBO atomic charges ( $Q_i$ ). aqueous Gibbs free energy ( $G_{\text{aq}}$ ). electronegativity ( $\chi$ ). hardness ( $\eta$ ) and softness ( $S$ ) in atomic units; volume of the solvent cavity ( $V_{\text{sc}}$ ) in  $\text{\AA}^3$ ; <sup>b</sup> Coefficients of the correlation equation  $\text{p}K_{\text{a,exp}} = A + B \cdot \text{Descriptor}$ ; A is the intercept and B is the slope; <sup>c</sup> in kcal/mol at the state of 1 mol.dm<sup>-3</sup>.

**Table S3:** Calculated descriptors at the B3LYP/6-311++G(d,p)-PCM(sp) level of theory and their correlation to the experimental aqueous pK<sub>a</sub> values.<sup>a</sup>

Acid	$\epsilon_{\text{HOMO-2}}$	$\epsilon_{\text{HOMO-1}}$	$\epsilon_{\text{HOMO}}$	$\epsilon_{\text{LUMO}}$	$\epsilon_{\text{LUMO+1}}$	$\epsilon_{\text{LUMO+2}}$	Q <sub>C</sub>	Q <sub>H</sub>	G <sub>aq</sub> <sup>c</sup>	V <sub>sc</sub>	$\chi$	$\eta$	S
1	-0.36387	-0.35878	-0.35117	-0.13358	-0.12808	-0.05695	-0.06659	0.37285	-541.88147	147.558	-0.242375	0.108795	9.19159888
2	-0.35372	-0.35256	-0.34928	-0.14508	-0.13861	-0.10523	0.15398	0.32993	-654.16102	156.845	-0.24718	0.1021	9.79431929
3	-0.34433	-0.31759	-0.31288	-0.12902	-0.12208	-0.04830	-0.02896	0.33593	-618.3688	161.714	-0.22095	0.09193	10.8778418
4	-0.35705	-0.35302	-0.34563	-0.13271	-0.12146	-0.01851	0.07159	0.32949	-581.2015	170.594	-0.23917	0.10646	9.39319932
5	-0.33052	-0.32064	-0.27860	-0.05788	-0.01816	-0.01044	-0.62555	0.29945	-380.59098	113.345	-0.16824	0.11036	9.06125408
6	-0.31753	-0.31356	-0.28637	-0.06125	-0.02285	-0.01295	-0.59836	0.29781	-490.18745	141.761	-0.17381	0.11256	8.88415068
7	-0.33213	-0.32216	-0.30078	-0.03327	-0.02729	-0.01071	-0.63549	0.30726	-534.44928	184.179	-0.167025	0.133755	7.47635602
8	-0.34245	-0.27949	-0.26429	-0.05100	-0.03248	-0.01180	-0.59311	0.29215	-462.53219	220.190	-0.157645	0.106645	9.37690468
9	-0.37597	-0.28392	-0.26952	-0.06221	-0.01969	-0.01185	-0.58085	0.25636	-345.82881	158.393	-0.165865	0.103655	9.64738797
10	-0.31782	-0.30405	-0.28107	-0.05197	-0.01287	-0.01143	-0.59713	0.27088	-460.37045	192.833	-0.16652	0.11455	8.72981231
11	-0.29019	-0.28709	-0.27810	-0.01680	-0.00737	0.00562	-0.57489	0.27681	-378.00755	126.865	-0.14745	0.13065	7.6540375
12	-0.32377	-0.31967	-0.29430	-0.02692	-0.01549	-0.01270	-0.58138	0.27312	-574.93883	230.979	-0.16061	0.13369	7.47999102
13	-0.36111	-0.30853	-0.23074	-0.03082	-0.00487	0.01453	-0.48905	0.25745	-194.09667	109.150	-0.13078	0.09996	10.0040016
14	-0.50402	-0.30446	-0.30422	0.00455	0.01167	0.01200	-0.23892	0.23892	-77.352504	73.877	-0.149835	0.154385	6.47731321
15	-0.35645	-0.30164	-0.26155	-0.01223	0.00214	0.01533	-0.18729	0.22533	-116.62617	76.884	-0.13689	0.12466	8.02181935
16	-0.45466	-0.43373	-0.41258	0.00666	0.03734	0.03982	0.92742	0.14268	-338.35446	71.995	-0.20296	0.20962	4.77053716
17	-0.34493	-0.26591	-0.25706	-0.02573	-0.01983	-0.01034	-0.59795	0.21561	-271.5447	164.262	-0.141395	0.115665	8.64565772
18	-0.42768	-0.35966	-0.28345	-0.01256	0.00188	0.02201	-0.37668	0.18834	-78.586828	86.010	-0.148005	0.135445	7.38307062
19	-0.36473	-0.29892	-0.29855	0.00034	0.01513	0.01610	-0.40802	0.20400	-117.87541	107.567	-0.149105	0.149445	6.69142494
20	-0.39532	-0.39443	-0.39441	-0.00524	0.03926	0.04119	-0.81342	0.20335	-40.509424	61.450	-0.199825	0.194585	5.13914228
21	-0.37025	-0.34555	-0.34546	-0.00552	0.01581	0.02827	-0.57688	0.19230	-79.804954	100.459	-0.17549	0.16997	5.88339119
<b>R<sup>2</sup></b>	<b>0.20</b>	<b>0.03</b>	<b>0.02</b>	<b>0.57</b>	<b>0.51</b>	<b>0.53</b>	<b>0.01</b>	<b>0.81</b>	<b>0.73</b>	<b>0.37</b>	<b>0.17</b>	<b>0.49</b>	<b>0.54</b>
A <sup>b</sup>	-40.70	-8.5	-0.01	31.80	25.27	21.0	16.84	92.20	45.10	49.10	56.01	-34.05	84.48
B <sup>b</sup>	-163.0	-83.0	-61.10	282.20	243.0	394.0	-5.01	-280.49	0.08	-0.22	212.82	408.23	-8.11

**Table S3 (Cont.)**

Base	$\epsilon_{\text{HOMO-2}}$	$\epsilon_{\text{HOMO-1}}$	$\epsilon_{\text{HOMO}}$	$\epsilon_{\text{LUMO}}$	$\epsilon_{\text{LUMO+1}}$	$\epsilon_{\text{LUMO+2}}$	$Q_{\text{C}}$	$G_{\text{req}}^e$	$V_{\text{sc}}$	$\chi$	$\eta$	S
1	-0.31812	-0.28936	-0.26859	-0.10240	-0.06046	-0.01513	0.02999	-541.4677	137.304	-0.185495	0.083095	12.0344184
2	-0.31987	-0.29721	-0.26993	-0.11934	-0.10907	-0.02468	0.26506	-653.74185	148.084	-0.194635	0.075295	13.2810944
3	-0.28748	-0.28309	-0.25154	-0.10423	-0.07016	-0.00402	0.04985	-617.93306	153.134	-0.177885	0.073655	13.5768108
4	-0.30679	-0.27562	-0.25580	-0.09695	-0.01869	-0.00644	0.18579	-580.76778	157.781	-0.176375	0.079425	12.5904942
5	-0.28688	-0.24154	-0.22475	-0.00776	-0.00570	0.01524	-0.57470	-380.14292	109.521	-0.116255	0.108495	9.21701461
6	-0.27859	-0.25542	-0.22039	-0.00822	-0.00290	0.01252	-0.54927	-489.73733	138.342	-0.114305	0.106085	9.42640336
7	-0.28373	-0.27417	-0.22513	-0.01155	-0.00966	0.00968	-0.62582	-533.99732	179.997	-0.11834	0.10679	9.36417268
8	-0.23937	-0.22673	-0.20294	-0.01261	-0.01067	0.01133	-0.52712	-462.07658	214.105	-0.107775	0.095165	10.50806649
9	-0.24935	-0.22739	-0.20523	-0.01495	-0.01149	0.01245	-0.53049	-345.36262	147.942	-0.11009	0.09514	10.5108262
10	-0.28290	-0.24289	-0.23273	-0.03172	-0.02387	0.01017	-0.59119	-459.9052	185.249	-0.132225	0.100505	9.94975374
11	-0.25925	-0.24944	-0.20919	-0.02260	0.00292	0.01504	-0.60607	-377.53136	120.625	-0.115895	0.093295	10.718688
12	-0.29034	-0.27405	-0.22637	-0.01992	-0.01319	0.00362	-0.60937	-574.46792	222.230	-0.123145	0.103225	9.68757568
13	-0.33704	-0.18823	-0.18748	-0.01228	0.01307	0.01740	-0.38269	-193.61839	112.947	-0.09988	0.0876	11.4155251
14	-0.25783	-0.25586	-0.25412	0.00203	0.03250	0.03397	-0.63355	-76.868545	45.027	-0.126045	0.128075	7.80792504
15	-0.26525	-0.21904	-0.19276	-0.00567	0.01473	0.01835	-0.51228	-116.11479	67.000	-0.099215	0.093545	10.6900422
16	-0.38264	-0.37096	-0.23600	0.02923	0.03564	0.03732	0.44985	-337.85983	58.690	-0.103385	0.132615	7.54062512
17	-0.24607	-0.22793	-0.11727	-0.00177	0.01258	0.01563	-0.74012	-271.095183	160.85	-0.05952	0.05775	17.316173
18	-0.33641	-0.24015	-0.18565	0.01082	0.03250	0.03397	-0.79547	-78.047781	53.513	-0.087415	0.098235	10.1796712
19	-0.26874	-0.26067	-0.15290	-0.00610	0.00438	0.01600	-0.88025	-117.32863	62.486	-0.0795	0.0734	13.6239782
20	-0.36123	0.35781	-0.19969	-0.00896	0.03786	0.03910	-1.51808	-39.96809	24.429	-0.104325	0.095365	10.4860274
21	-0.35735	-0.32803	-0.14919	-0.02006	0.00150	0.02118	-1.15766	-79.255763	41.308	-0.084625	0.064565	15.4882676
<b>R<sup>2</sup></b>	<b>0.09</b>	<b>0.11</b>	<b>0.54</b>	<b>0.35</b>	<b>0.46</b>	<b>0.53</b>	<b>0.41</b>	<b>0.73</b>	<b>0.50</b>	<b>0.56</b>	<b>0.02</b>	<b>0.06</b>
A <sup>b</sup>	-22.0	28.5	90.0	25.7	21.0	8.0	6.98	45.10	45.1	64.0	30.2	-1.5
B <sup>b</sup>	-137.0	42.8	336.0	262	341.0	817.0	-23.8	0.08	-0.22	378.0	-125.0	1.79

**Table S3 (Cont.)**

Acid/ Base	$\epsilon_{\text{LUMO}}^{\text{acid}} + \epsilon_{\text{HOMO}}^{\text{base}}$	$\epsilon_{\text{LUMO}}^{\text{acid}} - \epsilon_{\text{HOMO}}^{\text{base}}$	$\Delta Q_C$	$\Delta \chi$	$\Delta \eta$	$\Delta S$	$\Delta G_{\text{aq}}^{\text{c}}$	$\Delta G_{\text{aq}}^{\text{d}}$	$\Delta G_{\text{aq}}^{\text{e}}$	$\Delta V_{\text{sc}}$
1	-0.40217	0.13501	0.09658	0.05688	0.025700	2.842819	-12.64	-11.8	-11.24	-10.25
2	-0.41501	0.12485	0.11108	0.052545	0.026805	3.486775	-8.43	-7.61	-7.86	-8.76
3	-0.38056	0.12252	0.07881	0.043065	0.018275	2.698968	1.49	1.14	2.54	-8.58
4	-0.38851	0.12309	0.11420	0.062795	0.027035	3.197294	0.10	1.02	1.27	-12.81
5	-0.28263	0.16687	0.05085	0.051985	0.001865	0.155760	10.42	9.37	10.27	-3.82
6	-0.28164	0.15914	0.04909	0.059505	0.006475	0.542252	11.58	10.76	11.56	-3.42
7	-0.2584	0.19186	0.00967	0.048685	0.026965	1.887816	12.81	11.94	12.72	-4.19
8	-0.25394	0.15194	0.06599	0.049870	0.011480	1.131160	14.82	12.5	15.01	-6.08
9	-0.26744	0.14302	0.05036	0.055775	0.008515	0.863438	20.71	18.04	21.65	-10.45
10	-0.2847	0.18076	0.00594	0.034295	0.014045	1.219941	20.00	18.38	21.06	-7.58
11	-0.22599	0.19239	-0.03118	0.031555	0.037355	3.064650	27.84	25.68	27.92	-6.24
12	-0.25329	0.19945	-0.02799	0.037465	0.030465	2.207584	23.52	21.91	24.61	-8.75
13	-0.2183	0.15666	0.10636	0.030900	0.012360	1.411523	29.79	26.85	29.23	3.79
14	-0.24957	0.25867	-0.39463	0.023790	0.026310	1.330611	31.62	25.6	32.80	-28.86
15	-0.20499	0.18053	-0.32499	0.037675	0.031115	2.668222	50.16	47.72	50.00	-9.88
16	-0.22934	0.24266	-0.47757	0.099575	0.077005	2.770087	39.71	36.49	39.49	-13.30
17	-0.14300	0.09154	-0.14217	0.081875	0.057915	8.670515	61.01	59.05	60.58	-3.41
18	-0.19821	0.17309	-0.41879	0.060590	0.037210	2.796600	66.38	63.31	67.37	-32.50
19	-0.15256	0.15324	-0.47223	0.069605	0.076045	6.932553	70.86	67.59	72.22	-45.08
20	-0.20493	0.19445	-0.70466	0.095500	0.099220	5.346885	68.99	57.89	68.80	-37.02
21	-0.15471	0.14367	-0.58078	0.090865	0.105405	9.604876	73.15	67.26	73.73	-59.25
<b>R<sup>2</sup></b>	<b>0.69</b>	<b>0.12</b>	<b>0.83</b>	<b>0.31</b>	<b>0.67</b>	<b>0.49</b>	<b>0.96</b>	<b>0.95</b>	<b>0.96</b>	<b>0.54</b>
A <sup>b</sup>	67.8	10.8	10.2	-7.64	0.59	3.19	-0.91	-0.58	-1.34	5.95
B <sup>b</sup>	190.0	38.2	-62.3	469.0	499.0	5.0	0.67	0.72	0.67	-0.84

<sup>a</sup> Orbital energies ( $\epsilon_{\text{HOMO}+2}$  to  $\epsilon_{\text{LUMO}+2}$ ). NBO atomic charges ( $Q_i$ ). aqueous Gibbs free energy ( $G_{\text{aq}}$ ). electronegativity ( $\chi$ ). hardness ( $\eta$ ) and softness ( $S$ ) in atomic units; volume of the solvent cavity ( $V_{\text{sc}}$ ) in  $\text{\AA}^3$ ; <sup>b</sup> Coefficients of the correlation equation  $\text{p}K_{\text{a,exp}} = A + B \cdot \text{Descriptor}$ ; A is the intercept and B is the slope; <sup>c</sup> Calculated using eq 3a; <sup>d</sup> Calculated using eq 3b (HF/6-31G(d)); <sup>e</sup> Calculated using eq 4;  $\Delta G$  values in kcal/mol at the state of 1 mol.dm<sup>-3</sup>.

**Table S4:** Calculated descriptors at the B3LYP/6-311++G(d,p) level of theory and their correlation to the experimental aqueous  $pK_a$  values.<sup>a</sup>

Acid	$\epsilon_{LUMO}$	$Q_C$	$Q_H$	$\chi$	$\eta$	S	Base	$\epsilon_{HOMO}$	$Q_C$	$\chi$	$\eta$	S
1	-0.14447	-0.04488	0.27719	-0.25313	0.10866	9.20344	1	-0.11562	0.00252	-0.03295	0.08268	12.09555
2	-0.15839	0.1666	0.26293	-0.26269	0.10430	9.58819	2	-0.11838	0.24250	-0.04371	0.07467	13.39226
3	-0.13775	-0.00863	0.26578	-0.22806	0.09031	11.07358	3	-0.09357	0.03487	-0.02066	0.07292	13.71460
4	-0.13981	0.09345	0.25425	-0.24758	0.10777	9.27945	4	-0.10726	0.14629	-0.02565	0.08161	12.25340
5	-0.07309	-0.60296	0.26813	-0.17876	0.10567	9.46342	5	-0.05048	-0.56853	0.03031	0.08079	12.37854
6	-0.07146	-0.57444	0.27067	-0.18165	0.11019	9.07523	6	-0.06176	-0.54100	0.02453	0.08629	11.58950
7	-0.04249	-0.60876	0.26961	-0.16970	0.12721	7.86133	7	-0.05398	-0.62404	0.01698	0.07096	14.09245
8	-0.05398	-0.56839	0.24713	-0.15728	0.10330	9.68054	8	-0.04311	-0.50905	0.01918	0.06229	16.05394
9	-0.06099	-0.5635	0.23217	-0.16341	0.10242	9.76372	9	-0.03385	-0.51416	0.03095	0.06480	15.43329
10	-0.05519	-0.5813	0.24259	-0.16517	0.10998	9.09298	10	-0.04572	-0.58440	0.02291	0.06863	14.57089
11	-0.02659	-0.55389	0.2258	-0.14634	0.11975	8.35108	11	-0.02964	-0.58852	0.03484	0.06448	15.50989
12	-0.01780	-0.55839	0.24851	-0.14922	0.13142	7.60919	12	-0.03804	-0.59468	0.02167	0.05971	16.74902
13	-0.02899	-0.4659	0.23632	-0.12756	0.09857	10.14559	13	0.00423	-0.36389	0.06141	0.05718	17.48863
14	0.00392	-0.222	0.22200	-0.14872	0.15264	6.55158	14	-0.02868	-0.46037	0.04882	0.07750	12.90323
15	-0.01081	-0.1746	0.21123	-0.13538	0.12457	8.02794	15	0.01109	-0.37622	0.06940	0.05831	17.15119
16	-0.00627	0.93457	0.11453	-0.21169	0.20542	4.86819	16	-0.01244	0.47031	0.09259	0.10503	9.52154
17	-0.01717	-0.58828	0.21038	-0.13240	0.11523	8.67867	17	0.03167	-0.70512	0.06748	0.03581	27.9252
18	-0.01184	-0.36705	0.18352	-0.14607	0.13423	7.45018	18	0.03858	-0.65389	0.08582	0.04724	21.16850
19	-0.00104	-0.40294	0.20146	-0.14975	0.14871	6.72450	19	0.05171	-0.71057	0.08827	0.03656	27.35604
20	-0.00553	-0.80608	0.20152	-0.20045	0.19492	5.13044	20	0.06551	-1.39183	0.10273	0.03722	26.87089
21	-0.00580	-0.57439	0.19146	-0.17583	0.17003	5.88149	21	0.06952	-0.98622	0.09580	0.02628	38.05899
<b>R<sup>2</sup></b>	<b>0.60</b>	<b>0.02</b>	<b>0.63</b>	<b>0.19</b>	<b>0.52</b>	<b>0.57</b>		<b>0.88</b>	<b>0.28</b>	<b>0.82</b>	<b>0.83</b>	<b>0.62</b>
A <sup>b</sup>	32.3	16.7	104	53.1	-34.8	86.6		26.6	8.49	4.49	4.42	-16.9
B <sup>b</sup>	269	-5.8	-371	194	421	-8.22		300	-22.9	375	-383	2.04

**Table S4 (Cont.)**

Acid/ Base	$\epsilon_{\text{LUMO}}^{\text{acid}} + \epsilon_{\text{HOMO}}^{\text{base}}$	$\epsilon_{\text{LUMO}}^{\text{acid}} - \epsilon_{\text{HOMO}}^{\text{base}}$	$\Delta Q_{\text{C}}$	$\Delta \chi$	$\Delta \eta$	$\Delta S$	$\Delta G_{\text{gas}}^{\text{c}}$
1	-0.26009	-0.02885	0.04740	0.22018	0.02598	2.89211	287.15
2	-0.27677	-0.04001	0.07590	0.21898	0.02963	3.80407	290.84
3	-0.23132	-0.04418	0.04350	0.20740	0.01739	2.64102	304.05
4	-0.24707	-0.03255	0.05284	0.22193	0.02616	2.97395	301.18
5	-0.12357	-0.02261	0.03443	0.20907	0.02489	2.91511	319.21
6	-0.13322	-0.00970	0.03344	0.20618	0.02391	2.51427	316.57
7	-0.09647	0.01149	-0.01528	0.18668	0.05625	6.23112	321.34
8	-0.09709	-0.01087	0.05934	0.17646	0.04101	6.37340	325.46
9	-0.09484	-0.02714	0.04934	0.19436	0.03763	5.66957	336.15
10	-0.10091	-0.00947	-0.00310	0.18808	0.04135	5.47791	331.50
11	-0.05623	0.00305	-0.03463	0.18117	0.05527	7.15881	340.31
12	-0.05584	0.02024	-0.03629	0.17089	0.07172	9.13982	336.62
13	-0.02476	-0.03322	0.10201	0.18897	0.04139	7.34304	346.06
14	-0.02476	0.03260	-0.23837	0.1975	0.07514	6.35165	368.83
15	0.00028	-0.02190	-0.20162	0.20477	0.06626	9.12325	377.80
16	-0.01871	0.00617	-0.46426	0.30427	0.10039	4.65335	365.67
17	0.01450	-0.04884	-0.11693	0.199875	0.079415	19.2465	373.03
18	0.02674	-0.05042	-0.28684	0.23189	0.08699	13.71832	398.98
19	0.05067	-0.05275	-0.30763	0.23802	0.11216	20.63154	405.33
20	0.05998	-0.07104	-0.58575	0.30317	0.15770	21.74044	409.25
21	0.06372	-0.07532	-0.41183	0.27162	0.14375	32.17750	410.77
<b>R<sup>2</sup></b>	<b>0.79</b>	<b>0.20</b>	<b>0.77</b>	<b>0.34</b>	<b>0.83</b>	<b>0.70</b>	<b>0.95</b>
A <sup>b</sup>	30.5	11.8	10.3	-42.8	-7.66	0.37	-139
B <sup>b</sup>	154	-282	-78.8	285	420	1.99	0.45

<sup>a</sup> Orbital energies ( $\epsilon_{\text{HOMO-2}}$  to  $\epsilon_{\text{LUMO+2}}$ ). NBO atomic charges ( $Q_i$ ). gas-phase Gibbs free energy change ( $\Delta G_{\text{gas}}$ , in cal/mol). electronegativity ( $\chi$ ). hardness ( $\eta$ ) and softness ( $S$ ) in atomic units; volume of the solvent cavity ( $V_{\text{sc}}$ ) in Å<sup>3</sup>. <sup>b</sup> Coefficients of the correlation equation  $\text{p}K_{\text{a,exp}} = A + B \cdot \text{Descriptor}$ ; A is the intercept and B is the slope; <sup>c</sup> in kcal/mol.

**Table S5:**Results of the “Best subsets” regression analysis for two-, three- and four-variable correlation equations at different levels of theory.<sup>a</sup>

B3LYP/6-311++G(d,p)-PCM(opt)										
# Variables	R <sup>2</sup>	$\Delta G_{\text{aq}}$	Q <sub>H</sub>	$\epsilon_{\text{LUMO}}^{\text{acid}}$ + $\epsilon_{\text{HOMO}}^{\text{base}}$	$\epsilon_{\text{LUMO}}^{\text{acid}}$ - $\epsilon_{\text{HOMO}}^{\text{base}}$	$\Delta Q_{\text{C}}$	$\Delta\chi$	$\Delta\eta$	$\Delta S$	$\Delta V_{\text{sc}}$
2	0.982	X				X				
2	0.983	X						X		
3	0.983	X				X				X
3	0.990	X				X			X	
4	0.991	X	X			X			X	

B3LYP/6-311++G(d,p)-PCM(sp) {Using eq 3a for $\Delta G_{\text{aq}}$ }										
# Variables	R <sup>2</sup>	$\Delta G_{\text{aq}}$	$\epsilon_{\text{LUMO}}^{\text{acid}}$ + $\epsilon_{\text{HOMO}}^{\text{base}}$	$\epsilon_{\text{LUMO}}^{\text{acid}}$ - $\epsilon_{\text{HOMO}}^{\text{base}}$	$\Delta Q_{\text{C}}$	$\Delta\chi$	$\Delta\eta$	$\Delta S$	$\Delta V_{\text{sc}}$	
2	0.980	X			X					
2	0.972	X							X	
3	0.980	X			X				X	
3	0.987	X			X			X		
4	0.988	X			X		X	X		

B3LYP/6-311++G(d,p)-PCM(sp) {Using eq 4 for $\Delta G_{\text{aq}}$ }										
# Variables	R <sup>2</sup>	$\Delta G_{\text{aq}}$	Q <sub>H</sub>	$\epsilon_{\text{LUMO}}^{\text{acid}}$ + $\epsilon_{\text{HOMO}}^{\text{base}}$	$\epsilon_{\text{LUMO}}^{\text{acid}}$ - $\epsilon_{\text{HOMO}}^{\text{base}}$	$\Delta Q_{\text{C}}$	$\Delta\chi$	$\Delta\eta$	$\Delta S$	$\Delta V_{\text{sc}}$
2	0.980	X				X				
2	0.980	X						X		
3	0.980	X				X				X
3	0.987	X				X			X	
4	0.990	X	X			X			X	
4	0.991	X	X	X		X			X	

**Table S5** (Cont.)

B3LYP/6-311++G(d,p)

# Variables	R <sup>2</sup>	$\Delta G_{\text{gas}}$	$\epsilon_{\text{LUMO}}^{\text{acid}} + \epsilon_{\text{HOMO}}^{\text{base}}$	$\epsilon_{\text{LUMO}}^{\text{acid}} - \epsilon_{\text{HOMO}}^{\text{base}}$	$\epsilon_{\text{LUMO}}^{\text{acid}}$	$\Delta Q_{\text{C}}$	$\Delta \chi$	$\Delta \eta$	$\Delta S$	$\Delta V_{\text{sc}}$	Q <sub>H</sub>
2	0.966	X	X	X	X	X	X	X	X	X	X
2	0.960	X	X	X	X	X	X	X	X	X	X
3	0.969	X	X	X	X	X	X	X	X	X	X
4	0.975	X	X	X	X	X	X	X	X	X	X
4	0.973	X	X	X	X	X	X	X	X	X	X

<sup>a</sup> Orbital energies ( $\epsilon$ ) and NBO atomic charges (Q) in atomic units; aqueous Gibbs free energy change in kcal/mol; change the volume of the solvent cavity going from the acid to the conjugate base ( $\Delta V_{\text{sc}}$ ) in  $\text{\AA}^3$ .

**Table S6:**R<sup>2</sup> values of the intervariable correlations of selected descriptors calculated at different levels of theory.

B3LYP/6-311++G(d,p)-PCM(opt)

	$\Delta G_{\text{aq}}$	$\epsilon_{\text{LUMO}}^{\text{acid}} + \epsilon_{\text{HOMO}}^{\text{base}}$	$\epsilon_{\text{LUMO}}^{\text{acid}} - \epsilon_{\text{HOMO}}^{\text{base}}$	$\Delta Q_{\text{C}}$	$\Delta \chi$	$\Delta \eta$	$\Delta S$	$\Delta V_{\text{sc}}$
$\epsilon_{\text{LUMO}}^{\text{acid}} + \epsilon_{\text{HOMO}}^{\text{base}}$	0.81							
$\epsilon_{\text{LUMO}}^{\text{acid}} - \epsilon_{\text{HOMO}}^{\text{base}}$	0.01	0.05						
$\Delta Q_{\text{C}}$	0.77	0.48	0.14					
$\Delta \chi$	0.33	0.16	0.00	0.50				
$\Delta \eta$	0.55	0.27	0.01	0.74	0.75			
$\Delta S$	0.44	0.19	0.12	0.37	0.41	0.68		
$\Delta V_{\text{sc}}$	0.55	0.23	0.00	0.69	0.37	0.58	0.45	
Q <sub>H</sub>	0.82	0.80	0.14	0.70	0.34	0.45	0.20	0.33

**Table S6 (Cont.)**

**B3LYP/6-311++G(d,p)-PCM(sp)**

	$\Delta G_{\text{aq}}^{\text{a}}$	$\epsilon_{\text{LUMO}}^{\text{acid}} + \epsilon_{\text{HOMO}}^{\text{base}}$	$\epsilon_{\text{LUMO}}^{\text{acid}} - \epsilon_{\text{HOMO}}^{\text{base}}$	$\Delta Q_{\text{C}}$	$\Delta\chi$	$\Delta\eta$	$\Delta S$	$\Delta V_{\text{sc}}$
$\epsilon_{\text{LUMO}}^{\text{acid}} + \epsilon_{\text{HOMO}}^{\text{base}}$	0.82							
$\epsilon_{\text{LUMO}}^{\text{acid}} - \epsilon_{\text{HOMO}}^{\text{base}}$	0.03	0.06						
$\Delta Q_{\text{C}}$	0.74	0.44	0.19					
$\Delta\chi$	0.22	0.09	0.02	0.32				
$\Delta\eta$	0.55	0.29	0.01	0.71	0.58			
$\Delta S$	0.41	0.22	0.12	0.31	0.39	0.69		
$\Delta V_{\text{sc}}$	0.46	0.18	0.02	0.66	0.24	0.58	0.40	
$Q_{\text{H}}$	0.80	0.69	0.17	0.70	0.25	0.46	0.19	0.29

<sup>a</sup> Calculated using eq3a

**B3LYP/6-311++G(d,p)**

	$\Delta G_{\text{gas}}^{\text{a}}$	$\epsilon_{\text{LUMO}}^{\text{acid}} + \epsilon_{\text{HOMO}}^{\text{base}}$	$\epsilon_{\text{LUMO}}^{\text{acid}} - \epsilon_{\text{HOMO}}^{\text{base}}$	$\Delta Q_{\text{C}}$	$\Delta\chi$	$\Delta\eta$	$\Delta S$
$\epsilon_{\text{LUMO}}^{\text{acid}} + \epsilon_{\text{HOMO}}^{\text{base}}$	0.89						
$\epsilon_{\text{LUMO}}^{\text{acid}} - \epsilon_{\text{HOMO}}^{\text{base}}$	0.13	0.03					
$\Delta Q_{\text{C}}$	0.73	0.54	0.09				
$\Delta\chi$	0.25	0.10	0.25	0.65			
$\Delta\eta$	0.81	0.67	0.13	0.87	0.47		
$\Delta S$	0.69	0.56	0.38	0.48	0.22	0.75	
$Q_{\text{H}}$	0.58	0.50	0.02	0.66	0.43	0.53	0.24

**Table S7:**Reduced set of descriptors of the carbon acids of the Test set at different levels of theory.<sup>a</sup>

B3LYP/6-311++G(d,p)-PCM(opt)													
Acids	pK <sub>a</sub> <sup>b</sup>	$\epsilon_{\text{LUMO}}^{\text{acid}} + \epsilon_{\text{HOMO}}^{\text{base}}$	$\epsilon_{\text{LUMO}}^{\text{acid}} - \epsilon_{\text{HOMO}}^{\text{base}}$	Q <sub>H</sub> <sup>acid</sup>	$\Delta Q_{\text{C}}$	$\Delta G_{\text{aq}}$	$\Delta V_{\text{sc}}$	$\Delta\chi$	$\Delta\eta$	$\Delta\chi$	$\Delta\eta$	$\Delta\chi$	$\Delta S$
22	33	-0.17109	0.15709	0.23043	-0.35269	52.0	-25.3	0.044865	0.037045	3.75962600			
23	6.4	-0.23388	0.14568	0.28555	0.04386	18.6	-3.91	0.058595	0.011935	1.14798700			
24	5	-0.26576	0.14244	0.30074	0.06842	9.13	-9.2	0.05194	0.0138800	1.47695202			
25	3.57	-0.35539	0.13425	0.34924	0.16914	12.5	-0.235	0.050585	0.027225	3.25825700			
B3LYP/6-311++G(d,p)-PCM(sp)													
Acids	pK <sub>a</sub> <sup>b</sup>	$\epsilon_{\text{LUMO}}^{\text{acid}} + \epsilon_{\text{HOMO}}^{\text{base}}$	$\epsilon_{\text{LUMO}}^{\text{acid}} - \epsilon_{\text{HOMO}}^{\text{base}}$	Q <sub>H</sub> <sup>acid</sup>	$\Delta Q_{\text{C}}$	$\Delta G_{\text{aq}}^{\text{c}}$	$\Delta G_{\text{aq}}^{\text{d}}$	$\Delta G_{\text{aq}}^{\text{e}}$	$\Delta V_{\text{sc}}$	$\Delta\chi$	$\Delta\eta$	$\Delta\chi$	$\Delta S$
22	33	-0.17839	0.16257	0.22365	-0.35085	52.44	51.52	53.66	-25.3	0.041900	0.036690	3.62712262	
23	6.4	-0.26292	0.15512	0.28162	0.03034	19.08	17.36	18.69	-3.9	0.050375	0.007465	0.70804487	
24	5	-0.26832	0.1397	0.28928	0.06047	8.82	6.95	9.07	-9.2	0.054485	0.011935	1.27085132	
25	3.57	-0.37018	0.13476	0.30825	0.11800	8.08	9.44	9.34	-0.65	0.051875	0.027945	3.28265547	
B3LYP/6-311++G(d,p)													
Acids	pK <sub>a</sub> <sup>b</sup>	$\epsilon_{\text{LUMO}}^{\text{acid}} + \epsilon_{\text{HOMO}}^{\text{base}}$	$\epsilon_{\text{LUMO}}^{\text{acid}} - \epsilon_{\text{HOMO}}^{\text{base}}$	$\epsilon_{\text{LUMO}}^{\text{acid}}$	Q <sub>H</sub> <sup>acid</sup>	$\Delta Q_{\text{C}}$	$\Delta G_{\text{gas}}$	$\Delta\chi$	$\Delta\eta$	$\Delta\chi$	$\Delta\eta$	$\Delta\chi$	$\Delta S$
22	33	-0.02082	-0.01344	-0.01713	0.22765	-0.34698	368.49	0.181580	0.05329	8.3818			
23	6.4	-0.09285	-0.02693	-0.05989	0.25966	0.02707	330.56	0.198165	0.03720	5.3269			
24	5	-0.11352	-0.03790	-0.07571	0.25497	0.04426	322.60	0.215715	0.02354	3.0318			
25	3.57	-0.21311	-0.03915	-0.12613	0.24279	0.05617	308.54	0.226875	0.02692	3.1203			

<sup>a</sup> Orbital energies ( $\epsilon$ ) and NBO atomic charges (Q) in atomic units; aqueous Gibbs free energy change ( $\Delta G_{\text{aq}}$ ) in kcal/mol; change the volume of the solvent cavity going from the acid to the conjugate base ( $\Delta V_{\text{sc}}$ ) in Å<sup>3</sup>; <sup>b</sup> Ref. 37 and 38; <sup>c</sup> Calculated using eq 3a; <sup>d</sup> Calculated using eq 3b (HF/6-31G(d)); <sup>e</sup> Calculated using eq 4.  $\Delta G$  values in kcal/mol at the state of 1 mol.dm<sup>-3</sup>.

**PARTIE 3 :**

*Etude théorique des effets de solvants sur l'équilibre tautomère céto-énolique des composés  $\beta$ -dicarbonylés cycliques et de leurs analogues à chaînes ouvertes.*

*1<sup>er</sup> cas : Acide tétronique / Acétylacétate d'éthyle*

*2<sup>ème</sup> cas : Dimédone /Acétylacétone*

**Résumé**

---

*Les effets de solvants sur les équilibres tautomères céto-énols de l'acétylacétate d'éthyle (EAA) et l'acétylacétone (AcAc) et leurs analogues cycliques l'acide tétronique (TA) et la dimédone (DIM) sont théoriquement étudiés. La présente étude montre que le tautomère cétonique le plus stable de l'EAA et l'AcAc correspond à la forme dicéto de trans, E, Z tandis que le tautomère énol le plus stable correspond à la structure dans laquelle l'énolisation se produit à côté du groupement carbonyle. Nos calculs mettent également en évidence que le tautomère cétonique du TA et DA préfère le dicéto trans, E, E, alors que le tautomère d'énol le plus stable correspond à la structure dans laquelle l'énolisation a lieu à côté du groupement carbonyle. Les énergies libres de Gibbs calculées indiquent que, dans les solvants polaires, les équilibres céto-énols de l'EAA et de l'AcAc sont déplacés vers le tautomère cétonique, tandis que les équilibres céto-énols du TA et DIM sont déplacés vers le tautomère énol. Les tendances de la variation des constantes d'équilibre par rapport à la variation de la polarité du solvant sont bien reproduites par les calculs B3LYP et MP2. La présente étude montre que le terme enthalpique est prédominant dans la détermination des constantes d'équilibre calculées. L'effet du terme entropique sur les énergies libres de Gibbs calculées s'avère très faible sur les équilibres tautomères céto-énol étudiés.*

---

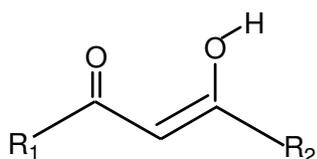
## 1. Introduction

La Tautomérisation est un cas particulier de l'isomérisation qui joue un rôle très important dans la chimie organique moderne, la biochimie, la chimie médicinale et la pharmacologie [1,2]. La compréhension des mécanismes réactionnels organiques et des processus biochimiques nécessite une compréhension avancée des phénomènes de tautomérisation. La tautomérisation se produit également dans les produits naturels comme dans le cas des acides aminés et des bases pyrimidines [3-4]. Pour cette raison, l'explication de la réactivité chimique et de l'activité biologique demeure plutôt difficile quand la tautomérisation a lieu [2-5]. En effet, il n'est pas toujours clair quelle est la forme tautomère responsable de la réactivité chimique ou de l'activité biologique quand le choix est entre un tautomère plus stable et un tautomère moins stable thermodynamiquement. Ce phénomène est remarqué dans le cas des propriétés acide-base ou redox, des propriétés électrophile-nucléophile ou même des propriétés géométriques.

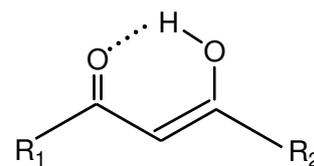
La tautomérisation céto-énolique est l'une des formes les plus étudiées des phénomènes prototropiques [1,6,7]. Ce réarrangement se produit dans différents systèmes tautomères contenant un ou plusieurs groupements carbonyles liés aux atomes de carbone  $sp^3$  avec un ou plusieurs atomes d'hydrogène. La proportion céto/énol dépend de divers facteurs tels que la température, la nature du solvant et des substituants [8, 9].

La tautomérisation céto-énol des composés 1,3-dicarbonylés à chaînes ouvertes de structure  $R_1C(O)-CH_2-C(O)R_2$  a suscité un grand intérêt au cours de ces dernières décennies [9-27]. La préférence du tautomère énolique ou cétonique dépend fortement de la nature des substituants  $R_1$  et  $R_2$  [11-13]. Selon des études spectroscopiques et théoriques en phase gazeuse, les deux tautomères énoliques sont prédominants lorsque  $R_1 = R_2 = H$  [14-16],  $CH_3$  [17,18],  $C(CH_3)_3$  [19] ou  $CF_3$  [20-21].

Il est important de noter que le tautomère énolique existe exclusivement sous la forme cis-énolique (b) qui correspond à une forme cyclique stabilisée par la liaison d'hydrogène intramoléculaire (voir **Schéma 1**) [22, 23].



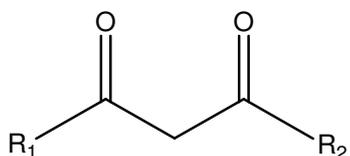
(a) open cis-enol,Z,Z



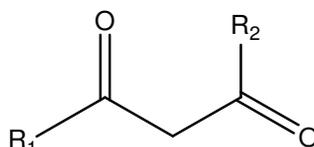
(b) closed cis-enol,Z,Z

**Schéma 1:** Formes énoliques des 1,3-dicarbonylés acycliques

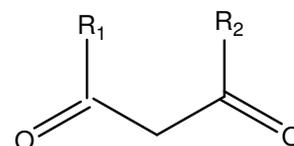
Inversement, les composés 1,3-dicarbonylés avec  $R_1 = R_2 = F$  [24], Cl [25],  $OCH_3$  [26] ou  $NH_2$  [27] existent préférentiellement sous la forme dicétonique. Selon les orientations relatives des groupements carbonyles, la forme dicétonique présente trois conformations possibles (**Schéma 2**). En absence des effets stériques des groupements  $R_1$  et  $R_2$ , la conformation cis-dicéto, Z,Z (forme U) possède les plus grandes répulsions coulombiennes des groupements carbonyles, suivie de la conformation trans-dicéto, E,Z (forme sickle) et la conformation trans-dicéto, E,E (forme W) devrait être la plus stable [28,29].



Cis-diceto , Z,Z  
(Forme U)



Trans diceto,E,Z  
(Forme Sickle)



Trans diceto,E,E  
(Forme W)

**Schéma 2:** les différentes formes dicétoniques

Les effets de solvation sur l'équilibre chimique ont été découverts en 1896, simultanément avec la découverte de la tautomérisation céto-énolique des composés 1,3-dicarbonylés [8-9]. L'étude de cet équilibre dans une série de solvants en utilisant la résonance magnétique nucléaire du proton (RMN) a été couramment utilisée comme une technique physico-chimique pendant de longues années [30-35]. Pour les composés  $\beta$ -dicarbonylés à chaînes ouvertes, les deux tautomères céto et énol sont fréquemment présents en solution dans différentes proportions [36-37] et le pourcentage de l'énolisation

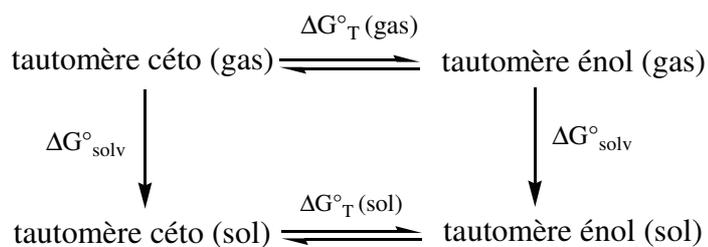
de ces composés est très sensible à la nature du solvant. Par exemple, Allen et al. [38] ont remarqué que le pourcentage d'énol du 2,4 pentandione (acétylacétone) est de 60% dans le diméthyle sulfoxyde (DMSO,  $\epsilon=46.7$ ), 75% dans l'acétone ( $\epsilon=20.7$ ), 94% dans  $\text{CCl}_4$  ( $\epsilon=2.23$ ) et 95% dans cyclohexane ( $\epsilon=2.02$ ). Il a été montré que les proportions d'énol augmentent dans les solvants non polaires aprotiques, c.-à-d., quand les liaisons d'hydrogène intermoléculaires avec le solvant sont absentes. En revanche, dans les solvants protiques polaires, les liaisons hydrogène intermoléculaires avec les molécules du solvant stabilisent la forme dicétonique.

De nombreuses études expérimentales et théoriques traitant la tautomérisation céto-énolique des composés 1,3 dicarboxylés cycliques peuvent être trouvées dans la littérature [39]. Le pourcentage de la forme énolique de ces  $\beta$ -dicarboxyles est fortement influencé par le solvant (en termes de polarité et en termes des propriétés donneur/accepteur de la liaison hydrogène [9]). Expérimentalement, il a observé que l'effet de solvant sur l'équilibre céto-énolique des  $\beta$ -dicarboxylés cycliques est opposé à celui de leurs analogues à chaînes ouvertes [39-44]. Notre but dans cette application est de mener une étude théorique de la structure moléculaire des  $\beta$ -dicarboxylés cycliques à savoir l'acide tétronique (TA) et la Dimédone (DIM) et de leurs composés analogues à chaînes ouvertes à savoir l'acide l'acétylacétate d'éthyle (EAA) et l'acétylacétone (AcAc) respectivement. Les effets de solvant sur l'équilibre céto-énolique de ces  $\beta$ -dicarboxylés sont mis en évidence par les énergies libres de Gibbs calculées avec deux méthodes quantiques (B3LYP et MP2) en utilisant le modèle de solvation PCM (Polarizable Continuum Model).

## 2. Méthodologie des calculs

Les calculs ont été effectués au moyen du programme G03W [45]. Les méthodes MP2 (Møller-Plesset de second ordre) [46] et B3LYP (Becke-3-parameter-Lee-Yang-Parr) [47] ont été utilisées. Les effets de solvants ont été étudiés en effectuant des calculs type

SCRf (Self-Consistent Reaction Field) en utilisant le modèle PCM sur les géométries optimisées en phase gazeuse. Ce modèle a été proposé par Tomasi et al. [48-50] et mis en application par Barone et al. [51-54]. Le cycle thermodynamique (**Schéma 3**) a été utilisé pour le calcul de la variation de l'énergie libre  $\Delta G^\circ_T$  et de la constante de l'équilibre tautomère  $K_T$ .



**Schéma 3:** Cycle thermodynamique utilisé pour le calcul de  $\Delta G^\circ_T$ .

avec :  $\Delta G^\circ_T(\text{solution}) = G^\circ_{\text{sol}}(\text{énol}) - G^\circ_{\text{sol}}(\text{céto})$  (1)

où  $G^\circ_{\text{sol}}(X) = G^\circ_{\text{gas}}(X) + \Delta G^\circ_{\text{solv}}(X)$  (2)

$$K_T = \exp(-\Delta G^\circ_T / RT) \quad (3)$$

avec  $R = 1.98709369 \text{ cal.mol}^{-1}.\text{K}^{-1}$  et  $T = 298.15 \text{ K}$

Dans l'équation 1,  $\Delta G^\circ_T(\text{solution})$  est calculée à partir de l'énergie libre de Gibbs en solution des espèces impliquées. Dans l'équation 2,  $G^\circ_{\text{sol}}(X)$  est directement calculée à partir des énergies libres en phase gazeuse de Gibbs ( $G^\circ_{\text{gas}}$ ) et des énergies libres de Gibbs de la solvation ( $\Delta G^\circ_{\text{solv}}$ ). La constante d'équilibre tautomérique  $K_T$  est définie dans l'équation 3.

### 3. Premier Cas: Etude des équilibres tautomères céto-énoliques de l'acétylacétate d'éthyle (EAA) et de l'acide tétronique (TA)

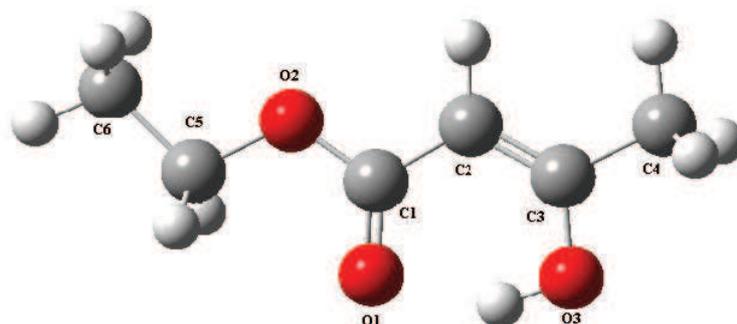
#### 3.1. Etude structurale de l'acétylacétate d'éthyle (EAA)

Le composé EAA contient deux groupements différents  $R_1=CH_3$  et  $R_2=OCH_2CH_3$ . Ceci rend les propriétés tautomères de ce composé très intéressantes. Selon les calculs B3LYP/6-311++G (d, p) et MP2/6-31+G(d,p), la forme cis-énol (b) dans laquelle l'énolisation a lieu à côté du groupement carbonyle  $CH_3C(OH)=CHC(O)OCH_2CH_3$  est plus stable de 10 kcal/mol que la forme cis-énol (a) dans laquelle l'énolisation est du côté du groupement ester (**Schéma 4**). Pour cette raison, seulement le tautomère énol le plus stable, représenté dans la **Figure 1a**, sera considéré dans cette présente étude théorique.

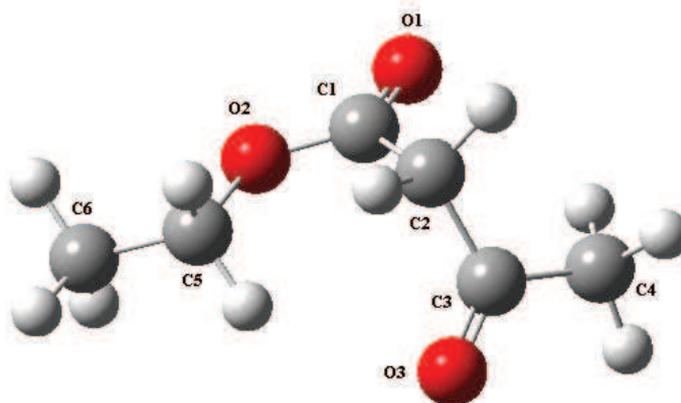


**Schéma 4:** les formes énoliques de l'acétylacétate d'éthyle

Pour déterminer toutes les conformations dicétoniques possibles, la surface d'énergie potentielle a été balayée au niveau de la théorie B3LYP/6-311++G (d, p). Les angles de torsion  $\theta$  ( $O_1C_1C_2C_3$ ) et  $\theta$  ( $O_3C_3C_2C_1$ ) ont été variés avec un pas de  $20^\circ$  et les autres paramètres ont été optimisés (voir **Figure 1b** pour la numérotation des atomes). Le minimum global sur la surface d'énergie potentielle a été obtenu pour les valeurs suivantes :  $\theta(O_1C_1C_2C_3) = 86^\circ$  et  $\theta(O_3C_3C_2C_1) = 99.5^\circ$ . La géométrie optimisée du composé EAA, donnée dans **Figure 1b**, montre que ce  $\beta$ -cétoester est stabilisé sous la forme sickle (en faucille). La même conformation sickle a été obtenue au niveau MP2/6-31+G(d,p).



(a) Le tautomère stable énol de l'acétylacétate d'éthyle



(b) Le tautomère stable énol de l'acétylacétate d'éthyle

**Figure 1.** Les géométries *B3LYP/6-311++G(d,p)* des tautomères stables de l'acétylacétate d'éthyle EAA

### 3.2. Etude théorique des effets de solvants sur le déplacement de l'équilibre céto-énol de l'acétylacétate d'éthyle (EAA)

Les résultats expérimentaux montrent qu'en présence de solvants polaires, EAA existe sous forme d'un mélange de deux tautomères dicéto et énol [55]. La constante d'équilibre  $K_T = [\text{énol}]/[\text{céto}] = 0.11$  mesurée dans l'éthanol montre une forte préférence pour la forme dicétonique. Cependant, dans les solvants non polaires et dans la phase gazeuse, le tautomère énol est prédominant (64% dans n-hexane par exemple) [56-58]. A partir des résultats expérimentaux donnés dans le **Tableau 1**, on peut conclure que les formes céto/énol du composé EAA sont à la fois présentes en solution avec des proportions mesurables. On note que la forme énolique est plus en plus favorisée en phase gazeuse et dans les solvants non polaires tel que l'hexane; tandis que la forme dicétonique est majoritaire dans l'éthanol et dans l'acétonitrile considérés comme milieux polaires.

**Tableau 1:** Variation des constantes d'équilibre expérimentales  $K_T$  de l'acétylacétate d'éthyle en fonction du solvant [56].

Medium	Gas phas ( $\epsilon = 1$ )	Benzène ( $\epsilon = 2.25$ )	Ethanol ( $\epsilon = 24.55$ )	Acétonitrile ( $\epsilon = 36.64$ )
Enol(%)/dicéto(%)	42.5/57.5	16.0/84.0	9.9/90.1	4.8/95.2
$K_T = [\text{énol}]/[\text{céto}]$	0.74	0.19	0.11	0.05

Afin de mettre en évidence les effets de solvation sur l'équilibre tautomère céto-énolique, nous avons considéré divers milieux avec des constantes diélectriques  $\epsilon$  variant de 1 (phase gazeuse) à 78.39 (eau). Le modèle de solvation PCM a été utilisé aux niveaux de la théorie B3LYP/6-311++G(d, p) et MP2/6-31+G(d, p). Les résultats sont récapitulés dans le **Tableau 2**.

**Tableau 2:** Les grandeurs thermodynamiques  $\Delta S_T$  (cal/mol. K<sup>-1</sup>),  $T\Delta S^\circ_T$  (kcal/mol),  $\Delta H^\circ_T$  (kcal/mol),  $\Delta G^\circ_T$  (kcal/mol) et la constante  $K_T$  de l'équilibre céto-énol de l'acétylacétate d'éthyle (EAA) calculées par la méthode B3LYP/6-311++G(d,p) et MP2/6-31+G(d,p) (entre parenthèses) dans divers solvants

	Gas phase ( $\epsilon=1$ )	Chloroform ( $\epsilon=4.90$ )	DMSO ( $\epsilon=46.70$ )	Water ( $\epsilon=78.39$ )
Acétylacétate d'éthyle				
$\Delta S^\circ_T$	-4.16 (-2.11)	-4.37 (-1.73)	-4.46 (-1.58)	-5.42 (-1.56)
$T\Delta S^\circ_T$	-1.24 (-0.63)	-1.30 (-0.51)	-1.32 (-0.47)	-1.58 (-0.47)
$\Delta H^\circ_T$	-10.73 (-7.28)	-8.64 (-5.36)	-7.72 (-4.53)	-5.18 (-2.22)
$\Delta G^\circ_{T^a}$	-9.49 (-6.66)	-7.34 (-4.84)	-6.40 (-4.06)	-3.60 (-1.75)
$\Delta G^\circ_{T^b}$	-9.49 (-6.66)	-8.21 (-4.65)	-7.57 (-3.74)	-5.08 (-0.63)
$K_T$	9.05 x 10 <sup>6</sup> (7.62 x 10 <sup>4</sup> )	1.04 x 10 <sup>6</sup> (0.26 x 10 <sup>4</sup> )	0.35 x 10 <sup>6</sup> (5.0 x 10 <sup>2</sup> )	0.53 x 10 <sup>4</sup> (2.90)

<sup>a</sup> calculée à partir de la formule  $\Delta G^\circ_T = \Delta H^\circ_T - T\Delta S^\circ_T$

<sup>b</sup> calculée à partir du cycle thermodynamique (Eq. (2)).

On remarque que les  $\Delta G^\circ_T$  B3LYP augmentent de -9.49 en phase gazeuse ( $\epsilon=1$ ) à -5.08 kcal/mol dans l'eau ( $\epsilon=78.39$ ).

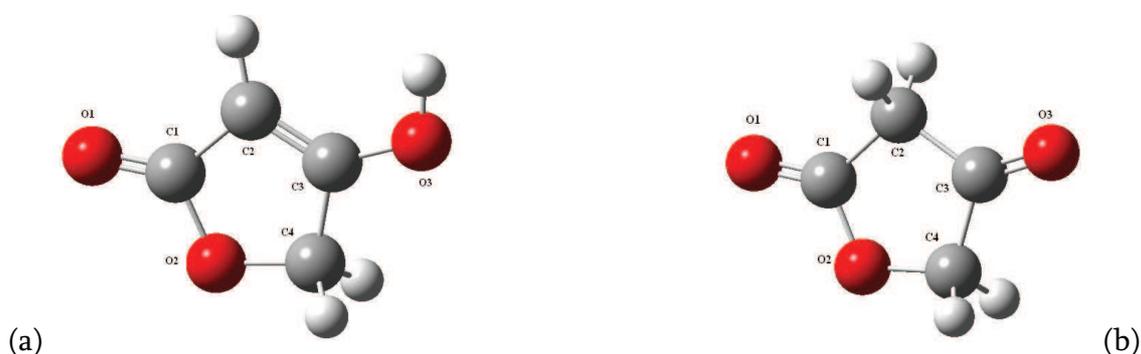
Les mêmes tendances ont été obtenues avec les calculs MP2. En effet, les énergies libres de Gibbs MP2 varient de -6.66 en phase gazeuse jusqu'à -0.63 kcal/mol dans l'eau.

Les valeurs des constantes d'équilibre calculées  $K_T$  (**Tableau 2**) diminuent en allant de la phase gazeuse vers la phase aqueuse. Ceci signifie que l'augmentation de la polarité du solvant est accompagnée d'un accroissement du pourcentage de la forme tautomère dicétonique. En conclusion, les enthalpies libres calculées (B3LYP et MP2) sont en bon accord avec les résultats expérimentaux qui stipulent que le tautomère dicéto est favorisé dans les solvants polaires alors que le tautomère énol est favorisé dans les milieux non polaires.

### 3.3. Etude structurale de l'acide tétronique (TA)

La forte énolisation du composé EAA a été attribuée à la stabilisation de la forme énolique par les liaisons hydrogène intramoléculaires. En revanche, ces interactions ne sont géométriquement possibles dans le cas de l'acide tétronique (TA). Par conséquent, dans la solution aqueuse, la forme énol de TA sera mieux stabilisée par la liaison hydrogène intermoléculaire avec les molécules du solvant, entraînant un déplacement de l'équilibre vers le tautomère énol. *Les résultats RMN de proton montrent que l'acide TA et ses dérivés simples existent seulement sous la forme trans-énol dans laquelle l'énolisation a lieu à côté du groupement carbonyle avec un rendement supérieur à 70% [42-43].* Les calculs théoriques B3LYP et MP2 indiquent que le tautomère trans-énol est favorisé par 6 kcal/mol environ en le comparant avec le tautomère énol dans laquelle l'énolisation est de côté du groupement ester (voir la **Figure 2a**). Pour cette raison, dans notre étude théorique seulement le tautomère énol le plus stable de l'acide tétronique TA (avec R=H) sera considéré.

Contrairement aux composés  $\beta$ -cétoesters à chaîne ouvertes, TA existe uniquement dans la conformation trans-dicéto, E,E (W shape). En fait, la géométrie optimisée au niveau B3LYP du TA (avec R=H), présentée dans la **Figure. 2b**, montre que ce  $\beta$ -cétoester est stabilisé sous la forme trans-dicéto avec les angles de torsion  $\theta(O_1C_1C_2C_3) = 179.94^\circ$  et  $\theta(O_3C_3C_2C_1) = 179.98^\circ$  (voir **Figure 2b** pour la numérotation des atomes).



**Figure 2:** Les géométries B3LYP/6-311++G(d,p) des tautomères stables trans-énol (a) et trans-dicéto (b) de l'acide tétronique TA.

### 3.4. Etude théorique des effets de solvants sur le déplacement de l'équilibre céto-énol de l'acide tétronique (TA)

Les acides tétroniques ( $\beta$ -cétoesters), qui sont un cas particulier des composés  $\beta$ -dicarboxylés cycliques, sont représentés traditionnellement, comme des dérivés de la forme énol [41-43]. En effet, le spectre UV de l'acide tétronique, dans l'éthanol aqueux, montre deux absorptions importantes  $\lambda_{\max}=223$  (C=O) et  $\lambda_{\max}=248$  nm (C=C) due à la présence du tautomère énol [44]. L'expérience [41-43,59-62] montre également que la tautomérisation de l'acide tétronique est accélérée par les solvants polaires. Ce processus pourrait devenir plus favorable dans les milieux condensés, en particulier dans les solvants fortement polaires. Les grandeurs thermodynamiques  $\Delta S_T$ ,  $\Delta H^\circ_T$ ,  $\Delta G^\circ_T$  et de la constante  $K_T$  de l'équilibre céto-énol de l'acide tétronique (TA) calculées par les méthodes B3LYP/6-311++G(d,p) et MP2/6-31+G(d,p) dans divers solvants sont donnés dans le **Tableau 3**.

**Tableau 3:** Les grandeurs thermodynamiques  $\Delta S_T$  (cal/mol. K<sup>-1</sup>),  $T\Delta S^\circ_T$  (kcal/mol),  $\Delta H^\circ_T$  (kcal/mol),  $\Delta G^\circ_T$  (kcal/mol) et la constante  $K_T$  de l'équilibre céto-énol de l'acide tétronique (TA) calculées par la méthode B3LYP/6-311++G(d,p) et MP2/6-31+G(d,p) (entre parenthèses) dans divers solvants

	Gas phase ( $\epsilon=1$ )	Chloroform ( $\epsilon=4.90$ )	DMSO ( $\epsilon=46.70$ )	Water ( $\epsilon=78.39$ )
Acide tétronique				
$\Delta S^\circ_T$	-2.42 (-2.98)	-2.40 (-2.99)	-2.36 (-3.00)	1.94 (2.49)
$T\Delta S^\circ_T$	-0.73 (-0.89)	-0.71 (-0.89)	-0.70 (-0.89)	0.58 (0.74)
$\Delta H^\circ_T$	4.30 (5.90)	3.06 (4.54)	2.45 (3.88)	2.80 (3.86)
$\Delta G^\circ_{T^a}$	5.03 (6.79)	3.77 (5.43)	3.15 (4.77)	2.22 (3.12)
$\Delta G^\circ_{T^b}$	5.03 (6.79)	3.68 (5.71)	2.92 (5.06)	2.34 (4.91)
$K_T$	$2.05 \times 10^{-4}$ ( $1.07 \times 10^{-5}$ )	$2.01 \times 10^{-3}$ ( $6.52 \times 10^{-5}$ )	$7.24 \times 10^{-3}$ ( $1.95 \times 10^{-4}$ )	$1.92 \times 10^{-2}$ ( $2.51 \times 10^{-4}$ )

<sup>a</sup> calculée à partir de la formule  $\Delta G^\circ_T = \Delta H^\circ_T - T\Delta S^\circ_T$

<sup>b</sup> calculée à partir du cycle thermodynamique (Eq. (2)).

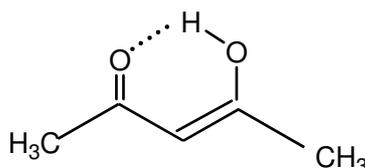
Les calculs B3LYP montrent que la variation d'énergie libre de Gibbs de l'équilibre céto-énoïl  $\Delta G^{\circ}_T$  est de l'ordre de 5.03 kcal/mol en phase gazeuse et de 2.34 kcal/mol en solution aqueuse. Par conséquent, les solvants polaires réduisent l'énergie  $\Delta G^{\circ}_T$ , indiquant que la polarité du solvant favorise l'énoïlisation du TA. Le calcul de la constante de l'équilibre  $K_T$  montre également que la forme énoïl est plus favorisée dans les solvants polaires. Les mêmes tendances sont bien reproduites par la méthode MP2.

Comme nous pouvons voir dans les **Tableaux 2 et 3**, les valeurs calculées du terme  $T\Delta S^{\circ}_T$  par la méthode B3LYP sont généralement faibles ( $\sim 0.70$  kcal.mol<sup>-1</sup> et 1.30 kcal.mol<sup>-1</sup> pour EAA et TA respectivement) par rapport aux valeurs de  $\Delta H^{\circ}_T$ . Ces résultats indiquent que l'effet de l'entropie sur l'énergie libre de Gibbs est très minime et le terme  $T\Delta S^{\circ}$  peut être négligé. Ainsi, le terme enthalpique  $\Delta H^{\circ}$  est prédominant dans la détermination des constantes des deux équilibres céto-énoïques étudiés.

#### ***4. Second Cas: Etude des équilibres tautomères céto-énoïques de l'acétylacétone (AcAc) et de la dimédone (DIM)***

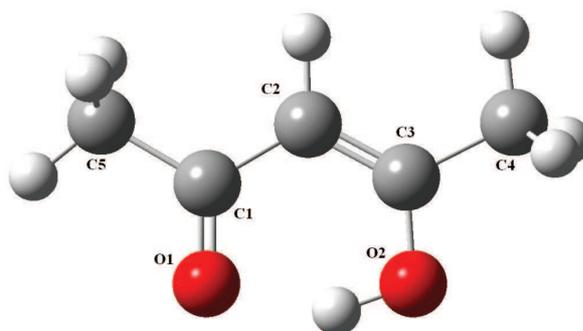
##### ***4.1. Etude structurale de l'Acétylacétone (AcAc)***

Le pentadione ou l'acétylacétone est un composé 1,3 dicarboxylé symétrique qui contient deux substituants identiques  $R_1 = R_2 = \text{CH}_3$ , ceci indique qu'une seule énoïlisation est possible  $\text{CH}_3\text{C}(\text{OH})=\text{CHC}(\text{O})\text{CH}_3$ . Les calculs B3LYP/6-311++G(d, p) et MP2/6-31+G(d, p) montrent que la forme cis-énoïl, représentée dans le **Schéma 5** est stabilisée par les liaisons hydrogène intramoléculaires (**Figure 3a**).

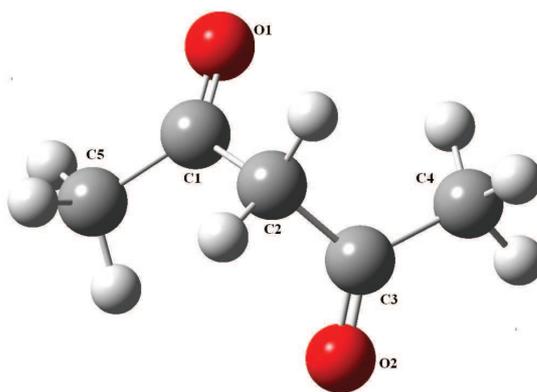


**Schéma 5:** le tautomère énoïl de l'acétylacétone

Pour trouver toutes les conformations dicétoniques possibles, la surface d'énergie potentielle a été balayée au niveau B3LYP/6-311++G(d,p). Les angles de torsion  $\theta$  ( $O_1C_1C_2C_3$ ) et  $\theta$  ( $O_3C_3C_2C_1$ ) ont été variés avec un pas de  $20^\circ$  et tous les autres paramètres ont été optimisés (voir la **Figure 3b** pour la numérotation des atomes). Le minimum global sur la surface d'énergie potentielle a été obtenu pour les valeurs suivantes :  $\theta(O_1C_1C_2C_3)=89.25^\circ$  et  $\theta(O_2C_3C_2C_1)=89.32^\circ$ . La géométrie optimisée de l'acétylacétone, donnée dans la **Figure 3b**, montre que ce  $\beta$ -dicarboxylé est stabilisé sous la forme sickle (en faucille). Une géométrie similaire a été obtenue par la méthode MP2/6-31+G(d, p).



(a) Le tautomère énol le plus stable de l'acétylacétone

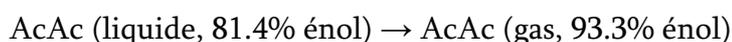


(b) Le tautomère dicéto le plus stable de l'acétylacétone

**Figure 3:** Les géométries B3LYP/6-311++G(d,p) optimisées des tautomères énol et dicéto de l'acétylacétone

#### 4.2. Etude théorique des effets de solvants sur le déplacement de l'équilibre céto-énol de l'Acétylacétone (AcAc)

Il est bien connu que l'acétylacétone (AcAc), ( $R_1=R_2=CH_3$ ) existe dans deux formes tautomères. Selon de nombreuses études RMN [63-65,22] le tautomère énol est prédominant en phase liquide pur avec un pourcentage de 81% [22] et de 91% dans la solution  $CCl_4$  [63]. Temprado [66] note que le tautomère énol est plus volatil et donc sa concentration dans la phase gazeuse devrait être encore importante que celle dans la phase liquide. En utilisant l'enthalpie expérimentale de vaporisation ( $10.0 \pm 0.1$  kcal/mol) à 25°C, Irving [67] a déterminé la composition tautomérique en phase gazeuse de l'acétylacétone suivante:



Les études  $^1H$ -NMR [68] montrent que la forme énol d'AcAc prédomine dans le liquide pur, en solution et dans la phase gazeuse à toutes les températures. La variation de l'énergie libre de Gibbs de l'équilibre tautomère en phase gazeuse est de l'ordre  $\Delta G^\circ_{298} = 2.20$  kcal/mol [68] qui correspond à un pourcentage de la forme dicéto de l'ordre 2.5% à 298 K. Les constantes d'équilibre tautomère céto-énol dans la phase gazeuse ont été obtenues à partir des spectres de photoélectrons à différentes températures [69]. Les résultats obtenus montrent que les concentrations en énol sont de 73% à 25°C, de 60% à 100°C et environ de 50% à 175°C.

A partir des résultats expérimentaux donnés dans le **Tableau 4**, on déduit que les formes de céto/énol du composé AcAc sont à la fois présentes en solution avec des proportions mesurables. La forme énolique est prédominante en phase gazeuse et dans les solvants non polaires tels que le benzène ( $\epsilon = 2.25$ , 89%) et l'hexane (95%); tandis que la forme dicétonique est plus en plus favorisée dans l'éthanol (17.8%) et dans l'acétonitrile (38.5%) considérés comme solvants polaires [56-58].

**Tableau 4:** Variation des constantes d'équilibre expérimentales  $K_T$  de l'acétylacétone en fonction du solvant [56].

Medium	Gaz ( $\epsilon=1$ )	Benzène ( $\epsilon=2.25$ )	Ethanol ( $\epsilon=24.55$ )	Acétonitrile ( $\epsilon=36.64$ )
Enol (%)/dicéto (%)	92.1/7.8	89.0/11.0	82.2/17.8	61.5/38.5
$K_T = [\text{énol}]/[\text{céto}]$	11.7	8.1	4.6	1.6

Afin de mettre en évidence les effets de solvant sur l'équilibre tautomère céto-énolique, nous avons considéré dans notre étude divers milieux avec des constantes diélectriques variant de  $\epsilon=1$  (phase gazeuse) à  $\epsilon=78.39$  (eau). Les grandeurs thermodynamiques  $\Delta S_T$ ,  $\Delta H^\circ_T$ ,  $\Delta G^\circ_T$  et de la constante  $K_T$  de l'équilibre céto-énol de l'acétylacétone (AcAc) calculées par les méthodes B3LYP/6-311++G(d,p) et MP2/6-31+G(d,p) dans divers solvants sont donnés dans le **Tableau 5**.

**Tableau 5:** Les grandeurs thermodynamiques  $\Delta S_T$  (cal/mol.  $K^{-1}$ ),  $T\Delta S^\circ_T$  (kcal/mol),  $\Delta H^\circ_T$  (kcal/mol),  $\Delta G^\circ_T$  (kcal/mol) et la constante  $K_T$  de l'équilibre céto-énol de l'acétylacétone (AcAc) calculées par la méthode B3LYP/6-311++G(d,p) et MP2/6-31+G(d,p) (entre parenthèses) dans divers solvants

	Gaz ( $\epsilon=1$ )	Chloroform ( $\epsilon=4.90$ )	DMSO ( $\epsilon=46.70$ )	Water ( $\epsilon=78.39$ )
Acétylacétone				
$\Delta S^\circ_T$	-8.57 (-6.56)	-8.97 (-5.47)	-9.25 (-5.31)	-9.57 (-5.44)
$T\Delta S^\circ_T$	-2.55 (-1.96)	-2.67 (-1.63)	-2.76 (-1.58)	-2.85 (-1.62)
$\Delta H^\circ_T$	-5.60 (-1.99)	-4.59 (-1.16)	-4.24 (-0.88)	-2.08 (0.90)
$\Delta G^\circ_{T^a}$	-3.04 (-0.03)	-1.92 (0.47)	-1.48 (0.70)	0.77 (2.52)
$\Delta G^\circ_{T^b}$	-3.04 (-0.03)	-2.76 (0.57)	-2.75 (0.72)	-0.76 (3.0)
$K_T$	169.22 (1.05)	105.49 (0.38)	103.72 (0.30)	3.6 (6.3 $10^{-3}$ )

<sup>a</sup> calculée à partir de la formule  $\Delta G^\circ_T = \Delta H^\circ_T - T\Delta S^\circ_T$

<sup>b</sup> calculée à partir du cycle thermodynamique (Eq. (2)).

Les énergies libres de Gibbs B3LYP augmentent de -3.04 en phase gazeuse ( $\epsilon=1$ ) à 0.77 kcal/mol en phase aqueuse ( $\epsilon=78.39$ ) (voir le **Tableau 5**). Les mêmes tendances ont été obtenues avec les calculs MP2 de l'énergie libre. En effet, les enthalpies MP2 varient de -0.03 en phase gazeuse jusqu'à 3.0 kcal/mol dans l'eau. Les constantes tautomères calculées  $K_T$  (**Tableau 5**) diminuent en passant de la phase gazeuse à la phase aqueuse. L'augmentation de la polarité du solvant est donc accompagnée par un accroissement du pourcentage du tautomère dicéto.

En conclusion, les énergies libres de Gibbs calculées (B3LYP et MP2) sont en bon accord avec les résultats expérimentaux qui stipulent que la forme dicétonique est favorisée dans les solvants polaires tandis que la forme énol est favorisée dans des milieux non polaires.

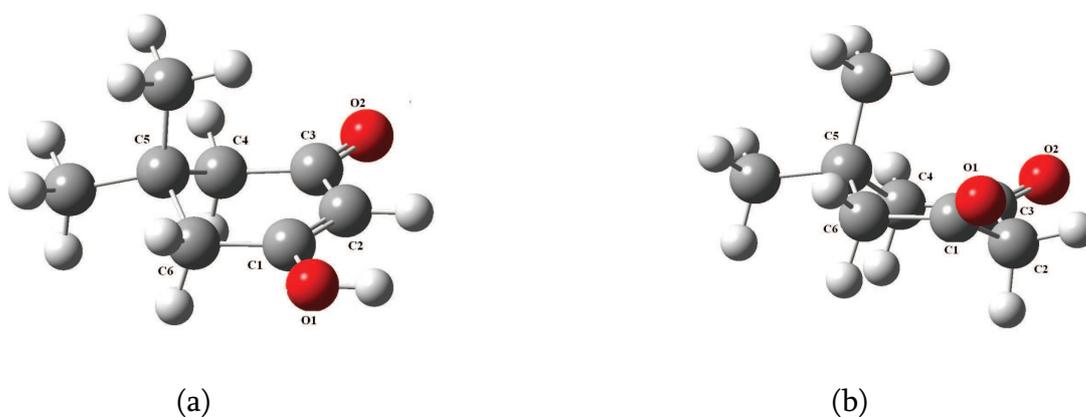
#### 4.3. Etude structurale de la dimédone

Dans les composés  $\beta$ -dicarbonylés cycliques tels que la dimédone (**Schéma 6**), les ponts hydrogène entre l'énol et le groupement carbonyle n'est pas possible pour des raisons géométriques et stériques. Par conséquent, dans la solution aqueuse, la forme énol du dimédone est mieux stabilisée par la liaison hydrogène intermoléculaire avec le solvant, entraînant un déplacement de l'équilibre vers le tautomère énol. Les résultats expérimentaux montrent que la dimédone existe seulement sous la forme trans-énol. (Voir la **Figure 4a**).

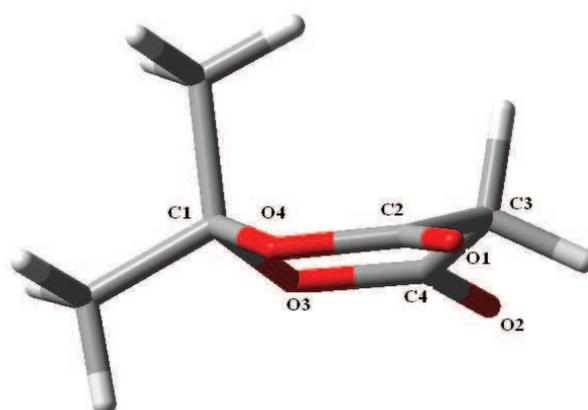


*Schéma 6*

La géométrie optimisée par la méthode B3LYP de la dimédone, présentée dans la **Figure 4b**, montre que le tautomère dicéto existe seulement sous la forme trans-dicéto, E, E avec les angles de torsion  $\theta(O_1C_1C_2C_3) = 138.13^\circ$  et  $\theta(O_2C_3C_2C_1)=138.13^\circ$  (voir la **Figure 4b** pour la numérotation des atomes). Les calculs B3LYP et MP2 effectués sur la structure de la dimédone, montrent également que ce 1,3 dicarbonylé possède la conformation chaise contrairement à l'acide de Meldrum, très similaire à la dimédone, qui préfère la forme bateau (**Figure 5**).



**Figure 4:** Les géométries B3LYP/6-311++G(d,p) optimisées des tautomères stables trans-enol (a) et trans-dicéto (b) de la dimédone



**Figure 5:** La géométrie B3LYP/6-311++G(d,p) du tautomère dicéto de l'acide de Meldrum

#### 4.4. Etude théorique des effets de solvants sur le déplacement de l'équilibre céto-énol de la dimédone

L'étude spectrale  $^1\text{H}$  RMN dans  $\text{CDCl}_3$  de la dimédone réalisée par Yogev et al. [37] montre un rapport céto/énol = 59/41. Tandis que Cremlyn et al. [70] ont trouvé un rapport de 64/36. Dans une solution de chloroforme  $\text{CHCl}_3$  ( $\epsilon=4.90$ ), le tautomère dicéto est prédominant [70]. Dans un milieu très concentré de la dimédone, la forme énolique peut exister sous forme de dimères ou de polymères.

Plusieurs études expérimentales [71-73] ont montré que dans une solution de DMSO ( $\epsilon=46.70$ ), la dimédone préfère exclusivement la forme énol. Les résultats  $^{13}\text{C}$ -RMN indique que la forme dicéto possède un pourcentage de l'ordre de 1.2%. Par contre une étude  $^1\text{H}$ -RMN révèle que le tautomère dicéto est pratiquement absent ( $< 1\%$ ) [70].

D'un autre coté il a été expérimentalement montré que la tautomérisation de la dimédone est accélérée par les solvants polaires. Ce processus devient plus favorable dans les solvants fortement polaires (voir le **Tableau 6**).

**Tableau 6:** Variation des constantes d'équilibre  $K_T$  de la dimédone en fonction du solvant [56,57].

Solvant	Cyclohexane	DMSO	$\text{H}_2\text{O}$
	( $\epsilon=2.02$ )	( $\epsilon=46.7$ )	( $\epsilon=78.39$ )
% Enol	< 2.0	95.0	> 95.0

Les grandeurs thermodynamiques  $\Delta S_T$ ,  $\Delta H^{\circ}_T$ ,  $\Delta G^{\circ}_T$  et de la constante  $K_T$  de l'équilibre céto-énol de la dimédone calculées par les méthodes B3LYP/6-311++G(d,p) et MP2/6-31+G(d,p) dans divers solvants sont donnés dans le **Tableau 7**.

**Tableau 7:** Les grandeurs thermodynamiques  $\Delta S_T$  ( $\text{cal}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$ ),  $T\Delta S^\circ_T$  (kcal/mol),  $\Delta H^\circ_T$  (kcal/mol),  $\Delta G^\circ_T$  (kcal/mol) et la constante  $K_T$  de l'équilibre céto-énol de la dimédone (DIM) calculées par la méthode B3LYP/6-311++G(d,p) et MP2/6-31+G(d,p) (entre parenthèses) dans divers solvants

	Gaz ( $\epsilon=1$ )	Chloroform ( $\epsilon=4.90$ )	DMSO ( $\epsilon=46.70$ )	Water ( $\epsilon=78.39$ )
Dimédone				
$\Delta S^\circ_T$	-1.24 (-1.0)	-1.27 (-0.56)	-1.27 (-0.42)	-1.37 (-0.02)
$T\Delta S^\circ_T$	-0.37 (-0.30)	-0.37 (-0.17)	-0.37 (-0.12)	-0.40 (-0.005)
$\Delta H^\circ_T$	2.90 (5.30)	2.30 (4.65)	2.05 (4.38)	1.40 (3.33)
$\Delta G^\circ_{T^a}$	3.26 (5.60)	2.67 (4.82)	2.42 (4.50)	1.80 (3.33)
$\Delta G^\circ_{T^b}$	3.26 (5.61)	2.52 (5.61)	2.12 (4.91)	1.20 (4.04)
$K_T$	$4.10 \times 10^{-3}$ ( $7.72 \times 10^{-5}$ )	$1.42 \times 10^{-2}$ ( $7.72 \times 10^{-5}$ )	$2.79 \times 10^{-2}$ ( $2.52 \times 10^{-4}$ )	$13.20 \times 10^{-2}$ ( $1.09 \times 10^{-3}$ )

<sup>a</sup> calculée à partir de la formule  $\Delta G^\circ_T = \Delta H^\circ_T - T\Delta S^\circ_T$

<sup>b</sup> calculée à partir du cycle thermodynamique (Eq. (2)).

Les calculs B3LYP prouvent que la variation d'énergie libre de Gibbs de l'équilibre céto-énol  $\Delta G^\circ_T$  est de l'ordre de 3.26 kcal/mol dans la phase gazeuse ( $\epsilon=1$ ) et de 1.20 kcal/mol dans l'eau ( $\epsilon=78.39$ ) (**Tableau 7**). Par conséquent, les solvants polaires réduisent l'énergie  $\Delta G^\circ_T$ , indiquant que la polarité du solvant favorise l'énolisation de la dimédone. La constante de l'équilibre calculée  $K_T$  indiquent également que la forme énol est favorisée dans les solvants polaires. Nous notons que les tendances de la variation des constantes d'équilibre par rapport au changement de la polarité du solvant sont bien reproduites par les deux méthodes B3LYP et MP2.

## 5. Conclusions

Etude structurale théorique sur L'acétylacétate d'éthyle (EAA) montre que le tautomère dicétonique le plus stable est le dicéto trans, E, Z (forme faucille); tandis que le tautomère énol le plus stable correspond à la structure cis-énol dans laquelle l'énolisation est de coté du groupement carbonyle.

Dans le cas de l'acide tétronique TA le tautomère dicétonique est de la forme trans, E, E tandis que le tautomère trans-énol le plus stable correspond à la structure dans laquelle l'énolisation est adjacente au groupement carbonyle.

Les calculs de la variation de l'enthalpie libre montrent que l'équilibre de céto-énol du composé EAA est déplacé vers le tautomère dicétonique dans les solvants polaires; tandis que l'équilibre céto-énol du TA est déplacé vers le tautomère énolique dans les solvants polaires. Ces tendances sont en bon accord avec des résultats expérimentaux. Nos calculs révèlent également que la le terme entropique dans le calcul de la variation l'énergie libre de Gibbs des équilibre de l'EAA et du TA est minime et par conséquent le terme enthalpique est prédominante dans la détermination de la constante d'équilibre.

Comme tous les composés  $\beta$ -dicarboxylés acyclique la forme dicétonique la plus stable de l'acétylacétone (AcAc) est la forme trans E, Z (forme faucille); tandis que l'énol le plus stable (par les liaisons hydrogènes intramoléculaires) correspond à la structure cis-énol.

La dimédone (DIM) existe sous deux tautomères stables : la forme dicétonique trans E, E et la forme trans-énol stabilisée par les liaisons hydrogènes intermoléculaires.

Les calculs de la variation de l'enthalpie libre montrent que l'équilibre de céto-énol du composé AcAc est déplacé vers le tautomère dicétonique dans les solvants polaires; tandis que l'équilibre céto-énol de la dimédone est déplacé vers le tautomère énolique dans les mêmes types de solvants polaires. Les tendances trouvées sont en bon accord avec des résultats expérimentaux.

## Bibliographie

- [1] (a) J. March, "Advanced Organic Chemistry, Reaction, Mechanism and Structure", Wiley, New York, 1993.
- (b) F.A. Carey, R.J. Sundberg, "Advanced Organic Chemistry", Plenum Press, New York, 1993.
- [2] (a) W. Saenger, "Principles of Nucleic Acid Structure", Springer, New York, 1994.
- (b) W.O. Foye, T.L. Lemke, D.A. Williams, "Principles of Medicinal Chemistry" Williams and Wilkins, Baltimore, MD, 1995.
- (c) H. Weinstein, D. Chou, C.L. Johnson, S. Kang, J.P. Green, *Mol. Pharmacol.* 14 (1976) 738.
- (d) K.J. Watling, "The RBI Handbook of Receptor Classification and Signal Transduction", RBI, Natick, MA, 1998.
- (e) A.F. Pozharskii, A.T. Soldatenkov, A.R. Katritzky, "Heterocycles in Life and Society" Wiley, New York, 1997.
- [3] (a) E.D. Raczyn'ska, M.K. Cyran'ski, M. Darowska, T. Rudka, *Targets Heterocycl. Chem.* 4 (2000) 327.
- (b) E.D. Raczyn'ska, M. Darowska, M. Makowski, *Trends Heterocycl. Chem.* 9 (2003) 167.
- (c) A.A. Bliznyuk, H.F. Schaefer, I.J. Amster, *J. Am.Chem. Soc.* 115 (1993) 5149.
- (d) R.M. Day, C. J. Thalhauser, J.L. Sudmeier, M.P. Vincent, , E.V. Torchilin, D.G. Sanford, C.W. Bachovchin, W.W. Bachovchin, *Protein Sci.* 12 (2003) 794.
- (e) P. Huda'ky, A. Perczel, *J. Phys. Chem. A* 108 (2004) 6195.
- (f) J. Rak, P.J.M. Skurski, Simons, Gutowski, *J. Am. Chem. Soc.* 123 (2001) 11695.
- [4] (a) J.S. Kwiatkowski, B. Pullman, *Adv. Heterocycl. Chem.* 18 (1975) 200.
- (b) J.S. Kwiatkowski, W.B. Person, "In Theoretical Biochemistry and Molecular Biology", D.L. Beveridge, R. Lavery, Eds, Academic Press, New York, p. 153, 1990.
- (c) A. R. Katritzky, M. Karelson, *J. Am. Chem. Soc.* 113 (1991) 1561.
- (d) G.H. Roehring, N.A. Oyler, L. Adamowicz, *Chem. Phys. Lett.* 225 (1994) 265.

- (e) J. Sjöponer, P. Hobza, *J. Phys. Chem.* 98 (1994) 3161.
- (f) M. Kabelac, M. Kratochvil, J. Sjöponer, P. Hobza, *J. Biomol. Struct. Dyn.* 117 (2000) 1077.
- [5] (a) K. Smith, "Porphyrins and Metalloporphyrins", Elsevier, Amsterdam, 1975.
- (b) *The Porphyrins*, Dolphin, D., Academic Press: New York, Vols. 1-7, 1978-1979.
- (c) J.L. Sessler, S.J. Weighorn, "Expanded, Contracted, and Isomeric Porphyrins", Elsevier, Oxford, 1997.
- (d) K.M. Kadish, K.M. Smith, R. Guilard, "The Porphyrin Handbook", Academic Press, San Diego, CA, Vol. 2, 1999.
- [6] (a) S. Patai, "The Chemistry of the C=O group", Wiley, London, 1966.
- (b) Z. Rappoport, "The Chemistry of Enols", Wiley, New York, 1990.
- [7] (a) A. Kresge, *J. Chem. Soc. Rev.* 25 (1996) 275.
- (b) M. Karelson, U. Maran, A.R. Katritzky, *Tetrahedron* 52 (1996) 11325.
- (c) Z. Rappoport, J. Frey, M. Sigalov, E. Rochlin, *Pure Appl. Chem.* 69 (1997) 1933.
- (d) A. Fontana, P. De Maria, G. Siani, M. Pierini, S. Cerritelli, R. Ballini, *Eur. J. Org. Chem.* 65 (2000) 1637.
- (e) E. Iglesias, *Curr. Org. Chem.* 8 (2004) 1.
- [8] L. Claisen, *Liebigs Ann. Chem.* 291 (1896) 25.
- [9] C. Reichard, "solvents and solvents effects in organic chemistry", New York, p.2, 1988.
- [10] J. Emsley, "Structure and Bonding", Springer, Berlin, Vol. 57, p. 147, 1984.
- [11] M. Huang, W. Graham Richards, Guy H. Grant, *J. Phys. Chem. A* 109 (2005) 5198.
- [12] N.V. Belova, G.V. Girichev, S.A. Shlykov, and H. Oberhammer, *J. Org. Chem.* 71 (2006) 5298.
- [13] (a) N.V. Belova, H. Oberhammer, G.V. Girichev, S.A. Shlykov, *J. Phys. Chem. A* 111 (2007) 2248.
- (b) N.V. Belova, V.V. Sliznev, H. Oberhammer, G.V. Girichev, *J. Mol. Struct.* 978 (2010) 282.

- [14] S. L. Baughcum, R.W. Duerst, W.F. Rowe, *J. Am. Chem. Soc.* 103 (1981) 6296.
- [15] S. L. Baughcum, Z. Smith, E.B. Wilson, Duerst, *J. Am. Chem. Soc.* 106 (1984) 2260.
- [16] P. Turner, S.L. Baughcum, S. L. Coy, Z. Smoth, *J. Am. Chem. Soc.* 106 (1984) 2265.
- [17] A.H. Lowrey, C. George, P. D'Antonio, J. Karle, *J. Am. Chem. Soc.* 93 (1971) 6399.
- [18] A.L. Andreassen, S.H. Bauer, *J. Mol. Struct.* 12 (1972) 381.
- [19] K. Iijima, A. Ohnogi, S. Shibata, *J. Mol. Struct.* 156 (1987) 111.
- [20] N.I. Giricheva, G.V. Girichev, S.B. Lapshina, N.P. Kuzmina, *Zh. Strukt. Khim.* 41 (2000) 58.
- [21] M. Moriyasu, A. Kato, Y. Hashimoto, *J. Chem. Soc. Perkin Trans. 2* (1986) 515.
- [22] J. Burdett, M. Rogers, *J. Am. Chem. Soc.* 86 (1971) 2105.
- [23] D.P. Chong, C-Han Hu, *J. Electron Spectrosc. Relat. Phenom.* 94 (1998) 181.
- [24] A.D. Jin, H.G. Mack, A. Waterfeld, H. Oberhammer, *J. Am. Chem. Soc.* 113 (1991) 7874.
- [25] H.G. Mack, H. Oberhammer, C.D. Della Vedova, *J. Mol. Struct.* 346 (1995) 51.
- [26] N.V. Belova, H. Oberhammer, G.V. Girichev, *J. Mol. Struct.* 689 (2004) 255.
- [27] Y. Kawashima, "10<sup>th</sup> European Symposium on Gas Electron Diffraction", St. Petersburg, Russia, 2003.
- [28] I. Lee, I.S. Han, C.K. Kim and H.W. Lee, *Bull. Korean Chem. Soc.* 24 (2003) 1141.
- [29] G.V. Perez, A.L. Perez, *J. Chem. Educ.* 77 (2000) 910.
- [30] (a) J. Toullec, "in *The Chemistry of Enols*", Z. Rappoport, Wiley, Chichester, U. K., p. 323, 1990.
- (b) S. Forse'n, M. Nilson, "in *The Chemistry of the Carbonyl Group*", J. Zabicky, Interscience Publishers, New York, Vol. 2, p. 198, 1970.
- (c) O. Neylands, "in *Structure and Tautomerism of  $\beta$ -Dicarbonyl Compounds*", E. Gudrience, Zinatne, Riga, Latvia, p.142, 1976.
- [31] (a) E. Iglesias, *J. Phys. Chem.* 100 (1996) 12592.
- (b) E. Iglesias, *J. Chem. Soc., Perkin Trans. 2* (1997) 431.
- (c) E. Iglesias, *New J. Chem.* 29 (2005) 457.

- (d) E. Iglesias, *J. Phys. Chem. B* 105 (2001) 10287.
- [32] (a) E. Iglesias, V. Ojea-Cao, L. Garcia-Rio, J.R. Leis, *J. Org. Chem.* 64 (1999) 3954.  
(b) E. Iglesias, *J. Org. Chem.* 65 (2000) 6583.  
(c) E. Iglesias, *J. Inclusion Phenom. Macrocyclic Chem.* 52 (2005) 55.
- [33] L.W. Reeves, *Can. J. Chem.* 35 (1957) 1351.
- [34] E.J. Drexler, K.W. Field, *J. Chem. Educ.* 53 (1976) 392.
- [35] E.W. Garland, J.W. Nibler, D.P. Shoemaker, "Experiments in Physical Chemistry", McGraw-Hill, New York, p. 453, 2003.
- [36] A.G. Williams, G.B. Butler, *J. Org. Chem.* 45 (1980) 1232.
- [37] D. Veierov, T. Bercovici, E. Fischer, Y. Mazur, A. Yogevev, *J. Am. Chem. Soc.* 99 (1977) 2723.
- [38] G. Allen, R.A. Dwek, *J. Chem. Soc. B* 161 (1966).
- [39] (a) C.H. Luchmuller, T. Maldacker, M. Cefola, *Anal. Chim. Acta* 48 (1969) 139.  
(b) E. Mendez et al., *J. Phys. Chem. C* 111 (2007) 3369.  
(c) F. Zuccarello, G. Buemi, C. Gandolfo, A. Contino, *Spectro. Acta Prat A* 39 (2003) 139.  
(d) J. Bojarski et al., "Advances in Heterocyclic Chemistry", Academic Press, New York, vol 38, p. 231, 1985.  
(e) M. Eckert-Maksic, L. Maksimovic, *J. Mol. Struct. (THEOCHEM)* 153 (1987) 121.  
(f) C.K. Skylaris, O. Igglessi-Markopoulou, A. Detsi, J. Markopoulos, *Chem. Phys.* 293 (2003) 355.
- [40] Schwarzenbach G, Felder E, *Helv Chim Acta* 27 (1944) 1044.
- [41] Pévet I, Meyer C, Cossy J, *Tetrahedron Lett* 42 (2001) 5215.
- [42] A.R. Rajaram, *L. Pu, Org. Lett.* 8 (2006) 2019.
- [43] C. Willis, E. Bodio, Y. Bourdreux, C. Billaud, T. Le Gall, C. Mioskowski, *Tetrahedron Lett.* 48 (2007) 6421.
- [44] Brodyck, J. L. Royles, *Chem. Rev.* 95 (1995) 1981.
- [45] (a) M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R.

Cheeseman, J. A. Montgomery, Jr., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J. J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, and J.A. Pople, Gaussian 03, revision E.01; Gaussian, Inc.: Wallingford CT, 2004.

(b) Frisch, Æ; Frisch, M.; Trucks, G. W. Gaussian 03 User's Reference; Gaussian Inc.: Wallingford, CT, 2003.

- [46] C. Møller, M.S. Plesset, *Phys. Rev.* 46 (1934) 681.
- [47] (a) W. Koch, Max C.A. Holthausen, "Chemist's Guide to Density Functional Theory", Wiley, VCH, Weinheim, Germany, 2001.
- (b) A.D. Becke, *J. Chem. Phys.* 98, (1993) 5648.
- (c) A.D. Becke, *Phys. Rev. A.* 38 (1988) 3098. (d) C. Lee, W. Yang, R.G. Parr, *Phys. Rev. B* 37, 785 (1988).
- (d) B. Miehlich, A. Savin, H. Stoll, H. Preuss, *Chem. Phys. Lett.* 157 (1989) 200.
- [48] S. Miertus, E. Scrocco, J. Tomasi, *Chem. Phys.* 55 (1981) 117.
- [49] S. Miertus, J. Tomasi, *Chem. Phys.* 65 (1982) 239.
- [50] J. Tomasi, M. Persico, *Chem. Rev.* 94 (1994) 2027.
- [51] V. Barone, M. Cossi, J. Tomasi, *J. Chem. Phys.* 107 (1997) 3210.
- [52] V. Barone, M. Cossi, *J. Phys. Chem. A* 102 (1998) 1995.
- [53] J. L. Pascual-Ahuir, E. Silla, I. Tunon, *J. Comput. Chem.* 15 (1994) 1127.

- [54] M. Cossi, B. Mennucci, R. Cammi, *J. Comput. Chem.* 17 (1996) 57.
- [55] R. Schreck, *J. Am. Chem. Soc.* 71 (1949) 1881.
- [56] P. Vogel, "Chimie Organique, Méthodes et Modèles", université DeBoeck, 1998.
- [57] G. Cook and M.P. Feltman, *J. Chem. Ed.* 84 (2007) 1827.
- [58] P. Umnahanant, J.S. Chickos, *J. Chem. Eng. Data.* 50 (2005) 1720.
- [59] S. Gelin, P. Pollet, *Tetrahedron Lett.* 21 (1980) 4491.
- [60] J. P. Hofmann, F. Duus, A.D. Bond, P.E. Hansen, *J. Mol. Struct.* 790 (2006) 80.
- [61] S. Gelin, D. Hartmann., *J. Heterocyclic Chem.* 13 (1976) 521.
- [62] K. Saito, T. Yamaguchi., *Bull. Chem. Soc. Japan* 51 (1978) 651.
- [63] M. Tanaka, T. Shono, K. Shinra, *Bull. Chem. Soc. Jpn.* 42 (1969) 3190.
- [64] B. L. Zhang, F. Mabon, M. L. Martin, *J. Phys. Org. Chem.* 6 (1993) 367.
- [65] H. J. Harries, G. Parry, J. Burgess, *Inorg. Chim. Acta.* 31 (1978) 233.
- [66] M. Temprado, M. V. Roux, P. Umnahanant, H. Zhao, J. S. Chickos, *J. Phys. Chem. B.* 109 (2005) 12590.
- [67] R. J. Irving, I. Wadso, *Acta Chem. Scand.* 24 (1970) 589.
- [68] M. M. Folkendt, B. E. Weiss-Lopez, J. P. Chauvel, N. S. True, *J. Phys. Chem.* 89 (1985) 3347.
- [69] A. Schweig, H. Vermeer, U. Weidner, *Chem. Phys. Lett.* 26 (1974) 229.
- [70] R. J. Cremlyn A. G. Osborne J.F. Warmsley, *Spectrochimica Acta. A,* 52 (1996) 1423
- [71] M.P. Sammes, P. N. Maini, *Magn. Reson. Chem.* 25 (1987) 372.
- [72] G. A. Olah, J. L. Grant, P. W. Westerman, *J. Org. Chem.* 40 (1975) 2102.
- [73] F. Imashiro, S. Maeda, K. Takegoshi, T. Tereo, A. Saika, *Chem. Phys. Lett.,* 92 (1982) 642.

**SOLVENT EFFECTS ON THE KETO-ENOL  
TAUTOMERIC EQUILIBRIUM OF TETRONIC  
AND ETHYL ACETOACETATE CARBON  
ACIDS. A THEORETICAL STUDY**

I. E. CHARIF and S. M. MEKELLECHE\*

*Département de Chimie, Faculté des Sciences  
Université A. Belkaid, BP 119, 13000, Tlemcen, Algeria  
\*sm\_mekelleche@mail.univ-tlemcen.dz*

D. VILLEMIN

*ENSICAEN, Université de Caen, LCMT  
UMR 6507 CNRS, 6 Boulevard Maréchal Juin  
14050, Caen, France*

Received 2 March 2010

Accepted 31 May 2010

The solvent effects on the keto-enol tautomeric equilibria of ethyl acetoacetate (EAA) and tetrionic acid (TA) are theoretically investigated. The present study shows that the most stable keto tautomer of EAA corresponds to the trans diketo, E, Z form; while the most stable enol tautomer corresponds to the structure in which the enolization takes place at the carbonyl group. Our calculations also put in evidence that the keto tautomer of TA prefers the trans diketo, E, E form, while the most stable enol tautomer corresponds to the structure in which the enolization takes place at the carbonyl group. The calculated free energies indicate that, in polar solvents, the keto-enol equilibrium of EAA is shifted towards the keto tautomer, whereas the keto-enol equilibrium of TA is shifted toward the enol tautomer. The trends of the change of equilibrium constants with respect to the change of solvent polarity are well reproduced by both B3LYP and MP2 calculations. The present study shows that the enthalpic term is predominant in the determination of the calculated equilibrium constants and the entropic effect on the calculated Gibbs free energies is found to be very small and has little influence on the studied keto-enol tautomeric equilibria.

*Keywords:* Ethyl acetoacetate; tetrionic acid; solvent effects; keto-enol equilibrium; theoretical calculations.

## 1. Introduction

Tautomerism is a particular case of isomerism which plays an important role in modern organic chemistry, biochemistry, medicinal chemistry, and pharmacology.<sup>1,2</sup>

\*Corresponding author.

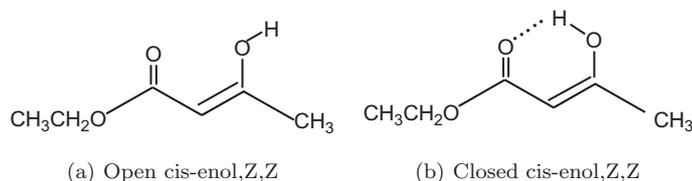
Understanding the mechanism of organic reactions and biochemical processes, in which a substrate or an active intermediate requires an advanced understanding of tautomerization phenomena. Tautomerism occurs also in natural products as in the case of amino acids and pyrimidine bases.<sup>3,4</sup> For this reason, an explanation of the chemical reactivity and the biological activity remains rather difficult when tautomerism occurs.<sup>2-5</sup> Indeed, it is not always clear which tautomeric form is responsible for chemical reactivity/biological activity when the choice is between the thermodynamically most stable tautomer and a less stable tautomer with particular acid-base, electrophilic-nucleophilic, redox, or even just geometric properties.

Keto-enol tautomerism is one of the most commonly studied forms of prototropy phenomena.<sup>1,6,7</sup> This interconversion occurs in different tautomeric systems containing one or more carbonyl groups linked to  $sp^3$  carbons bearing one or more hydrogen atoms. There are also polyfunctional tautomeric derivatives in which the carbonyl groups are separated by a conjugated system. The ratio of the keto/enol tautomers depends on various factors such as temperature, nature of substituents, and solvents.<sup>8,9</sup>

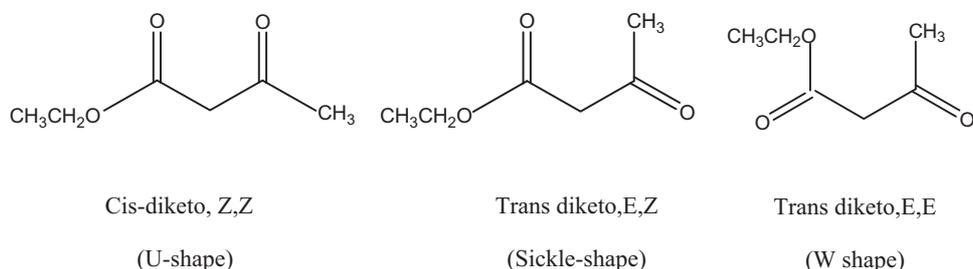
The keto-enol tautomerism of open chain 1,3-dicarbonyl compounds of the type  $R_1C(O)-CH_2-C(O)R_2$  has attracted great interest in the last decades.<sup>9-27</sup> The preference of the enol or the keto tautomer depends strongly on the nature of substituents  $R_1$  and  $R_2$ .<sup>11-13</sup> We note that two enol forms can occur, with the OH group close to  $R_1$  or close to  $R_2$  as in the case of ethyl acetoacetate (see Scheme 1).

According to gas phase spectroscopic and theoretical studies, the enol tautomer is predominant when both  $R_1$  and  $R_2$  are H,<sup>14-16</sup>  $CH_3$ ,<sup>17,18</sup>  $C(CH_3)_3$ <sup>19</sup> or  $CF_3$ .<sup>20,21</sup> It is of great importance to note that this enol tautomer exists exclusively in the cis-enolic form which corresponds to a cyclic form stabilized by intramolecular hydrogen bonding (see Scheme 1(b)).<sup>22,23</sup> On the other hand, the compounds with  $R_1 = R_2 = F$ ,<sup>24</sup>  $Cl$ ,<sup>25</sup>  $OCH_3$ <sup>26</sup> or  $NH_2$ <sup>27</sup> exist preferably in the diketo form. In this form, different conformations are feasible, depending on the relative orientations of the carbonyl groups (Scheme 2). In the absence of steric interactions, the cis-diketo,Z,Z form (U-shape) has the greatest coulombic repulsions, followed by the trans-diketo,E,Z form (sickle-shape). The trans-diketo,E,E form (W shape) should be the most stabilized structure.<sup>28,29</sup>

The solvent effects on the chemical equilibrium was discovered in 1896, simultaneously with the discovery of keto-enol tautomerism in 1,3-dicarbonyl



Scheme 1. Enolic forms of ethyl acetoacetate.



Scheme 2. Ketonic forms of ethyl acetoacetate.

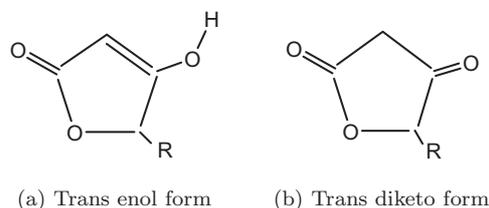
compounds.<sup>8,9</sup> The study of this equilibrium in a variety of solvents using proton nuclear magnetic resonance (NMR) has been widely used as a physical chemistry technique for a long time.<sup>30-35</sup> For open chain  $\beta$ -dicarbonyl compounds, both keto and enol tautomers are frequently present in solution in measurable proportions.<sup>36,37</sup> The percentage of enolization of  $\beta$ -dicarbonyls is very sensitive to the nature of the solvent. For instance, Allen *et al.*<sup>38</sup> have shown that the enol percentage of 2,4-pentanedione (acetylacetone) is only 60% in dimethyl sulfoxide (DMSO) in comparison to 75% in acetone, 94% in  $\text{CCl}_4$  and 95% in cyclohexane. Moreover, it is well established that the enol proportions increase in aprotic non polar solvents, i.e. when intermolecular hydrogen bonding with the solvent does not compete. By contrast, in polar protic solvents, the intermolecular hydrogen bond with the solvent stabilizes the keto form. This means that the keto-enol equilibrium is sensible to the nature of the solvent.

Numerous investigations on experimental and theoretical studies of the keto-enol tautomerism of cyclic chain 1,3-dicarbonyl compounds can be found in the literature.<sup>39</sup> The percentage of the enol forms of this cyclic  $\beta$ -diketone is strongly influenced by the solvent, both in terms of polarity and hydrogen bond acceptor/donor properties.<sup>9</sup> We note that, in water, the open chain  $\beta$ -diketones exist predominantly as diketo tautomers; while the closed  $\beta$ -diketones show an opposite behavior and exist predominantly as enol tautomers.<sup>39,40</sup>

Tetronic acids which are a particular case of cyclic  $\beta$ -ketoesters compounds, have traditionally been represented as derivatives of the enol form as opposed to the corresponding keto form.<sup>41-43</sup> Indeed, the ultraviolet spectrum of tetronic acid, in aqueous ethanol, shows two important absorptions at  $\lambda_{\text{max}} = 223$  (C=O) and  $\lambda_{\text{max}} = 248$  nm (C=C) due to the presence of the enol tautomer.<sup>44</sup>

According to experimental observations,<sup>41-44</sup> tetronic acid exists only in the trans enolic form in which the enolization takes place at the carbonyl group (Scheme 3). Contrary to open chain 1,3-dicarbonyl compounds, the diketo form of tetronic acid and other cyclic  $\beta$ -dicarbonyl compounds exist only in a trans diketo, E, E conformation.

Our aim in this work is to present a theoretical study of the molecular structure of tetronic cyclic acid (TA) and its analog acyclic acid, namely, ethyl acetoacetate



Scheme 3. Tautomeric forms of the tetric acid.

(EAA). The solvent effects on the keto-enol equilibrium of these  $\beta$ -ketoesters compounds are elucidated by free energy calculations performed at two correlated levels (B3LYP and MP2) of theory using the polarizable continuum solvation model (PCM).

## 2. Computational Details

All calculations were carried out by means of the G03W computational package.<sup>45</sup> Second-order Møller-Plesset (MP2)<sup>46</sup> and the Becke-3-parameter-Lee-Yang-Parr (B3LYP)<sup>47</sup> methods were employed to perform the quantum chemistry calculations. Solvent effects were studied by performing self-consistent reaction field (SCRF) calculations using the well-known polarizable continuum model (PCM) on the gas phase optimized geometry of each species, initially devised by Tomasi and co-workers<sup>48–50</sup> and progressively implemented by Barone *et al.*<sup>51,52</sup> In this model, the solute molecule is placed into a cavity formed by the envelope of spheres centered on the atomic groups,<sup>53,54</sup> in which the dielectric constant is the same as in the vacuum whereas outside it is that of the desired solvent. All the calculations using the PCM solvent model employ the UAHF (United Atom for Hartree-Fock) atomic radii when constructing the solvent cavity, as recommended in the Gaussian 03 user's reference when the "scfvac" keyword is used to obtain the free-energy of solvation  $\Delta G_{\text{solv}}^{\circ}$ , as is the case in this study.<sup>45b,51,52</sup>

The thermodynamic cycle used to calculate the tautomeric constant  $pK_{\text{T}}$  is described in Scheme 4.

$$\Delta G_{\text{T}}^{\circ}(\text{sol}) = G_{\text{sol}}^{\circ}(\text{enol}) - G_{\text{sol}}^{\circ}(\text{keto}) \quad (1)$$

where

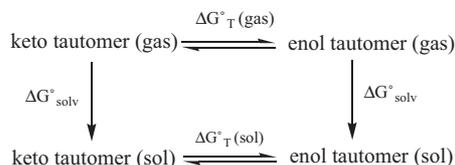
$$G_{\text{sol}}^{\circ}(X) = G_{\text{gas}}^{\circ}(X) + \Delta G_{\text{solv}}^{\circ}(X) \quad (2)$$

$$K_{\text{T}} = \exp(-\Delta G_{\text{T}}^{\circ}/RT) \quad (3)$$

with

$$R = 1.98709369 \text{ cal}\cdot\text{mol}^{-1} \cdot \text{K}^{-1} \text{ and } T = 298.15 \text{ K}$$

In Eq. (1),  $\Delta G_{\text{T}}^{\circ}(\text{sol})$ , which represents the Gibbs free energy change in solution of the considered tautomeric equilibrium, is calculated from the solution Gibbs free

Scheme 4. Thermodynamic cycle used in the calculation of  $\Delta G_{\text{T}}^{\circ}$ .

energy of the involved species. In Eq. (2),  $G_{\text{sol}}^{\circ}(X)$  is directly calculated from the gas-phase Gibbs free energies ( $G_{\text{gas}}^{\circ}$ ) and Gibbs free energies of solvation ( $\Delta G_{\text{solv}}^{\circ}$ ). The tautomeric constants  $K_{\text{T}}$  and  $pK_{\text{T}}$  are defined in Eq. (3).

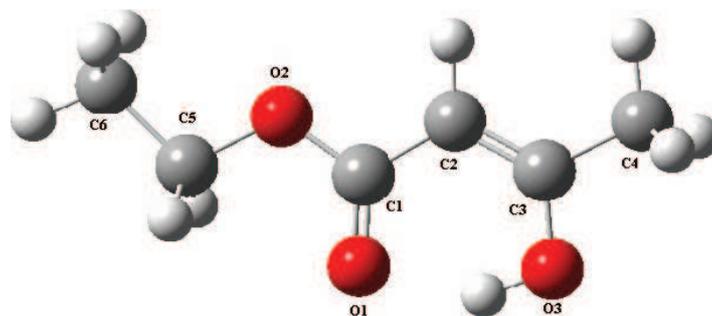
### 3. Results and Discussion

#### 3.1. Solvent effects on the tautomeric equilibrium of ethyl acetoacetate (EAA) compound

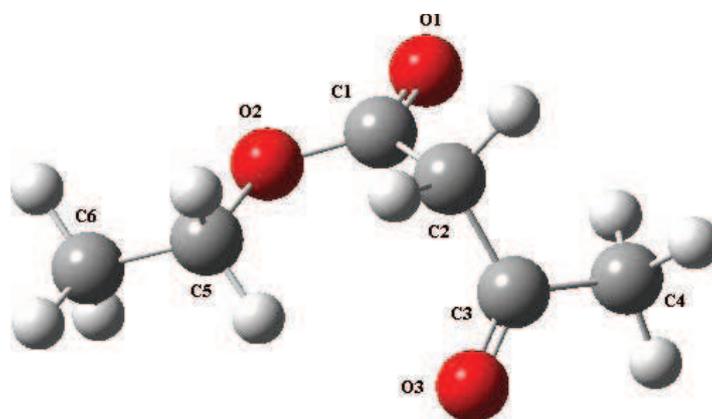
EAA compound contains two different groups  $R_1=\text{CH}_3$  and  $R_2=\text{OCH}_2\text{CH}_3$ . This makes the tautomeric properties of this compound highly interesting. According to B3LYP/6-311++G(d,p) and MP2/6-31+G(d,p) calculations, the cis-enol form in which the enolization takes place at the carbonyl group  $\text{CH}_3\text{C}(\text{OH})=\text{CHC}(\text{O})\text{OCH}_2\text{CH}_3$  is more stable by about 10 kcal/mol than the cis-enol form in which the enolization takes place at the ester group. For this reason, only the most stable enolic tautomer form, represented in Fig. 1(a), will be the subject of the present theoretical study.

To find all possible diketo conformers, the potential energy surface was scanned at the B3LYP/6-311++G(d,p) level of theory. The torsional angles  $\theta(\text{O}_1\text{C}_1\text{C}_2\text{C}_3)$  and  $\theta(\text{O}_3\text{C}_3\text{C}_2\text{C}_1)$  were changed in steps of  $20^\circ$  with full optimization of all other parameters (see Fig. 1(b) for atom numbering). Only one deep minimum on the potential energy surface was obtained with  $\theta(\text{O}_1\text{C}_1\text{C}_2\text{C}_3) = 86^\circ$  and  $\theta(\text{O}_3\text{C}_3\text{C}_2\text{C}_1) = 99.5^\circ$ . The optimized geometry of EAA compound, given in Fig. 1(b), show that this  $\beta$ -ketoesters is more stabilized in a sickle-shape. The same behavior of the potential energy surface was found at the MP2/6-31+G(d,p) computational level.

Experimental results show that, in the presence of polar solvents, EAA exists as a mixture of diketo/enol tautomers.<sup>55</sup> The measured equilibrium constant  $K_{\text{T}} = [\text{enol}]/[\text{keto}] = 0.11$  (in ethanol) implies a strong preference for the diketo form. However, in non polar solvents and in the gas phase, the enol tautomeric form is predominant (64% in n-hexane).<sup>56-58</sup> It turns out from experimental results given in Table 1, that the keto/enol forms of EAA are both present in solution in measurable proportions. We note that the percentage of the enol form is increased in gas phase; while the keto form is predominant in ethanol and acetonitrile polar solvents.



(a) Cis-enol tautomer



(b) Diketo tautomer (sickle-shape)

Fig. 1. B3LYP/6-311++G(d,p) optimized geometries of the two most stable tautomers of ethyl acetoacetate. Color online.

Table 1. Experimental keto-enol equilibrium constants  $K_T$  of ethyl acetoacetate in gas phase and in various solvents.<sup>56</sup>

	Gas Phase ( $\epsilon = 1$ )	Benzene ( $\epsilon = 2.25$ )	Ethanol ( $\epsilon = 24.55$ )	Acetonitrile ( $\epsilon = 36.64$ )
Enol(%) / diketo(%)	42.5 / 57.5	16.0 / 84.0	9.9 / 90.1	4.8 / 95.2
$K_T = [\text{enol}] / [\text{keto}]$	0.74	0.19	0.11	0.05

In order to put in evidence the solvent effects on the keto-enol tautomeric equilibrium, we have considered various mediums with dielectric constants varying from 1 (gas phase) to 78.39 (water). The polarized continuum model (PCM) was applied at the B3LYP/6-311++G(d,p) and MP2/6-31+G(d,p) levels of theory. B3LYP Gibbs free energies increase from  $-9.49$  in the gas phase ( $\epsilon = 1$ ) to  $-5.08$  kcal/mol in water ( $\epsilon = 78.39$ ) (see Table 2). The same trends are obtained for MP2 free energy calculations. Indeed, MP2 free energies increase from  $-6.66$  in the gas phase

Table 2. B3LYP/6-311+G(d,p) and MP2/6-31+G(d,p) (in parentheses) calculated values of  $\Delta S_T$  ( $\text{cal.mol}^{-1}.\text{K}^{-1}$ ),  $\Delta H_T^\circ$  ( $\text{kcal.mol}^{-1}$ ),  $\Delta G_T^\circ$  ( $\text{kcal.mol}^{-1}$ ), and  $K_T$  for keto-enol equilibriums of EAA and TA compounds. The data are given for gas phase and for various solvents.

	Gas Phase ( $\epsilon = 1$ )	Chloroform ( $\epsilon = 4.90$ )	DMSO ( $\epsilon = 46.70$ )	Water ( $\epsilon = 78.39$ )
Ethyl acetoacetate				
$\Delta S_T^\circ$	-4.16 (-2.11)	-4.37 (-1.73)	-4.46 (-1.58)	-5.42 (-1.56)
$\Delta H_T^\circ$	-10.73 (-7.28)	-8.64 (-5.36)	-7.72 (-4.53)	-5.18 (-2.22)
$\Delta G_{T^a}^\circ$	-9.49 (-6.66)	-7.34 (-4.84)	-6.40 (-4.06)	-3.60 (-1.75)
$\Delta G_{T^b}^\circ$	-9.49 (-6.66)	-8.21 (-4.65)	-7.57 (-3.74)	-5.08 (-0.63)
$K_T$	$9.05 \times 10^6$ ( $7.62 \times 10^4$ )	$1.04 \times 10^6$ ( $0.26 \times 10^4$ )	$0.35 \times 10^6$ ( $5.0 \times 10^2$ )	$0.53 \times 10^4$ (2.90)
Tetronic acid				
$\Delta S_T^\circ$	-2.42 (-2.98)	-2.40 (-2.99)	-2.37 (-3.00)	1.94 (2.49)
$\Delta H_T^\circ$	4.30 (5.90)	3.06 (4.54)	2.45 (3.88)	2.80 (3.86)
$\Delta G_{T^a}^\circ$	5.03 (6.79)	3.77 (5.43)	3.15 (4.77)	3.39 (4.60)
$\Delta G_{T^b}^\circ$	5.03 (6.79)	3.68 (5.71)	2.92 (5.06)	2.34 (4.91)
$K_T$	$2.05 \times 10^{-4}$ ( $1.07 \times 10^{-5}$ )	$2.01 \times 10^{-3}$ ( $6.52 \times 10^{-5}$ )	$7.24 \times 10^{-3}$ ( $1.95 \times 10^{-4}$ )	$1.92 \times 10^{-2}$ ( $2.51 \times 10^{-4}$ )

<sup>a</sup>Calculated using the formula  $\Delta G_T^\circ = \Delta H_T^\circ - T\Delta S_T^\circ$ .

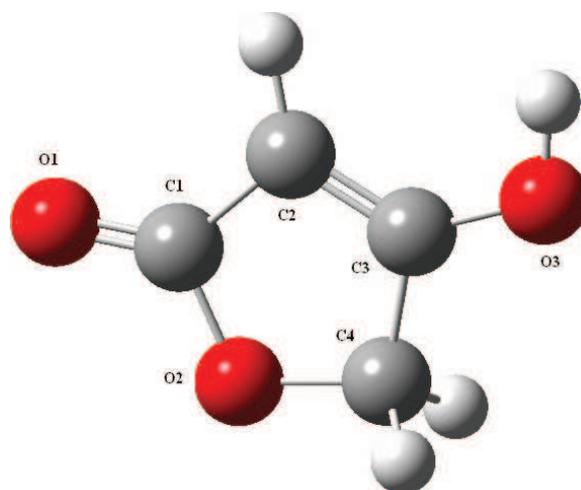
<sup>b</sup>Calculated from thermodynamic cycle (Eq. (2)).

to  $-0.63 \text{ kcal/mol}$  in water. The calculated tautomeric constants  $K_T$  (Table 2) decrease going from the gas phase to polar aqueous solution, indicating that, the increase of the polarity of the solvent is accompanied by an enhancement of the percentage of the keto tautomeric form. In conclusion, the calculated free energies carried out at the two correlated computational levels (B3LYP and MP2) are in good agreement with experimental findings which stipulate that the keto form is more favored in polar solvents while the enol form is more favored in non polar mediums.

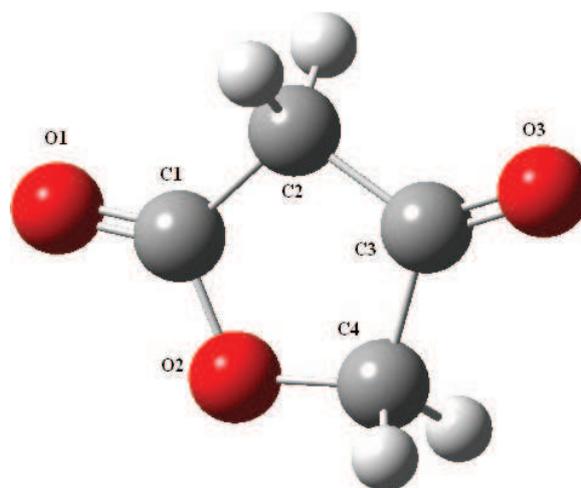
### 3.2. Solvent effects on the tautomeric equilibrium of tetronic acid (TA)

The high degree of enolization of EAA can be attributed to stabilization of the enol form by internal hydrogen bonding. By contrast, such interactions are geometrically not feasible in the case of TA. Consequently, in aqueous solution, the enol form of TA will be better stabilized by intermolecular hydrogen bonding, causing the equilibrium shift to the enol form. Proton NMR results show that TA and its simple derivatives exist only in one trans-enolic form in which the enolization takes place at the carbonyl group with an interest yield superior to 70%.<sup>42,43</sup> B3LYP and MP2 theoretical calculations reveal that the trans-enolic form is favored by about 6 kcal/mol in comparison with the other enolic form in which the enolization takes place at the ester group (see Fig. 2(a)). For this reason, only the most stable enolic tautomer of TA (with R=H) will be considered in the present theoretical study.

In opposite to open chain  $\beta$ -ketoesters compounds, TA exists only in a trans diketo, E, E conformation. Effectively, B3LYP optimized geometry of TA (with



(a) Trans-enol form



(b) Trans-diketo form

Fig. 2. B3LYP/6-311++G(d,p) optimized geometries of the two most stable tautomeric forms of the tetrionic acid. Color online.

R=H), given in Fig. 2(b), shows that this cyclic  $\beta$ -ketoesters compound is more stabilized in the trans-diketo form with the torsional angles  $\theta(O_1C_1C_2C_3) = 179.94^\circ$  and  $\theta(O_3C_3C_2C_1) = 179.98^\circ$  (see Fig. 2(b) for atom numbering).

Experiment<sup>41–43,59–62</sup> also indicates that the tautomeric interconversion of TA is accelerated by polar solvents. This process might become more favorable in condensed media, particularly in strong polar solvents. B3LYP calculations show that the free energy change,  $\Delta G_T^\circ$ , of keto-enol equilibrium is 5.03 kcal/mol in the gas

phase and 2.34 kcal/mol in water. Consequently, polar solvents reduce  $\Delta G_T^\circ$ , indicating that the polarity of solvent favors the enolization of TA. The calculated tautomeric constant  $K_T$  also indicate that the enol form is more favored in polar solvents as expected experimentally. We note that the trends of the change of equilibrium constants with respect to the change of the solvent polarity are well reproduced by both B3LYP and MP2 methods, although the free energies changes are overestimated. Exact estimation of the keto-enol equilibrium composition apparently requires methods more sophisticated than B3LYP or MP2.<sup>13</sup>

Entropy effects are sometimes neglected in quantum mechanical calculations. However, as we can see from Table 2, the B3LYP calculated entropy values are generally small (*ca.* 5 cal. mol<sup>-1</sup>.K<sup>-1</sup> and 2 cal. mol<sup>-1</sup>.K<sup>-1</sup> for EAA and TA respectively). These results indicate that the entropy effect on the Gibbs free energy is very small and the entropy term, that is, the  $T\Delta S^\circ$  value can be neglected for the tautomeric equilibrium of EAA and TA. Thus, the enthalpic term is predominant in the determination of the equilibrium constants.

#### 4. Concluding Remarks

The most important findings of our computational study are as follows:

- (1) Theoretical calculations show that the keto tautomer of EAA is stabilized in a trans diketo, E, Z form (sickle-shape); while the most stable enol tautomer, for these beta-ketoesters, corresponds to the structure in which the enolization takes place at the carbonyl group.
- (2) Theoretical calculations show that the keto tautomer of TA is stabilized in the trans diketo, E, E form; while the most stable enol tautomer corresponds to the structure in which the enolization takes place at the carbonyl group.
- (3) Free energy calculations show that EAA keto-enol equilibrium is shifted toward the keto tautomer in polar solvents; while the keto-enol equilibrium of TA is shifted toward the enol tautomer in polar solvents. These trends are in good agreement with experimental findings.
- (4) Our calculations show that the entropic term on the Gibbs free energy of EAA and TA is small and the enthalpic term is predominant in the determination of the equilibrium constant.

#### References

1. (a) March J, *Advanced Organic Chemistry, Reaction, Mechanism and Structure*, Wiley, New York, 1993. (b) Carey FA, Sundberg RJ, *Advanced Organic Chemistry*, Plenum Press, New York, 1993.
2. (a) Saenger W, *Principles of Nucleic Acid Structure*, Springer, New York, 1994. (b) Foye WO, Lemke TL, Williams DA, *Principles of Medicinal Chemistry*, Williams and Wilkins, Baltimore, MD, 1995. (c) Weinstein H, Chou D, Johnson CL, Kang S, Green JP, *Mol Pharmacol* **12**:738, 1976. (d) Watling KJ, *The RBI Handbook of Receptor Classification and Signal Transduction*, RBI, Natick, MA, 1998. (e) Pozharskii AF,

- Soldatenkov AT, Katritzky AR, *Heterocycles in Life and Society*, Wiley, New York, 1997.
- (a) Raczyn'ska ED, Cyran'ski MK, Darowska M, Rudka T, *Targets Heterocycl Chem* **4**:327, 2000. (b) Raczyn'ska ED, Darowska M, Makowski M, *Trends Heterocycl Chem* **9**:167, 2003. (c) Bliznyuk AA, Schaefer HF, Amster IJ, *J Am Chem Soc* **115**:5149, 1993. (d) Day RM, Thalhauser CJ, Sudmeier JL, Vincent MP, Torchilin EV, Sanford DG, Bachovchin CW, Bachovchin WW, *Protein Sci* **12**:794, 2003. (e) Huda'ky P, Perczel A, *J Phys Chem A* **108**:6195, 2004. (f) Rak J, Skurski P, Simons J, Gutowski M, *J Am Chem Soc* **123**:11695, 2001.
  - (a) Kwiatkowski JS, Pullman B, *Adv Heterocycl Chem* **18**:200, 1975. (b) Kwiatkowski JS, Person WB, in *Theoretical Biochemistry and Molecular Biology*, Beveridge DL, Lavery R (eds.), Academic Press, New York, p. 153, 1990. (c) Katritzky AR, Karelson M, *J Am Chem Soc* **113**:1561, 1991. (d) Roehring GH, Oyler NA, Adamowicz L, *Chem Phys Lett* **225**:265, 1994. (e) Sýponer J, Hobza P, *J Phys Chem* **98**:3161, 1994. (f) Kabelac M, Kratochvil M, Sýponer J, Hobza P, *J Biomol Struct Dyn* **117**:1077, 2000.
  - (a) Smith K, *Porphyrins and Metalloporphyrins*, Elsevier, Amsterdam, 1975. (b) The Porphyrins, Dolphin, D (ed.), Vols. 1–7, Academic Press: New York, 1978–1979. (c) Sessler JL, Weghorn SJ, *Expanded, Contracted, and Isomeric Porphyrins*, Elsevier, Oxford, 1997. (d) Kadish KM, Smith KM, Guillard R, *The Porphyrin Handbook*, Academic Press, Vol. 2, San Diego, CA, 1999.
  - (a) Patai S, *The Chemistry of the C=O Group*, Wiley, London, 1966. (b) Rappoport Z, *The Chemistry of Enols*, Wiley, New York, 1990.
  - (a) Kresge A, *J Chem Soc Rev* **25**:275, 1996. (b) Karelson M, Maran U, Katritzky AR, *Tetrahedron* **52**:11325, 1996. (c) Rappoport Z, Frey J, Sigalov M, Rochlin E, *Pure Appl Chem* **69**:1933, 1997. (d) Fontana A, De Maria P, Siani G, Pierini M, Cerritelli S, Ballini R, *Eur J Org Chem* **65**:1637, 2000. (e) Iglesias E, *Curr Org Chem* **8**:1, 2004.
  - Claisen L, *Ann Chem* **291**:25, 1896.
  - Reichard C, *Solvents and Solvents Effects in Organic Chemistry*, VCR, New York, p. 2, 1988.
  - Emsley J, *Structure and Bonding*, Springer, Berlin, Vol. 57, p. 147, 1984.
  - Huang M, Graham Richards W, Guy H Grant. *J Phys Chem A* **109**:5198, 2005.
  - Belova NV, Girichev GV, Shlykov SA, Oberhammer H, *J Org Chem* **71**:5298, 2006.
  - (a) Belova NV, Oberhammer H, Girichev GV, Shlykov SA, *J Phys Chem A* **111**:2248, 2007. (b) Belova NV, Sliznev VV, Oberhammer H, Girichev GV, *J Mol Struct* (2010), accepted manuscript. doi: 10.1016/j.molstruc.2010.02.070.
  - Baughcum SL, Duerst RW, Rowe WF, *J Am Chem Soc* **103**:6296, 1981.
  - Baughcum SL, Smith Z, Wilson EB, Duerst, *J Am Chem Soc* **106**:2260, 1984.
  - Turner P, Baughcum SL, Coy SL, Smoth Z, *J Am Chem Soc* **106**:2265, 1984.
  - Lowrey AH, George C, D'Antonio P, Karle J, *J Am Chem Soc* **93**:6399, 1971.
  - Andreassen AL, Bauer SH, *J Mol Struct* **12**:381, 1972.
  - Iijima K, Ohnogi A, Shibata S, *J Mol Struct* **156**:111, 1987.
  - Giricheva NI, Girichev GV, Lapshina SB, Kuzmina NP, *Zh Strukt Khim* **41**:58, 2000 (in Russian).
  - Moriyasu M, Kato A, Hashimoto Y, *J Chem Soc Perkin Trans* **2**:515, 1986.
  - Burdett J, Rogers M, *J Am Chem Soc* **86**:2105, 1971.
  - Chong DP, C-Han Hu, *J Electron Spectrosc Relat Phenom* **94**:181, 1998.
  - Jin AD, Mack HG, Waterfeld A, Oberhammer H, *J Am Chem Soc* **113**:7874, 1991.
  - Mack HG, Oberhammer H, Della Vedova CD, *J Mol Struct* **346**:51, 1995.
  - Belova NV, Oberhammer H, Girichev GV, *J Mol Struct* **689**:255, 2004.

27. Kawashima Y, 10th *European Symposium on Gas Electron Diffraction*, St. Petersburg, Russia, 2003.
28. Lee I, Han IS, Kim CK, Lee HW, *Bull Korean Chem Soc* **24**:1141, 2003.
29. Perez GV, Perez AL, *J Chem Educ* **77**:910, 2000.
30. (a) Toullec J, in *The Chemistry of Enols*, Rappoport Z, Wiley, Chichester, UK, p. 323, 1990. (b) Forse'n S, Nilson M, in *The Chemistry of the Carbonyl Group*, Zabicky J, Interscience Publishers, New York, Vol. 2, p. 198, 1970. (c) Neylands O, in *Structure and Tautomerism of  $\beta$ -Dicarbonyl Compounds*, Gudrience E, Zinatne, Riga, Latvia, p. 142, 1976.
31. (a) Iglesias E, *J Phys Chem* **100**:12592, 1996. (b) Iglesias E, *J Chem Soc, Perkin Trans* **2**:431, 1997. (c) Iglesias E, *New J Chem* **29**:457, 2005. (d) Iglesias E, *J Phys Chem B* **105**:10287, 2001.
32. (a) Iglesias E, Ojea-Cao V, Garcia-Rio L, Leis JR, *J Org Chem* **64**:3954, 1999. (b) Iglesias E, *J Org Chem* **65**:6583, 2000. (c) Iglesias E, *J Inclusion Phenom Macrocyclic Chem* **52**:55, 2005.
33. Reeves LW, *Can J Chem* **35**:1351, 1957.
34. Drexler EJ, Field KW, *J Chem Educ* **53**:392, 1976.
35. Garland EW, Nibler JW, Shoemaker DP, *Experiments in Physical Chemistry*, McGraw-Hill, New York, p. 453, 2003.
36. Williams AG, Butler GB, *J Org Chem* **45**:1232, 1980.
37. Veierov D, Bercovici T, Fischer E, Mazur Y, Yogev A, *J Am Chem Soc* **99**:2723, 1977.
38. Allen G, Dwek RA, *J Chem Soc B* **161**:1966.
39. (a) Luchmuller CH, Maldacker T, Cefola M, *Anal Chim Acta* **48**:139, 1969. (b) Mendez E *et al.*, *J Phys Chem C* **111**:3369, 2007. (c) Zuccarello F, Buemi G, Gandolfo C, Contino A, *Spectro Acta Prat A* **39**:139, 2003. (d) Bojarski J *et al.*, *Advances in Heterocyclic Chemistry*, Academic Press, New York, Vol. 38, p. 231, 1985. (e) Eckert-Maksic M, Maksimovic L, *J Mol Struct (THEOCHEM)* **153**:121, 1987. (f) Skylaris CK, Igglessi-Markopoulou O, Detsi A, Markopoulos J, *Chem Phys* **293**:355, 2003.
40. Schwarzenbach G, Felder E, *Helv Chim Acta* **27**:1044, 1944.
41. Pévet I, Meyer C, Cossy J, *Tetrahedron Lett* **42**:5215, 2001.
42. Rajaram AR, Pu L, *Org Lett* **8**:2019, 2006.
43. Willis C, Bodio E, Bourdreux Y, Billaud C, Le Gall T, Mioskowski C, *Tetrahedron Lett* **48**:6421, 2007.
44. Brodyck, Royles JL, *Chem Rev* **95**:1981, 1995.
45. (a) Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery Jr JA, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, and Pople JA, *Gaussian 03, Revision E.01*, Gaussian, Inc., Wallingford CT, 2004. (b) Frisch  $\text{\AA}$ , Frisch M, Trucks GW, *Gaussian 03 User's Reference*, Gaussian Inc., Wallingford, CT, 2003.
46. Møller C, Plesset MS, *Phys Rev* **46**:618, 1934.

47. (a) Koch W, Max Holthausen CA, *Chemist's Guide to Density Functional Theory*, Wiley, VCH, Weinheim, Germany, 2001. (b) Becke AD, *J Chem Phys* **98**:5648, 1993. (c) Becke AD, *Phys Rev A* **38**:3098, 1988. (d) Lee C, Yang W, Parr RG, *Phys Rev B* **37**:785, 1988. (e) Miehlich B, Savin A, Stoll H, Preuss H, *Chem Phys Lett* **157**:200, 1989.
48. Miertus S, Scrocco E, Tomasi J, *Chem Phys* **55**:117, 1981.
49. Miertus S, Tomasi J, *Chem Phys* **65**:239, 1982.
50. Tomasi J, Persico M, *Chem Rev* **94**:2027, 1994.
51. Barone V, Cossi M, Tomasi J, *J Chem Phys* **107**:3210, 1997.
52. Barone V, Cossi M, *J Phys Chem A* **102**:1995, 1998.
53. Pascual-Ahuir JL, Silla E, Tunon I, *J Comput Chem* **15**:1127, 1994.
54. Cossi M, Mennucci B, Cammi R, *J Comput Chem* **17**:57, 1996.
55. Schreck R, *J Am Chem Soc* **71**:1881, 1949.
56. Vogel P, *Chimie Organique, Méthodes et Modèles*, DeBoeck Université, p. 639, 1998.
57. Cook G, Feltman MP, *J Chem Ed* **84**:1827, 2007.
58. Umnahanant P, Chickos JS, *J Chem Eng Data* **50**:1720, 2005.
59. Gelin S, Pollet P, *Tetrahedron Lett* **21**:4491, 1980.
60. Hofmann JP, Duus F, Bond AD, Hansen PE, *J Mol Struct* **790**:80, 2006.
61. Gelin S, Hartmann D, *J Heterocyclic Chem* **13**:521, 1976.
62. Saito K, Yamaguchi T, *Bull Chem Soc Japan* **51**:651, 1978.

**PARTIE 4 :**

*Etude théorique des effets stéréoelectroniques sur l'acidité carbonée des composés  $\beta$ -dicarboxylés cycliques et de leurs analogues à chaînes ouvertes.*

**1. Introduction**

L'acidité carbonée aqueuse des composés cycliques tel que l'acide de Meldrum ( $pK_a = 4.83$ ), la dimédone ( $pK_a = 5.25$ ), l'acide barbiturique ( $pK_a = 4.01$ ) et l'acide tétronique ( $pK_a = 3.76$ ) [1] est comparable à celle de l'acide acétique ( $pK_a = 4.75$ ). L'acidité relativement élevée de ces composés a été donc attribuée aux hydrogènes acides attachés sur un carbone situé entre deux groupements carbonyliques. L'acidité carbonée des ces  $\beta$ -dicarboxylés est d'autant plus remarquable que les composés voisins à chaînes ouvertes qui ne sont pas aussi acides à comparer à le malonate diéthyle ( $pK_a = 13.3$ ), acétylacétone ( $pK_a = 9.0$ ), le malonamide ( $pK_a = 12.5$ ) et acétylacétate d'éthyle ( $pK_a = 10.7$ ). Par conséquent ces acides ont attiré l'attention due à leur acidité exceptionnellement élevée [1].

Arnett et Harrelson [2] ont suggéré que l'acidité élevée de l'acide de Meldrum par rapport au malonate de méthyle résulte de la rotation restreinte autour des liaisons d'ester dans le cycle à six chaînons de l'acide de Meldrum. Ces auteurs ont conclu que les acidités sont rapidement diminuées en allant d'un cycle à six chaînons vers un cycle à dix chaînons jusqu'à ce que le cycle à treize chaînons ait le même  $pK_a$  que le malonate de méthyle. D'autre part, Wang et Houk [3] ont suggéré que l'acidité élevée de l'acide de Meldrum est originaire des différences de répulsions stériques et électrostatiques (dipôle-dipôle) entre les conformers d'ester E et Z des molécules neutres et anioniques en utilisant l'acétate de

méthyle comme un composé modèle. De même, Wiberg et Laidig [4] ont montré théoriquement que l'acidité surprenante de l'acide de Meldrum ayant une conformation d'ester bis(E) peut être expliquée par la différence dans l'acidité entre les rotamères Z et E de l'acétate de méthyle. Les effets de solvant sur les acidités des esters de Z et E ont été étudiés par Evanseck et col. [5]. Les calculs ont montré une stabilisation de l'ester E de 3.0 kcal/mol dans l'eau et de 2.7 kcal/mol dans l'acétonitrile par rapport de l'ester E. Les énergies libres de solvation de l'anion de la forme E sont également plus stables que celles de la forme Z avec 2.3 kcal/mol dans l'eau et 1.5 cal/mol dans l'acétonitrile.

En comparant les résultats trouvés par Evanseck aux données expérimentales sur le malonate diméthyle et l'acide Meldrum, la conversion de deux esters Z en deux esters E augmentera l'acidité de 10.6 kcal/mol, une valeur qui très proche à celle obtenue expérimentalement (11.7 kcal/mol) [2].

La différence d'acidité en phase aqueuse entre de l'acide de Meldrum et le malonate a été expliquée par la conversion de deux groupements esters Z en deux groupements esters E [5,6]. Cependant, l'acidité élevée de l'acide de Meldrum ne peut pas être simplement expliquée en raison de la conversion de deux esters Z en deux esters. Gao et al.[6] ont montré que les effets de solvant sont plutôt faibles et la grande stabilisation de l'anion énolate est due aux effets stéréoélectroniques appelés également effets anomériques. Ces effets représentent un facteur important pouvant être à l'origine de l'acidité surprenante de l'acide de Meldrum. Le but de cette partie est d'apporter une contribution à la compréhension de l'origine de l'acidité remarquablement élevée des acides de Meldrum, dimédone, barbiturique et tétronique en mettant en évidence les effets stéréoélectroniques en utilisant l'analyse des populations électroniques NBO (Natural Bond Orbital) [7-10].

## 2. Résultats et discussions

### 2.1. Calculs des énergies libres de déprotonation $\Delta G^\circ$

Les énergies de déprotonation  $\Delta G^\circ$  à 298 K en phase gazeuse et en phase aqueuse calculées à l'aide de la méthode B3LYP/6-311++G(d, p) sont récapitulées dans le **Tableau 1**

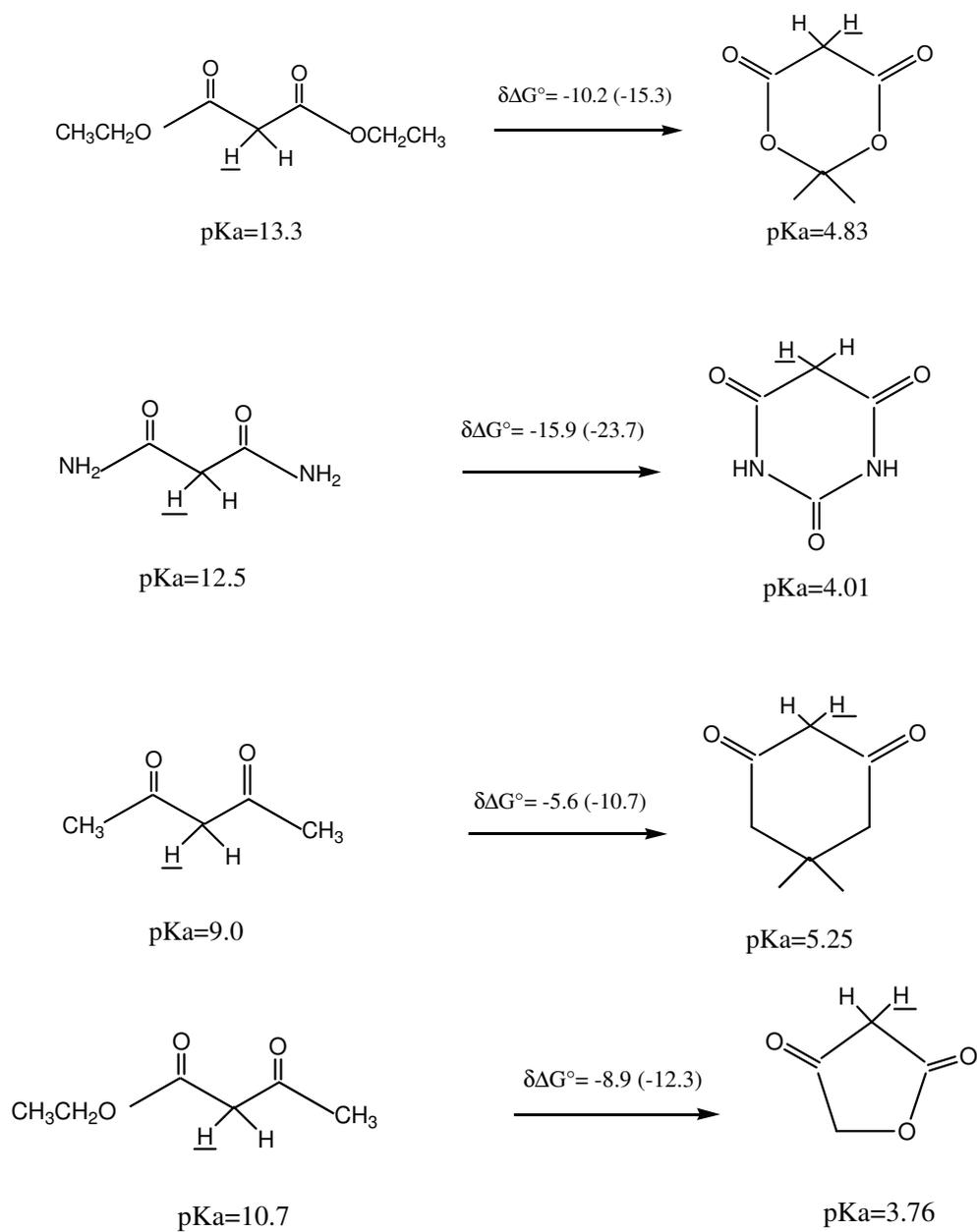
**Tableau 1** : Les énergies de déprotonation  $\Delta G^\circ$  (kcal/mol) (B3LYP/6-311++G(d, p)) en phase gazeuse  $\Delta G^\circ_{\text{gas}}$  et en phase aqueuse  $\Delta G^\circ_{\text{aq}}$  des composés  $\beta$ -dicarboxylés cycliques et acycliques.

Acide	$\Delta G^\circ_{\text{gas}}$	$\Delta G^\circ_{\text{aq}}$	pK <sub>a</sub> (exp)
Malonate diéthyle	336.62	22.31	13.30
Acide de Meldrum	321.34	12.10	4.83
Malonamide	340.31	27.22	12.50
Acide Barbiturique	316.57	11.31	4.01
Acétylacétate d'éthyle	331.50	19.31	10.70
Acide tétronique	319.21	10.44	3.76
Acétylacétone	336.15	20.18	9.00
Dimédone	325.46	14.56	5.25

Les valeurs relatives calculées  $\delta \Delta G^\circ_{\text{aq}}$  au même niveau B3LYP/6-311++G(d, p) sont données dans la **Figure 1**. On remarque que :

- $\Delta G^\circ_{\text{aq}}$  de l'acide de Meldrum (cyclique) est inférieur de 10.2 kcal/mol que celle du malonate diéthyle (acyclique).
- $\Delta G^\circ_{\text{aq}}$  de l'acide tétronique (cyclique) est 8.9 kcal/mol plus faible que celle de son composé analogue acétoacétate d'éthyle (acyclique).

Cependant, les tendances des variations des valeurs de  $\Delta G^\circ$  des différents composés montrent que l'acidité augmente ( $\Delta G^\circ$  diminue) d'une manière remarquable en passant d'un composé  $\beta$ -dicarboxylé cyclique vers son analogue à chaînes ouvertes.



**Figure 1:** Les valeurs relatives des énergies de déprotonation en phase aqueuse et en phase gazeuse des acides carbonés dicarbonylés cycliques et acycliques.

## 2.2. L'analyse NBO

Plusieurs méthodes ont été utilisées pour analyser la contribution des orbitales localisées aux propriétés moléculaires [7-9]. En plus des effets de stabilisation, les interactions stéréoelectroniques fournissent également la manière de transmettre l'information entre les différentes parties de la molécule. Par exemple, la méthode NBO a été employée pour établir les échanges électroniques, les réactions de transfert entre des composés donneur-accepteur et les interactions de hyperconjugaison [10-13].

L'analyse de la manière dont l'information entre les orbitales est transmise dans toute la molécule permet d'interpréter comment une interaction peut affecter une propriété moléculaire localisée dans la même région ou dans une région différente de la molécule.

Dans ce travail, nous présentons une méthode basée sur l'approche NBO pour l'évaluation de la contribution des interactions de délocalisation électronique à l'acidité. N'importe quelle propriété qui dépend d'une fonction d'onde polyatomique décrivant le système moléculaire (par exemple, énergie, densité de charge électronique, charges atomiques...) est susceptible d'être analysée, permettant une évaluation numérique de la contribution des interactions stéréoelectroniques de la propriété.

- **Les interactions donneur-accepteur.**

Dans l'analyse NBO, chaque liaison NBO de valence  $\sigma_{AB}$ , est appariée avec une NBO de valence antiliante correspondante  $\sigma^*_{AB}$ . Les NBOs de Lewis type  $\sigma$  (donneur) sont en interaction avec les NBOs non-Lewis type  $\sigma^*$  (accepteur) qui sont normalement vides. Puisque l'effet non-covalent de délocalisation sont associés aux interactions de  $\sigma \rightarrow \sigma^*$  entre les orbitales occupées (donneur) et les orbitales vacantes (accepteur), il est normal de les décrire en tant qu'un transfert de charge ou un modèle généralisé « base de Lewis – acide de Lewis ».

Dans la méthode NBO, les interactions donneur-accepteur sont prises en compte en examinant des interactions possibles entre les NBOs occupées type-Lewis (donneur) et les NBOs vides type-non-Lewis (accepteur) et puis en estimant leurs énergies par la

théorie des perturbations à second ordre. Pour chaque donneur NBO(i) et accepteur NBO(j), l'énergie de stabilisation  $E^{(2)}$  est associée à la délocalisation  $i \rightarrow j$ . Elle est explicitement estimée par l'équation suivante:

$$E_{i \rightarrow j^*}^{(2)} = -n_i^{(0)} \frac{\langle \hat{\phi}_i^{(0)} | \hat{F} | \hat{\phi}_{j^*}^{(0)} \rangle^2}{\epsilon_{j^*}^{(0)} - \epsilon_i^{(0)}}$$

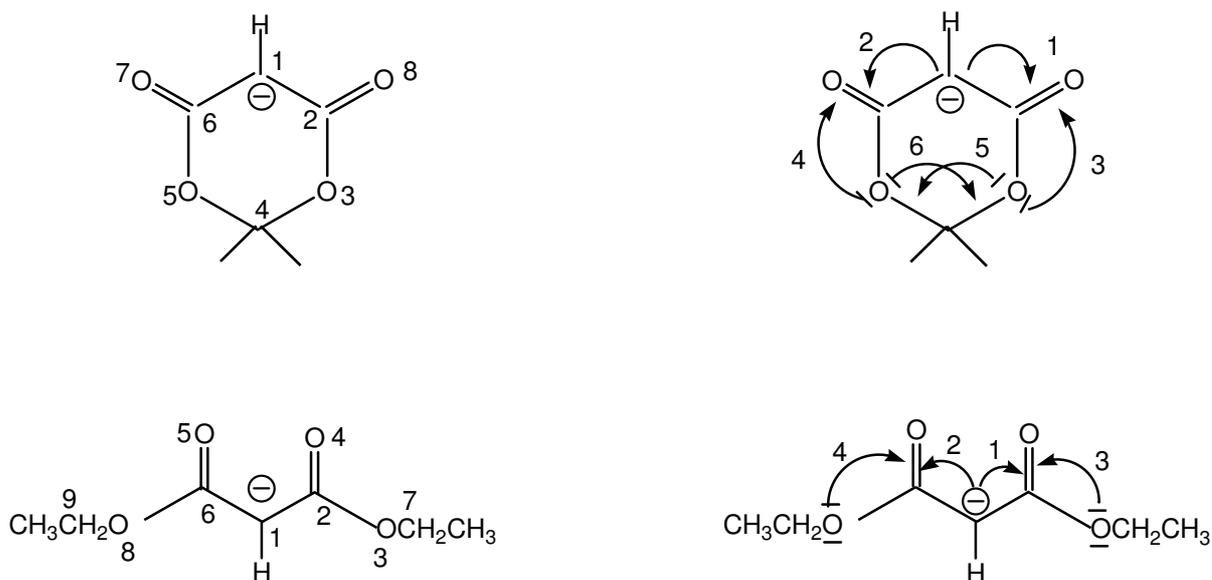
### 2.3. Etude des effets stéréoélectroniques sur l'acidité carbonée des acides $\beta$ -dicarboxylés.

Afin d'expliquer la différence d'acidité entre les acides carbonés dicarboxylés cycliques et leurs analogues à chaînes ouvertes, nous allons examiner toutes les interactions orbitales et les effets de délocalisation électronique dans les bases conjuguées de ces acides.

#### *Cas n° 1 : Acide de Meldrum / Malonate diéthyle:*

Pour simplifier la notation, des nombres ont été attribués aux interactions stéréoélectroniques les plus importantes dans l'anion de l'acide de Meldrum et le malonate diéthyle (voir **Schéma 1**). Dans le cas de l'acide de Meldrum, les interactions 1 et 2, comprenant l'orbitale antiliante du groupement carbonyle de  $\pi^*_{C=O}$ , sont équivalentes pour des raisons de symétrie. Même remarque est valable pour les interactions 3 et 4 mettant en jeu les oxygène ester.

Schéma 1

Tableau 2: Energies de stabilisation  $E(2)$  des principales interactions donneur-accepteur

<i>Acide de Meldrum</i>		<i>Malonate diéthylique</i>	
<i>Interaction</i>	<i><math>E(2)</math> (kcal/mol)</i>	<i>Interaction</i>	<i><math>E(2)</math> (kcal/mol)</i>
1 $n_{C1} \rightarrow \pi^*_{C2=O8}$	105.06	1 $n_{C1} \rightarrow \pi^*_{C2=O4}$	132.35
2 $n_{C1} \rightarrow \pi^*_{C6=O7}$	105.06	2 $n_{C1} \rightarrow \pi^*_{C6=O5}$	132.19
3 $n_{O3} (2) \rightarrow \pi^*_{C2=O8}$	22.68	3 $n_{O3} \rightarrow \pi^*_{C2=O4}$	29.70
4 $n_{O5} (2) \rightarrow \pi^*_{C6=O7}$	22.68	4 $n_{O8} \rightarrow \pi^*_{C6=O5}$	36.22
5 $n_{O3} (1) \rightarrow \sigma^*_{C4-O5}$	4.00		
$n_{O3} (2) \rightarrow \sigma^*_{C4-O5}$	6.80		
6 $n_{O5} (1) \rightarrow \sigma^*_{C4-O3}$	4.00		
$n_{O5} (2) \rightarrow \sigma^*_{C4-O3}$	6.80		

La charge négative dans les formes anioniques de l'acide de Meldrum et le malonate diéthyle  $G_1$  est fortement délocalisée sur les deux groupements carbonyles ( $C=O$ ) par des interactions de transfert de charge vicinal  $n_C \rightarrow \pi^*_{C=O}$ . En fait la l'acidité de ces composés dicarbonylés est attribuée au hydrogènes situés entre de groupements  $C=O$  et ceci est confirmé par les grandes valeurs de l'énergie de stabilisation ou de délocalisation  $E(2)$  de ces interactions  $n_C \rightarrow \pi^*_{C=O}$  dans l'acide de Meldrum et le malonate diéthyle (105 et 132 kcal/mol respectivement). Les deux composés présentent également les mêmes interactions type  $n_O \rightarrow \pi^*_{C=O}$  entre la paire libre de l'oxygène ester et la  $\pi^*$  du groupe carbonyle. L'énergie  $E(2)$  est de l'ordre 22.68 kcal/mol dans l'acide de Meldrum et de 29.70 et 36.22 kcal/mol dans malonate diéthyle (**Tableau 2**).

Dans le **Tableau 3** nous avons regroupé les occupations électroniques, calculées par la méthode NBO, des orbitales intervenants dans les différentes interactions donneur-accepteur dans l'acide de Meldrum et le malonate diéthyle.

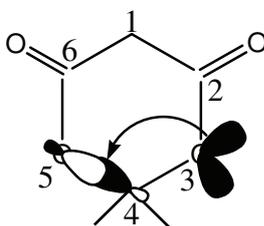
**Tableau 3** : les occupations des orbitales calculées par la méthode NBO

<i>Orbitale</i>	<i>Occupation</i>
<b>Acide de Meldrum</b>	
$n_{C1}$	1.350
$\pi^*_{C2=O8}$	0.405
$\pi^*_{C6=O7}$	0.405
$n_{O3(2)}$	1.856
$n_{O5(2)}$	1.856
$\sigma^*_{C4-O5} / \sigma^*_{C4-O3}$	0.100
<b>Malonate diéthyle</b>	
$n_{C1}$	1.370
$\pi^*_{C2=O4}$	0.440
$\pi^*_{C5=O6}$	0.431
$n_{O3(2)}$	1.855
$n_{O8(2)}$	1.836

L'analyse de ces occupations, pour les deux composés, montrent que les orbitales  $\pi^*_{C=O}$ , qui sont 'normalement' vacantes, possèdent une occupation de 0.40-0.44 électrons et l'occupation de la charge négative portée par le carbone C1 est inférieure à 2 électrons (1.35 et 1.37 électrons dans le cas de l'acide de Meldrum et le malonate diéthyle respectivement). Les deuxièmes paires libres des atomes d'oxygènes ester  $n_o$  sont également diminuées (1.836-1.865 e) à cause des interactions de transfert de charge. Ces résultats confirment bien les valeurs  $E(2)$  précédemment trouvées (**Tableau 2**).

En plus des interactions  $nc \rightarrow \pi^*_{C=O}$  et  $no \rightarrow \pi^*_{C=O}$ , l'acide de Meldrum présente d'autres interactions (qui sont absentes dans le malonate diéthyle) donneur-accepteur de type  $no \rightarrow \sigma^*_{C-O}$  entre les paires non liantes de l'oxygène ester et l'orbitale antiliante  $\sigma^*$  de la liaison vicinale. Ces interactions sont appelées stabilisation anomérique ou effets anomériques (**Schéma 2**). Ces effets sont très fréquents dans les composés cycliques bilactones tel que l'acide de Meldrum. L'énergie  $E(2)$  de l'ensemble ces interactions (4 orbitales des paires libres des deux oxygènes + les deux orbitales  $\sigma^*$  des liaisons  $\sigma^*_{C4-O5}$  et  $\sigma^*_{C4-O3}$ ) sont de l'ordre 21.6 kcal/mol (**Tableau 2**). Gao et col.[6] ont montré que la dissociation de l'acide de Meldrum est accompagnée par l'accroissement de l'énergie  $E(2)$  des effets anomériques. L'analyse des occupations des orbitales antiliantes  $\sigma^*_{C4-O5} / \sigma^*_{C4-O3}$  confirment ces interactions car ces orbitales sont occupées par 0.10 électrons.

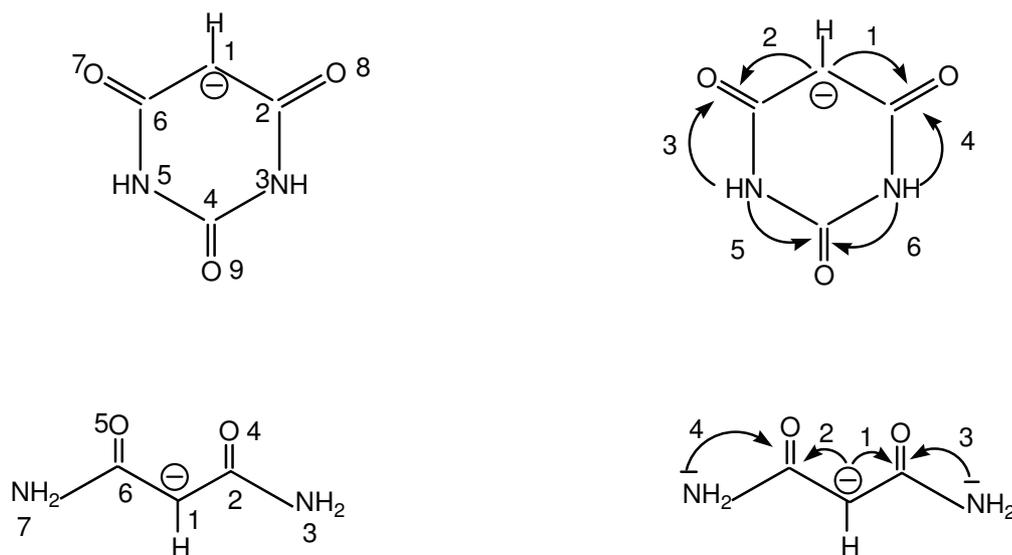
**Schéma 2**



Ces effets anomériques présents dans l'anion de l'acide de Meldrum, et qui sont totalement absents dans le malonate diéthyle, stabilisent d'avantage la base conjuguée de l'acide de Meldrum et par conséquent apportent une contribution importante dans l'accroissement de l'acidité du composé cyclique (acide de Meldrum) par rapport au composé analogue acyclique (le malonate diéthyle).

**Cas n° 2 : Acide barbiturique /Malonamide:**

Les principales interactions donneur-accepteur (et leurs énergies  $E(2)$ ) dues aux effets de délocalisation de l'acide barbiturique ( $pK_a=4.01$ ) et son composé analogue acyclique le malonamide ( $pK_a=12.5$ ) sont représentés ci-dessous (**Schéma 3, Tableau 4**).

**Schéma 3****Tableau 4:** Energies de stabilisation  $E(2)$  des principales interactions donneur- accepteur

<i>Acide barbiturique</i>		<i>Malonamide</i>	
<i>Interaction</i>	<i><math>E(2)</math> (kcal/mol)</i>	<i>Interaction</i>	<i><math>E(2)</math> (kcal/mol)</i>
1 $n_{C1} \rightarrow \pi^*_{C2=O8}$	79.07	1 $n_{C1} \rightarrow \pi^*_{C2=O4}$	97.78
2 $n_{C1} \rightarrow \pi^*_{C6=O7}$	78.29	2 $n_{C1} \rightarrow \pi^*_{C6=O5}$	97.78
3 $n_{N5} \rightarrow \pi^*_{C6=O7}$	49.60	3 $n_{N3} \rightarrow \pi^*_{C2=O4}$	19.68
4 $n_{N3} \rightarrow \pi^*_{C2=O8}$	50.39	4 $n_{O7} \rightarrow \pi^*_{C6=O5}$	19.68
5 $n_{N5} \rightarrow \pi^*_{C4=O9}$	37.42		
6 $n_{N3} \rightarrow \pi^*_{C4=O9}$	37.36		

La charge négative dans les formes anioniques, de l'acide barbiturique et le malonamide sur l'atome C<sub>1</sub> et fortement conjuguées sur les deux groupements carbonyles (C=O) par des interactions de transfert de charge vicinales  $n_{C1} \rightarrow \pi^*_{C=O}$  (notées 1 et 2). Ces interactions sont traduites par les grandes valeurs de l'énergie de stabilisation E(2) dans les deux composés (de l'ordre de 79 et 97.8 kcal/mol respectivement). L'acide barbiturique et le malonamide présentent également les mêmes interactions type  $n_N \rightarrow \pi^*_{C=O}$  entre la paire libre de l'azote et l'orbitale  $\pi^*$  du groupe carbonyle adjacent à la charge négative. L'énergie E(2) est de l'ordre 50.4 kcal/mol dans l'acide barbiturique et de 19.68 kcal/mol dans le malonamide.

Dans le **Tableau 5** suivant nous avons regroupé les occupations électroniques, calculées par la méthode NBO, des orbitales intervenants dans les différentes interactions donneur-accepteur dans le cas de l'acide barbiturique et son analogue le malonamide.

**Tableau 5 :** les occupations des orbitales calculées par la méthode NBO

<i>Orbitale</i>	<i>Occupation</i>
<i>Acide barbiturique</i>	
$n_{C1}$	1.379
$\pi^*_{C2=O8}$	0.423
$\pi^*_{C6=O7}$	0.422
$n_{N3}$	1.664
$n_{N5}$	1.663
$\pi^*_{C4=O9}$	0.339
<i>Malonamide</i>	
$n_{C1}$	1.348
$\pi^*_{C2=O4}$	0.420
$\pi^*_{C6=O5}$	0.420
$n_{N3}$	1.880
$n_{N5}$	1.880

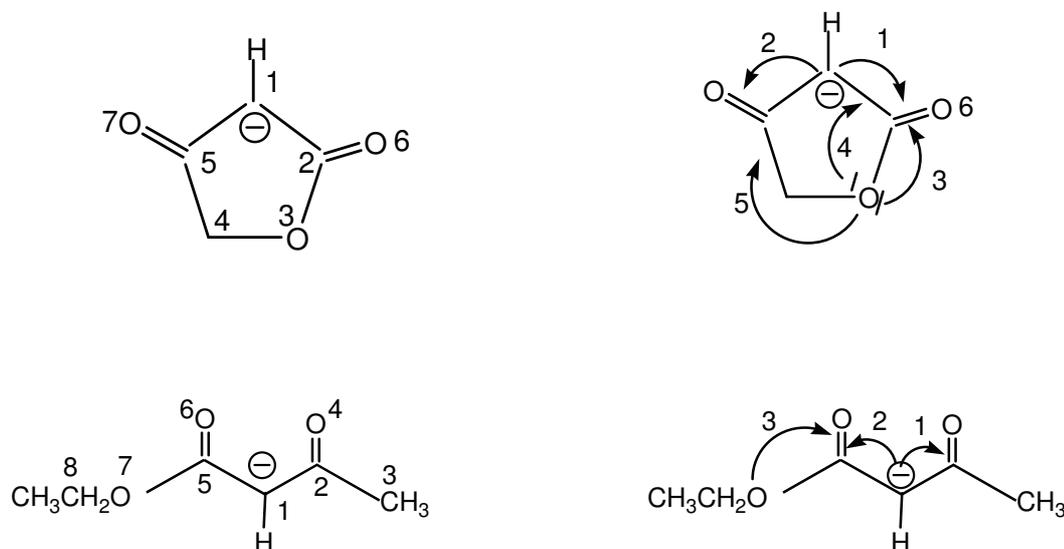
L'analyse des occupations des deux composés, montrent que les orbitales  $\pi^*_{C=O}$ , ne sont totalement vides et possèdent une occupation de 0.42 électrons et l'occupation de la charge négative portée par le carbone C1 est inférieure à 2 électrons (1.379 et 1.345 électrons dans le cas de l'acide barbiturique et malonamide respectivement). Les paires libres des atomes d'azotes  $n_N$  sont également diminuées (1.66-1.88 électrons) due aux interactions de transfert de charge  $n_N \rightarrow \pi^*_{C=O}$ . Ces résultats sont en bon accord avec le calcul des énergies d'interaction donneur-accepteur E(2) (**Tableau 5**).

Le calcul B3LYP/6-311++G\*\* des énergies libres de déprotonation donne une différence en phase aqueuse  $\delta\Delta G^\circ = \Delta G^\circ(\text{barbiturique}) - \Delta G^\circ(\text{malonamide}) = -23.5$  kcal/mol. La forte acidité de l'acide barbiturique par rapport au malonamide peut être expliquée par les fortes délocalisations entre les doublets libres des azotes et le carbonyle C4=O9 situé entre les deux azotes, désignées dans le **Schéma 3** par les interactions 5 et 6 qui sont inexistantes dans le malonamide. Ces interactions donnent une forte stabilisation supplémentaire de la base conjuguée et par conséquent augmentera l'acidité de l'acide barbiturique. L'énergie E(2) de stabilisation de ces interactions  $n_{N3}$  (ou  $n_{N5}$ )  $\rightarrow \pi^*_{C4=O9}$  sont de l'ordre 37.4 kcal/mol. Ces interactions stabilisantes sont analysée également par la forte occupation de l'orbitale antiliante  $\pi^*_{C4=O9}$  qui est de l'ordre de 0.339 électrons (**Tableau 5**).

### ***Cas n° 3 : Acide tétronique /Acétylacétate d'éthyle:***

Les principales interactions (et leurs énergies E(2)) due aux effets de délocalisation de l'acide tétronique  $pK_a=3.76$  et son composé analogue acyclique l'acétylacétate d'éthyle  $pK_a=10.7$  sont représentés dans le **Schéma 4** (**Tableau 6**).

Schéma 4

Tableau 6: Energies de stabilisation  $E(2)$  des principales interactions donneur-accepteur

Acide tétronique		Acétylacétate d'éthyle	
Interaction	$E(2)$ (kcal/mol)	Interaction	$E(2)$ (kcal/mol)
1 $n_{C1} \rightarrow \pi^*_{C2=O6}$	155.35	1 $n_{C1} \rightarrow \pi^*_{C2=O4}$	133.21
2 $n_{C1} \rightarrow \pi^*_{C5=O7}$	134.0	2 $n_{C1} \rightarrow \pi^*_{C5=O6}$	162.75
3 $n_{O3(2)} \rightarrow \pi^*_{C2=O6}$	72.44	3 $n_{O7} \rightarrow \pi^*_{C5=O6}$	41.04
4 $n_{O3(1)} \rightarrow \sigma^*_{C1-C2}$	6.40		
5 $n_{O3(1)} \rightarrow \sigma^*_{C4-C5}$	2.04		

La charge négative dans les formes anioniques de l'acide tétronique et de l'acétylacétate d'éthyle  $C_1$  est fortement délocalisée sur les deux groupements carbonyles ( $C=O$ ) par des interactions de transfert de charge vicinales  $nc \rightarrow \pi^*_{C=O}$ .

Les valeurs de l'énergie de stabilisation ou de délocalisation  $E(2)$  de ces interactions  $nc \rightarrow \pi^*_{C=O}$  dans l'acide tétronique et l'acétylacétate d'éthyle sont 155 et 162 kcal/mol respectivement. Les deux composés présentent les mêmes interactions type  $no \rightarrow \pi^*_{C=O}$  entre la paire libre de l'oxygène ester et la  $\pi^*$  du groupe carbonyle. L'énergie

E(2) est de l'ordre 72.55 kcal/mol dans l'acide tétronique et de 41 kcal/mol dans l'acétylacétate d'éthyle.

Dans le **Tableau 7** nous avons regroupé les occupations électroniques, calculées par la méthode NBO, des orbitales intervenants dans les différentes interactions donneur-accepteur dans l'acide tétronique et l'acétylacétate d'éthyle.

**Tableau 7** : les occupations des orbitales calculées par la méthode NBO

<i>Orbitale</i>	<i>Occupation</i>
<i>Acide tétronique</i>	
n <sub>C1</sub>	1.320
n <sub>O3(2)</sub>	1.794
$\pi^*_{C2=O6}$	0.458
$\pi^*_{C5=O7}$	0.457
n <sub>O3(2)</sub>	1.794
$\sigma^*_{C4-C5}$	0.063
$\sigma^*_{C1-C2}$	0.050
<i>Acétylacétate d'éthyle</i>	
n <sub>C1</sub>	1.327
$\pi^*_{C2=O4}$	0.398
$\pi^*_{C5=O6}$	0.450
n <sub>O3(2)</sub>	1.833

L'analyse de ces populations électroniques, pour les deux composés, montrent que les orbitales  $\pi^*_{C=O}$ , qui sont normalement vacantes, possèdent une occupation de 0.39-0.46 électrons. On note également que l'occupation de la paire électronique portée par le carbone C<sub>1</sub> (i.e. la charge négative) est inférieure à 2 électrons (1.32 et 1.327 électrons dans le cas de l'acide tétronique et l'acétylacétate d'éthyle respectivement). La deuxième paire libre des atomes d'oxygènes ester n<sub>O</sub> sont également diminuées (1.794-1.833 électrons) à

cause des interactions de transfert de charge. Ces résultats sont en bon accord avec le calcul des énergies E(2) (**Tableau 6**).

Comme dans le cas de l'acide de Meldrum et de l'acide barbiturique, l'acide tétronique présente d'autres interactions donneur-accepteur  $n_{\text{O}} \rightarrow \sigma^*$  entre les paires non liante de l'oxygène ester et des orbitales moléculaires antiliante  $\sigma^*$ . L'énergie E(2) de l'ensemble de ces interactions 4 et 5 ( $n_{\text{O3(1)}} \rightarrow \sigma^*_{\text{C1-C2}} + n_{\text{O3(1)}} \rightarrow \sigma^*_{\text{C4-C5}}$ ) est de l'ordre 8.5 kcal/mol. L'analyse des occupations des orbitales antiliantes  $\sigma^*_{\text{C1-C2}} / \sigma^*_{\text{C4-C5}}$  confirment ces interactions car ces orbitales sont occupées par 0.050/0.063 électrons. Ces interactions présents dans l'anion de l'acide tétronique, et qui sont totalement absentes dans l'acétoacétate d'éthyle, augmenteront les effets de délocalisation dans le cycle à 5 chaînons et stabilisent d'avantage l'anion de l'acide tétronique et par conséquent apportent une contribution importante dans l'accroissement de l'acidité du composé cyclique par rapport au composé analogue acyclique ( $\delta pK_a = pK_a(\text{TA}) - pK_a(\text{EAA}) = 6.96$ ).

### 3. Conclusion

Dans la partie 4, nous avons présenté une étude théorique basée sur l'analyse NBO pour expliquer l'acidité remarquablement élevée des acides de Meldrum, barbiturique et tétronique en la comparant avec celle des composés analogues à chaînes ouvertes. Nous avons rationalisé le rôle des effets stéréoelectroniques sur la stabilité des bases conjuguées des acides  $\beta$ -dicarboxylés étudiés et nous avons remarqué que les bases conjuguées des acides  $\beta$ -dicarboxylés cycliques présentent des interactions stabilisantes dites donneur-accepteur. Ces interactions stabilisantes, qui sont totalement absentes dans les composés analogues à chaînes ouvertes, apportent une contribution importante dans l'accroissement de l'acidité des composés  $\beta$ -dicarboxylés cycliques par rapport à leurs analogues acycliques.

## Bibliographie

- [1] (a) M.C. Rezende, *Tétrahedron* 57 (2001) 5923.  
(b) E.P. Serjeant, B. Dempsey, « Constantes de l'ionisation d'Acides Organiques dans Solution aqueuse », IUPAC Série des Données Chimique, Pergamon Press, New York, 1979.  
(c) R.G. Pearson, R.L. Dillon, *J. Est. Chem. Soc.* 75 (1953) 2439.  
(d) F.G. Bordwell, *Acc. Chem. Loi.* 21 (1988) 456, 463.
- [2] E. M Arnett, J. A. Harrelson, *J. Am. Chem. Soc.* 109 (1987) 809.
- [3] X.Wang, K. N. Houk, *J. Am. Chem. Soc.* 110 (1988) 1870.
- [4] K. B. Wiberg, K. E. Laidig, *J. Am. Chem. Soc.* 110 (1988) 1872.
- [5] J. D. Evanseck, K. N. Houk, J. M. Briggs, W. L. Jorgensen, *J. Am. Chem. Soc.* 116 (1994)10630.
- [6] K. Byun, Y. Mo, J. Gao, *J. Am. Chem. Soc.* 123 (2001) 3974.
- [7] J. P. Foster, F. Weinhold, *J. Am. Chem. Soc.* 102 (1980) 7211.
- [8] A. E. Reed, F. Weinhold, *J. Chem. Phys.* 78 (1983) 4066.
- [9] A. E. Reed, R. B. Weinstock, F. Weinhold, *J. Chem. Phys.* 83 (1985) 735.
- [10] A. E. Reed, F. Weinhold, *J. Chem. Phys.* 83 (1985) 1736.
- [11] (a) Schindler, M. Kutzelnigg, *W. J. Chem. Phys.* 76 (1982) 1919.  
(b) Schindler, M. Kutzelnigg, *W. J. Am. Chem. Soc.* 105 (1983) 1360.  
(c) M. Schindler, W. Kutzelnigg, *Mol. Phys.* 48 (1983) 781.  
(d) W. Kutzelnigg, "In Nuclear Magnetic Shieldings and Molecular Structure"; NATO ASI Series C; Tossel, J. A., Ed.; Kluwer: Boston, Vol. 386, 1993.
- [12] J. A. Bohmann, F. Weinhold, T. C. Farrar, *J. Chem. Phys.* 107 (1997) 1173.
- [13] J. E. Peralta, R. H. Contreras, J. P. Snyder, *Chem. Commun.* (2000) 2025.

## ***CONCLUSION GENERALE***

Le travail présenté dans ce travail est articulé sur quatre parties:

- i) L'élaboration des corrélations quantitatives des constantes d'acidité aqueuse avec les énergies libres de Gibbs de déprotonation des acides carbonés étudiées.
- ii) La prédiction des constantes d'acidité aqueuse ( $pK_a$ ) des acides carbonés en développant des modèles de corrélation quantitatives structure-acidité (QSPR) en utilisant les méthodes de régression linéaire simple et multiple.
- iii) L'étude théorique des effets de solvant sur l'équilibre tautomère céto-énolique des composés  $\beta$ -dicarboxylés cycliques et de leurs analogues à chaînes ouvertes.  
1<sup>er</sup> cas : Acide tétronique / Acétylacétate d'éthyle  
2<sup>ème</sup> cas : Dimédone /Acétylacétone
- iv) Etude théorique des effets stéréoelectroniques sur l'acidité carbonée des acides  $\beta$ -dicarboxylés à l'aide de la méthode NBO (Natural Bond Orbital).

Dans la 1<sup>ère</sup> partie, nous avons élaboré des corrélations linéaires satisfaisantes entre les constantes d'acidité ( $pK_a$ ) expérimentales aqueuses d'une variété d'acides carbonés et les énergies libres de Gibbs de déprotonation calculées en phase gazeuse et en phase aqueuse. La corrélation linéaire obtenue en phase gazeuse est légèrement meilleure ( $R^2=0.969$ ,  $SD=3.27$ ) en la comparant avec celle obtenue en phase aqueuse ( $R^2=0.934$ ,  $SD=4.80$ ).

La différence d'acidité des composés de la série d'étude peut être interprétée par les effets inductifs et les effets de résonance. L'acidité élevée de certains acides carbonés a été rationalisée en calculant l'énergie de transfert de charge  $n_C \rightarrow \pi^*$  entre l'orbitale non-liante localisée sur l'atome de carbone de la base conjuguée (la charge négative  $n_C$ ) et l'orbitale antiliante vacante adjacente  $\pi^*$  des groupements C=C, C=O ou N=O.

Dans la 2<sup>ème</sup> partie, plusieurs modèles QSPR faisant intervenir des descripteurs quantiques des acides et de leurs bases conjuguées ont été élaborés pour prédire les valeurs expérimentales de  $pK_a$  aqueuses des acides carbonés. Quatre composés non inclus dans la série d'étude ont été utilisés pour tester le pouvoir prédictif des meilleurs modèles QSPR obtenus. Dans le cas des calculs B3LYP/6-311++G(d,p)-PCM(opt), le modèle QSPR le plus

simple comporte l'énergie libre aqueuse de déprotonation ( $\Delta G_{aq}$ ) et la variation de la charge NBO sur le carbone basique de la base conjuguée ( $\Delta Q_C$ ) avec une estimation de l'ordre de 1.61 unité de  $pK_a$ . En ajoutant un troisième descripteur  $\Delta\eta^{-1}$  (la variation de la mollesse en allant de l'acide à sa base conjuguée) au modèle précédent, on remarque une légère amélioration de la prédiction estimée à 1.52 unité  $pK_a$  par rapport aux valeurs expérimentales. Dans le cas des calculs B3LYP/6-311++G(d,p)-PCM(sp) en utilisant l'énergie  $E_{aq}$  dans le calcul de  $\Delta G_{aq}$  (l'équation 4), le modèle QSPR le plus simple comporte l'énergie libre aqueuse de déprotonation ( $\Delta G_{aq}$ ) et la variation de la charge NBO ( $\Delta Q_C$ ) avec une estimation de l'ordre de 1.50 unité de  $pK_a$ . En ajoutant un troisième descripteur  $\Delta\eta^{-1}$  au modèle précédent, on remarque une amélioration de la prédiction estimée à 1.37 unité  $pK_a$  par rapport aux valeurs expérimentales. Les meilleures équations QSPR trouvées pourraient être utiles pour prédire les valeurs aqueuses de  $pK_a$  des acides carbonés très forts ( $pK_a < pK_a(H_3O^+) = -1.74$ ) et très faibles ( $pK_a > pK_a(H_2O) = 15.74$ ) pour lesquels il est difficile d'obtenir des données expérimentales suffisamment précises.

Dans la partie 3, nous avons étudié théoriquement la structure de l'acétylacétate d'éthyle (EAA). Les calculs ont montré que le tautomère dicétonique le plus stable est le dicéto-trans, E, Z (forme faucille); tandis que le tautomère énol le plus stable correspond à la structure cis-énol dans laquelle l'énolisation est de côté du groupement carbonyle. Dans le cas de l'acide tétronique TA le tautomère dicétonique est de la forme trans, E, E tandis que le tautomère trans-énol le plus stable correspond à la structure dans laquelle l'énolisation est adjacente au groupement carbonyle. Les calculs de la variation de l'enthalpie libre montrent que l'équilibre de céto-énol du composé EAA est déplacé vers le tautomère dicétonique dans les solvants polaires; tandis que l'équilibre céto-énol du TA est déplacé vers le tautomère énolique dans les solvants polaires. Ces tendances sont en bon accord avec des résultats expérimentaux. Nos calculs révèlent également que le terme entropique dans le calcul de la variation l'énergie libre de Gibbs des équilibre de l'EAA et du TA est minime et par conséquent le terme enthalpique est prédominant dans la détermination de la constante d'équilibre. Comme tous les composés  $\beta$ -dicarbonylés acyclique la forme dicétonique la plus stable de l'acétylacétone (AcAc) est la forme trans

E, Z (forme faucille); tandis que l'énol le plus stable (par les liaisons hydrogènes intramoléculaires) correspond à la structure cis-énol. La dimédone existe sous deux tautomères stables : la forme dicétonique trans E,E et la forme trans-énol stabilisée par les liaisons hydrogènes intermoléculaires. Les calculs de la variation de l'énergie libre de Gibbs montrent que l'équilibre de céto-énol composé AcAc est déplacé vers le tautomère dicétonique dans les solvants polaires; tandis que l'équilibre céto-énol de la dimédone est déplacé vers le tautomère énolique dans les milieux polaires. Les tendances trouvées sont en bon accord avec des résultats expérimentaux.

Dans la partie 4, nous avons présenté une étude théorique basée sur l'analyse NBO (Natural Bond Orbital) pour expliquer l'acidité remarquablement élevée des acides de Meldrum, barbiturique et tétronique en la comparant avec celle des composés analogues à chaînes ouvertes. Nous avons rationalisé le rôle des effets stéréoelectroniques sur la stabilité des bases conjuguées des acides  $\beta$ -dicarboxylés étudiés et nous avons remarqué que les bases conjuguées des acides  $\beta$ -dicarboxylés cycliques présentent des interactions stabilisantes dites donneur-accepteur. Ces interactions stabilisantes, qui sont totalement absentes dans les composés analogues à chaînes ouvertes, apportent une contribution importante dans l'accroissement de l'acidité des composés  $\beta$ -dicarboxylés cycliques par rapport à leurs analogues acycliques.

Comme perspectives à ce travail, nous envisageons :

- Elaboration des modèles QSPR faisant intervenir des descripteurs quantiques pour prédire les constantes d'acidité et de basicité pour d'autres familles de composés.
- Etude des mécanismes réactionnels des équilibres céto-énoliques des composés  $\beta$ -dicarboxylés cycliques et acycliques en utilisant la théorie de l'état de transition.
- Approfondir l'étude théorique basée sur l'approche NBO, en utilisant la version 5, pour mettre en évidence tous les effets (transfert de charge, énergie électrostatique, énergie de déformation) sur l'acidité des acides  $\beta$ -dicarboxylés cycliques et leurs analogues à chaînes ouvertes.

## ملخص

- يهدف العمل المقدم في هذه الأطروحة إلى:
- 1 - إيجاد علاقات كمية بين ثوابت الحموضة المائية للأحماض الكربونية مع المؤشرات النظرية مثل الطاقات الحرة للتشرد.
  - 2 - تنبؤ ثوابت الحموضة المائية للأحماض الكربونية عن طريق بلورة علاقات كمية بنية-حموضة (دراسة QSPR).
  - 3 - الدراسة النظرية لتأثير المذيبات على التوازن سيتون- إينول للمركبات الكربونية  $\beta$ -dicarbonyl الحلقية ونظيراتها ذات السلاسل المفتوحة.  
الحالة 1: اثيل استيل اسيتات / حمض تيترونيك  
الحالة 2: استيل اسيتون / ديميدون
  - 4 - الدراسة النظرية للأفعال الفراغية-الالكترونية على الحموضة المائية للأحماض الكربونية  $\beta$ -dicarbonyl الحلقية ونظيراتها ذات السلاسل المفتوحة باستعمال طريقة NBO.

قمنا بهذه الدراسة النظرية باستعمال مقاربات نظرية إحصائية وكمية  
قمنا بالحسابات باستعمال برنامج Gaussian و بالطرق DFT/B3LYP/6-311++G\*\* و MP2/6-31+G\*\*

**الكلمات المفتاحية:** الحموضة الكربونية، ثابت الحموضة المائية، مؤشرات نظرية، دراسة QSPR، المركبات  $\beta$ -dicarbonyl الحلقية و اللاحلقية، توازن سيتون- إينولي، تأثير المذيب، نموذج PCM، تحليل NBO، الأفعال الفراغية-الالكترونية.

## Résumé

Le travail présenté dans cette thèse a pour objectifs :

- 1- Elaboration des corrélations quantitatives entre les constantes d'acidité aqueuse et les énergies libres de déprotonation des acides carbonés étudiés.
- 2- Prédiction des constantes d'acidité aqueuse ( $pK_a$ ) des acides carbonés en développant des modèles de corrélation quantitatives structure-acidité (QSPR: Quantitative Structure-property relationships) en utilisant les méthodes de régression linéaires.  
- Etude théorique des effets de solvant sur l'équilibre tautomère céto-énolique des composés  $\beta$ -dicarbonylés cycliques et de leurs analogues à chaînes ouvertes.  
1<sup>er</sup> cas : Acide tétronique / Acétylacétate d'éthyle  
2<sup>ème</sup> cas : Dimédone /Acétylacétone
- 4- Etude théorique des effets stéréoelectroniques sur l'acidité carbonée des acides  $\beta$ -dicarbonylés cycliques et de leurs analogues à chaînes ouvertes à l'aide de la méthode NBO (Natural Bond Orbital).

Le travail a été mené à l'aide des approches statistiques, en l'occurrence la régression linéaire simple et multiple et des approches quantiques. Les calculs ont été effectués avec le programme *Gaussian* en utilisant les méthodes DFT/B3LYP/6-311++G\*\* et MP2/6-31+G\*\*.

**Mots-Clés :** Acidité carbonée ;  $pK_a$  aqueux ; Descripteurs théoriques ; QSPR ; Composés  $\beta$ -dicarbonylés cycliques et acycliques ; Equilibre céto-énolique ; Effet de solvant ; Modèle PCM ; Analyse NBO, effets stéréoelectroniques.

## Abstract

The goal of the work presented in this thesis is:

- 1- Elaboration of quantitative correlations between the aqueous acidity constants and deprotonation free energies of carbon acids under study.
- 2 - Prediction of aqueous acidity constants ( $pK_a$ ) of carbon acids by developing QSPR (Quantitative Structure-property relationships) models using linear regression methods.
- 3- Theoretical study of the solvent effects on the tautomeric keto-enol equilibrium of cyclic  $\beta$ -dicarbonyl compounds and their analogues with open chains.  
1st case: Tetric acid/ethyl Acetoacetate  
2nd case: Dimédone /Acetylacetone
- 4- Theoretical study of the stereoelectronic effects on carbon acidity of the cyclic  $\beta$ -dicarbonyl compounds and their analogues with open chains using NBO (Natural Bond Orbital) method.

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\*\* and MP2/6-31+G\*\* methods.

**Keywords:** Carbon acidity ; aqueous  $pK_a$  ; Theoretical descriptors ; QSPR ; cyclic and acyclic  $\beta$ -dicarbonyl compounds ; Keto-Enol tautomeric equilibrium ; Solvent effects ; PCM model ; NBO analysis ; stereoelectronic effects.