Prediction of the equilibrium properties in food and biological systems with the COSMO-RS model

Oumar Toure

UNIVERSITE BLAISE PASCAL **UNIVERSITE D'AUVERGNE**

N° D. U. 2432 **ANNEE : 2014**

ECOLE DOCTORALE SCIENCES POUR L'INGENIEUR

N° d'ordre : EDSPIC 639

Thèse

Soutenue publiquement le 10 janvier 2014

Par

Oumar TOURE

Ingénieur en Génie Physique

Pour l'obtention du grade de

DOCTEUR D'UNIVERSITE

(SPECIALITE: GENIE DES PROCEDES)

Prédiction des propriétés d'équilibre dans les milieux biologiques et alimentaires par le modèle COSMO-RS

Devant le jury composé de :

Directeur de thèse :

M. DUSSAP Claude-Gilles, Professeur, Université Blaise Pascal, Clermont- Ferrand.

Président :

M. LEBERT André, Professeur, Université Blaise Pascal, Clermont- Ferrand

Rapporteurs :

M. JOULIA Xavier, Professeur, INP-ENSIACET, Toulouse.

M. LEGRAND Jack, Professeur, Université de Nantes, Saint Nazaire.

Examinateurs :

M. KLAMT Andreas, Professeur, COSMOlogic GmbH & Co., Leverkusen, Allemagne

M. DE HEMPTINNE Jean-Charles, Professeur, IFP Energies Nouvelles, Paris

M. BAUDOUIN Olivier, Directeur des procédés, ProSim S.A.

M. GROS Jean-Bernard, Professeur émérite, Université Blaise Pascal, Clermont- Ferrand.

Institut Pascal, Axe Génie des Procédés, Energétique et Biosystèmes - Université Blaise Pascal - CNRS UMR 6602

UNIVERSITE BLAISE PASCAL **UNIVERSITE D'AUVERGNE**

N° D. U. 2432 **YEAR : 2014**

ECOLE DOCTORALE SCIENCES POUR L'INGENIEUR

Order no: EDSPIC 639

Thesis

Defended publicly on January, 10, 2014

By

Oumar TOURE

Ingénieur en Génie Physique

To obtain the degree of

DOCTOR OF PHILOSOPHY

(SPECIALITY: CHEMICAL ENGINEERING)

Prediction of the equilibrium properties in food and biological systems with the COSMO-RS model

---

Members of the defense jury:

Thesis director:

Mr. DUSSAP Claude-Gilles, Professor, Université Blaise Pascal, Clermont-Ferrand

President:

Mr. LEBERT André, Professor, Université Blaise Pascal, Clermont- Ferrand

Reporters:

Mr. JOULIA Xavier, Professor, INP-ENSIACET, Toulouse

Mr. LEGRAND Jack, Professor, Université de Nantes, Saint Nazaire

Examiners:

Mr. KLAMT Andreas, Professor, COSMOlogic GmbH & Co., Leverkusen, Germany

Mr. DE HEMPTINNE Jean-Charles, Professor, IFP Energies Nouvelles, Paris

Mr. BAUDOUIN Olivier, Process Manager, ProSim S.A., Toulouse

Mr. GROS Jean-Bernard, Emeritus Professor, Université Blaise Pascal, Clermont- Ferrand

Institut Pascal, axe Génie des Procédés, Energétique et Biosystèmes -
Université Blaise Pascal - CNRS UMR 6602

I dedicate this work to

- ✓ all my family,
- ✓ all my friends,
- ✓ and all Malian people.

I hope that you will find here

the expression of my sincere gratitude

for your daily support.

Je dédie ce travail à

- ✓ toute ma famille,
- ✓ tous mes a mi(e)s,
- ✓ et à tout le peuple malien.

J'espère que vous trouverez ici

l'expression de ma sincère gratitude

pour votre soutien quotidien.

Remerciements / Acknowledgements

First of all, I would like to thank **God** for providing me the health, the courage, the patience and the perseverance without which it would be impossible to undertake this work.

I sincerely express my gratitude to the French National Research Agency (ANR) and Auvergne council who provide trough the "Na⁻" project the financial support for this thesis.

I sincerely thank my supervisor Prof. Claude Gilles Dussap, Director of Axe GePEB of Institut Pascal, for his support, his useful advices and his availability; even though he was very busy he manages out the time to enable me to achieve this thesis work. Likewise, I would like to thank Prof. André Lebert, Coordinator of the "Na⁻" project and through him all the partners of this project who give to me advices and their constant support during all this PhD period. I acknowledge Prof. Jean-Bernard Gros for his constant availability even though he is retired, he did not manage any effort for providing useful advices to me.

I am very grateful to Prof. Xavier Joulia from INP-ENSIACET (Toulouse) and Prof. Jack Legrand from Université de Nantes (Saint Nazaire) who have both kindly accepted to report this work even if they are very busy. My sincere gratitude goes to the other members of the jury namely: Prof. Andreas Klamt from COSMOlogic (Leverkusen, Germany); Prof. Jean-Charles De Hemptinne from IFP Energies Nouvelles (Paris), and Dr. Olivier Baudouin from ProSim S.A. (Toulouse), for accepting to evaluate this thesis.

My heartfelt thanks to all faculty members of Axe GePEB: Axe GePEB was like a home far away from home to me.

My parents, siblings, and friends with their constant support, encouragement and prayers have helped me to move on all these years.

Résumé étendu

La simulation des procédés et leur optimisation nécessitent la connaissance des propriétés physico-chimiques des corps purs et des mélanges qui y sont mis en œuvre. Comme les mélanges contiennent la plupart du temps un nombre élevé de constituants, il est nécessaire que ces propriétés soient modélisées. C'est le champ d'étude de la thermodynamique appliquée qui permet, à partir d'une représentation moléculaire des corps purs et des mélanges, d'assurer la cohérence des données expérimentales et de garantir la robustesse de la représentation. Les logiciels de simulation sont donc développés en parallèle avec des bases de données relatives aux corps purs et une banque de modèles thermodynamiques pour estimer les propriétés des solutions.

Dans les procédés des industries biologiques ou alimentaires, on trouve systématiquement des milieux aqueux complexes contenant des espèces dont la diversité est importante (eau, solvants organiques, solides dissous, gaz dissous, espèces ioniques, macromolécules). La modélisation de tels procédés doit donc tenir compte de la complexité de ces milieux et nécessite de pouvoir traiter simultanément des équilibres chimiques (dissociation, complexation, oxydoréduction,…) et des équilibres physiques (liquide-vapeur, solubilité,…). Il est démontré que ceci est possible si l'on dispose de données d'enthalpies libres, d'enthalpies et de capacités calorifiques fiables.

Le potentiel chimique ou enthalpie libre molaire de formation est la donnée indispensable pour modéliser les équilibres. La base de la modélisation thermodynamique repose sur la fiabilité et la cohérence du calcul du potentiel chimique des différents constituants à partir de l'enthalpie libre totale G.

$$\mu_i = g_i = \left(\frac{\partial G}{\partial n_i}\right)_{T,p,n_{j\neq i}} = \mu_i^0 + RT\ln\left(\gamma_i x_i\right)$$

Cette équation montre que la détermination du potentiel chimique μ_i d'un constituant dépend de la connaissance de 2 variables : l'enthalpie libre de formation μ_i^0 dans l'état de référence choisi, et le coefficient d'activité γ_i qui dépend aussi de l'état de référence choisi (étant donné que la température et la composition du mélange sont connues). La connaissance des enthalpies et des capacités calorifiques partielles revient à déterminer l'enthalpie libre de formation et le coefficient d'activité en fonction de la température.

Le travail de modélisation thermodynamique revient donc au calcul prédictif et/ou à la collecte de données de propriétés de formation et à l'utilisation et/ou au développement d'un modèle prédictif de coefficient d'activité.

Modélisation des coefficients d'activité :

Les coefficients d'activité permettent de traduire l'écart à l'idéalité des solutions et d'accéder à l'ensemble des propriétés thermodynamiques (activité de l'eau, coefficient osmotique)

et d'équilibres entre phases (équilibres liquide-liquide, liquide-vapeur et liquide-solide), à partir des relations générales de la thermodynamique.

Après une revue de la littérature sur les outils existants (chapitre I), notre choix s'est porté sur le modèle COSMO-RS qui est un modèle entièrement prédictif combinant une approche de thermodynamique statistique et une approche de chimie quantique pour prédire les propriétés physico-chimiques des solutions aqueuses. L'objectif de ce travail est d'étendre les performances de ce modèle (largement utilisé en génie chimique) au traitement des milieux biologiques et alimentaires. On peut résumer la présentation de la méthode COSMO-RS comme suit (chapitre II).

Le point de départ est la simulation (par des outils de calculs quantiques) de la molécule immobilisée, d'une part dans le vide et, d'autre part, dans un état hypothétique qui est celui d'un conducteur parfait. Il est important de souligner que la précision de la méthode COSMO dépend beaucoup de la méthode de calcul quantique choisie. Afin d'effectuer ces calculs, les rayons-COSMO de chaque atome doivent être connus. Pour les atomes les plus courants dans les molécules organiques (H, C, N, O, F, P, S, Cl, Br ou I) les rayons-COSMO ont été déterminés à partir des valeurs des rayons de van der Waals (vdW) disponibles dans la littérature (Bondi 1964, Mantina et al. 2009). En général, ceux-ci sont multipliés par un facteur (1,17) pour obtenir les rayons-COSMO (Klamt 1995, 1998, 2005).

Dans le conducteur parfait (état COSMO), les lois de l'électrostatique stipulent que le potentiel électrostatique global (ϕ) sur la surface d'une molécule individuelle doit être nul. Par contre, en supposant que chaque élément de surface est isolé, le système est équivalent à un ensemble d'éléments de surface (s_i) avec des potentiels locaux (ϕ_i) et des charges (q_i). Le modèle COSMO-RS (COSMO Real Solvation) repose sur cette image. Le traitement avec des méthodes de physique statistique d'ensembles d'éléments chargés en interaction de contact électrostatique permet de déterminer les propriétés d'excès. Pour une molécule donnée, chaque élément de surface a donc une densité surfacique de charge, encore appelée charge de polarisation ($\sigma_i = q_i/s_i$). Afin de déterminer cette densité, les calculs de chimie quantique sont effectués afin d'établir la structure électronique de la molécule (structure et répartition des électrons), en prenant en compte les atomes qui la constituent et sa forme. Dans le modèle COSMO-RS, une molécule est donc considérée comme un ensemble de surfaces identiques ayant chacune leurs densités surfaciques de charge. Dans ce contexte, une molécule donnée peut être décrite par la répartition des charges à sa surface. Cette dernière prend la forme d'une densité de probabilité dénommée σ-profile qui relie la valeur de la densité surfacique de charge σ (en e-/Å²) au nombre d'éléments de surface $p(\sigma)$ ayant la densité de charge σ. Ainsi, l'écart énergétique entre la molécule dans un conducteur parfait (état COSMO) et la molécule dans un solvant réel (état liquide) peut être exprimé. Le processus qui consiste à faire

passer la molécule d'une position fixe dans l'état gaz parfait à la même position dans la solution est dénommé ***solvatation*** (Ben Naim 1987). Le modèle COSMO-RS est d'ailleurs un très bon modèle de solvatation ; le liquide est considéré comme un ensemble de surfaces chargées (interagissant deux à deux). Ce modèle permet donc de prendre en compte les interactions physiques dites de courte portée entre les molécules grâce à l'étude des contacts entre les surfaces chargées qui les constituent par une approche de thermodynamique statistique.

L'élimination des couches de conducteur qui n'existent pas dans la vraie solution permet d'obtenir le même état de référence pour toutes les molécules, ce qui renforce la robustesse du modèle. La contribution énergétique résultante de cette phase « d'élimination des couches de conducteur » est dénommée E_{misfit}. Cependant, il apparaît indispensable de traiter les autres types d'interactions. Ce traitement est assuré dans la méthode COSMO-RS avec une approche similaire à celle adoptée pour les interactions de nature électrostatique, c'est-à-dire des interactions entre paires de surfaces chargées, en particulier pour prendre en compte la liaison hydrogène. Les interactions de type van der Waals sont également prises en compte *a posteriori* dans le modèle.

Enfin la prise en compte de ces différentes interactions énergétiques dans les calculs de thermodynamique statistique sur le système permet de déterminer le potentiel chimique (μ_i) de chaque molécule *i* présente dans la solution, et donc d'en déduire toutes les propriétés d'excès (activité de l'eau a_w, coefficients d'activité γ_i des constituants, coefficient osmotique Π, propriétés d'équilibre liquide-liquide, *etc.*).

Quand la solution contient des électrolytes (c'est-à dire des espèces chargées capables de se dissocier dans l'eau), la seule prise en compte des interactions de courte portée ne suffit plus pour traduire la réalité de la solution. Nous avons développé lors de ce travail de thèse, un nouveau modèle permettant d'étendre les performances du modèle COSMO-RS aux solutions aqueuses contenant des électrolytes (plus spécifiquement des binaires eau-sel). Le modèle résultant dénommé COSMO-RS-PDHS (présenté en détails dans le chapitre III) est une combinaison du modèle COSMO-RS (pour prendre en compte les interactions physiques de courte portée ou SR) avec un terme PDH (pour prendre en compte les interactions physiques de longue portée ou LR) ainsi qu'un terme chimique prenant en compte la solvatation fixe des ions par les molécules d'eau. Des travaux antérieurs menés au sein de l'Axe GePEB de l'Institut Pascal ont montré que la prise en compte robuste du phénomène d'hydratation (ou de solvatation) des ions est indispensable pour prédire correctement les coefficients d'activité des ions et l'activité de l'eau (Achard 1992, Ben Gaïda 2007). D'un point de vue modélisation (moléculaire), la représentation de la solvatation fixe des ions par les molécules d'eau passe par la détermination de la structure moléculaire du cluster (ensemble ion + n_{hi} molécules d'eau) ainsi que par l'analyse de l'effet de l'hydratation sur les propriétés d'excès des constituants de la solution.

Pour remédier à l'absence de données concernant les rayons des cations et ainsi déterminer leurs structures hydratées à dilution infinie dans l'eau (en particulier la valeur maximale *n* du nombre d'hydratation n_{hi}), nous avons développé 4 procédures complémentaires aboutissant à la détermination du nombre d'hydratation d'un ion aussi bien dans l'état gaz (3 procédures) qu'à dilution infinie dans l'eau (1 procédure). Il est important de préciser que pour effectuer les calculs dans l'état gaz, aucune donnée concernant les rayons-COSMO n'est requise. Partant de là, nous avons déterminé les distances ion-eau dans les différentes structures hydratés d'un ion donné pour les comparer avec celles disponibles dans la littérature. Il est apparu ainsi que les distances $d_{ion\text{-}eau}$ calculées dans notre étude sont en très bon accord avec les valeurs données par Marcus (1988, 1997 et 2012) ; ceci nous a amené à choisir les rayons ioniques proposés par Marcus pour déterminer les rayons COSMO des ions monoatomiques étudiés (en les multipliant par le facteur 1,17). Ces valeurs ont donc été utilisées par la suite pour effectuer les calculs COSMO et prédire les valeurs maximales du nombre d'hydratation des cations à dilution infinie dans l'eau. Une démarche similaire peut être développée pour les anions.

Connaissant les valeurs du nombre d'hydratation et la structure des espèces hydratées, nous avons pu tester les performances prédictives du modèle COSMO-RS-PDHS sur quelques solutions binaires (eau-sel). Les résultats obtenus avec cette approche entièrement prédictive (chapitre III) sont très satisfaisants pour des molalités de sels allant jusqu'à 6 moles/kg d'eau. Ces résultats très encourageants, peuvent être encore améliorés si l'on dispose de suffisamment de données pour traiter les équilibres chimiques entre les différentes espèces hydratées et non hydratées d'un ion, et ainsi prendre en compte l'hydratation variable des ions. Ben Gaïda (2007) a montré qu'une telle approche pourrait étendre le domaine de validité du modèle jusqu'à 20 moles/kg en molalité de sel. Cependant, le développement d'un tel modèle requiert la connaissance simultanée des nombres d'hydratation et des constantes d'équilibre des réactions d'hydratation (K) donc des enthalpies libres de réactions à dilution infinie. Ould-Moulaye (1998) a vérifié que ces dernières propriétés de réaction correspondent à des différences de propriétés de formation.

Modélisation des propriétés de formation:

La base de la représentation des propriétés thermodynamiques des solutions repose sur le choix d'un état de référence (ER) pour chacun des constituants de la solution. En ce qui concerne la température et la pression de référence, il n'y a pas de difficultés particulières (en général nous prendrons $T_r = 25\,°C = 298{,}15\,K$, et $p_r = 1\,atm = 1{,}01325.10^5\,Pa$), mais il en est autrement pour le choix de l'état physique de référence du composé (corps pur solide, liquide ou gazeux, état de référence à dilution infinie dans l'eau pour les ions). Rappelons que les propriétés de formation d'une molécule sont définies comme étant les propriétés de la réaction de formation de la molécule

à partir des éléments de référence (pris dans leurs états de référence respectifs) qui la constituent. Classiquement en thermodynamique, les molécules de type $C_xH_yO_zN_t$ sont traitées en prenant comme éléments de référence le carbone (C solide, sous forme graphite), l'hydrogène (H_2 gaz), l'oxygène (O_2 gaz) et l'azote (N_2 gaz).

A l'état gaz parfait, il existe différentes méthodes pour déterminer les propriétés de formation (en particulier l'enthalpie) à partir des données d'énergie issues de calculs quantiques (Irikura et al. 1998). Lorsqu'on connait les propriétés de formation à l'état gaz parfait des molécules d'intérêt, il est impératif de déterminer les même propriétés à dilution infinie dans l'eau (car c'est l'état de référence choisi pour les sucres et les ions,…) afin de traiter les réactions chimiques en phase aqueuse, dès que l'on dispose d'assez d'informations sur le cycle suivant:

$$\begin{array}{lcc} & \Delta G_r^{gaz} & \\ \text{Dans l'état gaz:} & \text{reactifs} \rightleftarrows & \text{produits} \\ & \Delta G_{\text{solv}}^{\text{react}} \downarrow & \downarrow \Delta G_{\text{solv}}^{\text{prod}} \\ \text{Dans l'eau:} & \text{reactifs} \rightleftarrows & \text{produits} \\ & \Delta G_r^{\infty,eau} & \end{array}$$

Pour ce faire, nous avons développé une nouvelle approche (dont la procédure « générale » est détaillée dans le chapitre IV) pour déterminer les propriétés de formation à l'état gaz parfait, et en déduire (en particulier en utilisant les valeurs d'enthalpies libres de formation) la valeur de $\Delta G_r^{\text{gaz}} = \Delta G_f^{\text{produits}} - \Delta G_f^{\text{reactants}}$. La méthode COSMO-RS, étant un modèle de solvatation développé pour les solutions aqueuses, joue un rôle critique dans l'étape de transition entre les phases (gaz et solution aqueuse). Nous l'avons utilisé avec succès pour déterminer les valeurs des enthalpies libres de solvatation des réactifs ($\Delta G_{\text{solv}}^{\text{react}}$) et des produits ($\Delta G_{\text{solv}}^{\text{prod}}$), et ainsi déduire la valeur de l'enthalpie libre de réaction dans l'eau (sachant que $\Delta G_r^{\infty,eau} = \Delta G_r^{gaz} + \Delta G_{\text{solv}}^{\text{prod}} - \Delta G_{\text{solv}}^{\text{react}}$).

Cependant, le chemin de la solvatation ne décrit que partiellement le processus de passage de l'état gaz à la phase condensée. Il nous a semblé nécessaire de revisiter ce chemin pour développer une vision complémentaire et cohérente au plan physique d'un chemin de solvatation qui passe par l'étape virtuelle des éléments de surface chargée. Ceci est présenté dans le chapitre V. Le but est de montrer sans identification supplémentaire qu'il est possible de prédire les enthalpies et enthalpies libres de vaporisation d'une grande variété de corps (70 molécules neutres environ) sur une très large gamme de température. Pour démontrer la robustesse et la consistance de ce nouveau chemin thermodynamique, nous avons été amenés à traiter des problèmes à l'interface entre la

thermodynamique classique, la mécanique quantique, l'électrostatique et la physique statistique. Cela a permis entre autres de démontrer dans l'ensemble grand canonique les relations fondamentales aboutissant au calcul du potentiel chimique avec la méthode COSMO-RS en faisant apparaître des axes de généralisation de ces équations, de mettre en évidence des étapes importantes lors du passage de l'état COSMO à l'état gaz voire à l'état liquide, et d'intégrer la prise en compte des mouvements de translation dans le traitement des propriétés thermodynamiques des corps purs (en particulier à leurs températures normales d'ébullition).

Quelques applications spécifiques des résultats de la thèse dans le projet Na⁻

Pour décrire la transformation d'un produit agricole brut en un aliment plus ou moins complexe, il est nécessaire de décrire en fonction du temps et de l'espace:

- les transferts de chaleur et de matière (eau, sels, …) qui se produisent entre le produit et son environnement ou à l'intérieur du produit. Ces transferts se traduisent par l'apparition de gradients de température, de teneur en eau et en solutés, donc par des modifications de la composition chimique en chaque point du produit. Ils sont caractérisés par des coefficients d'échange qui dépendent des conditions environnementales (température, humidité relative, …) et par des coefficients de diffusion qui dépendent des conditions physico-chimiques (température, activité de l'eau, pH, …).
- les cinétiques chimiques, biochimiques ou microbiologiques qui ont lieu sur la surface ou dans le produit. Ces cinétiques induisent aussi des modifications locales de la composition chimique ;
- les modifications locales de composition chimique qui sont induites. Celles-ci se traduisent par des changements des potentiels chimiques des différentes espèces chimiques présentes;
- l'évolution des propriétés mécaniques, nutritionnelles, hygiéniques et sensorielles du produit en fonction des transferts et des cinétiques. Ces propriétés sont le résultat des évolutions décrites précédemment.

L'objectif du projet « Na⁻», intitulé « Amélioration raisonnée des procédés de fabrication de produits carnés salés crus ou cuits en lien avec la réduction de la teneur en sel et en sodium», et financé par l'Agence Nationale de la Recherche (ANR), est de développer des connaissances sur les phénomènes physiques (transferts de matière, de chaleur, etc.) et biologiques (évolution enzymologique, microbiologique, etc.) qui ont lieu dans la viande de porc lors du procédé de fabrication de jambons crus et cuits, ceci afin de développer des outils de modélisation pour assister les professionnels dans leur démarche de réduction du taux de sel et en particulier du taux de sodium. Cette conception raisonnée vise à proposer des produits « moins salés » tout en maintenant leurs qualités sanitaires et organoleptiques et en assurant une meilleure homogénéité de la production. Cette démarche devrait déboucher sur des procédés innovants et/ou des produits nouveaux qui devraient être aussi moins énergivores et plus respectueux de l'environnement.

Dans le cadre du projet Na-, ce travail de thèse a permis une étude qualitative de l'effet du salage sur les propriétés physico-chimiques de la viande. A cet effet, la composition chimique du muscle a été approximée par sa composition moyenne en eau, en protéines (donc en acides aminés) et en lipides (donc en acides gras). Ensuite, certaines données d'entrée du modèle COSMO-RS, en particulier la distribution surfacique des charges (σ-profile) ont été utilisées pour simuler (d'un point de vue qualitatif) les phénomènes mis en jeu lors du procédé. Il apparaît ainsi que le σ-profile (qui est une propriété additive) permet de comprendre l'effet de l'ajout de sel sur les courbes de sorption et de titration des viandes de bœuf et de porc, et d'explorer des substituants pouvant remplacer le sel de cuisine (NaCl). Nos travaux ont ainsi montré que l'on peut utiliser la simulation moléculaire (COSMO-RS) pour simuler un échantillon de muscle (SM) et déterminer les propriétés physico-chimiques. Il apparaît ainsi que la courbe de titration de la viande de porc est identique à celle de la viande de bœuf, que la courbe de titration de la viande ne varie pas après l'ajout de sel alors que la courbe de sorption varie. Ces différents phénomènes ont été validés expérimentalement, ce qui confirme la performance prédictive du modèle. De plus, les simulations effectuées suggèrent que lors du processus de salage il serait plus judicieux d'injecter des saumures (en particulier des sels dissous) pour améliorer les transferts de sel et d'eau dans le muscle. Ce mode de salage est déjà adopté dans l'industrie de la viande cuite, et la question demeure pertinente pour la fabrication du jambon sec.

Il apparait ainsi que, contrairement aux idées reçues, des outils existent maintenant pour prédire les propriétés physico-chimiques de produits biologiques et alimentaires contenant de nombreuses molécules d'intérêt (sels, acides organiques, acides aminés, peptides, protéines, sucres, …) avec une bonne précision. Combinés avec des modèles de transfert de chaleur et de matière, ainsi qu'avec des modèles d'évolution de la qualité, ils constituent des simulateurs de procédés alimentaires performants. En effet, le modèle thermodynamique permet de prédire l'évolution spatio-temporelle des propriétés physico-chimiques (pH, a_W, pOR) qui dépendent de la composition du milieu. Ces propriétés sont nécessaires pour comprendre et prédire l'évolution de la population bactérienne dans une approche globale de prédiction de la qualité et de la sécurité microbiologique d'un produit alimentaire (tel qu'un jambon sec ou cuit). Ce travail démontre également qu'il est possible de traiter l'effet des substituts de sels dans une démarche de prédiction du pH et de l'a_W. Cela permet de déterminer les qualités d'un nouveau produit sans recourir à l'expérimentation lors de la phase d'optimisation du procédé à cause de la capacité prédictive du modèle. A cet effet, l'utilisateur final doit juste indiquer la liste des additifs (ainsi que leurs compositions respectives) ajoutés lors du processus de fabrication du jambon comme données d'entrée du modèle.

Table of Contents

Table of illustrations

(figures and tables)

List of Figures :

List of Tables :

Introduction

The knowledge of physico-chemical properties of fluids is essential in chemical engineering, especially for the design of many kinds of industrial equipments. These properties are generally determined through experimental measurements. However, measurements are sometimes expensive and time consuming and several measured data are often inconsistent with each other ones. Furthermore, measurements just yield numbers, while theoretical calculations ideally give physical insight providing thus answers to questions like why a parameter has a certain value, and which chemical variation of the compound may lead to a certain desired change of a property of the system. Moreover, in current research practices, it is very common to encounter situations where the properties of a new compound are needed even if this compound is not yet synthesized.

In this context, the conception of new products by new processes generally requires the use of numerical simulation. The latter is a method enabling the analysis of a process, a phenomenon, or the behaviour of a system thanks to a model. If the modelling consists in the establishment of a model, simulation uses the model that has a similar behaviour as the real system, in order to study several scenarios which are synthesized in order to take a decision. During the manufacturing of a new product, the simulation should be considered as an experimentation, the term "computer experimentations" is often used, *i.e.* a tool of exploration of the possible cases. The use of computers enables to increase the number of studies, to expand the studied cases and usually to decompose the factors that have strong interactions between them.

The process simulation tools include thermodynamic models, which are combined with properties databanks of pure components in order to estimate or predict the properties of mixtures. These tools are mainly developed for the treatment of chemical and petrochemical processes. In comparison to petrochemistry solutions which are mainly non-electrolytes solutions, foods and biochemical solutions are generally treated as aqueous mixtures that can be very complex. Indeed, the latter contain mainly water and a wide diversity of species and the presence of several phases of different natures (liquid, gaseous and solid). One of the research subject of the "Génie des Procédés Énergétique et Biosystèmes" (GePEB) team of the Institut Pascal laboratory (UMR CNRS 6602 Université Blaise Pascal) is to extend the application areas of these software up to the treatment of biological and food systems. Modelling such processes must therefore take into account this complexity of these environments. This requires handling simultaneously, chemical equilibria (*e.g.* dissociation, complexation and redox) and physical equilibria (*e.g.* liquid-liquid, vapour-liquid, solid-liquid, vapour-solid and solubility). To do so, it is necessary to select and/or develop predictive models that are thermodynamically consistent to treat the above mentioned equilibria. This requires a generalization of existing models of aqueous systems. Ideally the selected model should use a molecular approach in order to handle more easily the diversity of components in the studied systems. Ould Moulaye (1998) has verified that reliable free enthalpies, enthalpies and heat

capacities data have to be determined in order to characterize these equilibria. In particular, the chemical potential or molar free enthalpy of formation is essential for modelling equilibrium data. The main thermodynamic modelling task is therefore the reliable and consistent calculation of the chemical potential of the individual components from the total free enthalpy G. In order to predict a property such as the chemical potential, it is necessary to generalize the existing models up to the prediction of equilibrium properties in a chosen reference state (generally the infinite dilution in water), and the prediction of the non-ideality (*i.e.* the deviation from Raoult's law) for concentrated solutions. This splits the main thermodynamic modelling tasks to two main problems, which are: *i-)* the prediction (and/or the collection) of formation properties data and, *ii-)* the development of a predictive model of activity coefficients. Because the existing tools are developed generally for the treatment of the most common neutral compounds in multicomponent environments (including water), our main concerns are to extend their performances to the treatment of food and biological systems which generally also contain salts (*i.e.* electrolytes) and sugars.

Among the previous works achieved on these topics in the GePEB axis of Institut Pascal, the work of Achard (1992) has enabled to select a predictive model that can determine the equilibrium properties from the knowledge of the excess properties of the different components, thanks to the activity coefficient. The resulting model, denoted as ULPDHS, combines a chemical treatment of ions solvation with two activity coefficient models namely one based on the Debye-Hückel theory (Pitzer 1980, Pitzer 1973) which treats long range physical interactions and the UNIFAC-Larsen model (Larsen et al. 1987) which takes into accounts short range physical interactions. The ULPDHS model assumes a constant hydration of ionic species. In this context, ions were considered as UNIFAC independent groups and the solvation of charged species giving clusters were taken into account by means of a hydration number for each ion at infinite dilution. With this approach two interaction parameters (water-anion and water-cation interactions) were sufficient to characterize a water-salt system (Achard et al. 1994). The resulting model (ULPDHS) has been adopted by ProSim S.A. and is available in the software 'Simulis Thermodynamics'. It is applicable for multi-electrolytes solutions without adding any new interaction coefficient and it satisfactorily predicts water activity, osmotic coefficients and salting-out effects in aqueous mixtures of two or three electrolytes within less than 5% even for saturated solutions. The performance of the ULPDHS model was tested on aqueous mixtures containing up to 6 moles.kg^{-1} of electrolytes. The predicted results were very satisfying. The work of Catté (1994) has permitted to validate the performances of the ULPDHS model for the treatment of systems containing sugars and dissolved gases even at high pressures.

In order to handle the presence of components that can form complexes in solution or those involved in redox reactions, Ould Moulaye (1998) has provided the thermodynamic relations that

are necessary to extend the solution models like ULPDHS to the treatment of such systems. Furthermore, this previous work has permitted to estimate the thermodynamic properties of several components of the metabolism cell (amino acids, carboxylic acids, nucleotides, etc.). This task was done in a unique reference state in order to conciliate dissociation constants (pK_a) values, complexation constant (pK_c) values, and concentration ratios measured *in vitro* for enzymatic reactions as well as the corresponding Gibbs energy of reactions.

The work of Ben Gaïda (2007) has enabled to extend the performance of the ULPDHS to the accurate prediction of thermodynamic properties in concentrated solutions through the handling of a variable solvation of ionic species. The resulting model (denoted as ULPDHSV) predicts accurately activity coefficients in very concentrated solutions (molalities up to 20 moles.kg^{-1} of salts). The predictive behaviour of the ULPDHSV model was also evidenced by the accuracy of its water activity results in aqueous systems containing only a mixture of electrolytes as well as in those containing a mixture of electrolytes and neutral compounds. In particular, the performance of this model was validated on several food systems (apple juice and honey).

The results of these previous works done by the IP-GePEB team have contributed to the progress in the modelling of food and biological systems. However, the identification thus the collection and the conciliation of data remain to be continued. For instance, for several components like amino acids containing sulphur, phosphate glucose, or the ascorbic acid, the UNIFAC group interactions parameters are not yet available in the literature. However these parameters are required in the ULPDHS and ULPDHSV models in order to treat more complex systems containing for instance amino acids, fatty acids or macromolecules (proteins, polysaccharides, etc.) which are always present in food systems like meats and milk. The objective here is to update the previous approaches with recent thermodynamic models of solution. This is the reason of the participation to the "Na$^-$" project.

The "Na$^-$"project is entitled "Reasoned improvement of raw and cooked cured meats production in relation with the reduction of the salt and sodium content". The participation of IP-GePEB team provides the financial support of the present PhD study. As mentioned in its title, the main objective of this project is to reduce the salt content in meat products and especially in processed ham products. The industrial practices for reducing salt content in these products mostly remain very empirical, with tests realized in industrial situations. The ambition of the project is to give to meat industry an effective simulation tool permitting to reduce the salt content in the processed products as well as to ensure a better homogeneity of the production and a cost allowing an increased competitiveness for the companies. Finally, the products or the processes stemming from this project work should be less energy consuming and more environmental friendly. To achieve this ambitious goal, the project is voluntarily located in the field of optimization and design

of products and processes. It requires the use of numeric simulation tools, especially those that can provide a comprehensive description of the following points versus time and space:

- the heat and mass transfer that occur between the product and its environment or inside the product;
- the chemical, biochemical or microbiological transformations that occur at the surface or in the product;
- The local modifications of chemical composition that result in changes of chemical potentials of the different present chemical species, therefore of the activities (to the thermodynamic sense) of these, such as pH, a_w or redox potential;
- The evolution of the mechanical properties linked to the structural, nutritional, microbial and sensorial properties of the product according to the transfers and the kinetics. These properties are the results of the evolutions described previously.

In a general way, the simulation tools will permit the description of the spatial and time evolution of physico-chemical properties (temperature, water content, activities of the media compounds) in a given product. The knowledge of these properties is mandatory to describe and understand the (bio)chemical or microbiological kinetics occurring in a food or at its surface.

The "Na-" project gathers various scientific fields such as chemical engineering, applied thermodynamics, biochemistry, and microscopy to get a finest description of the phenomena occurring in ham production. This combination, with coordination of the multidisciplinarity of the partners constitutes the originality of the project.

In this context, this PhD work is focused on the study of equilibrium thermodynamic properties of food and biological aqueous media. It aims at extending the performance of the existing thermodynamic models up to the prediction of the physico-chemical properties of such complex systems (containing simultaneously electrolytes and non-electrolytes, solvated and un-solvated substances, etc.). Thus, this work deals with the same problematic as the above mentioned previous works of the IP-GePEB team.

In Chapter I, we will first present in more details the interest of physico-chemical properties in food and biological systems with a voluntary focus on meat products to fit the needs of the Na- project. Then, the basic thermodynamic knowledge required for understanding the modelling of physico-chemical properties is shortly summarized. An overview of the existing tools is performed before selecting the most suitable model for this work.

For the treatment of short range effects, our choice leads to the COSMO-RS method (Klamt 1995, 1998, 2005 and 2011) that is an *a priori* fully predictive method widely used in chemical engineering. COSMO-RS is also an excess Gibbs energy method like UNIFAC, but with the advantage that it does not require identification of interactions parameters between chemical groups. This model appears as a promising alternative predictive tool, especially when the UNIFAC

model cannot be used because of the unavailability of the required interaction parameters. Chapter II provides a detailed description of the COSMO-RS method, which combines concepts of electrostatics, quantum chemistry and statistical thermodynamics to predict the chemical potential of liquid systems.

COSMO-RS was developed to treat aqueous mixtures containing mainly neutral compounds. When electrolytes are considered, the system is characterized by the presence of both molecular species and ionic species, resulting in three different types of interactions: ion-ion, molecule-molecule and ion-molecule. Ion-ion interactions are governed by electrostatic forces between ions that have a much longer range than other intermolecular forces. The excess Gibbs energy of systems containing electrolytes thus appears as the sum of two terms, one related to long-range (LR) forces between ions and the other to short-range (SR) forces between all the species. In chapter III, we introduce a new tool able to predict activity coefficients of electrolytes in binary aqueous electrolyte systems (water-salt). This tool is based on an extension of the COSMO-RS method towards the representation of thermodynamic equilibrium properties of charged species. The long range interactions between ions are taken in accounts by the Pitzer term based on the Debye-Hückel theory. The prediction of the hydration number that is the key parameter for taking into account the solvation of the ion by water molecules is investigated in this chapter, where a new set of COSMO-radii is also proposed for all studied ionic elements. The resulting model called "COSMO-RS-PDHS" is fully predictive with no identified parameter.

As mentioned earlier, food and biological systems also contain a large number of species that are involved in chemical equilibria (*e.g.* dissociation, complexation, solvation and redox). To handle the prediction of these properties, it is necessary to know the chemical potential of formation at infinite dilution in water. The evaluation of the performance of the COSMO-RS method in predicting the pK_a of amino-acids and peptides is discussed in chapter IV. COSMO-RS uses quantum calculated "free" energies to predict pK_a using a linear free energy relationship (LFER). A new method of prediction of formation properties of several molecules of interest using the same input data as those of the COSMO-RS model (*i.e.* results of quantum DFT/COSMO calculations), is also introduced in Chapter IV. This kind of approach is fully predictive and allows treatment of conformations. This is its main advantage compared to classical group contribution methods. However, the biological solutions studied within this work involve an important variety of species. Thus, it is necessary to convert gas phase data to their corresponding values at infinite dilution in water. This transition is done by using several facilities included in the COSMO-RS algorithm like the calculation tool of solvation free energies data. In chapter IV, a comparison between the formation properties results determined from the use of the solvation data predicted by the COSMO-RS model and those obtained using the results of the thermodynamic relations of phase

equilibria is performed. By combining these data with the gas phase data and the solvation data, it becomes possible to predict the chemical potential values obtained at infinite dilution in water for a large set of neutral compounds including free radicals. The latter chemical potential values can be used to predict the Pourbaix diagram of several compounds (including both pK_a and standard redox potentials data). Furthermore, since the approximate composition of a meat muscle is known, a section is devoted to the evaluation of the performance of the COSMO-RS model as a qualitative understanding tool for phenomena appearing in complex mixtures, especially in the context of the Na^- project.

The previous results have provided a thorough experience of linking consistently gas phase and condensed phase properties. It appears that the solvation concept used in the COSMO-RS algorithm, although very efficient, is limited to the treatment of neutral compounds and allows only a partial description of the process of transition between gas and condensed phases. When we are dealing with charged species (that are ubiquitous in food and biological systems), the solvation Gibbs free energies provided by COSMO-RS are not accurate enough. Thus we decided to revisit the roadmap of the whole study and to propose new developments on a consistent view of a solvation process that passes through the virtual step of charged surfaces molecules in both phases. This is the purpose of Chapter V where we have to handle simultaneously concepts of classical thermodynamics, quantum physics, electrostatics and statistical physics. The goal is to determine consistently step-by-step, from the gas phase to the COSMO state and finally from the COSMO state to the condensed phase, energies and chemical potentials of the different species. Basically, it is shown that energies are given by quantum mechanics; chemical potentials (partial entropies) are obtained by statistical thermodynamics. An important result is obtained for considering the virtual transformation to COSMO state with a special treatment of the cavity of the conductor. The reasoning is validated by the predictive determination of vapour pressures and enthalpies of vaporization using COSMO-RS parameters without any new parameter regression. This opens the door to a complete revisit of a thermodynamically consistent treatment of thermo-chemical reaction data and activity data, including solvation, dissociation and complexation.

A general conclusion ends the manuscript summarizing the different results and putting the different highlights in a global perspective.

Chapter I:Scope and roadmap of the study

I.1. Equilibrium properties in food and biological processes

Processed foods (like cooked or dry meats, milk, honey, or juices), medicines, and biochemical solutions are all a myriad of interacting dissolved solute molecules. Water is generally the main component in most of these systems and equilibrium properties of water are considered as reference properties (*e.g.* water activity) in food processing (Gros and Dussap 2003). Thus, foods and biological solutions are generally treated as aqueous mixtures that can be very complex as illustrated in Figure I-1. Indeed, these media can contain a wide variety of components including organic compounds (like sugars, lipids and organic acids), salts or electrolytes (*i.e.* aqueous ions or charged species), dissolved gases and macromolecules (proteins, polysaccharides and polymers), undissolved particles which can be solid or semi-solid. Thus, due to the presence of different phases (liquid, solid and gaseous) the modelling of such systems requires a homogeneous treatment of both physical equilibria and chemical equilibria (dissociation, hydration, complexation, and redox) as mentioned by Achard (1992), Ould-Moulaye (1998) and Ben Gaïda (2007). All of these equilibria are characterized by specific measurable physico-chemical properties that can be used in process engineering to control or regulate the process.

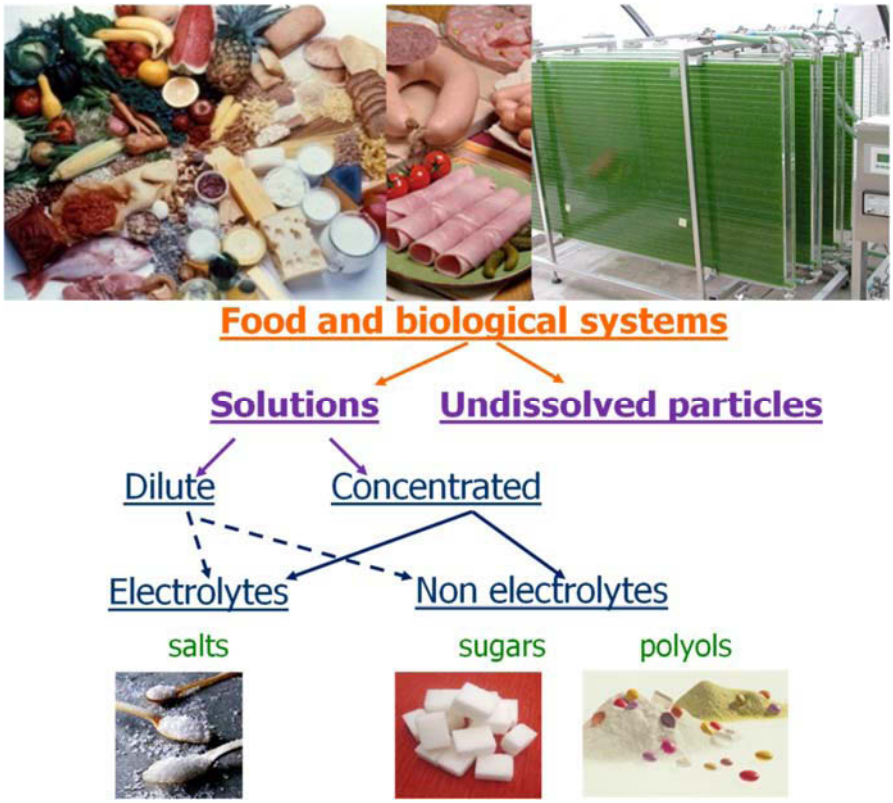


***Figure I-1**: Illustration of the thermodynamics modelling requirements for foods and biological systems.*

During a food process, environmental conditions may vary as a function of time or space in the food: for example, in many preservation techniques, temperature varies (cooling, cooking, pasteurization), pH goes down (acidification due to bacterial metabolism, fermentation), water activity (a_w) decreases (drying process, addition of solutes) or redox potential (E_h) falls as food is reduced. As a consequence, phenomena of diffusion (water or solutes transfer) as well as heat transfer occur in the product, indicating that in a food, gradients of temperature, pH, a_w and E_h appear. Consequently, gradients of oxidation or of bacterial population in the product thickness can be observed because many biological processes depend on the physico-chemical properties.

A combined modelling approach (Figure I-2) can be used to predict the different qualities of a food medium during a process. In general, most variables can be measured and, in the best cases, models exist to predict them.

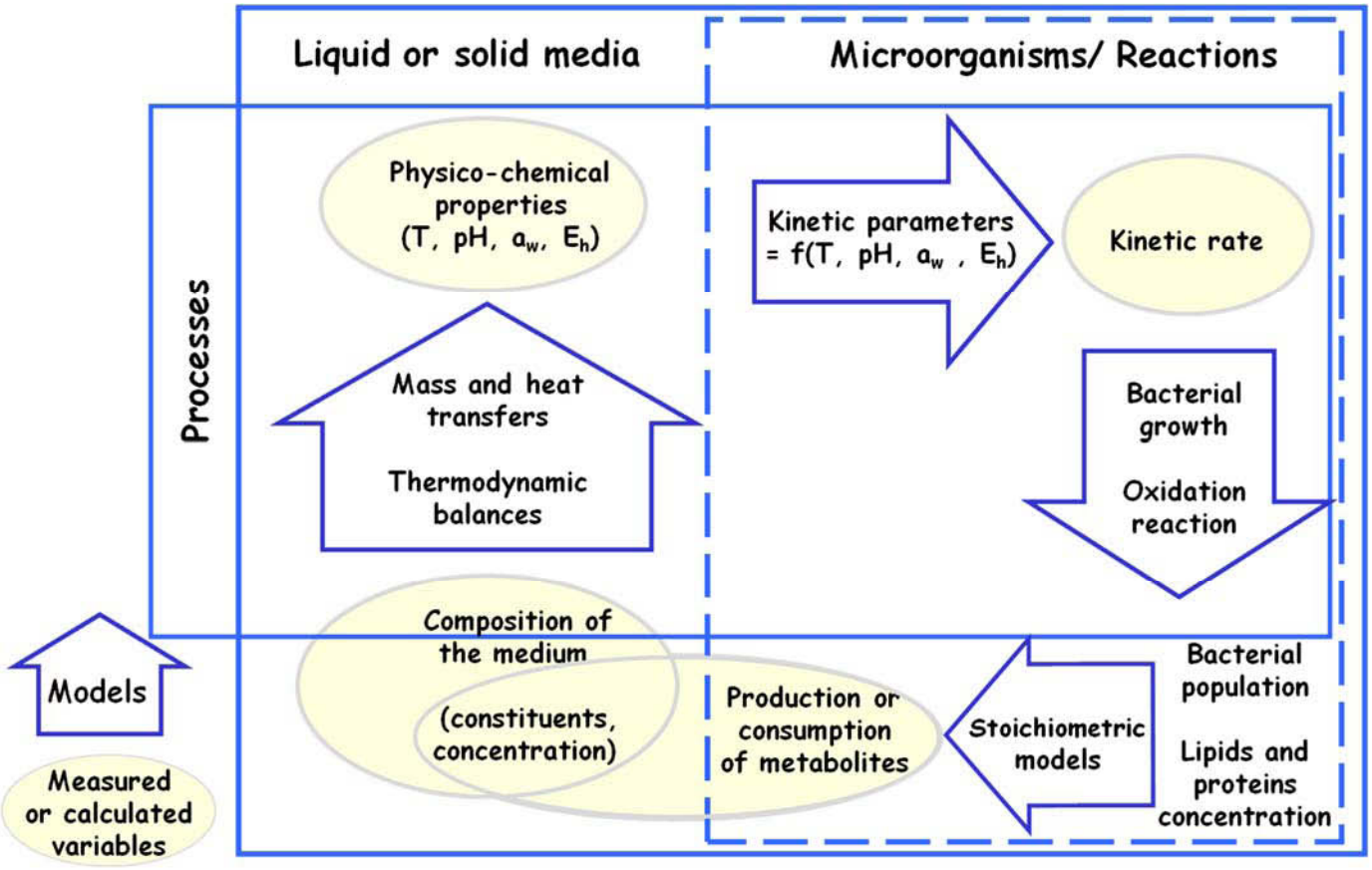


Figure I-2: Combined modelling approach (Adapted from Desnier-Lebert 2004).
This figure points out the complex interactions between different scientific areas, even if there are existing modelling tools at each step.

However, these models have been developed separately in order to take into account only some specific aspects of the global problem. For this purpose, several assumptions that limit the usage of the specific model are made during the development of a given model. For example, many predictive bacterial models exist that predict growth parameters or bacterial kinetics. Concerning the mass transfer (water or solutes) in the product, Fick's law is mainly used to predict the diffusion

of water and solutes inside the product. But using such a model with a constant diffusivity may lead to erroneous predictions of the properties in the product: for example the water diffusivity is varying with the water content in the product. Indeed, a variable diffusivity has to be taken into account in the Fick equation to correctly predict a_w in the product. In the case of heat transfer, the Fourier equation is a convenient tool when a variable heat diffusivity coefficient is taken into account.

In this global methodology, the water or solute concentration is calculated by the water transfer model and is used to calculate the local physico-chemical properties (pH, a_w, E_h) at the surface or inside the product using the thermodynamic model. The kinetic parameters then are estimated and the evolution of the technological and bacterial properties can be made. This global approach had already proved its efficiency in the prediction of the growth of *Listeria innocua* at the surface of a gelatine (solid) medium subjected to air drying (Desnier-Lebert 2004). A similar approach was adopted in the Na^- project which aims at reducing the salt content in meat products and especially in ham products. In this context, it is important for the process engineer to have enough means of action (such as the physico-chemical properties) in order to have a great control of the process and limit defects in processed products.

I.2. Interest of physico-chemical properties in meat products

Because this work was done in the framework of the "Na^-" project, *i.e.* the reduction of salt ratio in meat products, the illustrative examples will be focused on this kind of applications, even if there are numerous other applications of the physico-chemical properties in food science, as well as in processes encountered in many other scientific areas (distillation, extraction, and other phase-contacting operations, etc.).

The organoleptic qualities combine all of the sensory properties to the origin of sensations of pleasure associated with the consumption of meat and meat products: colour, flavour, juiciness, tenderness. Technological qualities reflect the abilities of a product to be processed. In meat and meat products, they are often associated with the concept of water holding capacity. Meats with a poor technological quality have low water retention and make exudations, water loss during cooking and therefore lower technological yields. The quality of cutting is also appreciated as a technological quality of interest.

Technological quality of cooked ham is usually defined as its ability to keep or bind water during processing. In the industry, brine is injected in the muscles. Sodium chloride, used in cured processing, is an active water binding compound which retains water, improving cooking yields and enhancing juiciness and textural palatability of meat products (Offer and Knight 1988). Curing affect the degree of swelling of the myofibrillar system (Offer and Trinick 1983) and allows more water to be bound within the tissue. The induced-curing increases the ionic strength of muscle tissue, and leads to the denaturation and solubilisation of proteins, changing their chemical composition and consequently the cellular structure. These chemical changes depend on animal species, myofibre characteristics and at a smaller scale, tissue composition and cell environment (connective tissue, membranes, fat, muscle contractile proteins, electrolytes ...) (Parsons and Knight 1990). Curing also brings about gaps and merge of muscle cells which suggests membrane and extracellular matrix solubilisations (Astruc et al. 2008). However, the way of salting dry ham is different than that for cooked ham.

The meat intended for dry products is usually salted with dry salt before drying, inhibiting microorganism grows and leading to high ionic strength and strong micro-structure changes (Larrea et al. 2007). Enzymatic degradation of muscle proteins, influenced by processing (salt, temperature, drying time; Buscailhon et al. 1994), participates in the development of the texture and the flavour of dry cured ham (Toldra 1998). Reducing the salt content in the dry ham promotes this proteolysis, increasing the risk to obtain unpleasant flavour, pasty texture (Tabilo et al. 1999) and therefore, affects the acceptability and slicing qualities. For both cooked and dry cured ham, the characteristics of raw meat used for processing are of great importance, as discussed below.

I.2.1. Temperature, pH/pK_a and water activity a_w in meat processing

Heating causes shrinkage of the endomysium, which led to the expulsion of intracellular water in the extracellular space, which is finally evacuated in the way of cooking losses. The addition of sodium chloride improves the water holding capacity by inducing the swelling of filament lattice, increasing juiciness and tenderness, reducing the cooking losses, and therefore the removal of micronutrients.

The pK_a of a compound is an important property in both life sciences and chemistry since the propensity of a compound to donate or accept a proton is fundamental to understanding chemical and biological processes. It is also closely related to the concepts of *pH* (acidity of solution) and *log(P)* (the partition coefficient between immiscible liquids) (Perrin et al. 1981). The pK_a value of a molecule also determines the amount of protonated and deprotonated species at a specific *pH*. More often, phospholipid membranes easily absorb neutral molecules, while ionized

molecules tend to remain in the plasma or the gut before being excreted. Many biological systems also use proton-transfer reactions to communicate between the intra- and extracellular media, and the rate of the proton-transfer reaction depends, in-part, on the pK_a values of the species involved. While experimental methods continue to become more sophisticated and refined, it is often desirable to predict dissociation constants for "virtual compounds," *i.e.*, those that have been described by a compound designer (chemist or modeller) but that have not yet been synthesized. In another area, microorganisms are inhibited by the non-dissociated forms of weak organic acids. The knowledge of pK_a values is then of great importance in microbiology previsional models (Lebert et al. 2005).

The speed of pH drop and the ultimate pH affect the organoleptic (colour, texture, juiciness, flavour) and technological (water holding capacity) qualities of processed foods. Moreover the muscle microstructure, connective tissue and lipids distribution influence the diffusion of sodium chloride. The texture and the water holding capacity are closely related to the structure of muscle tissue. The physical constraints of processes lead to morphological and chemical changes in meat products that have positive or negative impact on their quality. For instance, the acetic acid CH_3COOH which is added to improve the organoleptic and hygienic qualities of a food product does not stay in its neutral form and dissociates to form CH_3COO^- and H^+. The proportion of each form depends on the value of the dissociation constant (pK_a) of the acid. Many studies have demonstrated that the non-dissociated form of the acid has an effect on the inhibition of the bacterial growth (Brocklehurst et al. 1993). Rosso and his coworkers have demonstrated that the minimum pH of growth is linked to the pK_a of the weak acids (Rosso et al. 1997). Their study was focused on mono-acids only. In the case of acids that have two pK_a values, like citric acid, what are the respective effects of each form of the acid (AH, A^- and A^{2-}) on a micro-organism? Does the neutral form have always the most inhibition strength? How do the two other species behave in this mixture? In practice, in these disciplines, the study is focused only on the global composition of the solution. In order to understand the effect of each species, one has to analyze the detailed composition of the mixture and then study the chemical equilibria between the species in presence.

In microbiology it is widely accepted that the effect of water on microorganisms is accurately taken into account by the use of the water activity, which represents the availability of water, instead of the water concentration inside the products (Van den Berg and Bruin 1981; Simatos 2002). During a drying process, a_w controls the flux of water evaporation, which gives a new concentration of water. At the surface of the product, the a_w prediction requires a great deal of exchange between different phenomenological models, since there is no direct measurement technique. However, a specific device was recently designed to estimate indirectly the time-course variation of a_w at the surface of a cheese or a fermented sausage subjected to a low-velocity airflow

(Le Page et al. 2010, 2012), but it is not a direct measurement and this device cannot be used in all cases of drying processes. It is known that bacteria are sensitive to a_w in the high humidity range above 85%. When the surface of a product is exposed to air flow conditions, the surface characteristics of the food, *i.e.*, the water content and the surface a_w, are modified. Even if the average water content of a food sample is easily measurable, it does not reflect the water content (especially the a_w) at the surface of the products.

In summary, to describe the evolution of a food during processes, it appears important to measure or predict the evolution of the physico-chemical properties such as the temperature T, the dissociation constant pK_a, the pH and the water activity a_w. These four parameters are currently the means of action to limit defects in processed meat. However, these 'traditional' physico-chemical parameters, that can be measured accurately and are widely used by food scientists in the design, the optimization and the control of their processes, are not sufficient and it appears essential to act on a fifth parameter: the redox potential (E_h).

I.2.2. Redox potential E_h in foods science and biological systems

The redox potential (also known as oxydo-reduction potential, oxidation/reduction potential denoted ORP, pE or E_h) is a measurement of the tendency of a chemical species to acquire electrons and thereby be reduced; it is expressed in volts. Each species has its own intrinsic redox potential; the more positive the potential, the greater the species' affinity for electrons and tendency to be reduced. E_h is a common measurement for water quality. Indeed, E_h has an important role in life sciences by acting on many enzymatic reactions and on microorganisms. The measure and the control of this parameter in the processed meat products manufacturing could better manage the growth of technological flora and thus reduce the sensory and microbiological alteration of products. The muscle has a very high energy metabolism (glycolysis, oxidative phosphorylation, transport of oxygen by heme). Thus meat is a matrix rich in oxygen and reactive nitrogen oxygen species. The redox balance of a food matrix is reflected in the evolution of its E_h. Its measurement and its control in the processed meat products manufacturing could better manage redox state of meat matrix. With the traditional physico-chemical parameters measured, E_h would allow to understand and control the product evolution in terms of colour and microbiological alteration and act on the conservation of product.

Thus, many enzymatic reactions are oxidation-reduction reactions in which one compound is oxidized and another compound is reduced: the ability of an organism to carry out oxidation-reduction reactions depends on the oxidation-reduction state of the environment, or its reduction potential. Strictly aerobic microorganisms are generally active at positive values, whereas strict

anaerobes are generally active at negative values. Redox affects the solubility of nutrients, especially metal ions. There are organisms that can adjust their metabolism to their environment, such as facultative anaerobes. Facultative anaerobes can be active at positive E_h values, and at negative E_h values in the presence of oxygen bearing inorganic compounds, such as nitrates and sulphates.

A brief state of art focused on the role of E_h in the study of microorganisms and oxidation phenomena in meat products, as well as the measurement of E_h, is given below.

I.2.2.1. E_h and microorganisms

Bacterial starter cultures are widely used to drive the fermentation, reduce the variability in the quality, limit the growth of spoilage and pathogenic bacteria, and improve the sensorial qualities of fermented sausages. Bacterial starters are often made up of a balance between Lactic Acid Bacteria (LAB), mainly *Lactobacillus sakei*, and coagulase negative staphylococci (CNS), mainly *Staphylococcus xylosus*. LAB degrades carbohydrates mainly into acids involved in the taste. Their acidification and capacity to produce bacteriocin also contribute to the inhibition of spoilage and pathogenic bacteria. The CNS contributes *via* the nitrate reductase activity to the colour and *via* their antioxidant capacities and amino acid catabolism to the flavour (Talon and Leroy 2011). The physiology of these starters is greatly dependant of various environmental factors such as the well studied pH, temperature, water activity but also the less characterized redox potential. E_h depends on the composition of the food, the pH, the temperature and the concentration of dissolved oxygen (Martin et al. 2012). This parameter plays an important role in the physiology of microorganisms: growth, enzyme expression, and thermal resistance (Green and Paget 2004). The effect of E_h has been particularly characterized on a fermented dairy product, yogurt (Martin et al. 2009). It has been shown that E_h affects the texture: whey separation, production of exopolysaccharide, viscosity and the aroma: acetaldehyde, diacetyl, sulphur compounds (Martin et al. 2012).

E_h in food can be modulated by using gas which enables the products characteristics to be maintained. As examples, the decrease of E_h by gas improved microbial quality and quality of dairy products (Martin et al. 2009, 2010). Gas applications in the food industry are numerous, in particular the modified atmosphere packaging (MAP) using different gases: oxygen, nitrogen, carbon dioxide. There are no safety issues for the product and their use is authorized at European level.

Even if the literature has underlined the role of E_h in the food quality, much remains to be done on the understanding of the adaptive response of the bacteria to this parameter. In particular in the meat industry that uses extensively gaseous environment to preserve meat or meat products, no

data are available on the physiology of LAB or CNS used as starter culture under different redox conditions.

The microbial activities can contribute to the variation of E_h in food products. Their implication depends on their aerobic, facultative anaerobic or obligate anaerobic status. Anaerobes have high reducing capacities; they can decrease the E_h from -200 to -600 mV (Michelon et al. 2010). Among LAB, main studies have focused on dairy species. *Lactobacillus delbrueckii subsp. bulgaricus* and *Streptococccus thermophilus*, the main starters for yogurt, have low reducing capacities whereas *Lactococcus lactis* strains are able to decrease E_h to -200 mV (Martin et al. 2012).

The mechanisms involved in the reducing activity of *L. lactis* have recently been proposed. This includes reduction mechanisms of oxidizing compounds in the medium, such as oxygen, and mechanisms responsible for the stabilization of the reducing E_h in an anaerobic environment (Tachon et al. 2009; Michelon et al. 2010, 2012). In *L. lactis,* proteins in the cell envelope exposing exofacial thiol groups could be the cause of the reducing E_h measured at anaerobic medium condition (Michelon et al. 2010). Menaquinones and NADH dehydrogenases NoxA and NoxB have also been shown to be involved in the reduction of an aerobic medium (Tachon et al. 2009). However, the main meat starter LAB, *L. sakei*, has a quite different oxidative metabolism from *L. lactis*. Contrary to this species, *L. sakei* lacks cytochrome and menaquinones and is one of sole LAB species unable to respire even if heme and menaquinones are provided (Chaillou et al. 2005, Brooijmans et al. 2009, Pedersen et al. 2012). *L. sakei* possesses a heme dependant catalase and one superoxide dismutase, a catalase being absent in *L. lactis*. Moreover, a wide diversity in oxidant properties within this species has been reported (Guilbaud et al. 2012). It is thus likely that mechanisms involved in reducing activities differ between *L. lactis* and *L. sakei*. On the other side, *Staphylococcus xylosus*, the main CNS starter culture harbors one SOD and three catalases (Planchon et al. 2007, Barrière et al. 2002) and have high antioxidant properties (Barrière et al 2001a,b).

I.2.2.2. E_h and oxydations

To achieve good sensory properties and safety, the meat must be fresh and of high quality (Rahman 2007). The meat is generally sold raw, with no heat processing. Meat colour is an extremely important sensory characteristic according to which consumers make judgments of meat quality. It is influenced by the pigment content, the chemical form of pigment, and the meat structure (Lindahl et al. 2001). Some residual blood may also be present in meat, but it is generally minimal and is of little practical importance in considerations of meat colour. The degree of meat pigmentation is directly related to myoglobin content. In general, myoglobin concentration within a

given muscle will differ according to the species or age and is dependent on muscle fibber distribution (Lawrie 1985). Muscle composed predominantly of red fibbers contains more myoglobin than muscles with high white fibber content. Meat that does not have the acted cherry red colour is discriminated. This implies that meat colour determines the shelf life of meat. Cuts of meat that are darker due to too much MetMb will be viewed as old and undesirable for consumption.

The myoglobin form, which predominates in the surface of meat, determines the perceived colour. A typical scenario for colour expression of the various myoglobin forms can be observed during the cutting of meat. The deep portion of a fresh piece of meat is anoxic, and when sliced, will reveal an interior that is purplish-red in colour. Following exposure to air for 20-30 min, Mb will oxygenate to form cherry red MbO. As display time increases, MbO will oxidize to MetMb and the portion of meat displaying undesirable brownish discoloration will increase. The rate of MetMb formation is dependent on several factors, including the specific muscle, display temperature, type and intensity of lighting, and bacterial load. The pH drop during postmortem glycolysis results in increased internal reflectance (Bendall and Swatland 1988), increased lightness of the meat (Joo et al. 1999), decreased penetration depth of light, and changes in the selective light absorption through chromospheres like myoglobin and hemoglobin. The oxidation of heme iron and the chemical group bond at the sixth site will determine the meat colour. The relative proportions of the three myoglobin forms, Mb, MbO, and MetMb, affect the colour of fresh meat. The relative amounts of Mb, MbO, and MetMb in the meat depend on the oxygen availability, the autoxidation rate of myoglobin, and the MetMb-reducing capacity.

Colour deterioration and lipid oxidation may be linked, although the precise mechanisms are still unclear (Renerre 1990). Some control over increased susceptibility to oxidation can be attained by feeding higher levels of vitamin E, as an antioxidant active in meat (Asghar et al. 1991). The delaying of myoglobin oxidation is accomplished in a variety of ways. These include storage and display of meat under refrigerated conditions, hygienic preparation of meat cuts, and selective use of lighting. In addition, the application of antioxidants, such as ascorbic acid, citric acid, or α-tocopherol, may extend colour shelf life (Renerre et al. 1996, 1999, Gatellier et al. 2001, 2009).

Although lipid in meat contributes significantly to flavour, its oxidation will result in the production of free radicals, which lead to the formation of rancid odours and off-flavour. Oxidation might also play a role in controlling proteolytic activity of enzymes and could be linked to meat tenderness. The oxidative stability of meat depends upon the balance between anti- and prooxidants, including the concentration of polyunsaturated fatty acids (Mercier et al. 2004). It has been demonstrated that dietary fat and vitamin E supplementation can influence the antioxidant enzyme activities in meat (Renerre 1990). Meat fatty acid composition is influenced by a number of factors,

including muscle type and its oxidation (Geay et al. 2001). Factors that make muscle lipids susceptible to oxidation are either intrinsic to the products or related to technological process (Gatellier et al. 2009). Oxidation of muscle lipids produces primary and secondary products such as hydroperoxides, free radical, endoperoxides, malondialdehyde (MDA), epoxides, alkanes, hydrocarbons, alcohol, thiobarbituric acid reactive substance, and also acids that may be toxic to humans.

Initial storage conditions may affect subsequent lipid stability of frozen meat regardless of the storage temperature. At a frozen temperature of -20°C, usually used for domestic storage, MDA content increases with increased storage time. Low level of hydroperoxides in fresh meat increases rapidly to reach a maximum after several months of freezing. The threshold value for rancidity is about 1-2 mg MDA/kg of meat. However, consumers are unlikely to detect off-flavour at values below a threshold of about 0.5 mg MDA/kg (Gray and Pearson 1987). The double bonds located within polyunsaturated fatty acids are sites of chemical activity. Oxygen is a key element for lipid oxidation and may react with these sites to form peroxides, which lead to rancidity. Polyunsaturated fatty acids are susceptible to rancidity due to the double bonds. Meat with high concentrations of polyunsaturated fatty acids can develop rancid flavour faster than meat with less polyunsaturated fatty acids. The interaction of oxygen with polyunsaturated fatty acids is a no enzymatic process. Vacuum packaging of meat products therefore provides longer shelf life by excluding oxygen from the packaging.

Air oxygen penetrates the food material from outer atmosphere; therefore, it is useful to protect food surfaces. Meat or fish products are often protected by packaging materials, impregnated with antioxidants or by application of suitable antioxidants on the surface of materials, even when they are not packed. When packed, the packaging material is then of great importance (Terada and Breene 1989). If the material is permeable to oxygen, antioxidants may be added to the packaging to inhibit the diffusion of oxygen. These antioxidants may enter into the packed food, especially high-fat foods (Haigh 1986). Obviously, the only antioxidants that can be used in packaging are those that are allowed in foods, and they should be applied only in such amounts that the content in foods does not exceed the legal limits.

I.2.2.3. Measurement of E_h

In the same way as pH characterises the acid-base characteristics, E_h characterises the oxidoreductive properties of a medium. Using redox sensors as recommended by Clark and Hewitt in the middle of the last century, the redox state can be measured directly and monitored on line (Clark 1960). The use of sensors to measure the pH of a medium is very easy to implement

nowadays, however measuring E_h is more complex and requires a specific metrological approach (Galster 2000, Abraham et al. 2007).

The principle of measuring the E_h of a medium consists in measuring the electromotive force between a metal electrode (platinum, gold) and a reference electrode immersed in the medium being measured. In practice, two references are used: the calomel electrode and the silver/silver chloride electrode (Ag/AgCl). The reference electrode and the measuring electrode are combined and the oxidoreduction potential (E_m) is measured in volts. Oxidoreduction potentials must always be reported relative to the standard hydrogen electrode, consequently the potentials measured (E_m) must be adjusted according to the reference value (E^r) in relation to the hydrogen electrode: $E_h = E^r + E_m$. For example, the reference Ag/AgCl is 207 mV at 25°C ($E_h = 207 + 0.8\ (25 - T)$). Microelectrodes are preferred for heterogeneous systems, E_h gradients have been highlighted in cheese matrices such as Camembert (Abraham et al. 2007) and Cheddar (Topcu et al. 2008).

It has been shown that the evolution of the redox potential depends on the cheese manufacturing process implemented and the lactic starters used. In the manufacturing of a semi-hard cheese, Abraham and his co-workers have shown that during renneting, mixing, coagulum cutting and lactose removal phases, the medium remains oxidising due to persistent levels of dissolved oxygen between 20% and 60% from saturation, then progressively becomes reducing during pressing when no oxygen is incorporated thus allowing the lactic acid bacteria to express their reducing activity (Abraham et al. 2007). It thus seems possible to modulate E_h during cheese manufacturing via the action of LAB. Gas or reducing molecules can also be used.

Nevertheless, for meat applications, the redox sensors need to be adapted; the conventional sensors are glass combined sensors, with a diameter from 6 to 12 mm. The meat industry needs robust sensors with low diameters that can be used for measurement in meat carcass but also in individual meat products such as sausages. In addition, the state parameters that modify E_h in meat are not described in literature. For this reason, identification of such parameters and modelling of their effects on E_h values could be of great interest. Hence, in an objective of the recent concepts of Process Analytical Technology and Quality by Design, modelling and predictions of E_h would be intended to support innovation and efficiency in development of meat products, manufacturing, and quality assurance.

I.2.2.4. Concluding remarks about the practical use of physico-chemical properties in food science and biological systems

In summary, to describe the evolution of a food during processing, it appears important to predict the evolution of the physico-chemical properties, particularly E_h which is difficult to measure and to model. An integrated modelling approach would allow the prediction of different

physico-chemical properties (pH, a_w, E_h) and compensates for the lack of measurement techniques. However, models that can predict the physico-chemical properties of aqueous media are few and often complex. The simplified ones cannot be used to predict pH or a_w in a wide range of values, that is, they can only predict the properties of diluted aqueous solutions. The thermodynamic modelling of these physico-chemical properties is based on aqueous solution theory and on applied thermodynamics. In order to propose a complete modelling scheme of redox potential in complex media, allowing a reliable prediction of redox status in liquid solution and semi-solid products such as meat, it is necessary to develop a similar software tool as those that are involved for pH prediction in complex media and has been already proposed by the GePEB-IP team (Achard 1992). The measurable properties result from different twinned phenomena: chemical equilibrium and intermolecular forces, at short and long ranges. Ould-Moulaye has pointed out that the chemical potentials of each species are the only reliable and consistent data required to determine all of these physico-chemical properties (Ould-Moulaye 1998). An outline of the main thermodynamics knowledge required for understanding the key steps of the thermodynamics modelling is given below.

I.3. Thermodynamics foundation concepts

As discussed above, mixtures encountered in foods and biological systems are very complex and their chemical and thermodynamic parameters are not easily estimated from first principles nor easily determined experimentally. Although this situation has led to empiricism, nonetheless, knowledge of physico-chemical principles is essential to enhance our understanding of unit operations and behaviour of foods or any other mixture of molecules.

"When an operation is described, frequently it is helpful to envision a physical boundary around the operation. The element contained within the boundary is called a ***system" (Karel and Lund 2013)***. If the system exchanges both mass and energy with its environment, it is called an ***open*** system. If no mass or energy crosses the boundary, the system is ***isolated***. The system is said to be ***closed*** if no mass crosses the boundary, but it exchanges energy with its environment. A system with no heat flow across the boundary is ***adiabatic***, whereas one without work transfer is ***anergic***. If the pressure does not change, the system is ***isopiestic*** or ***isobaric*** and if the temperature of the system does not change, the system is ***isothermal*** (Karel and Lund 2003).

"An appreciation for thermodynamics of multicomponent mixtures is extremely important in food systems since food scientists are often interested in separating components, optimizing fragrance of products, or controlling gases or water in complex systems like foods "(Karel and Lund 2003). The final, or equilibrium, phase compositions depend on several variables, such as

temperature and pressure, and on the chemical nature and concentrations of the substances in the mixture. "Phase-equilibrium thermodynamics seeks to establish the relations among the various properties (in particular, temperature, pressure and composition) that ultimately prevail when two or more phases reach a state of equilibrium wherein all tendencies for further change has ceased" (Prausnitz et al. 1999). Below, we will study the behaviour of ideal multicomponent systems with ultimate application to real food systems.

I.3.1. Solution, partial molar functions and activity

A solution is a particular mixture containing one substance called the ***solvent*** that is liquid at the temperature of application in which other substances called the ***solutes,*** can be dissolved or mixed with to yield a homogeneous isotropic liquid. Water is of course the most widely used solvent because of its availability, low cost, nontoxicity, and safety as well as its ability to dissolve a great variety of substances, including electrolytes and polar organic substances. Certain substances are more soluble in polar organic solvents, which also have other properties that make them useful and desirable solvents. On the other hand, substances of low polarity that are essentially insoluble in water are often readily dissolved in nonpolar organic solvents. In many of these applications the solvents are employed in substantially pure form, as so-called neat solvents, and may have various impurities removed before use. For certain applications, however, neat solvents fall short of the mark as far as their dissolving power or other properties are concerned. It is then expedient to use ***solvent mixtures***, which may range from ***binary mixtures*** involving two solvents to ***ternary*** (three solvents) or even higher ***multicomponent mixtures***. Some of the admixed substances may not be liquids at the temperature of application hence mixtures involving them are not properly called solvent mixtures (Marcus 2002).

For pure substances, it is acceptable to use the ordinary thermodynamic functions (G, H, S, U, A, C_p, C_v…). However, for solutions the thermodynamic theory is expressed in terms of partial molar functions. This can best be explained by considering a solution so that adding 1 mol of X_1 or X_2 does not change the concentration appreciably, then we can measure the increase in volume (or any other thermodynamic quantity) when 1 mol (*i.e.* the amount of substance that contains an equal number of elementary entities as there are atoms in 12 g of the isotope of carbon-12) is added at constant temperature T and pressure p. By the way, the ***mole*** symbolized by ***mol*** is the SI unit for the amount of substance (n) that characterizes the number of particles as a multiple of the Avogadro number ($\mathcal{N} = 6.0221367\ 10^{23}\ mol^{-1}$). $\mathcal{N}$ relates the molar mass of an ***amount of substance*** (*i.e.* a standards-defined quantity that measures the size of an ensemble of elementary entities, such as atoms, molecules, electrons, and other particles) to its mass. The amount of substance is a macroscopic property which is sometimes referred to as ***chemical amount*** (Karel and Lund 2003).

The variation in volume (discussed above) is called the partial molar volume of the component in the solution at the specified temperature, pressure, and composition (Karel and Lund 2003). It is denoted by the symbol v_i *or* v_{X_i} and is written as:

$$v_i = v_{X_i} = \left(\frac{\partial V}{\partial n_i} \right)_{T,p,n_{j \neq i}} \tag{I-1}$$

To simplify notations, the subscript X_i will systematically be replaced by the subscript *i* (*i.e.* the X will be omitted).

These partial molar functions are necessary because adding pure components together to form solutions does not result in properties of the solution equal to the sum of the properties in pure state. For example, if 100 cm^3 of ethanol are mixed at 25°C with 100 cm^3 of water, the volume of the mixture is 190 cm^3. It can be shown that a property of partial molar functions is that the total value of the function is the sum of the product of moles of the component and its partial molar function (Karel and Lund 2003). For example, the total volume of a mixture from n_1 mol of X_1 and n_2 mol of X_2 is:

$$V = n_1 v_1 + n_2 v_2 \tag{I-2}$$

Partial molar functions can be defined for all state functions. For example,

$$s_i = \left(\frac{\partial S}{\partial n_i} \right)_{T,p,n_{j \neq i}} \; ; \quad h_i = \left(\frac{\partial H}{\partial n_i} \right)_{T,p,n_{j \neq i}} \; ; \; g_i = \left(\frac{\partial G}{\partial n_i} \right)_{T,p,n_{j \neq i}} = \mu_i \tag{I-3}$$

In addition, all thermodynamic relations that applied to state functions for pure component also apply to partial molar functions (Karel and Lund 2003). For example,

$$\left(\frac{\partial g_i}{\partial p} \right)_{T,n_i,n_{j \neq i}} = \left(\frac{\partial \mu_i}{\partial p} \right)_{T,n_i,n_{j \neq i}} = v_i \tag{I-4}$$

$$\left(\frac{\partial \mu_i}{\partial T} \right)_{p,n_i,n_{j \neq i}} = -s_i \tag{I-5}$$

$$\left(\frac{\partial \left(\frac{\mu_i}{T} \right)}{\partial T} \right)_{p,n_i,n_{j \neq i}} = -\frac{h_i}{T^2} \tag{I-6}$$

$$\left(\frac{\partial H_i}{\partial T}\right)_{p,n_i,n_{j\neq i}} = C_{P_i} \tag{I-7}$$

To illustrate its application, let us consider the formation of a binary solution of n mol containing n_1 mol of X_1 and n_2 mol of X_2 (*i.e.*, $n = n_1 + n_2$), similarly to equation I-2:

$$G_{sol} = n_1 g_1 + n_2 g_2 \tag{I-8}$$

At a given T and P, the ΔG, for the mixing process is:

$$\Delta G = n_1 g_1 + n_2 g_2 - n_1 g_1^0 - n_2 g_2^0 = n_1 (g_1 - g_1^0) + n_2 (g_2 - g_2^0) \tag{I-9}$$

where g_1^0 and g_2^0 denote the molar free energies of the pure components (Karel and Lund 2003).

Exactly analogous equations can be written for any extensive thermodynamic state variable (U, H, S, V, A, C_v, C_p, etc.) ultimately defining quantities such as enthalpy of solution, entropy of solution, etc.

The partial molar Gibbs free energy g_i is equivalent to the chemical potential μ_i. Instead of chemical potential, it is often convenient to use a related function, the absolute activity λ_i defined by:

$$\mu_i = RT \ln \lambda_i \tag{I-10}$$

Relations expressed for μ_i also apply to λ_i (Karel and Lund 2003). For example, the condition for equilibrium of a given component X_i between gas and liquid phases is:

$$\mu_i^g = \mu_i^l \quad or \quad \lambda_i^g = \lambda_i^l \tag{I-11}$$

For solutions, it is convenient also to express the difference between the value of μ_i in the solution and its value in some reference state. That is,

$$\mu_i - \mu_i^r = RT \ln\left(\frac{\lambda_i}{\lambda_i^r}\right) = RT \ln\left(a_i\right) \tag{I-12}$$

where the ratio of absolute activity to the activity in some reference state defines a relative activity a_i.

For non-electrolyte solutions of liquids, a convenient reference state is the pure liquid at a pressure p = 1atm and the temperature specified for the solution. If the reference state is identified at μ_i^0 and λ_i^0 Eq. I-9 becomes:

$$\mu_i - \mu_i^0 = RT \ln\left(\frac{\lambda_i}{\lambda_i^0}\right) = RT \ln\left(a_i\right) \tag{I-13}$$

and

$$a_i = \frac{\lambda_i}{\lambda_i^0} \tag{I-14}$$

Defined in this way, the relative activity is called simply ***activity*** of the species X_i (Karel and Lund 2003). Of course, if the species X_i is the water molecule, thus:

$$a_i = a_{H_2O} = a_W \tag{I-15}$$

Likewise, the quantity pH is the negative logarithm of the activity of the solvated proton H_3O^+,

$$pH = -\log_{10}\left(a_{H_3O^+}\right) \tag{I-16}$$

By convention, activity is treated as a dimensionless quantity, although its actual value depends on customary choices of standard state for the species. The activity of pure substances in condensed phases (solid or liquids) is normally taken as unity (the number 1). Activity depends on temperature, pressure and composition of the mixture, among other things. For gases, the effective partial pressure is usually referred to as fugacity f_i .

A chemical reaction is a process that leads to the transformation of one set of chemical substances called ***reactants*** A_i to another set called ***products*** B_j. This operation is written in the form of an equation of chemical reaction:

$$\alpha_1 A_1 + \alpha_2 A_2 + \ldots \rightleftharpoons \beta_1 B_1 + \beta_2 B_2 + \ldots \tag{I-17}$$

where α_i and β_j are appropriate integers that denote the respective stoichiometric numbers of the reactant A_i and the product B_j. Different symbols are used to connect the reactants and products with the following meanings: $=$for a stoichiometric relation; $\rightarrow$ for a net forward reaction; $\rightleftarrows$ for a reaction in both directions; $\rightleftharpoons$ for equilibrium (Prausnitz et al. 1999).

I.3.2. Acid/base reaction

As defined by Brönsted, an acid is ‘a species having a tendency to lose a proton’ while a base is ‘a species having a tendency to add on a proton’. Hence for every acid, AH, there is a conjugated base, A^- and for every base, B, there is a conjugated acid, BH^+.

If AH (or BH^+) is a strong acid, *i.e.* it has a great tendency to lose protons, it follows that its conjugated base A^- (or B) is a weak base, *i.e.* has only a small tendency to accept protons.

In aqueous solution, acids react with water acting as a base:

$$AH\,(or\,BH^+) \rightleftharpoons H^+ + A^-\,(or\,B) \tag{I-18}$$

$$AH\,(or\,BH^+) + H_2O\,(base) \rightleftharpoons H_3O^+ + A^-\,(or\,B) \tag{I-19}$$

and bases react with water acting as an acid:

$$A^-\,(or\,B) + H_2O\,(acid) \rightleftharpoons AH\,(or\,BH^+) + OH^- \tag{I-20}$$

In dilute aqueous solution where almost all measurements are made, water is thus the solvent and its activity is taken as unity. Then, the acidic dissociation (or ionization) constant K_a is calculated with the following equation:

$$K_a = \prod_i a_i^{\nu_i} = \frac{a_{H^+} a_B}{a_A} = \frac{a_{H_3O^+} a_B}{a_A} \tag{I-21}$$

B and A represent respectively the base and acid species. This equation can be written in the form:

$$pK_a = pH + \log\left(\frac{a_A}{a_B}\right) \tag{I-22}$$

where pK_a is the negative logarithm of K_a, and is equal to the *pH* at which the activities of A and B are equal.

I.3.3. Oxidation-reduction (redox) reaction

To determine the equilibrium properties of a redox reaction, it is necessary to know the stoichiometry and the electrostatic potential of the solution, expressed in volts (V).

$$ox + ne^- \rightleftharpoons red \tag{I-23}$$

Ould Moulaye (1998) has verified that the calculation of the redox potential corresponds to the same information as the determination of the chemical potential of the electron:

Indeed, by definition the reference properties of the electron are calculated from the normal hydrogen electrode data (NHE):

$$\begin{cases} H^+ + e^- \rightleftharpoons \dfrac{1}{2} H_2 \\ E = 0 : m_{H^+_s} = m_{e^-_s} = 1\,mol.kg^{-1};\; p_{H2} = 1\,atm, \forall T \end{cases} \tag{I-24}$$

The chemical potential of the electrode should be interpreted as information about the chemical potential of the electron, since the reaction Gibbs free energy is expressed as:

$$\Delta\mu = \frac{1}{2}\mu_{H_2} - \mu_{H^+} - \mu_{e^-} \tag{I-25}$$

At thermodynamic equilibrium (where $\Delta\mu = 0$), by definition the chemical potential of the electron is linked to the redox potential (1 Volt = 1Joule /Coulomb) by the relation:

$$\mu_{e^-} = -\mathcal{F}E \tag{I-26}$$

At equilibrium, the Gibbs free energy of reaction is given by:

$$\Delta\mu = \frac{1}{2}\mu_{H_2} - \mu_{H^+} + \beta_{e^-}\mu_{e^-} = 0 \tag{I-27}$$

where β_{e^-} is the stoichiometric coefficient of the electron ($\beta_{e^-} = -1$ for the standard hydrogen electrode) and $\mathcal{F} = \mathcal{N}*e = 6.02205\,10^{23} * 1.60919\,10^{-19} = 96485\,C.mol^{-1}$ is the constant of Faraday. Thus,

$$\mu_{e^-} = -\frac{1}{\beta_{e^-}}\left(\mu_{H_2} - \mu_{H^+}\right) \tag{I-28}$$

If we used H_3O^+ instead of H^+ in the previous equation (being aware of the stoichiometric equilibrium), then:

$$\begin{cases} H_3O^+ + e^- \rightleftharpoons \frac{1}{2}H_2 + H_2O \\ \mu_{e^-} = -\frac{1}{\beta_{e^-}}\left(\mu_{H_2} + \mu_{H_2O} - \mu_{H_3O^+}\right) \end{cases} \tag{I-29}$$

Similarly, the study of the equilibrium properties of the redox reaction (equation I-23) leads to:

$$\Delta\mu = \mu_{red} - \mu_{ox} - n\mu_{e^-} \tag{I-30}$$

Thus,

$$\mu_{e^-} = \frac{1}{n}\left(\mu_{red} - \mu_{ox}\right) = \frac{1}{n}\left(\left(\mu^0_{red} + RT\ln a_{red}\right) - \left(\mu^0_{ox} + RT\ln a_{ox}\right)\right) \tag{I-31}$$

The value of μ_{e^-} corresponds to the standard potential of the redox couple ox/red, denoted E^0. The latter is defined in reference conditions, *i.e.* at T=25°C for a hypothetical ideal solution where both the oxidizer and the reducer are taken at the reference molality of 1 mol.kg^{-1}.

$$E^0 = -\frac{\mu_{e^-}}{\mathcal{F}} = \frac{1}{n\mathcal{F}}\left(\mu^0_{red} - \mu^0_{ox}\right) \tag{I-32}$$

The redox potential in solution is given by:

$$E_h = E^0 + \frac{RT}{n\mathcal{F}} \ln \frac{a_{ox}}{a_{red}} \qquad \text{(I-33)}$$

It appears clearly that, to determine the redox potential in solution, one must be able to determine at the same time the standard redox potential *i.e.* the difference between the chemical potentials μ^0_{red} and μ^0_{ox} in reference conditions, and the respective activities of both reducer and oxidizer.

I.3.4. Chemical potential in solu tion and its derivative properties

All the physico-chemical properties discussed above can be determined if one is able to predict the chemical potential of any species X_i in solution, which is an essential and fundamental parameter for characterizing the above mentioned equilibria:

$$\mu_i = \left(\frac{\partial G}{\partial n_i}\right)_{T,\,p,\,n_{j\neq i}} = \mu_i^0 + RT\ln\left(a_i\right) = \mu_i^0 + RT\ln\left(\gamma_i x_i\right) \qquad \text{(I-34)}$$

This equation shows that the chemical potential μ_i of a given compound X_i depends mainly on 2 parameters: its Gibbs free energy of formation μ_i^0 and its activity coefficient γ_i which are both linked to the chosen reference state (since temperature and mixture composition are known). Then, the main thermodynamic modelling task is the prediction and/or the collection of formation properties data and/or the development of a predictive model of activity coefficients. To do so, a predictive approach is required to determine the chemical potentials of each compound of the mixture and all related thermodynamics properties.

The activity coefficient can be defined as the ratio of the activity to the mole fraction x_i :

$$\gamma_i = \frac{a_i}{x_i} \; or \; a_i = \gamma_i x_i \qquad \text{(I-35)}$$

where γ_i is the activity coefficient of the species X_i, by definition $\gamma_i = 1$ for an ideal solution.

As discussed in details by Prausnitz et al. (1999), an ideal gas is one in which there are no cohesive forces acting on the molecules and hence the internal pressure is zero. For an ideal solution, the cohesive forces are equal and uniform between all species. For example, if there are two components X_1 and X_2, the intermolecular forces between X_1 and X_1; X_2 and X_2; and X_1 and X_2 are all the same. Just as the ideal gas law has served a useful purpose in treating practical cases even when large deviations from ideality occurs, the concept of an ideal solution serves the same

purpose. As a result, the properties of ideal mixtures can be expressed directly in terms of simple concentrations or partial pressures of the substances present *e.g.* Raoult's law:

$$y_i \, p = x_i \, P_i^s \tag{I-36}$$

where y_i, p and P_i^s denote respectively the gas phase molar fraction, the pressure and the saturation (vapour) pressure of pure liquid X_i at temperature T (Prausnitz et al. 1999).

Deviations from ideality are accommodated by modifying the concentration by an activity coefficient. Analogously, expressions involving gases can be adjusted for non-ideality by scaling partial pressures by a fugacity coefficient.

Similarly to the activity, the activity coefficient γ_i is also a dimensionless quantity that relates the activity to a measured amount fraction x_i (or y_i in the gas phase), molality m_i or amount concentration c_i :

$$a_i = \gamma_i^{(x)} x_i = \gamma_i^{(m)} \frac{m_i}{m_{is}} = \gamma_i^{(c)} \frac{c_i}{c_{is}} \tag{I-37}$$

The division by the standard molality m_{is} or the standard amount concentration c_{is} is necessary to ensure that both the activity and the activity coefficient are dimensionless as is conventional. The activity is the same regardless of the ways to express composition and the standard state chosen, so the above expressions are equal. For ideal solutions, the activity coefficient is equal to unity, thus the activity can be substituted with the appropriate dimensionless measure of composition x_i , $\frac{m_i}{m_{is}}$ or $\frac{c_i}{c_{is}}$. The composition of mixtures are differently characterized regarding the study areas, for instance in chemistry, the molar fraction scale is used to establish the thermodynamic models, while in biology (microbiology or biochemistry) it is common to use molar or mass concentrations. This characterization depends also on the specificities of the studied mixtures; it is more appropriate to use the molality scale to characterize the solutions containing ions. The equations allowing the conversion of any of these composition scales from any other are given in Table I-1. For a given compound X_i (denoted by the subscript i); x_i , w_i , c_i , m_i , M_i and ρ_i denote respectively (in this table) its molar fraction, mass fraction, molar concentration, molality, molar mass and density.

Table I-1: Relations between the different scales of composition

Conversion of / To	x_i	w_i	m_i	c_i
x_i	$x_i = x_i$	$x_i = \dfrac{\frac{w_i}{M_i}}{\sum \frac{w_j}{M_j}}$	$x_1 = \dfrac{\frac{1}{M_1}}{\frac{1}{M_1} + \sum m_j}$ $x_i = \dfrac{m_i}{\frac{1}{M_1} + \sum m_j}$	$x_i = \dfrac{c_i}{\sum c_j}$
w_i	$w_i = \dfrac{x_i M_i}{\sum x_j M_j}$	$w_i = w_i$	$w_1 = \dfrac{1}{1 + \sum M_j m_j}$ $w_i = \dfrac{M_i m_i}{1 + \sum M_j m_j}$	$w_i = \dfrac{M_i c_i}{1 + \sum M_j c_j}$
m_i	$m_i = \dfrac{x_i}{M_1 x_1}$	$m_i = \dfrac{w_i}{M_i w_1}$	$m_i = m_i$	$m_i = \dfrac{c_i}{M_1 c_1}$
c_i	$c_i = \dfrac{\rho_{mix} x_i}{\sum M_j x_j}$	$c_i = \dfrac{\rho_{mix} w_i}{M_i}$	$c_1 = \dfrac{\rho_{mix}}{M_1 \left(1 + \sum M_j m_j\right)}$ $c_i = \dfrac{m_i \rho_{mix}}{1 + \sum M_j m_j}$	$c_i = c_i$

It is also important to notice that the value of the activity coefficient strongly depends on the chosen reference state which must be clearly defined in order to avoid confusion.

Two conventions are generally used for this purpose, the symmetric convention and the asymmetric convention. In the symmetric convention, all the solutes and solvent are taken in the pure component reference state. In the asymmetric convention, a pure component reference state is chosen for the solvent molecules and all the other solutes are taken at infinite dilution.

In the asymmetric convention, two composition scales are generally used, the molarity scale (or molar fraction scale) and the molality scale. The molarity or molar fraction scale is generally used to determine the thermodynamic properties of non-electrolytes solutions, for which a pure component reference state is usually adopted in calculations. The molality scale is generally used to determine the properties of electrolytes solutions for which the infinite dilution in water reference state is usually adopted in calculations and experiments.

The relations allowing the conversion of data from one reference state/convention to any other are given in Table I-2 for compositions, Table I-3 for activity coefficients and Table I-4 for the reference chemical potentials of both solutes (subscript *i*) and solvent (subscript *1*).

***Table I-2* : Relations between the compositions in the different reference states.**

The symbol M denotes the molar mass in kg.mol^{-1}.

To / Conversion of	Reference state: pure compound (PC)	Reference state: Infinite dilution (DI)	Reference state: Molality scale (m)
Reference state: pure compound (PC)	x_1 x_i	x_1 x_i	$z_1 = \exp\left(-\frac{1-x_1}{x_1}\right)$ $m_i = \frac{x_i}{x_i M_1}$
Reference state: Infinite dilution (ID)	x_1 x_i	x_1 x_i	$z_1 = \exp\left(-\frac{1-x_1}{x_1}\right)$ $m_i = \frac{x_i}{x_i M_1}$
Reference state: Molality scale (MS)	$x_1 = \frac{1}{1-\ln z_1}$ $x_i = x_1 M_1 m_i$	$x_1 = \frac{1}{1-\ln z_1}$ $x_i = x_1 M_1 m_i$	z_1 m_i

***Table I-3* : Relations between the activities coefficients in the different reference states.**

γ_{is} denotes the activity coefficient of the species X_i in a hypothetical ideal solution of molality m_{is} ($\gamma_{is} = \gamma_i^{\infty}$).

To / Conversion of	Reference state: pure compound (PC)	Reference state: Infinite dilution (DI)	Reference state: Molality scale (m)
Reference state: pure compound (PC)	$\gamma_1 = \gamma_1$ $\gamma_i = \gamma_i$	$\gamma_1^* = \gamma_1$ $\gamma_i^* = \frac{\gamma_i}{\gamma_i^{\infty}}$	$\gamma_1^m = \gamma_1 x_1 \exp\left(\frac{1-x_1}{x_1}\right)$ $\gamma_i^m = x_1 \frac{\gamma_i}{\gamma_{is}}$
Reference state: Infinite dilution (ID)	$\gamma_1 = \gamma_1^*$ $\gamma_i = \gamma_i^* \gamma_i^{\infty}$	$\gamma_1^* = \gamma_1^*$ $\gamma_i^* = \gamma_i^*$	$\gamma_1^m = \gamma_1^* x_1 \exp\left(\frac{1-x_1}{x_1}\right)$ $\gamma_i^m = x_1 \gamma_i^*$
Reference state: Molality scale (MS)	$\gamma_1 = \frac{\gamma_1^m}{x_1}\exp\left(-\frac{1-x_1}{x_1}\right)$ $\gamma_i = \frac{\gamma_i^m}{x_1}\gamma_{is}$	$\gamma_1^* = \frac{\gamma_1^m}{x_1}\exp\left(-\frac{1-x_1}{x_1}\right)$ $\gamma_i^* = \frac{\gamma_i^m}{x_1}$	$\gamma_1^m = \gamma_1^m$ $\gamma_A^m = \gamma_A^m$

Table I-4 : Relations between the reference chemical potentials in the different reference states.

Conversion of \ To	Reference state: pure compound (PC)	Reference state: Infinite dilution (DI)	Reference state: Molality scale (MS)
Reference state: pure compound (PC)	μ_1^{0L} μ_i^{0L}	$\mu_1^* = \mu_1^{0L}$ $\mu_i^* = \mu_i^{0L} + RT\ln\gamma_i^\infty$	$\mu_1^m = \mu_1^{0L}$ $\mu_i^m = \mu_i^{0L} + RT\ln\left(\gamma_i^\infty M_1 m_{is}\right)$
Reference state: Infinite dilution (ID)	$\mu_1^{0L} = \mu_1^*$ $\mu_i^{0L} = \mu_i^* - RT\ln\gamma_i^\infty$	μ_1^* μ_i^*	$\mu_1^m = \mu_1^*$ $\mu_i^m = \mu_i^* + RT\ln\left(M_1 m_{is}\right)$
Reference state: Molality scale (MS)	$\mu_1^{0L} = \mu_1^m$ $\mu_i^{0L} = \mu_i^m - RT\ln\left(\gamma_i^\infty M_1 m_{is}\right)$	$\mu_1^* = \mu_1^m$ $\mu_i^* = \mu_i^m - RT\ln\left(M_1 m_{is}\right)$	μ_1^m μ_i^m

All the relations expressed in this section show that the calculation of the physico-chemical parameters depends on the consistent calculation of the chemical potentials of each species in solution. The latter can be splitted to two kinds of information: the reference chemical potential of the species (which could ideally correspond to one of its formation properties, the Gibbs free energy of formation) and the activity coefficient (which is an excess property).

For the most common organic molecules, the formation properties data are generally available in databases like DIPPR, the NIST databases and the CRC handbooks. However for free radicals, and a wide variety of ions and neutral molecules of interest in foods and biological systems, the formation properties are not available in the literature, thus one has to determine a predictive tool to predict this property.

The activity coefficients are generally predicted using an excess Gibbs free energy model using one arbitrary fixed reference state for each substance in solution. Then any related equilibrium property can be deduced using the thermodynamic relations already linking this property to the chemical potential, as illustrated on Figure I-3.

The equilibrium condition between two given phases, for instance a gas phase (superscript ***g***) and a liquid phase (superscript ***l***), requires the equality of the chemical potential of any species X_i in both phases, *i.e.* $\mu_i^g = \mu_i^l$. Thus, the determination of the chemical potential of each compound of the mixture allows the characterization of both physical equilibria (SLE, LLE and VLE) and chemical equilibria (dissociation, hydration, complexation and redox) as mentioned by Achard (1992), Ould-Moulaye (1998) and Ben Gaïda (2007).

In summary, the main thermodynamic modelling task is the prediction and/or the collection of formation properties data and/or the development of a predictive model of activity coefficients. This task must be achieved consistently in order to determine accurately the physico-chemical properties discussed earlier. Indeed, the pH is calculated from the activity of the proton, the pK_a can be calculated either from a ratio of activities or from a difference of chemical potentials of formation. However, to be able to predict the redox potential E_h, it is necessary to have a robust predictive model that can determine both the activity coefficient and the formation properties of each species in the chosen reference state. To do so, a predictive approach is required to generalize the existing models up to the simultaneous prediction of equilibrium properties in a chosen reference state (generally the infinite dilution in water), and the prediction of the non-ideality (*i.e.* the deviation from Raoult's law) for concentrated solutions.

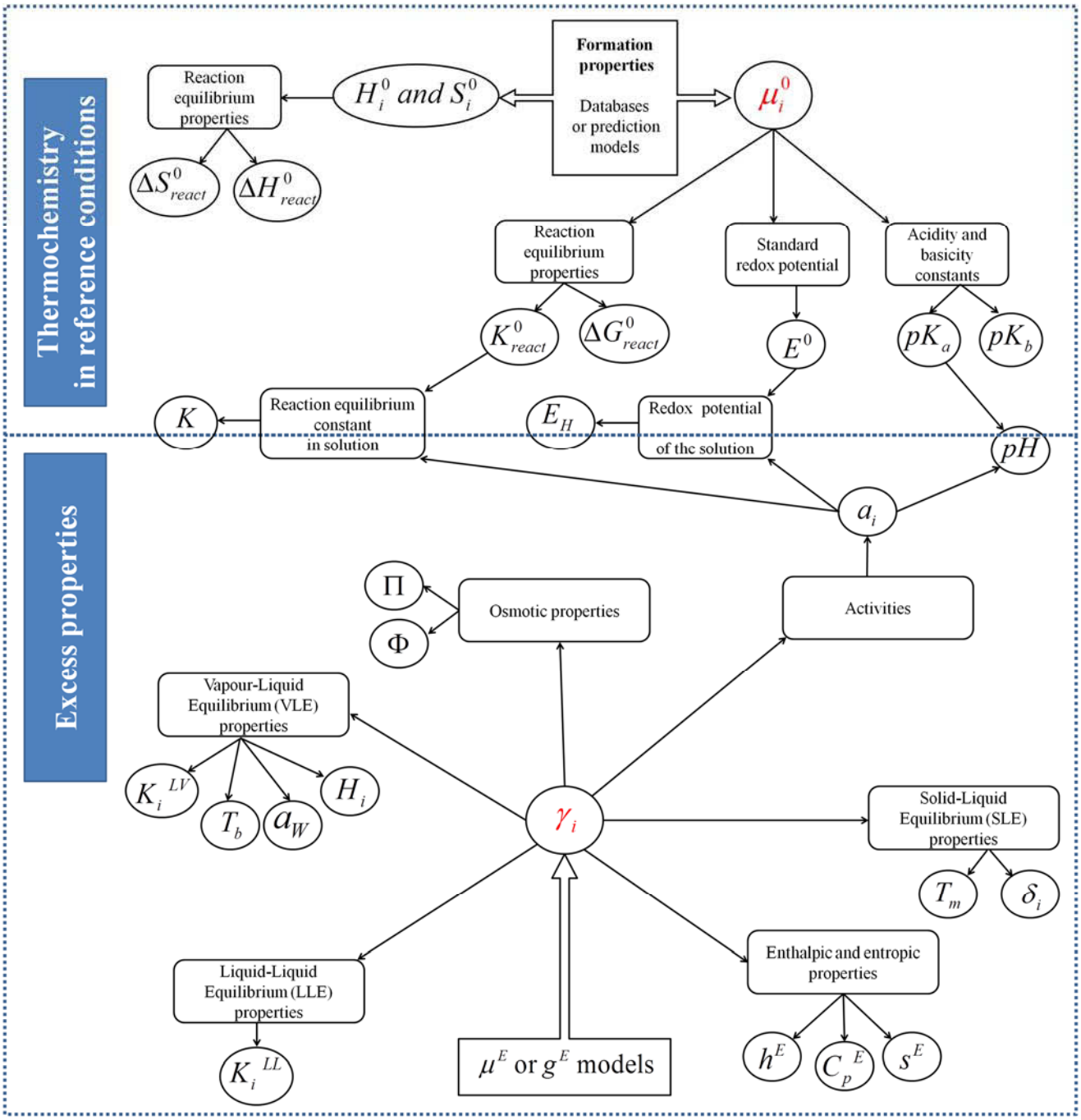


Figure I-3: Thermodynamic properties related to the chemical potential of the species in solution.

The calculation of the physico-chemical parameters depends on the consistent calculation of the chemical potentials of each species in solution potential expressed as $\mu_i = \mu_i^0 + RT \ln(\gamma_i x_i)$. *Thus, the modelling task is splitted to the determination of the reference chemical potential of the species (which should correspond to one of its formation properties, the Gibbs free energy of formation in order to perform thermochemistry calculations) and of the activity coefficient (which is an excess property).*

However, it is well established that the thermodynamic properties of a mixture depend on the interaction forces that exist between the species in the mixture (Israelachvili 2011). The

scientific description of these interactions touches on a very broad area of phenomena in physics, chemistry, chemical engineering, and biology, in which there have been tremendous advances that can be viewed in isolation within each discipline or within a broader multidisciplinary framework (Israelachvili 2011). The latter approach is adopted in this work which is focused on the prediction of thermodynamics equilibrium properties (the kinetic aspect is not treated herein). The physical properties of every pure substance depend directly on the nature of the molecules of which it consists. The ultimate generalization of physical properties of fluids will require a complete understanding of molecular behaviour, which we do not yet have. Since the nineteenth century many pieces of the puzzle of molecular behaviour have fallen into place, and a useful, though incomplete, generalization has been developed.

I.4. Brief historic of intermolecular forces concepts in thermodynamics

The following discussion is inspired from the books of Israelachvili (2011), Klamt (2005), Prausnitz et al. (1999) and Reid et al. (1977).

The laws of Charles and Gay-Lussac were combined with Avogadro's hypothesis to form the gas law, PV = nRT, which was perhaps the first important correlation of properties. Deviations from the ideal-gas law, though often small, were tied to the fundamental nature of the molecules. The equation of van der Waals, the virial equation, and other equations of state express these quantitatively, and have greatly facilitated progress in the development of a basic molecular theory. Though its origins are ancient, the molecular theory was not generally accepted until the nineteenth century (Reid et al. 1977, Israelachvili 2011).

Indeed, in the first half of the nineteenth century, there was a controversy about whether matter was a continuum or made up of molecules (atoms, particles, and corpuscles), with the balance slowly tilting toward the continuum picture. By the mid-nineteenth century, Maxwell (1831-1879) stated that no one believed in molecules any longer because anyone had shown that they did not exist but because "All those who believe in molecules are now dead" (Israelachvili 2011).

The latter half of the nineteenth century saw the return of molecules, thanks to the successes of the new kinetic theory and the van der Waals equation of state, both of which required attracting molecules to explain them. Clausius (1822-1888), Maxwell (1831-1879), van der Waals (1837-1923) and Boltzmann (1844-1906) were the main players in unifying continuum theories such as thermodynamics and mean-field theories with molecular theories through the new field of statistical

mechanics. But the origin of the forces themselves remained a mystery until the advent of quantum theory in the 1920s (Israelachvili 2011).

An important theoretical advance of the nineteenth century, thanks to mathematical developments of the previous century, was the use of energy to analyze interactions and the properties of many-body systems; this allowed progress to be made in various areas, including thermodynamics. William Rankine (1820-1872) is generally considered to have pioneered the science of energetics whereby the vector forces and motions of Archimedes, Galileo, and Newton were replaced by scalar energy functions. In many cases the two approaches are equivalent, but where time -and rate-dependent interactions are involved, they are not. In other words, the force is not always simply the derivative of the energy-distance function. New terms were coming into usage in addition to simple energy and force: internal energy, available energy, energy or force per unit length, per unit area, per unit volume (energy density), internal pressure, applied pressure, and more, and it remains an intellectual challenge to this day to be able to clearly distinguish among the many different types of forces and energies. During the nineteenth century, it was believed that one simple universal force law or potential energy function, similar to Newton's law for the gravitational force, would eventually be found to account for all intermolecular attractions. To this end, a number of interaction potentials were proposed that invariably contained the masses of the molecules, attesting to the belief at the time that these forces are related to gravitational forces. In the latter part of the nineteenth century, hopes for an all-embracing force law dwindled as it became increasingly apparent that no suitable candidate would be forthcoming to explain the multitude of phenomena quantitatively. However, by this time the modern concept of surface tension forces was becoming established, as was the recognition that these forces are the same as those that hold molecules together in solids and liquids, and that in both cases they arise from interactions acting over very short distances. It was also shown how these very short-range surface tension forces can account for such macroscopic phenomena as capillarity, the shapes of liquid drops on surfaces, the contact angle between coalescing soap bubbles, and the breakup of a jet of water into spherical droplets. Thus it was established that very short-range forces can lead to very long-range (*i.e.*, macroscopic) effects. It is therefore wrong to associate long-range effects with long-range forces. For what is more important is the strength of the interaction, short-range forces tend to be stronger than long-range forces (Israelachvili 2011).

By the beginning of the twentieth century, coinciding (not unexpectedly) to the end of classical era of physics and chemistry, a number of conceptual changes had occurred. The period that had started with Newton and ended with van der Waals, Boltzmann, Maxwell, and Gibbs saw the abandonment of the purely mechanistic view of intermolecular forces and the adoption of thermodynamic and probabilistic concepts such as free energy and entropy. It was now appreciated

that heat is also a form of energy but different from mechanical energy and not due to some substance or intermolecular potential - for example that the pressure of an ideal gas does not arise from any particular repulsive intermolecular potential between the molecules; indeed, it arises even when there is no interaction between the molecules and even for molecules of arbitrarily small size. Accompanying these conceptual developments, it became apparent that there is a big gap between knowing the pair potential or force law between any two molecules and understanding how an ensemble of such molecules will behave. For example, the mere knowledge that air molecules attract one another does not mean that they will condense into a liquid or a solid at any given temperature or pressure. Even today, there is no ready recipe for deriving the properties of condensed phases from the intermolecular potentials, and vice versa (Prausnitz et al. 1999, Israelachvili 2011).

Only with the elucidation of the electronic structure of atoms and molecules and the development of the quantum theory in the 1920s was it possible to understand the origin of intermolecular forces and to derive expressions for their interaction potentials. It was soon established that all intermolecular forces are essentially electrostatic in origin. This is encapsulated in the Hellman-Feynman theorem, which states that once the spatial distribution of the electron clouds has been determined by solving the Schrödinger equation, the intermolecular forces may be calculated on the basis of straightforward classical electrostatics. This theorem greatly simplified notions of the nature of intermolecular forces. Thus, for two charges, we have the familiar inverse-square Coulomb force, while for moving charges we have electromagnetic forces, and for the complex fluctuating charge distributions occurring in and around atoms, we obtain the various interatomic and intermolecular bonding forces familiar to physics, chemistry, and biology (Israelachvili 2011).

This seems marvellously simple. Unfortunately, exact solutions of the Schrödinger equation are not easy to come by. In fact, it is even too difficult to solve (exactly) something as simple as two hydrogen atoms, let alone two water molecules, interacting in a vacuum. For this reason, it has been found useful to classify intermolecular interactions into a number of seemingly different categories, even though they all have the same fundamental origin. Thus, such commonly encountered terms as ionic bonds, metallic bonds, van der Waals forces are a result of this classification, often accompanied by further divisions into strong and weak interactions and short-range and long-range forces. Such distinctions can be very useful, but they can also lead to confusion- for example, when the same interaction is "counted twice" or when two normally distinct interactions are strongly coupled. As more and more information and interest is accumulating on microscopic, mesoscopic, and nanoscopic (molecular-scale) systems and processes, there is a natural desire to understand and control these phenomena by manipulating the operative forces (Israelachvili 2011).

In 1929, Dirac wrote that: "The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble." In the 1950s, theoretical chemists started to take advantage of evolving computer technology and were able to achieve approximate solutions for the wave functions and electron densities of other small molecules. The rapid development of computer capabilities as well as of efficient theoretical approximations and of computer codes, allowed the science of quantum chemistry to make enormous progress in the following decades. "As representative of a large number of brilliant researchers, John A. Pople and Walter Kohn were awarded the Nobel Prize for their achievement in quantum chemistry in 1998. However, despite rapid progress, quantum chemistry remained an academic science for about 40 years. Even by using large supercomputers it was only possible to calculate the well-known properties of two or three atomic molecules with acceptable accuracy, while real predictions for new molecules of industrially relevant size were impossible. The assessment that quantum chemistry is unable to treat real-world problems became so manifest during these years that many in the chemical community did not realize the change in this situation that took place around 1990. This change can be mainly attributed to the development of gradient corrected density functionals, which allowed the application of density functional theory (DFT) to molecular systems (Hohenberg and Kohn 1964; Becke 1998), after it had been applied successfully in solid-state theory about ten years before. At about the same computational cost as for the mean-field theory of Hartree and Fock (HF) (Hartree 1928; Fock 1930), DFT takes the electron correlation into account quite accurately and thus provides a large increase in accuracy over the HF level. In its most efficient implementations, DFT today allows for good-quality ground state geometry optimizations and property calculations of molecules up to 40 atoms within a day or less on inexpensive PC hardware. By efficient combination of DFT geometry optimizations and frequency calculations with higher-level single-point (SP) calculations, *e.g.* coupled cluster calculations, reaction enthalpies can be calculated for industrially relevant compounds, with chemical accuracy in a few days on standard parallel computer systems. Thus, quantum chemistry has achieved a state where it can provide useful information about new molecules at lower cost and in shorter time than experiments can do. It has the potential to be a predictive tool, and the speed of computer development will rapidly expand its predictive power in the foreseeable future" (Klamt 2005).

"Despite the tremendous progress made in this field, there is still a severe drawback. The quantum chemistry tools developed by theoretical chemists are primary suited for isolated molecules in vacuum or in a dilute gas, where intermolecular interactions are negligible. Another class of quantum codes that has been developed mainly by solid-state physicists is suitable for crystalline

systems, taking advantage of the periodic boundary conditions. However, most of industrially relevant chemical processes, and almost all of biochemistry do not happen in the gas phase or in crystals, but mainly in a liquid phase or sometimes in an amorphous solid phase, where the quantum chemical methods are not suitable. On the one hand, the weak intermolecular forces, which play an important role in condensed systems, require a very high level of quantum theory for an appropriate representation, and DFT is not sufficiently accurate for this. On the other hand, the representation of condensed liquid systems by large ensembles of molecules converges only very slowly with an increase in ensemble size, even if artificial periodic boundary conditions are applied to avoid artifacts arising from the surface. Together with the fact that DFT calculations still scale at least quadratically with the system size, and that higher-order methods are even more demanding, there is no realistic prospect that quantum chemistry alone will be able to solve practical problems of liquid-phase chemistry in the near future. Even the well-known Car-Parrinello method (Car and Parrinello 1985), which is a combination of approximate DFT and molecular dynamics simulation with periodic boundary conditions, still requires external constraints, and yields only short snapshots of relatively small condensed ensembles of molecules. However, since this calculation will require very high computational costs, this approach may be helpful in generating some qualitative insight into mechanisms acting in liquid systems, but it is far from being a predictive tool for practical chemical questions" (Klamt 2005).

"The most obvious way to achieve a theoretical description of molecules in the liquid state is the path from single molecules, through small and large clusters of molecules, to very large ensembles of molecules. Even at times when realistic quantum chemical simulations were impossible for molecular sizes of practical relevance, people started in this direction by developing classical approximations for the intra- and intermolecular interactions, often called force-field (FF) or molecular mechanics (MM) methods, instead of the much more appropriate, but impractical quantum theoretical description. The large number of parameters appearing in the sophisticated functional expressions of such force fields for bond lengths, bond angles, torsion angles, electrostatic forces, dispersive forces, and hydrogen bond interactions have been adjusted mainly to experimental data, and sometimes even to high-level quantum chemical results for small molecules in the last decades"(Klamt 2005).

Liquid states are, to a large degree stabilized by the entropy of configurational disorder, and fluid-phase simulations must take this disorder into account, either by averaging the relevant properties over a sufficiently long time in a molecular dynamic (MD) simulation of a system that is representative for the liquid system (Jorgensen 1991), or by taking thermodynamics averages over a large number of randomly chosen configurations of such a system, the so called Monte Carlo (MC) simulations (Metropolis et al. 1953, van Gunsteren and Berendsen 1990). For thermodynamic

equilibrium simulations, where dynamic information is not requested, the latter technique has turned out to be more efficient. Powerful algorithms have been developed to sample the thermodynamically relevant configurations efficiently. Thus, using large supercomputers, very reliable MC simulations can now be performed for alkanes and other non-polar liquids and mixtures. However, for most classes of other compounds such simulations still suffer from the inaccuracies of the classical force fields required for the evaluation of the intermolecular interactions (Errington and Panagiotopoulos 1998; Martin and Siepmann 1998; Nath et al. 1998, Prausnitz et al. 1999, Klamt 2005, Israelachvili 2011).

Recent theoretical trends have therefore focused on obtaining semi rigorous equations that are nevertheless adequate for treating multimolecular and multicomponent systems, but even then one may need to resort to a computer to analyze such complex systems. Indeed, it is now not uncommon for a Monte Carlo or Molecular Dynamics computer simulation to result in a totally new equation that accurately describes, and can make predictions on, a particularly complex system or dynamic process. In the past, equations were obtained either from observation (experiment) or derived analytically or from other equations. To some, the insights gained by "computer experiments" do not carry the same intellectual satisfaction as a traditional scientific discovery or break-through, but they have become the main and often only reliable tool for analyzing highly complex processes (Israelachvili 2011).

The subject has become so broad that a tendency has developed for different disciplines to adopt their own concepts and terminology, to emphasize quite different aspects of interactions that are essentially the same or even make "discoveries" that are well known in other fields. For example, in chemistry and biology, emphasis is placed almost entirely on the short-range force fields around atoms and molecules, rarely extending more than one or two atomic distances. The language of the present-day molecular biologist is full of terms such as molecular packing, specific binding sites, lock and key mechanisms, and docking, all of which are essentially short range. In the different though closely related area of colloid science, the emphasis is quite often on the long-range forces, which may determine whether two surfaces or particles are able to get close enough in the first place before they can interact via the types of short-range forces mentioned above. In this discipline one is more likely to hear about electric double-layer forces, van der Waals forces, steric polymer interactions, and so forth - all of which are essentially long ranged. But this situation is also changing: pressured by the sheer complexity and all-embracing nature of our subject, the barriers between certain areas of physics, chemistry biology, and engineering are rapidly disappearing, with a "return" to a state where science, engineering, and mathematics are no longer regarded as independent disciples (Israelachvili 2011). A sample illustration of this statement is given by Prausnitz, Lichtenthaler and de Azevedo in their reference book (Prausnitz et al. 1999) where they

argued that ‘‘careful study has shown that for accurate work both physical and chemical forces must be taken into account. However, the dividing line between physical and chemical forces cannot easily be determined with rigor and as a result, it is often necessary to make an essentially arbitrary decision on where that line is drawn’’.

Since the need for predictions of thermodynamic properties in the fluid phase is so strong, several more pragmatic physical and organic chemists, and also some chemical engineers, have tried other pathways for property predictions, especially in the liquid state (Prausnitz et al. 1999). A brief (but not exhaustive) overview of these tools is given below.

I.5. Existing tools

As discussed above, since the early years of physical chemistry, thousands of articles have been written in an effort to understand the behaviour of mixed fluids. To construct a theory of liquid mixture, two kinds of information are required: the structure of liquids (the way the molecules in a liquid are arranged in space) and the intermolecular forces between like and unlike molecules. Unfortunately, information of either kind is inadequate and, as a result, all theories must make simplifying assumptions to overcome this disadvantage (Prausnitz et al. 1999). More theoretical work has been concerned with mixtures of liquids whose molecules are non-polar and spherical: for example, the regular solution theory of Scatchard and Hildebrand frequently provides a good approximation for mixtures of hydrocarbons (Prausnitz et al. 1999).

Since the liquid state is in some sense intermediate between the crystalline state and the gaseous state, it follows that there are two types of approach of a theory of liquids (Prausnitz et al. 1999).

The first considers liquids to be gas-like; a liquid is pictured as a dense and highly nonideal gas whose properties can be described by some equation of state (EOS); that of van der Waals is the best known example. An equation of state description of pure liquids can readily be extended to liquid mixtures as was done by van der Waals and by some of his disciples like van Laar and later by many others. This approach enables to take into account the properties of the pure components that are manifested in a P-V-T behaviour or in the equation of state. In general pure fluids have different free volumes, *i.e.* different degrees of expansion. Thus if they are mixed that difference contributes to the excess function. Differences in free volumes must be taken into account, especially for mixtures of liquids whose molecules differ greatly in size. For example in a solution of a polymer in a chemically similar solvent of low molecular weight, there is little dissimilarity in intermolecular interactions but the free volume dissimilarity is significant. Advances in statistical thermodynamics have brought to the forefront tangent-sphere models of chain-like fluids. These

models are generally suitable for polymers which are pictorially treated as freely-jointed tangent-spheres where non-bonded spheres interact through a specified intermolecular potential. This family of method includes the statistical-associated-fluid theory (SAFT) developed by Chapman and co-workers (Chapman et al. 1989, 1990), and its recent improvements (Prausnitz et al. 1999).

The second approach considers a liquid to be solid-like, in a quasicrystalline state, where the molecules do not translate fully in a chaotic manner as in a gas, but where each molecule tends to stay in a small region, a more or less fixed position in space about which it vibrates back and forth. The quasicrystalline picture of the liquid state supposes molecules to sit in a regular array in space, called a lattice, and therefore liquid and liquid mixture models based on this simplified picture are called lattice models. These theories are described in detail elsewhere (Prausnitz et al. 1999 and references therein) and their proper study requires familiarity with the methods of statistical mechanics. Since the lattice theory of liquids assumes that molecules are confined to lattice positions (sometimes called cages), calculated entropies (disorder) are low by what is called the "communal entropy". While this is a serious deficiency, it tends to cancel when lattice theory is used to calculate excess properties of liquid mixtures which are the focus of the present work. Molecular considerations suggest that deviations from ideal behaviour in liquid solutions are due primarily to the following effects: First, forces of attraction between unlike molecules are quantitatively different from those between like molecules, giving rise to a non-vanishing enthalpy of mixing; second, if the unlike molecules differ significantly in size or shape, the molecular arrangement in the mixture may be appreciably different from that for the pure liquids, giving rise to a nonideal entropy of mixing; and finally, in a binary mixture, if the forces of attraction between one of three possible pair interactions are very much stronger (or very much weaker) than those of the other two, there are certain preferred orientations of the molecules in the mixture that, in extreme cases, may induce thermodynamic instability and demixing (incomplete miscibility) (Prausnitz et al. 1999).

The so-called theories of solutions express the excess properties of mixture as function of the intermolecular forces and of the fundamental structure of liquid (ordering and configuration of molecules in space). In this approach, it is important to have predictive methods, which describe the energetic contributions between molecules in terms of activity coefficients. These methods are available and present advantages and drawbacks or limitations for their application in process design. For a better interpretation and correlation of the thermodynamic behaviour of aqueous solutions, Lei and his collaborators have reviewed the almost of these "predictive molecular thermodynamic models" (Lei et al. 2008). In this context, they stated that it is necessary to provide molecular structures or physical properties of pure components in the mixture in order to describe phase equilibria properties. To describe the main characteristics of non-electrolytes liquid mixtures,

at thermodynamic equilibrium conditions, this simpler approach is generally sufficient. However, as discussed above, food and biological systems are generally treated as multicomponent aqueous mixtures that generally contain simultaneously water, salts or electrolytes, sugars. When electrolytes are considered, the system is characterized by the presence of both molecular species and ionic species, resulting in three different types of interactions: ion-ion, molecule-molecule and ion-molecule. Ion-ion interactions are governed by electrostatic forces between ions that have a much longer range than other intermolecular forces. Molecule-molecule and ion-molecule interaction forces are known to be short-range in nature. The excess Gibbs energy of systems containing electrolytes can be considered as the sum of two terms, one related to long-range (LR) forces between ions and the other to short-range (SR) forces between all the species.

In this section, a short overview of the main equations of state (EOS) is given, as well as their semi empirical extension to more complicated molecules or mixtures. Then the different theories and existing models for the representation of thermodynamic properties of solutions to take into account respectively Short Range (SR) and Long Range (LR) effects are also discussed. The SR predictive molecular thermodynamic models can be classified into two categories:

- the models with relation to experimental data; in which the interaction parameters are correlated from experimental data and prediction of thermodynamic properties is made on the basis of existing parameters for each constituent of the molecules present in the mixture; this category includes semi-predictive models and group contributions methods;
- and the models with no relation to experimental data (or called *a priori* predictive models), in which only atom-specific parameters are required and prediction of thermodynamic properties is made on the basis of unimolecular quantum chemical calculations that provide the necessary information for the evaluation of molecular interactions in liquids.

I.5.1. Equations of state and their extensions to liquid systems

Equations of state are ubiquitous in petrochemistry and chemistry, and play a key role in the conception of separation and synthesis units. These equations relate the temperature, pressure, amount of substances and the volume occupied by a system. These equations are expressed in the form: $E(T,P,V,x) = 0$, where x denotes the composition matrix (moles number), and $x=1$ for a pure compound.

The use of EOS for the estimation of thermodynamics properties of pure compounds and mixtures has been investigated by different scientists since a century, as discussed above. Indeed, since van der Waals had proposed his well-known cubic equations of state in 1873, numerous papers were published with a special focus on gas phase properties. This is mainly due in the one hand to the availability of experimental data concerning the P, V, T parameters, in the other hand, to theoretical

considerations about the constitution of atoms and molecules as well as the fact that interactions are relatively weak in the case of gases (Prausnitz et al. 1999).

The origin of the interest in equations of state is their aptitudes, as though simplified, to allow the calculation of a wide number of thermodynamic properties of pure compounds, particularly in phase equilibria. These calculations are performed in a coherent way and the founded results verify the general law linking the temperature, pressure, and other thermodynamic functions. In order to use these equations efficiently, it is important to define properly the studied compounds, the experimental or calculations conditions and the investigated properties. A non-exhaustive overview of the main equations of state encountered in the thermodynamic process engineering softwares like "Simulis Thermodynamics" commercialized by ProSim S.A., is given in Table I-5.

Table I-5 : Several equations of state from literature (from Prausnitz et al. 1999)

	Equations	Remarks
Ideal gas law	$PV = nRT$	Where n denotes the mole number, and R is a constant.
Nonideal gas law	$Pv = ZRT$	Z: Compressibility factor which is a measure of nonideality
Virial equation	$Z = \frac{Pv}{RT} = 1 + \frac{B}{v} + \frac{C}{v^2} + ...$	B, C,..., denote respectively the second, third,..., coefficient of the Viriel
van der Waals (cubic EOS)	$P = \frac{RT}{v-b} - \frac{a}{v^2}$	a and b are coefficients determined at the critical point
General form of the cubic equations of state	$P = \frac{RT}{v-b} - \frac{a(T)}{v^2 + ubv + wb^2}$	a and b are coefficients determined at the critical point. ➢ u=0, w=0 for the van der Waals EOS ➢ u=1, w=0 for the Soave EOS or the Redlich-Kwong EOS ➢ u=2, w=-1 for the Peng-Robinson EOS
Dieterici EOS	$P = \frac{RT}{v-b}\exp\left(-\frac{a}{RTv}\right)$	a and b are coefficients determined at the critical point
Beattie-Bridgeman	$P = \frac{RT}{v} + \frac{\beta}{v^2} + \frac{\gamma}{v^3} + \frac{\delta}{v^4}$ Where : $\begin{cases} \beta = RTB_0 - A_0 - \frac{Rc}{T^2} \\ \gamma = RTB_0 b - A_0 a - \frac{RcB_0}{T^2} \\ \delta = \frac{RTB_0 bc}{T^2} \end{cases}$	All of the parameters coefficients are determined from critical properties

However, equations of state are not able to satisfy all the needs of process designers. Indeed, to apply EOS, it is important to:

- Define the application domain of these equations (what compounds, what properties and what domains of pressure and temperature). The main problems encountered in these cases, are usually the extrapolation of these equations to temperatures higher than the critical temperature, or to the lowest temperatures.

- Get a large set of experimental data about the critical properties of pure compounds to apply EOS, as though these data are not always available in the literature, and are hypothetical for the almost of the compounds like sugars which do never exist in the liquid state since they are denatured by the fusion so that there is no experimental way to determine the critical point of these compounds.
- Overcome the complexity of the mathematical expression of these EOS that could induce several inconsistencies which could limit the studied domain of the chosen equation of state, or its extrapolation.

The utility of EOS is greatly increased when they can be made applicable to mixtures. Two approaches are generally used for this purpose. In the first approach, the parameters of the EOS of the mixture are evaluated from its critical properties with the same equations that are applicable to pure substances. Critical properties of random mixtures rarely are known from experiment and are not easy to estimate, however it is fortunate that superior methods can be based on certain ***pseudocritical properties*** of which the simplest are mol fraction weighted sums of the corresponding properties of the components of the mixture. In the other approach, a ***direct combination*** of individual parameters on the basis of composition is usually preferred to the use of pseudocritical properties in many EOS.

For instance, in the van der Waals equation, this can be accomplished by expressing the parameters of the mixture in terms of the composition and the parameters of the pure substances, using the ***combining rules*** proposed by Lorentz and Berthelot. The latter, also known as Lorentz-Berthelot rules, are:

$$\begin{cases} \sqrt{a} = \sum_{i=1}^{N_c} y_i \sqrt{a_i} \\ b = \sum_{i=1}^{N_c} y_i b_i \end{cases} \quad \text{(I-38)}$$

Where, y_i denotes the molar fraction of specie i in the gas phase. Their use in the statistical mechanics of mixtures is illustrated by Rowlinson and Swinton (1982).

The most recent researches undertaken in order to adapt EOS to the treatment of liquids mixtures, are generally focused on the treatment of complex systems like polymers (Lei et al. 2008; Vega and Jackson 2011; and references therein). In this context, the Statistical Associating Fluid Theory (SAFT), that is an equation of state based on statistical mechanics for use in an engineering context, was first published in Fluid Phase Equilibria over 20 years ago (Chapman et al. 1989, 1990). The SAFT approach was intended, from the very beginning, to be used to describe the thermodynamic properties and phase behaviour of complex fluids (associating systems, oligomers and polymers) for which the ubiquitous and popular cubic equations of state fail to provide an adequate description; in some quarters, its mathematical complexity when compared with simpler

equations of state was considered to be a handicap for a standard implementation. The first developments of SAFT focused on the application of the approach to a wide variety of fluids and fluid mixtures (Huang and Radosz 1990, 1991) or on simplifying the equation of state to make it more tractable. These simplified versions allowed for straightforward programming and computation, though often at the price of a poorer description of the experimental systems of interest. In the late 1990s and early in the twenty-first century, efforts were directed towards refining the theoretical description, essentially improving the accuracy of the term used for the reference fluid.

This led to the different incarnations of the equation of state which are in common use today such as soft-SAFT (Blas and Vega 1998), SAFT-VR (Gil-Villegas et al. 1997) and PC-SAFT (Gross and Sadowski 2002), as well as the related cubic plus association CPA equation of state (Kontogeorgis et al. 2006, Kontogeorgis and Folas 2009). The key to the popularity of SAFT-based equations of state from a scientific point of view is their firm statistical-mechanics basis, which allows for a rigorous physical interpretation of the contributions due to the various intermolecular interactions. This provides a framework from which the effects on the thermodynamic properties of the various molecular features can be separated and quantified. Today, the success of SAFT in providing a quantitative description of the fluid phase equilibria of complex mixtures, the extension of the approach to explore other regions of the phase diagram, and the wide range of applications for which it has been used are reflected in the numerous papers published since the original work.

Vega and Jackson have reported the most recent successful extensions of the approach and applications to challenging systems: electrolytes and ionic liquids; aqueous systems; polyelectrolytes and amino acids; inhomogeneous properties; the critical region, group contribution methods; transport and derivative properties; and process modelling (Vega and Jackson 2011). They also discussed the need for an extension of the theory from its basis on molecular fluids with segments interacting through the standard Lennard-Jones or square-well potential to a form with a variable repulsive (and/or attractive) interaction such as the Mie potential.

However, even if the equations of states concepts well represent the behaviour of the non-polar or compounds with low polarities, in a wide range of pressure, these equations are not adapted for the treatment of polar compounds in which the interactions in the liquid state are important (chemical or electrostatic interactions). The latter case is the most encountered in biological and food systems, where the liquid phase is preponderant and mainly constituted by water. Furthermore, due to the difficulties encountered during the application of these equations, this methodology is out of the scope of this study where the treatment of the thermodynamic systems will be handled by the use of excess Gibbs free energy models.

I.5.2. Models suitable for the treatment of short range effects

I.5.2.1. SR models based on the physical theory of solution

The physical theory of solution is mainly based on the lattice model of solutions and its improvements in order to take into account a more realistic treatment of interactions between liquids. For a better understanding of such theoretical models, it is interesting to have a general idea about the lattice models, the two-liquid theory, before presenting the most used models based on these theoretical concepts (Prausnitz et al. 1999).

I.5.2.1.1. Lattice model and two-liquid theory

The lattice theory of liquids assumes that molecules are confined to lattice positions (sometimes called cages). Let us consider a mixture of two simple liquids X_1 and X_2. Molecules of types X_1 and X_2 are small and spherically symmetric and the ratio of their sizes is close to unity. We suppose that the arrangement of the molecules in each pure liquid is that of a regular array as indicated in Figure I-4; all the molecules are situated on lattice points that are equidistant from one another (Prausnitz et al. 1999).

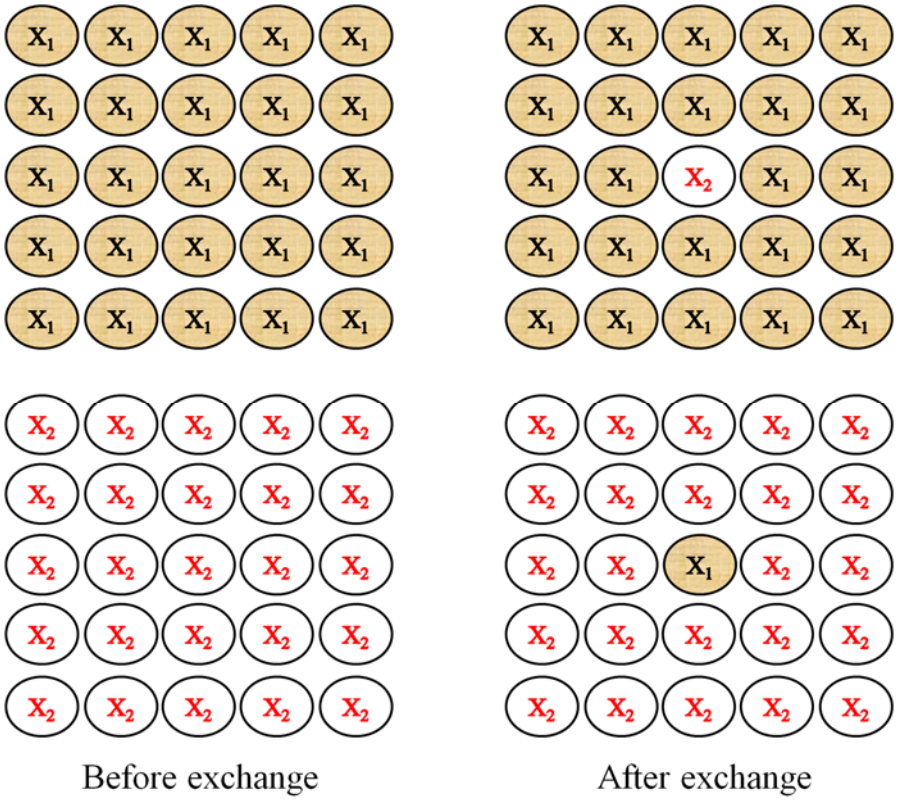


Figure I-4 : Physical significance of interchange energy.
The energy absorbed in the process above is 2w (Adapted from Prausnitz et al. 1999).

Molecular motion is limited to vibrations about equilibrium positions and is not affected by the mixing process. We suppose further that for a fixed temperature, the lattice spacings for the two

pure liquids and for the mixture are the same, independent of composition (*i.e* , $v^E = 0$). To derive an expression for the potential energy of a liquid, pure or mixed, we assume that the potential energy is pairwise additive for all molecular pairs and that only nearest neighbours need to be considered in the summation. This means that the potential energy of a large number of molecules sitting on a lattice is given by the sum of the potential energies of all pairs of molecules that are located immediately next to one another. For uncharged nonpolar molecules, intermolecular forces are short range and therefore we assume in this simplified discussion that we can neglect contributions to the total potential energy from pairs that are not nearest neighbours. Let us consider that each of N_1 molecules of type X_1 and N_2 molecules of type X_2 has z nearest (touching) neighbours. z is called the ***coordination number*** and may have a value between 6 and 12 depending on the type of packing, *i.e.*, the way in which the molecules are arranged in three-dimensional space; empirically, for typical liquids at ordinary conditions, z is close to 10 (Prausnitz et al. 1999). The total number of nearest neighbours is $z/2\left(N_1 + N_2\right)$ and there are three types of nearest neighbours: X_1-X_1, X_2-X_2 and X_1-X_2. Let N_{11} be the number of nearest-neighbour pairs of type 11, N_{22} that of type 22, and N_{12} that of type 12. These three numbers are not independent; they are restricted by the following conservation equations:

$$\begin{cases} zN_1 = 2N_{11} + N_{12} \\ zN_2 = 2N_{22} + N_{12} \end{cases} \tag{I-39}$$

The total potential energy of the lattice U_t is given by:

$$U_t = N_{11}\Gamma_{11} + N_{22}\Gamma_{22} + N_{12}\Gamma_{12} \tag{I-40}$$

where, Γ_{11} is the potential energy of a X_1-X_1 pair, Γ_{22} that of a X_2-X_2 pair, and Γ_{12} that of a X_1-X_2 pair. Substitution of N_{11} and N_{22} from equation I-39 gives:

$$U_t = \frac{z}{2}N_1\Gamma_{11} + \frac{z}{2}N_2\Gamma_{22} + \frac{w}{2}N_{12} \tag{I-41}$$

where w denotes the interchange energy, and is defined by

$$w \equiv z\left[\Gamma_{12} - \frac{1}{2}\left(\Gamma_{11} + \Gamma_{22}\right)\right] \tag{I-42}$$

After thermodynamic properties calculations in the canonical ensemble, it has been demonstrated elsewhere (Prausnitz et al. 1999) that for a regular solution (*i.e.* $v^E = s^E = 0$) the excess Gibbs energy, the excess Helmholtz energy, the excess enthalpy (or enthalpy of mixing) and the excess energy (or energy of mixing), are all equal:

$$g^E = a^E = h^E = u^E = \mathcal{N}wx_1x_2 \tag{I-43}$$

where $\mathcal{N}$ is the Avogadro's constant. The derived activity coefficients are given by:

$$\begin{cases} \ln \gamma_1 = \dfrac{w}{kT} x_2{}^2 \\ \ln \gamma_2 = \dfrac{w}{kT} x_1{}^2 \end{cases} \tag{I-44}$$

Because the interchange energy *w* is related to the potential energies, it should be possible to obtain a numerical value for *w* from information on potential functions. Various attempts to do this have been reported (Prausnitz et al. 1999, and references therein).

One of the important assumptions made previously in the lattice concept was that when two molecules of two components are mixed, the arrangement of the molecules is completely random; *i.e.*, the molecules have no tendency to segregate either with their own kind or with the other kind of molecules. In a completely random mixture, a given molecule shows no preference in the choice of its neighbours (Prausnitz et al. 1999).

Because intermolecular forces operate between molecules, a completely random mixture in a two-component system of equal sized molecules can only result if these forces are the same for all three possible molecular pairs. In that event, however, there would also be no energy change upon mixing. Strictly, then, only an ideal mixture can be completely random (Prausnitz et al. 1999).

In the lattice theory (for *w* independent of T), entropy is a measure of randomness; the entropy of mixing for a completely random mixture is always larger than that of a mixture that is incompletely random, regardless of whether non-randomness is due to preferential formation of X_1-X_2 or X_1-X_1 (or X_2-X_2) pairs. Excess entropy due to ordering (*i.e.*, non-randomness) is always negative (Prausnitz et al. 1999).

Guggenheim (1966) has constructed a lattice theory for molecules of equal size that form mixtures that are not necessarily random. This theory is not rigorous but utilizes a simplification known as the quasichemical approximation. The excess Gibbs energy based on the quasichemical approximation (given in equation I-45) is not very different from equation I-43 based on the assumption of random mixing (Prausnitz et al. 1999).

$$\frac{g^E}{RT} = \left(\frac{w}{kT}\right) x_1 x_2 \left[1 - \frac{1}{2}\left(\frac{2w}{zkT}\right) x_1 x_2 + \ldots\right] \tag{I-45}$$

$$\frac{h^E}{RT} = \left(\frac{w}{kT}\right) x_1 x_2 \left[1 - \left(\frac{2w}{zkT}\right) x_1 x_2 + \ldots\right] \tag{I-46}$$

$$\frac{s^E}{R} = -\left(\frac{w}{kT}\right) x_1 x_2 \left[\frac{1}{2}\left(\frac{2w}{zkT}\right) x_1 x_2 + \ldots\right] \tag{I-47}$$

These equations show that in the quasichemical approximation, all excess functions are symmetric in x, however excess Gibbs energy is no longer equal to excess enthalpy and excess entropy is no longer zero. Only, in the limit as $\left(\frac{2w}{zkT}\right) \to 0$, we obtain $g^E \to h^E$ and $s^E \to 0$ as expected.

Extension of a corresponding-states theory to mixtures is based on the fundamental idea that a mixture can be considered to be a hypothetical pure fluid whose characteristic molecular size and potential energy are composition averages of the characteristic sizes and energies of the mixture's components (one-fluid theory). In macroscopic terms, effective critical properties (pseudocritical properties) are composition averages of the component critical properties. However, this fundamental idea is not limited to one hypothetical pure fluid; it can be extended to include more than one hypothetical fluid, leading to m-fluid theories. These theories use as a reference a suitable (usually mole fraction) average of the properties of the m hypothetical pure fluids. For example, two-fluid theories, illustrated in Figure I-5, use two pure reference fluids (Prausnitz et al. 1999).

For simple mixtures, one-fluid and two-fluid theories give similar results when compared with experiment. Because one-fluid theory is easier to use, it is usually preferred. However, two-fluid theory provides a useful point of departure for deriving semi-empirical equations to represent thermodynamic excess functions for highly non-ideal mixtures (Prausnitz et al. 1999).

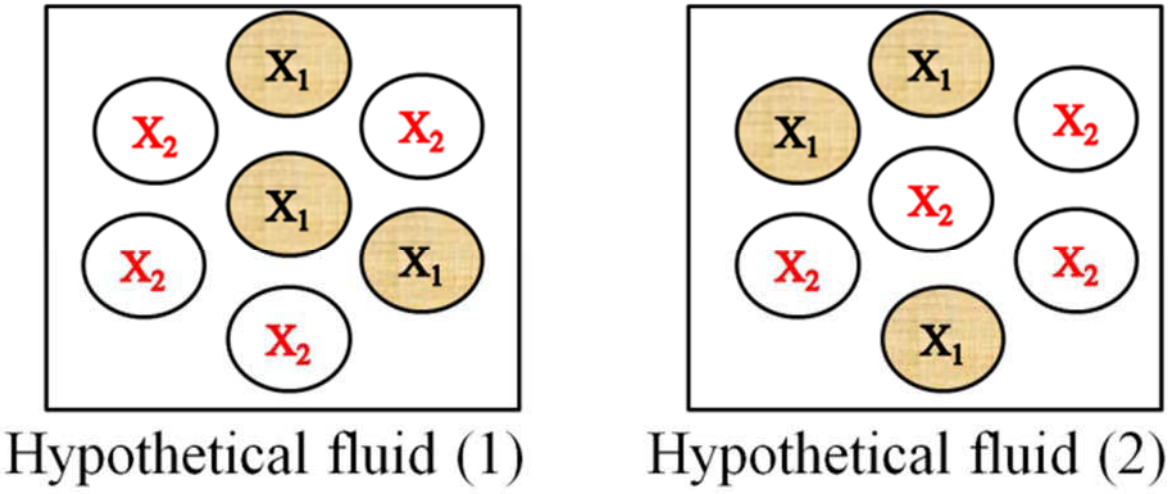


Figure I-5 : Essential idea of the two-fluid theory of binary mixtures.

Hypothetical fluid (1) has a molecule X_1 at the centre. Hypothetical fluid (2) has a molecule X_2 at the centre. (Adapted from Prausnitz et al. 1999)

I.5.2.1.2. Semi-predictive activity coefficients models

A short description of the main semi-predictive models (*i.e.* equations proposed for correlating activity coefficients with composition and to a lesser extent with temperature) that are used in process engineering, is given in Table I-6, for binary systems. In this table, the expression of the excess Gibbs energy, the parameters that are generally derived from experimental data that they

must fit accurately and the respective expressions of the activity coefficients of the solvent (1) and the solute (2) are given (Prausnitz et al. 1999).

One should notice that the two-suffix Margules equations are of the same form as the equations calculated in the lattice model even if the interchange energy *w* which has a physical sense was replaced by a mathematical parameter in the two-suffix Margules equations.

The equations for binary mixtures given in Table I-6, have two parameters each, except the symmetrical Margules equation which has one parameter, and the NRTL and four-suffix Margules equations that have both three parameters. Four of these equations are applicable to multicomponent mixtures without any parameters beyond those of the constituent pairs. These multicomponent parameters are listed in Table I-7.

The semi-empirical methods are generally used to establish a computerized technique for calculating multicomponent vapour-liquid equilibria (VLE) as required, for example, in the design of separation equipment such as distillation columns. As though these equations do not have a precise theoretical basis, they appear to be of a form that is particularly useful for solutions containing one or more polar components (Prausnitz et al. 1999).

However, in the Wilson, NRTL, and UNIQUAC models, the experimental data must be given to correlate the binary-pair parameters. Therefore, they are not purely predictive.

Table I-6 : Some models for the Excess Gibbs Energy and subsequent activity coefficients for binary systems (from Prausnitz et al. 1999).

Names	Excess Gibbs energy g^E	Binary parameters	Activity coefficients
Scatchard-Hildebrand	$g^E = \dfrac{(\delta_1 - \delta_2)^2}{\left(\dfrac{1}{v_1 x_1} + \dfrac{1}{v_2 x_2}\right)}$	$\delta_1,\ \delta_2$	$RT \ln \gamma_1 = V_1 (1-\Phi_1)^2 (\delta_1 - \delta_2)^2$ $RT \ln \gamma_2 = V_2 \Phi_1^{\,2} (\delta_1 - \delta_2)^2$ where : $\Phi_1 = \dfrac{v_1 x_1}{v_1 x_1 + v_2 x_2}$
Two-suffix (or symmetrical) Margules	$g^E = A x_1 x_2$	A	$RT \ln \gamma_1 = A x_2^{\,2}$ $RT \ln \gamma_2 = A x_1^{\,2}$
Three-suffix Margules	$g^E = x_1 x_2 \left[A + B(x_1 - x_2) \right]$	A, B	$RT \ln \gamma_1 = (A + 3B) x_2^{\,2} - 4B x_2^{\,3}$ $RT \ln \gamma_2 = (A - 3B) x_1^{\,2} - 4B x_1^{\,3}$
van Laar	$g^E = \dfrac{A x_1 x_2}{x_1 \left(\dfrac{A}{B}\right) + x_2}$	A, B	$RT \ln \gamma_1 = A \left(1 + \dfrac{A x_1}{B x_2}\right)^{-2}$ $RT \ln \gamma_2 = B \left(1 + \dfrac{B x_2}{A x_1}\right)^{-2}$
Four-suffix Margules	$g^E = x_1 x_2 \left[A + B(x_1 - x_2) + C(x_1 - x_2)^2 \right]$	A, B, C	$RT \ln \gamma_1 = (A + 3B + 5C) x_2^{\,2} - 4(B + 4C) x_2^{\,3} + 12C x_2^{\,4}$ $RT \ln \gamma_2 = (A - 3B + 5C) x_1^{\,2} + 4(B - 4C) x_1^{\,3} + 12C x_1^{\,4}$

Names	Excess Gibbs energy g^E	Binary parameters	Activity coefficients
Wilson	$\frac{g^E}{RT} = -x_1 \ln\left(x_1 + \Lambda_{12}x_2\right) - x_2 \ln\left(x_2 + \Lambda_{21}x_1\right)$	$\Lambda_{12}, \Lambda_{21}$	$\ln\gamma_1 = -\ln\left(x_1 + \Lambda_{12}x_2\right) + x_2\left(\frac{\Lambda_{12}}{x_1 + \Lambda_{12}x_2} - \frac{\Lambda_{21}}{\Lambda_{21}x_1 + x_2}\right)$ $\ln\gamma_2 = -\ln\left(x_2 + \Lambda_{21}x_1\right) - x_1\left(\frac{\Lambda_{12}}{x_1 + \Lambda_{12}x_2} - \frac{\Lambda_{21}}{\Lambda_{21}x_1 + x_2}\right)$
Non random two liquid (NRTL)	$\frac{g^E}{RT} = x_1x_2\left(\frac{\tau_{21}G_{21}}{x_1 + x_2G_{21}} + \frac{\tau_{12}G_{12}}{x_2 + x_1G_{12}}\right)$ Where $\tau_{12} = \frac{\Delta g_{12}}{RT} \quad \tau_{21} = \frac{\Delta g_{21}}{RT}$ $\ln G_{12} = -\alpha_{12}\tau_{12} \quad \ln G_{21} = -\alpha_{12}\tau_{21}$	$\Delta g_{12}, \Delta g_{21}, \alpha_{12}$	$\ln\gamma_1 = x_2{}^2\left(\tau_{21}\left(\frac{G_{21}}{x_1 + x_2G_{21}}\right)^2 + \frac{\tau_{12}G_{12}}{\left(x_2 + x_1G_{12}\right)^2}\right)$ $\ln\gamma_2 = x_1{}^2\left(\tau_{12}\left(\frac{G_{12}}{x_2 + x_1G_{12}}\right)^2 + \frac{\tau_{21}G_{21}}{\left(x_1 + x_2G_{21}\right)^2}\right)$
Universal quasi chemical (UNIQUAC)	$g^E = g^E(\text{combinatorial}) + g^E(\text{residual})$ $\frac{g^E(\text{combinatorial})}{RT} = x_1 \ln\frac{\Phi_1}{x_1} + x_2 \ln\frac{\Phi_2}{x_2} + \frac{z}{2}\left(q_1x_1 \ln\frac{\theta_1}{\Phi_1} + q_2x_2 \ln\frac{\theta_2}{\Phi_2}\right)$ $\frac{g^E(\text{residual})}{RT} = -q_1x_1 \ln\left[\theta_1 + \theta_2\tau_{21}\right] - q_2x_2 \ln\left[\theta_2 + \theta_1\tau_{12}\right]$ $\Phi_1 = \frac{x_1r_1}{x_1r_1 + x_2r_2} \quad \theta_1 = \frac{x_1q_1}{x_1q_1 + x_2q_2}$ $\ln\tau_{21} = -\frac{\Delta u_{21}}{RT} \quad \ln\tau_{12} = -\frac{\Delta u_{12}}{RT}$ where r and q are pure-component parameters and coordination number $z = 10$	$\Delta u_{12}, \Delta u_{21}$	$\ln\gamma_i = \ln\frac{\Phi_i}{x_i} + \frac{z}{2}q_i \ln\frac{\theta_i}{\Phi_i} + \Phi_j\left(l_i - \frac{r_i}{r_j}l_j\right) - q_i \ln\left(\theta_i + \theta_j\tau_{ji}\right)$ $+ \theta_jq_i\left(\frac{\tau_{ji}}{\theta_i + \theta_j\tau_{ji}} - \frac{\tau_{ij}}{\theta_j + \theta_i\tau_{ij}}\right)$ where $\begin{cases} i = 1, \ j = 2 \\ or \ i = 2, \ j = 1 \end{cases}$ and $\begin{cases} l_i = \frac{z}{2}\left(r_i - q_i\right) - \left(r_i - 1\right) \\ l_j = \frac{z}{2}\left(r_j - q_j\right) - \left(r_j - 1\right) \end{cases}$

Table I-7 : Some models for the Excess Gibbs Energy and subsequent activity coefficients for multicomponent systems (from Prausnitz et al. 1999).

Names	Excess Gibbs energy g^E	Parameters	Activity coefficients
Scatchard-Hildebrand	$g^E = \sum_i x_i v_i \left(\delta_i - \bar{\delta}\right)^2$	δ_i	$RT \ln \gamma_i = v_i \left(\delta_i - \sum_j \frac{x_j v_j \delta_j}{\sum_k x_k v_k} \right)^2$
Wilson	$\frac{g^E}{RT} = -\sum_i x_i \ln\left(\sum_j x_j \Lambda_{ij} \right)$	$\Lambda_{ij} = \frac{V_j^L}{V_i^L} \exp\left(\frac{\lambda_{ij}}{RT} \right)$ $\Lambda_{ii} = \Lambda_{jj} = 1$	$\ln \gamma_i = -\ln\left(\sum_{j=1}^m x_j \Lambda_{ij} \right) + 1 - \sum_{k=1}^m \left(\frac{x_k \Lambda_{ki}}{\sum_{j=1}^m x_j \Lambda_{kj}} \right)$
Non random two liquid (NRTL)	$\frac{g^E}{RT} = \sum_i x_i \left(\sum_j \frac{\tau_{ji} G_{ji} x_j}{\sum_k G_{ki} x_k} \right)$	$\tau_{ji} = \frac{\Delta g_{ji} - \Delta g_{ii}}{RT}$ $G_{ji} = \exp\left(-\alpha_{ji} \tau_{ji}\right)$ $\tau_{ii} = \tau_{jj} = 0$ $G_{ii} = G_{jj} = 1$	$\ln \gamma_i = \frac{\sum_{j=1}^m \tau_{ji} G_{ji} x_j}{\sum_{l=1}^m G_{li} x_l} + \sum_{j=1}^m \frac{x_j G_{ij}}{\sum_{l=1}^m G_{lj} x_l} \left(\tau_{ij} - \frac{\sum_{n=1}^m x_n \tau_{nj} G_{nj}}{\sum_{l=1}^m G_{li} x_l} \right)$
Universal quasi chemical (UNIQUAC)	$\frac{g^E(\text{combinatorial})}{RT} = \sum_i x_i \ln \frac{\Phi_i}{x_i} + \frac{z}{2} \sum_i q_i x_i \ln \frac{\theta_i}{\Phi_i}$ $\frac{g^E(\text{residual})}{RT} = -\sum_i q_i x_i \ln\left(\sum_{j=1}^m \theta_j \tau_{ji} \right)$ where r and q are pure-component parameters and coordination number $z = 10$	$\tau_{ij} = \exp\left(-\frac{\Delta u_{ji} - \Delta u_{ii}}{RT} \right)$ $\tau_{ii} = \tau_{jj} = 1$ $l_i = \frac{z}{2}(r_i - q_i) - (r_i - 1)$	$\ln \gamma_i = \ln \gamma_i^c + \ln \gamma_i^R$ $\ln \gamma_i^c = \ln \frac{\Phi_i}{x_i} + \frac{z}{2} q_i \ln \frac{\theta_i}{\Phi_i} + l_i - \frac{\Phi_i}{x_i} \sum_j x_j l_j$ $\ln \gamma_i^R = q_i \left[1 - \ln\left(\sum_{j=1}^m \theta_j \tau_{ji} \right) - \sum_{j=1}^m \frac{\theta_j \tau_{ij}}{\sum_{k=1}^m \theta_k \tau_{kj}} \right]$

I.5.2.1.3. Groups contributions models (GCMs)

For engineering purposes, it is often necessary to make some estimate of activity coefficients for mixtures where only fragmentary data or no data at all, are available. For vapour-liquid equilibria, such estimates can be made using a group-contribution method. In this approach, a molecule is divided (somewhat arbitrarily) into functional groups, as illustrated in Table I-8 and Figure I-6 in the case of glycerol and several sugars.

Table I-8 : Decomposition in functional groups of the glycerol and several sugars molecules.

		Number of functional groups				
Molecules	Formulae	Alkanes CH_2	Alkanes CH	Alkanes C	Alcohols OH	Ethers CH-O-
Glucose	$C_6H_{12}O_6$	1	4	0	5	1
Fructose	$C_6H_{12}O_6$	2	2	1	5	1
Maltose	$C_{12}H_{22}O_{11}$	2	7	0	8	3
Sucrose	$C_{12}H_{22}O_{11}$	3	5	1	8	3
Glycerol	$C_3H_8O_3$	2	1	0	3	0

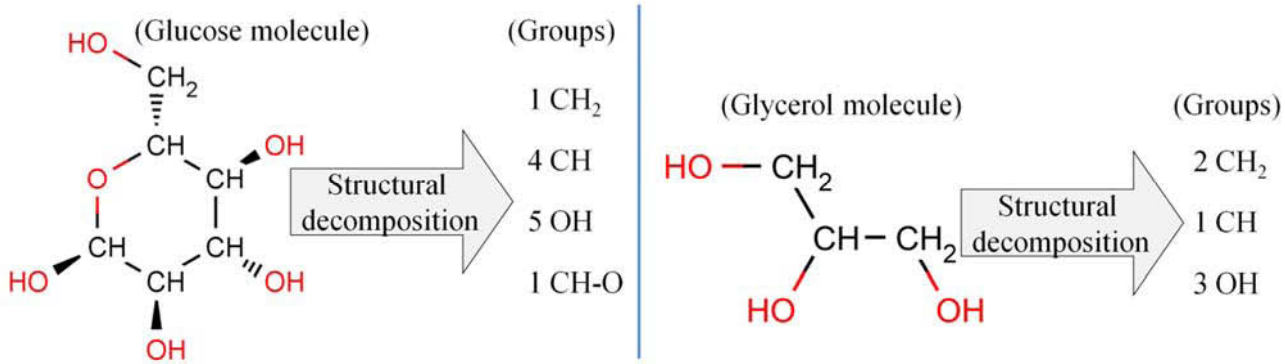


Figure I-6 : Illustration of the group-contribution principle in the cases of glucose and glycerol molecules.

Molecule-molecule interactions are considered to be properly weighted sums of group-group interactions. Therefore, for a multifunctional component in a multicomponent system, GCMs assume that each functional group behaves in a manner independent of the molecule in which it appears. Once quantitative information on the necessary group-group interactions is obtained from reduction of experimental data for binary systems, it is then possible to calculate molecule-molecule interactions (and therefore phase equilibria) for molecular pairs where no experimental data are available. The fundamental advantage of this procedure is that when attention is directed to typical mixtures of non-electrolytes, the number of possible distinct functional groups is much smaller than the number of

distinct molecules or, more directly, the number of distinct group-group interactions is very much smaller that the number of possible distinct molecule-molecule interactions (Prausnitz et al. 1999).

The main group contribution models are the ASOG model (Kojima and Togichi 1979) and the UNIFAC model (Fredenslund et al. 1975; Larsen et al. 1987). These methods are daily used in chemical industry and new developments continue to be made in the case of UNIFAC improving the range and accuracy of the methods.

The earliest group-contribution method for predicting activity coefficients is the ASOG (analytical solution of groups) model that was proposed by Derr and Deal and used the Wilson model to represent the group activity coefficients (Derr and Deal 1969, 1973, Prausnitz et al. 1999). The ASOG model treats a solution as a mixture of various structural groups, as opposed to a mixture of two or more distinct compounds. This treatment has the inherent advantage that there are significantly fewer structural groups than compounds; hence, fewer parameters are needed for a large number of binary and multicomponent mixtures. However, by far, the group interaction parameters for only 43 groups and 341 group pairs are available in the parameter matrix of ASOG model, which frequently cannot meet the requirement for the systems in which the group interaction parameters are missing (Kojima and Togichi 1979).

An original UNIFAC model that combines the functional group concept with a model for activity coefficients based on an extension of the quasi chemical theory of liquid mixtures (UNIQUAC) was proposed by Fredenslund et al. in 1975. This model can be applied at infinite dilution and finite concentrations and was the most widely used before several revisions and extensions were developed. The UNIFAC model is especially suitable for simple solvent molecules with low molecular weight. No binary data are required in the UNIFAC model. That is the main reason for which it is a widely used predictive model. This model is currently very popular and can be used to predict the liquid phase activity coefficient of binary or multicomponent systems, even when the experimental phase equilibrium data are unavailable. It has several advantages over the Wilson, NRTL, and UNIQUAC equations:

- Size and binary interaction parameters are available for a wide range of types of function groups (more than 100 function groups);
- Extensive comparisons with experimental data are available;
- It is an open system, and more function groups and more parameters will be filled in the UNIFAC list in the future.
- With the view to multicomponent mixtures, the UNIFAC model is more advantageous than the Wilson, NRTL, and UNIQUAC models in saving the measurement time since experimental measurements of vapour-liquid phase equilibrium are very time-consuming and therefore expensive. For example, if measurements are performed for a 10-component system at just one constant pressure (*e.g.*, atmospheric pressure) in 10% mole

steps and an average number of 10 data points can be experimentally determined daily, the measurements (in total, 92 378 data points) will take more than 37 years.

These are several reasons illustrating why the UNIFAC model was very popular and desirable in the synthesis, design and optimization of separation processes over the past few years.

The UNIFAC model is still developing, and by far, there are three versions for solvent-solvent systems with low molecular weights. In the UNIFAC method, the activity coefficient is expressed as functions of composition and temperature. The model has a combinatorial contribution to the activity coefficient $\ln \gamma_i^{C}$ that is essentially due to differences in size and shape of the molecules and a residual contribution $\ln \gamma_i^{R}$ that is essentially due to energetic interactions as shown in Table I-9.

UNIFAC has been successfully used for the design of distillation columns (including azeotropic and extractive distillation) where the required multicomponent activity coefficients were estimated because of a lack of experimental information. Separate UNIFAC correlations have been proposed for liquid-liquid equilibria (Magnussen et al. 1981), but these tend to be less accurate than those for vapour liquid equilibria and, therefore, are not used widely (Prausnitz et al. 1999).

For engineering design, correlations (in particular, group-contribution methods) are attractive because it is almost always easier and faster to make a calculation than to perform an experiment. The GCMs are based on the assumption of group independence; this assumption says, for example, that the properties of a carbonyl (C=O) group in a monoketone are the same as those in a diketone, and that the properties of this carbonyl group in a linear ketone (e.g. hexanone) are the same as those in a cyclic ketone (e.g. cyclohexanone). Also a group contribution method does not distinguish between 1-hexanone and 3-hexanone; the position of the carbonyl group in the molecule is not considered. These assumptions are, of course, not correct. It is possible to relax these assumptions by introducing more groups, e.g. by distinguishing between a carbonyl group in a linear molecule and that in a cyclic molecule. Agreement with experiment can then be improved but at a high cost: more parameters are required; to obtain them we need more experimental data. In the limit, as we define more and more groups, the advantage of the group-contribution method is lost. A useful GCM is a compromise between on the one hand, using many different groups as suggested by our knowledge of molecular structure, and on the other, using only a limited number of groups, as required by a limited bank of experimental data (Prausnitz et al. 1999).

Table I-9: Some group contribution models for the Excess Gibbs Energy and subsequent activity coefficients for multicomponent systems (from Prausnitz et al. 1999).

Names	Parameters	Activity coefficients
ASOG (Analytical Solution of Groups) Derr and Deal 1969, 1973	a_{kL}	$\ln\gamma_i = \ln\gamma_i^S + \ln\gamma_i^G$ $\ln\gamma_i^S = 1 + \ln r_i - r_i$ where $r_i = \frac{\nu_i}{\sum_j x_j \nu_j}$ $\ln\gamma_i^G = \sum_k \nu_{ki}\left(\ln\Gamma_k - \ln\Gamma_k^{(i)}\right)$ where $\begin{cases}\nu_i = \sum_k \nu_{ki} \text{ when } H_2O, CH \text{ and } C \text{ are absent}\\ \nu_{H_2O} = 1.6; \nu_{CH} = 0.8; \nu_C = 0.5\end{cases}$ and $\begin{cases}\ln\Gamma_k = 1 - \ln D_k - C_k\\ \ln\Gamma_k^{(i)} = 1 - \ln D_k^{(i)} - C_k^{(i)} : \text{in each pure substances}\\ \ln\Gamma_k^{(i)} = 0 : \text{if only one kind of molecule is present in the molecule}\end{cases}$
UNIFAC (UNIQUAC Functional Group Activity Coefficients) Fredenslund et al. 1975	Structural parameters of each group k R_k, Q_k Interaction parameters between any couples of groups k and l $a_{k,l}$	$\ln\gamma_i = \ln\gamma_i^C + \ln\gamma_i^R$ $\ln\gamma_i^C = \ln\frac{\Phi_i}{x_i} + \frac{z}{2} q_i \ln\frac{\theta_i}{\Phi_i} + l_i - \frac{\Phi_i}{x_i}\sum_j x_j l_j$ where $\begin{cases} r_i = \sum_{k=1}^{NG} \nu_{k,i} R_k \equiv \text{ volume parameter of molecule } i\\ q_i = \sum_{k=1}^{NG} \nu_{k,i} Q_k \equiv \text{ surface parameter of molecule } i\end{cases}$ $z = 10 \equiv$ coordination number $\ln\gamma_i^R = \sum_k \nu_{k,i}\left(\ln\Gamma_k - \ln\Gamma_{k,i}\right)$ where $\ln\Gamma_k = Q_k\left[1 - \ln\left(\sum_{j=1}^{NG}\Theta_j\Psi_{j,k}\right) - \sum_{j=1}^{NG}\frac{\Theta_j\Psi_{k,j}}{\sum_{m=1}^{NG}\Theta_m\Psi_{m,j}}\right]$ and $\begin{cases}\Theta_k = \frac{x_k Q_k}{\sum_{m=1}^{NG} x_m Q_m} \equiv \text{surface fraction of the group } k\\ x_l = \frac{\sum_j x_j \nu_{l,j}}{\sum_j x_j \sum_k \nu_{k,j}} \equiv \text{molar fraction of the group } l\\ \Psi_{k,l} = \exp\left(\frac{-a_{k,l}}{RT}\right) \equiv \text{Interaction term}\end{cases}$

Names	Parameters	Activity coefficients
Modified UNIFAC (UNIQUAC Functional Group Activity Coefficients) UNIFAC-Dortmund (Weidlich and Gmehling 1987, Gmehling et al. 1993)	Structural parameters of each group k R_k, Q_k Interaction parameters between any couples of groups *k* and *l* $a_{k,l}, b_{k,l}$, and $c_{k,l}$	Modification 1: $\tau_{k,l} = \exp\left[\frac{-A_{k,l}}{T}\right]$ $A_{k,l} = a_{k,l} + b_{k,l}T + c_{k,l}T^2$ Modification 2: $\ln \gamma_i^C = \ln \frac{w_i}{x_i} + 1 - \frac{w_i}{x_i} - \frac{zq_i}{2}\left[1 - \frac{\Phi_i}{\theta_i} + \ln \frac{\Phi_i}{\theta_i}\right]$ where $w_i = \frac{x_i r_i^{3/4}}{\sum_j x_j r_j^{3/4}}$
Modified UNIFAC (UNIQUAC Functional Group Activity Coefficients) Larsen et al. 1987	Structural parameters of each group k R_k, Q_k Interaction parameters between any couples of groups *k* and *l* $a_{k,l}, b_{k,l}$, and $c_{k,l}$	Modification 1: $\tau_{k,l} = \exp\left[\frac{-A_{k,l}}{T}\right]$ $A_{k,l} = a_{k,l} + b_{k,l}\left(T - T_0\right) + c_{k,l}\left(T \ln \frac{T_0}{T} + T - T_0\right)$ $T_0 = 298.15 K$ Modification 2: $\ln \gamma_i^C = \ln \frac{w_i}{x_i} + 1 - \frac{w_i}{x_i}$ where $w_i = \frac{x_i r_i^{2/3}}{\sum_j x_j r_j^{2/3}}$

However, because UNIFAC (and similar methods) provide only rough approximations, it is necessary to obtain at least a few number of reliable experimental results, either from the literature or from the laboratory. On the other hand, many experimental results reported in the literature are of poor quality and in that event, a calculated activity coefficient may be more reliable.

An exciting method for improving accuracy and for increasing the range of applicability of group-contribution methods is provided by quantum-chemical calculations to give group-group interaction parameters from first principles. Initial efforts toward that end have been reported by Sandler and co-workers (Wu and Sandler 1991, Wolbach and Sandler 1997). The UNIFAC model still has a problem; that is, the ions (cation and anion groups) are not complete in the UNIFAC menu, as well as several chemical groups like those including sulphur.

In summary, the objective of group-contribution methods is to use existing phase equilibrium data to predict phase equilibria for systems where no data are available. While such predictions may be used for preliminary design purposes, group-contribution methods yield only approximate vapour-

liquid equilibria. Whenever reliable experimental data are available, these should be used instead of group-contribution predictions. While the Raoult's law provides a zeroth approximation for phase equilibria, group-contribution methods often provide no more than a first approximation (Prausnitz et al. 1999).

I.5.2.1.4. COSMO-RS models

The COSMO-RS model, *i.e.* the extension of COSMO (conductor-like screening) model to real solvents (RS), is a novel and efficient method for the *a priori* prediction of thermophysical data, developed since 1994. It is based on a physically founded model and, unlike GCMs, uses only atom-specific parameters, which can be used to predict the thermodynamic properties of solvent-ionic liquid systems. Therefore, it is anticipated that this model is, at least qualitatively, able to describe structural variations correctly (Lei et al. 2008). Since Chapther II is devoted to a more detailed description of the COSMO-RS model, only a short description of the COSMO-RS methods are given in this section.

First, using quantum calculations tools, each of the individual molecules are moved from vacuum to the conductor (and this state denoted as "COSMO state" is chosen as COSMO-RS reference state). Then, the individual molecular surfaces are split in segments, histograms are formed (σ-profiles) and a statistical thermodynamics calculation is done on the segments, resulting in chemical potentials (free energies) of segments. These are transferred back to the molecular surfaces, and the chemical potential of a molecule is calculated as the sum of the segment chemical potentials on its surface (plus a combinatorial correction).

Instead of assuming the surrounding molecules in focus (solutes) to be an electrical conductor, the COSMO-RS method separates the surface of the solute molecule into portions of given area and compares the screening charges with those of a second molecule (the solvent), which is treated in the same manner. The screening charges represent the electrostatic interaction potential of the molecules and enable the calculation of one component's chemical potential using a statistical mechanics approach (Lei et al. 2008).

Liquid in COSMO-RS is considered to be an ensemble of almost closely packed ideally screened molecules, and the interactions of the molecules are expressed as pairwise interactions of the screening charges. This includes electrostatic interactions as well as hydrogen bonding. By this reduction of molecular interactions to surface contacts, the statistical thermodynamics is reduced to a simple set of equations, which are similar to, but even somewhat more accurate than, the UNIQUAC type of equations used in the UNIFAC models (Lei et al. 2008).

During the statistical thermodynamics calculations, it is assumed that ideal electrostatic behaviour means a complete neutrality of the charges of both molecules. Every deviation from this, called a "misfit", leads to activity coefficients (which are derived from the knowledge of the chemical

potential) differing from unity. Furthermore, the energy to transport a molecule into an electrical conductor is a measure of the vapour pressure. Moreover, hydrogen bonding and van der Waals interactions are also taken into accounts in the calculation of the chemical potential of each component in the mixture. As a result, all kinds of thermodynamical data can be calculated. The model even works for multicomponent systems. For instance, the activity coefficient of component *i* is related to chemical potential and given as follows:

$$\gamma_i = \frac{1}{x_i} \exp\left(\frac{\mu_i - \mu_i^0}{RT} \right) \tag{I-48}$$

where μ_i is the chemical potential of component *i* in the mixture and μ_i^0 is the chemical potential in the pure liquid substance, *R* is the gas constant and *T* is the system temperature .

The accuracy of COSMO-RS depends strongly on the quantum chemical method used. Furthermore, the method carries internal parameters because only electrostatic interactions are taken into account. The following atom-based internal constants are obtained using this method:

- Radius of the elements (used for cavity construction; only 17% larger than bond radii);
- Dispersion constants c_{vdW} and τ_{vdW} (one per element; vdW energy contributions expressed by element-specific parameters);
- Effective contact area a_{eff} (determines the number of independent neighbors for a molecule);
- Electrostatic misfit energy coefficient α' ("self" energy of a single segment of a surface divided by the surface charge density);
- Hydrogen-bonding constants c_{hb} and σ_{hb} (σ_{hb} is the threshold for hydrogen bonding and c_{hb}, the strength coefficient);
- Ring correction coefficient, ω;
- Coefficient, λ, for the combinatorial part of the chemical potential; and
- Transfer constant, η (connects reference states in gas and solution).

These constants have been determined once and have since been improved in several revisions (Lei et al. 2008). The popularity of the COSMO-RS approach in chemical engineering has caused the development of a number of reimplementations, while the original COSMO-RS had been developed in the COSMOtherm program since 1999. Although they slightly differ with respect to the parameterization and the details of the implementation, most of them are named as COSMO-RS in the literature. Some confusion was caused by the fact that the first reimplementation of COSMO-RS was published as 'COSMO-SAC' and initially was claimed to be an independent model, but meanwhile it is generally accepted that COSMO-SAC is just a reimplementation of the COSMO-RS concept. All reimplementations made so far are less complete and detailed as compared with COSMOtherm. All available comparisons indicate a higher accuracy of COSMOtherm as compared with the reimplementations namely the COSMO-SAC model proposed by Lin and Sandler (2002, 2004) and COSMO-RS(Ol) developed by Grensemann and Gmehling (2005).

Within the COSMOtherm software, a number of additional applications of COSMO-RS have been developed by the COSMOlogic team, including the prediction of dissociation constants in aqueous and nonaqueous solvents, the prediction of the free energy of molecules at liquid-liquid and liquid-vapour interfaces, and the prediction of the free energies and of the partitioning of solutes in polymers, micellar systems, and biomembranes. Furthermore, a set of QSAR descriptors, the so-called σ-moments, has been derived from the COSMO-RS theory, which can be used to regress almost any kind of partition property even in complex cases as blood-brain partitioning, soil sorption, adsorption to activated carbon, and many more. With respect to the native purpose of solvation models, *i.e.*, with respect to the prediction of the free energies of solvation, COSMO-RS in its COSMOtherm implementation has currently proved to be most accurate. The COSMO-RS model can also be used to evaluate the separation ability of ionic liquids for a given separation task. It is composed of three steps: conformational analysis, COSMO calculation, and COSMO-RS calculation.

I.5.2.2. Short Range (SR) models based on the chemical theory of solution

The models discussed above treat the non-ideal behaviour of solutions in terms of physical intermolecular forces not specific. These models relate the activity coefficients to physical quantities that reflect the size of the molecules and the physical forces operating between them. However, these considerations are not always viable in some cases where several non-ideal behaviours remain, even after the use of physical models. Such deviations can be alternatively treated by using another approach, the so-called chemical theory in which it is assumed that in solution, solute and solvent molecules interact to form new chemical species, the non-ideal behaviour is thus characterized by the resulting chemical reactions. There are essentially two types of chemical reactions, association and solvation (Prausnitz et al. 1999).

Association reactions usually lead to the formation of chemical aggregates or dimers, trimers, etc. consisting of identical monomers. These reactions can be represented by reactions of the type:

$$A_n + A_m \rightleftharpoons A_{n+m} \tag{I-49}$$

$$B_n + B_m \rightleftharpoons B_{n+m} \tag{I-50}$$

where A and B denote the monomer and n, m the degrees of association (or polymerization). A common case of association is the dimerization ($n+m=2$) which are due to hydrogen bonding that is responsible for the most common form of association in liquid solutions (Prausnitz et al. 1999).

Solvation reactions refer to the formation of chemical aggregates of two or more molecules that are not all identical, represented by the general equation:

$$A_n + B_m \rightleftharpoons A_n B_m \tag{I-51}$$

Here too, the formation of new species is due to hydrogen bonding or to electron exchanges between molecules (Prausnitz et al. 1999).

The chemical theory of solutions postulates existence of chemically distinct species in solution that are assumed to be in chemical equilibrium. In its original form, the theory then assumes that these chemically distinct substances form an ideal solution. According to these assumptions, the observed non-ideality of a solution is only an apparent one because it is based on an apparent, rather than a true, account of the solution's composition (Prausnitz et al. 1999).

Below, we will only give a short insight to the chemical solvation model.

Let us consider the simple case of a binary solution where complexes form according to

$$A + B \rightleftharpoons AB \tag{I-52}$$

The equilibrium constant is related to the activities of the three species by

$$K = \frac{a_{AB}}{a_A a_B} \tag{I-53}$$

where a_A and a_B denote the respective activities of the monomer molecules A and B, and a_{AB} the activity of the dimer molecules AB.

If the solution is formed by n_1 moles of A and n_2 moles of B, and if at equilibrium n_{AB} moles of complex are formed, the true mole fraction x of A, B, and AB are:

$$x_A = \frac{n_1 - n_{AB}}{n_T} = \frac{n_1 - n_{AB}}{n_1 + n_2 - n_{AB}} \tag{I-54}$$

$$x_B = \frac{n_2 - n_{AB}}{n_T} = \frac{n_2 - n_{AB}}{n_1 + n_2 - n_{AB}} \tag{I-55}$$

$$x_{AB} = \frac{n_{AB}}{n_T} = \frac{n_{AB}}{n_1 + n_2 - n_{AB}} \tag{I-56}$$

where $n_T = n_1 + n_2 - n_{AB}$ denotes the total mole number.

The chemical theory of solutions considers that the three species are in equilibrium, *i.e.*, A, B and AB form an ideal solution (Prausnitz et al. 1999). Therefore the activity of each species is equal to its true mole fraction.

The apparent mole fractions of the two components are z_1 (for A) and z_2 (for B). They are given by:

$$z_1 = \frac{n_1}{n_1 + n_2} \tag{I-57}$$

$$z_2 = \frac{n_2}{n_1 + n_2} \tag{I-58}$$

If all the species in solution are well defined, it is possible to establish an expression of the excess molar Gibbs energy g^E of the mixture as a function of the concentrations of the species and the equilibrium constant K. The activity coefficients of these components are given by:

$$\gamma_1 = \frac{a_A}{z_1} \tag{I-59}$$

$$\gamma_2 = \frac{a_B}{z_2} \tag{I-60}$$

Algebraic rearrangement then gives for the activity coefficients:

$$\gamma_1 = \frac{kz_1 - 2 + 2\left(1 - kz_1 z_2\right)^{1/2}}{kz_1{}^2} \tag{I-61}$$

$$\gamma_2 = \frac{kz_2 - 2 + 2\left(1 - kz_1 z_2\right)^{1/2}}{kz_2{}^2} \tag{I-62}$$

where $k \equiv \frac{4K}{K+1}$. If $K = 0$, $\gamma = 1$, because in this case, no complex AB is formed, and therefore, by assumption, there is no deviation from ideal behaviour. At the other extreme, when $K = \infty$, activity coefficients of both components (γ_1 and γ_2) go to zero at the midpoint ($x_1 = x_2 = \frac{1}{2}$) because at this particular composition all molecules are complexed and no uncomplexed molecules A or B remain.

During the last years, many authors have pointed out a good representation of VLE experimental data by the chemical models (Sandler 1994). This success is not surprising since one of the main advantages of the chemical models is their uniqueness for a wide variety of molecules, and then it is not necessary to give no parameters or additional data when the initial geometries are known. Moreover, the theoretician is free of choosing which chemical equilibria he would like to study, then affects an arbitrary equilibrium constant to each studied chemical equilibrium reaction. However, this arbitrariness in deciding what "true" chemical species are present in the solution, and the inability to assign equilibrium constants to the postulated equilibria without experimental data on the solution under consideration consist the great disadvantage of the chemical models. The chemical theory of solutions, therefore, has little predictive value; it can almost never give quantitative predictions of solution behaviour from pure-component data alone. However, for those solutions where chemical forces are dominant, the chemical theory has much qualitative and interpretative value; if some data on such a solution are available, they can often be interpreted along reasonable chemical lines, and therefore, the chemical theory can serve as a tool for interpolation and cautious extrapolation of limited data. A chemical rather than a physical view is frequently used for correlation of solution non-

ideal behaviours in a class of chemically similar mixtures like alcohols in paraffinic solvents (Prausnitz et al. 1999).

I.5.3. Models suitable for the treatment of long range effects

Contrarily to non-electrolytes solutions where interactions are mainly of short range nature, the non-ideal behaviour of electrolytes solutions are described as a sum of two types of interactions: electrostatic interactions (attractive or repulsive) of long range nature, and the short range interactions between ions and solvent molecules. This deviation from ideality is encountered in electrochemistry, as well as in several phase equilibria calculations (salts precipitation in saturated solution, separation of two liquid phases, vapour-liquid equilibria).

The equations of section I.4 define the activity and the activity coefficient of non-dissociating solute. However in solution of an electrolyte, the solute dissociates into cations and anions. Cations and anions are not independent components because of the electroneutrality. For example, when one mole of a strong electrolyte like NaCl is dissolved to a one-molal solution of positively charged sodium ions and a one-molal solution of negatively charged chloride ions. Ordinarily thermodynamic measurements give properties not of individual ionic species, but of the neutral electrolytes formed by cations and anions. In a solution of an electrolyte, electroneutrality imposes the condition that the number of moles of the individual ionic species cannot be varied independently. In aqueous NaCl there are three species but only two (not three) components (Prausnitz et al. 1999).

When dissolved in a high-dielectric-constant solvent like water, an electrically neutral electrolyte $C_{\nu_+}A_{\nu_-}$ is dissociated into ν_+ positive ions (cations C) each with a charge z_+ and ν_- negative ions (anions A) each with a charge z_-. Charges are given in normalized units where $z_+=1$ for a proton. Electrolytic dissociation is represented by

$$\nu_+ C^{z_+} + \nu_- C^{z_-} \rightleftharpoons C_{\nu_+} A_{\nu_-} \tag{I-63}$$

Electro-neutrality requires that

$$\nu_+ z^+ + \nu_- z^- = \nu_+ z^+ - \nu_- \left| z^- \right| = 0 \tag{I-64}$$

Thus the mean ionic molality and the mean activity coefficient of the electrolyte $C_{\nu_+}A_{\nu_-}$ are given by

$$m_\pm = \left(m_+^{\nu_+} m_-^{\nu_-} \right)^{\frac{1}{\nu_+ + \nu_-}} \tag{I-65}$$

$$\gamma_\pm = \left(\gamma_+^{\nu_+} \gamma_-^{\nu_-} \right)^{\frac{1}{\nu_+ + \nu_-}} \tag{I-66}$$

Before discussing the main activity coefficients models of electrolyte solutions, it is important to notice that:

- the logarithm of the activity coefficient $\ln \gamma_{\pm}$ is a sum of two terms : a short range term $\ln \gamma_{\pm}^{SR}$ and a long range term $\ln \gamma_{\pm}^{LR}$;
- we choose as reference the infinite dilute solution where by definition $\gamma_{\pm}(i) = 1$;
- the long range term describes the phenomena appearing in dilute solutions, where solute molecules are relatively located far from each other, and therefore the activity coefficient is mainly dominated by the influence of charges. This long range term was originally proposed by Debye and Hückel in 1923 and 1924, and the activity coefficient models based on the calculation of this term are satisfactory only at low ionic strength;
- the short range term describes what happens in concentrated solutions, where solute molecules are very close to each other. This term is generally a sum of ion-ion and ion-undissociated molecule terms. In almost all of the practical cases, the main interactions considered in aqueous electrolyte solutions are the cation-anion interactions (Sandler 1994).

I.5.3.1. Debye-Hückel theory (dilute ionic solutions)

At fixed concentration of ions, electrolytes containing ions with multiple charges have a stronger effect on the activity coefficients of ions than electrolytes containing only singly-charged ions. To express this dependence it is useful to introduce the (molal) ionic strength of the solution, I, defined by

$$I(\text{mol.kg}^{-1}) = \frac{1}{2}\sum_{i=1}^{nI} m_i z_i^{\,2} \qquad \text{(I-67)}$$

where m_i and z_i denote respectively the molality and the charge of a given ion i while nI denotes the number of charged species in solution.

The starting point of the model proposed by Debye and Hückel is to consider the ions included in a dilute solution, as punctual charges which interact with coulombic forces (Prausnitz et al. 1999). Thus, if we choose one punctual ion i as origin, the electrostatic potential $\Psi_i(\mathrm{r})$ around this ion (assuming that it is alone in solution) is given by

$$\Psi_i(\mathrm{r}) = \frac{-z_i e}{4\pi\varepsilon_0\varepsilon_r}\frac{1}{r} \qquad \text{(I-68)}$$

where $(z_i e)$ denotes the charge of the particle i (in Coulomb), ε_r and ε_0 denote respectively the dielectric constant of the solvent and the vacuum permittivity.

At infinite dilution, the distribution of ions in solution can be considered completely random because the ions are too far apart to exert any significant influence on each other. In this case, the mean ion activity coefficient of the electrolyte is unity. However, for dilute (not infinitely dilute) solutions, where the ions are no longer “blind” to each other, Coulombic forces become important; in the neighbourhood of a negative ion, the local concentration of positive ions is slightly higher than that for the bulk solution (Prausnitz et al. 1999).

A slightly positive atmosphere around anion *i*, and a slightly negative atmosphere around cation *j*, produce a decrease in attraction between *i* and *j*. This decrease due to preferential spatial distribution of ions gives a shielding effect. To account for shielding, the theory of Debye-Hückel shows that r^{-1} in Coulomb's potential should be multiplied by a "damping factor",

$$r^{-1} \to \left(r^{-1}\right)\exp\left(-r\kappa\right) \tag{I-69}$$

Here κ^{-1} is the shielding length commonly called the Debye length which is a characteristic distance of interaction; it plays an important role in the Debye-Hückel theory. When κ^{-1} is very large, $r\kappa$ is small and the exponential in equation I-69 is close to unity; in that event, we recover the original Coulomb potential that is for two isolated charges in a continuous medium characterized by ε_r (Prausnitz et al. 1999).

The Debye length is defined as

$$\kappa^{-1} = \left(\frac{\varepsilon_0 \varepsilon_r RT}{2 d_s \mathcal{N}^2 e^2 I}\right)^{\frac{1}{2}} \tag{I-70}$$

where ε_0 is the vacuum permittivity ($\varepsilon_0 = 8.85419\ 10^{12}\ C^2N^{-1}m^{-2}$), ε_r is the relative permittivity or dielectric constant, d_s is the solvent density ($g.cm^{-3}$), $\mathcal{N}$ is Avogadro's constant, e is the electronic charge ($e = 1.60218\ 10^{-19}$ C) and I is the ionic strength. This equation tells us that the Debye length decreases with rising concentration; the higher the concentration of ions (ionic strength), the more effective the shield.

For dilute aqueous solutions near ambient temperature, there is no significant difference between molality and molarity. The activity coefficients of ions, expressed in molality scale, are given by

$$\ln \gamma_i^{(m)} = -A_\gamma z_i^2 I^{1/2} \tag{I-71}$$

where A_γ is given by

$$A_\gamma = \left(\frac{e^2}{\varepsilon_0 \varepsilon_r RT}\right)^{\frac{3}{2}} \frac{\mathcal{N}^2}{8\pi} \left(2 d_s\right)^{\frac{1}{2}} \tag{I-72}$$

Equation I-71 gives the activity coefficient of a given ion *i*.

For electrolyte $C_{\nu+}A_{\nu-}$, upon substitution of equation I-71 in equation I-66, $\gamma_\pm$ is given by

$$\ln \gamma_\pm^{(m)} = -A_\gamma \left|z_+ z_-\right| I^{1/2} \tag{I-73}$$

where $\left|z_+ z_-\right|$ is the absolute value of the product of the charges.

Likewise, a similar derivation yields for the osmotic coefficient

$$\Phi - 1 = -A_\Phi \left|z_+ z_-\right| I^{1/2} \tag{I-74}$$

where the Debye-Hückel constant A_ϕ is directly related to constant A_γ given by:

$$A_\Phi = \frac{1}{3} A_\gamma \tag{I-75}$$

Equation I-73 is the Debye-Hückel limiting law, useful for interpreting the properties of electrolyte solutions. This law is applicable only for dilute solutions (with low ionic strength to 0.01 mole.kg^{-1}).

I.5.3.2. Pitzer model

Zemaitis et al. (1986) have reviewed several models based on the Debye-Hückel theory that link the activity coefficients to several solution properties like the ionic strength, the volumic mass or density, the dielectric constant of the solvent and the ionic charges.

Among these methods, in the model proposed by Pitzer, the so-called Pitzer Debye-Hückel (PDH) model, the excess Gibbs energy can be expressed as a function of the molar fractions of the species in solution, in the infinite dilution reference state, by

$$g_{PDH}^{E*} = -\sum_k x_k \sqrt{\frac{1000}{M_s}} \frac{4 A_\Phi I_x}{\rho} \ln\left(1 + \rho\sqrt{I_x}\right) \tag{I-76}$$

where:

- the summation concerns all the (molecular and ionic) species in solution;
- M_s is the molar mass of the solvent (in g.mol^{-1}), M_s=18.0153 g.mol^{-1} for water;
- ρ is the closest approach parameter that characterizes the minimal distance between two ions of opposed charges, its value had been adjusted to experimental results and a constant value of 14.9 is commonly applied to a variety of salts (Pitzer 1980);
- I_x is the ionic strength of the solution expressed in the molar fraction scale, given by

$$I_x = \frac{1}{2} \sum_{i=1}^{nI} z_i^2 x_i \tag{I-77}$$

  where x_i and z_i denote respectively the molar fraction and the charge of a given ion i while nI denotes the number of charged species in solution, I_x is linked to the ionic strength expressed in molarity scale by

$$I_x = \frac{I_c}{55.5084 + I_c} \tag{I-78}$$

- and A_Φ is the Debye-Hückel parameter:

$$A_\Phi = \frac{1}{3} \sqrt{\frac{2\pi \mathcal{N} \rho_s}{1000}} \left(\frac{e}{\sqrt{D_s kT}} \right)^3 \tag{I-79}$$

where: ρ_s denotes the density of the solvent (in g.cm^{-3}); D_s is the dielectric constant of the solvent, for water D_s = 78.03 at T=25°C; k is the Boltzmann constant and $\mathcal{N}$ is Avogadro's constant. Thus, A_Φ= 0.391 for an aqueous solution at 25°C.

The activity coefficient of a given ion *i*, can be expressed as a function of the molar fractions of the species in solution, in the infinite dilution reference state, by

$$\ln \gamma^*_{i,PDH} = -\sqrt{\frac{1000}{M_s}} A_{\Phi k} \left[\frac{2 z_i^{\,2}}{\rho} \ln\left(1+\rho\sqrt{I_x}\right) + \frac{2 z_i^{\,2} \sqrt{I_x} - 2\left(I_x\right)^{\frac{3}{2}}}{1+\rho\sqrt{I_x}} \right] \tag{I-80}$$

Since this study is focused on aqueous solutions, it is useful to get rigorous formulae for the solvent density ρ_s and the dielectric constant D_s of water at different temperature.

For this purpose, using the correlation proposed in Ould-Moulaye 1998, based on data from Archer and Wang 1990, the dielectric constant of water is given by:

$$D_s = 78.30 \left(\frac{T}{298.15} \right)^{-1.4577} \tag{I-81}$$

Likewise, by correlating the data from Sengers and Watson 1986 and Haar et al. 1984, Ould-Moulaye has proposed a correlation for the water density:

$$\rho_s = -3.56410^{-6} T^2 + 1.877510^{-3} T + 0.7536 \,(\text{in g.cm}^{-3}) \tag{I-82}$$

Elsewhere, the temperature dependence of the constant A_Φ is given by Zemaitis et al. (1986)

$$\begin{aligned} A_\Phi = & -61.44534 \exp\left(\frac{T-273.15}{273.15} \right) + 2.864468 \left(\exp\left(\frac{T-273.15}{273.15} \right) \right)^2 \\ & + 183.5379 \ln\left(\frac{T}{273.15} \right) - 0.6820223 \left(T - 273.15 \right) \\ & + 0.0007875695 \left(T^2 - \left(273.15\right)^2 \right) + 58.95788 \left(\frac{273.15}{T} \right) \end{aligned} \tag{I-83}$$

It is important to notice that in the PDH model, the non ideal behaviour is due to electrostatic interactions that depend only on the charge of the ion, thus this model does not require any ion specific parameter.

I.5.4. Activity coefficients models for non-electrolyte solutions containing electrolytes

Electrolytes mixtures come in various forms and add another dimension to the normal complexities of non-electrolyte solutions: entirely new species can be formed in water, some of which

are not obvious; components can precipitate; soluble components can affect vapour pressure of the solution very significantly. In industrial applications, the solutions are often highly concentrated and encounter high pressures and temperatures. Therefore expertise in electrolyte systems has become increasingly critical in oil and gas exploration and production, as well as in the more traditional chemical industry operations. A variety of correlations are available that can solve the problems that are encountered in industry.

Models for electrolyte activity coefficients result largely from that of Debye-Hückel for the long-range ion-ion interaction contribution, like that of Pitzer 1973. Nevertheless, in all cases, Robinson and Stokes (1959), then Achard et al. (1994) clearly identified the need to define the ionic entity in terms of its degree of hydration. Finally, in the food industry, Lebert and Richon (1984), Le Maguer (1992), Achard (1992), Catté (1994), Catté et al. (1995), Lebert et al. (2005) have made a pioneer work on aqueous carbohydrate systems. UNIFAC equation is also used to predict the retention of aroma compounds by food (Sancho et al. 1997) or a_w and pH of bacterial growth media (Coroller et al. 2001; Lebert et al. 2005, 2006). The structure of the models permits predictive approaches for complex solutions over a large range of concentration, and even ionic activity modelling, i.e. solution pH modelling (Ben Gaïda et al. 2006, 2010).

GePEB-IP team has demonstrated many years ago that 3 types of interactions must be considered in a complementary way to develop a robust thermodynamic modelling that could be used for any type of solution (Achard et al. 1994). For predicting activity coefficients of strong electrolytes, Achard 1992 and Achard et al. 1994, have combined a chemical treatment of ions solvation with two activity coefficient models namely one based on the Debye-Hückel theory (Pitzer 1980, Pitzer 1973) which treats LR interactions and the UNIFAC-Larsen model (Larsen et al. 1987) which takes in accounts SR interactions.

In the resulting model, called “ULPDHS”, the activity coefficient of each compound *i* is given by:

$$\ln \gamma_i = \ln \gamma_i^{SR} + \ln \gamma_i^{LR} \tag{I-84}$$

The latter equation is suitable for calculating the activity coefficients of both solvents and ions. An extended form of the Debye-Hückel law given by Pitzer, is used to account for LR electrostatic forces, whereas the UNIFAC model is for SR physical interactions. Additionally, the solvation equations accounting for hydration of ions in aqueous solution (formation of water clusters) are incorporated into the UNIFAC model. Structural parameters for hydrated components for aqueous solution are the following:

$$R_k^H = R_k + N_{h_k} R_1 \tag{I-85}$$

$$Q_k^H = Q_k + N_{h_k} Q_1 \tag{I-86}$$

where R_1 and Q_1 refer to water and N_{h_k} denotes the infinite dilution hydration number of group k.

The ions in the mixture are considered to be hydrated by water (1), except for organic solvents, whose hydration numbers are set to be zero. Moreover, it is also assumed that solvation phenomena between organic solvents and ions are nonexistent. In this case, the mole fractions of water and ionic components are, respectively,

$$x_1^H = \frac{x_1 - \sum_{j=2}^{N} N_{h_j} x_j}{1 - \sum_{j=2}^{N} N_{h_j} x_j} \tag{I-87}$$

$$x_i^H = \frac{x_i}{1 - \sum_{j=2}^{N} N_{h_j} x_j} \tag{I-88}$$

Activity coefficients of water and ionic components are, respectively,

$$\gamma_1^{SR} = \gamma_1^{SR,H} \frac{x_1^H}{x_1} \tag{I-89}$$

$$\gamma_i^{SR} = \gamma_i^{SR,H} \frac{x_i^H}{x_i} \left(\gamma_1^{SR,H} x_1^H \right)^{-N_{h_i}} \tag{I-90}$$

where $\gamma_i^{SR,H}$ are calculated with the UNIFAC model using mole fractions and structural parameters of hydrated components given above. There is no influence of hydration on activity coefficients of organic solvents.

The group interaction parameters were estimated between ions (Li^+, Na^+, K^+, Mg^{2+}, Ca^{2+}, Co^{2+}, Ba^{2+}, Sr^{2+}; NH_4^+, Cl^-, Br^-, F^-, NO_3^- and SO_4^{2-}) and solvent groups (CH_2, OH, CH_3OH, H_2O and CH_3CO). The group interaction parameters between the solvent groups are the same as those in the UNIFAC model. It was shown that the ULPDHS model represents VLE for solvent-water-salt mixtures with an average accuracy of the total pressures around 4% and the vapour-phase mole fractions around 4%. This model (ULPDHS) has been adopted by ProSim S.A. and is available in the software 'Simulis Thermodynamics'. It can be used for multi-electrolytes solutions without adding any new interaction coefficient and satisfactorily predicts water activity, osmotic coefficients and salting-out effects in aqueous mixtures of two or three electrolytes within less than 5% even for saturated solutions. The ULPDHS model was also successfully used to study the equilibrium thermodynamic properties of aqueous systems containing sugars, polyols, mineral salts, weak acids, etc. (Catté 1994, Achard 1992 , Ben Gaïda 2007).

However, the main drawbacks of such solution models (using UNIFAC, water solvation and Pitzer-Debye-Hückel contributions) are to get the value of the parameters required to feed the model

because of the interdependence of the different contributions. This model is not able to distinguish neither isomers nor conformers, due to the inherent assumption of the group contribution concept. Moreover, for several components of interest in food science (like amino acids containing sulphur, phosphate glucose, or the ascorbic acid) the UNIFAC group interactions parameters are not available. Furthermore, as discussed above, the UNIFAC model (alone) cannot allow the prediction of the properties derived from thermo-chemistry like the pK_a of acids or bases, which are required to get an efficient model for redox potential prediction (as illustrated on Figure I-3). Moreover there are not enough data in the literature concerning the formation properties of ions, radicals (which are rarely isolated in aqueous solution to allow experimental measurements without any theoretical assumptions that should biases the experimental results).

To overcome these difficulties, we decided to explore a new way based on the recent progress of quantum chemistry and statistical thermodynamics. This model should ideally be used in the same modelling structure as that of the ULPDHS model. Our choice leads to the COSMO-RS method that is also an excess Gibbs energy method based on the treatment of short range effects, which does not require any interactions parameters. This model appears as a promising alternative predictive tool, especially when the UNIFAC model cannot be used. It is discussed in detail in chapter II where a state-of-the art of the COSMO-RS method and its application ranges are introduced.

Chapter II:The COSMO-RS model

II.1. General presentation of the COSMO-RS method

The COSMO-RS method is a combination of the COnductor-like Screening MOdel (COSMO, a variant of dielectric continuum models) with a statistical thermodynamics treatment of interacting surfaces for Realistic Solvation (the RS part of the COSMO-RS model). It is a widely used predictive method for predicting thermodynamic equilibrium of fluids and liquid mixtures. COSMO-RS uses the intermediate results from quantum chemistry (QC) calculations on individual molecules to predict thermodynamic properties of mixtures of these molecules, as illustrated in Figure II-1.

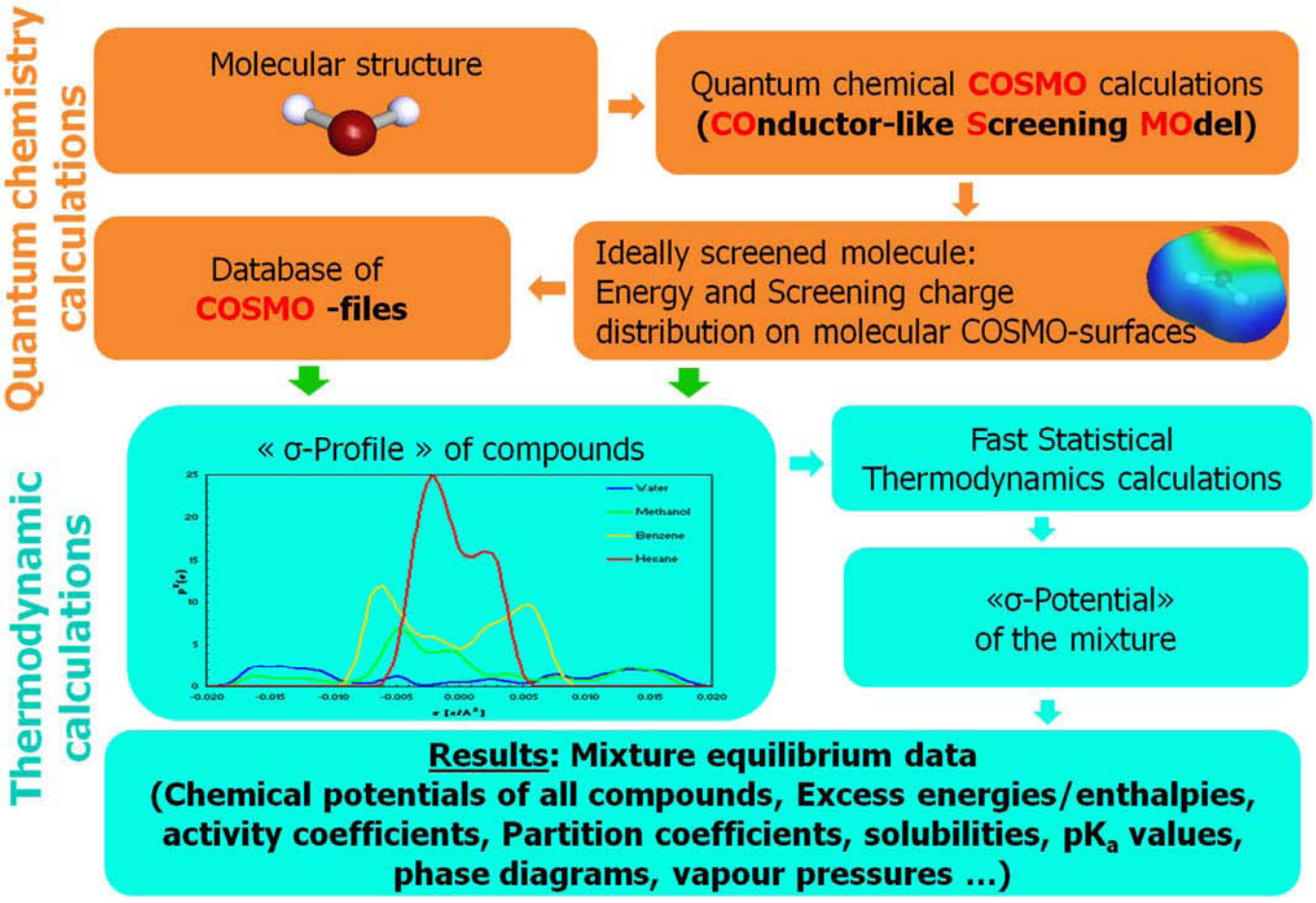


Figure II-1 : Flowchart of the COSMO-RS method (Adapted from COSMOthermX 2011).

COSMO-RS is described as the most accurate quantum chemically based model for the prediction of solvation energies (Klamt 2005, 2011). It also allows the prediction of all kind of thermodynamic equilibrium properties of liquids. There are a fair number of reports of accurate prediction by COSMO-RS of thermodynamic properties in general in the literature. Many of these have been written by Klamt and his coworkers (see Klamt 1995, 2005 and 2011; and references therein).

A global presentation of the two main steps of the COSMO-RS algorithm, namely quantum COSMO calculations and thermodynamic calculations, are given below. The first step includes the cavity construction, electrostatic energy computations, and determination of the surface charge density. The second step includes the thermo-statistical treatment of pairwise segment interactions, the concept of σ-profiles, and ends up with the calculation of the chemical potentials of each molecule. The latter can be used to calculate the activity coefficients and all of its derivative properties, as discussed in Chapter I.

II.1.1. Quantum COSMO calculations

Continuum models were introduced more than a century ago, in very simplified versions, giving results of remarkable importance using computational instruments no more complex than a slide rule. These models have been used for more than 50 years, and in more detailed versions they are still in use, but the essential step, opening new perspectives for the study of solvent effects, has reached its merging with the QM descriptions of molecules. Continuum models are in fact the ideal conceptual framework to describe solvent effects within the QM approach. The initial stimulus was provided by the recognition that the QM description of the electrostatic potential generated by the charge distribution of a molecule could represent a valid analytic and interpretative tool to study intermolecular interactions. A continuum model in computational molecular sciences can be defined as a model in which a number of the degrees of freedom of the constituent particles (a large number, indeed) are described in a continuous way, usually by means of a distribution function (Tomasi et al. 2005).

The COSMO model belongs to the family of the continuum solvation model (CSM).

II.1.1.1. Electrostatic energy calculations

II.1.1.1.1. Basic idea and its development

The inverse-square Coulomb force between two charged atoms, or ions, is by far the strongest of the physical forces that are considered in electrostatic energy calculations (Israelachvili 2011).

The electric field $\left\|\overrightarrow{E_1}\right\|$ (expressed in V.m^{-1}) at a distance r away from a charge q_1 is defined by

$$\left\|\overrightarrow{E_1}\right\| = \frac{q_1}{4\pi\varepsilon_0\varepsilon r^2} \qquad \text{(II-1)}$$

where ε is the dielectric permittivity or constant of the medium. This field, when acting on a second charge q_2 at r, gives rise to a force F (expressed in N) known as the Coulomb force or Coulomb law:

$$F(r) = q_2 \left\| \overrightarrow{E_1} \right\| = \frac{q_1 q_2}{4\pi\varepsilon_0 \varepsilon r^2} \tag{II-2}$$

The free energy *w(r)* (expressed in J) of the Coulomb interaction between two charges q_1 and q_2 is therefore given by

$$\begin{aligned} w(r) &= \int_{\infty}^{r} -F(r)\,dr = -\int_{\infty}^{r} -\frac{q_1 q_2}{4\pi\varepsilon_0 \varepsilon r^2}\,dr = +\left[\frac{q_1 q_2}{4\pi\varepsilon_0 \varepsilon r}\right]_{\infty}^{r} \\ &= \frac{q_1 q_2}{4\pi\varepsilon_0 \varepsilon r} = \frac{z_1 z_2 e^2}{4\pi\varepsilon_0 \varepsilon r} \end{aligned} \tag{II-3}$$

where the "reference state" of zero energy is taken to be at $r=\infty$. The expression on the right of this equation is commonly used for ionic interactions in aqueous solutions where the magnitude and sign of each ionic charge is given in terms of the elementary electron charge (e=1.602*10^{-19} C) multiplied by the ionic valence z (Israelachvili 2011).

It is important to always have the right reference state in mind when considering intermolecular interactions. Any value for the energy is not very meaningful unless referred to some state with which it is being compared. Thus, when ions come together to form a condensed phase from the gaseous state, the reference state is at $r=\infty$, the interaction occurs in a vacuum (ε=1) and the dielectric constant effect can be omitted in the expression of the Coulomb interaction. On the other hand if two ions are interacting in a condensed liquid medium, the reference state is also at $r=\infty$, but the dielectric constant now appears in the denominator of any expression for the Coulomb interaction, since the interaction is now occurring entirely within the solvent medium (Israelachvili 2011).

When a single ion is in a vacuum or in a medium, even though it may not be interacting with other ions, it still has an electrostatic free energy associated with it. This energy is equal to the electrostatic work done in forming the ion that in a vacuum is referred to simply as the self-energy, while in a medium it is referred to as the Born or solvation energy of the ion. The Born energy is an important quantity since it determines, among other things, the extent to which ions will dissolve and partition in different solvents (Israelachvili 2011).

For describing Born energies, let us imagine the process of charging an atom or sphere of radius R^{ion} by gradually increasing its charge from zero to its full charge Q^{ion}. At any stage of this process let the ionic charge be q and let this be incremented by dq. The work done in bringing this additional charge from infinity to $r= R^{ion}$ is therefore, from equation II-3, putting $q_1=q$ and $q_2=dq$, and r= R^{ion},

$$dw = \frac{qdq}{4\pi\varepsilon_0\varepsilon R^{ion}} \tag{II-4}$$

so that the total free energy of charging the ion, μ_i, the Born energy, is

$$\mu_i = \int dw = \int_0^{Q^{ion}} \frac{qdq}{4\pi\varepsilon_0\varepsilon R^{ion}} = \frac{\left(Q^{ion}\right)^2}{8\pi\varepsilon_0\varepsilon R^{ion}} = \frac{\left(ze\right)^2}{8\pi\varepsilon_0\varepsilon R^{ion}} \tag{II-5}$$

The Born energy gives the electrostatic free energy of an ion in a medium of dielectric constant ε. It is positive because the energy is unfavourable; that is, it is the energy of keeping a net charge Q^{ion} distributed on the surface of a sphere against its own electrostatic repulsion (Israelachvili 2011).

The Born energy can also be obtained from the energy of the electric field of the ion. From basic electrostatic theory (Israealachvili 2011 and references therein), the free energy density of an electric field $\left\|\vec{E}\right\|$ arising from a charge or any distribution of charges is $\frac{1}{2}\varepsilon_0\varepsilon\left\|\vec{E}\right\|^2$ per unit volume. Thus, in general,

$$\mu_i = \frac{1}{2}\varepsilon_0\varepsilon\int\left\|\vec{E}\right\|^2 dV \tag{II-6}$$

And by integrating the energy density of an ion over all of space, we immediately obtain the Born energy by a different way as above:

$$\mu_i = \frac{1}{2}\varepsilon_0\varepsilon\int_{R^{ion}}^{\infty}\frac{\left(Q^{ion}\right)^2}{\left(4\pi\varepsilon_0\varepsilon r^2\right)^2}4\pi r^2 dr = \frac{\left(Q^{ion}\right)^2}{8\pi\varepsilon_0\varepsilon R^{ion}} \tag{II-7}$$

The change in free energy $\Delta\mu_i$ (expressed in J) on transferring an ion from a medium of low dielectric constant ε_1 to one of high dielectric constant ε_2 is negative, that is it is energetically favourable and equal to:

$$\begin{aligned}\Delta\mu_i &= -\frac{\left(ze\right)^2}{8\pi\varepsilon_0 R^{ion}}\left[\frac{1}{\varepsilon_1}-\frac{1}{\varepsilon_2}\right] \\ &= -\frac{28z^2}{R^{ion}}\left[\frac{1}{\varepsilon_1}-\frac{1}{\varepsilon_2}\right]kT\,(\text{per ion at } T\text{=300K})\end{aligned} \tag{II-8}$$

or

$$\Delta G = \mathcal{N}\Delta\mu_i = -\frac{69z^2}{R^{ion}}\left[\frac{1}{\varepsilon_1}-\frac{1}{\varepsilon_2}\right] \tag{II-9}$$

where R^{ion} is given in nanometers and ΔG is the molar free energy expressed in kJ.mol^{-1}. The latter equation provides the basis for calculating the partitioning of ions between different

solvents. Note that the Born energy does not include the energy expended by solvents in forming the cavities for accommodating the ions, these are generally small compared to the large Born energy (Israelachvili 2011).

The ion-induced dipole interaction is related to the Born energy of an ion in a medium. For this purpose, let us first compute the total interaction energy of an ion in a medium with all the surrounding solvent molecules (Israelachvili 2011). For an ion of radius R^{ion}, this is given as:

$$\mu_i = -\frac{\rho\alpha\left(ze\right)^2}{8\pi\varepsilon_0{}^2\varepsilon^2 R^{ion}} \tag{II-10}$$

where ρ is the number of solvent molecules per unit volume (the number density). To proceed further, we have to make some connection between the molecular and continuum properties of the solvent. This requires finding a relation between the molecular polarizability α and the dielectric constant ε of a medium. The problem is not still completely understood. However, the value of $(\rho\alpha)/\varepsilon_0$ in the latter equation may, in a first approximation, be associated with the electric susceptibility χ of a medium. This is the polarizability per unit volume of a medium and is related to the dielectric constant ε by

$$\chi = \left(\varepsilon - 1\right) \tag{II-11}$$

The change in free energy $\Delta\mu_i$ when an ion goes from a medium of dielectric constant ε_1 to one of ε_2 is therefore (since $d\chi=d\varepsilon$)

$$\Delta\mu_i = -\int_{\chi_1}^{\chi_2} \frac{\left(ze\right)^2}{8\pi\varepsilon_0\varepsilon^2 R^{ion}} d\chi = -\int_{\varepsilon_1}^{\varepsilon_2} \frac{\left(ze\right)^2}{8\pi\varepsilon_0\varepsilon^2 R^{ion}} d\varepsilon = -\frac{\left(ze\right)^2}{8\pi\varepsilon_0 R^{ion}}\left[\frac{1}{\varepsilon_1}-\frac{1}{\varepsilon_2}\right] \tag{II-12}$$

which is the change in Born energy, equation II-8.

Thus, the Born energy can be derived either from a continuum analysis or from a molecular approach. Furthermore, the molecular approach has the added advantage of providing insight into the limitations of the Born equation.

To understand the role of the solvent in Coulomb interactions, we must investigate how a solvent affects the electric fields around dissolved ions. Ultimately, we shall have to consider the origin of the dielectric permittivity ε since it defines the solvent in all equations for electrostatic interactions. As already mentioned above, one can analyze such phenomena in terms of continuum and/or molecular theories. It was shown that the Born energy may be obtained by integrating the electric field energy over all of space (equation II-7). The derivation of the latter equation provides some important insights:

- First, the electrostatic self-energy is seen not to be concentrated on the ion itself, but rather it is spread out over the whole space around the ion.

- Secondly, it is shown that the Coulomb interaction in a medium can also be derived from the change in the electric field energy, integrated over the whole of space, when two charges are brought together. The Coulomb interaction can therefore be considered as the change in the Born energies of two charges as they approach each other. This is conceptually important, since it shows that the Coulomb interaction in a medium is not determined by the dielectric constant in the region between two charges but by its value in the region surrounding (as well as between) charges. It is for this reason that the strength of the Coulomb interaction of two oppositely charged ions will be reduced in a solvent medium even if the ions still remain in contact that is even before there are any solvent molecules between them.

It has been now established how a solvent medium affects the electrostatic Born and Coulomb energies of ions. The value of ε of the locally surrounding medium is important; if the standard expressions for the Born and Coulomb energies (and other interaction energies that depend on ε) are to apply, the bulk value of ε must be attained already within the first shell of surrounding solvent molecules. For ε = 1 (in vacuum or in gas phase) there is no polarization field. For ε>>1 (in water or in a conducting medium), the polarization field almost balances the applied field. Electrostatic forces are therefore much reduced in media that have high dielectric constants. As could be expected, the dielectric constant ε is related to the polarizability α, higher values of α resulting in higher values for ε. This relation connects the molecular and continuum electric properties of matter (Israelachvili 2011).

As illustrated in Figure II-2, the charge distribution is evaluated in different ways, using a linear distribution, a surface distribution or a volume distribution.

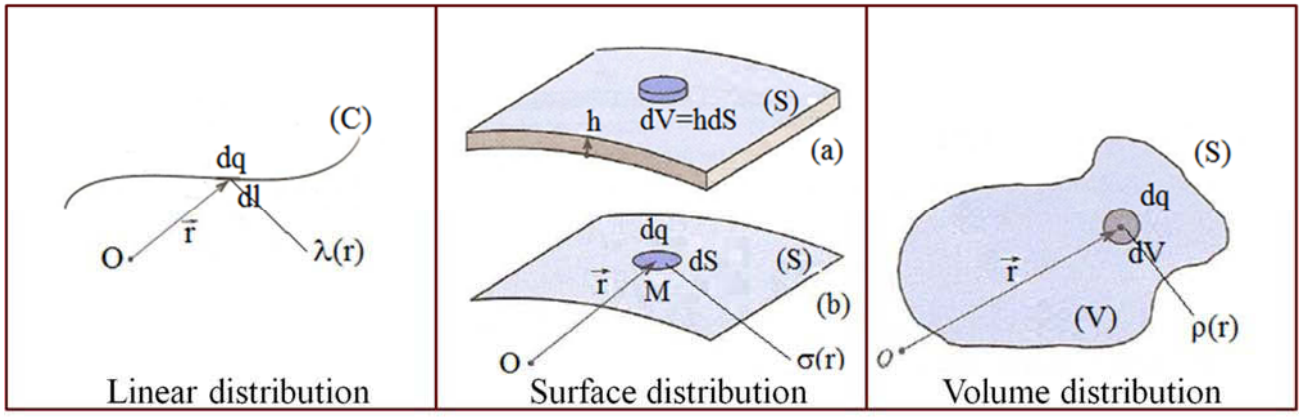


Figure II-2 : Illustration of the different types of charge distribution.

The linear charge density (λ) is the ratio of an infinitesimal electric charge *dq* (expressed in C) to an infinitesimal line element *dl* (expressed in m),

$$\lambda = \frac{dq}{dl} \qquad \text{(II-13)}$$

Similarly, the surface charge density (σ) uses a surface area element *dS*,

$$\sigma = \frac{dq}{dS} \tag{II-14}$$

And the volume charge density (ρ) uses a volume element *dV*,

$$\rho = \frac{dq}{dV} \tag{II-15}$$

Integrating these definitions gives the total charge *Q* of a region according to line integral of the charge density over a line of length *L*, or the surface integral of the surface charge density over a surface *S*, or the volume integral of the volume charge density over a volume *V*.

$$Q = \int_L \lambda(r)\,dl = \iint_S \sigma(r)\,dS = \iiint_V \rho(r)\,dV \tag{II-16}$$

All these repartitions give access to the electric field $\vec{E}$:

$$\vec{E} = -\overrightarrow{grad}\Phi \tag{II-17}$$

where ϕ denotes the electrostatic potential. The energy *w(r)* of such charges distribution, which is also denoted as *E*, is expressed as:

$$E = w(r) = \frac{1}{2}\iiint_V \rho\,\Phi\,dV = \frac{1}{2}\iint_S \sigma\,\Phi\,dS = \frac{1}{2}\int_L \lambda\,\Phi\,dl \tag{II-18}$$

These electrostatic energies can be used to calculate the external potential in quantum calculations in a medium owning a dielectric solvent ε. The continuum models, which were developed in order to describe such electrostatic behaviour of solvents, have their origin in simple physical considerations. The attention has been focused since the very beginning on a microscopic description of one component of the system (*i.e.* the "solute"). For this purpose, it is assumed that forces are quantum mechanical in origin and amenable to a host of theoretical treatments of varying complexity (Israelachvili 2011). Their origin may be understood intuitively as follow:

- for a non-polar atom such as helium, the time average of its dipole moment is zero, but at any instant there exists a finite dipole moment given by the instantaneous positions of the electrons about the nuclear protons;
- this instantaneous dipole generates an electric field that polarizes any nearby neutral atom, inducing a dipole moment in it;
- the resulting interaction between the two dipoles gives rise to an instantaneous attractive force between the two atoms, and the time average of this force is finite.

For a simple semi-quantitative understanding of how these forces arise, we may consider the following model based on the interaction between two Bohr atoms (Israelachvili 2011).

In the Bohr atom, an electron is pictured as orbiting around a proton. The smallest distance between the electron and proton is known as the first Bohr radius a_0 and is the radius at which the Coulomb energy $e^2/4\pi\varepsilon_0 a_0$ is equal to $2h\nu$, that is

$$a_0 = \frac{e^2}{8\pi\varepsilon_0 h\nu} = 0.053\,\text{nm} \qquad \text{(II-19)}$$

where h is the Plank constant and ν the orbiting frequency of the electron. In the literature $h\nu$ is often expressed as $\hbar\omega$, where $\hbar=h/2\pi$ and $\omega=2\pi\nu$. For a Bohr atom, ν = 3.3*10^{15} s^{-1}; so that $h\nu$= 2.2*10^{-18} J. This is the energy of an electron in the first Bohr radius and is equal to the energy needed to ionize the atom, *i.e.* the ionization potential I (Israelachvili 2011).

The Bohr atom has no permanent dipole moment. However, at any instant there exists an instantaneous dipole moment, $u = a_0 e$, whose field will polarize a nearby neutral atom, giving rise to an attractive interaction that is entirely analogous to the dipole induced dipole (Debye) interaction. The energy of this interaction in a vacuum will therefore be given by:

$$E = w(r) = -\frac{u^2\alpha_0}{\left(4\pi\varepsilon_0\varepsilon\right)^2 r^6} = -\frac{\left(a_0 e\right)^2\alpha_0}{\left(4\pi\varepsilon_0\right)^2 r^6} \qquad \text{(II-20)}$$

where α_0 is the electronic polarizability of the second Bohr atom, which is approximately equal to $4\pi\varepsilon_0 a_0^3$. Using this expression for α_0 and the value of a_0, we immediately find that the preceding interaction energy can be written approximatively as

$$E = w(r) \approx -\frac{{\alpha_0}^2 h\nu}{\left(4\pi\varepsilon_0\right)^2 r^6} \qquad \text{(II-21)}$$

which is similar to the expression derived by London in 1930 using quantum mechanical perturbation theory. The latter has been superseded by more exact, though more complicated expressions, but it can relied upon to give fairly accurate values for interactions in a vacuum although these are usually lower than more rigorously determined ones (Israelachvili 2011).

The basic electrostatic concepts discussed above are used in the classical electrostatics view of interactions; in the quantum calculation tools (like COSMO) another unit-convention is used to compute the analogue of the Born energy as discussed below.

II.1.1.1.2. Conductor-like screening model (COSMO)

The expressions given by Born in 1920 and Bell in 1931 for the classical interaction energy of a simple solute with a medium represented as a continuous dielectric were formally extended by Kirkwood to quantum descriptions of solutes in 1934, without limits in the complexity of the system. The decisive contribution of Onsager in 1936 was to provide an interpretative tool, used by

chemists for many years: the simplicity of the formal expressions successively elaborated has stimulated application to various solvent effects (Klamt 2005). Reviews on such continuum solvation models (CSMs) have been given by Tomasi and Persico (1994), Cramer and Truhlar (1999), and Tomasi et al. (2005). The Conductor-like Screening Model COSMO introduced by Klamt and Schüürmann in 1993, in principle, is a more efficient variant of these CSMs.

CSMs usually require the solution of the complex boundary conditions for a dielectric in order to obtain the screening charges. COSMO instead uses the much simpler boundary condition of vanishing electrostatic potential for a conductor,

$$\Phi^{tot} = 0 \tag{II-22}$$

This represents an electrostatically ideal solvent with ε=∞ (Klamt 2005).

When decomposing the surface into a set of surface elements, the value of the electrostatic potential is specific of each element, which in turns corresponds to a set of values of electrostatic potentials. Adopting a matrix notation, the total electrostatic potential on the cavity surface segments is determined by the solute potential Φ^{sol}, which consist of the electronic and the nuclear parts, and the contribution of the screening charges q,

$$\Phi^{tot} = \Phi^{sol} + Aq = 0 \tag{II-23}$$

where *A* is the Coulomb interaction matrix of the surface segments, *i.e.*,

$$A_{ij} = \frac{1}{\left\| \vec{t}_i - \vec{t}_j \right\|} \text{ and } A_{ii} \cong 1.07 s_i^{-0.5} \tag{II-24}$$

with $\vec{t}_i$ and $\vec{t}_j$ being the respective positions of segments *i* and *j*, thus $\left\| \vec{t}_i - \vec{t}_j \right\| = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$ if *x*, *y* and *z* denotes the Cartesian coordinates.

For a conductor, the boundary condition $\Phi^{tot} = 0$ defines the screening charges as

$$q = -A^{-1}\Phi^{sol} \tag{II-25}$$

Since matrix *A* and its Cholesky factorisation, which is needed for an efficient solution of Eq. (II-23), only have to be calculated once for each geometry, the additional computational costs of COSMO in the quantum chemical algorithms are small. In addition, COSMO may even slightly speed up the quantum chemical self-consistency part. Thus, COSMO calculations are usually no more expensive than gas phase quantum chemistry. At the end of the quantum chemical self-consistency and geometry optimisation loops, a self-consistent state is reached, *i.e.* energy, density, and geometry of the solute X are as if embedded in a dielectric of permittivity ε. In addition, the screening charge density σ supplied by the continuum on each position of the molecular contact surface and the dielectric interaction energy of the solute with the continuum is known now. The

dielectric energy is an excellent measure for the polarity of molecules, much better than dipole moment, because it subsumes all ways of electrostatic interactions of a molecule with its surrounding (Klamt 2005).

The COSMO theory avoids the complicated solution of the dielectric boundary conditions by approximating the screening charges q of a dielectric medium of permittivity ε with the scaled screening charges of a conductor, where the scaling factor is

$$f(\varepsilon) = \frac{\varepsilon - 1}{\varepsilon + \frac{1}{2}} \tag{II-26}$$

$$q^* = f(\varepsilon) q \tag{II-27}$$

Thus, only the much simpler boundary condition of a conductor appears in the equations. The COSMO approximation is exact in the limit of $\varepsilon = \infty$ and it is within 0.5% accuracy for strong dielectrics like water ($\varepsilon = 80$). Even in the lower dielectric limit of solvents, which typically is $\varepsilon = 2$, COSMO can be shown to coincide with the exact dielectric model within 10% in all relevant cases. Due to its inherent technical advantages, the COSMO approach is going to become the standard CSM in quantum chemical codes (Klamt 2005).

The dielectric energy (E_{diel}), *i.e.* the free electrostatic energy gained by the solvation process, is half of the solute-solvent interaction energy.

$$E_{diel} = \frac{1}{2} f(\varepsilon)\, q\, \Phi^{sol} \tag{II-28}$$

The total free energy of the solvated molecule is the sum of the energy of the isolated system calculated with the solvated wave function and the dielectric energy

$$E = E\left(\Psi^{solv}\right) + E_{diel} \tag{II-29}$$

A COSMO energy calculation starts with the construction of the cavity surface grid. Within the Self Consistent Field (SCF) procedure, the screening charges are calculated in every cycle and the potential generated by these charges is included into the Hamiltonian. This ensures a variational optimization of both the molecular orbitals and the screening charges and allows the evaluation of analytic gradients (Klamt 2005).

II.1.1.2. Cavity construction

The cavity is a basic concept in all continuum solvation models. The model is composed of a molecule that is the solute, put into a void cavity within a continuous dielectric medium mimicking the solvent. The efficient construction of proper and sufficiently accurate segmentations of a molecular-shaped cavity is an important technical aspect of apparent surface charge models,

because it has a strong influence on both the accuracy and speed of the calculations (Tomasi et al. 2005).

Given the coordinates and atomic numbers of all N atoms of a molecule, the usual concept for the definition of a continuum solvation model (CSM) cavity starts with the assignment of minimum distances, R_i, for all atoms i. This assignment can either be according to the atomic numbers, using element-specific radii, or according to more details of the chemical environment, *i.e.*, by the introduction of atom-type-specific radii. The latter procedure obviously introduces more adjustable parameters, with the potential of a more accurate reproduction of experimental data, but also with the danger of loss of the predictive power of the model. Some workers even introduce a dependence of the radii upon the partial charge of the atom or the total charge of the molecule, but this leads to obvious consistency problems (Tomasi et al. 2005, and references therein).

The shape and size of the cavity are differently defined in the various versions of the continuum models. As a general rule, a cavity should have a physical meaning, such as that introduced by Onsager, and not be only a mathematical artifice as often happens in other descriptions of solvent effects. In particular, the cavity should exclude the solvent and contain within its boundaries the largest possible part of the solute charge distribution. The optimal size of the cavity has thus been a subject of debate, and several definitions have been proposed. The adopted definitions are the result of a trade-off between conflicting physical requirements. The shape of the cavity has also been the object of many proposals. It is universally accepted that the cavity shape should reproduce as well as possible the molecular shape. Cavities not respecting this condition may lead to deformations in the charge distribution after solvent polarization, with large unrealistic effects on the results, especially for properties. Here, once again, there is a trade-off between computational exigencies and the willing for better accuracy. Computations are far simpler and faster when simple shapes are used, such as spheres and ellipsoids, but molecules are often far from having a spherical or ellipsoidal shape. Quantum mechanical calculations of the molecular surface can give a direct *ab initio* definition of the cavity (Tomasi et al. 2005).

Molecules often have an irregular shape, and the occurrence of small portions of space on their periphery where solvent molecules cannot penetrate is not a rare event. This intuitive consideration is at the basis of two definitions, those of solvent-excluding and solvent-accessible surface (SES and SAS, respectively). Both introduce in the surface (and in the volume) changes to the vdW description related, in a different way, to the finite size of the solvent molecules. In both cases, the solvent molecule is reduced to a sphere, with a volume equal to the vdW volume (other definitions of this radius have been used, but this seems to be the most consistent definition). The positions assumed by the centre of a solvent sphere rolling on the vdW surface of the solute define the SA surface, that is, the surface enclosing the volume in which the solvent centre cannot enter.

The same sphere used as a contact probe on the solute surface defines the SE surface, that is, the surface enclosing the volume in which the whole solvent molecule cannot penetrate. In the literature, the SES is also called "smooth molecular surface" or "Connolly surface", due to Connolly's fundamental work in this field. Indeed, the SE surface developed by Connolly can be considered to be the prototype for the computational study of molecular surfaces. A schematic drawing of the different surfaces for the same molecule is given in Figure II-3 (Tomasi et al. 2005).

After assigning the radii, the union of the corresponding atom-centred spheres is considered as the interior of the molecule.

In the COSMO implementation of the TURBOMOLE software (that is used in this work to perform quantum calculations) a double grid approach and segments of hexagonal, pentagonal, and triangular shapes is used in order to ensure a sufficiently accurate and efficient segmentation of the molecular shaped cavity. The construction of this cavity starts with a union of spheres of radii $R_i=R_i^{vdW}+R^{solv}$ for all atom i; where R_i^{vdW} denotes the van der Waals radius of this atom and R^{solv} is the solvent accessible radius which describes the geometry of the solvent molecules. The following steps are done for each atom, in order to ensure that each segment of the segment accessible surface (SAS), schematized in Figure II-4, is connected to a single solute atom (Klamt and Schüürmann 1993, Klamt 2005).

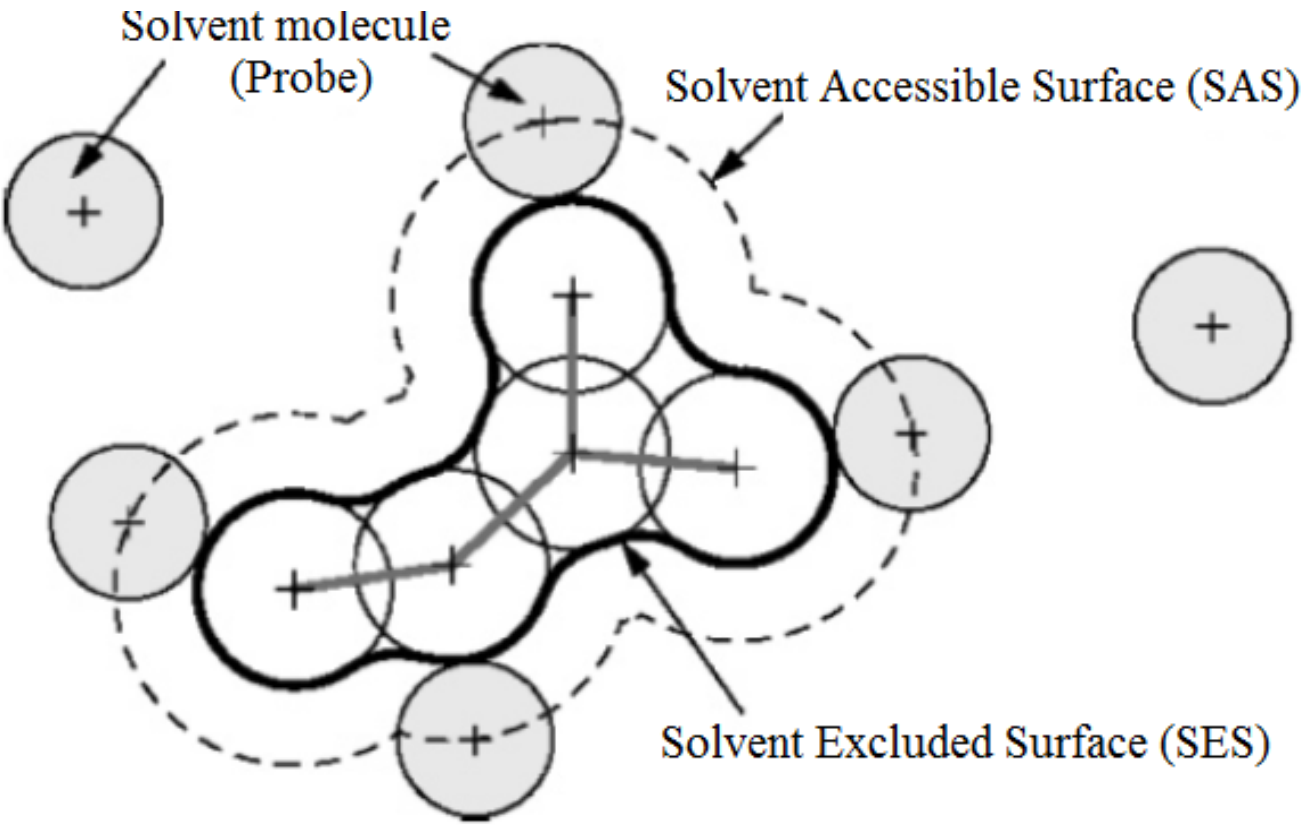


Figure II-3 : Solvent accessible surface (SAS) traced out by the centre of the probe representing a solvent molecule (Adapted from Tomasi et al. 2005).

The solvent excluded surface (SES) is the topological boundary of the union of all possible probes that do not overlap with the molecule.

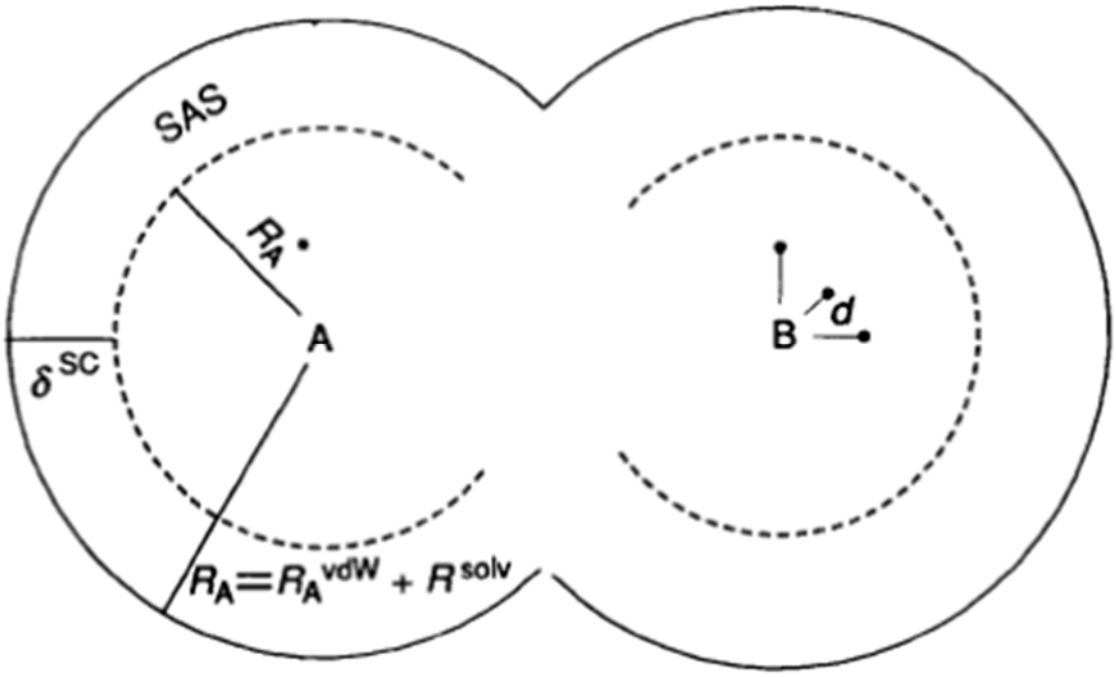


Figure II-4 : Schematic construction of the Solvent accessible surface (SAS) (Adapted from Klamt 2005).

Solid circles indicate the surface accessible to the centres of solvent molecules, dashed lines the surface accessible to solvent charges, where R_i^{vdW} denotes the van der Waals radius of this atom and R^{solv} is the solvent accessible radius which describes the geometry of the solvent molecules.

II.1.1.3. More details about segmentations

As a next step, the resulting surface is divided into segments. For the subsequent calculation of the electrostatic interactions, it is necessary to know the area, shape and position of each segment. It is achieved reasonably homogeneous segment areas in order to reduce the total number, *M*, of segments on the molecule, because the time requirements of the algorithms for the solution of the CSM boundary equations typically scale with the third power of M and the memory-demand scales with M² (Klamt 2005).

The cavity construction in the original COSMO implementation in MOPAC93 (called OLDCAV, according to MOPAC keyword), starts with a union of spheres of radius $R_i = R_i^{vdW} + R^{solv}$ for all atom *i*. Then each sphere is triangulated, starting from a regular icosahedron with 20 triangles as an initial triangular grid. Then two different refinement steps are used. The first is the addition of the triangle edge-midpoints as new vertices. This increases the number of triangles by a factor of four, as in the most widely used method for CSM cavity construction, *i.e.* the GEPOL algorithm

(Klamt 2005). The second is the addition of the triangle centres as new vertices, leading to a new set of triangles and increasing the number by a factor of three. One should notice that it is possible to generate triangulations with $k = 20*3^i*n^2$, where *i* being 0 or 1, and *n* any positive integer. The lower values of *k* thus are 20, 60, 80, 180, 240, 320...(Klamt 2005).

Finally, the triangles are not considered as segments, but the corresponding hexagons (and 12 pentagons), which are got when each vertex of the triangular grid is considered as a centre, and all triangle midpoints of the six or five surrounding triangles of the vertex are connected. Thus, we use the 12 faces of the dodecahedron as segments in the case of *k*=20 triangulation, and the 32 faces of the buckyball for *k*=60. Indeed, the explicit construction of the hexagons and pentagons is not required, because only the knowledge of the positions of their centres and the corresponding surface area are necessary to do so. At the same corner-centre distance, the number of hexagons and pentagons is $k' = \frac{k}{2} + 2$, *i.e.*, approximately half of the number of triangles, leading to sphere segment numbers of $k' = 10*3^i*n^2 + 2 = 12, 32, 42, 92, 122, \ldots$ In addition, during the evaluation of the electrostatic interaction matrix the centre approximation is better justified for these near-circular segments than for triangular segments. In order to achieve a fine sampling of the surface, a basis grid of *NPPA* (number of points per full atom) equal to 1082 is initially projected on each of the atom spheres of radius $R_i^{vdW} + R^{solv}$. Then all points in the interior of another sphere are discarded. Next, all remaining points are projected downward to the sphere of radius R_i. By this construction, no points end up in the problematic smoothing regions. Next, the points on each sphere are connected to form larger segments. For this a grid of *NSPA* (number of segments per full atom) segment centres is projected on the inner sphere where *NSPA* is one of the values $k' = 10*3^i*n^2 + 2$. Now the points from the basis grid are associated to their nearest grid centres, and those grid centres getting zero points are discarded. Then all segment centres are redefined as the centre of area of their associated points. Finally, this association procedure is repeated once in order to achieve nearest-neighbour association after the move of the centres. Finally, each segment is characterized by a centre position $\vec{t}_i = (x_i, y_i, z_i)$, a total area s_i, and a set of basis grid points with associated positions and areas (Klamt 2005).

Then the Coulomb interaction matrix elements, A_{ij}, are calculated as a sum of the corresponding partial contributions of the associated basis grid points of segments *i* and *j*, if their distance is below a threshold *DISEX*. The centre approximation is used otherwise. The diagonal elements A_{ii} are calculated in the same way using the basis grid points, and even for self-energies of the basis grid points a reasonable approximation of $1.07*s^{-0.5}$ is used, where *s* denotes the area represented by a basis grid point. In this way the A-matrix for the segments is evaluated with very high precision,

while keeping the number of segments *M*, which is finally crucial for the numerical cost of the COSMO algorithm, at a relatively low level of *~N*NSPA/3*. In semi-empirical quantum methods a default value of *NSPA=42* is suggested for non-hydrogen atoms. For hydrogen atoms, a coarser segmentation with a default value of 12 is considered. Due to the rigor used for the sphere segments, a crude approximation was made for the smoothing region: any segments are not put into these regions (which are kept electrostatically open). Indeed, this worked surprisingly well and generated good correlations of the calculated solvation energies with the experimental results. Nevertheless, Klamt and his co-workers later realized the need for a surface closure, because some small segments could get extremely high values of s if they were isolated by open regions from the rest of the cavity. To avoid such artefacts, a surface-closure algorithm has been developed that is the standard for the COSMO cavities (NEW-CAV) today. No changes were made for the spherical part. For the surface closure, first all pairs of intersecting solvent-increased spheres generate a ring. If this ring is partially within other spheres, the corresponding angle regions are marked as invalid. Then R_{solv} projects the ring downward toward each of the two atom centres, leading to two opposing rings. The valid regions of these rings are now filled with a series of triangles, each having two corners on one of the rings and one on the other. Finally, the sole corner of each triangle moves a little toward the centre of the ring that holds the two other corners of the triangle. By that procedure the triangles become slightly inclined, mimicking the shape of the Connolly surface. The degree of inclination is a function of the solvent radius, the two atom radii, and their distance. In this way, a continuous and differentiable behaviour of the surface between two atoms is achieved, even in the difficult case close to their complete dissociation. Since the change of the surface according to a change of the distance between the atoms can be approximated reasonably well by a stretching and change of inclination of these triangles, the derivative of the surface area with respect to geometry can be calculated from the analytic behaviour of these triangles. In the final step, the triangular regions belonging to triple points, *i.e.*, intersections of three spheres get paved with additional triangle. Each of the triangles generated in this way is considered as an individual segment (Klamt 2005).

For *NSPA*=92 and *NSPAH*=32, which is the present recommendation for DFT calculations, the surface-closure unfortunately enlarges the initial number of segments by approximately 70%. Nevertheless, the corresponding increase in calculation time is considered as justified by the resulting increase in physical consistency (Klamt 2005).

Since all segments can be considered to move approximately with their nearest atoms and area-changes have to be taken into account for the closure triangles, the derivative of the **A**-matrix with respect to the geometry is quite simple. This derivative is needed for the calculation of analytic derivatives of the COSMO interaction energy. In addition to good analytic gradients, a stable cavity

construction algorithm is required for a good convergence of the geometry optimization algorithm. Artificial fluctuations of the total energy caused by small changes of the geometry must be kept as small as possible, although they can never be completely avoided in a discretized cavity model. The COSMO algorithm presented here, works satisfactorily in this regard, achieving geometry convergence in about the same number of steps as required for gas-phase geometry optimizations (Klamt 2005).

II.1.1.4. Concluding remarks about the COSMO calculations in the COSMO-RS framework

As discussed above, for each molecule or more precisely for each type of molecules, the first step of the COSMO-RS flowchart is to perform a COSMO calculation in a virtual conductor environment owning a dielectric constant $\varepsilon = \infty$. During this calculation, the COSMO cavities are defined in the usual way by smoothed spheres with atomic radii, which are about 20% larger than the usual van der Waals radii. Such cavities have proved to give the most reliable results in dielectric continuum models. For the following discussion it is important to note that the volume of such COSMO cavities corresponds approximately to the liquid molar volume of the components in standard conditions (Klamt 2005).

In the virtual conductor (COSMO) environment, the solute molecule induces a polarization charge density σ on the interface between the molecule and the conductor, *i.e.* on the molecular surface. These charges act back on the solute and generate a more polarisable electron density than in vacuum. During the quantum chemical (QC) self-consistency algorithm, the solute molecule is thus converged to its energetically optimal state in a conductor with respect to electron density. The molecular geometry is optimized using the same quantum methods as those used for calculations in vacuum (ε=1), as illustrated in Figure II-5. As a result of these COSMO calculations the total energies of the molecules are calculated in the virtual conductor, and the distribution of conductor polarization charges, *i.e.*, the polarization charge density, σ, on each point of the COSMO surface is determined (Klamt 2005).

When combined with accurate quantum chemistry methods like density functional theory (DFT) or correlated *ab initio* methods, CSMs prove to give very reasonable results for Henry law constants for air/water and air/alkane systems (Tomasi and Persico 1994). Successful application to the prediction of partition coefficients has been reported as well (Klamt 2005 and references therein). But despite this remarkable success, CSMs are not capable of a general description of the activity of molecules in solvents, because they are missing any idea of statistical thermodynamics and thus, cannot simulate general solvents, different temperatures, and mixtures. Even more, the application of the concept of the linear response theory of dielectrics to the situation of polar solutes in polar

solvents is untenable because the interactions in such systems are much too strong to be treated by with a linear response theory (Klamt 2005).

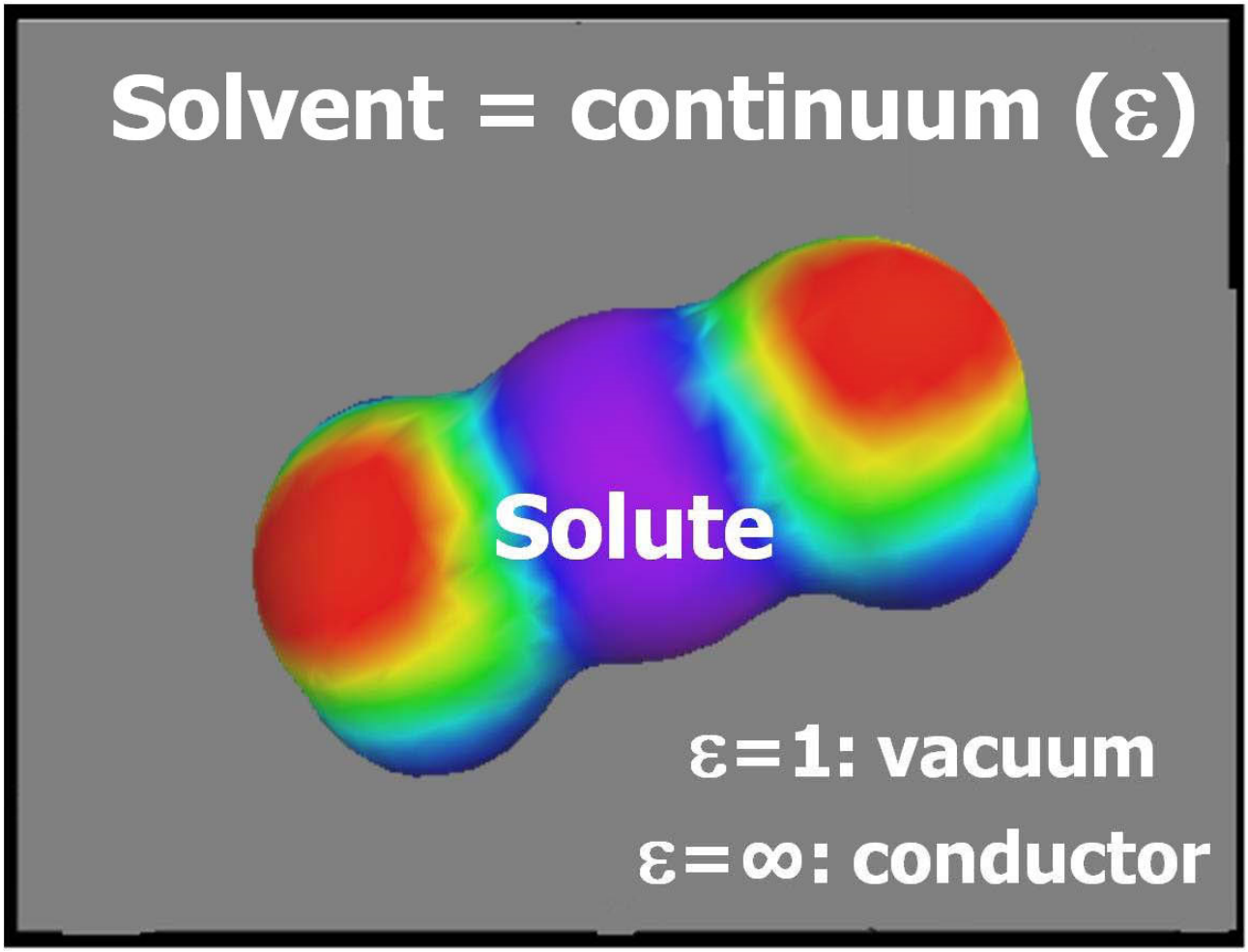


Figure II-5 : Illustration of the principle of the COSMO calculation.

In order to overcome these limitations of CSMs, the premises of the COSMO-RS interaction energy concept are threefold:

- CSMs like COSMO give very reasonable results for solvation in water and alkanes;
- The dielectric theory used in CSMs is untenable; and
- Concepts of temperature and mixture are missing in CSMs.

In this situation, COSMO-RS uses the self-consistent state of molecules embedded in a virtual conductor ($\varepsilon = \infty$), as it is readily achievable by COSMO calculations, as a new reference state for molecules in solution, without assuming that this state is physically existing in reality. There can be no doubt that this state of ideal screening is much closer to the real situation of molecules in solution, at least in polar solvents, than the usually chosen reference state of molecules in vacuum, because it takes account of the electrostatic screening of molecules by their surrounding, and even of the resulting back-polarization of the solutes. It is important to be able to calculate the state of the

molecules in vacuum using the same quantum chemistry method because this will enable to calculate the energy difference between the gas phase and the condensed phase, which is needed for vaporization data (Klamt 2005).

II.1.2. Thermo-statistical calculations

The polarization charge density of the COSMO calculation, also called screening charge density, is a good local descriptor of the molecular surface polarity. It is used to extend the dielectric continuum (COSMO) model towards "Real solvents" (COSMO-RS). The quantity σ can thus be used to quantify and colour-code molecular polarity on the surface (Klamt 2005). Such colour-coded-σ-surfaces of several molecules are given in Figure II-6.

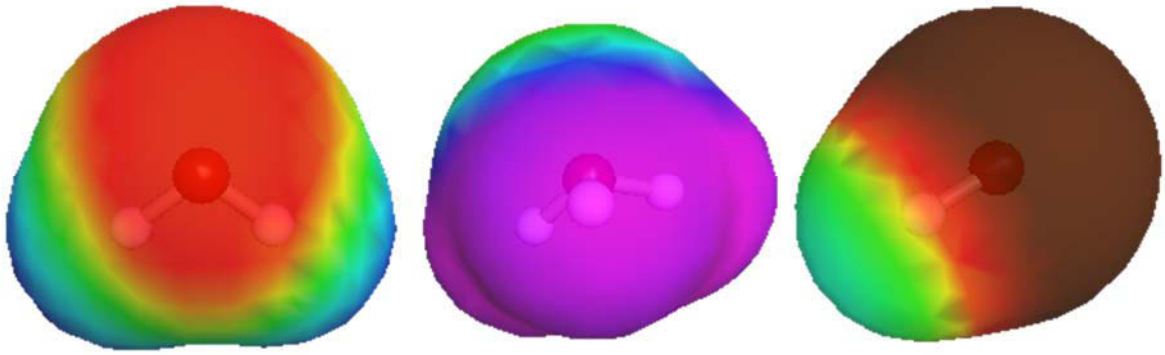


Figure II-6 : COSMO-surfaces of several molecules (water, hydronium ion H_3O^+ and hydroxide OH^-) colour coded by the polarization charge density σ.
Red areas denote strongly negative parts of the molecular surface and hence strongly positive values of σ. Deep blue marks denote strongly positive surface regions (strongly negative σ) and green denotes non-polar surface.

II.1.2.1. σ-averaging

In principle, the charge density q_i/s_i from COSMO could be used as the value of σ on a segment *i*. However, the COSMO segments have areas that range from $5*10^{-3}$ nm² down to about $5*10^{-5}$ nm². Thus the COSMO charge densities are much more local than the average values on the effective contact surfaces, a_{eff} =0.0767 nm². Hence it appears to be reasonable to use values of σ that are averaged over larger areas. For this purpose, an averaging radium r_{av} was introduced in the COSMO-RS algorithm (Klamt 2005). Then the polarization charge density can be defined as a local average given by:

$$\sigma_i = \frac{\sum_j \frac{q_j}{s_j + s_{av}} \exp\left(-\frac{-d_{ij}^{\,2}}{r_{av}^{\,2}}\right)}{\sum_j \frac{s_j}{s_j + s_{av}} \exp\left(-\frac{-d_{ij}^{\,2}}{r_{av}^{\,2}}\right)} \qquad \text{(II-30)}$$

where s_{av} is the area of a circle of radius r_{av}. Compared to the most trivial averaging, this slightly more complicated formula better takes into account the very different finite areas s_j of the segments j contributing to the average. The optimum value $r_{av} \cong 0.05\,\text{nm}$ is used in all COSMO-RS parameterization since 1998 (Klamt et al. 1998, Klamt 2005). The other parameters are directly taken or calculated from the COSMO-file which contains all the details about the surface segments of a given molecule.

II.1.2.2. σ-profiles

Since the interaction energies of the surfaces depend only on the local polarization charge-densities σ, only the net composition of the surface of a molecule X with respect to σ is of importance for the statistical thermodynamics of local pair-wise surface interactions (even if this is just a quite good approximation that has been improved in the COSMO-RS algorithm by Klamt (2005) and his co-workers). Thus, the full 3D information about σ on molecular surface is converted to a histogram, the so-called ***σ -profile*** of the molecule X_i, denoted as $p^{X_i}(\sigma)$. The σ-profile tells us how much of a surface we find in a polarity interval [σ-$d\sigma/2$, σ+$d\sigma/2$] (Klamt 2005). As an example, let us consider water. Its surface polarization charge density and the resulting σ-profile are shown in Figure II-7.

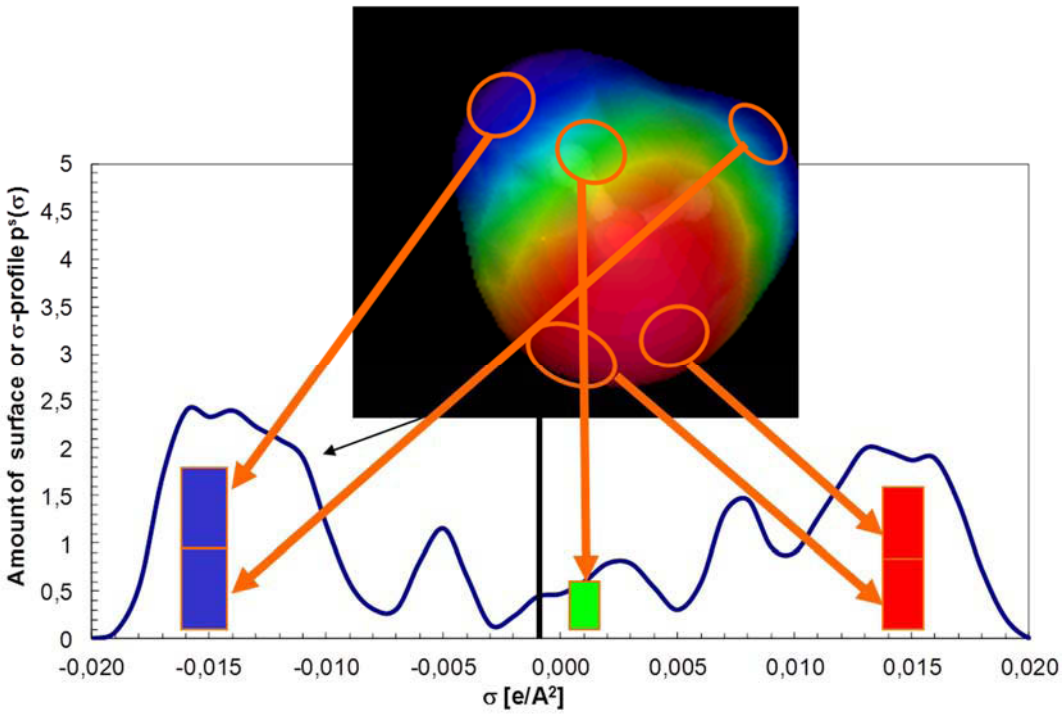


***Figure II-7*: Illustration of the conversion of the 3D coloured surface to σ-profile, for water (Adapted from Klamt 2005).**

Red areas denote strongly negative parts of the molecular surface and hence strongly positive values of σ. Deep blue marks denote strongly positive surface regions (strongly negative σ) and green denotes non-polar surface.

The entire σ-profile of water spans the range of σ values generally found for stable organic and inorganic molecules, including most ions. It is dominated by two major peaks arising from the strongly negative polar regions of the electron lone-pairs of the oxygen atom and from the strongly positively polar hydrogen atoms, respectively. In the colour coding of the surfaces these regions can be recognized as clearly as deep red and deep blue. Note that owing to the sign inversion of the polarization charge density σ, compared to the molecular polarity, the peak from the negative lone-pairs is located on the right side of the σ-profile at about 1.5e/nm², while the peak arising from the positively polar hydrogens is located on the left side, at about -1.5e/nm². One important feature of the σ-profile of water is its remarkable symmetry with respect to σ. There is about the same amount of strongly positive and equally strong negative surface area (Klamt 2005).

The polarization charge densities and σ-profiles of several molecules are given in Figure II-8 to Figure II-10, where the σ-profile of water is also given since the latter acts as a type of reference.

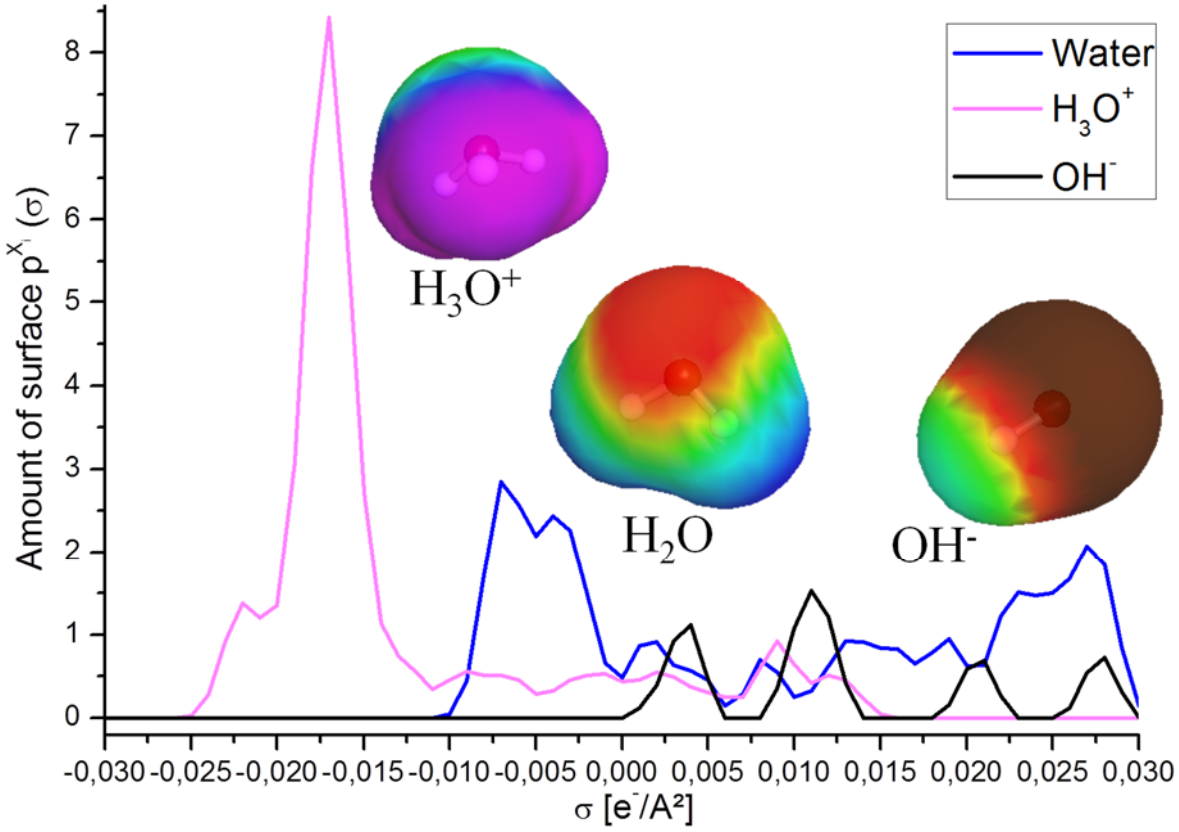


Figure II-8 : σ-profiles of water (H_2O), Hydronium ion (H_3O^+) and hydroxide (OH^-).

Red areas denote strongly negative parts of the molecular surface and hence strongly positive values of σ. Deep blue marks denote strongly positive surface regions (strongly negative σ) and green denotes non-polar surface.

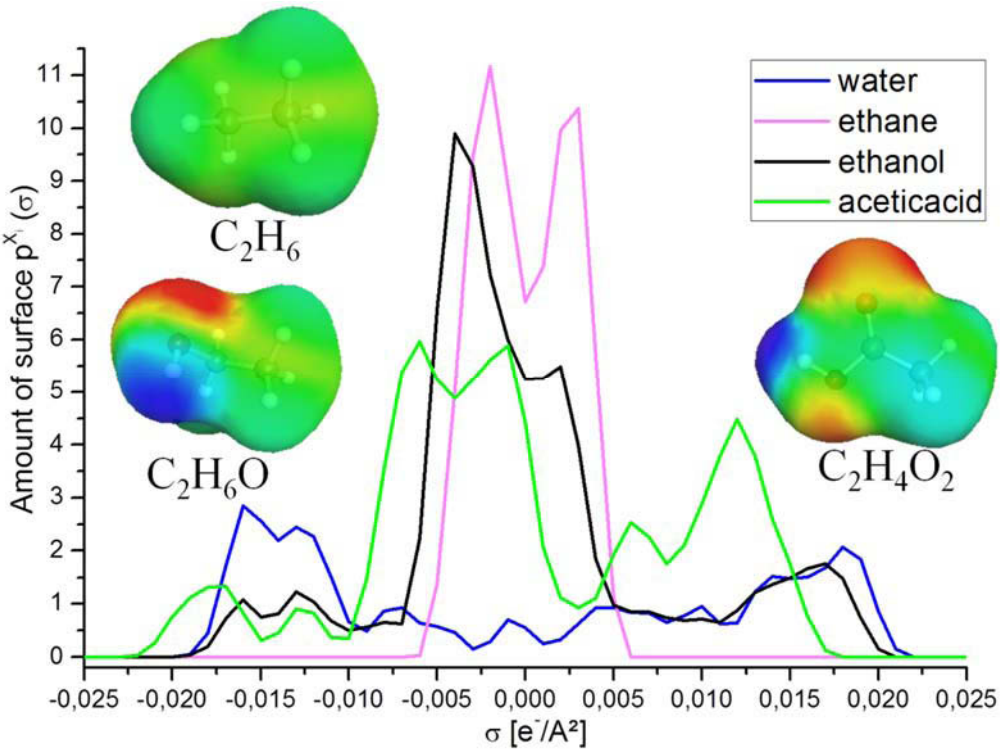


Figure II-9 : σ-profiles of water (H_2O), ethane (C_2H_6) and ethanol (C_2H_6O)

Red areas denote strongly negative parts of the molecular surface and hence strongly positive values of σ. Deep blue marks denote strongly positive surface regions (strongly negative σ) and green denotes non-polar surface.

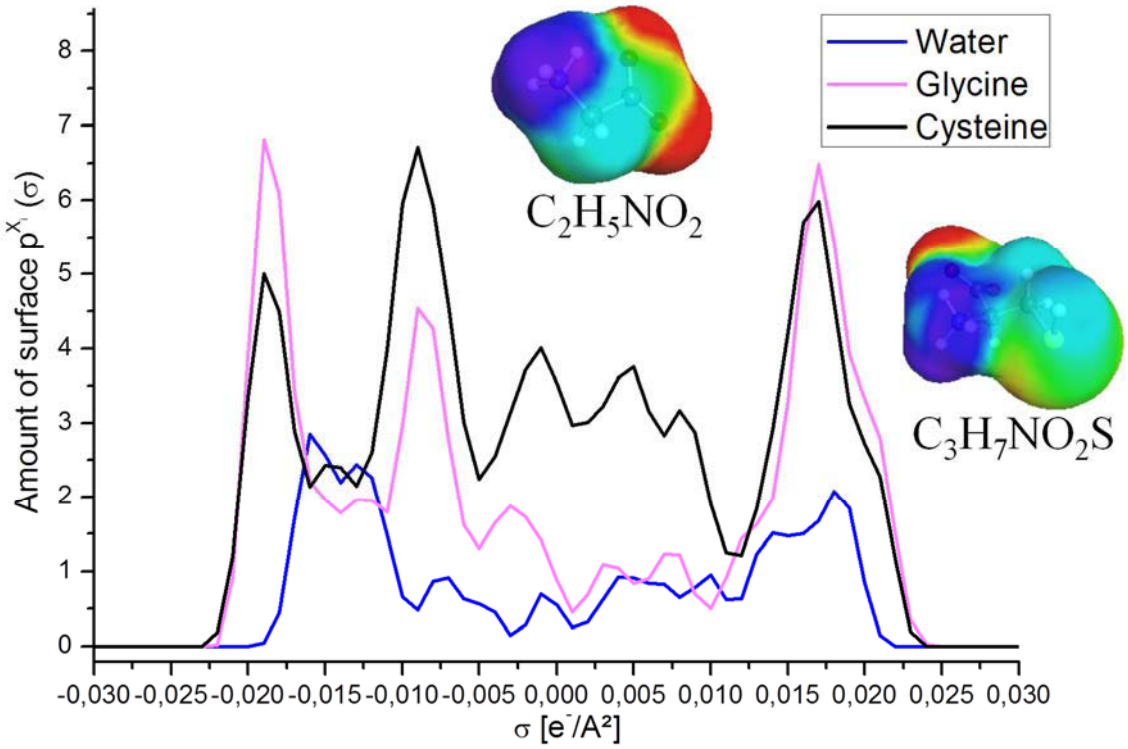


Figure II-10 : σ-profiles of water (H_2O), glycine ($C_2H_5NO_2$) and cysteine ($C_2H_7NO_2S$)

Red areas denote strongly negative parts of the molecular surface and hence strongly positive values of σ. Deep blue marks denote strongly positive surface regions (strongly negative σ) and green denotes non-polar surface.

The 3D polarization density distribution on the surface of each molecule X_i, is converted into a distribution-function, the so called σ-profile $p^{X_i}(\sigma)$, which gives the relative amount of surface with polarity σ on the surface of the molecule (Klamt 2005). The σ-profile is an additive molecular property, thus, the σ-profile of the entire solvent of interest S, which might be a mixture of several compounds, $p^S(\sigma)$ can be built by adding the $p^{X_i}(\sigma)$ of the components weighted by their mole fraction x_i in the mixture. To simplify notations the value of the σ-profile $p^{X_i}(\sigma)$ of a given molecule X_i will be denoted as $p_i(\sigma)$ in the upcoming discussions. Thus, the σ-profile of the entire liquid system or solvent of interest S is

$$p^S(\sigma) = \sum_{i \in S} x_i p^i(\sigma) \tag{II-31}$$

where x_i denotes the molar fraction of the compound X_i.

In order to perform thermodynamics calculation, equation II-31 must be used in a normalized form given as:

$$\overline{p^S(\sigma)} = \sum_{i \in S} \frac{x_i p^i(\sigma)}{x_i A_i} \tag{II-32}$$

Thus, from simple statistical considerations, all values of the probability of finding a surface of density σ in the mixture are expected to be less or equal to unity. Then, the "normalized" σ-moments of such distribution can be easily calculated using the following equations:

$$\begin{cases} \overline{M_k(X_i)} = \sum_{j \in X_i} \overline{p^j(\sigma)} f_k(\sigma) \\ f_k(\sigma) = \sigma^k \qquad \text{for } k=0,1,2,3,... \end{cases} \tag{II-33}$$

The zero-order normalized σ-moment $\overline{M_0(X_i)}$ is the sum of the probability of finding all segments of the molecule X_i and is equal to 1 by definition due to the normalization. Without normalizing, this moment would be the total surface area of the molecule X_i (Klamt 2005).

Without normalizing, the first-order σ-moment $\overline{M_1(X_i)}$ is the total COSMO polarization charge on the surface of the molecule X_i (Klamt 2005). Thus, $\overline{M_1(X_i)} = 0$ for neutral compounds like H_2O, 1 for monovalent cations like H_3O^+ and -1 for monovalent anions like OH^-. These findings are observed by analyzing Figure II-8 where the σ-profile of water is highly symmetric while those of H_3O^+ and OH^- are asymmetric with a large part of their respective peaks being in the corresponding polarity zones of the charge of the species. Similarly the σ-profiles of the other neutral molecules shown in Figure II-9 and Figure II-10 are all highly symmetric.

Without normalizing, the second-order σ-moment $\overline{M_2(X_i)}$ is highly correlated with the total COSMO polarization energy, and hence it is a measure of the overall ability of the solute molecule X_i to interact electrostatically with a polarisable continuum (Klamt 2005).

Without normalizing, the third-order σ-moment $\overline{M_3(X_i)}$ lacks a simple physical analogy and represents a kind of "asymmetry" of the σ-profile of the solute (Klamt 2005).

The σ-moments can be used in a multivariate linear regression analysis and even more advanced statistical and mathematical techniques (like artificial neural networks) to find quantitative structure-property relationships (QSAR) that are regularly used in computational chemistry (Klamt 2005).

The chemical potential of a unit piece of surface of kind σ, *i.e.* the so-called "σ-potential", in a given liquid S (pure compound or mixture) can be represented by a Taylor-series with respect to σ (Klamt 1995, 2005):

$$\mu^S(\sigma) = \sum_k^m c_k^S \overline{M_k(X_i)} \qquad \text{(II-34)}$$

where c_k^S are the σ-moment coefficients (SMC's) describing the liquid system S.

However, this kind of empirical correlation for estimating the chemical potential of a given piece of segment would introduce the same problems (like experimental data availability, etc.) as those appearing in group-contribution methods like UNIFAC. Thus, for developing the COSMO-RS theory, it was chosen to develop a more physical way (based on thermo-statistical calculations) for calculating the interaction energies between a given pairwise of interacting segments σ and σ' in order to deduce the corresponding value of the σ-potential. The details of this development are given in the "COSMOSPACE" paper (Klamt et al. 2002). Below, the main concepts beyond this thermodynamic approach are introduced.

II.1.2.3. Interactions energies between two surface segments

Within the COSMO-RS approach of segments interaction, there are three types of interactions to take into account, namely the van der Waals, electrostatic misfit and hydrogen bonding interactions (Klamt 2005).

Before discussing how these interactions are taken into account, it is important to notice the fundamental difference between the choices of the reference state. The COSMO-reference state is a hypothetical or virtual reference state where the molecule is assumed to be screened in a perfect conductor environment as though there is no conductor in the real case. Thus this virtual conductor environment has to be removed (this is done by introducing the concept of electrostatic misfit

interactions). Then the COSMO reference state would be comparable to the "ideal liquid" solution. In order to give a more realistic treatment of the interactions between two liquids, it is important to treat the hydrogen bond interactions as well as the dispersive interactions. The hydrogen bond interactions are treated using a kind of quantum-chemical based treatment which can be considered as an approximate but useful treatment. To preserve the coherence with the original reasoning of Klamt 2005, these interactions are discussed in the following order: vdW energies, misfit energies, and HB energies.

II.1.2.3.1. van der Waals (vdW) interactions

The dispersive or van der Waals (vdW) interactions are composed of an attractive force that arises from a sophisticated quantum-chemical short-range interaction of electrons, and an even shorter range, *i.e.*, almost hard sphere repulsion of the cores of the atomic electron distributions, the so-called Pauli repulsion. The first term is due to the correlation in the motion of electrons that leads to a lowering of the energy of electrons. It decreases with r^{-6} with an increase in the distance, r, of two electrons. Electrons in the outer region of the electron cloud of a molecule in vacuum lack correlation partners in half of the space directions. The resulting strong attraction is balanced by the extremely strong Pauli repulsion, which avoids a strong interpenetration of the electron clouds of non-bonded molecules. As a consequence, the molecules experience a very steep energy valley if they are about at vdW-distance from each other. Only 1 Å further out, the vdW energy gain is almost lost. Therefore, molecular surfaces experience a kind of binary decision: "Either get to intimately close contact with the surface of another molecule, or stay away and enjoy its freedom" (Klamt 2005). This is the reason why such a clear first-order phase transition is usually observed between gas phase and condensed phases, and not just a continuous increase of the average distance of the molecules with increasing temperature. Therefore, it may be assumed that in liquid system almost each piece of molecular surface has a vdW interaction partner in about the vdW-distance. The strength of the vdW interaction can be expected to be roughly proportional to the electronic polarizability of the partners. Since the electronic polarizability in a first-order approximation can be considered as a constant in organic molecules, corresponding to an optical dielectric constant of $\varepsilon_\infty = n^2 \cong 2$, the amount of vdWs interaction per contact area is about -40 kJ/mol/nm², or about -20 kJ/mol/nm² per molecule surface area, if we take into account the fact that in each contact there are two molecular surfaces of the same size. To a second order, there will be variations, and these are known to depend mainly on the elements involved in a contact (Klamt 2005). Therefore, to the second order we can express the vdWs surface energy between two elements, e and e', in the form

$$e_{vdW}\left(e,e'\right) \cong 2e_{vdW}^{0} + \delta e_{vdW}\left(e\right) + \delta e_{vdW}\left(e'\right) \equiv \tau_{vdW}\left(e\right) + \tau_{vdW}\left(e'\right) \quad \text{(II-35)}$$

Unfortunately, the exact values of the element-specific vdW surface energies $\tau_{vdW}(e)$ are not simple to calculate. Therefore, it is necessary to assume that these are known from some fit to experimental data. The vdW energy is - to the second order of approximation- just the sum of two independent contributions of the surfaces involved in the interaction. Thus it can be assumed that this vdWs energy is gained individually by each molecule when it is brought from vacuum to the virtual continuum, which is chosen (within the COSMO-RS approach) as reference state for molecule in the liquid state (Klamt 2005).

Hence, it is assumed that this continuum behaves like a conductor with respect to electrostatics, and like an average organic vdW partner with respect to dispersion. In this way, it is not necessary to consider vdWs interactions explicitly beyond the continuum representation, although they are qualitatively and quantitatively a very important part of the interactions in the liquid phase. Then, the vdW energy gain of a molecule X_i during the transfer from the gas phase to any solvent is given by the integral of $\tau_{vdW}(e)$ over the molecular surface. Corrections beyond this picture are integrated in the COSMOtherm software, in the form

$$E_{vdW}^{X_i} = \sum_{k \in X_i} a_k^{X_i} \tau_{vdW}\left(e(k)\right) \qquad \text{(II-36)}$$

where $a_k^{X_i}$ denotes the molecular surface on atom *k*, and $\tau_{vdW}\left(e(k)\right)$ indicates the element of atom *k*. The vdW energy $E_{vdW}^{X_i}$ is dependent only on the element type of the atoms that are involved in surface contact. It is spatially nonspecific. $E_{vdW}^{X_i}$ is an additional term to the energy of the reference state in solution. Currently nine of the vdW parameters $\tau_{vdW}(e)$ for elements H, C, N, O, F, S, Cl, Br, and I have been optimized. For the majority of the remaining elements reasonable guesses are available (Klamt et al. 1998).

Because this vdW contribution is independent of any neighbourhood relations, it is not really an interaction energy, but it may be considered as an energy contribution to the reference state in solution. Thus, the state of a molecule embedded in a vdW interacting conductor is considered as reference state. Within this approximation, vdW interactions do not contribute to the free energies of transfer between two different liquid states, but only to the liquid-gas transfer, *i.e.* to vaporisation (Klamt 2005).

With no further approximations an ensemble of molecules representing the components of the mixture is considered as swimming around individually in the infinitely extended virtual conductor. Because the electric field of the molecules is perfectly screened off by the conductor, there is no interaction of the molecules. The total energy of this ensemble is known by just

summing the total energies of the individual molecules in this virtual, but distinguished state, which is also called the COSMO reference state illustrated in Figure II-11.

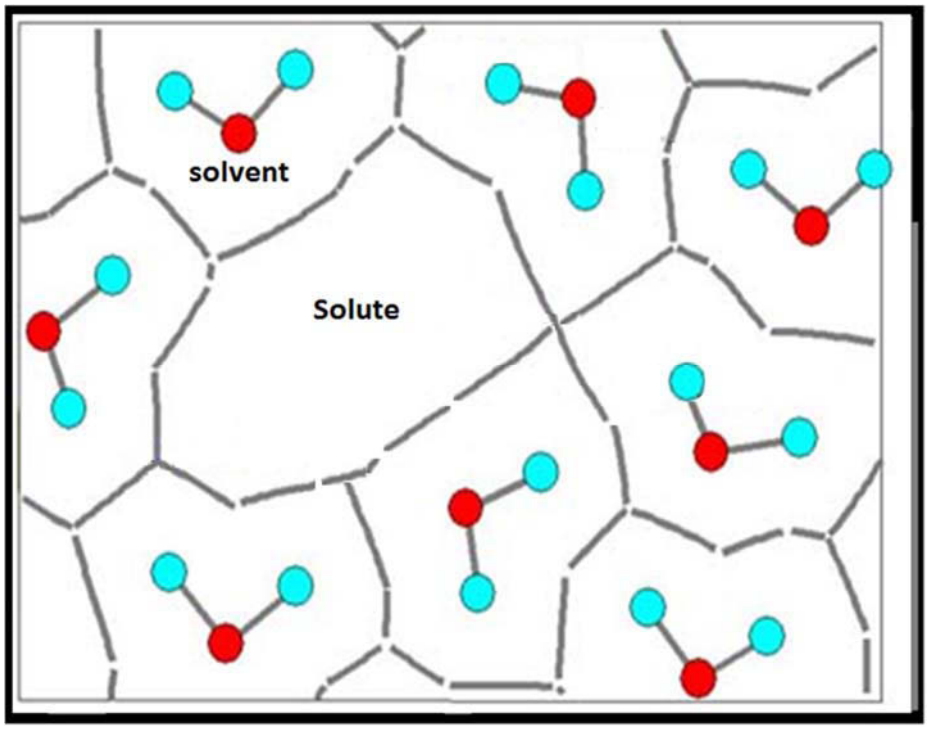


Figure II-11 : Illustration of the packing of liquid molecules in the COSMO reference state. *The liquid molecules are schematized for this purpose as contacting molecular cavities. (Adapted from Klamt 2005 and Klamt and Eckert 2000)*

II.1.2.3.2. Electrostatic-misfit interactions

Obviously, the interactions in a real fluid are much complicated than in the ideal condition of the COSMO state. In order to come closer to a realistic picture, we now start to squeeze out the conductor from between our molecules, until we finally end with a system having about the right liquid density. Let us assume that during this process there is always at least an infinitesimal thin film of conductor left between two molecules. Such a thin film is sufficient to ensure perfect screening of the electric fields of the molecules. Since there are no interactions between the molecules in the conductor, the squeezing of the conductor does not result in any change of energy as long as the molecules can keep their original shape and cavity. Finally, we need to slightly deform the individual molecules by pressing them together in order to get a close packing with about the right density. Let us assume that the cavities are slightly deformable if the entire volume is conserved under the deformation. Such deformation will cause the cavity to get a little closer to the atom centres at some part of the cavity, but a little further away in other parts, in order to conserve the volume. As a result, the dielectric interaction energy will increase slightly in some parts, but decrease in neighbouring parts. In summary, we may assume that this kind of deformation costs only very little energy, and that on average the energies and polarization charges of the molecules stay the same as in the conductor. Because the volumes of the original COSMO cavities

correspond with the molar volumes in the liquid phase, there is almost no conductor left in the system when we finally reach the right density. The molecular cavities are closely nestled against each other, with a thin film of conductor separating them (Klamt 2005).

This situation is shown schematically in Figure II-12, where the grey lines may represent the thin film of conductor.

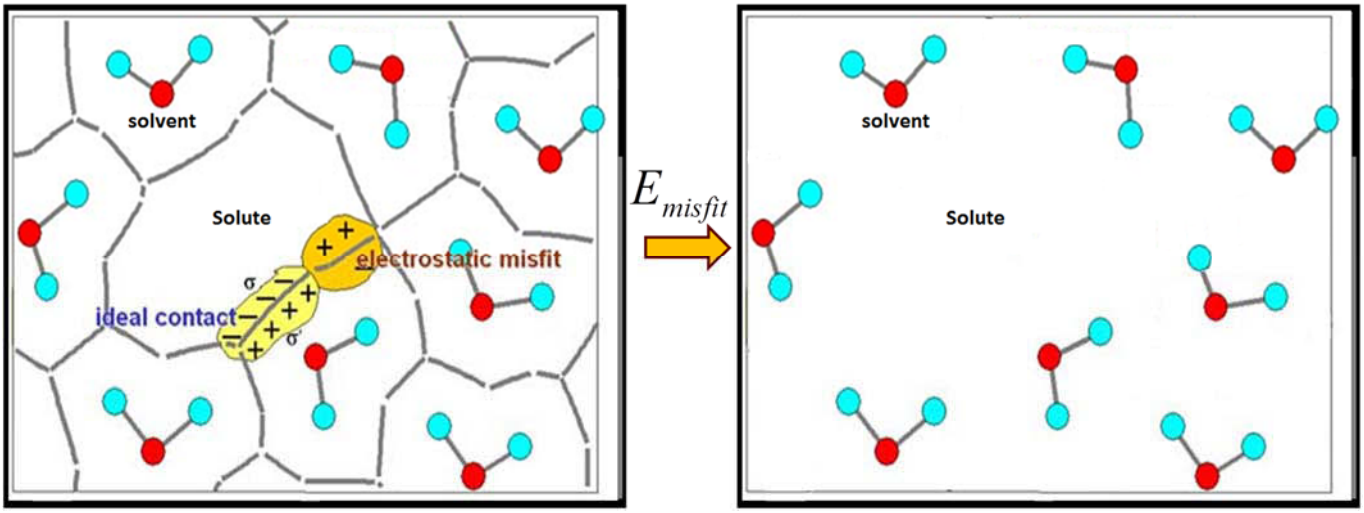


Figure II-12 : Illustration of the "misfit interactions" concept in the COSMO-RS method.

With the approximations mentioned above, this system is assumed to still have the same total energy as the systems of molecules swimming in the bulk conductor (Klamt 2005).

Although the system now looks much more like a liquid system, the thin film of conductor separating the molecules is an artefact. In nature there is no conductor between molecules, and hence we have to get rid of the conductor to get closer to the real situation. For doing this, it is helpful to assume that all polarization charges are now frozen, which does not make a real difference but may help our imagination during the following steps. Let consider what happens if we take away the conductor on a small piece of molecular contact of size a_{eff}. Let the average polarization charge densities arising from the two neighbouring molecules on this piece of contact surface be σ and σ'. In this situation, the net charge density on the contact under consideration is just σ + σ'. If σ + σ' is zero, as shown for one segment in Figure II-12, we have an electrostatically ideal contact. The conductor does not need to contribute to the electrostatic screening on this piece of contact, because the two molecules would already screen each other in the same way. Therefore, the conductor is taken away in this situation without any electrostatic energy difference. Although such pairing of opposite σ-values is energetically optimal, the normal situation will obviously not involve such ideal pairing of exactly opposite polarities because, in reality, thermal fluctuations, steric hindrance, or even the non-availability of the right partner, will cause some electrostatic misfit, as shown for another contact in Figure II-12. Since the electrostatic potential on each piece of molecular surface is zero, as long as it has the conductor present, the removal of the conductor on

a piece of surface corresponds to bringing the countercharge density of σ + σ' from infinity to this surface area. In a non-polarisable environment this would cost an electrostatic energy:

$$E^{0}_{misfit}\left(\sigma,\sigma'\right)=a_{eff}\,\frac{\alpha}{2}\left(\sigma+\sigma'\right)^{2} \tag{II-37}$$

The self-energy coefficient can be calculated simply from basic electrostatics. It depends on the size of the contact, and slightly on the shape, but assuming nearly circular contacts should be a good approximation. Since the electrostatic misfit energy is better approximated if we take this into account by reducing the energy by a portion $f_{\text{dielectric}}\left(\varepsilon_{\infty}\cong 2\right)\cong 0.4$. Thus, we obtain the expression:

$$e_{misfit}\left(\sigma,\sigma'\right)=\left(1-f\left(\varepsilon_{\infty}\right)\right)\frac{\alpha}{2}\left(\sigma+\sigma'\right)^{2}\equiv\frac{\alpha'}{2}\left(\sigma+\sigma'\right)^{2} \tag{II-38}$$

for the misfit energy density per unit of contact area. The functional dependence on the polarization charge densities is beyond any question and for the coefficient α' we have a good physical estimate as soon as we know the size of the contact area, a_{eff}. Thus, we may assume that the electrostatic misfit surface energy density is known and calculated as:

$$E_{misfit}\left(\sigma,\sigma'\right)=a_{eff}\,\frac{\alpha'}{2}\left(\sigma+\sigma'\right)^{2} \tag{II-39}$$

Because we have already included the vdW energies in the continuum energy up to second order, only higher-order contributions would arise from the replacement of the continuum by a real molecular partner. Thus, we can neglect vdW interactions in this step of local conductor replacement (Klamt 2005).

II.1.2.3.3. Hydrogen bonding (HB) interactions

There is a third kind of important interaction, which needs to be taken into account, *i.e.*, the hydrogen-bond (HB) interactions. We shall consider a contact of a very polar hydrogen atom, an ***HB-donor***, with a strongly charged electron lone-pair, an ***HB-acceptor***, as shown in the upper right of Figure II-13.

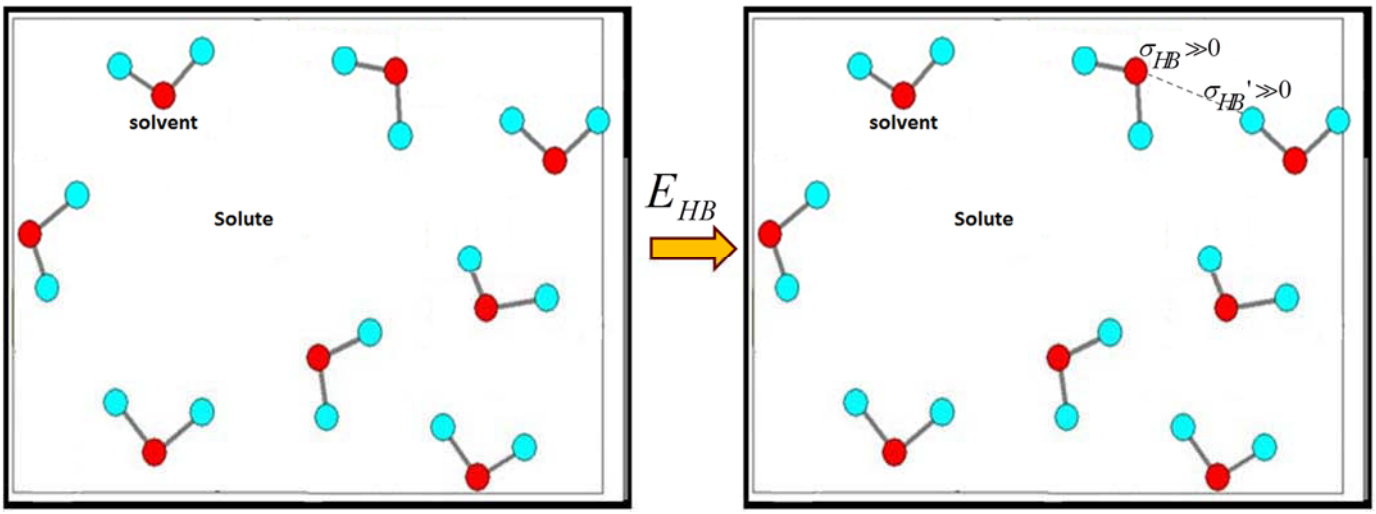


Figure II-13 : Illustration of the "hydrogen bond" concept in the COSMO-RS method.

After removing the conductor, the HB-donor will start to penetrate in the HB-acceptor density. This goes along with a large energy gain and a specific re-orientation of the donor relative to the acceptor. Since the latter causes a considerable loss of entropy compared with less specific normal interactions we have to express the HB-interaction as a free energy gain. Any attempt to quantify the HB-interaction energy by quantum-chemical means would require tremendous effort, because a very high methodological level is needed to achieve a good description of hydrogen bonding, even in vacuum. Therefore, Klamt and co-workers have tried to parameterize it empirically with the information already available in the COSMO-RS model. Hydrogen bonding only appears if two surfaces segments with strong and opposite polarity get in contact, and the HB energy should be stronger as both partners become more polar (Klamt 2005). Such behaviour can be described reasonably by the formula:

$$e_{HB}(\sigma,\sigma') = c_{HB}(T)\min\left\{0, \min\left(0, \sigma_{don} + \sigma_{HB}\right)\max\left(0, \sigma_{acc} - \sigma_{HB}\right)\right\} \quad \text{(II-40)}$$

This expresses the HB free energy per unit surface area gained in a contact of two molecular surfaces with polarization-charge densities σ and σ'. Other, simpler or more complicated formulae can be used as well, with only small differences in the results. The form given in equation II-41 is the one presently used in the current COSMO-RS parameterization, where

$$\begin{aligned} E_{HB}(\sigma,\sigma') &= a_{eff} e_{HB}(\sigma,\sigma') \\ &= a_{eff} c_{HB}(T)\min\left\{0, \min\left(0, \sigma_{don} + \sigma_{HB}\right)\max\left(0, \sigma_{acc} - \sigma_{HB}\right)\right\} \end{aligned} \quad \text{(II-41)}$$

In this expression, σ_{don}= min *(σ, σ')* denotes the more negative σ, *i.e.*, the more positive polarity of the two pieces of surfaces. This surface would obviously take the part of the HB-donor. Only if this donor is more polar than a certain threshold, *i.e.*, if its σ is more negative than σ_{HB}, can it act as a donor in a hydrogen bond. Conversely, its partner must overcome a certain threshold polarity to act

as an acceptor, *i.e.*, σ_{acc} = max $(\sigma, \sigma') > \sigma_{HB}$ is required. The parameters c_{HB} and σ_{HB} and the detailed temperature dependence cannot be derived from theoretical arguments. They have been determined by fitting to experimental data (Klamt 2005).

II.1.2.3.4. Concluding remarks about interactions between two surface-segments

We have thus now expressed the energy difference resulting from the local removal of the conductor between two molecules as a function of the two surface polarities σ and σ', of the interacting molecular surfaces. All relevant kinds of intermolecular interactions, *i.e.*, vdW interactions, electrostatic interactions, and hydrogen bonding are taken into account. Summing the local contributions up over all contact surfaces should finally lead us to the total energy difference of the real system without a conductor to the ensemble of perfectly screened molecules, which was chosen as the energetic reference point of the COSMO-RS model. Since we know the energy of the latter from the initial quantum COSMO calculations, we then have the total energy of a realistic ensemble representing our liquid system (Klamt 2005).

$$e_{\text{int}}(\sigma,\sigma') = e_{misfit}(\sigma,\sigma') + e_{HB}(\sigma,\sigma') \qquad \text{(II-42)}$$

The total interaction energy of a liquid-like ensemble of molecules is calculated starting from COSMO calculations and taking into account deviations from the simple conductor-like electrostatic interactions as well as hydrogen bonding as local pair-wise interactions of molecular surfaces. As discussed above, vdW interactions do not contribute to the free energies of transfer between two different liquid states, but only to the liquid-gas transfer, *i.e.* to vaporisation data.

This is a very different way of quantifying the total energy than the ways usually used in all kinds of molecular mechanics (MM) or quantum mechanics (QM) calculations, where the total energy is evaluated by demanding summations of non-local pair-wise interactions of atoms, or of nuclei and electrons. Although chronologically incorrect, since Klamt argued that he did not realize these relations at the time of development of the COSMO-RS method, it is worth mentioning the similarity between the COSMO-RS view of liquid-phase interactions to that of the widely used empirical liquid-phase thermodynamics models, like the Flory-Huggins theory, Guggenheim's quasi-chemical theory, UNIQUAC, UNIFAC and many others. All these theories work with a kind of local group interaction parameters, or more precisely local group surface interaction parameters, but all of them treat these parameters as empirical fit parameters to be adjusted to experiment. In some way, the COSMO-RS picture of first bringing the molecular surfaces close to each other within a conductor, and finally removing the conductor for the first time, provides the chance of connecting the models for surface interactions with real molecular interactions, and of calculating the surface interaction parameters from molecular structures. But the COSMO-RS picture allows for a much more detailed picture of the local surface interactions, since it does not require an

averaging of the interactions over entire functional groups that is required in group contribution methods for a limitation of the degrees of freedom of the fit (Klamt 2005).

Nevertheless, we are left with an important problem at this point: the energy of a single configuration of an ensemble is almost worthless for the evaluation of fluid-phase thermodynamics, because the total free energy must be evaluated as a statistical average of a large number of different configurations in order to take into account all the different ways in which molecules can contact and interact with each other. MD or MC techniques in the context of QM or MM calculations usually solve this question. Unfortunately, the COSMO-RS concept only provides the concept of how to evaluate the energy of a given liquid-like configuration, but not of generating such a configuration because, owing to the missing distance dependencies, there is no concept of the forces in this model. Having no forces, it is impossible to apply molecular dynamics, and even finding reasonable configurations required for Monte Carlo statistics is not feasible. Hence, if the aim were to follow the classical pathways of statistical molecular mechanics, it would be necessary to combine COSMO-RS with a kind of force-field MC approach to generate configurations, which would then be energetically quantified by COSMO-RS. In this way there is danger of losing most of the benefits achieved previously with the novel way of description of interactions in the liquid phase. Therefore, Klamt had decided to try another way out, which finally proved to be very successful as described below.

II.1.2.4. From surface contact interactions to fluid phase thermodynamics

Let us assume that the energy of a liquid ensemble S of molecules compared to its conductor reference state, *i.e.*, compared to state in which all molecules X_i of the ensemble are individually embedded in a perfect conductor, is expressed as the sum over all surface contacts k of the molecules of our ensemble (Klamt 2005):

$$E_{tot}^{S} \cong \sum_{k \in S} a_k e_{\text{int}}\left(\sigma_k, \sigma_k^{'}\right) \cong \sum_{k \in S} a_{eff} e_{\text{int}}\left(\sigma_k, \sigma_k^{'}\right) \qquad \text{(II-43)}$$

By the second part of Eq. (II-43) we make the simplifying assumption that in average the contact areas of the pair contacts k can be well described by an effective contact area a_{eff}. This equation looks simple, but its evaluation requires that we know all contacts k, *i.e.*, all the pairs of polarization charge densities σ_k and σ'_k, which are in close molecular contact with each other. Obviously, in addition to the problem that these contacts in a liquid system are subject to permanent fluctuations, we do not even know them at a single time. In this way, the problem of the calculation of solvation energies of molecules has been converted into the problem of calculating the correct

average of the total energy of a liquid, *i.e.*, permanently fluctuating ensemble *S*. Thus we end up with statistical thermodynamics (Klamt 2005).

Given all the geometric constraints implied by the neighbourhood of segments on molecular surfaces, *i.e.*, the fact that the existence of a certain contact (σ, σ') has strong implications on the possibility or probability of contacts of their neighbour segments, the evaluation of Eq. (II-43) would require an exhaustive or at least a sufficient sampling of the possible mutual arrangements of the molecules *X* of our ensemble *S*. Fortunately, empirically it has turned that for almost all liquid systems, the explicit consideration of these geometric constraints is not required for a good evaluation of the thermodynamic averages. This independently was found during the development of COSMO-RS as well as much earlier within the simpler and more empirical thermodynamic models widely used in chemical engineering (Lei et al 2008 and references therein). With neglect of geometrical constraints, the statistical thermodynamics of the liquid ensemble thus reduces to the much simpler evaluation of the statistical thermodynamics of the corresponding ensemble of pairwise interacting surface segments (Klamt 2005). This problem is equivalent to finding the σ-potential $\mu^S(\sigma)$, *i.e.*, the chemical potential of an effective surface segment of area a_{eff} and polarity σ in the ensemble *S*, from the equation:

$$\mu^S(\sigma) = -\frac{RT}{a_{eff}} \ln\left(\int \overline{p^S(\sigma')} \exp\left(\frac{a_{eff}}{RT}\left(\mu^S(\sigma') - E_{misfit}(\sigma,\sigma') - E_{HB}(\sigma,\sigma') \right) \right) d\sigma' \right) \quad \text{(II-44)}$$

with

$$\overline{p^S(\sigma)} = \sum_{i \in S} \frac{x_i p_i(\sigma)}{x_i A_i} \quad \text{(II-45)}$$

being the composition histogram of the ensemble *S* with respect to the surface polarization charge density σ. This again is based on the mole fractions x_i of the individual molecular components X_i of the ensemble, as well as on the individual molecular surface areas A_i and molecular σ-histograms $p_i(\sigma)$ of the components. The latter are usually named as σ-profiles within COSMO-RS. Equation (II-44) has to be solved iteratively, usually starting with $\mu^S(\sigma') = 0$ on the right hand side. After convergence, the σ-potential $\mu^S(\sigma)$ provides a solvent specific function that describes the affinity of solvent *S* for molecular surface of polarity σ. The final step of COSMO-RS consists in the evaluation of the chemical potential of any solute molecule X_i in solvent *S* by integrating the solvent σ-potential $\mu^S(\sigma)$ over the surface of solute X_i as

$$\mu_i^S = \int p_i(\sigma)\mu^S(\sigma)d\sigma + RT \ln\left(x_i \gamma_i^{combi,S} \right) \quad \text{(II-46)}$$

Please note, that in the expression of the chemical potential of Eq. II-46, since the term appearing in the integral is a ''pseudo-chemical potential'', the $RT \ln x_i$ term has to be added in order to get the

value of the standard chemical potential (Ben-Naim 1987). Thus, the logarithmic term takes into account the trivial dependence of the chemical potential of X_i on its mole fraction x_i, and a correction for the relative sizes of solutes and solvents known as 'combinatorial term'. Approximate expressions based on the surface areas and volumes of solutes and solvents are generally used. Since molecular surfaces and volumes are readily available within COSMO-RS from the COSMO cavities, a similar expression for $\gamma_i^{combi,S}$ is adopted. As a result, all information required for the evaluation of the chemical potential of a given compound X_i, *i.e.*, σ-profiles and σ-potentials, surface areas and volumes, are derived from the COSMO calculations of the involved components at their reference state, *i.e.* in the conductor (Klamt 2005).

By Eq. (II-46), COSMO-RS provides the chemical potential of any solute X in almost any pure or mixed solvent S as function of temperature and concentration, and by that it gives access to almost the entire fluid phase equilibrium thermodynamics, *i.e.*, to almost all kinds of activity coefficients, partition coefficients, solubilities, free energies, heats and entropies of solvation, liquid-liquid or vapour-liquid equilibria, and many more. Hence, the COSMOtherm program and its graphical interface COSMOthermX, both based on the COSMO-RS theory of interacting molecular surface charges, allow the treatment of mixture thermodynamics at variable temperatures and thus, open the door for many application areas that were inaccessible for the continuum models before, especially in chemical engineering (Klamt 2005).

II.1.3. Conformations/ isomers

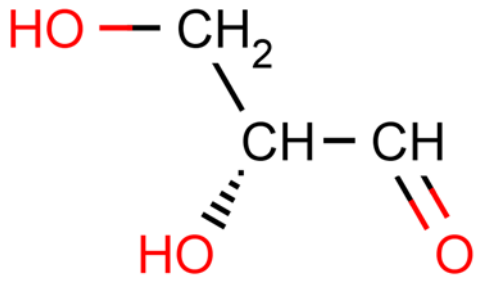
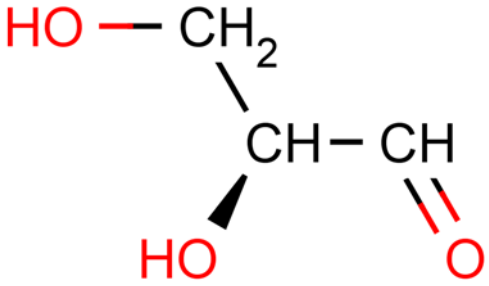
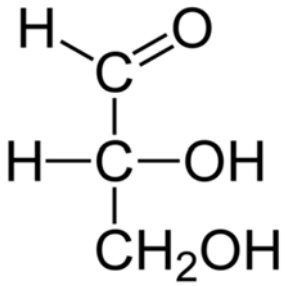
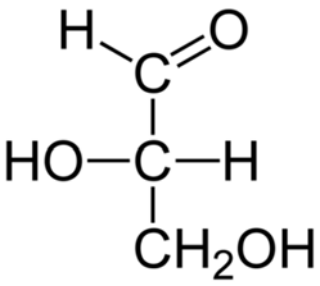
In chemistry, isomers (from Greek *isomers, isos* = "equal", *méros* = "part") are molecules with the same molecular formula but different chemical structures. By definition, the isomers that possess identical constitution, but which differ in the arrangement of their atoms in space are called ***stereoisomers***. Two ***enantiomers*** are defined as one of a pair of molecular entities, which are mirror images of each other and non-superposable. The ***conformer*** is one of a set of stereoisomers, each of which is characterized by a ***conformation*** (*i.e.* a spatial arrangement of the atoms) corresponding to a distinct potential energy minimum (Moss 1996).

For example, glyceraldehyde is the simplest carbohydrate and a possible prebiotic precursor of more complex sugars. It is a fundamental molecule in molecular chirality chemistry; the *D* or *L* configuration of sugars is determined by referring to the configuration of glyceraldehyde (Belitz et al. 2009). Indeed, glyceraldehyde is a triose monosaccharide with chemical formula $C_3H_6O_3$. It is the simplest of all common aldoses. Its name comes from combining glycerine and aldehyde, as glyceraldehyde is merely glycerine with one hydroxymethylene group changed to an aldehyde. Glyceraldehyde has a chiral centre and therefore exists as two different enantiomers with opposite optical rotation: *R* from Latin *rectus* meaning "right", or *S* from Latin *sinister* meaning "left". In the

D/L system, glyceraldehyde is used as the configurational standard for carbohydrates. Monosaccharides with a conformation identical to (*R*)-glyceraldehyde at the last stereo-centre, for example C5 in glucose, are assigned the stereo-descriptor *D*-. Those similar to (*S*)-glyceraldehyde are assigned an *L*- (Belitz et al. 2009).

The Fisher projection and the skeletal formula of the two *D/L* conformers of glyceraldehydes are given in Table II-1.

Table II-1 : 2D-structures of the D/L conformers of glyceraldehyde

	D-glyceraldehyde *R*-glyceraldehyde (+)-glyceraldehyde	*L*-glyceraldehyde *S*-glyceraldehyde (-)-glyceraldehyde
Skeletta formula		
Fisher projection		

With the exception of glycine, where *R* = H, all of the naturally occurring amino acids are optically active. Using the same *D/L* convention that Fischer developed for sugars, all of the chiral amino acids are classified as *L*-amino acids. Figure II-14 compares the Fischer projection of *D*-glyceraldehyde to that of a chiral amino acid.

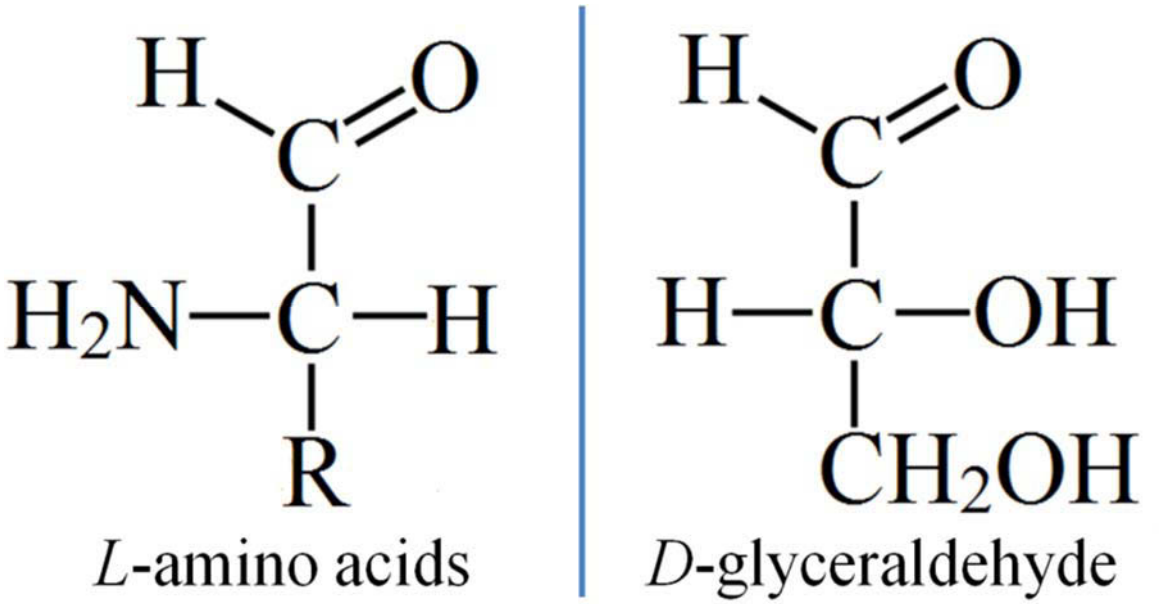


Figure II-14 : Stereo-chemical designation of amino acids, where R represents any of the more than 20 substituents that are found in naturally occurring amino acids.
The (D) form of glyceraldehyde has been used to illustrate more easily the stereo-isomerism with the amino-acids which are generally of the (L) form.

Since molecules often can adopt more than one conformation, it is important to be able to handle the conformations equilibria in mixture. The COSMO-RS method is able to treat such equilibria, as discussed below.

Until now, in the COSMO-RS theory, we have treated molecules as rather stiff entities having a well-defined total energy, COSMO cavity and COSMO polarization charges. While this is a good approximation for many simple chemical compounds, *e.g.* methane, ethane, propane, benzene, toluene, methanol, etc., most of the more complex molecules have more than one relevant conformation, *i.e.*, they have relevant meta-stable energy minima in addition to the total energy minimum. Thus, the practical assumption that a single energetic minimum can be used as representative for all possible geometries is a crude approximation, especially in liquid phase where the lowest free energy structure might depend on the solvent composition. Indeed, conformers have different electrostatics (*e.g.* dipoles), different hydrogen bonding capacity and different van der Waals properties (Klamt 2005).

In the COSMO-RS theory, this translates to different σ -profiles (as illustrated in Figure II-15), different σ-potentials and changing relevance in different solvents.

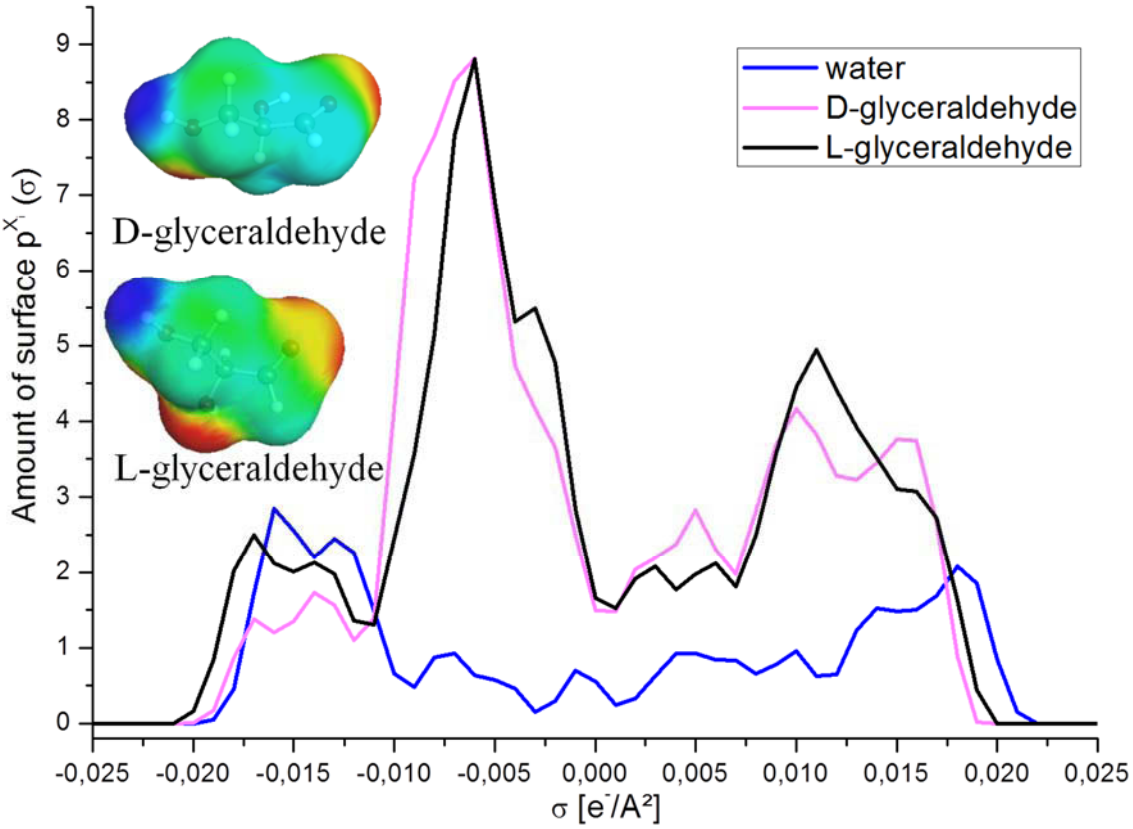


Figure II-15 : σ-profiles of the two (D and L) conformers of glyceraldehyde ($C_3H_6O_3$)

These differences in σ-profiles can induce considerable solvent-dependent shifts demonstrating that for molecules with such pronounced conformational differences, the equilibration of conformations has to be taken into account in the calculation of thermodynamic phase-equilibrium data. In this context, a compound X is represented by a set of COSMO-files for the conformers X_i. At the thermodynamic equilibrium the chemical potentials of all conformers are identical, at a given temperature T. If it is assumed that the mixture is an ideal solution of the conformers, then:

$$\mu_i^S = \mu_i^0 + RT \ln\left(x_i\right) \qquad \text{(II-47)}$$

where x_i denotes the molar fraction of a conformer i in the mixture (solvent S) (Klamt 2005).

In order to enable a consistent treatment of conformational equilibria, an automated conformation equilibration scheme is implemented in the COSMO-RS algorithm. This scheme is based on a physical statistics approach using the Maxwell Boltzmann statistics that describes the average distribution of non-interacting material particles over various energy states in thermal equilibrium (*i.e.* the temperature is kept constant). If a multiplicity ω_i is assigned to each conformer X_i based on geometrical degeneration aspects, the population $\pi^S(X_i)$ (in percentage) of a conformer i, in a solvent S can also be calculated as

$$\pi^S(X_i) = \frac{\omega_i \exp\left\{-\frac{E_{COSMO}(i) + \mu_i^S}{RT}\right\}}{\sum_j \omega_j \exp\left\{-\frac{E_{COSMO}(j) + \mu_j^S}{RT}\right\}} \quad \text{(II-48)}$$

according to the Boltzmann distribution between states of different free energies at the given temperature *T*. If the compound X_i is itself a relevant part of the solvent *S*, the chemical potentials μ_i^S, themselves depend on the conformation population. Therefore, in general, the latter equation has to be iterated to self-consistency, starting from an initial population guess based on μ_i^S =0. The calculated chemical potential of each conformer is used to take into account the equilibrium between the solvent *S* (*i.e.* all the conformers X_i weighted by their respective population $\pi^S(X_i)$ in a thermodynamical sense) and any other solvent *S'* by using classical concepts of the solution theory (Klamt 2005).

II.2. Computational details of the COSMO-RS calculations

COSMOtherm is a command line/file driven program, which can be run directly from a UNIX or DOS shell. It allows for the calculation of any solvent or solvent mixture and solute or solute system at variable temperature and pressure. COSMOtherm uses the chemical potentials derived from the COSMO-RS theory to compute all kinds of equilibrium thermodynamic properties or derived quantities (*e.g.* vapour pressure, free energy of solvation, activity coefficients, partition coefficients, solubility and solid-liquid equilibria (SLE), liquid-liquid equilibrium (LLE) and vapour-liquid equilibrium (VLE), phase diagrams, azeotropes, miscibility gaps, excess enthalpies and excess free energies, pK_a of acids and bases, various QSPR models, reaction constants, liquid extraction equilibria) (COSMOthermX 2011).

COSMOthermX is a Graphical User Interface to the COSMOtherm command line program. It allows for the interactive use of the COSMOtherm program, i.e. selection of compounds, preparation of property input, program runs and display of calculation results.

The input for the compounds is read from the COSMO files, identified by the extensions ".cosmo" or ".ccf", which are result files from quantum chemical COSMO calculations. COSMOtherm extracts the relevant information directly from the COSMO files. The compressed COSMO files (.ccf) use significantly less disk space than conventional COSMO files. At least one COSMO file or compressed COSMO file has to be selected as compound input (COSMOthermX 2011).

Because the quality, accuracy, and systematic errors of the electrostatics resulting from the underlying COSMO calculations depend on the quantum chemical method as well as on the basis set, COSMOtherm needs a special parameterization for each method / basis set combination. All of these parameterizations are based on molecular structures quantum chemically optimized at the given method / basis set level. COSMO files shipped with COSMOtherm are available on various quantum chemical levels (COSMOthermX 2011).

Recommendations for which method to use depend upon the required quality and the later usage of the predictions. The application of COSMOtherm in chemical and engineering thermodynamics (*e.g.* prediction of binary VLE or LLE data, activity coefficients in solution or vapour pressures) typically requires high quality of property predictions of mixtures of small to medium sized molecules (up to 25 non-Hydrogen atoms). The recommended quantum chemical method for such a problem is a full TURBOMOLE BP-RI-DFT COSMO optimization of the molecular structure using the large TZVP basis set, in the following denoted BP-TZVP, and the corresponding parameter file (BP_TZVP_C21_0111.ctd) (COSMOthermX 2011). This is the parameterisation used for all calculations performed within this study.

For this purpose, first, one has to select the COSMO-files of all of the studied, or to generate the COSMO-files of the compounds that are not available in the databases. In the latter case, the calculation can be done within COSMOthermX using the "New molecule" tool, or externally using TURBOMOLE (or its graphical interface TmoleX).

Once all the molecules contained in the mixture are available and selected in the compound list, the user can specify which kind of property calculation he would like to perform from the larger section of the main window that offers a selection of property cards. Inside each card it is possible to adjust parameters like temperature, mole fraction, etc. Input settings from the property cards are transferred to the Property Selection panel with the ADD button. The COSMOtherm calculation is started from the RUN button in the Property section, from the RUN menu or from the shortcut bar.

By default, COSMOtherm produces two sorts of output files: The COSMOtherm output file filename.out and a file filename.tab, which contains the calculated property information in tabulated form. These files will automatically pop up after the calculation has finished. By default, only the average property of the compound is printed to the output and table file. The global keyword "*wconf*" toggles the printing of all conformer thermodynamic properties as well as conformer weights for all temperatures and mixtures to the output file (COSMOthermX 2011).

Additional output files will be written if the corresponding options in EXTRAS/GLOBAL OPTIONS are activated. These output files might contain σ-moments (.mom), atomic σ-moments (.moma), σ-profiles (.prf), or σ-potentials (.pot) (COSMOthermX 2011).

The phase definition subsections in all property panels have a context menu. With a right mouse button click, phase compositions can be copied and pasted into another property panel. More details about the computational details (explanations, examples and results) are given in the COSMOthermX Tutorial. The flowchart of a mixture property calculation within COSMOthermX is illustrated in Figure II-16.

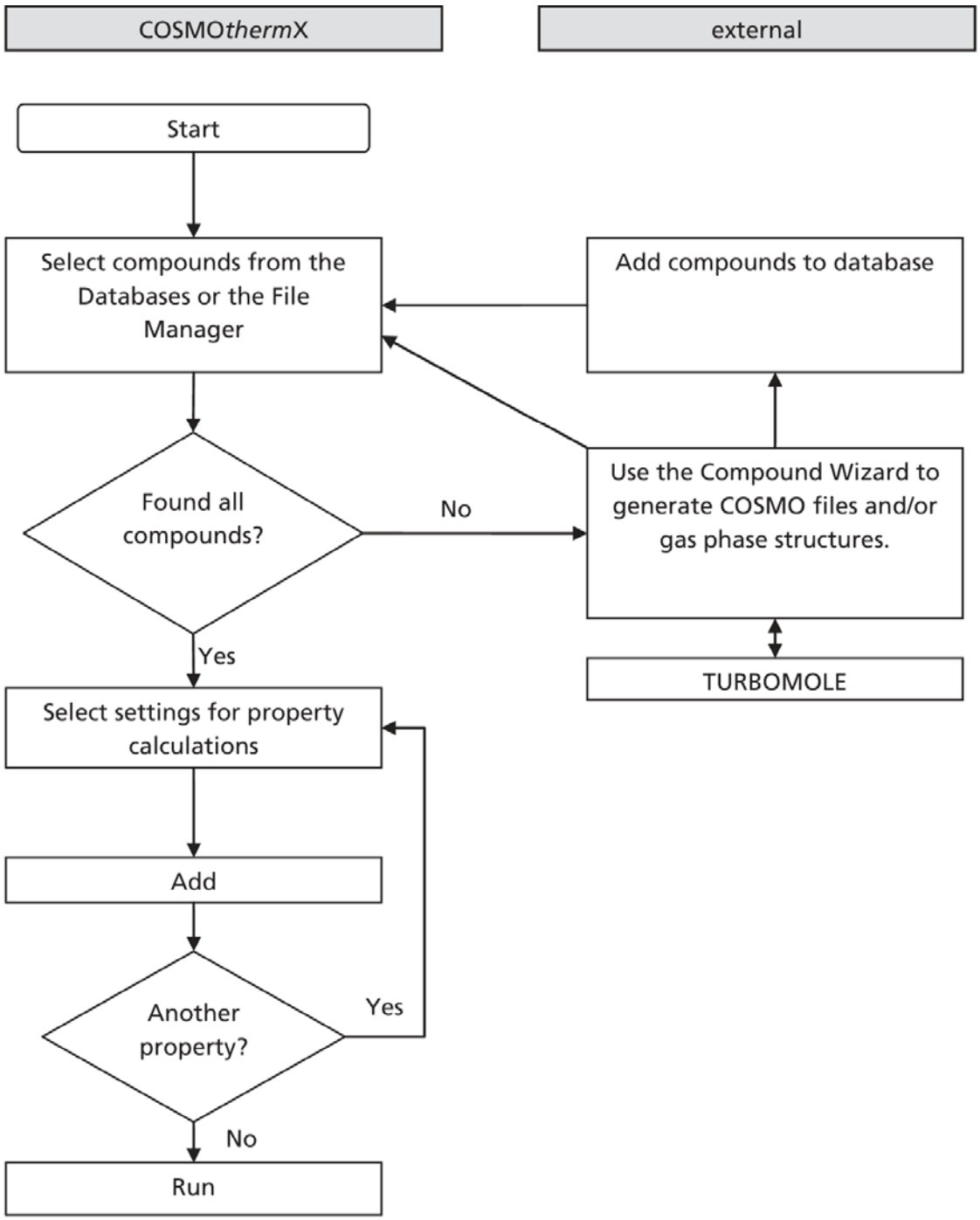


Figure II-16 : Flowchart of a property calculation with the COSMOthermX software. *TURBOMOLE is the software used to perform the quantum COSMO calculations (From COSMOthermX 2011)*

II.3. Concluding remarks about the COSMO-RS model

In summary, the COSMO-RS model uses input data (*i.e.* COSMO files) calculated using quantum calculations tools like the TURBOMOLE software or its graphical interface TmoleX. These data are used to perform thermo-statistical calculations using COSMOtherm (or its graphical interface COSMOthermX) in order to compute the equilibrium physical chemical properties of a given mixture, as illustrated in Figure II-17.

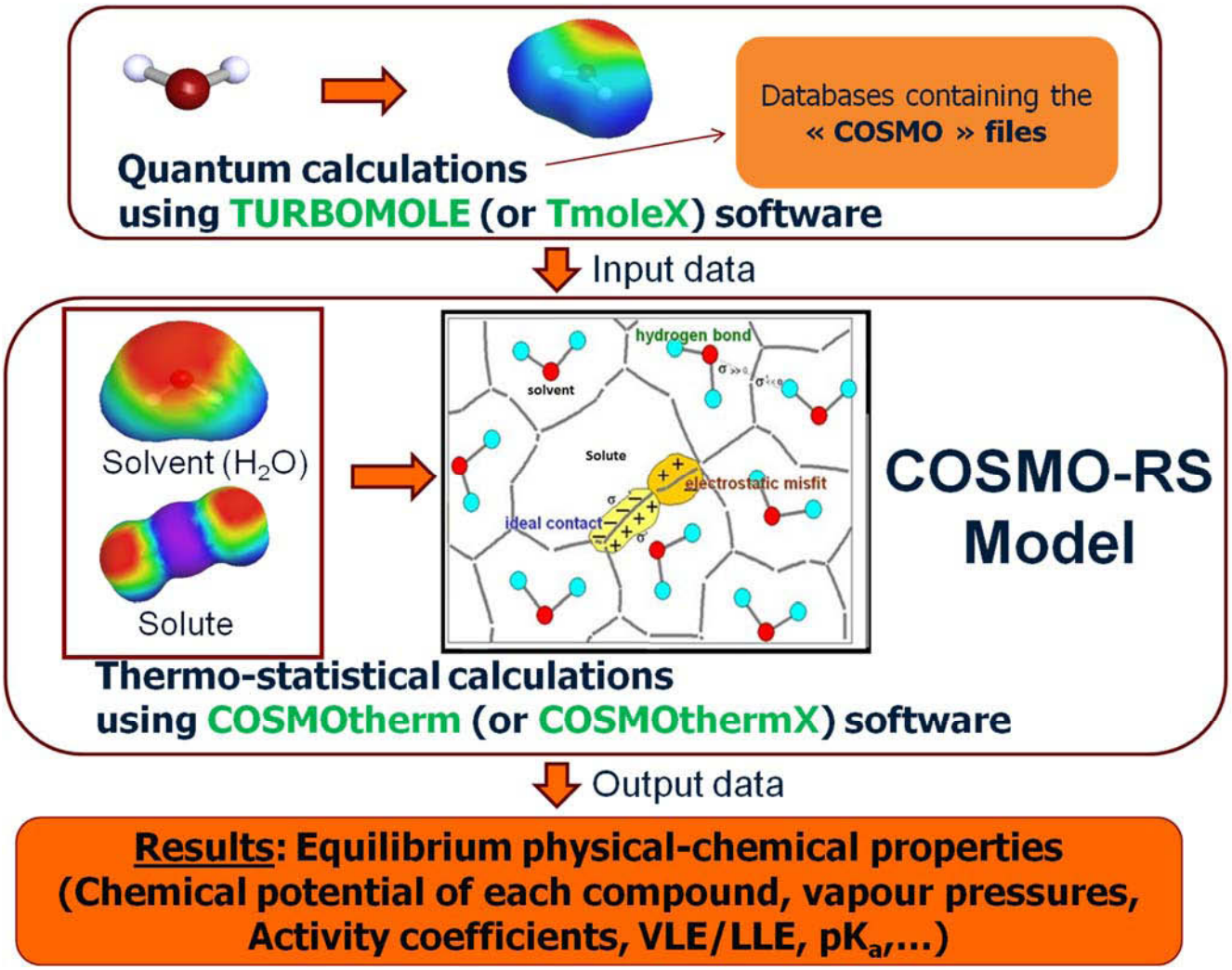


Figure II-17 : Summary of the COSMO-RS algorithm.

Chapter III:Prediction of activity coefficients of aqueous systems containing electrolytes

III.1. Introduction

When electrolytes are considered, the system is characterized by the presence of both molecular species and ionic species, resulting in three different types of interactions: ion-ion, molecule-molecule and ion-molecule. Ion-ion interactions are governed by electrostatic forces between ions that have a much longer range than other intermolecular forces. Molecule-molecule and ion-molecule interaction forces are known to be short-range in nature. The excess Gibbs energy of systems containing electrolytes can be considered as the sum of two terms, one related to long-range (LR) forces between ions and the other to short-range (SR) forces between all the species.

For predicting activity coefficients of strong electrolytes, Achard (1992) and Achard et al. (1994) have combined a chemical treatment of ions solvation with two activity coefficient models namely one based on the Debye-Hückel theory (Pitzer 1980, Pitzer 1973) which treats LR interactions and the UNIFAC-Larsen model (Larsen et al. 1987) which takes in accounts SR interactions. In this context, ions were considered as UNIFAC independent groups and the solvation of charged species giving clusters were taken into account by means of a hydration number for each ion at infinite dilution. With this approach two interaction parameters (water-anion and water-cation interactions) were sufficient to characterize a water-salt system (Achard et al. 1994). The resulting model (ULPDHS) has been adopted by ProSim S.A. and is available in the software 'Simulis Thermodynamics'. It can be used for multi-electrolytes solutions without adding any new interaction coefficient and satisfactorily predicts water activity, osmotic coefficients and salting-out effects in aqueous mixtures of two or three electrolytes within less than 5% even for saturated solutions.

In the present study, the predictive power of the COSMO-RS model developed by Klamt (1995, 2005 and 2011) is used to determine short range mean activity coefficients of salts in water-electrolyte binary mixtures. Contrarily to the work of Ingram et al. (2012) who have developed an extension of the COSMO-RS method to electrolytes solutions by optimizing several of its parameters before adding a PDH term, this work aims at introducing a fully predictive approach like that used in the ULPDHS model which should be more efficient in predicting activity coefficients of complex aqueous systems containing electrolytes.

To predict equilibrium properties of complex systems containing water, salts, sugars and other organic compounds, it is mandatory to predict activity coefficients in the simplest cases, *i.e.* a water-salt (like NaCl) binary system (Achard 1992; Achard et al. 1994; Ben Gaïda 2007). The activity coefficient modelling work described in this paper is focused on the

prediction of water-salt binary systems and should be considered as the first step of this challenge. For this purpose, as illustrated in Figure III-1, a consistent treatment of each species actually present in the solution must be undertaken.

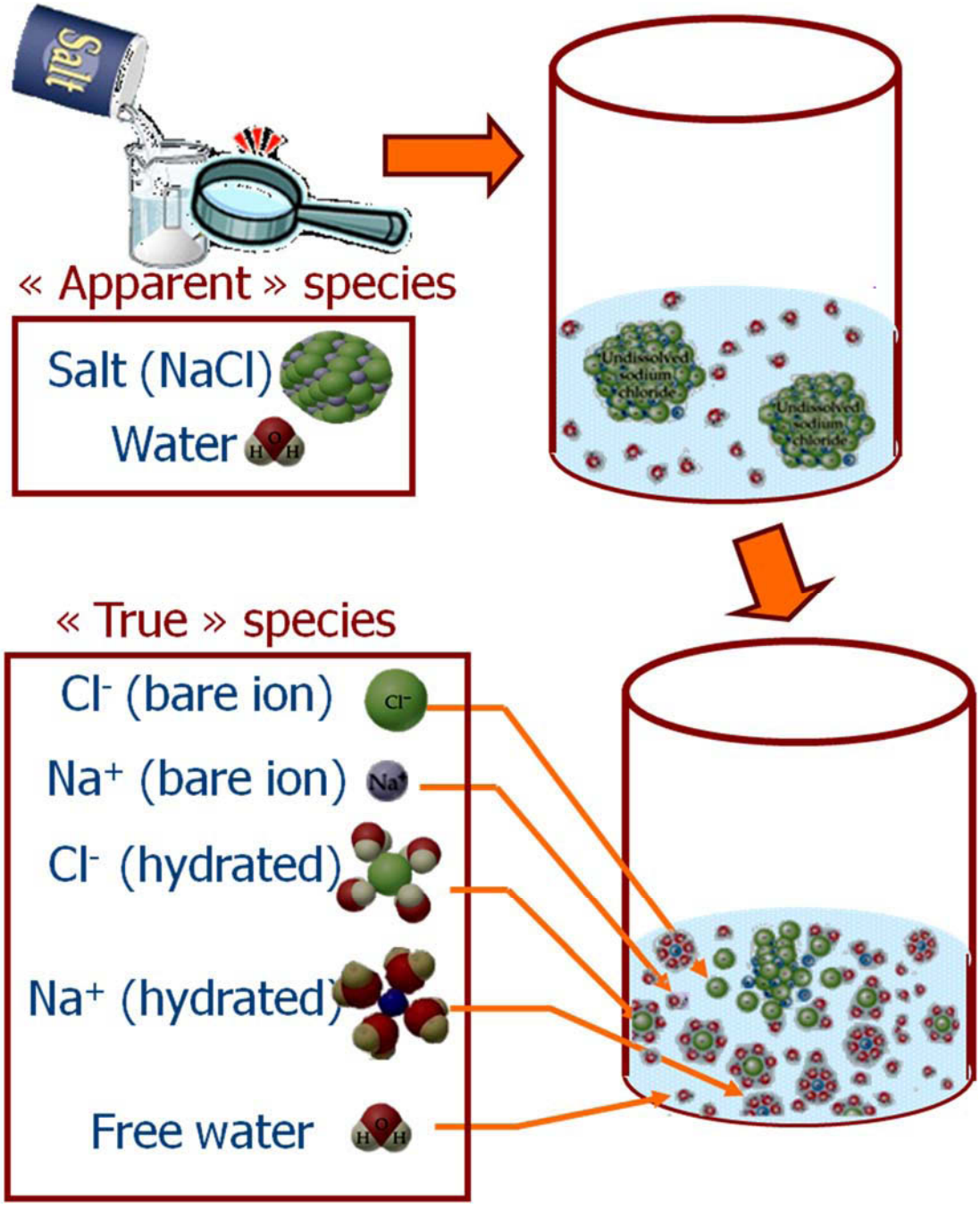


Figure III-1 : List of species that have to be taken into account in the illustrative case of a binary mixture water-NaCl.

Indeed, when a salt is mixed with water, it will get dissolved (partially or totally) and the electrolyte will dissociate to form ionic species that are hydrated by water molecules, as schematized in Figure III-1. Thus, if apparently the mixture contains only one salt and water molecules, in reality the species in solution include "free" species *i.e.* bare ions (cation and anion), hydrated clusters (formed by one ion and several water molecules linked to this ion)

and "free" water molecules. The simplifying assumptions used by Achard (1992) are used for this study. An average fixed number of water molecules solvating each ionic species is fixed. It accounts for a complex chemical equilibrium between several solvated species, the proportion if which depends on water activity. This kind so-called "variable solvation model" has been derived by Ben Gaïda (2007) for extending the domain of validity of ionic activity coefficients. As the structure of such a model is much more complex than a fixed solvation model, this will not be considered here. To characterize this equilibrium, several assumptions are made and adopted in this study:

- Electrolytes are assumed to be fully dissociated in the range of composition of interest (0- 6 moles of salt /kg of water, in molality scale) ;
- Partial dissociation (weak acids or bases) is accounted by solving the equilibrium equation with the activities of the ionic species.
- The hydrated ionic clusters contain a fixed number (n_{h_i}) of water molecules linked to the ion and anions are less hydrated than cations (Achard 1992, Ben Gaïda 2007).
- The COSMO-RS-PDHS model is used to compute $\gamma_i^{(x)}$ of water and individual ionic species. The knowledge of the latter allows the calculation of the water activity ($a_W = \gamma_i^{(x)} x_{H_2O}$) as well as the mean activity coefficient $\gamma_{\pm}^{(x)}$ of the salt *CA*. Thus, the molality based mean activity coefficient $\gamma_{\pm}^{(m)}$ of salt which is necessary to compare calculated and experimental data like those of Hamer and Wu (1972), is deduced using equations (III-1), where M_k denotes the molar mass of the solvent and m_{CA} denotes the molality of the salt.

$$\gamma_{\pm}^{(m)} = \frac{\gamma_{\pm}^{(x)}}{1 + \nu_+ m_{CA} M_k + \nu_- m_{CA} M_k} \text{ ; where } \gamma_{\pm}^{(x)} = \left(\gamma_+^{\nu_+} \gamma_-^{\nu_-}\right)^{1/(\nu_+ + \nu_-)} \qquad \text{(III-1)}$$

- If not mentioned otherwise, all the activity coefficients are assumed to be expressed in the mole fraction scale (*i.e.* $\gamma_i^{(x)}$), for clarity the superscript will be systematically omitted in this chapter. It is worth mentioning that the conversions of these values in the molality scale have been already discussed in Chapter I.

All the activity coefficients are defined with the mole fraction based activity coefficients $\gamma_i^{(x)}$.

III.2. Structure of the COSMO-RS-PDHS model

As illustrated in Figure III-2, the COSMO-RS-PDHS method is a combination of the COSMO-RS method (that takes into account short range physical interactions) with the un-symmetric Pitzer Debye-Hückel model (that treats long range physical interactions) and a chemical treatment of the solvation or clustering of ions (to take into account short range chemical interactions).

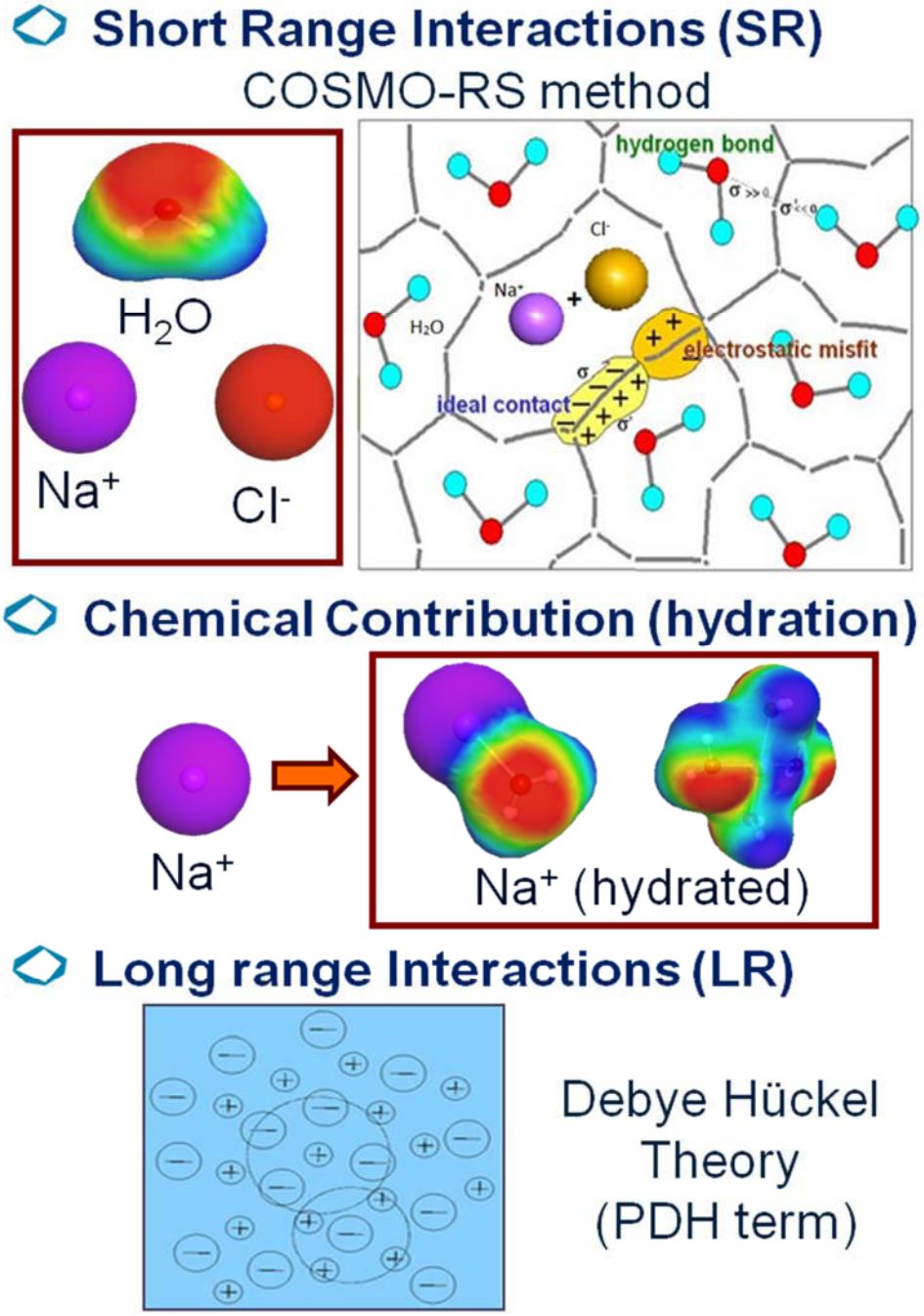


***Figure III-2** : List of interactions in solution that have to be taken into account in the illustrative case of a binary mixture water-NaCl.*

The goal of this predictive method is to use the predictive power of the COSMO-RS model to determine the short range (SR) activity coefficient $\gamma_i^{SR,H}$ (in the hydrated reference state) of a given compound *i,* as illustrated in the upper left side of Figure III-3. For this purpose, the ionic species are replaced by their corresponding clusters in COSMO-RS calculations and the molar fractions of each compound of the mixture are calculated using the equations developed in the chemical part of the ULPDHS model described in details by Achard (1992) and Achard et al. (1994), and discussed in the next section of this chapter.

Then, as illustrated in the upper right side of Figure III-3, long range (LR) interactions between ions of opposite charge are dominated by electrostatic forces and are accounted for by a PDH term γ_i^{LR} (in the non-hydrated reference state) since it is assumed that repulsive forces between ions of like charge are extremely large.

The activity coefficient γ_i of the combined model is calculated using Equation (III-2), which states that $\ln(\gamma_i)$ can be considered as the sum of two terms, one related to LR forces between ions and the other to SR forces between all the species,

$$\ln\left(\gamma_i\right) = \ln\left(\gamma_i^{SR}\right) + \ln\left(\gamma_i^{LR}\right) \qquad \text{(III-2)}$$

In this context, it is mandatory to establish the relations between the activity coefficients $\gamma_i^{SR,H}$ (in the hydrated reference state) and γ_i^{SR} (in the non-hydrated reference state) for two main reasons:

- To calculate the global activity coefficient resulting from the SR and LR terms, it is necessary to express both activity coefficients (γ_i^{SR} and γ_i^{LR}) in the same reference state;
- The activity coefficients and the molar fractions in the non-hydrated reference state are used in all thermodynamic relations related to the physical-chemical equilibrium properties of fluids, as discussed in Chapter I.

Thus, it is necessary to establish the relations to convert $\gamma_i^{SR,H}$ in a non-hydrated reference state corresponding to the classical molar fractions. The solvation (especially the hydration) phenomena depend on the concentration are concentration-dependent; when the ionic strength increases, the solvation of ions *i.e.* the (molar) number of water inside a cluster, decreases. The volume and the surface of the clusters can thus vary. However, this study treats only constant solvation (hydration) phenomena. This approximation is reliable if the electrolyte concentration is not very important; for very concentrated solutions, the number of water molecules included in the clusters could be overestimated and the correction becomes too high. Nevertheless, in the usual range of molality (0 to 6 mol/kg of water), this approximation is perfectly justified.

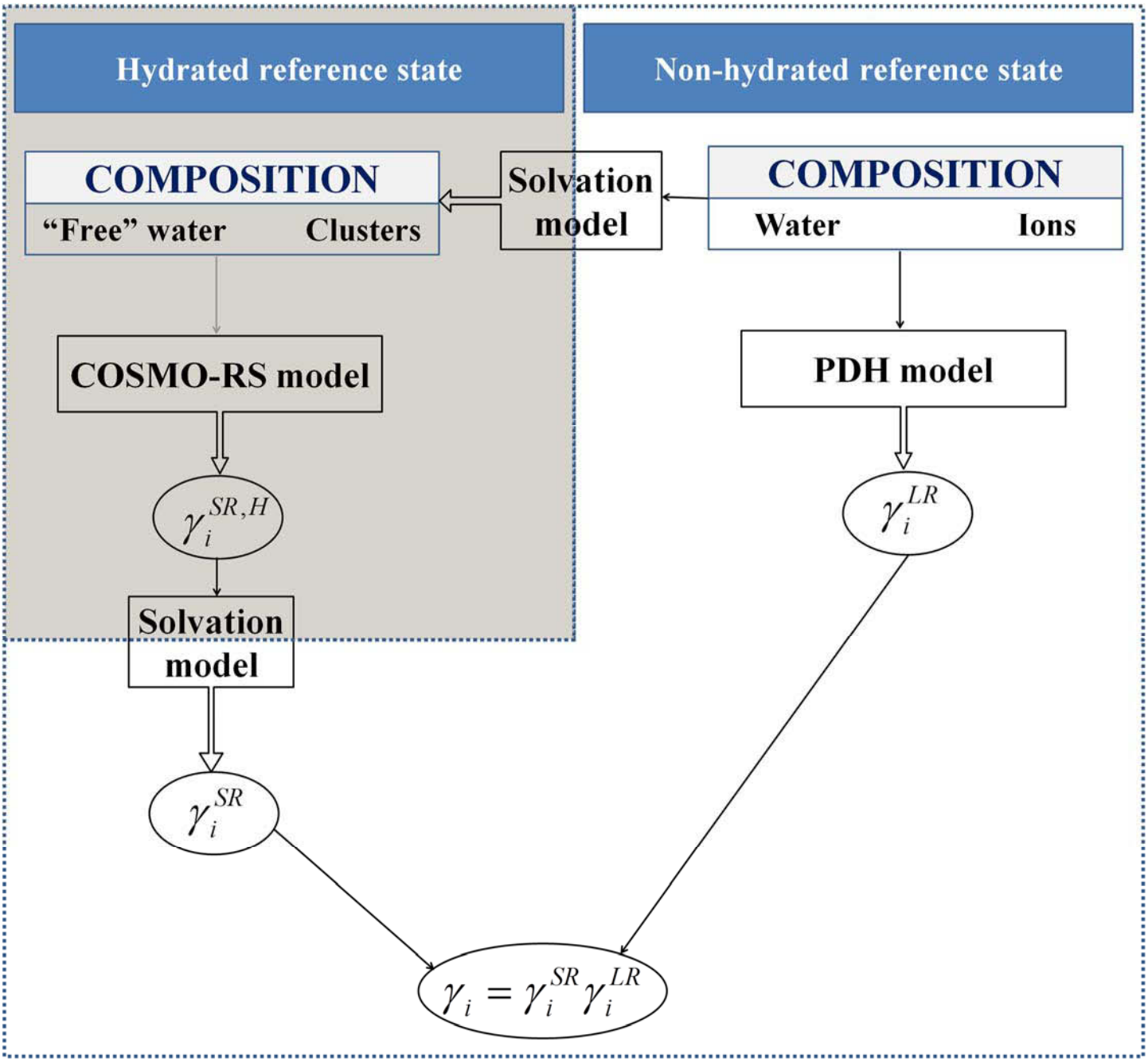


Figure III-3 : Principle of the determination of the activity coefficients using the solvation of charged particles.

The global activity coefficient γ_i is calculated as the product of SR and LR terms (since $\ln\gamma_i = \ln\gamma_i^{SR} + \ln\gamma_i^{LR}$). For this purpose, the hydrated composition of the mixture is calculated from the non-hydrated one using the solvation model. Then, the COSMO-RS model is used to predict the SR term in the hydrated reference state $\gamma_i^{SR,H}$. Finally, the solvation model is used (once again) to convert the latter to its corresponding value (γ_i^{SR}) in a non-hydrated reference state where it is combined to the LR term γ_i^{LR} resulting from the Pitzer-Debye-Hückel (PDH) model.

III.2.1. Determination of molar fraction in the hydrated reference state (x_i^H)

Let us consider a mixture containing N compounds each having a molar number (n_i). For water (i=1) and all the N-1 remaining species can be hydrated by water (1). The total molar number in solution is given by:

$$n_T = \sum_{i=1}^{N} n_i \tag{III-3}$$

The classical molar fractions x_i are calculated using the following equation:

$$x_i = \frac{n_i}{n_T} \tag{III-4}$$

When considering the solvation (hydrated reference state), the molar numbers of water in solution denoted as n_1^H, is not equal to n_1. Indeed, several water molecules are retained by the clusters and must be taken into account in the overall balance:

$$n_1^{\ H} = n_1 - \sum_{i=2}^{N} n_{h_i} n_i \tag{III-5}$$

Where n_{h_i} denotes the hydration number of the species i.

For the other species i ($i \geq 2$),

$$n_i = n_i^{\ H} \ , \text{for } i \geq 2 \tag{III-6}$$

Thus, the total molar number is modified in the hydrated reference state and given by

$$n_T^{\ H} = \sum_{i=1}^{N} n_i^{\ H} = \sum_{i=1}^{N} n_i - \sum_{i=2}^{N} n_{h_i} n_i \tag{III-7}$$

n_T^H can also be expressed as,

$$n_T^{\ H} = n_T - \sum_{i=2}^{N} n_{h_i} n_i \tag{III-8}$$

The molar fraction in the hydrated reference state can be calculated as:

$$\begin{cases} x_1^{\ H} = \dfrac{x_1 - \sum_{i=2}^{N} n_{h_i} x_i}{1 - \sum_{i=2}^{N} n_{h_i} x_i}, \text{for water } (i=1) \\ x_i^{\ H} = \dfrac{x_i}{1 - \sum_{i=2}^{N} n_{h_i} x_i}, \text{for the other species } (i \geq 2) \end{cases} \tag{III-9}$$

The model has a theoretical significance when $x_1 \geq \sum_{i=2}^{N} n_{h_i} x_i$ because x_1^H must be positive. This is the main limitation of the fixed solvation (hydration) assumption.

III.2.2. Determination of activity coefficients (γ_i^H) in the hydrated reference state

In order to establish analytic relation between $\gamma_i^{SR,H}$ and γ_i^{SR} , we use several thermodynamic relations discussed earlier in Chapter I.

For this purpose let us consider a binary mixture containing two species (1 and 2), 1 being able to hydrate 2. The total Gibbs energy of the solution is thus given by:

$$G_{sol} = n_1\mu_1 + n_2\mu_2 \tag{III-10}$$

Where μ_1 and μ_2 denote the respective chemical potentials of species (1) and (2); the latter are linked to activity values (a_1 and a_2) by equations (I-34) and (I-35), thus,

$$G_{sol} = n_1\mu_1^0 + n_2\mu_2^0 + RT\left(n_1 \ln a_1 + n_2 \ln a_2\right) \tag{III-11}$$

where μ_i^0 denotes the reference chemical potential of the species *i*, in its non-hydrated state.

For the hydrated species, the expression of G_{sol} can be reformulated as

$$G_{sol} = n_1^H \mu_1^H + n_2^H \mu_2^H \tag{III-12}$$

In this case, since species (2) is solvated by species (1), we have:

$$\begin{cases} n_1^H = n_1 - n_{h_2} n_2 \\ n_2^H = n_2 \end{cases} \tag{III-13}$$

For the solvent (species 1), the chemical potential μ_1^H is palpably identical to μ_1 since the chemical potential of the solvent cannot change with the model; their respective expressions can be rewritten as:

$$\mu_1 = \mu_1^0 + RT \ln a_1 \tag{III-14}$$

$$\mu_1^H = \mu_1^{0,H} + RT \ln a_1^H \tag{III-15}$$

Similarly, $\mu_1 = \mu_1^H$, leading thus to the equality between the activity coefficients values:

$$a_1 = a_1^H \tag{III-16}$$

However, for the ion (species 2), the chemical potentials μ_2 and μ_2^H are different since they are related to the bare ion and the solvated ion, respectively. Thus,

$$\mu_2 = \mu_2^0 + RT \ln a_2 \tag{III-17}$$

$$\mu_2{}^H = \mu_2^{0,H} + RT \ln a_2{}^H \tag{III-18}$$

The total Gibbs energy of the solution can be expressed in two different ways using equation (III-10) or equations (III-11 to III-13):

$$G_{sol} = \left(n_1 - n_{h_2} n_2\right)\mu_1^0 + n_2 \mu_2^{0,H} + RT\left[\left(n_1 - n_{h_2} n_2\right)\ln a_1{}^H + n_2 \ln a_2{}^H\right] \tag{III-19}$$

After identifications, we get:

$$\mu_2^{0,H} = \mu_2^0 + n_{h_2}\mu_1^0 \tag{III-20}$$

$$\ln a_2 = \ln a_2{}^H - n_{h_2} \ln a_1{}^H \tag{III-21}$$

Or,

$$a_2 = \frac{a_2{}^H}{\left(a_1{}^H\right)^{-n_{h_2}}} \tag{III-22}$$

Equation (III-18) enables to characterize the reference state of the hydrated species 2, relations III-15 and III-21 can be used to calculate the activity coefficients of species 1 and 2 as function of these of the hydrated species, *i.e.* the activities which are calculated with the solution model. These relations can be rewritten in terms of activity coefficients:

$$\gamma_1 = \gamma_1{}^H \frac{x_1{}^H}{x_1} \tag{III-23}$$

$$\gamma_2 = \gamma_2{}^H \frac{x_2{}^H}{x_2}\left(\gamma_1{}^H x_1{}^H\right)^{-n_{h_2}} \tag{III-24}$$

where $x_1{}^H$ and $x_2{}^H$ are given by equation III-9. It can be demonstrated that when $\gamma_1{}^H$ and $\gamma_2{}^H$ satisfy the Gibbs-Duhem relation, this is also the case for γ_1 and γ_2.

The latter relations (III-23 and III-24) can be generalized to a mixture containing *N* species (including N-1 species that can be hydrated) as:

$$\begin{cases} \gamma_1 = \gamma_1{}^H \dfrac{x_1{}^H}{x_1}, \text{for water } (i=1) \\ \gamma_i = \gamma_i{}^H \dfrac{x_i{}^H}{x_i}\left(\gamma_1{}^H x_1{}^H\right)^{-n_{h_i}}, \text{for the other species } (i \geq 2) \end{cases} \tag{III-25}$$

In the COSMO-RS-PDHS model, these equations are used to determine the respective compositions of “free” water and cluster that have to be defined as input in the COSMO-RS method (that takes into account short range physical interactions).

As discussed in Chapter I and II, several atom-based internal constants including the COSMO-radius of the elements (used for cavity construction; only 17% larger than Bondi radii) and the dispersion constants c_{vdW} and τ_{vdW} (one per element; vdW energy contributions expressed by element-specific parameters) are required in the COSMO-RS algorithm. For the most common elements, these constants have been determined once and have since been improved in several revisions. However, for the almost of ionic species encountered in the aqueous electrolytes systems, this parameterization remains to be performed though reasonable guesses are proposed in the original parameterization of the COSMO-RS method (Klamt et al. 1998) and its latest improvements (Klamt 2005, 2011). Furthermore, contrarily to organic molecules, which have a well defined structure (*e.g.* water molecule contains two hydrogen atoms covalently bound to a single oxygen atom) that can be ubiquitously determined in any quantum calculations, the structures of hydrated ionic clusters (formed by a given ion and a fixed number (n_{h_i}) of water molecules) remain to be established in literature. Indeed, the value of n_{h_i} depends on hydration properties of the mixture (concentrations, water availability, activity coefficients of each species, etc.), which in turn depend on n_{h_i}.

To overcome this problem, we have decided to investigate in a first step the maximum value (n) of the number n_{h_i}. obtained at infinite dilution in water in the primary hydration shell of the studied ion (*i.e.* by assuming that there are always enough water molecules in the mixture that can be inserted in the hydration shell of a given ion, if the first hydration shell is full this water molecule will be inserted in the second hydration shell). Finally, to use these ionic clusters in the COSMO-RS algorithm, all the elements specific parameters (COSMO-radii and dispersion constants) are required even if there are not available in the literature.

Since in the original parameterization of the COSMO-RS model, reasonable guesses are proposed for the dispersion constants, the following study will be focused on the determination of the COSMO radii of the elements of IA and II groups, as well as the prediction of the value of the hydration number n. In this context, a new prediction method (including four different and complementary procedures to determine n) has been developed, as discussed below.

III.3. Determination of the structures (hydration numbers and COSMO-radii) of several cations

In aqueous solutions, electrolytes *i.e.* charged species are often hydrated by water molecules that take, near ions, orientations to form generally a structure of coordination, where hydrogen or oxygen atoms of the water molecule are pointing in direction of the anions or cations, respectively (Israelachvili 2011). The formed ensemble "ion- water" is called an aggregate or a cluster. The averaged number of water molecules that are fixed to the studied ion, forming the cluster is called the hydration number *n*. At infinite dilution in water, this number has of course its highest value. It decreases when the ionic strength increases due to the fact that steric hindrance between clusters becomes important (Israelachvili 2011). This indicates that the solvation of ions depends on the electrolytes concentration. However, it is important to notice that these bound water molecules are not completely immobilized and they do move and exchange with bulk water, albeit more slowly (Israelachvili 2011). So there are several definitions of the ions hydration number in the literature (Ohtaki and Radnai 1993, Chipot and Pohorille 1998, Zavitsas 2001, Israelachvili 2011, Bockris and Reddy 1970, Partanen and Minkkinen 1993). There are also disparities regarding the experimental methods or techniques used (Israelachvili 2011, Ohtaki and Radnai 1993, Bockris and Reddy 1970, Partanen and Minkkinen 1993, Buchner et al. 1999, Kiriukhin and Collins 2002) and theoretical calculations used (Pye et al. 1996, Hashimoto and Kamimoto 1998, Nielsen et al. 1999, Driesner et al. 2000, Pratt and Rempe 1999, Rempe et al. 2000, Whitfield et al. 2007, Varma and Rempe 2006, Beglov and Roux 1994, Aaqvist 1990, Tunell and Lim 2006).

For instance, Zavitsas (2001) defines the hydration number of an ion as the average number of water molecules that are sufficiently linked to the solute until being part of the latter. That enables to reduce the number of water molecules constituting the "bulk" water, or the free solvent, while the number of solute particles does not change in the mixture. Other authors (Chipot and Pohorille 1998, Ohtaki and Radnai 1993) define hydration number as the average number of oxygen atoms in water located at a given distance from the solute or as the number of nearest neighbour water molecules around an ion.

Experimentally hydration numbers are deduced from measurements of the viscosity, diffusion, compressibility, conductivity, solubility, and various thermodynamic and spectroscopic properties of electrolytes solutions (*e.g.* activity coefficients) (Israelachvili 2011). Conductance and diffusion studies, dispersion techniques, dielectric relaxation, neutron and X-ray diffraction, X-ray absorption, infrared and nuclear magnetic resonance

(NMR) spectroscopy are several experimental techniques used to perform such measurements (Ohtaki and Radnai 1993). There are disparities between the definitions of hydration numbers used in these different techniques that induce disparities between the experimental values of the hydration number for a given ion.

For instance, conductance and diffusion experimental studies, define the hydration number as the number of water molecules that lose their translational freedom and move with the ion (Bockris and Reddy 1970). In dispersion techniques, water seems to hydrate ions in their quasi-spherical shell; these techniques give mainly information about the primary shell of hydration (Partanen and Minkkinen 1993) and define the number of water molecules in this shell as the hydration number. Dielectric relaxation techniques define the hydration number as the average number of water molecules that are bonded non-rotationally to the solute and do not contribute to the solvent relaxation process (Buchner et al. 1999).

Kiriukhin and Collins (2002) have studied the dynamic hydration of ions after defining the apparent dynamic hydration number as the number of water molecules strongly linked that have to emphasis with the ion to give its apparent molecular weight in a size excluding cell. Neutron and X-ray diffraction techniques determine differential scattering cross-sections that are proportional to weighted sums of partial structure factors in polyatomic systems, like salt solutions (Ohtaki and Radnai 1993, Bockris and Reddy 1970). These partial structure factors can yield accurate partial pair-distribution functions between ions and water oxygen atoms. The latter also called radial distribution functions *g(r)* should be integrated up to their first absolute minima and then define the nearest volume of coordinating solvent molecules to obtain hydration numbers of ions (Ohtaki and Radnai 1993, Bockris and Reddy 1970).

Likewise, theoretical calculations make distinction between primary and secondary hydration shells (Pye et al. 1996, Hashimoto and Kamimoto 1998, Nielsen et al. 1999). Hydration numbers can be calculated using *ab initio* molecular dynamic (AIMD) simulations which can produce radial distribution profiles that clearly define inner coordination shells. The integration of these curves up to their respective first minima allows the determination of hydration numbers of ions (Driesner et al. 2000). Using quantum mechanical methods, Driesner et *al.*(2000) have studied with the density functional theory (DFT) the hydration complexes of a set of ions (including Li^+, Na^+, K^+ and Mg^{2+}) from ambient to supercritical conditions, in order to understand how the process of hydration occurs in aqueous solutions and to understand the fractionation of oxygen and hydrogen isotopes in electrolytes solutions. For this purpose, they have reported several hydration numbers values of the studied ions and have optimized the structures of each complex before computing harmonic vibrationnal frequencies using the results to calculate thermo-chemical properties. These properties were

used to explore the molecular causes of oxygen and hydrogen isotopes "salt effects" that are an important isotope fractionation mechanism in a variety of geo-chemical processes (Driesner et al. 2000). Another method to determine the hydration number is to use the quasi-chemical organization of solution theory (Pratt and Rempe 1999). This approach has been successfully used to determine the aqueous hydration number of cations, such as Li^+ (Pratt and Rempe 1999, Rempe et al. 2000) and K^+ (Whitfield et al. 2007).

In summary, the hydration of group IA and IIA cations has been investigated in a plethora of theoretical and experimental studies (Israelachvili 2011, Tunell and Lim 2006, Ohtaki and Radnai 1993, Whitfield et al. 2007, Varma and Rempe 2006, Rempe et al. 2000, Zavitsas 2001, Hancock et al. 2004, Bockris and Saluja 1972, Boda et al. 2012, Marcus 2012, Marcus 1997, Pratt and Rempe 1999, Driesner et al. 2000). To illustrate the disparities between theoretical and experimental methods results which are not completely well understood, several of these literature data are summarized in Table III-1.

***Table III-1*: Some hydration numbers values collected from literature data.**

**Number of water molecules forming a stoichiometric complex with the ion, e.g. $[Mg(H_2O)_6]^{2+}$. The data are taken respectively from [a]Ohtaki and Radnai 1993, [b]Bockris and Reddy 1970, [c]Varma and Rempe 2006, [d]Tunell and Lim 2006, [e]Driesner et al. 2000, [f]Kiriukhin and Collins 2002, [g]Pratt and Rempe 1999, [h]Rempe et al. 2000, [i]Whitfield et al. 2007, [j]Hancock et al. 2004, [k]Israelachvili 2011, [l]Boda et. al. 2012, [m]Marcus 2012. [n] N.A denotes 'Not Available' in the cited literature data.*

	Experimental results prior to 1996	Experimental results From 1996 to 2006	Results from theoretical approaches prior to 2006	Most recent values 2011
H_3O^+	3[a], 2[a]	N.A.[n]	N.A.[n]	3 (±1)[k]
Li^+	3-6 [a,b,c]	4[d], 4-6[e], 6[c,f]	4-6[d,e], 4[c,g,h]	5 (±1)[k]
Na^+	4-8 [a,b,c]	4[c, d], 5-6[e], 6[d,f]	5-6[e], 4-6[c,d]	4 (±1)[k]
K^+	3-6 [a,b,c]; 8 [b,c]	4-8[e], 4[d], 5[c], 6[d,f]	5-7.8[e], 4-8[c,i], 4-6[d]	3 (±1)[k]
Rb^+	3.5 [a]	3.5[d]	4-6[d,e]	N.A.[n]
Cs^+	5.3-8.2 [d]	N.A.[n]	N.A.[n]	1 (±1)[k]
Be^{2+}	4 [a], 6 [a]	4[d,f]	4[d]	4* (±1)[k]
Mg^{2+}	6 [a]; 10 [b]	6[d,e,f]	6[d,e,j]	6* (±1)[k]
Ca^{2+}	6 [a]; 9 [b]	10[f], 6-8[d]	6-8[d,e]	6 (±1)[k]
Sr^{2+}	8.2-9.8 [a,]	10[b], 8[d]	7-8[d]	8[l]
Ba^{2+}	9.5 [a]; 9.7 [a]	10[b], 9.5[d]	7[d], 9[d]	10.3[m]

However, most of these experimental or theoretical calculation methods are computer-time consuming and there are lots of factors and assumptions that should limit the reliability

of the measured or calculated hydration number values. Indeed, Varma and Rempe (2006) have demonstrated that molecular dynamics employing force fields (like CHARMM 27 package (Beglov and Roux 1994) and OPLS-AA package (Aaqvist 1990) which are widely used in theoretical biology) are suitable for simulating the energetic properties of ion solvation but not for predicting structural properties such as radial distributions of ligand molecules around ions. Obviously these parameterized force fields should only be used to predict the properties of ions solvation (hydration free energies and ionic radii) for which they were originally parameterized, and should not be used for predicting other solvation properties like hydration numbers. Varma and Rempe (2006) also pointed out lots of difficulties that make diffraction measurements on liquids particularly challenging and data analysis efforts generally complicated, even though the latest developments have significantly improved these methods both in terms of data gathering and data analysis (Aaqvist 1990). *Ab initio* molecular dynamic simulations involve generation of dynamical trajectories, sampling and system size are thus two issues that can potentially affect the reliability of the *ab initio* (quantum) investigations of ion solvation.

As discussed in chapter II, the COSMO-RS method is a combination of the COnductor-like Screening Model (COSMO, a variant of dielectric continuum models) with a statistical thermodynamic treatment of interacting surfaces for Realistic Solvation (the RS part of the COSMO-RS model). It uses only atom-specific parameters and can be used to predict the thermodynamic properties of solvent liquid systems (including ionic liquids). Therefore, it is anticipated that this model is, at least qualitatively, able to describe structural variations correctly. COSMO-RS is described as the most accurate quantum chemically based model for the prediction of solvation energies (Klamt 2011). It also permits the prediction of all kind of thermodynamic equilibrium properties of liquids (Klamt 2005 and 2011).

However, the accuracy of COSMO-RS depends strongly on the quantum chemical method used (*i.e.* the so-called COSMO calculations). In order to perform these calculations, the COSMO-radii of each atom of a given molecule must be defined. For the common atoms encountered in the most common organic molecules (namely H, C, N, O, F, P, S, Cl, Br or I), the COSMO-radii values are calculated from van-der-Waals radii (Bondi 1964, Mantina et al. 2009). Klamt (1998, 2005, 1993 and 2011) stated that the optimum COSMO-radii values are about 1.17 times the van-der-Waals radii. However for the elements studied here, only reasonable guesses are proposed in the Turbomole user manual (Turbomole 2007), and in COSMO or COSMO-RS related papers (Eckert and Klamt 2010, COSMOthermX 2011, Eckert and Klamt et al. 2002, Klamt et al. 1998, Klamt 2005, Klamt and Schüürmann 1993, Klamt 2011). These values are only based on the covalent radii suggested by Sutton (1965)

for Li, Na, Be and Mg atoms, and no value is proposed for K, Rb, Cs, Ca, Sr or Ba. One of the main goals of this study is to determine a suitable set of COSMO-radii for the main elements of the IA and IIA groups.

For this purpose, a new prediction method of hydration numbers requiring a little computational effort is introduced and used to compare the hydration numbers determined in the gas phase and those computed at infinite dilution in water is performed. To do so, the predicting power of the COSMO-RS method is used for simulating hydrated species, and then to determine the hydration number *n*, *i.e.* the number of water molecules bound to solvated ions or hydrated ions, at infinite dilution in water. Likewise, several procedures based on the analysis of the quantum results obtained from DFT calculations using the same parameterization level like the COSMO-RS model are introduced and then used to deduce the corresponding hydration number of a given cation in the gas phase. For the studied cations, this number (*n*) is also called coordination number. The present study is focused on the determination of the hydration numbers of several cations of interest in foods and biological science (namely Li^+, Na^+, K^+, Mg^{2+} and Ca^{2+}), as well as other cations like Rb^+, Cs^+, Be^{2+}, Sr^{2+} and Ba^{2+} to complete the IA and IIA columns and then to prove the predictive power of our method. The specific case of the proton H^+ replaced by its most stable form (*i.e.* the hydronium ion H_3O^+) is also treated.

III.3.1. Materials and methods

III.3.1.1. Determination of the COSMO-radii of the cations

The strategy used in this study to determine the COSMO-radii of all studied cations starts with the gas phase calculations where DFT results (mainly energies and optimized geometries) are independent of ion radii. Thus, by analyzing the calculated mean distances between the oxygen atom of one water molecule and a given ion (inside the cluster), it becomes possible to choose a set of radii that are in agreement with these values. Then, the scaling factor (Klamt 2011, Eckert and Klamt 2010, Klamt 2005) of 1.17 times the radii is applied to the chosen values in order to determine COSMO radii of ionic species.

III.3.1.2. Hydration number prediction procedures

The hydration number prediction method developed in this work could be synthesized in several steps described below.

First, one has to generate the gas phase geometric structures of all the successive hydrated complexes formed by this ion and *j* water molecules that are bound to it. All hydrated species of ions are considered as clusters in which water molecules are bonded to the studied ion.

These clusters are also called the stoichiometric complexes formed by a number of water molecules and the ion. This task was done using TmoleX software (Steffen et al. 2010), a graphical interface of TURBOMOLE (Turbomole 2007).

In this software, it is possible to build the geometry of any molecule, especially clusters in this case; the default values of bond lengths and angles were used to generate initial geometries using the 'clean up molecule' facility (Steffen et al. 2010). Then, the gas phase energy and the geometry optimizations were performed until convergence is reached. The optimized gas phase geometry, corresponding also to the minimum energy, is thus considered as the most probable structure of the studied cluster; it is then used to perform a single point COSMO energy calculation in order to generate the COSMO-file that will be used to compute the aqueous phase properties using the COSMO-RS method.

All the quantum level calculations of the present study were done at the large TZVP auxiliary basis sets (Eichkorn et al. 1997), in the following denoted BP-TZVP (Eichkorn et al. 1997) (*i.e.* B88-VWN-P86 functional and 'def-TZVP' basis set), using a full TURBOMOLE BP-RI-DFT COSMO optimization of the molecular structure, (COSMOthermX 2011, Steffen et al. 2010, Turbomole 2007). It has been demonstrated that careful optimization of the auxillary basis sets leads to considerable gains in performance and only marginal sacrifices in accuracy as compared to the conventional technique based on four-centre integrals. For instance, Eichkorn et *al.* (1997) have demonstrated that errors resulting from the RI-J approximation (used in TmoleX software) could be made sufficiently small in a study of auxilliary basis sets for the atoms H to At - excluding the Lanthanides - optimized for an efficient treatment of molecular electronic Coulomb interactions. They presented accuracies and timings which demonstrate that the RI-J method allows for the treatment of molecules with up to 300 atoms and/or 2500 basis functions in moderate symmetry on work stations (Eichkorn et *al.* 1997).

As illustrated on Figure III-4 (for the case of Mg^{2+} ion) the hydration number prediction method introduced in this paper, starts from gas phase calculations (where three complementary procedures are proposed) up to the prediction of hydration number in aqueous solution at infinite dilution, using the predictive power of the COSMO-RS method.

III.3.1.2.1. Gas phase mean ion-water distances analysis procedure

This procedure is based on the analysis of the evolution of the averaged mean distance ion-water $d_{ion\text{-}water}$ inside the cluster $[Ion, (H_2O)_j]^{q+}$ formed by a given ion (of charge q^+) and j water molecules. For a given ion, by plotting the (arithmetic) mean value of $d_{ion\text{-}water}$ versus the number j of water molecules inside the cluster, a clear distinction between the first and second hydration shells can be observed. It thus becomes possible to count the maximum

number of water molecules included in the first shell that is defined by several authors as the hydration number of the studied ion. To avoid confusions in this paper, the latter will be called 'volume-based' hydration number and denoted as $n_{volumic}$.

III.3.1.2.2. Gas phase structures analysis procedure

Since the molecular structures are available (as results of the quantum calculations), one can look for the number of water molecules that are enough strongly bound to a given ion, allowing thus the determination of another hydration number value that should be less or equal to the volume-based one. Indeed, when one water molecule is not directly bound to the studied ion inside its hydrated cluster, this water molecule forms a hydrogen bond with at least another water molecule which is covalently bound to the ion.

The gas phase binding energy ($\Delta E_{binding,\,j}$) for the formation of the j^{th} cluster is given by:

$$\Delta E_{\text{binding,j}} = E_{gas}\left(\left[Ion,\left(H_2O\right)_j\right]^{q+}\right) - E_{gas}\left(\left[Ion,\left(H_2O\right)_{j-1}\right]^{q+}\right) \qquad \text{(III-26)}$$

where E_{gas} denotes the DFT quantum calculated energy in vacuum.

Thus, the combination of the study of the successive binding energies values and the fact that a covalent bond is stronger than a hydrogen bond allows the estimation of a reliable value of the hydration number of a given ion. To avoid confusions, this 'molecular structure-based' hydration number will be denoted as $n_{structure,gas}$ in this study.

III.3.1.2.3. Gas phase thermodynamic reaction properties analysis procedure

This procedure is based on the mixture properties analysis *i.e.* the thermodynamic properties of the successive hydration reactions of formation of the complex $[Ion, (H_2O)_j]^{q+}$ formed by a given ion (of charge q^+) and j water molecules. For this purpose, the Gibbs free energy ΔG_j^{*react} of each hydration reaction is computed using the following equations.

The first hydration reaction is given by:

$$Ion^{q+} + H_2O \rightleftharpoons \left[Ion,\left(H_2O\right)\right]^{q+} \qquad \text{(III-27)}$$

In a more general form, the successive hydration reactions are given by:

$$\left[Ion,\left(H_2O\right)_{j-1}\right]^{q+} + H_2O \rightleftharpoons \left[Ion,\left(H_2O\right)_j\right]^{q+} \text{ for } j = 1,2,...,n \qquad \text{(III-28)}$$

where q^+ denotes the charge of the cation and j the number of water molecules inside the cluster, *i.e.* $j = 0$ for bare ion and $j = 1, 2, ...$ for the successive hydrated complexes. The hydrated cationic species are defined herein as clusters having a covalent bond between the ion atom and the oxygen atom of each water molecule included in any hydrated cluster. This

assumption seems to be reliable since the water molecules (in the primary shell) are strongly linked to the ion (solute), until being part of the latter.

Finally, an analysis of the incremental Gibbs energies of the successive hydration reaction ΔG_f^{gas} of each reaction j, calculated by using equation III-29, allows the determination of the gas phase hydration number of the studied ion. To avoid confusions, the latter will be called 'gas phase thermo-chemistry based' hydration number and denoted as $n_{thermo,\,gas}$.

$$\Delta G_j^{react,gas} = \sum_k \nu_k \Delta G_k^{gas} \; (Hess\,law) = \sum_k \nu_k \left(\Delta E_k^{elec,f} - T\Delta S_k^{gas,f} \right) \qquad \text{(III-29)}$$

For a given molecule k, ν_k, ΔG_f^{gas}, $\Delta E_k^{elec,f}$ and $\Delta S_k^{gas,f}$ denote, respectively, its stoichiometric coefficient (positive for the products and negative for the reactants), its Gibbs free energy (in the gas phase), its electrostatic energy of formation and its entropy of formation. T denotes the temperature (within this study, T=298.15 K). These data were calculated using the reference elements (crystalline forms) recommended by the NIST (Irikura et al. 1998), and the computed values of the absolute entropies $S_k°$ of the clusters. The latter is calculated as a sum of translational, rotational and vibrationnal terms (Irikura et al. 1998).

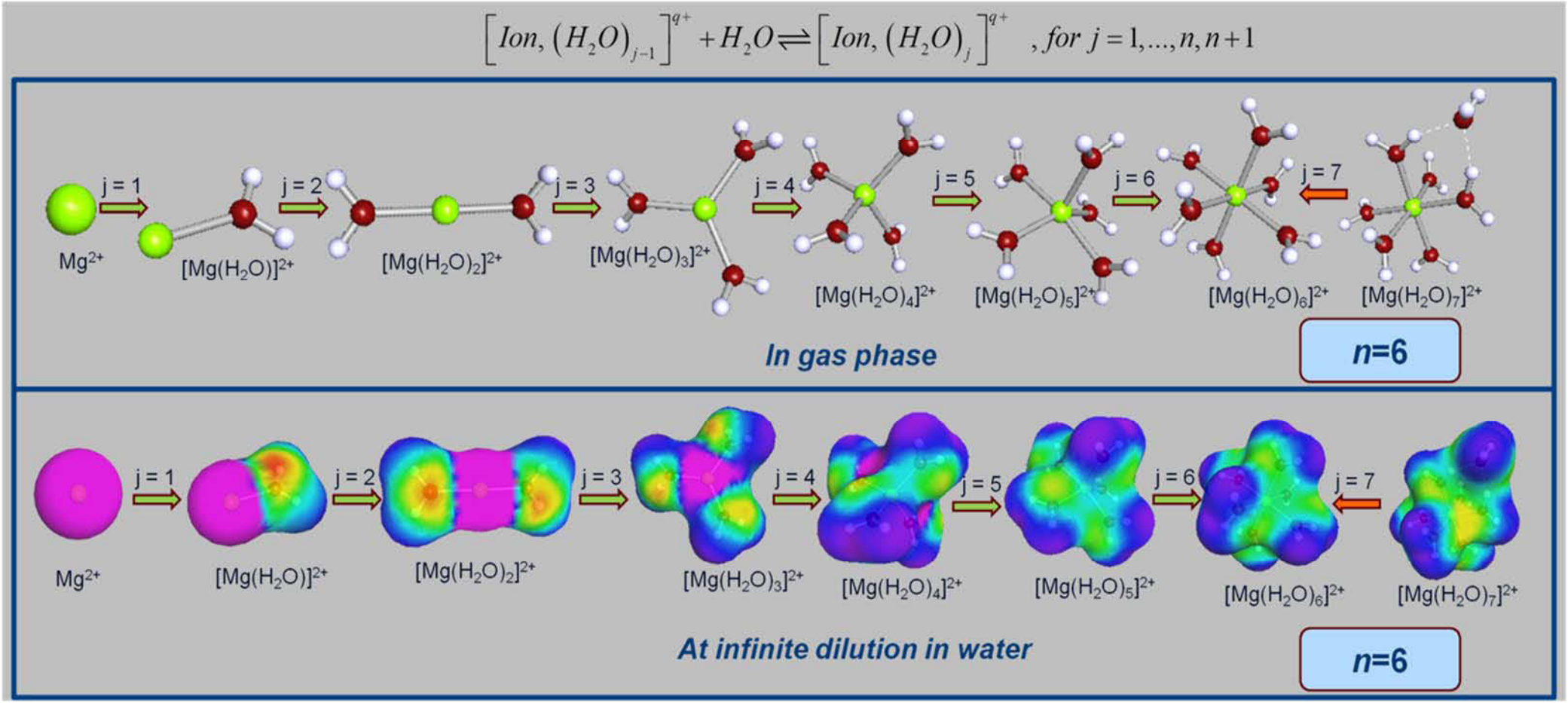


Figure III-4 : Different species observed during the successive hydration reactions of Mg^{2+} ion.

j = 0 corresponds to the bare ion, that reacts with one water molecule to form the first complex $[Mg(H_2O)]^{2+}$. The others complexes $[Mg(H_2O)_j]^{2+}$ where j denotes the number of water molecules hydrating the ion are also shown.

III.3.1.2.4. Aqueous phase thermodynamic reaction properties analysis procedure

The COSMO-RS method uses a robust statistical thermodynamics procedure to compute the chemical potential μ_i^S of a given molecule *i* in solvent *S* (Klamt et al. 1998, Klamt 2005). In reality μ_i^S is a pseudo-chemical potential that should be denoted as μ_i^{*S} (as defined by Ben-Naim (1987 and 2001) to which a term $RTln(x_i)$ must be added to get the absolute chemical potential values (Klamt 2005). Because this study is performed at infinite dilution in water, this term cancels out between two successive hydrated species of the same ion. If $\mu_i^{*\infty}$ denotes the infinite dilution pseudo-chemical potential of a given compound in solution, the aqueous phase binding energy (at infinite dilution in water) for the formation of the jth cluster is given by:

$$\Delta\mu^*_{binding,j} = \mu^{*\infty}\left(\left[Ion,\left(H_2O\right)_j\right]^{q+}\right) - \mu^{*\infty}\left(\left[Ion,\left(H_2O\right)_{j-1}\right]^{q+}\right) \quad \text{(III-30)}$$

which gives a reasonable indication about the occurrence of the reaction of formation of species *j*.

The whole process is no more than an exchange of two water molecules and cannot result in any overall change of free energy. However, a water molecule in bulk water is in a different state from a water molecule near an ion (Israelachvili 2011). $\Delta\mu_{binding,j}^*$ could thus be considered as the free energy required for adding one extra water molecule to the hydrated cluster $[Ion, (H_2O)_{j-1}]^{q+}$. At long distances compared to the radius of the ion, the water molecules have to be randomly oriented relative to the ion, and then, the interaction energy has to be zero; in other words, when $\Delta\mu_{binding,j}^*$ becomes positive (higher than the ion-water interaction energy in bulk water), the formation of the jth cluster would scarcely take place. Thus, the hydration number of the ion is the number of water molecules (*n*) contained in the reactant (solvated form number $n = j-1$ or $[Ion, (H_2O)_n]^{q+}$) of this ion.

To avoid confusions, the value of *n* determined using this procedure, will be called 'infinite dilution in water thermo-chemistry based' hydration number and abbreviated as n_{thermo}^{∞}.

III.3.2. Results and discussion

III.3.2.1. Gas phase results

III.3.2.1.1. Gas phase mean ion-water distances analysis results

The averaged distances between the ions and the oxygen of water molecules $d_{ion\text{-}water}$ values from literature data and those calculated in this study are given in Table III-2. The evolution of the mean distance ion-water (values given in Table III-2) compared to the number j of water molecules inside the cluster is shown in Figure III-5. For each of the studied ions, this figure points out the difference between its first and second hydration shells, as well as the corresponding values of the hydration number $n_{volumic}$. As results one can conclude that the maximum number of water molecules that can be inserted in the primary hydration shell of the studied monovalent ions are respectively $n_{volumic}$= 3 for H_3O^+, 4 for K^+, Rb^+ and Cs^+, and 6 for Li^+ and Na^+. Likewise, for divalent ions, the estimated hydration numbers are: $n_{volumic}$= 4 for Be^{2+}, 6 for Mg^{2+}, 8 for Ca^{2+} and Sr^{2+} and 9 for Ba^{2+}. These results are in good agreement with the literature data (shown in Table III-1).

Table III-2: Mean distances between ions and the oxygen of water molecules $d_{ion\text{-}water}$ values from literature data and those calculated in this study.
The corresponding hydration number values $n_{structure,gas}$ and coordination number (CN) by Tunell and Lim(2006) are given in brackets. The first column contains the ionic radii values, published by Marcus (1988 and 2012). These values have to be multiplied by the sacling factor of 1.17 to get the COSMO-radii. [a] Values estimated from the radii values using the relation $d_{ion\text{-}water} = R_{ion} + R_{water}$ with R_{water}= 1.38 nm given in Marcus (1997). [b] N.A denotes 'Not Available' in the cited literature data.

	Ionic radii R_{ion} From Marcus 1997 and Marcus 2012 [/nm]	$d_{ion\text{-}water}$ From Marcus 2012 and Marcus 1988 [/nm]	$d_{ion\text{-}water}$ From Ohtaki and Radnai 1993 [/nm]	$d_{ion\text{-}water}$ (CN) From Tunell and Lim 2006 [/nm]	$d_{ion\text{-}water}$ ($n_{structure,gas}$) Calculated in this study [/nm]
H_3O^+	1.30	2.755 ± 0.015	2.44	N.A [b]	2.550 (3)
Li^+	0.69	2.080 ± 0.070	1.94-2.28	1.946 (4)	1.991 (4)
Na^+	1.02	2.356 ± 0.060	2.30-2.50	2.269 (4)	2.340 (4)
K^+	1.38	2.798 ± 0.081	2.60-2.95	2.673 (4)	2.768 (4)
Rb^+	1.49	2.89	N. A. [b]	2.870 (4)	2.905 (3)
Cs^+	1.70	3.139 ± 0.076	2.95-3.21	N.A [b]	3.126 (4)
Be^{2+}	0.35	1.73[a]	1.75	1.647 (4)	1.663 (4)
Mg^{2+}	0.72	2.090 ± 0.041	2.00-2.15	2.098 (6)	2.126 (6)
Ca^{2+}	1.00	2.422 ± 0.052	2.39-2.54	2.432-2.513 (7)	2.455 (7)
Sr^{2+}	1.13	2.64	N.A [b]	2.630 (8)	2.652 (8)
Ba^{2+}	1.36	2.74[a]	N.A [b]	2.725-2.911 (9)	2.917 (9)

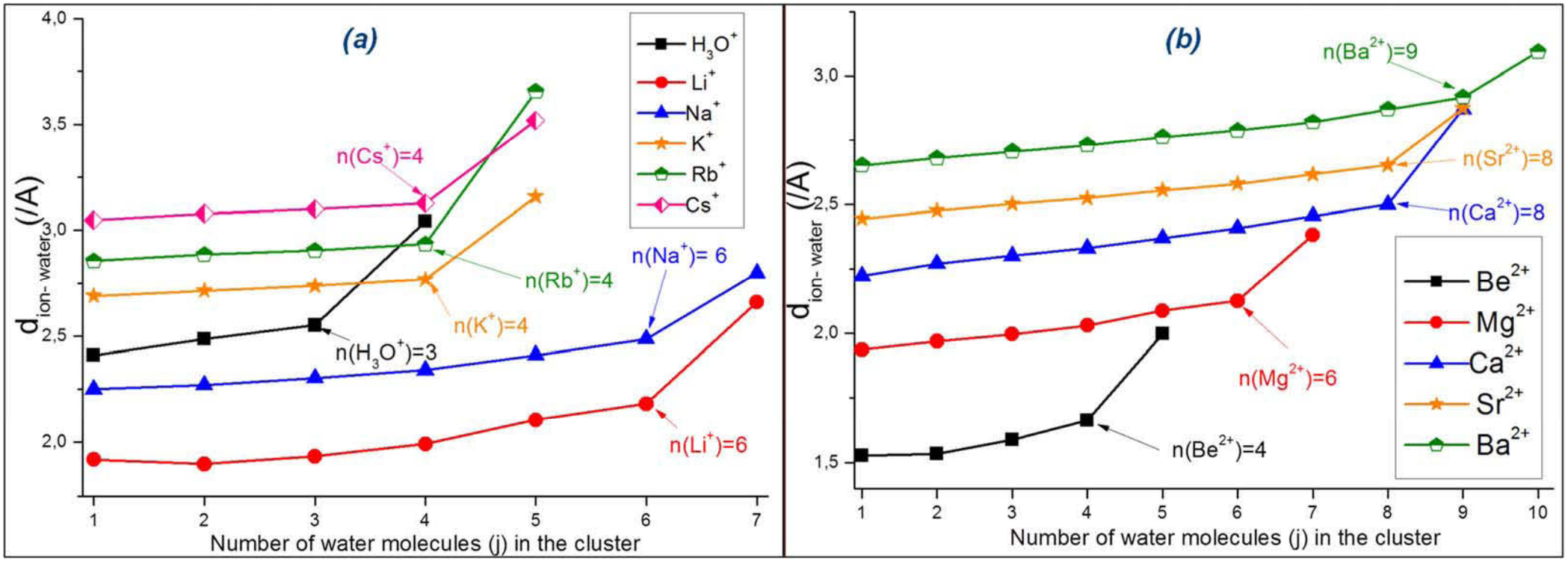


***Figure III-5** : Evolution of the averaged internuclear ion-water distances ($d_{ion\text{-}water}$) of the successive stoichiometric complexes formed by hydrated (a) monovalent cations (H_3O^+, Li^+, Na^+, K^+, Rb^+ and Cs^+) and (b) divalent cations(Be^{2+}, Mg^{2+}, Ca^{2+}, Sr^{2+} and Ba^{2+}) and j water molecules, versus the number (j) of water molecules inside the cluster.*

Here, n denotes the maximum number of waters molecules in the primary hydration shell ($n_{volumic}$). The subscript is omitted for clarity on the figure.

III.3.2.1.2. Gas phase structures analysis results and determination of COSMO-radii

The evolution of the binding energies ($\Delta E_{binding,\ j}$) versus the number j of water molecules inside the j^{th} hydrated cluster is shown in Figure III-6 and Figure III-7. On these figures, the dotted and dashed lines represent respectively the binding energy for adding one extra water molecule to the bulk water in its first shell $\Delta E^{1^{st}\ shell}_{binding,H_2O}$ and in its second hydration shell $\Delta E^{2^{nd}\ shell}_{binding,H_2O}$. Thus when there are no hydrogen bonds in the cluster, if $\Delta E_{binding,j} > \Delta E^{1^{st}\ shell}_{binding,H_2O}$, then $n_{structure,gas} = j\text{-}1$. If there is a hydrogen bond inside the j^{th} cluster (except for H_3O^+) $n_{structure,gas}$ is equal to the total number of water molecules that are directly linked to the ion. In the case of H_3O^+, the same methodology is applied on hydrogen bonds contained in the first hydration shell. Moreover, if there are hydrogen bonds in the j^{th} cluster and if $\Delta E_{binding,j} > \Delta E^{2^{nd}\ shell}_{binding,H_2O}$, then $n_{structure,gas} = j\text{-}1$.

The values of the 'molecular structure-based' hydration number $n_{structure,gas}$ are given in Table III-2 which also contains the hydration numbers values given by Tunell and Lim (2006) who called them coordination number (CN). One can conclude that the 'molecular structure-based' hydration number of the studied monovalent ions are respectively $n_{structure,gas}$ = 3 for H_3O^+, 4 for Li^+, Na^+ and K^+, 3 for Rb^+ and 4 for Cs^+.

Likewise, for divalent ions, the estimated hydration numbers are: $n_{structure,gas}$ = 4 for Be^{2+}, 6 for Mg^{2+}, 7 for Ca^{2+}, 8 for Sr^{2+} and 9 for Ba^{2+}. These results are in good agreement with the literature data (shown in Table III-1).

The first column of Table III-2 contains the ionic radii (R_{ion}) values, published by Marcus (1988, 1997 and 2012). The values of $d_{ion\text{-}water}$ calculated in this study are consistent with the sum $R_{ion}+R_{water}$ given by Marcus where R_{water} = 1.38 nm. This leads to the conclusion that the ionic radii values given by Marcus can be used as reference values for calculating COSMO-radii of cations by applying the scaling factor 1.17 times R_{ion}.

III.3.2.1.3. Thermodynamic reaction properties analysis results

The calculated Gibbs free energies of each of the successive hydration reactions are respectively reported in Table III-3 (for Li^+ and Na^+) and Table III-4 (for all the studied ions). After analyzing the variations of these reactions data (in the gas phase), one can deduce the value of the hydration number ($n_{thermo,gas}$) for each ion of interest.

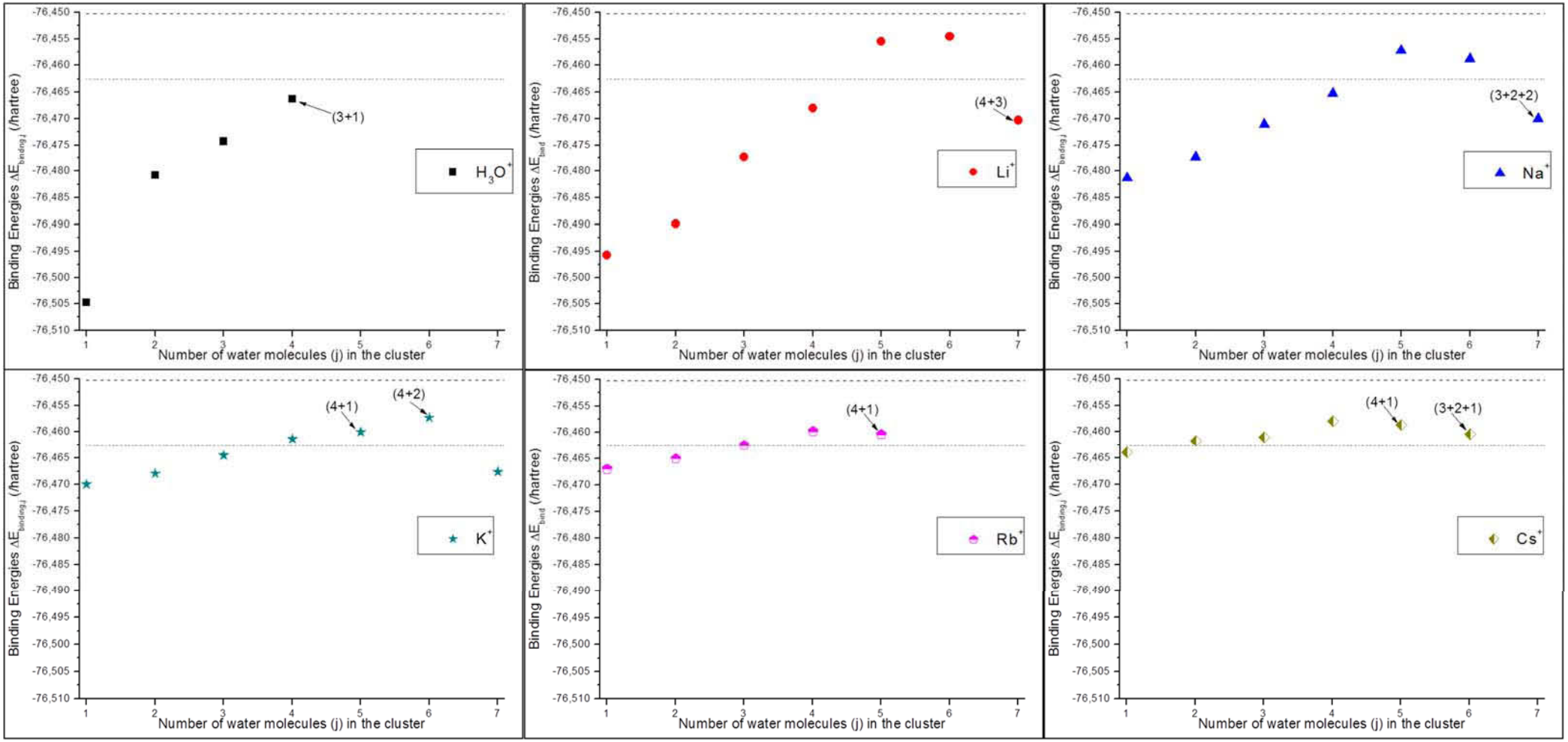


Figure III-6 : Evolution of the gas phase binding energies $\Delta E_{binding;j}$ of monovalent cations (H_3O^+, Li^+, Na^+, K^+, Rb^+ and Cs^+) and j water molecules, versus the number (j) of water molecules inside the cluster.

The dotted and dashed lines represent respectively the binding energy for adding one extra water molecule to the bulk water in its first shell $E^{1stshell}{}_{binding;H2O}$ and in its second hydration shell $E^{2ndshell}{}_{binding;H2O}$.

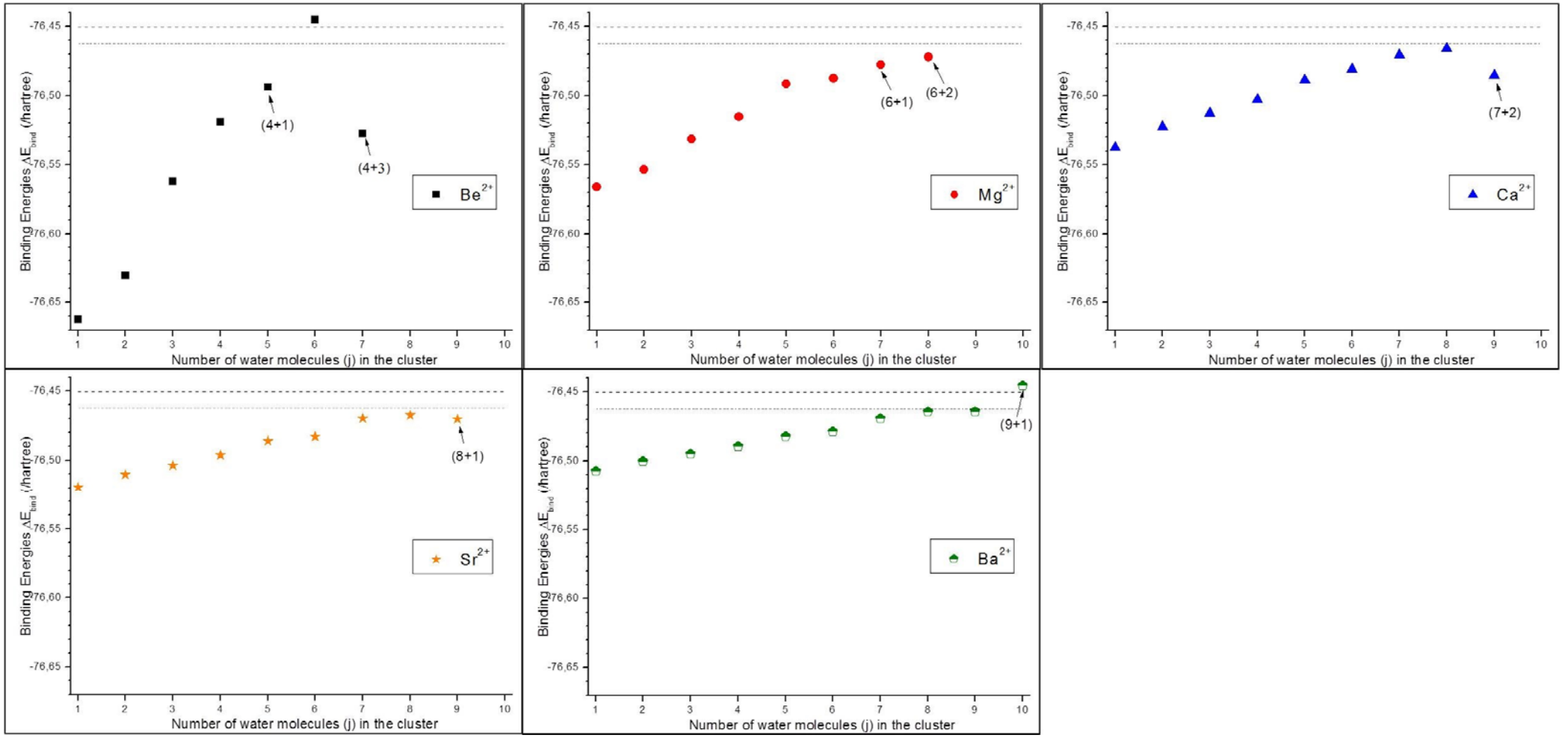


Figure III-7 *: Evolution of the gas phase binding energies $\Delta E_{binding;j}$ of divalent cations (Be^{2+}, Mg^{2+}, Ca^{2+}, Sr^{2+} and Ba^{2+}) and j water molecules, versus the number (j) of water molecules inside the cluster.*

The dotted and dashed lines represent respectively the binding energy for adding one extra water molecule to the bulk water in its first shell $E^{1stshell}{}_{binding;H2O}$ and in its second hydration shell $E^{2ndshell}{}_{binding;H2O}$.

For reaction numbers j = 1 to 4, it is shown in Table III-4 that for the Li^+, Na^+ and K^+ ions, the Gibbs free energy of reaction is negative. When j = 5, the Gibbs free energy of reaction becomes positive that means that the 5th reaction ($\left[Ion,\left(H_2O\right)_4\right]^+ + H_2O \rightleftharpoons \left[Ion,\left(H_2O\right)_5\right]^+$; where *Ion=Li, Na or K*) is oriented in a way that is unfavorable to the formation of the 5th hydrated complex. It can be deduced that the corresponding hydration number is $n_{thermo,gas}$ = 4 for each of the first three alkali metal ions namely Li^+, Na^+ and K^+. Likewise, the hydration numbers of Rb^+ and Cs^+ are respectively $n_{thermo,gas}$ = 3 and 2. Similarly, the hydration numbers of the divalent cations are respectively $n_{thermo,gas}$ = 4 for Be^{2+}, 6 for Mg^{2+}, 7 for Ca^{2+}, 8 for Sr^{2+} and 9 for Ba^{2+}. These values are also in good agreement with literature data (given in Table III-1), as discussed later.

As there are many experimental and theoretical data about gas phase incremental hydration reactions free energies, especially for alkali metal ions (Dzidic and Kebarle 1970, Kebarle 1977, Pratt and Rempe 1999, Rempe et al. 2000, Tunell and Lim 2006), it is interesting to compare these data to those predicted using the COSMO-RS method. The results obtained for Li^+ and Na^+ ions are given in Table III-3, and are compared to the incremental hydration reactions free energies data calculated by Tunell and Lim (2006), and several experimental data taken from Dzidic and Kebarle (1970) and Kebarle (1977). The gas phase hydration numbers predicted here, as well as the incremental hydration reactions free energies are all in good agreement with these published data.

Table III-3: Comparison between gas phase computed and experimental incremental free energies (in kJ.mol^{-1}) for the reaction:

$\left[Ion,\left(H_2O\right)_{j-1}\right]^{q+} + H_2O \rightleftharpoons \left[Ion,\left(H_2O\right)_j\right]^{q+}$ ***(j = 1 to 6).***

Litterature computed data are taken from Tunell and Lim 2006 and experimental data, in parentheses, are given by Dzidic and Kebarle (1970) and Kebarle (1977).

$j \rightarrow j+1$	Calculated data (this work)		Litterature data	
	Li+ (kJ/mol)	Na+ (kJ/mol)	Li^+ (kJ/mol)	Na^+ (kJ/mol)
$0 \rightarrow 1$	-105.55	-68.66	-122.15 (-94.12)	-78.23 (-73.63)
$1 \rightarrow 2$	-83.90	-59.07	-94.54 (-79.06)	-64.42 (-55.22)
$2 \rightarrow 3$	-54.73	-35.29	-58.57 (-55.64)	-41.83 (-38.90)
$3 \rightarrow 4$	-30.31	-23.27	-33.56 (-31.37)	-30.54 (-26.35)
$4 \rightarrow 5$	14.92	11.90	17.15	19.66
$5 \rightarrow 6$	18.23	14.38	18.41	1.0
Gas phase Hydration number $n_{thermo,gas}$	4	4	4	4

***Table III-4*: Gas phase incremental Gibbs free enthalpies of reaction (ΔG_j^{react}) values for the successive hydration reaction of the studied ions where j denotes the reaction number.**

The hydration number $n_{thermo,gas}$ of the studied ion is also given. The parentheses denote the formation of structures containing hydrogen bonds (for instance when j=4+2, the cluster contains 4 inner and 2 outer water molecules). [a] *Since all values of ΔG_j^{react} are negative for H_3O^+, we chose $n_{thermo,gas} = n_{volume}$ for this ion.*

$j \rightarrow j+1$	H_3O^+	Li^+	Na^+	K^+	Rb^+	Cs^+	Be^{2+}	Mg^{2+}	Ca^{2+}	Sr^{2+}	Ba^{2+}
0 → 1	-190.84	-105.55	-68.66	-40.90	-33.58	-26.14	-540.02	-288.47	-214.40	-168.17	-136.50
1 → 2	-130.73	-83.90	-59.07	-31.52	-25.42	-18.14	-448.00	-249.36	-164.23	-140.35	-112.31
2 → 3	-112.01	-54.73	-35.29	-29.17	-13.29	3.55	-264.21	-192.84	-154.28	-128.35	-106.75
3 → 4	-94.49	-30.31	-23.27	-7.74	-16.93	-35.52	-153.79	-145.02	-115.91	-94.76	-84.53
4 → 5		14.92	11.90					-80.17	-74.91	-67.06	-63.74
4 -→ 4+1				(-7.21)		(15.06)	(-86.43)				
4 -→ 3+2					(-9.61)						
5 → 6		18.23	14.38					-65.51	-59.12	-65.84	-48.04
4+1 -→ 4+2				33.56			41.09				
6 → 7									-25.76	-31.53	-29.93
6-→ 6+1								(-51.07)			
6-→ 4+3		(-6.63)	(-24.91)								
7→ 8										-20.02	-9.12
7-→ 7+1									(-14.74)		
8→ 9											-0.58
8-→ 8+1										(-18.92)	
8-→ 7+2									(-62.31)		
9-→ 9+1											63.04
Hydration number ($n_{thermo,gas}$)	3[a]	4	4	4	3	2 or 4	4	6	7	8	9

III.3.2.2. Aqueous phase analysis results

The calculated binding pseudo-chemical potentials $\Delta\mu^{*}_{binding,j}$ of each successive hydration reaction for each studied ion are reported in Table III-5. After analyzing the variations of these reactions data (at infinite dilution in water), one can deduce the value of the hydration number ($n_{thermo,\infty}$) for each ion of interest. For reactions numbers *j=1* to 4, it is shown in Table III-5 that for the Li^+, Na^+ and K^+ ions, the Gibbs free energy of reaction is negative ($\Delta\mu^{*}_{binding,j}$<0). When *j=5*, $\Delta\mu^{*}_{binding,j}$ becomes positive that means the 5th reaction ($\left[Ion,\left(H_2O\right)_4\right]^+ + H_2O \rightleftharpoons \left[Ion,\left(H_2O\right)_5\right]^+$; where *Ion=Li, Na or K*) is oriented in a way that is unfavourable to the formation of the 5th hydrated complex. It can be deduced that the corresponding hydration number is $n_{thermo,\infty}$ = 4 for each of the first three alkali metal ions namely Li^+, Na^+ and K^+. Likewise, the hydration numbers of Rb^+ and Cs^+ are respectively $n_{thermo,\infty}$ = 3 and 4. For H_3O^+, the results clearly show that the most stable form of proton at infinite dilution in water is H_3O^+ since all the $\Delta\mu^{*}_{binding,j}$ values are positive and greater than thermal energy (about 2.478 kJ.mol^{-1}).

Similarly, the hydration numbers of the divalent cations are respectively $n_{thermo,\infty}$ = 4 for Be^{2+}, between 4 and 6 for Mg^{2+}, 6 or 7 for Ca^{2+}, between 6 and 8 for Sr^{2+}, and between 7 and 9 for Ba^{2+}. These values are also in good agreement with literature data (given in Table III-1), as discussed later.

***Table III-5**: Hydration equilibrium pseudo-chemical potentials of binding ($\Delta\mu^{*}_{binding,j}$) values for the successive hydration reaction of the studied ions where j denotes the reaction number.*

The hydration number $n_{thermo,\infty}$ of the studied ion is also given in the last line. The parentheses denote the formation of structures containing hydrogen bonds.

$j \rightarrow j+1$	H_3O^+	Li^+	Na^+	K^+	Rb^+	Cs^+	Be^{2+}	Mg^{2+}	Ca^{2+}	Sr^{2+}	Ba^{2+}
0 → 1	9.55		-65.68	-24.63	-19.40	-14.85				-150.82	-85.55
1 → 2	8.92	-175.41	-69.20	-24.36	-18.97	-13.96		-799.32		-148.56	-82.25
2 → 3	3.37	-25.86	-30.09	-19.10	-17.24	-14.43	-94.36	-139.86	-242.68	-193.92	-89.91
3 → 4	4.51	-2.38	-10.46	-15.72	-12.02	-10.85	-17.57	-77.55	-136.14	-120.03	-91.11
4 → 5		6.64	4.95					3.33	-8.39	-27.60	-23.88
4 -·→ 4+1				-3.52		-5.56	20.95				
4 -·→ 3+2					3.45						
5 → 6		4.69	4.16					-1.80	-20.66	-17.99	-36.12
4+1 -·→ 4+2				5.98			17.92				
6 → 7									10.13	3.92	5.94
6 -·→ 6+1								7.07			
6 -·→ 4+3		1.37	-5.99								
7 → 8										4.45	2.08
7 -·→ 7+1									8.53		
8 → 9											7.01
8 -·→ 8+1										3.37	
8 -·→ 7+2									-0.82		
9 -·→ 9+1											-23.38
Hydration number ($n_{thermo,\infty}$)	0	4	4	4	3	4	4	4 or 6	6 or 7	6, 7 or 8	8 or 9

III.3.3. Discussion

Ab initio methods suggest that Li^+, Na^+ and K^+ have strong coordination with exactly four water molecules (Varma and Rempe 2006). This is confirmed by *ab initio* molecular dynamics (AIMD) simulations that invariably identify additional 'loosely' coordinated water molecules (Varma and Rempe 2006), that lead to the most recent experimentally determined hydration numbers of Na^+ to be 5 and K^+ to be 6 (Varma and Rempe 2006). Likewise, the present study shows that 4 water molecules are strongly coordinated to each of these three alkali ions.

Experimental and simulation data (Ohtaki and Radnai 1993, Bockris and Reddy 1970, Driesner et al. 2000, Varma and Rempe 2006) on the hydration number of Li^+ in aqueous solutions vary from 4 to 6. Ohtaki and Radnai (1993) pointed out that the most frequently appearing coordination number of ions probably depends on the concentration of the ions, meanwhile its value for Li^+ becomes $n = 4$ or 5 in a system with high concentration and $n = 6$ in a system with lower concentration. The latter is also reported by Kiriukhin and Collins (2002) and is in good agreement with a more recent value reported by Israelachvili (2011) ($n = 5\pm1$ for Li^+). The calculated value $n_{volumic} = 6$ from steric constraints, is consistent with these literature data. However, since the gas phase binding energies for the formation of $[Li, (H_2O)_5]^+$ or $[Li, (H_2O)_6]^+$ requires more energy than the binding of one water molecule to another water molecule of the first hydration shell (*i.e.* $\Delta E_{binding,j} > \Delta E^{1^{st}\,shell}_{binding,H_2O}$ for j= 5 or 6), the gas phase hydration number of Li^+ is $n_{structure,gas} = 4$. Finally, the combination of all the results obtained within this study suggests that the most probable hydration number of Li^+, both in gas phase and at infinite dilution in water, is $n = 4$.

In a molecular dynamics simulation study of ionic hydration, Driesner et al. (2000) reported that the hydration number (at infinite dilution) of the Na^+ ion remains essentially constant around 5.5. Other authors (Kiriukhin and Collins 2002, Driesner et al. 2000, Varma and Rempe 2006) reported that the hydration number of Na^+ is somewhere between 4 and 6. These values are a bit less than $n_{volumic} = 6$. However, since the gas phase binding energies for the formation of $[Na,(H_2O)_5]^+$ or $[Na, (H_2O)_6]^+$ requires more energy than the binding of one water molecule to another water molecule of the first hydration shell (*i.e.* $\Delta E_{binding,j} > \Delta E^{1^{st}\,shell}_{binding,H_2O}$ for j= 5 or 6), the gas phase hydration number of Na^+ $n_{structure,gas}$ is between 4 and 5. Moreover the combination of all the results obtained within this study suggests that the most probable hydration number of Na^+, both in gas phase and at infinite dilution is $n = 4$.

Ohtaki and Radnai (1993) pointed out that the potassium ion K^+ is one of the most difficult ions to determine the hydration structure by the X-ray and neutron diffraction methods. For this ion, experimental hydration numbers vary from 4 to 8 (Bockris and Reddy 1970, Ohtaki and Radnai 1993, Kiriukhin and Collins 2002, Driesner et al. 2000, Varma and Rempe 2006) and simulations suggest that the hydration number of K^+ is somewhere between 4 and 7.8 (Driesner et al. 2000, Varma and Rempe 2006, Whitfield et al. 2007, Tunell and Lim 2006). First shell and second shell interactions may be more important around the K^+ ion than around Na^+ ion because the residence time of water molecules in the hydration shell of the larger K^+ ion is smaller, that should explain the experimental difficulties reported by Ohtaki and Radnai (1993). Another possible explanation (Marcus 1988) of these experimental difficulties is the fact that the mean distance ion-water is equal to the mean distance water-water, as shown in Table III-2. Nevertheless, all the independent procedures developed in this study converge to the conclusion that the hydration number of K^+ is $n = 4$ both in gas phase and at infinite dilution in water.

For Rb^+, Tunell and Lim (2006) have predicted a gas phase hydration number of 4 or 6 (rather than the values 4 and 5 found in their previous DFT calculations). They stated that Rb^+ does not exhibit a strong preference for a specific coordination number. In the present study, these statements are confirmed by the predicted hydration number of 3 or 4, both in the gas phase and at infinite dilution in water. This result is also in a very good agreement with the value of 3.5 determined by NMR measurement in aqueous solution.

There are few studies treating the hydration number of Cs^+: Israelachvili (2011) has reported that $n = 1\pm1$, Ohtaki and Radnai (1993) reported $n = 5.3$ - 8.2. In this study, the maximum number of water molecules inside the primary hydration shell of Cs^+ is $n_{volumic}= 4$ in the gas phase. However, this value changes to reach $n_{thermo,\ gas}= 2$ or $n_{structure,\ gas} = 4$ if the respective analysis of the gas phase incremental Gibbs free energies or the gas phase structure are performed. The comparison of the gas phase binding energies to that between one water molecule and another water molecule of the first hydration shell, leads to the conclusion that $n = 2$ in good agreement with the value reported by Israelachvili (2011). However, in aqueous solution, the predicted value of the hydration number of Cs^+ becomes similar to those of Li^+, Na^+ or K^+, *i.e.* $n_{thermo,\infty} = 4$.

For H_3O^+, the maximum number of water molecules inside the primary hydration shell is $n_{volumic}= 3$ in this study, in agreement with the results reported by Israelachvili (2011) for this ion. However, in solution the most stable form of the proton is H_3O^+ $n_{thermo,\infty}(H_3O^+) = 0$ *i.e.* $n_{thermo,\infty}(H^+) = 1$. This should explain why the use of H_3O^+ as reactant instead of H^+ in the

proton exchange reaction, leads to accurate pK_a values (Toure et al. 2013, Eckert and Klamt 2006).

Asthagiri and Pratt (2003) used both the quasi-chemical theory and the *ab initio* molecular dynamics (AIMD) simulations for stating that Be^{2+} is tetra-hydrated in aqueous solution. They also stated that using a Car-Parrinello approach, the hexa-hydrated cluster is quickly dissociated to give the tetra-hydrated one. All these findings are confirmed within the present study, where we got $n = 4$, also found by Tunell and Lim (2006), both in the gas phase and at infinite dilution in water. Moreover, it is clearly shown in Figure III-7 that the cluster $[Be,(H_2O)_6]^{2+}$ having six water molecules has a binding energy greater than that between one water molecule and another water molecule of the second hydration shell (*i.e.* $\Delta E_{binding,j} > \Delta E^{2^{nd}\ shell}_{binding,H_2O}$ for $j = 6$). This also confirms that the hexa-hydrated form of Be^{2+} has few chances to stay stable in solutions.

For Mg^{2+}, the hydration number predicted in this study *i.e.* $n = 6$ seems extremely well established in the literature data (Israelachvili 2011, Ohtaki and Radnai 1993, Kiriukhin and Collins 2002, Driesner et al. 2000, Hancock et al. 2004, Tunell and Lim 2006).

Tunell and Lim (2006) stated that Ca^{2+} does not exhibit a strong preference for a specific coordination number. The hydration number values $n = 6$, 7 or 8 determined in this study were also reported by Ohtaki and Radnai (1993) according to most X-ray diffraction data, and by several other authors (Bockris and Reddy 1970, Hancock et al. 2004). Moreover, Tunell and Lim (2006) stated that the maximum number of water molecules in the first hydration shell of Ca^{2+} is 8, which is exactly the value of its $n_{volumic}$. They also suggested that this number should decrease when salt concentration increases in aqueous solutions, in agreement with the values of $n_{thermo,\infty} = 6$ or 7.

For Sr^{2+}, Tunell and Lim (2006) found that the maximum number of water molecules inside the primary hydration shell is $n = 8$, that is exactly the value of $n_{volumic}$ (also determined by X-ray diffraction methods in aqueous solution, and single high-resolution crystal structure of a Sr^{2+} aqua-complex) (Tunell and Lim 2006). At infinite dilution in water, this value could decrease as low as $n_{thermo,\infty} = 6$ or 7.

For Ba^{2+}, the maximum number of water molecules inside the primary hydration shell is $n_{volumic} = n_{thermo,\ gas} = 9$, in agreement with the findings of Tunell and Lim (2006) from binding energies analysis. At infinite dilution in water, this ion does not exhibit a preferred hydration number value since we found $n_{thermo,\infty} = 7$, 8 or 9.

To summarize, the predicted results are in good agreement with the latest literature data especially those collected by Israelachvili (2011) particularly for Be^{2+} and Mg^{2+} which

are the only ions in Table III-1 for which the stoichiometric hydration number is clearly defined and given. One should retain that each hydration number determined within this study is what Israelachvili (2011) called the number of water molecules forming a stoichiometric complex with the studied ion.

III.3.4. Concluding remarks abou t the determination of hydrated structures of cations

In this study, the hydration number is defined as the number of water molecules forming a stoichiometric complex with the studied ion (covalent bonds except for H_3O^+). Four complementary procedures, one based on the analysis of the mean ion-water distances, one based on the gas phase structures analysis, one based on the study of the gas phase thermodynamic properties of the successive hydration reactions and another based on the analysis of the pseudo-chemical potentials of binding (at infinite dilution in water) of each hydrated complex versus the number of water molecules inside the cluster, are described and used to predict the corresponding hydration number values $n_{volumic}$, $n_{thermo,\ gas}$, $n_{structure,\ gas}$ and $n_{thermo,\infty}$. The discrepancies between the gas phase hydration numbers obtained in different studies are explained by subtle differences between the definitions of the hydration number. In aqueous solutions, the hydration numbers are more precisely defined even if a salt concentration dependence is reported by several authors. This study concerns infinite dilution in water.

Importantly, the values of $d_{ion\text{-}water}$ calculated in this study are consistent with the sum $R_{ion}+R_{water}$ given by Marcus (1997 and 2012) where $R_{water} = 1.38$ nm. This justifies the choice of the ionic radii values given by Marcus (1997 and 2012) as reference values for calculating COSMO-radii of cations and applying the scaling factor of 1.17 times R_{ion}. Since hydration number is more of a qualitative indication of the degree to which ions bind water rather than an exact value, the present study shows that the DFT/QM level used in COSMO-RS is enough accurate to determine the hydration numbers of ions. Moreover, compared to the other experimental and theoretical calculation methods used in the literature, this method requires a much smaller computational time and data analysis efforts (several minutes per cluster molecule for the quantum calculations, then it takes less than 1 minute to run thermodynamics calculation in COSMOthermX).

This study shows that four water molecules strongly bind each of the first three alkali metal ions namely Li^+, Na^+ and K^+. Likewise, the hydration numbers of Rb^+ and Cs^+ are $n = 3$ and 4 respectively. The most stable form of proton in aqueous solution is H_3O^+, *i.e.* $n = 0$ for

H_3O^+ at infinite dilution in water (confirming the use of this species in pK_a prediction algorithms). Similarly, the hydration numbers of Be^{2+}, Mg^{2+}, Ca^{2+}, Sr^{2+} and Ba^{2+} ions are respectively n = 4, 4 or 6, 6 or 7, between 6 and 8, and between 7 and 9. These results are in good agreement with the latest published data.

This study provides insights for describing the physico-chemical phenomena that occur during hydration processes of ions as well as possible, in order to take them in account in the development of a predictive thermodynamics model adapted to aqueous complex systems containing electrolytes (*i.e.* charged species).

However the latter study was focused only on the determination of the maximum value (n) of n_{h_i}. In order to take into account the concentration dependence of the hydration number, it is necessary to use a predictive model like COSMO-RS-PDHS to predict activity coefficients data that can be measured experimentally. It is worth mentioning that the knowledge of n_{h_i} is also required to use such model.

Finally, each of the hydrated cluster $[Ion, (H_2O)_j]^{q+}$ where j= 0, 1,2,..,n water molecules (inside the cluster) are successively used in the COSMO-RS-PDHS program (an in-house Matlab program developed during this thesis) to predict activity coefficients of a binary mixture water-salt that is in reality a ternary mixture water (i = 1), one cation (i = 2) and one anion (i = 3). For a given ion (*i.e.* if i = 2 or 3), it is expected that only the right value of n_{h_i} will lead to realistic predicted activity coefficients data, as discussed in section III-4.

III.4. Application of the COSMO-RS-PDHS model to the prediction of mean salt activity coefficient and water activity of several binary aqueous-electrolyte mixtures

In this section, it is assumed that we have a fixed hydration of ions (i.e. n_{h_i} is not salt dependent, and its value has to be determined only once for each ion).

III.4.1. Results obtained by usin g COSMO-radii of ionic elements suggested in th e current state-of-art of the COSMO-RS method

The COSMO-radii of several atomic elements are given in Table III-6.

Table III-6: Default values of the COSMO-radii of several atomic elements used in the current implementation of COSMO-RS model during COSMO calculations in the Turbomole software.

[a]For non-metal ions, the COSMO-radii are approximatively equal to 1.17 times the radii proposed by Mantina et al. (2009) who also used data from Bondi (1964).

[b] for metal ions the COSMO-radii are approximatively equal to 1.17 times the radii proposed by Sutton (1965).

[c] Unparameterized elements, for which by default R_{COSMO}=2.223 A in Turbomole.

Non-metallic elements		Metallic elements	
Elements	COSMO-radii R_{COSMO} [in A] 1A=10^{-10}m	Elements	COSMO-radii R_{COSMO} [in A] 1A=10^{-10}m
H	1.300[a]	Li	1.570[b]
O	1.720[a]	Na	1.800[b]
Cl	2.050[a]	K	2.223[c]
Br	2.160[a]	Mg	1.638[b]
I	1.720[a]	Ca	2.223[c]
F	1.300[a]	Rb	2.223[c]

It is interesting to illustrate the need of a solvation model in the COSMO-RS-PDHS modelling framework in the case of the binary water-NaCl used as illustrative example. In this context, Figures III-8 and III-9 show the results obtained without hydrating the ions, and Figures III-10 and III-11 show the results obtained by hydrating Na^+ with two water molecules and Cl^- with one water molecule. Figures III-12 and III-13 show the results obtained by hydrating Na^+ with three water molecules and Cl^- with one water molecule. The results obtained on other binary mixtures (water-salt) are also given in Figures III-14 to III-29. All these results are in good agreement with literature data, and suggest that n_{h_2} = 1 for H^+, 3 for Na^+, 4 for Li^+, n_{h_3} =1 for Cl^-, Br^- and I^-.

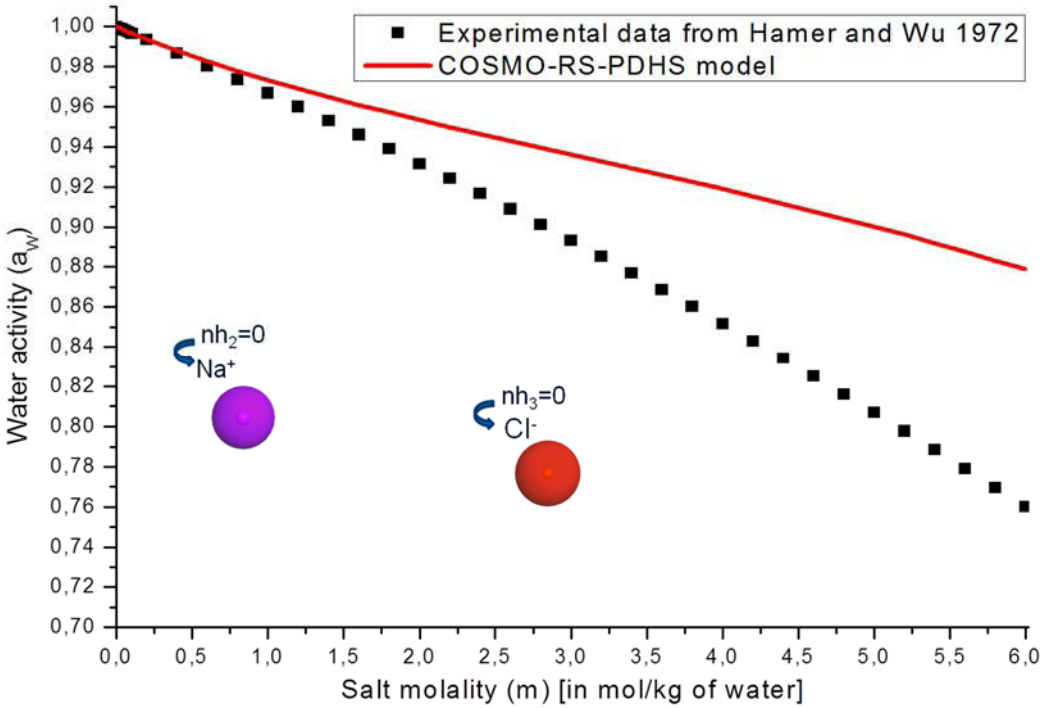


Figure III-8 : Water activity results predicted by the COSMO-RS-PDHS model on the binary water -NaCl, without hydrating the ions (i.e. $n_{h_2} = n_{h_3} = 0$).

The COSMO-radii of atomic elements (Na, Cl, H and O) are given in Table III-6.

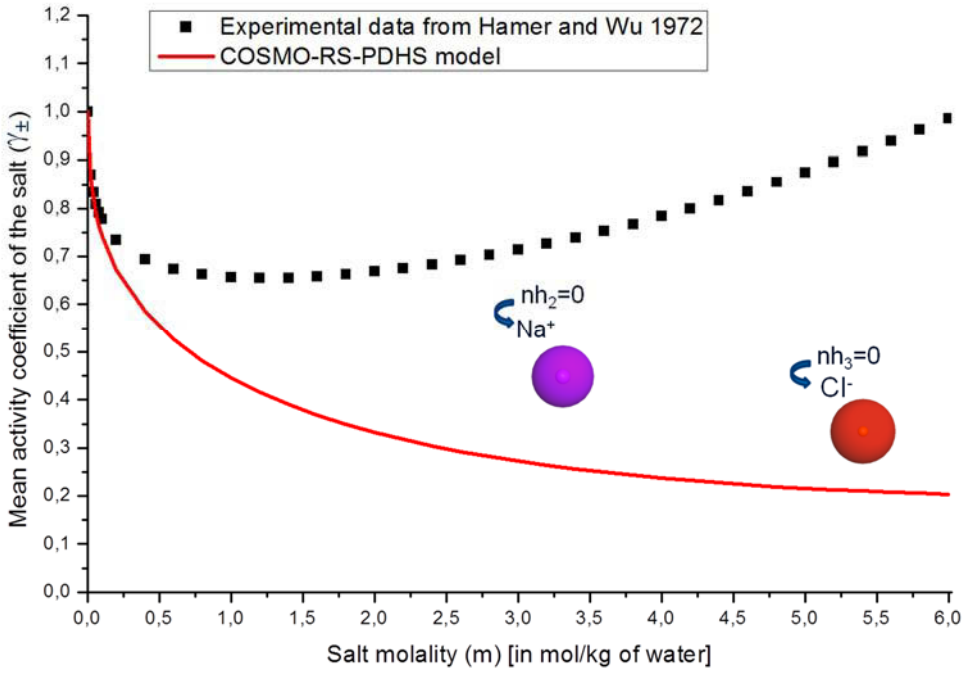


Figure III-9 : $\gamma_\pm^{(m)}$ results predicted by the COSMO-RS-PDHS model on the binary water -NaCl, without hydrating the ions (i.e. $n_{h_2} = n_{h_3} = 0$).

$\gamma_\pm^{(m)}$ denotes the mean activity coefficient of the salt is expressed in molality scale; the COSMO-radii of atomic elements are given in Table III-6.

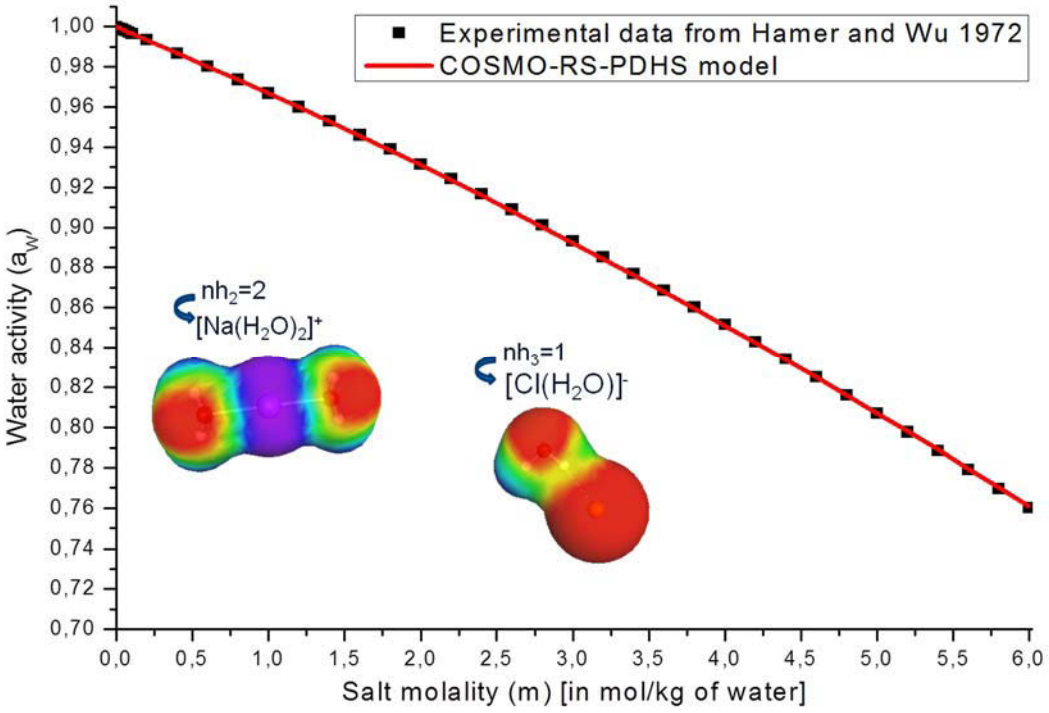


Figure III-10 : Water activity results predicted by the COSMO-RS-PDHS model on the binary water -NaCl, after hydrating the ions (n_{h_2}=2 for Na^+ and n_{h_3}=1 for Cl^-). *The COSMO-radii of atomic elements (Na, Cl, H and O) are given in Table III-6.*

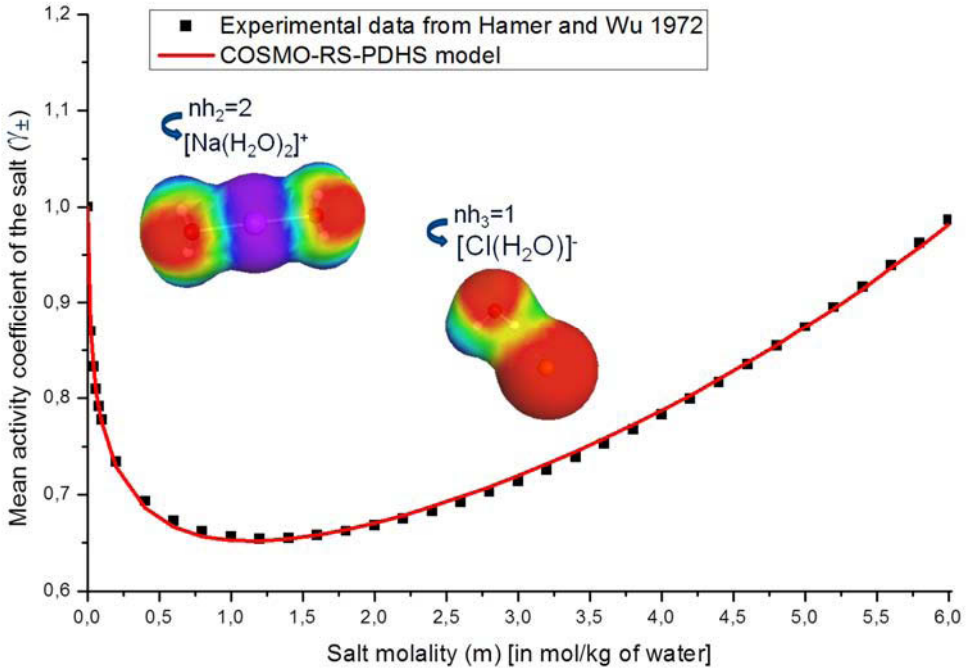


Figure III-11 : $\gamma_\pm^{(m)}$ results predicted by the COSMO-RS-PDHS model on the binary water -NaCl, after hydrating the ions (n_{h_2}=2 for Na^+ and n_{h_3}=1 for Cl^-).

$\gamma_\pm^{(m)}$ denotes the mean activity coefficient of the salt is expressed in molality scale; the COSMO-radii of atomic elements are given in Table III-6.

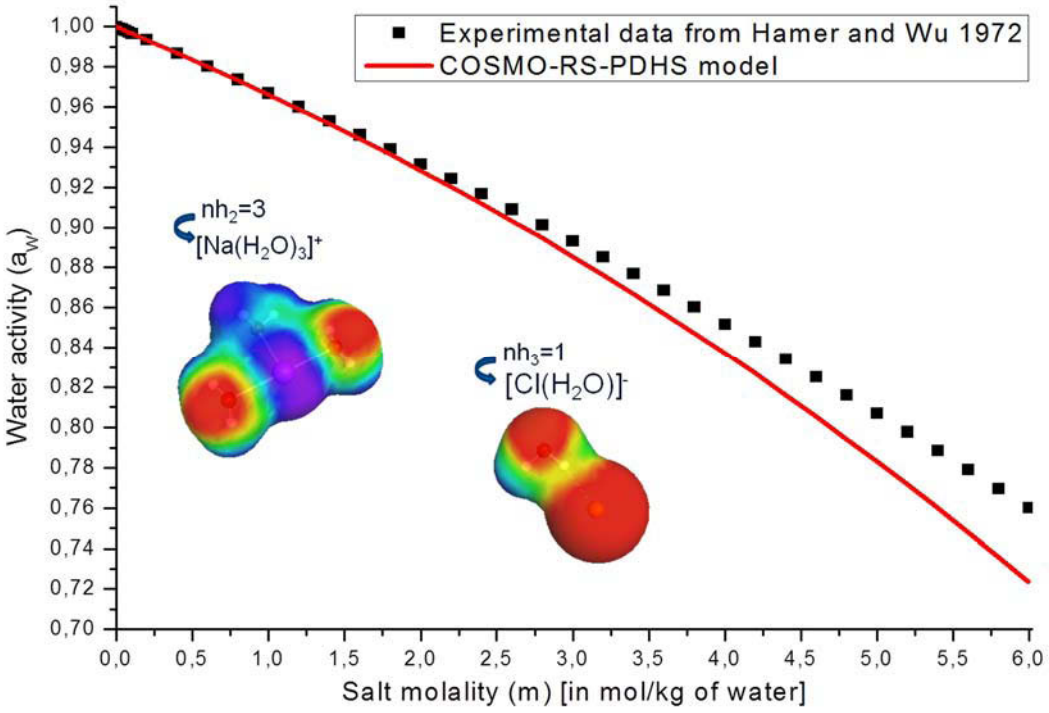


Figure III-12 : Water activity results predicted by the COSMO-RS-PDHS model on the binary water -NaCl, after hydrating the ions (n_{h_2}=3 for Na^+ and n_{h_3}=1 for Cl^-). *The COSMO-radii of atomic elements (Na, Cl, H and O) are given in Table III-6.*

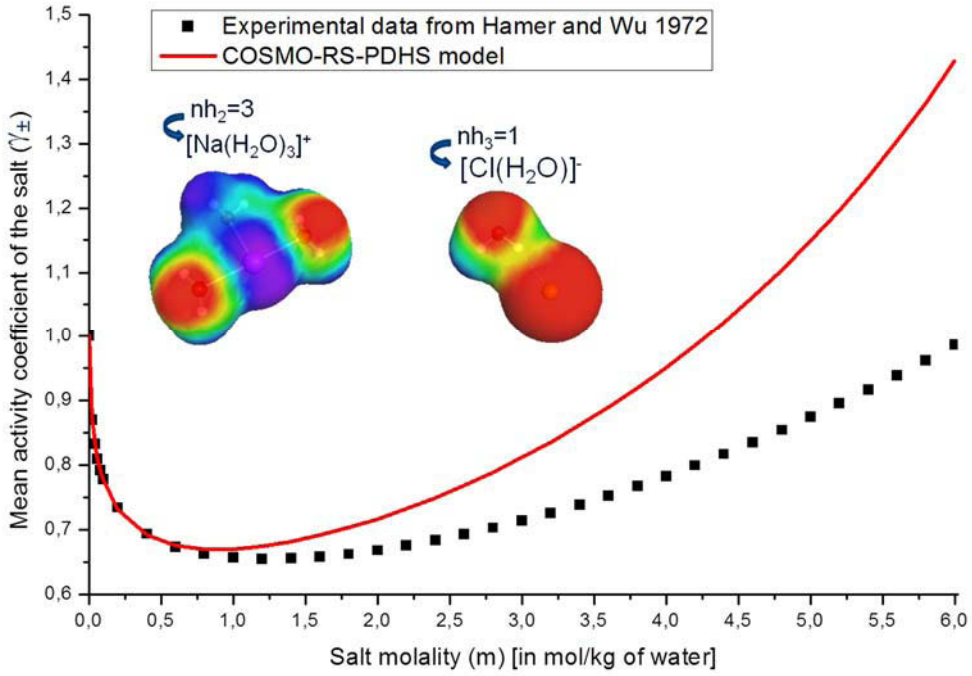


Figure III-13 : $\gamma_{\pm}^{(m)}$ results predicted by the COSMO-RS-PDHS model on the binary water -NaCl, after hydrating the ions (n_{h_2}=3 for Na^+ and n_{h_3}=1 for Cl^-).

$\gamma_{\pm}^{(m)}$ denotes the mean activity coefficient of the salt is expressed in molality scale; the COSMO-radii of atomic elements are given in Table III-6.

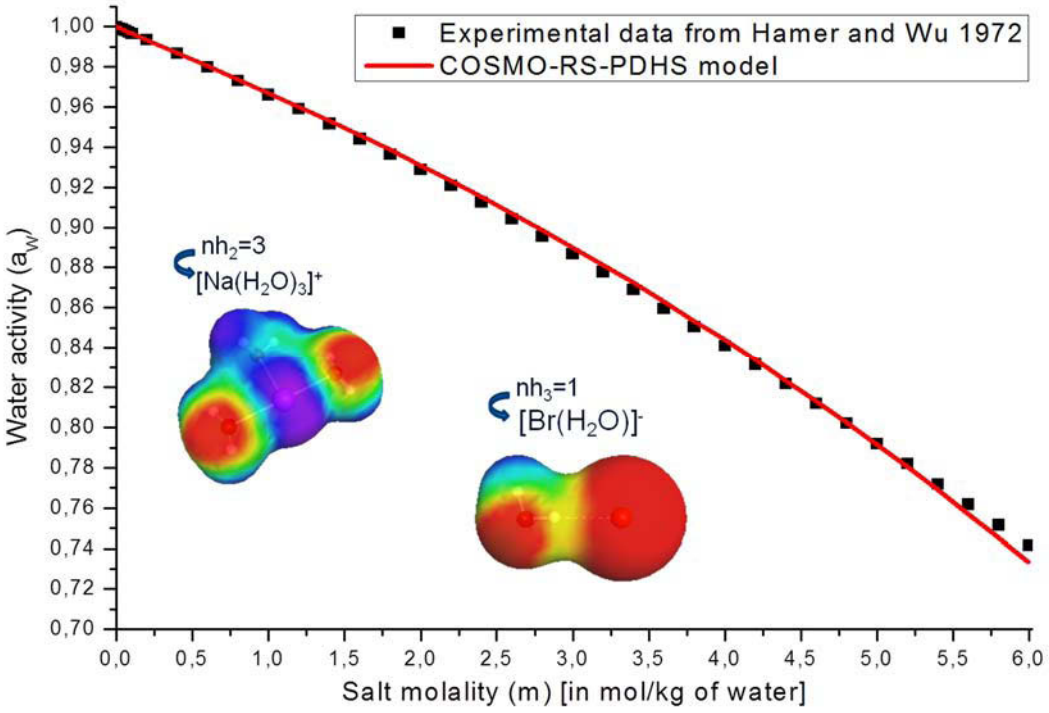


Figure III-14 : Water activity results predicted by the COSMO-RS-PDHS model on the binary water -NaBr, after hydrating the ions (n_{h_2}=3 for Na^+ and n_{h_3}=1 for Br^-).

The COSMO-radii of atomic elements (Na, Br, H and O) are given in Table III-6.

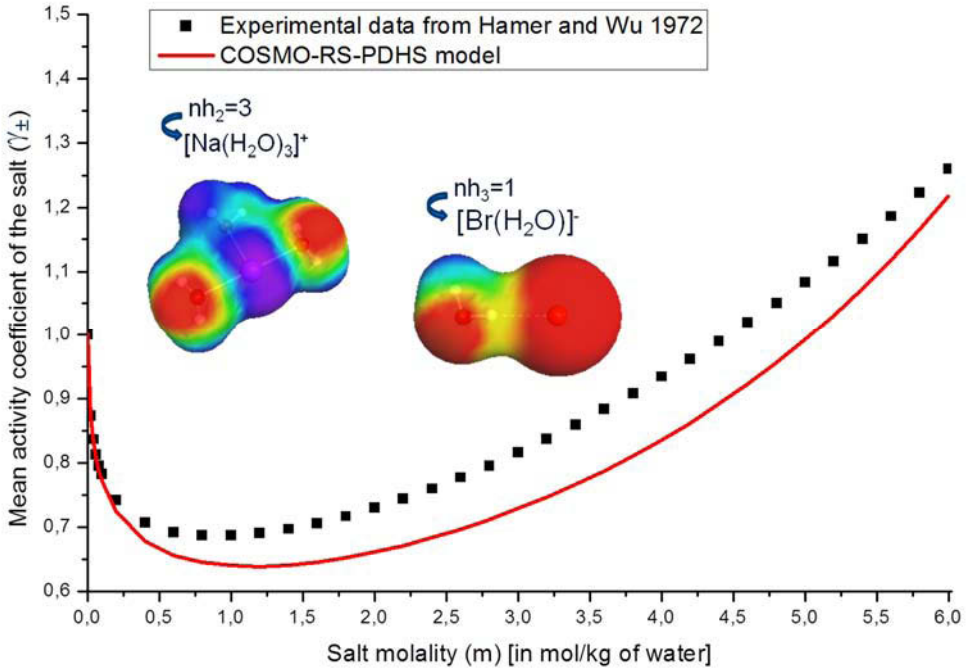


Figure III-15 : $\gamma_{\pm}^{(m)}$ results predicted by the COSMO-RS-PDHS model on the binary water -NaBr, after hydrating the ions (n_{h_2}=3 for Na^+ and n_{h_3}=1 for Br^-).

$\gamma_{\pm}^{(m)}$ denotes the mean activity coefficient of the salt is expressed in molality scale; the COSMO-radii of atomic elements are given in Table III-6.

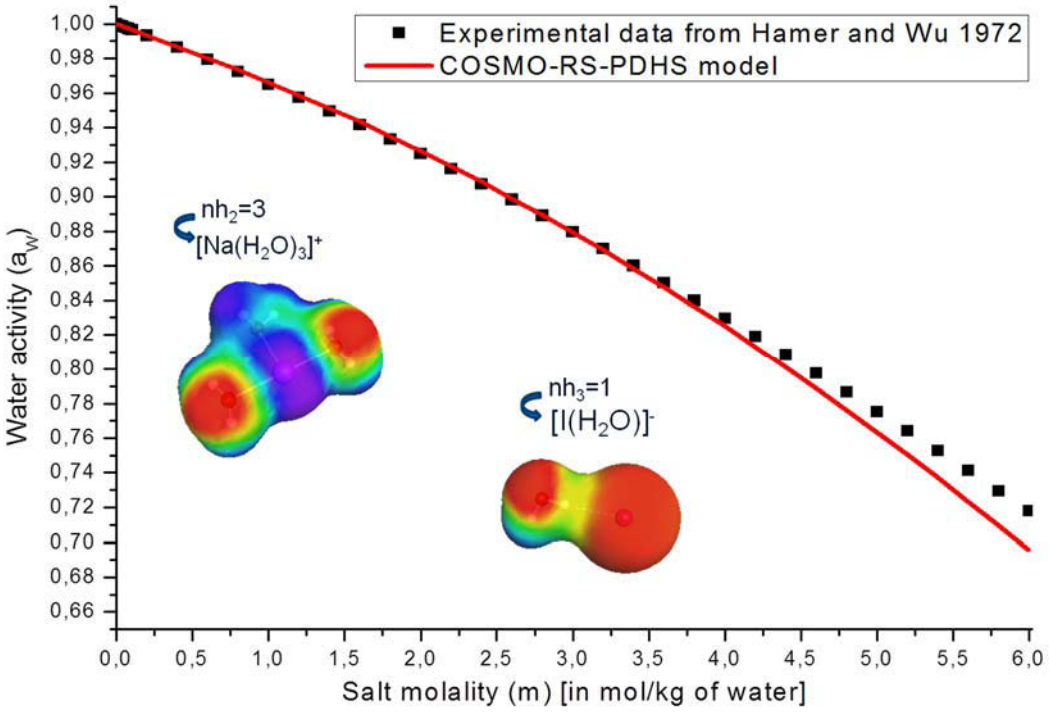


Figure III-16 : Water activity results predicted by the COSMO-RS-PDHS model on the binary water -NaI, after hydrating the ions (n_{h_2} =3 for Na^+ and n_{h_3} =1 for I^-).

The COSMO-radii of atomic elements (Na, I, H and O) are given in Table III-6.

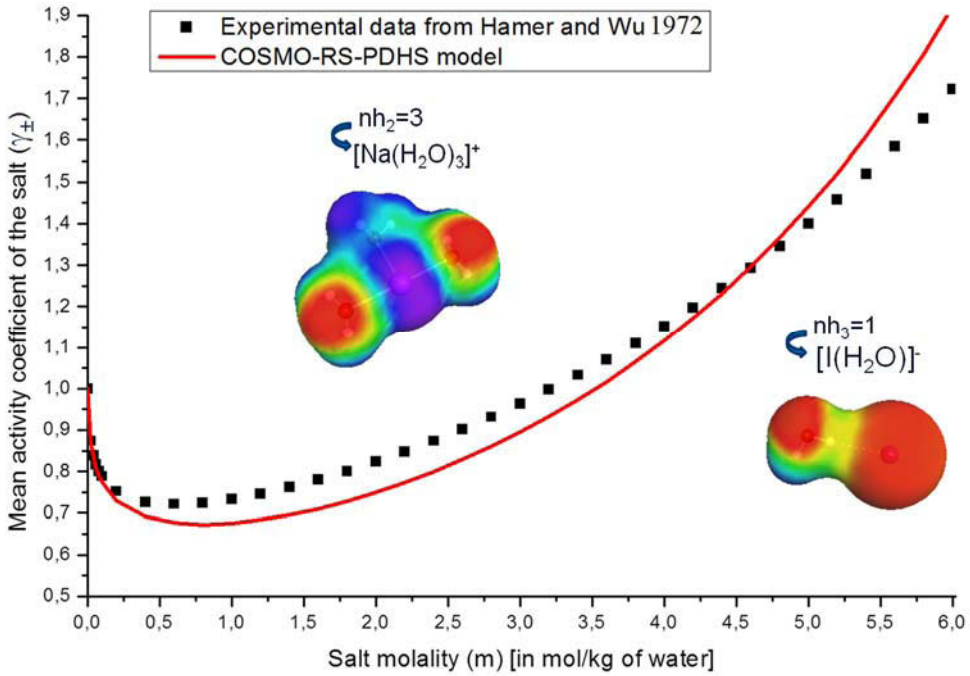


Figure III-17 : $\gamma_{\pm}^{(m)}$ results predicted by the COSMO-RS-PDHS model on the binary water-NaI, after hydrating the ions (n_{h_2} =3 for Na^+ and n_{h_3} =1 for I^-).

$\gamma_{\pm}^{(m)}$ denotes the mean activity coefficient of the salt is expressed in molality scale; the COSMO-radii of atomic elements are given in Table III-6.

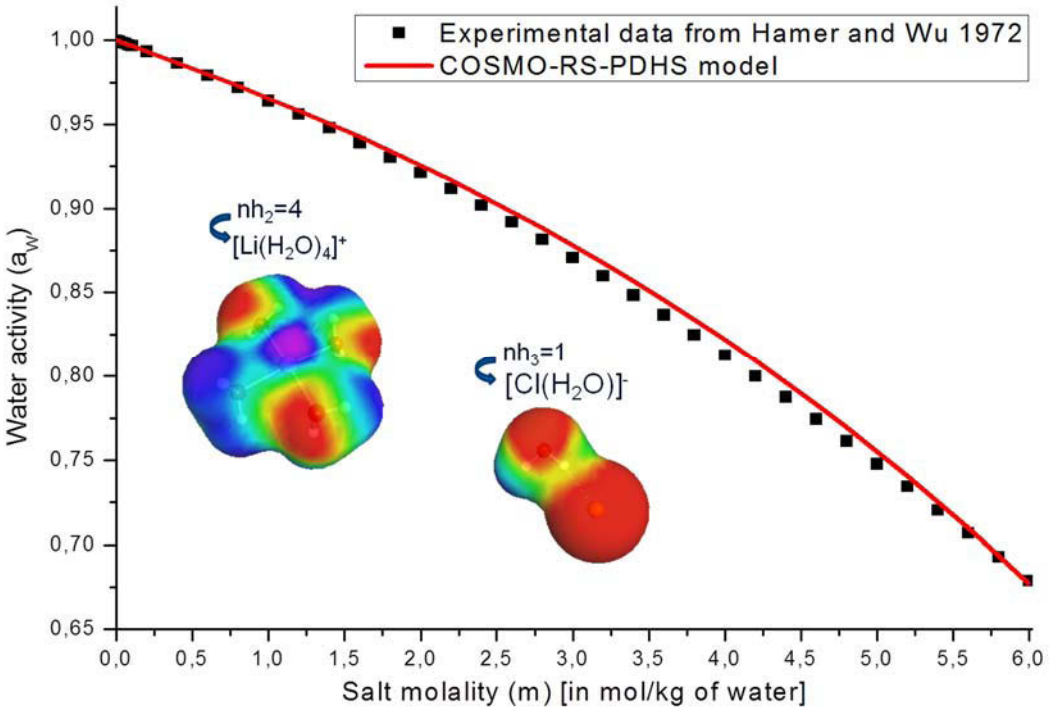


Figure III-18 : Water activity results predicted by the COSMO-RS-PDHS model on the binary water -LiCl, after hydrating the ions (n_{h_2} =4 for Li^+ and n_{h_3} =1 for Cl^-).

The COSMO-radii of atomic elements (Li, Cl, H and O) are given in Table III-6.

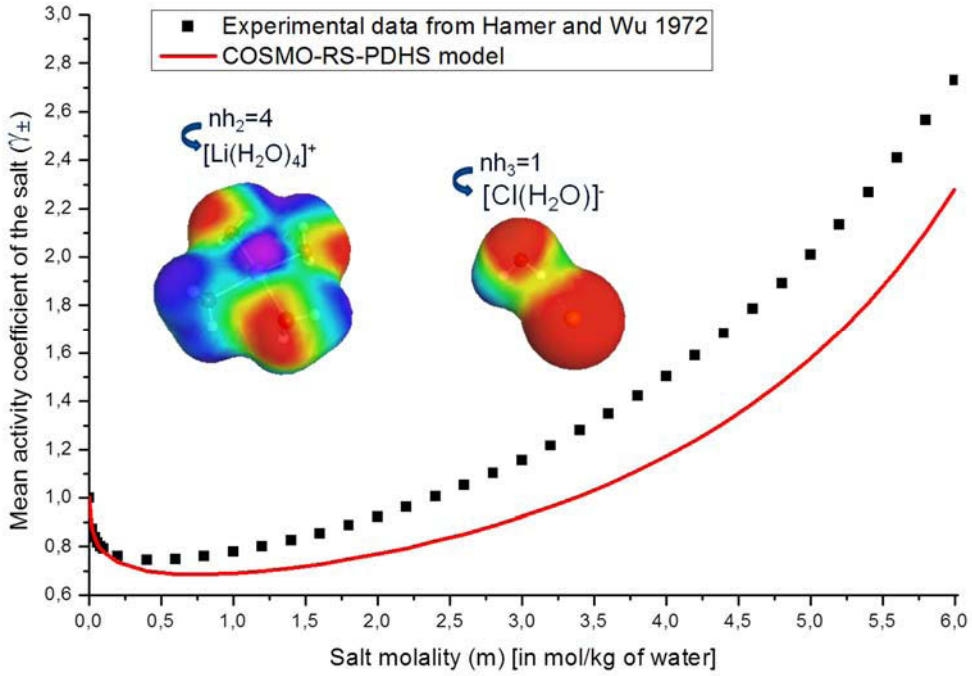


Figure III-19 : $\gamma_{\pm}^{(m)}$ results predicted by the COSMO-RS-PDHS model on the binary water- LiCl, after hydrating the ions (n_{h_2} =4 for Li^+ and n_{h_3} =1 for Cl^-).

$\gamma_{\pm}^{(m)}$ denotes the mean activity coefficient of the salt is expressed in molality scale; the COSMO-radii of atomic elements are given in Table III-6.

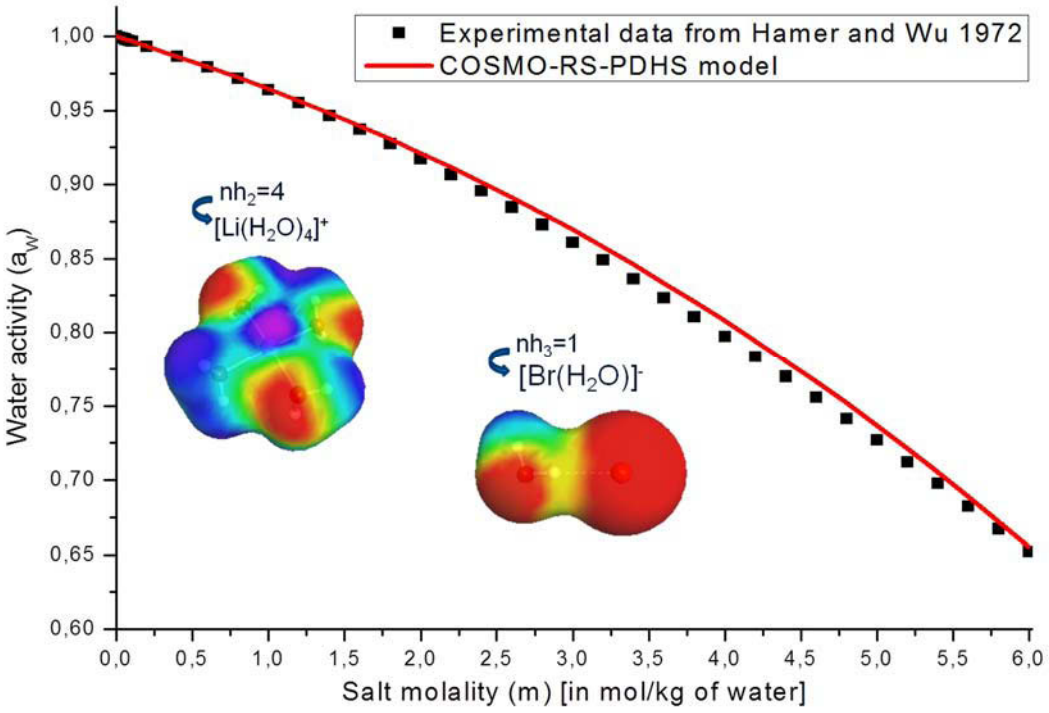


Figure III-20 : Water activity results predicted by the COSMO-RS-PDHS model on the binary water -LiBr, after hydrating the ions (n_{h_2}=4 for Li^+ and n_{h_3}=1 for Br^-).

The COSMO-radii of atomic elements (Li, Br, H and O) are given in Table III-6.

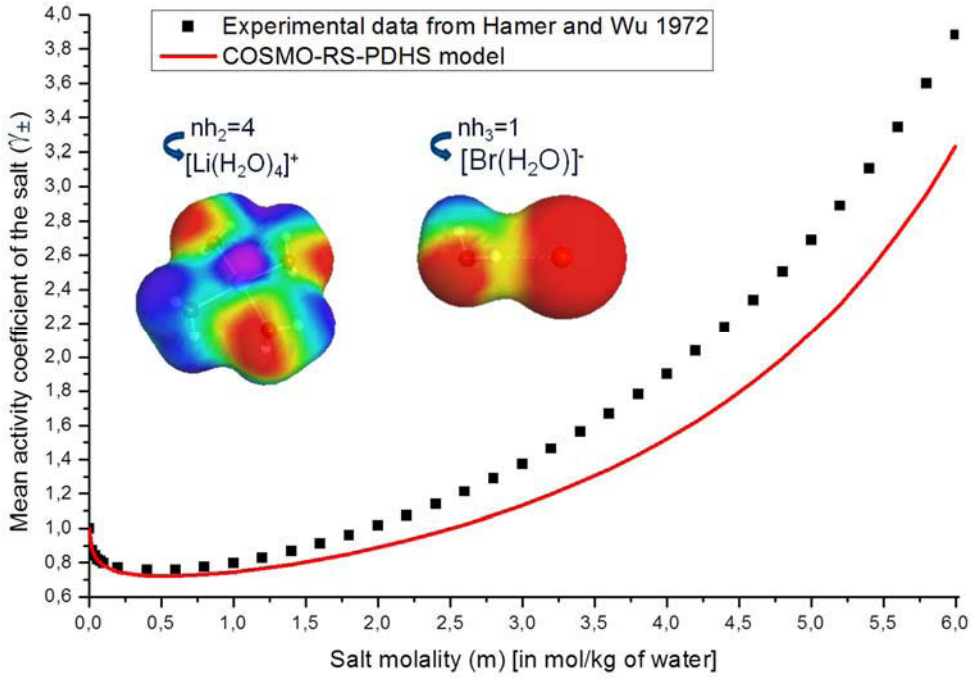


Figure III-21 : $\gamma_{\pm}^{(m)}$ results predicted by the COSMO-RS-PDHS model on the binary water- LiBr, after hydrating the ions (n_{h_2}=4 for Li^+ and n_{h_3}=1 for Br^-).

$\gamma_{\pm}^{(m)}$ denotes the mean activity coefficient of the salt is expressed in molality scale; the COSMO-radii of atomic elements are given in Table III-6.

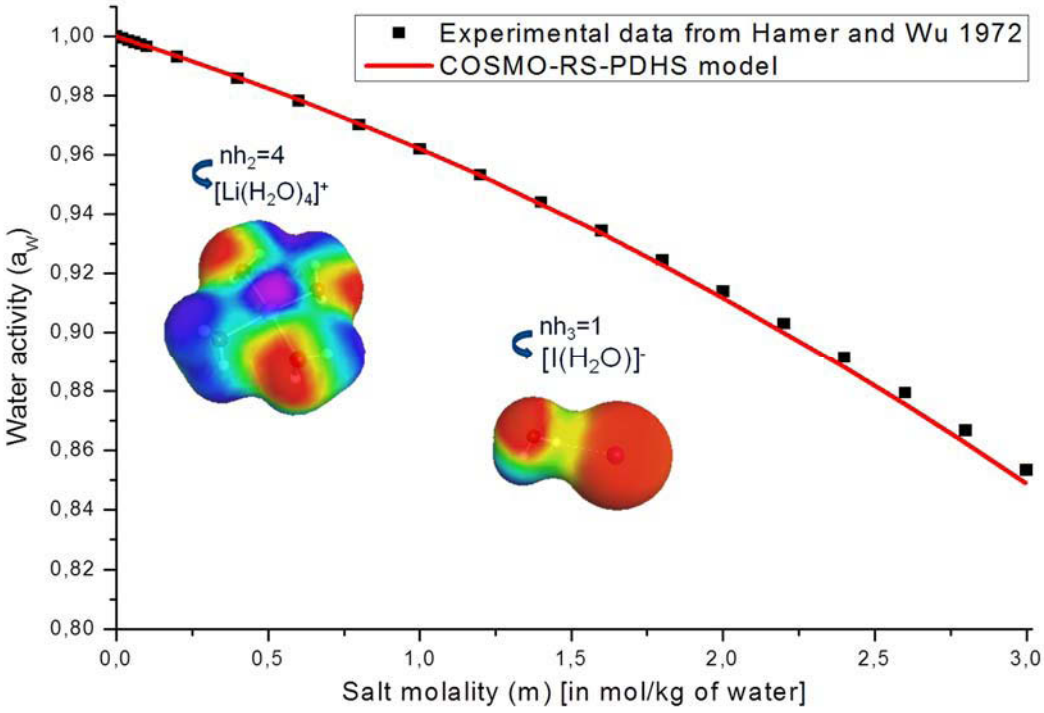


Figure III-22 : Water activity results predicted by the COSMO-RS-PDHS model on the binary water -LiI, after hydrating the ions (n_{h_2} =4 for Li^+ and n_{h_3} =1 for I^-).

The COSMO-radii of atomic elements (Li, I, H and O) are given in Table III-6.

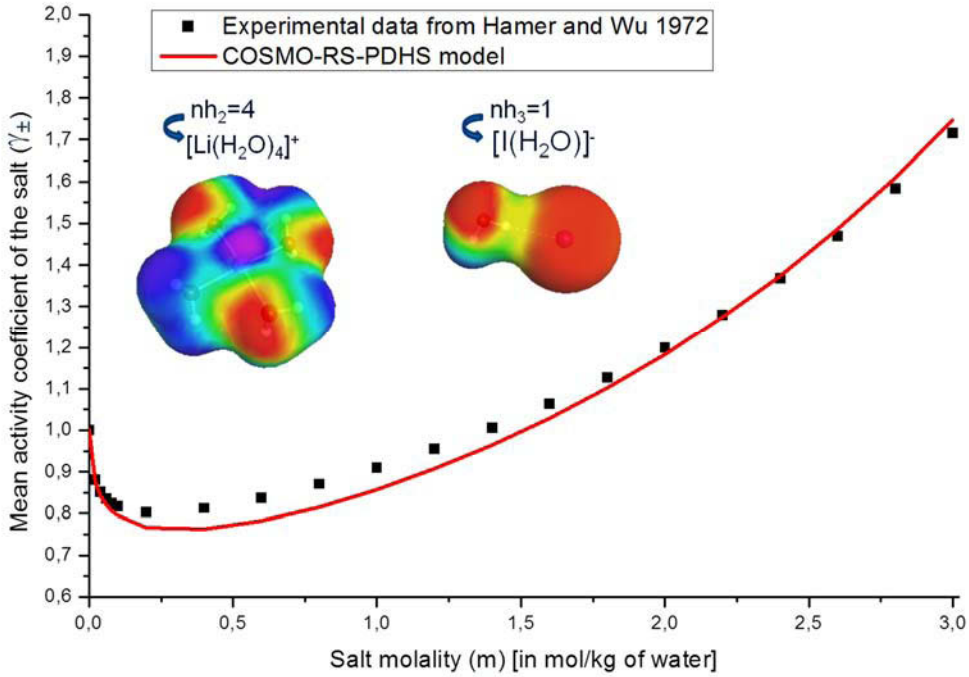


Figure III-23 : $\gamma_{\pm}^{(m)}$ results predicted by the COSMO-RS-PDHS model on the binary water- LiI, after hydrating the ions (n_{h_2} =4 for Li^+ and n_{h_3} =1 for I^-).

$\gamma_{\pm}^{(m)}$ denotes the mean activity coefficient of the salt is expressed in molality scale; the COSMO-radii of atomic elements are given in Table III-6.

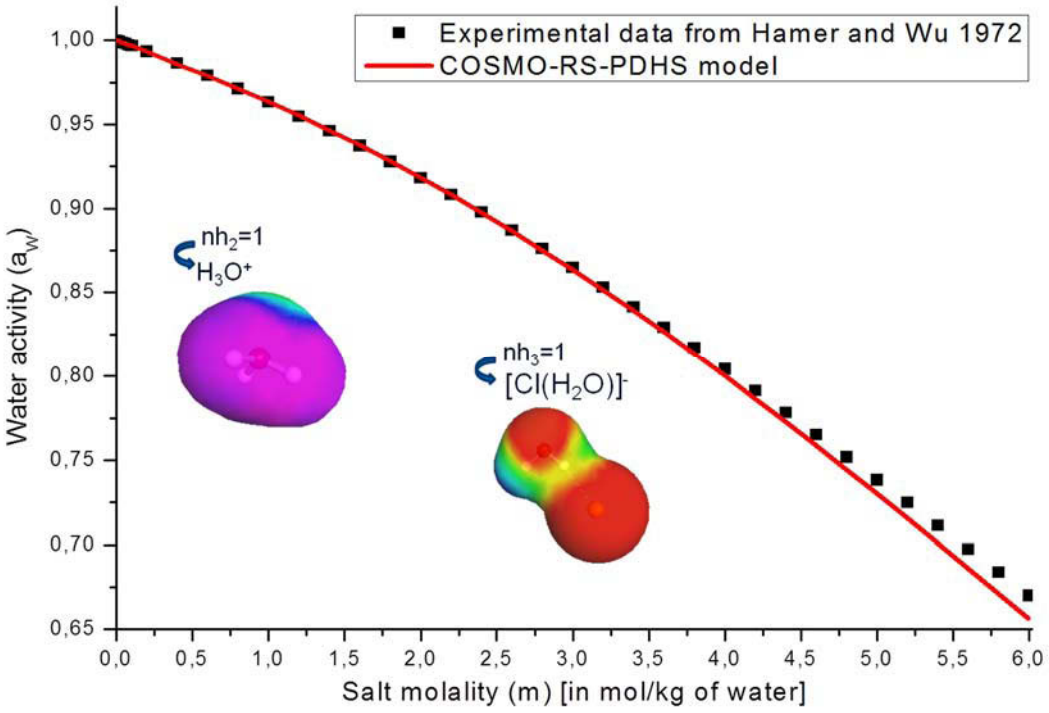


Figure III-24 : Water activity results predicted by the COSMO-RS-PDHS model on the binary water -HCl, after hydrating the ions (n_{h_2} =1 for H^+ and n_{h_3} =1 for Cl^-).

The COSMO-radii of atomic elements (Cl, H and O) are given in Table III-6.

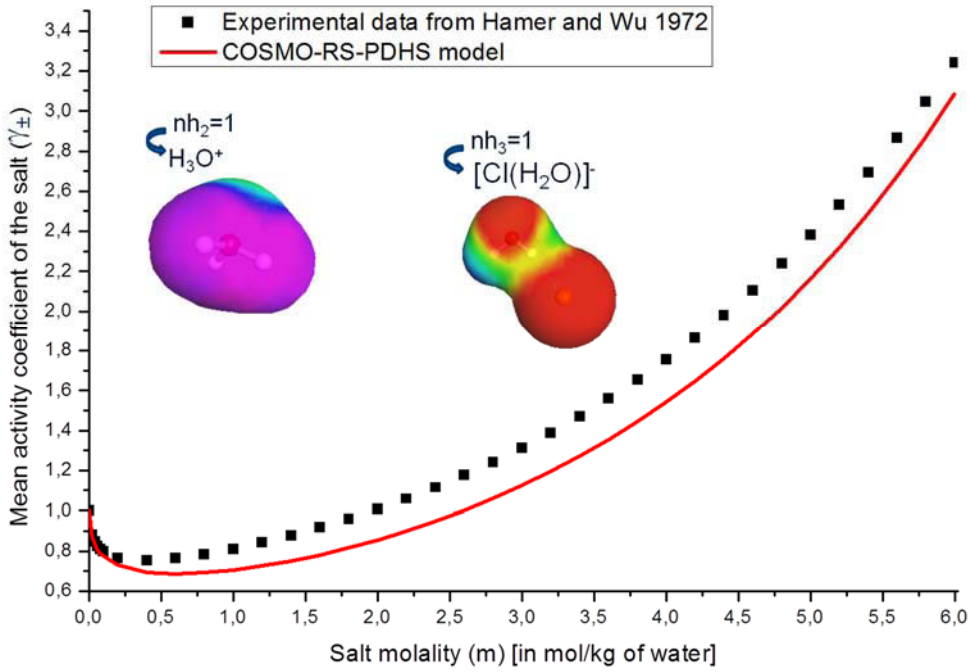


Figure III-25 : $\gamma_{\pm}^{(m)}$ results predicted by the COSMO-RS-PDHS model on the binary water- HCl, after hydrating the ions (n_{h_2} =1 for H^+ and n_{h_3} =1 for Cl^-).

$\gamma_{\pm}^{(m)}$ denotes the mean activity coefficient of the salt is expressed in molality scale; the COSMO-radii of atomic elements are given in Table III-6.

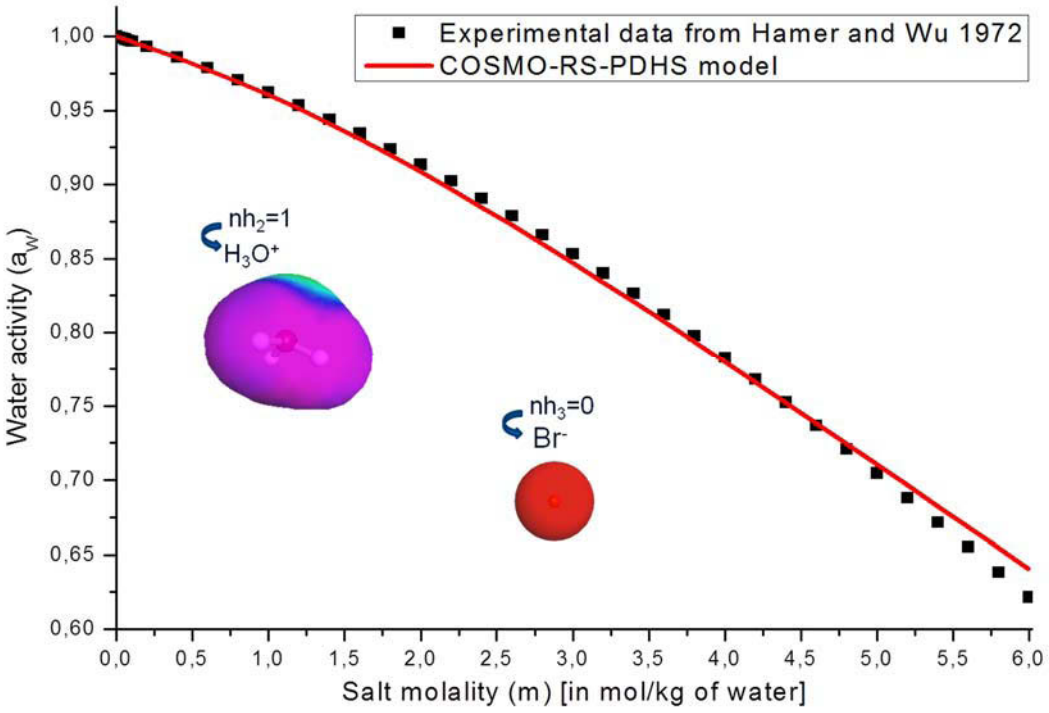


Figure III-26 : Water activity results predicted by the COSMO-RS-PDHS model on the binary water -HBr, after hydrating the ions (n_{h_2} =1 for H^+ and n_{h_3} =0 for Br^-).

The COSMO-radii of atomic elements (Br, H and O) are given in Table III-6.

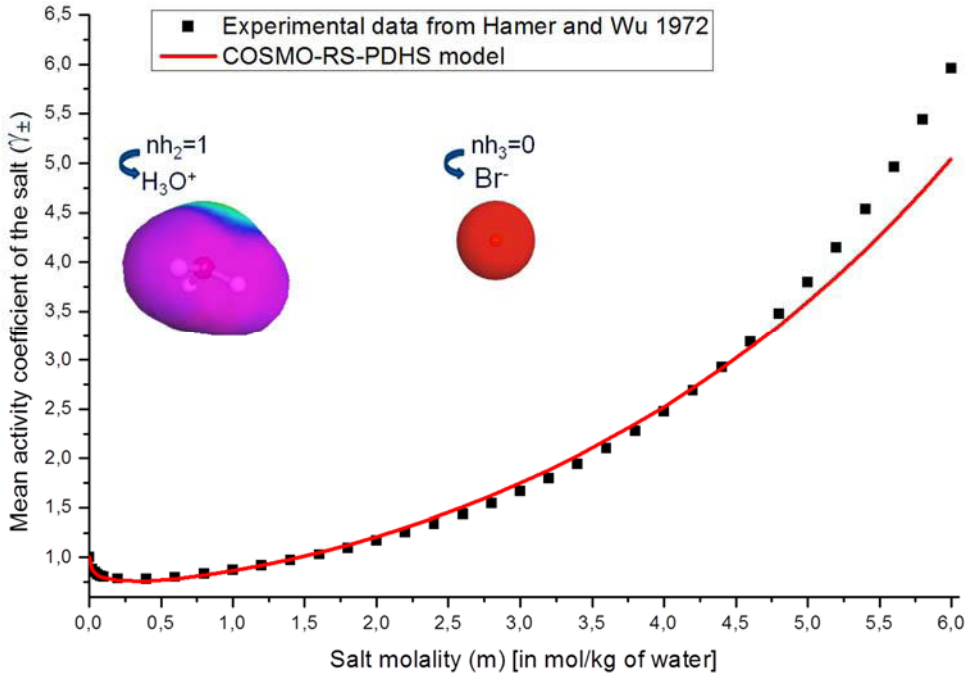


Figure III-27 : $\gamma_{\pm}^{(m)}$ results predicted by the COSMO-RS-PDHS model on the binary water-HBr, after hydrating the ions (n_{h_2} =1 for H^+ and n_{h_3} =0 for Br^-).

$\gamma_{\pm}^{(m)}$ denotes the mean activity coefficient of the salt is expressed in molality scale; the COSMO-radii of atomic elements are given in Table III-6.

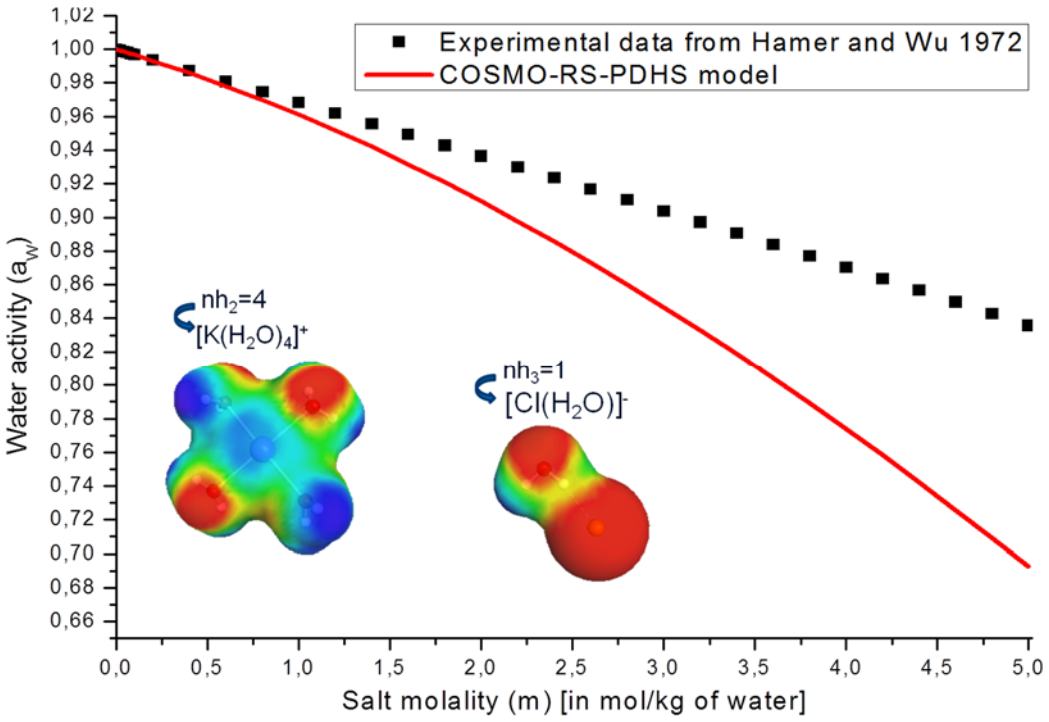


Figure III-28 : Water activity results predicted by the COSMO-RS-PDHS model on the binary water -KCl, after hydrating the ions (n_{h_2} =4 for K^+ and n_{h_3} =1 for Cl^-).

The COSMO-radii of atomic elements (K, Cl, H and O) are given in Table III-6.

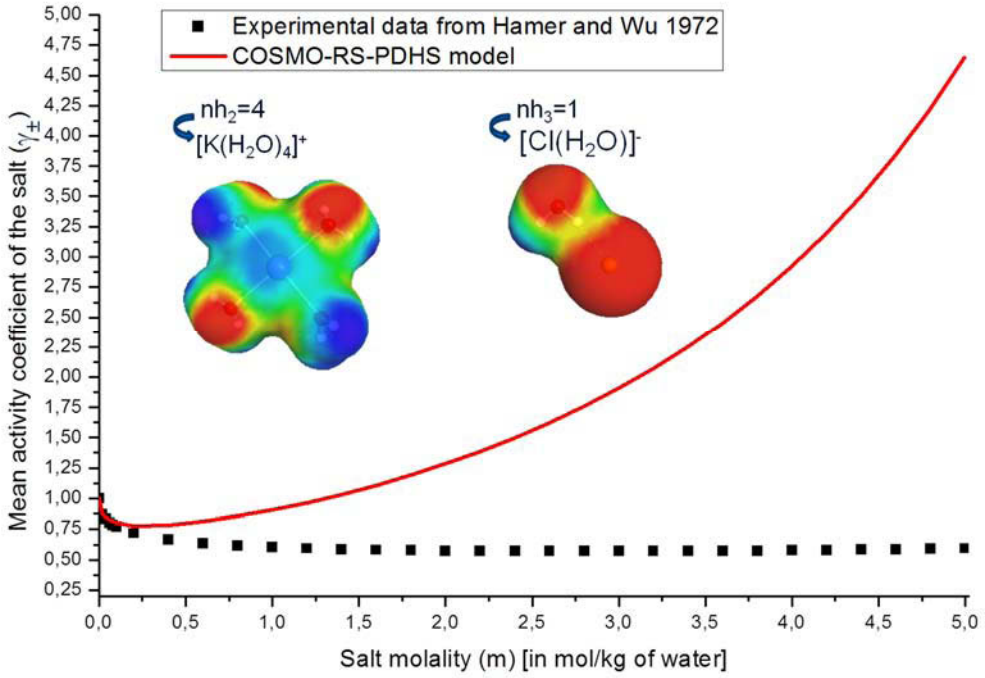


Figure III-29 : $\gamma_{\pm}^{(m)}$ results predicted by the COSMO-RS-PDHS model on the binary water- KCl, after hydrating the ions (n_{h_2} =4 for K^+ and n_{h_3} =1 for Cl^-).

$\gamma_{\pm}^{(m)}$ denotes the mean activity coefficient of the salt is expressed in molality scale; the COSMO-radii of atomic elements are given in Table III-6.

However, as illustrated in Figures III-28 and III-29, any attempt of prediction of the activity coefficients of systems containing K^+ has failed when we have used one of the COSMO-radii of K^+ suggested in COSMO-RS related papers (Klamt et al. 1998 and 2000, Klamt 2005 and 2011) or in the suggested literature (Mantina et al. 2009, Bondi 1964, Sutton 1965).

III.4.2. Results obtained by using COSMO-radii of ionic elements suggested in this thesis

As discussed earlier, the consistency between the values of $d_{ion\text{-}water}$ calculated in section III.2 and those given by Marcus (1997 and 2012) justifies the choice of the ionic radii values given by Marcus as reference values for calculating COSMO-radii of cations (elements of the IA and IIA groups) by applying the scaling factor of 1.17 times R_{ion}. By analogy, the same scaling factor can be applied for anions. This leads to a set of COSMO-radii of anions that are consistent with those calculated from van der Waals radii and used to get the values given in Table III-6 even if there are slight differences. The COSMO-radii calculated from the ionic radii given by Marcus (1997 and 2012) are given in Table III-7.

Table III-7: COSMO-radii of monovalent atomic ions calculated from the ionic radii suggested by Marcus (1997 and 2012).

[a]The COSMO-radii are approximately equal to 1.17 times the radii proposed by Marcus (1997 and 2012) for these monovalent ions.

Monovalent Anions		Monovalent cations	
Elements	COSMO-radii R_{COSMO} [in A] ($1A=10^{-10}m$)	Elements	COSMO-radii R_{COSMO} [in A] ($1A=10^{-10}m$)
Cl	2.118[a]	Li	0.810[a]
Br	2.293[a]	Na	1.190[a]
I	2.574[a]	K	1.610[a]
F	1.556[a]	Rb	1.740[a]
		Cs	1.990[a]

It is interesting to test the prediction performance of COSMO-RS-PDHS on several binary (water-salt) mixtures using these radii. The results obtained are illustrated in Figures III-30 to III-51.

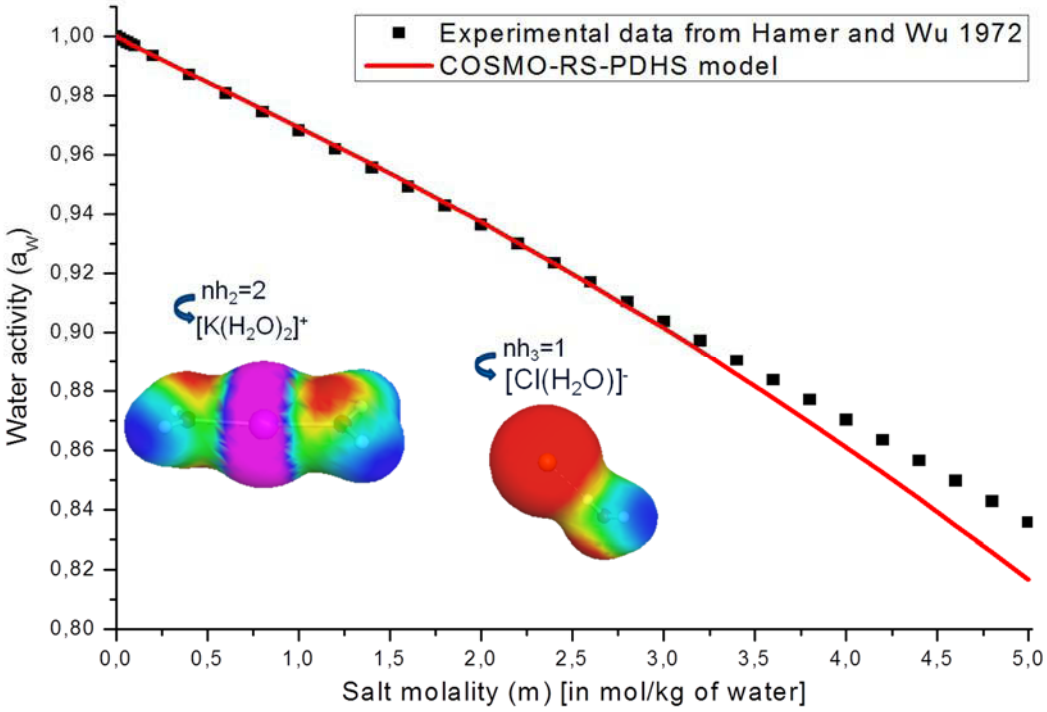


Figure III-30 : Water activity results predicted by the COSMO-RS-PDHS model on the binary water -KCl, after hydrating the ions (n_{h_2} =2 for K^+ and n_{h_3} =1 for Cl^-).

The COSMO-radii of elements are given in Table III-7 for ionic atoms (K and Cl) and Table III-6 for the other elements (H and O).

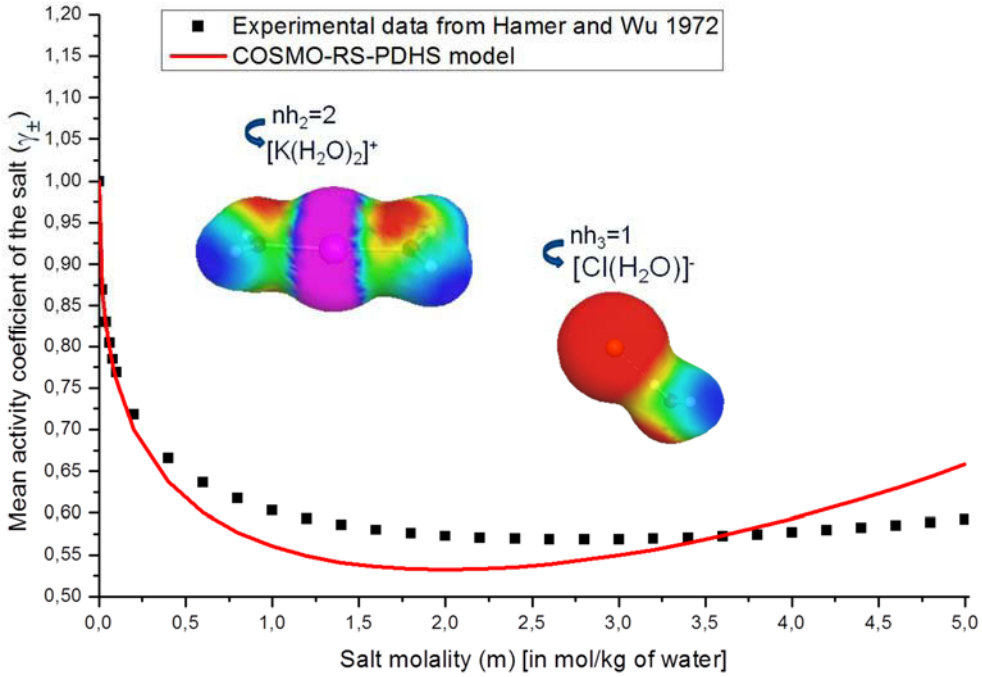


Figure III-31 : $\gamma_\pm^{(m)}$ results predicted by the COSMO-RS-PDHS model on the binary water- KCl, after hydrating the ions (n_{h_2} =2 for K^+ and n_{h_3} =1 for Cl^-).

$\gamma_\pm^{(m)}$ denotes the mean activity coefficient of the salt is expressed in molality scale; the COSMO-radii of elements are given in Table III-7 for ionic atoms (K and Cl) and Table III-6 for the other elements (H and O).

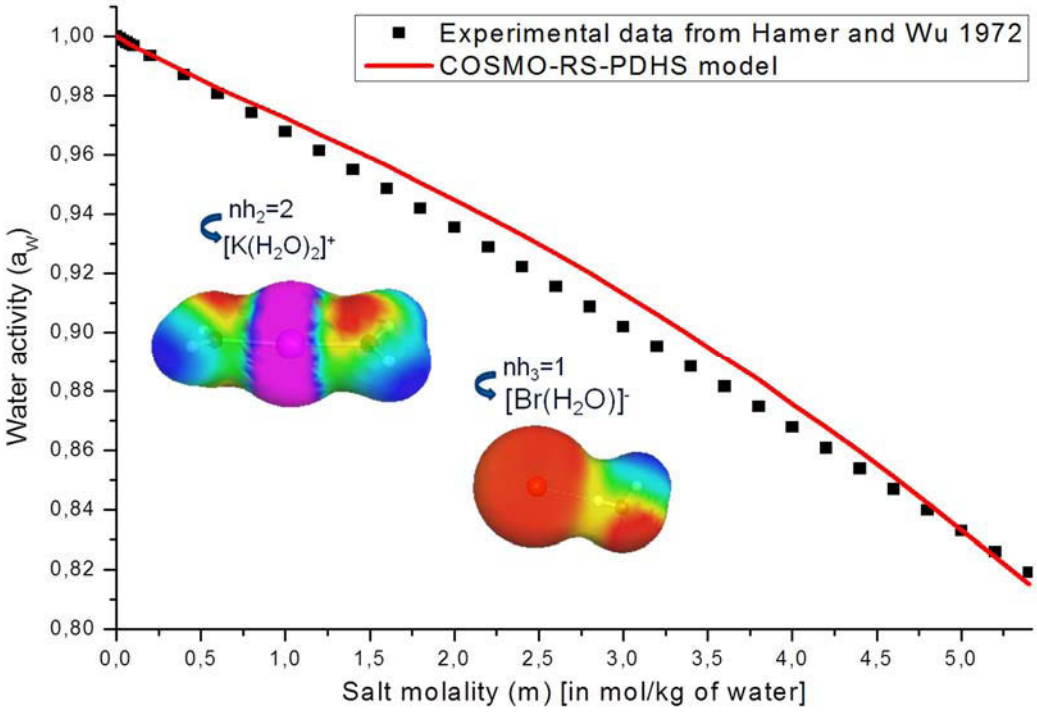


Figure III-32 : Water activity results predicted by the COSMO-RS-PDHS model on the binary water -KBr, after hydrating the ions (n_{h_2} =2 for K^+ and n_{h_3} =1 for Br^-).

The COSMO-radii of elements are given in Table III-7 for ionic atoms (K and Br) and Table III-6 for the other elements (H and O).

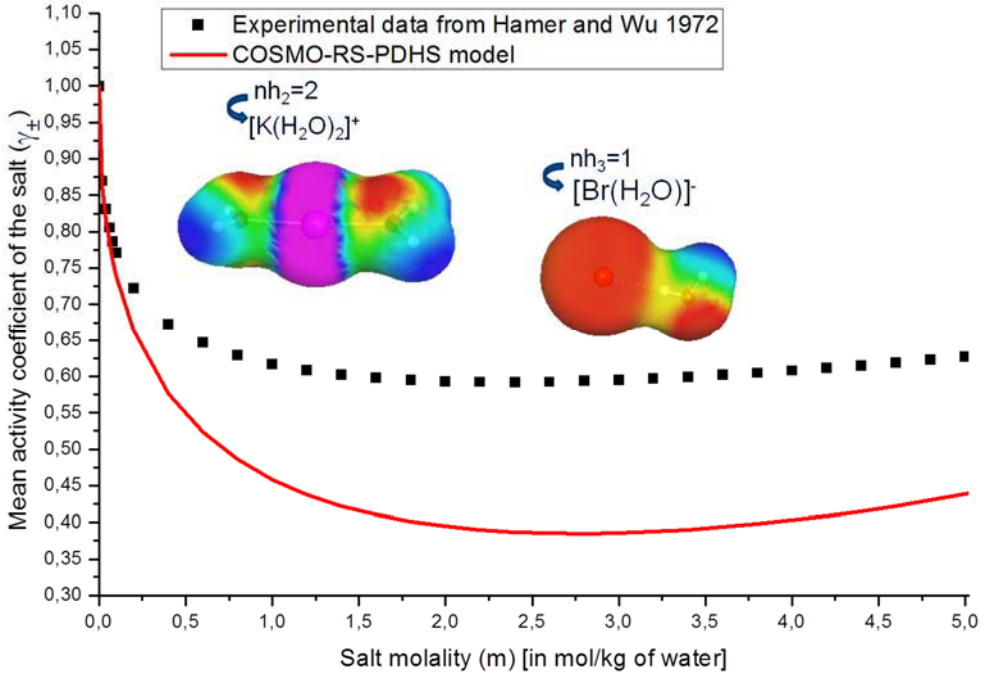


Figure III-33 : $\gamma_{\pm}^{(m)}$ results predicted by the COSMO-RS-PDHS model on the binary water- KBr, after hydrating the ions (n_{h_2} =2 for K^+ and n_{h_3} =1 for Br^-).

$\gamma_{\pm}^{(m)}$ denotes the mean activity coefficient of the salt is expressed in molality scale; the COSMO-radii of elements are given in Table III-7 for ionic atoms (K and Br) and Table III-6 for the other elements (H and O).

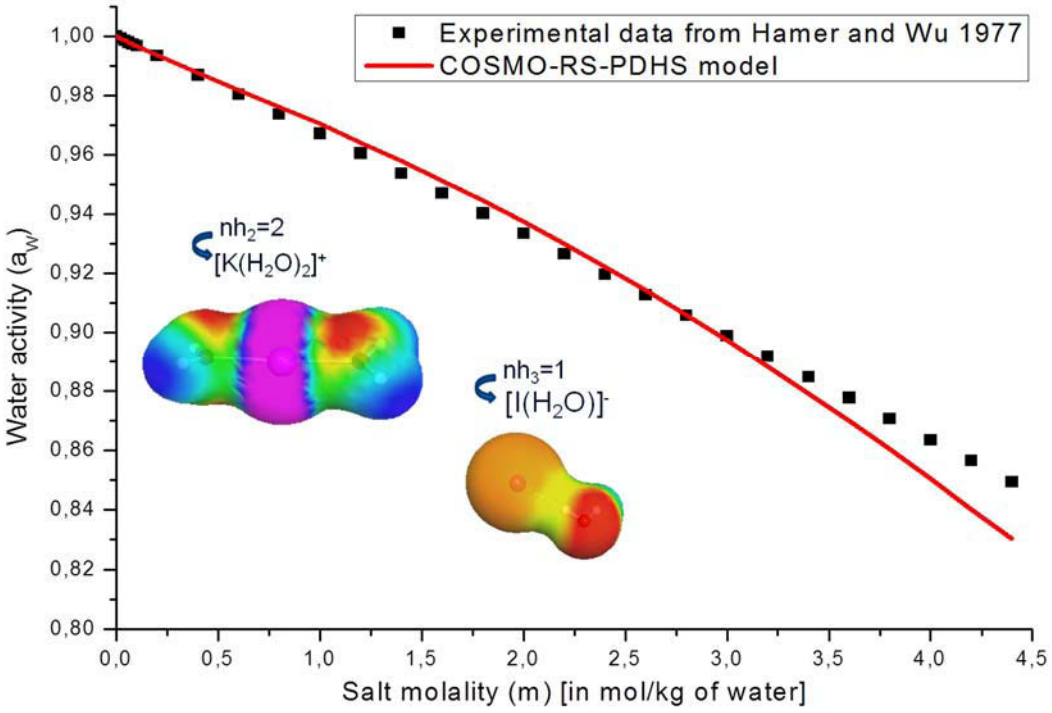


Figure III-34 : Water activity results predicted by the COSMO-RS-PDHS model on the binary water -KI, after hydrating the ions (n_{h_2}=2 for K^+ and n_{h_3}=1 for I^-).

The COSMO-radii of elements are given in Table III-7 for ionic atoms (K and I) and Table III-6 for the other elements (H and O).

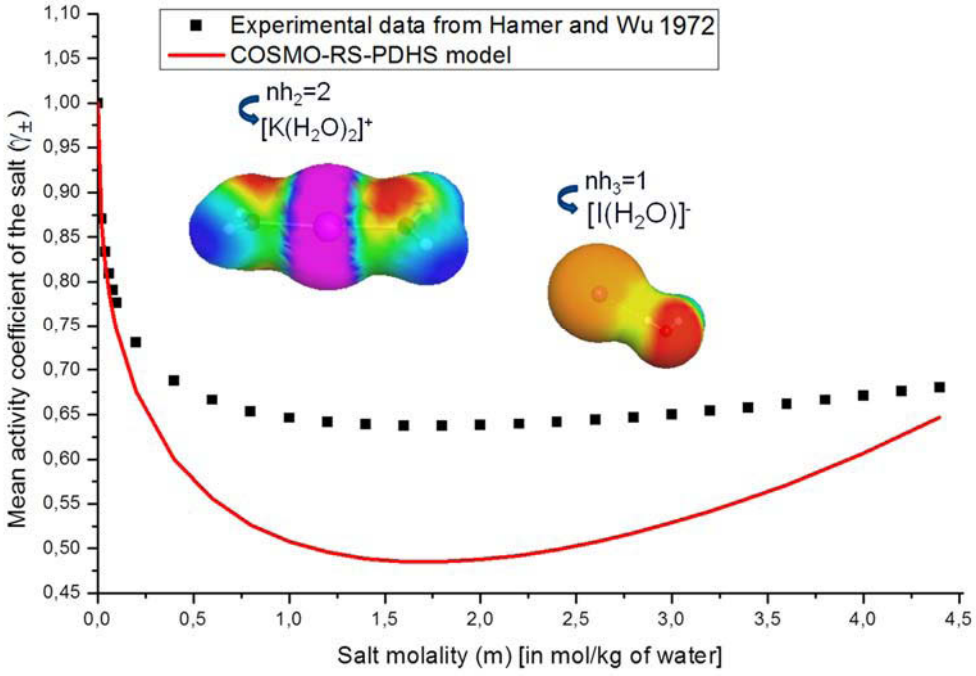


Figure III-35 : $\gamma_{\pm}^{(m)}$ results predicted by the COSMO-RS-PDHS model on the binary water- KI, after hydrating the ions (n_{h_2}=2 for K^+ and n_{h_3}=1 for I^-).

$\gamma_{\pm}^{(m)}$ denotes the mean activity coefficient of the salt is expressed in molality scale; the COSMO-radii of elements are given in Table III-7 for ionic atoms (K and I) and Table III-6 for the other elements (H and O).

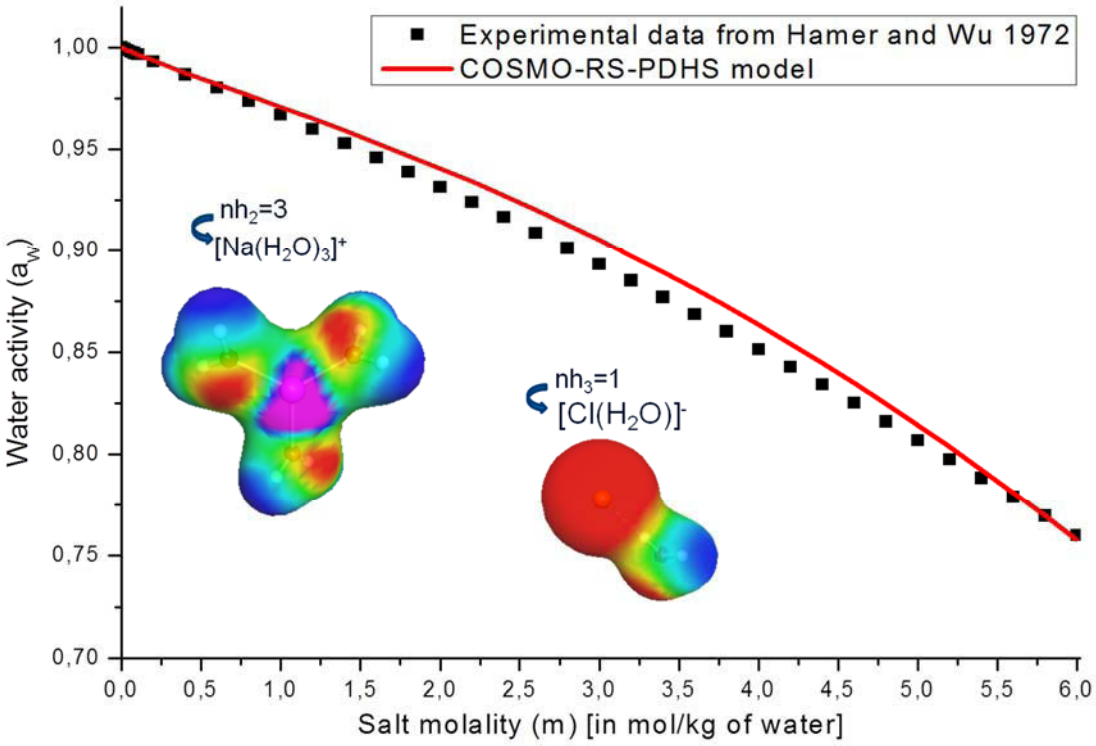


Figure III-36 : Water activity results predicted by the COSMO-RS-PDHS model on the binary water -NaCl, after hydrating the ions (n_{h_2} =3 for Na^+ and n_{h_3} =1 for Cl^-).

The COSMO-radii of elements are given in Table III-7 for ionic atoms (Na and Cl) and Table III-6 for the other elements (H and O).

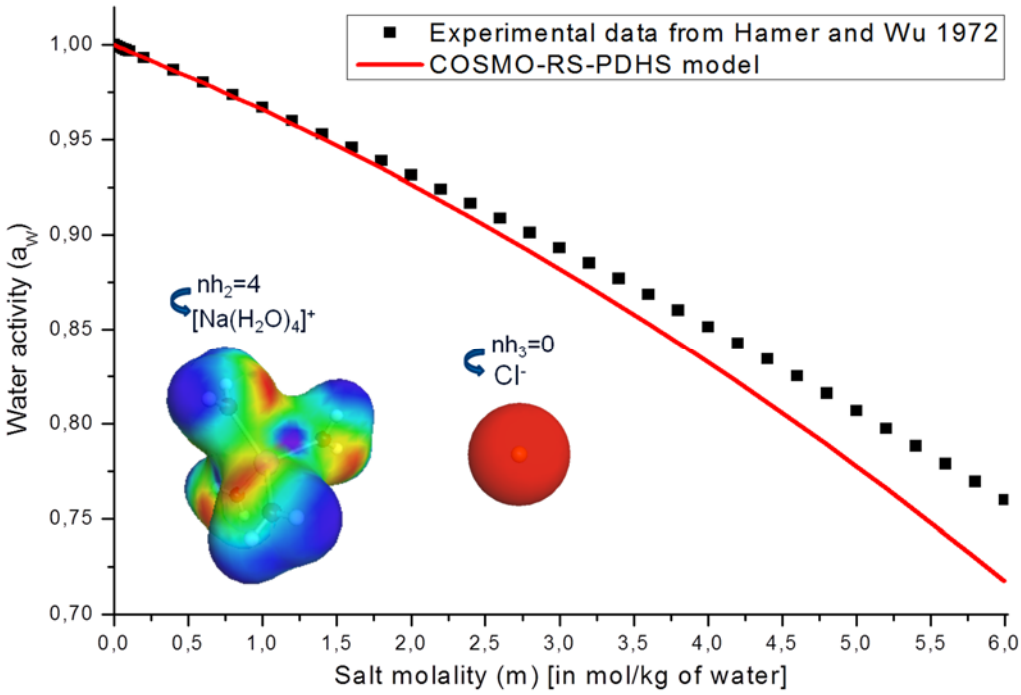


Figure III-37 : Water activity results predicted by the COSMO-RS-PDHS model on the binary water-NaCl, after hydrating the ions (n_{h_2} =4 for Na^+ and n_{h_3} =0 for Cl^-).

The COSMO-radii of elements are given in Table III-7 for ionic atoms (Na and Cl) and Table III-6 for the other elements (H and O).

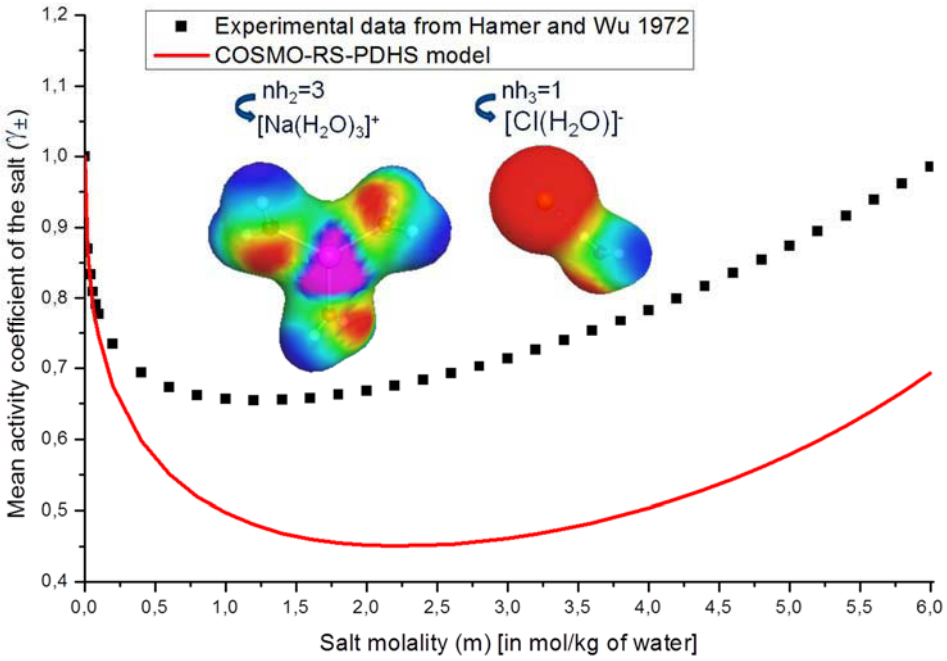


Figure III-38 : $\gamma_{\pm}^{(m)}$ results predicted by the COSMO-RS-PDHS model on the binary water -NaCl, after hydrating the ions (n_{h_2}=3 for Na^+ and n_{h_3}=1 for Cl^-).

$\gamma_{\pm}^{(m)}$ denotes the mean activity coefficient of the salt is expressed in molality scale; the COSMO-radii of elements are given in Table III-7 for ionic atoms (Na and Cl) and Table III-6 for the other elements (H and O).

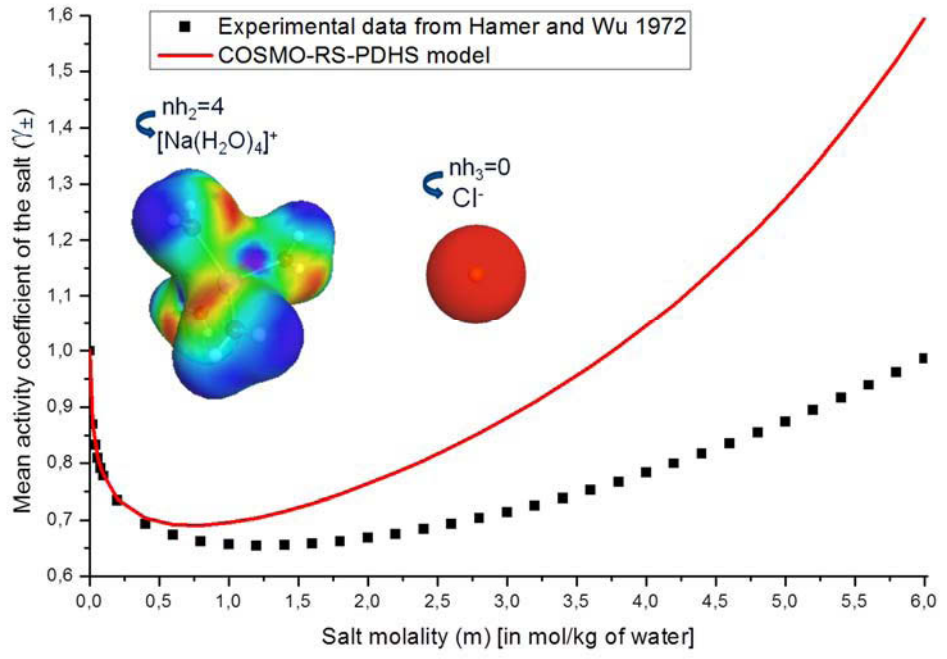


Figure III-39 : $\gamma_{\pm}^{(m)}$ results predicted by the COSMO-RS-PDHS model on the binary water-NaCl, after hydrating the ions (n_{h_2}=4 for Na^+ and n_{h_3}=0 for Cl^-).

$\gamma_{\pm}^{(m)}$ denotes the mean activity coefficient of the salt is expressed in molality scale; the COSMO-radii of elements are given in Table III-7 for ionic atoms (Na and Cl) and Table III-6 for the other elements (H and O).

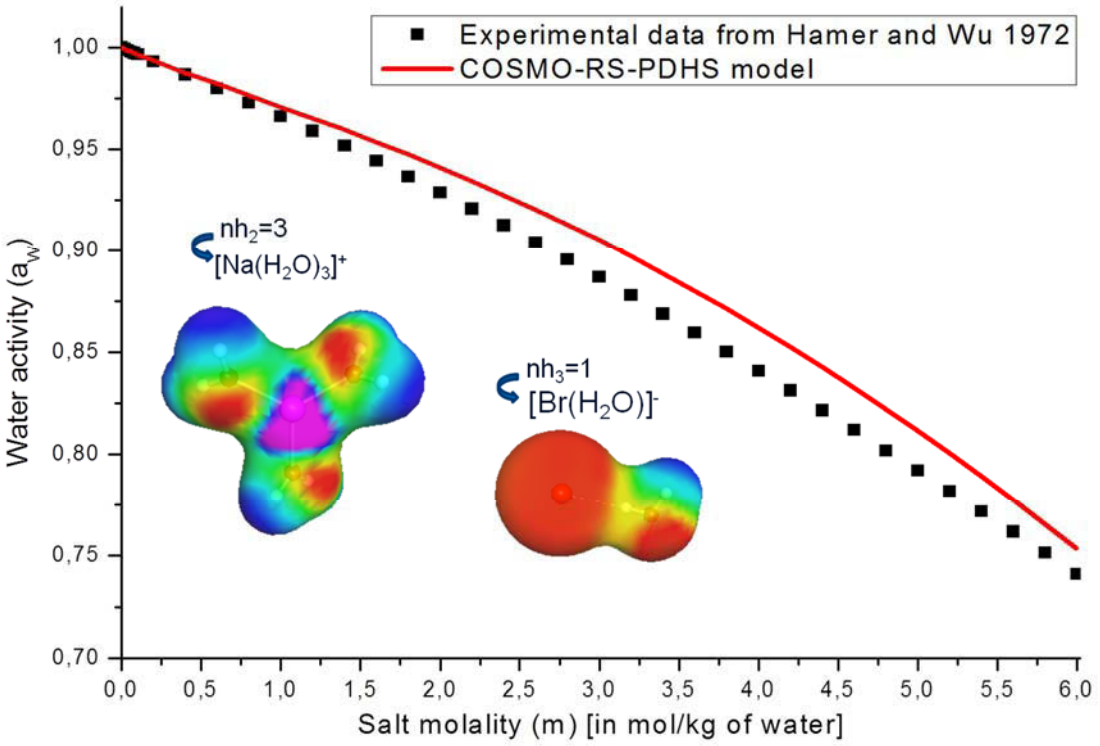


Figure III-40 : Water activity results predicted by the COSMO-RS-PDHS model on the binary water -NaBr, after hydrating the ions (n_{h_2}=3 for Na^+ and n_{h_3}=1 for Br^-).

The COSMO-radii of elements are given in Table III-7 for ionic atoms (Na and Br) and Table III-6 for the other elements (H and O).

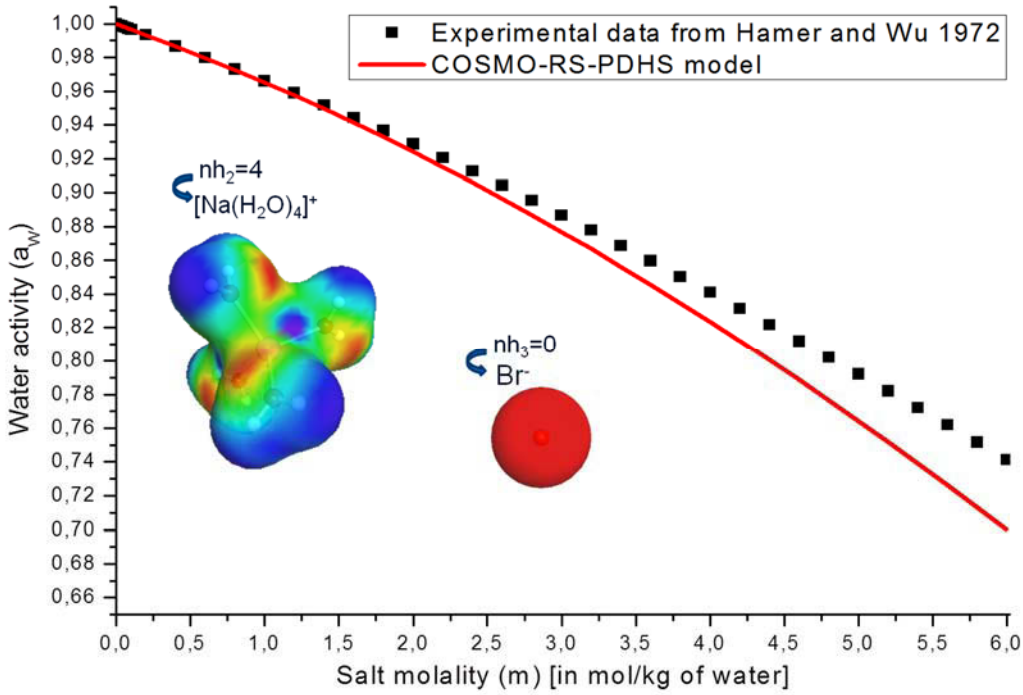


Figure III-41 : Water activity results predicted by the COSMO-RS-PDHS model on the binary water-NaBr, after hydrating the ions (n_{h_2}=4 for Na^+ and n_{h_3}=0 for Br^-).

The COSMO-radii of elements are given in Table III-7 for ionic atoms (Na and Br) and Table III-6 for the other elements (H and O).

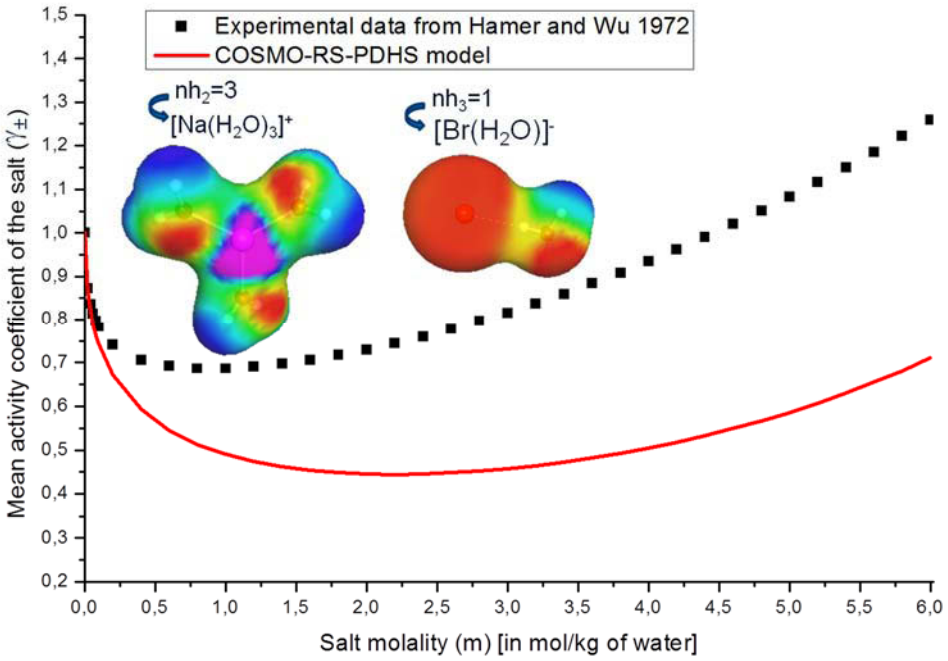


Figure III-42 : $\gamma_{\pm}^{(m)}$ results predicted by the COSMO-RS-PDHS model on the binary water -NaBr, after hydrating the ions (n_{h_2}=3 for Na^+ and n_{h_3}=1 for Br^-).

$\gamma_{\pm}^{(m)}$ denotes the mean activity coefficient of the salt is expressed in molality scale; the COSMO-radii of elements are given in Table III-7 for ionic atoms (Na and Br) and Table III-6 for the other elements (H and O).

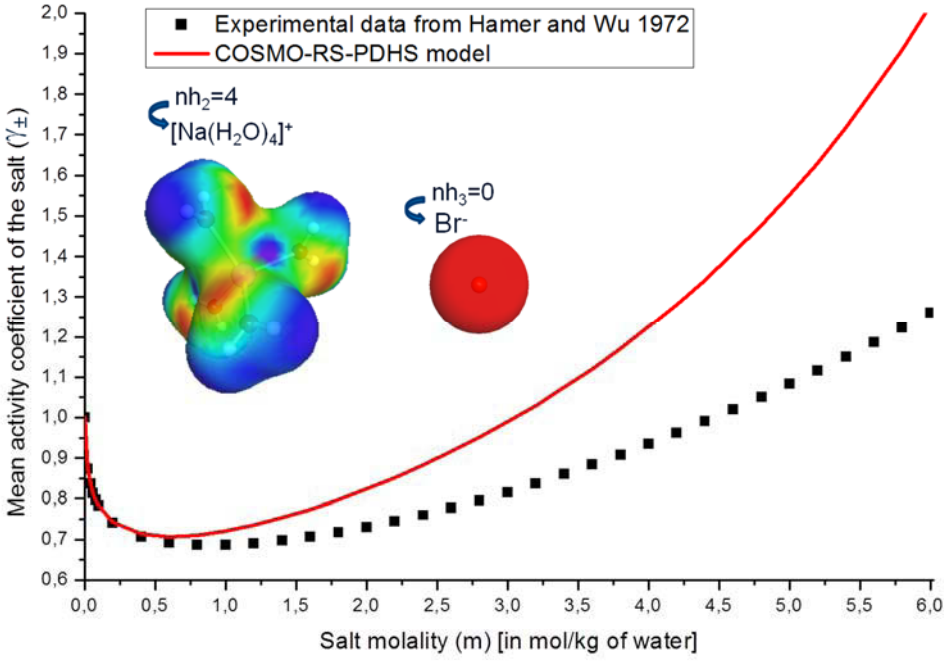


Figure III-43 : $\gamma_{\pm}^{(m)}$ results predicted by the COSMO-RS-PDHS model on the binary water-NaBr, after hydrating the ions (n_{h_2}=4 for Na^+ and n_{h_3}=0 for Br^-)..

$\gamma_{\pm}^{(m)}$ denotes the mean activity coefficient of the salt is expressed in molality scale; the COSMO-radii of elements are given in Table III-7 for ionic atoms (Na and Br) and Table III-6 for the other elements (H and O).

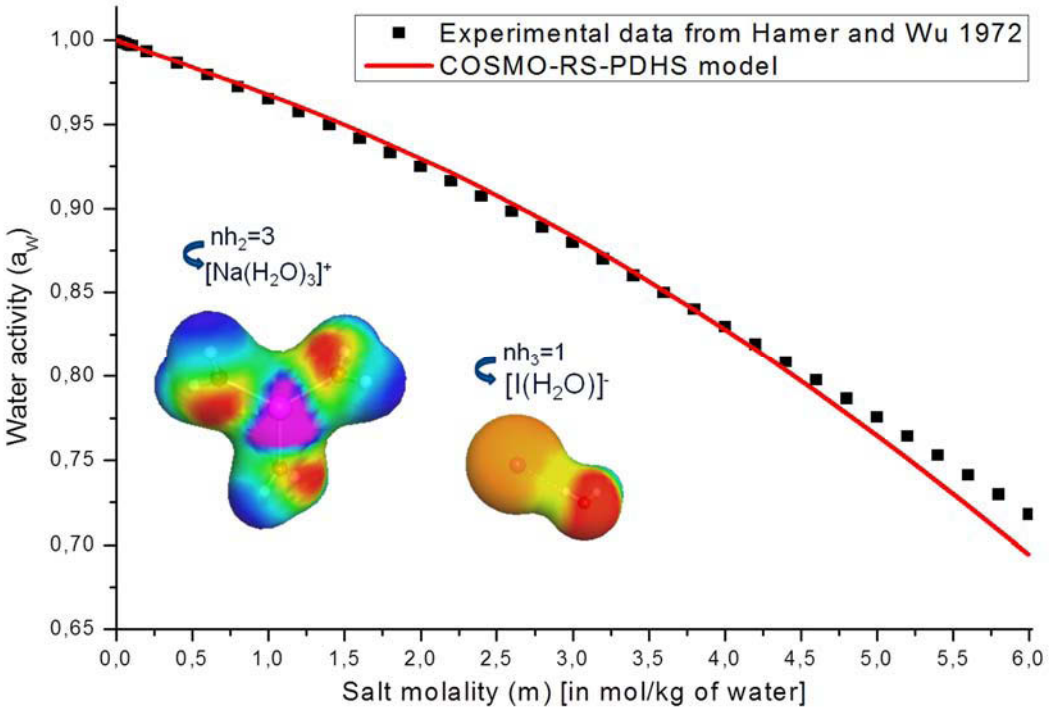


Figure III-44 : Water activity results predicted by the COSMO-RS-PDHS model on the binary water -NaI, after hydrating the ions (n_{h_2}=3 for Na^+ and n_{h_3}=1 for I^-).

The COSMO-radii of elements are given in Table III-7 for ionic atoms (Na and I) and Table III-6 for the other elements (H and O).

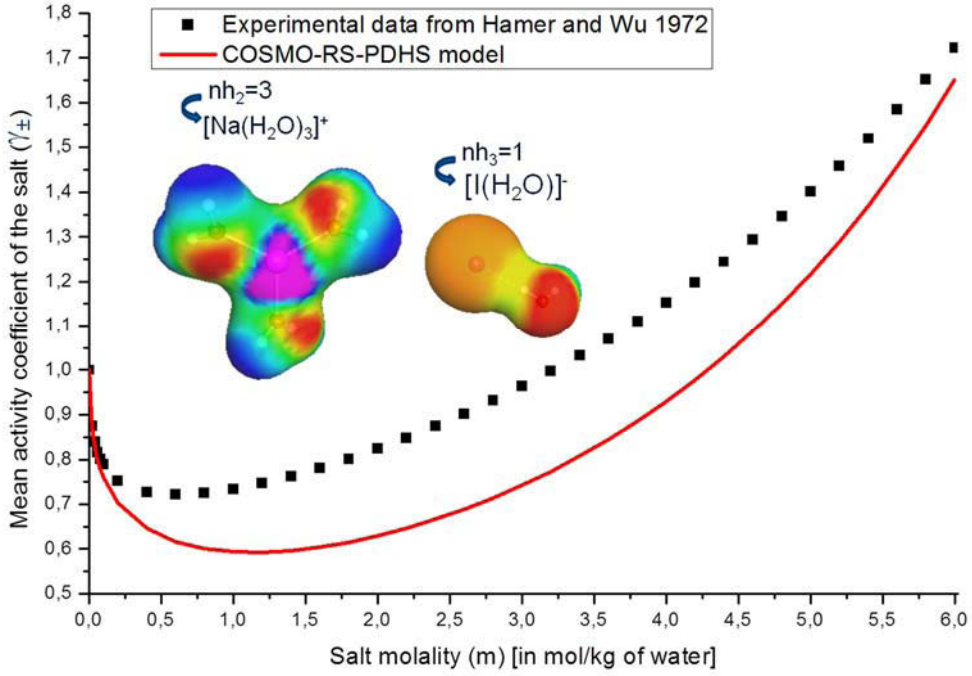


Figure III-45 : $\gamma_{\pm}^{(m)}$ results predicted by the COSMO-RS-PDHS model on the binary water -NaI, after hydrating the ions (n_{h_2}=3 for Na^+ and n_{h_3}=1 for I^-).

$\gamma_{\pm}^{(m)}$ denotes the mean activity coefficient of the salt is expressed in molality scale; the COSMO-radii of elements are given in Table III-7 for ionic atoms (Na and I) and Table III-6 for the other elements (H and O).

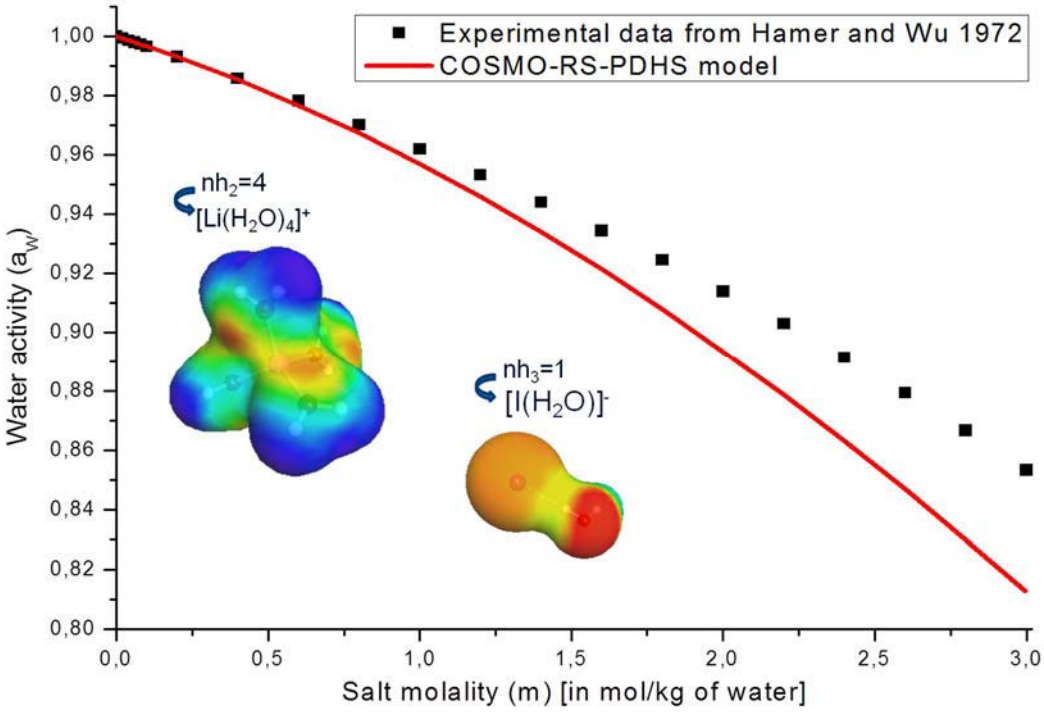


Figure III-46 : Water activity results predicted by the COSMO-RS-PDHS model on the binary water -LiI, after hydrating the ions (n_{h_2} =4 for Li^+ and n_{h_3} =1 for I^-).

The COSMO-radii of elements are given in Table III-7 for ionic atoms (Li and I) and Table III-6 for the other elements (H and O).

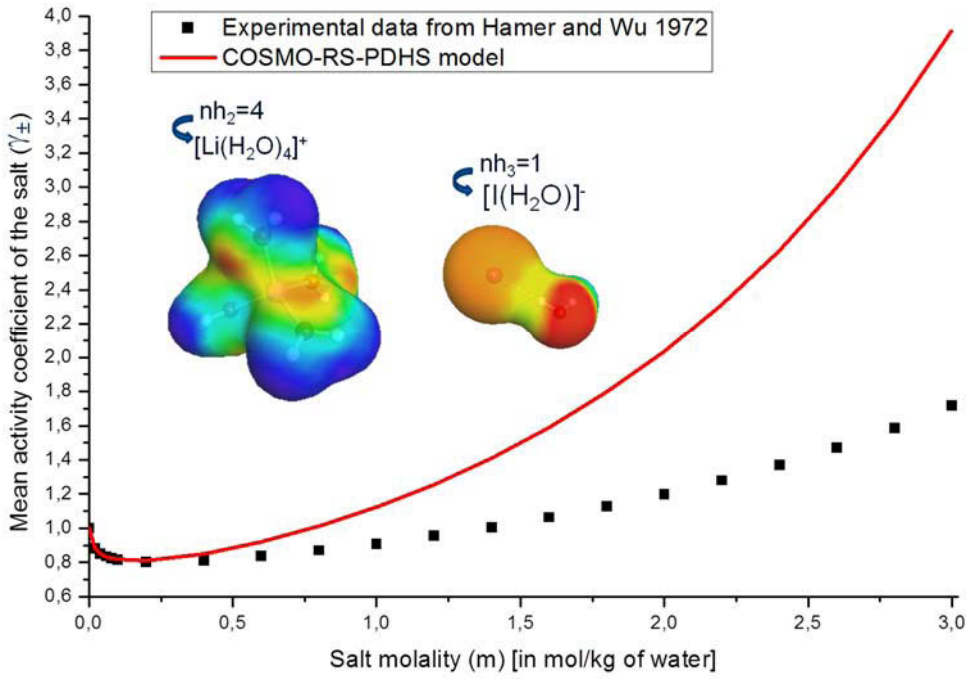


Figure III-47 : $\gamma_{\pm}^{(m)}$ results predicted by the COSMO-RS-PDHS model on the binary water -LiI, after hydrating the ions (n_{h_2} =4 for Li^+ and n_{h_3} =1 for I^-).

$\gamma_{\pm}^{(m)}$ denotes the mean activity coefficient of the salt is expressed in molality scale; the COSMO-radii of elements are given in Table III-7 for ionic atoms (Li and I) and Table III-6 for the other elements (H and O).

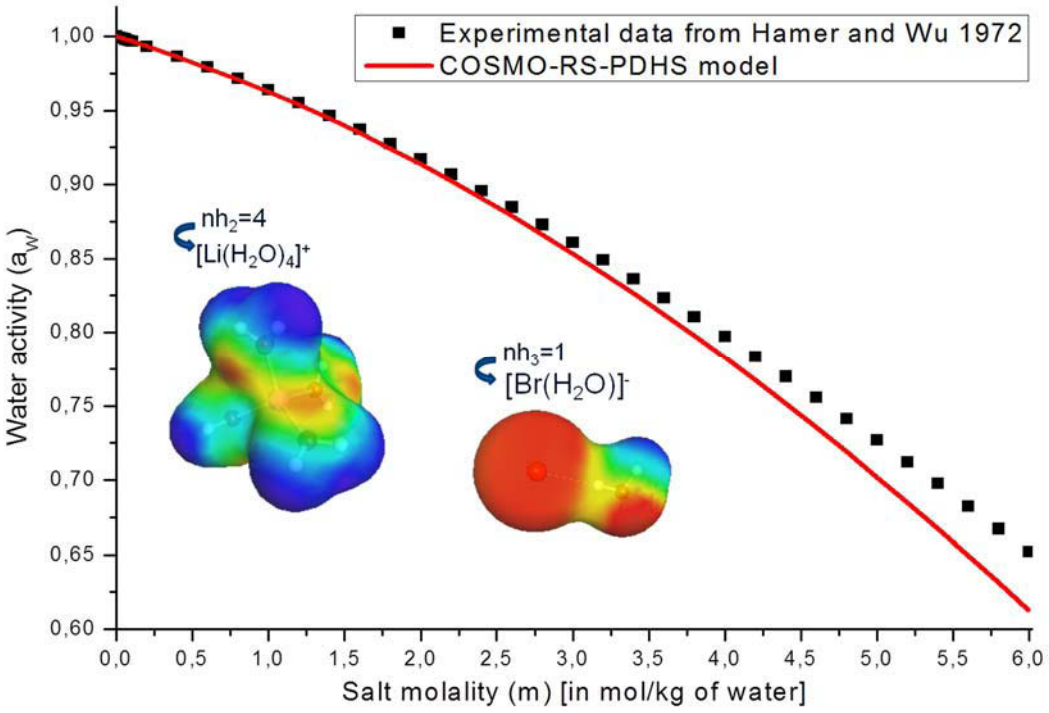


Figure III-48 : Water activity results predicted by the COSMO-RS-PDHS model on the binary water -LiBr, after hydrating the ions (n_{h_2}=4 for Li^+ and n_{h_3}=1 for Br^-). *The COSMO-radii of elements are given in Table III-7 for ionic atoms (Li and Br) and Table III-6 for the other elements (H and O).*

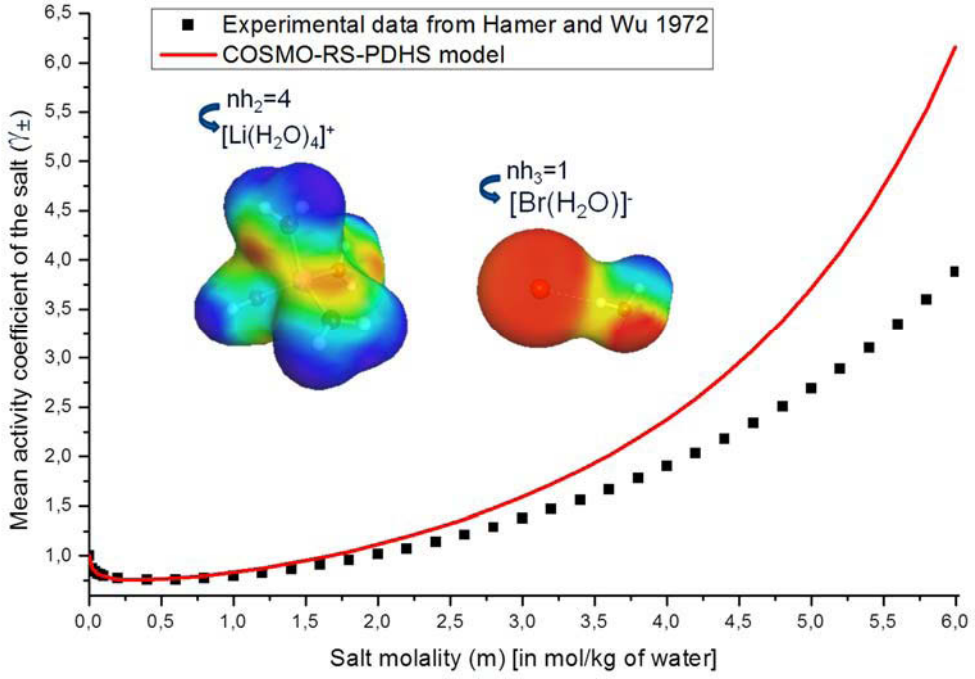


Figure III-49 : $\gamma_{\pm}^{(m)}$ results predicted by the COSMO-RS-PDHS model on the binary water -LiBr, after hydrating the ions (n_{h_2}=4 for Li^+ and n_{h_3}=1 for Br^-).

$\gamma_{\pm}^{(m)}$ denotes the mean activity coefficient of the salt is expressed in molality scale; the COSMO-radii of elements are given in Table III-7 for ionic atoms (Li and Br) and Table III-6 for the other elements (H and O).

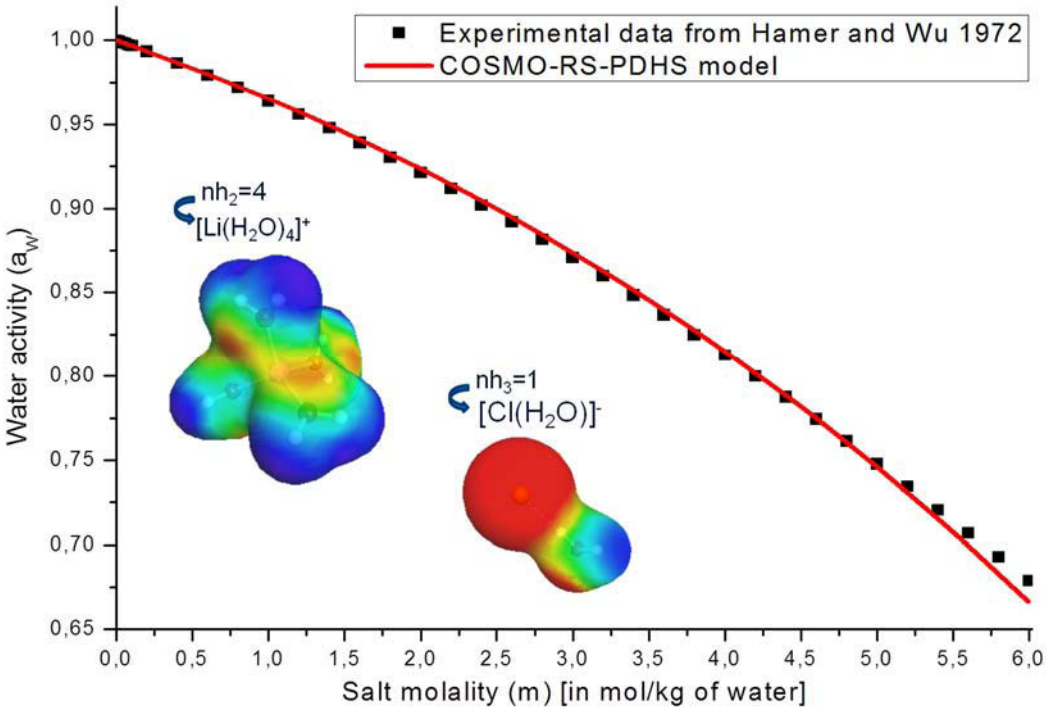


Figure III-50 : Water activity results predicted by the COSMO-RS-PDHS model on the binary water -LiCl, after hydrating the ions (n_{h_2}=4 for Li^+ and n_{h_3}=1 for Cl^-). *The COSMO-radii of elements are given in Table III-7 for ionic atoms (Li and Cl) and Table III-6 for the other elements (H and O).*

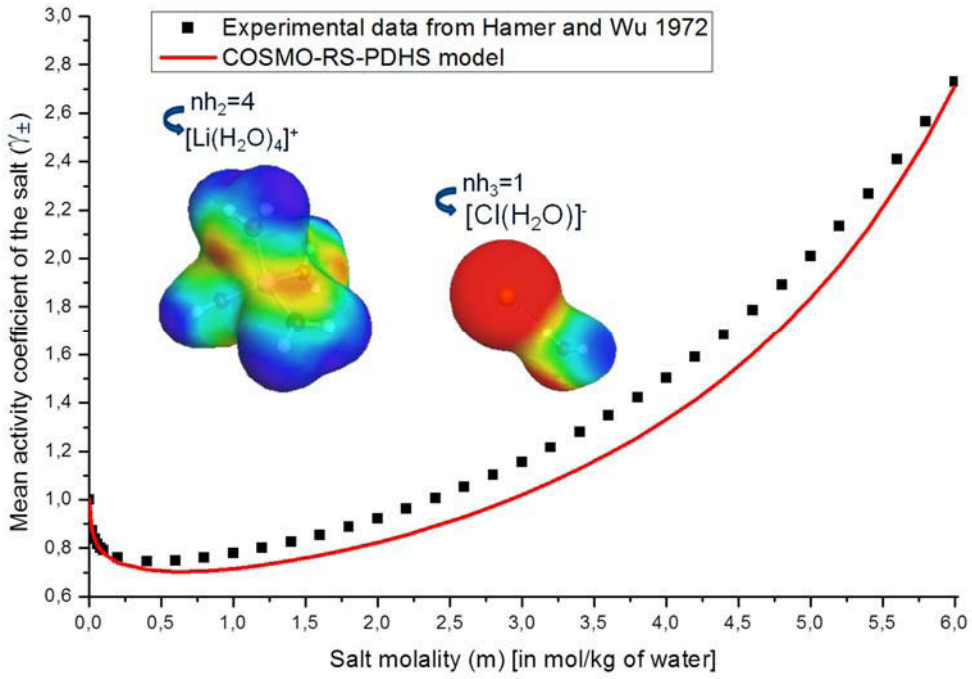


Figure III-51 : $\gamma_{\pm}^{(m)}$ results predicted by the COSMO-RS-PDHS model on the binary water -LiCl, after hydrating the ions (n_{h_2}=4 for Li^+ and n_{h_3}=1 for Cl^-).

$\gamma_{\pm}^{(m)}$ denotes the mean activity coefficient of the salt is expressed in molality scale; the COSMO-radii of elements are given in Table III-7 for ionic atoms (Li and Cl) and Table III-6 for the other elements (H and O).

III.5. Concluding remarks about the prediction of activity coefficients of systems containing electrolytes

From the above, it is proved that the combination of the COSMO-RS method, the un-symmetric Pitzer Debye-Hückel model and a chemical treatment of the solvation or clustering of ions (*i.e.* the COSMO-RS-PDHS model) is very promising for predicting water activities and salt mean activity coefficients.

As illustrated above, the predicting power of the COSMO-RS-PDHS model is better for calculating water activity (that is an important property in food and biological systems, as discussed in chapter I) than for the prediction of mean ionic activity coefficients. Such behaviour is classical, knowing that, via Gibbs Duhem's relationship, activity coefficients of ionic species are derived from water activity data and are much more sensitive to uncertainties and discrepancies between model and experimental determination. However the structure of the present model shows sufficient flexibility so that it is expected that these results can be improved if adequate parameterization of ions size properties is done.

In this context, a new prediction method of the maximum value of the hydration number (that is the key parameter for taking into account the solvation of the ion by water molecules) has been introduced. For this purpose, the hydration number is defined as the number of water molecules forming a stoichiometric complex with the studied ion. Four complementary procedures (three procedures at the gas phase and the last one at infinite dilution in water) have been described and used to predict the corresponding hydration number values. In order to perform COSMO calculations the COSMO-radii had to be provided. Because this is not the case in the literature regarding the almost of the studied ionic elements, a strategy of estimation of radii has also been introduced in Chapter III. This strategy starts with the gas phase calculations where DFT results (mainly energies and optimized geometries) are independent of ion radii. Thus, by analyzing the calculated mean distances between the oxygen atom of one water molecule and a given ion (inside the cluster), it becomes possible to choose a set of radii that are in agreement with these values. Importantly, the values of the distance ion-water ($d_{ion\text{-}water}$) calculated in this study are consistent with the sum $R_{ion}+R_{water}$ given by Marcus (1988, 1997 and 2012) where R_{water} = 1.38 nm. This justifies the choice of the ionic radii values given by Marcus as reference values for calculating COSMO-radii of cations and applying the scaling factor of 1.17 times R_{ion}. This study of hydration processes has also demonstrated that the DFT/QM level used in COSMO-RS is enough accurate to determine the hydration numbers of ions.

In order to enlarge the domain of validity of the COSMO-RS-PDHS model, it will be also necessary to use a variable hydration of ionic species. It has been demonstrated that this can be performed if one has an accurate treatment of the different chemical equilibria between the different hydrated clusters formed by water molecules and a given ion. To do so, it is necessary to compute the equilibrium constant of each hydration reaction (Ben Gaïda 2007). The latter can be calculated in reference conditions, from the formation properties of the species in solution. However, there are not enough data in the literature regarding the formation properties of ionic clusters. To overcome this problem, it is necessary to develop a predictive model for predicting formation properties (ideally at infinite dilution in water). This new prediction method must be validated on a set of molecules with well-known formation properties before being used as a predictive tool.

In this context, the use of the quantum calculated input data of the COSMO-RS that can already be used to predict pK_a values (which are strongly linked to chemical potential of formation data) appears as an interesting starting point, as discussed in chapter IV.

Chapter IV:Prediction of formation properties, pK_a and redox potentials

IV.1. Introduction

In chapter I, it has been demonstrated that the formation properties, especially the chemical potential of formation μ_i^0, can be used to compute several equilibrium physico-chemical properties such as the dissociation constant (pK_a) and the standard redox potential E^0.

Regarding the prediction of pK_a, several existing prediction tools have been developed (Harding et al. 2009, Manchester et al. 2010, Lee and Crippen 2009, Perrin 1965 and 1972, Szegezdi and Csizmadia 2004 and 2007, Klamt et al. 2003, Eckert et al. 2010). Because the formation properties of the ionized and un-ionized atoms/molecules are not always available in databases, almost of these tools are based on empirically calculated physico-chemical parameters that are obtained from experimental data using specific regression equations (*e.g.* the ChemAxon method, or the ACD/Labs method), or using a linear free energy relationship (LFER) from quantum calculated free energies of dissociation (*e.g.* the COSMO-RS method). Since in the COSMO-RS algorithm, the LFER parameters have been determined on a dataset including only 64 organic and inorganic acids (including any peptides, Klamt et al. 2003), this work concerns the test of the predictive power of this model in pK_a prediction for a set of molecules including amino acids and several peptides of interest in foods and biological systems. To do so, a comparison between the pK_a values predicted using respectively COSMO-RS, ChemAxon and ACD/Labs methods, is performed and discussed in section IV-1.

There are few studies regarding the prediction of the standard-redox potential E^0 in existing models which are generally based on the use of the equation introduced by Hammett (1937) and its extended forms (Hansch et al. 1991 and references therein) that have been established as the most general and simplest approach to describe structure-property relations. However the Hammett methodology has been criticized by theoreticians due to its empirical basis even if the Hammett constants can be easily obtained from the study of the ionization of organic acids in solution (Hoge and Bader 2007).

To overcome this problem, it is necessary to consistently predict the formation properties in the appropriate reference state (a useful one is the infinite dilution in water reference state). This practical choice was adopted within this study to determine the formation properties in solution, especially $\mu_i^{0,aq,\infty}$ which is directly linked to E^0. As discussed in section IV-2, such data can be provided from quantum calculations if a rigorous definition of the reference states is established, and when consistent set are available from converting data obtained in one reference state to the other one. Indeed, because quantum calculations are performed

generally in gas phase, it is necessary to deduce the chemical potential values that would be determined at infinite dilution in water from those predicted or collected in the gas phase. For this purpose, it is necessary to develop a new predictive model that can predict the gas phase formation properties in order to complete the existing databases. Then, the next modelling task consists to provide accurate data to convert gas phase data to infinite dilution properties either by using established thermodynamic relations (Ould Moulaye 1998), or by using the solvation properties as defined by Ben Naim (1987 and 2001). Because the COSMO-RS model belongs to the family of solvation models, it can be used to compute the Gibbs free energy required in the solvation process (*i.e.* the transport of a given molecule from a fixed position in gas phase to the same position at infinite dilution in water). Finally, the combination of the gas phase chemical potential of formation with the solvation free energies of several compounds appearing in the redox mechanism of polyphenols are predicted to deduce the corresponding $\mu_i^{0,aq,\infty}$, then the standard-redox potential E^0.

The COSMO-RS method can also be used as a qualitative understanding tool of phenomena appearing during food processes, especially in the context of the "Na-" project that is focused on the reduction of salt ratio in meat products. In this context, an approximate composition of food is used to simulate its corresponding σ-profile that will be used to explain several phenomena observed experimentally in meat processes, as discussed in section IV-3.

IV.2. Determination of the pK_a of several compounds of interest in food systems

As presented in chapter I, the pK_a of a compound determines the amount of protonated and deprotonated species at a specific *pH*. It is an important property to understand the chemical, food and biological processes. Such systems contain mainly water and a wide variety of compounds that can be charged or neutral. For instance, bovine or pork meats generally contain water (70 to 85%), protein, fat (acids) and carbohydrates. The proteins from meat products are 91-100% digestible. Similarly, bovine milk contains mainly water (902 g/L), lactose (49 g/L), proteins (32 g/L), lipids (38 g/L) and many salts (9 g/L) (Gros and Dussap 2003). One of the most important reasons to understand amino acid structure and properties is to be able to understand protein structure and properties.

Thus, amino acids that make up 75% of the human body are vital to every part of human function. Amino acids play central roles both as building blocks of proteins and as

intermediates in metabolism. The twenty amino acids that are found within proteins convey a vast array of chemical versatility. The amino acid content, and the sequence of those amino acids, of a specific protein, is determined by the sequence of the bases in the gene that encodes that protein. The chemical properties of the amino acids of proteins determine the biological activity of the protein. Proteins not only catalyze all (or most) of the reactions in living cells, they control virtually all the cellular process. In addition, proteins contain within their amino acid sequences the necessary information to determine how that protein will fold into a three dimensional structure, and the stability of the resulting structure. The field of protein folding and stability has been a critically important area of research for years, and remains today one of the great unsolved mysteries. It is, however, being actively investigated, and progress is being made every day (White and Wimley 1999).

Humans can produce ten amino acids, namely alanine, asparagine, aspartic acid, cysteine, glutamic acid, glutamine, glycine, proline, serine and tyrosine. The others must be supplied by the food. Failure to obtain enough of even one of the ten essential amino acids, results in the degradation of the body's proteins (muscle and so forth) to obtain the one amino acid that is needed. Unlike fat and starch, the human body does not store excess amino acids for a later use, the amino acids must be in the food every day. For instance, tyrosine is produced from phenylalanine, so if the diet is deficient in phenylalanine, tyrosine will be required as well. The essential amino acids are arginine (required for the young, but not for adults), histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine (Belitz et al. 2009).

Therefore, amino acids appear as biologically important organic compounds that are structurally made from amine (-NH2) and carboxylic acid (-COOH) functional groups, along with a side-chain (-R) specific to each amino acid, as illustrated in Figure IV-1 *:* . The key elements of an amino acid are carbon, hydrogen, oxygen, and nitrogen, though other elements are found in the side-chains of certain amino acids (Belitz et al. 2009).

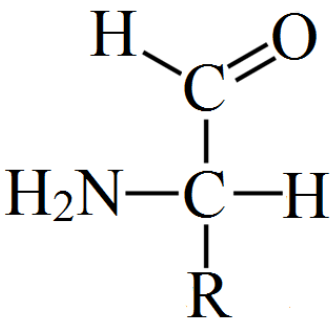


Figure IV-1 : The generic structure of an alpha amino acid in its un-ionized form.

Amino acids are the structural units (monomers) that make up proteins. They join together to form short polymer chains called peptides or longer chains called either polypeptides or proteins. These polymers are linear and un-branched, with each amino acid within the chain attached to two neighbouring amino acids (Belitz et al. 2009).

In this context, it appears essential to get a valuable idea about the predictive power and the performance of the COSMO-RS method in pK_a predictions, especially for amino acids and peptides. In the following study (Touré et *al.* 2013), the pK_a values predicted using this method are compared to those obtained by other semi-empirical tools like ACD/Labs and ChemAxon methods.

Dossier

This paper is a part of the hereunder thematic dossier published in OGST Journal, Vol. 68, No. 2, pp. 187-396 and available online here

Cet article fait partie du dossier thématique ci-dessous publié dans la revue OGST, Vol. 68, n°2, pp. 187-396 et téléchargeable ici

DOSSIER Edited by/Sous la direction de : **Jean-Charles de Hemptinne**

InMoTher 2012: Industrial Use of Molecular Thermodynamics
InMoTher 2012 : Application industrielle de la thermodynamique moléculaire

Oil & Gas Science and Technology – Rev. IFP Energies nouvelles, Vol. 68 (2013), No. 2, pp. 187-396

Oil & Gas Science and Technology – Rev. IFP Energies nouvelles, Vol. 68 (2013), No. 2, pp. 281-297

DOI: 10.2516/ogst/2012094

Dossier

InMoTher 2012 - Industrial Use of Molecular Thermodynamics
InMoTher 2012 - Application industrielle de la thermodynamique moléculaire

Comparison of Predicted pK_a Values for Some Amino-Acids, Dipeptides and Tripeptides, Using COSMO-RS, ChemAxon and ACD/Labs Methods

O. Toure*, C.-G. Dussap and A. Lebert

Institut Pascal (Axe GePEB), Université Blaise Pascal, Polytech'Clermont-Ferrand,
24 avenue des Landais, BP 206, 63174 Aubière Cedex - France
e-mail: oumar.toure@polytech.univ-bpclermont.fr - claude-gilles.dussap@polytech.univ-bpclermont.fr - andre.lebert@univ-bpclermont.fr

* Corresponding author

Résumé — **Comparaison des valeurs de pK_a de quelques acides aminés, dipeptides et tripeptides, prédites en utilisant les méthodes COSMO-RS, ChemAxon et ACD/Labs** — Les valeurs de constantes d'acidité (pK_a) jouent un rôle très important, en particulier dans l'industrie alimentaire. Les propriétés chimiques des molécules dépendent significativement de leurs états d'ionisation. La plupart des molécules sont capables de gagner et/ou perdre un proton dans les solutions aqueuses. Ce transfert de proton apparaît la plupart du temps entre l'eau et un atome ionisable de la molécule organique. La réponse de la molécule à la protonation ou à la déprotonation dépend significativement du site concerné par le transfert de proton. La distribution partielle des charges dans la molécule varie également en fonction des sites actifs pour la protonation du couple acide/base. Par conséquent on peut l'utiliser pour déterminer le pK_a d'une molécule.

Dans un premier temps, nous avons utilisé la méthode COSMO-RS, une combinaison du modèle de solvatation diélectrique (COSMO) et d'un traitement de thermodynamique statistique pour des solvants plus réels (RS), pour prédire les constantes de dissociation de 50 molécules environ (des acides aminés, des dipeptides et des tripeptides). Les résultats de pK_a obtenus ont été comparés aux valeurs expérimentales, ainsi qu'aux valeurs de pK_a prédites par deux autres méthodes. Nous avons utilisé respectivement la méthode ChemAxon, utilisant un programme basé sur le calcul des charges partielles des atomes d'une molécule, et la méthode ACD/Labs qui permet de déterminer des valeurs de pK_a pour chaque centre de dissociation en considérant que le reste de la molécule est neutre, en utilisant une base de données internes contenant des structures chimiques ainsi que leurs valeurs expérimentales de pK_a.

L'écart-type moyen des valeurs prédites vaut respectivement 0,596 pour la méthode COSMO-RS, 0,445 pour la méthode ChemAxon et 0,490 pour la méthode ACD/Labs. Au vu de ces résultats, la méthode COSMO-RS apparaît comme une méthode prometteuse pour prédire les valeurs de pK_a de molécules d'intérêt dans l'industrie alimentaire pour lesquelles peu de données de pK_a sont disponibles comme les peptides, d'autant plus que les méthodes ACD/Labs et ChemAxon ont été paramétrées en utilisant un grand nombre de données expérimentales (incluant certaines des molécules étudiées dans cet article) alors que la méthode COSMO-RS a été utilisée d'un point de vue purement prédictif.

L'objectif final de cette étude est d'utiliser ces valeurs de pK_a dans un modèle thermodynamique prédictif pour des produits d'intérêt dans l'industrie alimentaire. Pour ce faire, les effets de

certains facteurs (comme le traitement des conformations dans les calculs COSMO-RS, l'influence de la force ionique) pouvant influencer la comparaison entre les données expérimentales et les données prédites, seront discutés.

Abstract — Comparison of Predicted pK_a Values for Some Amino-Acids, Dipeptides and Tripeptides, Using COSMO-RS, ChemAxon and ACD/Labs Methods — *Liquid-phase pK_a values play a key role in food science. Chemical properties of molecules depend largely on whether they are ionized or not. Most organic molecules are capable of gaining and/or losing a proton in aqueous solutions. Proton transfer most frequently occurs between water and any ionizable atom of the organic molecule. The molecule's response to protonation or deprotonation depends significantly on the site that was disturbed by proton transfer. Partial charge distribution in the molecule also varies with protonation of the acid/base active sites. Then it can be used to determine the pK_a of a molecule.*
First, we use the COSMO-RS method, a combination of the quantum chemical dielectric continuum solvation model COSMO with a statistical thermodynamics treatment for more Realistic Solvation (RS) simulations, for the direct prediction of pK_a constants of about 50 molecules (amino-acids, dipeptides and tripeptides). Then, we compare our results with experimental data and the pK_a values predicted using two other methods. We used respectively the ChemAxon method using a program based on the calculation of partial charge of atoms in the molecule and the ACD/Labs method that enables to calculate single pK_a values for all possible dissociation centers when the rest of the molecule is considered neutral, using an internal database containing chemical structures and their experimental pK_a values.
The averaged Root Mean Square Error (RMSE) of the predicted pK_a values for each method compared to experimental results were respectively 0.596 for COSMO-RS, 0.445 for ChemAxon and 0.490 for ACD/Labs. While ACD/Labs and ChemAxon are parameterized using a large set of experimental data (including several of the studied molecules), the COSMO-RS method was used in a fully predictive way. Regarding these results, COSMO-RS appears as a promising method to predict the pK_a values of molecules of interest in food science with scarce available pK_a values such as peptides. The final goal of this study is to use the pK_a values in a predictive thermodynamics model for products of interest in food industry. For this purpose, the effects of several factors (like conformations set treatment in COSMO-RS calculations, ionic strength effect) that can affect the comparison between observed and predicted pK_a data are discussed.

INTRODUCTION

Foods and biochemical media are generally treated as aqueous mixtures that can be very complex, containing mainly water and other varieties of components that can:

- have different molecular sizes (organic acids and minerals, amino-acids, peptides, proteins, etc.) at the temperature (T) and pressure (p) of the system;
- be liquid (*e.g.* alcohols, polyols), solid (*e.g.* sugars, salts) or gaseous (*e.g.* aromatic volatile compounds) at T, p;
- be charged (*e.g.* ions, carboxyl radicals, amines) or neutral (*e.g.* sugars, polyholosides).

A large number of the molecules of interest in food science and biochemistry contains acidic and or basic groups which govern many of their chemical, physical and biological properties. So, most of these organic molecules are capable of gaining and/or losing a proton in aqueous solutions [1].

As defined by Brönsted, an acid is 'a species having a tendency to lose a proton' while a base is 'a species having a tendency to add on a proton'. Hence for every acid, AH, there is a conjugated base, A^- and for every base, B, there is a conjugated acid, BH^+.

If AH (or BH^+) is a strong acid, *i.e.* it has a great tendency to lose protons, it follows that its conjugated base A^- (or B), is a weak base, *i.e.* has only a small tendency to accept protons.

In aqueous solution, acids react with water acting as a base:

$$\mathrm{AH\ (or\ BH^+) + H_2O\ (base) \rightleftarrows H_3O^+ + A^-\ (or\ B)} \quad (1)$$

and bases react with water acting as an acid:

$$\mathrm{A^-\ (or\ B) + H_2O\ (acid) \rightleftarrows AH\ (or\ BH^+) + OH^-} \quad (2)$$

In dilute aqueous solution where almost all measurements are made, water is thus the solvent and its activity

is taken as unity [1, 2]. Then, the acidic dissociation (or ionization) constant K_a is calculated with the following equation:

$$K_a = \prod_i a_i^{\nu_i} = \frac{(H^+ \text{ or } H_3O^+)(B)}{(A)} \quad (3)$$

where, parentheses denote activities, B and A represent respectively base and acid species. This equation can be written in the form:

$$pK_a = pH + \log\left(\frac{(A)}{(B)}\right) \quad (4)$$

where pK_a is the negative logarithm of K_a, and is equal to the pH at which the activities of A and B are equal [2].

The molecule's response to protonation or deprotonation depends significantly on the site that was disturbed by proton transfer. Partial charge distribution in the molecule also varies with protonation of the acid/base active sites. Since the partial charge distribution is very sensitive to the protonation-deprotonation process (both near and far from the disturbed site) [3, 4], it can be used to determine the pK_a of a molecule, which is a measure of the tendency of a molecule or ion to keep a proton, at its ionization center(s).

The more likely ionization occurs, the more likely a species will be taken up into aqueous solution, because water is a very polar solvent [2] (macroscopic dielectric constant $\varepsilon_r = 80$). If a molecule does not readily ionize, then it will tend to stay in a non-polar solvent such as cyclohexane ($\varepsilon_r = 2$) or octanol ($\varepsilon_r = 10$).

Due to the fact that proton transfer most frequently occurs between water and any ionizable atom of the organic molecule, dissociation constants (pK_a) values play a key role in food science and other process industries.

Indeed, the pK_a of a compound is an important property [5] in both life sciences and chemistry since the propensity of a compound to donate or accept a proton is fundamental for the understanding of chemical and biological processes. In biological terms, pK_a is thus an important concept in determining whether a molecule will be taken up by aqueous tissue components or the lipid membranes. It is also closely related to the concepts of pH (acidity of solution) and log(P) (the partition coefficient between immiscible liquids) [2]. As the pK_a value of a molecule also determines the amount of protonated and deprotonated species at a specific pH, for example at physiological pH, knowing the pK_a of a molecule gives insight into pharmacokinetic properties. The latter includes the rate at which a molecule will diffuse across membranes and other physiological barriers, such as the blood brain barrier. More often, phospholipid membranes easily absorb neutral molecules, while ionized molecules tend to remain in the plasma or the gut before being excreted. Many biological systems also use proton-transfer reactions to communicate between the intra- and extracellular media, and the rate of the proton-transfer reaction depends, in part, on the pK_a values of the species involved.

In another area, microorganisms are inhibited by the non-dissociated forms of weak organic acids. The knowledge of pK_a values is then of great importance in microbiology previsional models [6].

Furthermore, in efforts to take greater control over the 'design-make-test' cycle typically implemented in modern drug discovery efforts [7], considerable attention has been given to providing accurate pK_a measurements with good throughput. Increasing attention given to pK_a during drug discovery is evidenced by the development of high-throughput methods for rapid pK_a determination. While experimental methods continue to become more sophisticated and refined, it is often desirable to predict dissociation constants for "virtual compounds", *i.e.*, those that have been described by a compound designer (chemist or modeler) but that have not yet been synthesized [7].

Many different algorithms [5, 7-10] for predicting pK_a values have been developed, and a few have been packaged into commercial computer software applications.

The main objective of the present study is to look for one (or more) reliable pK_a prediction method(s) that can enable to determine the dissociation constants of some components of interest in food sciences. For this purpose, we used a training set of molecules (amino-acids, dipeptides and tripeptides) having known experimental pK_a values. Then, we used 3 different predictions methods namely ChemAxon [3, 4] (Marvin version 5.4.1.1), ACD/Labs [11] (version 10.01, Release 10.00) and COSMO-RS [12, 13] (COSMOtherm [8, 9], version C2.1, Release 01.11) to predict the pK_a values of these molecules, and compare each predicted value to the corresponding experimental value.

1 MATERIALS AND METHODS

1.1 Experimental Data

The experimental pK_a values (at room temperature and atmospheric pressure) used in this study are taken from "*Dissociation constants of organic bases in aqueous solutions*", Perrin [14] (1965) and "*Dissociation constants of organic bases in aqueous solutions – Supplement*", Perrin [15] (1972). In these books, the experimental information related to each pK_a values is given.

1.2 The ChemAxon Method

The ChemAxon method [3, 4, 16, 17] is based on empirically calculated physico-chemical parameters (mainly partial charges) that are obtained from ionization site-specific regression equations. For a given molecule, it uses three types of calculated parameters (intramolecular interactions, partial charges and polarizabilities) to determine the micro ionization constants pK_a of monoprotic molecules [16]:

$$pK_a = a \times Q + b \times P + c \times S + d \qquad (5)$$

where, Q and P denote respectively the partial charge and the polarizability increments, S is the sum of the structures specific (steric strain or/and hydrogen bond) increments; a, b, c and d are regression coefficients specific to the ionization site. All of these pK_a increments are calculated from ionization-site specific regression equations.

Then, the ratio of microspecies is calculated to assign calculated pK_a values to the atoms of the submitted molecule. Finally, macro pK_a values are obtained from the theoretical relations that hold between macro-micro pK_a values. When a molecule contains more than one ionizable atom (*i.e.* multiprotic compound), one has to distinguish between micro and macro acidic dissociation constants. The micro acidic dissociation constant is obtained from the equilibrium concentration of the conjugated acid-base pairs. The macro acidic dissociation constant is obtained from the global mass and charge conservation law. The pK_a of the active groups at a given pH can be calculated according to this relation [16, 17]:

$$K_{a,i} = \frac{\sum_j c_j^i}{\sum_k c_k^{i-1}} [H^+] \qquad (6)$$

where, $[H^+]$ denotes the proton concentration of the aqueous solution, c_j^i is the concentration of the j-th microspecies that released i protons from the fully protonated molecule, c_k^{i-1} is the concentration of the k-th microspecies that released (i-1) protons from the fully protonated molecule. Ratio of c_j^i and c_k^{i-1} also called microspecies distributions are calculated from the micro ionization constants.

As an illustration case, the pK_a calculation for threonine (pT) is described below. This molecule has 4 different ionic forms shown below (*Fig. 1*).

In the present study, only the two first pK_a values for this molecule are calculated (the corresponding experimental dissociation constants of threonine are $pK_{a1} = 2.09$; $pK_{a2} = 8.81$).

By plotting the titration curves *i.e.* the evolutions of the ratio of the ionized and neutral forms *versus* the *pH* (*Fig. 2*), one can identify the ionic species which are present in the mixture at each *pH* values, and because $pK_a = pH$ where the ratio of two different forms are equal, it is possible to identify all the pK_a values predicted in the range of *pH* specified ($0 \leq pH \leq 14$ for this study).

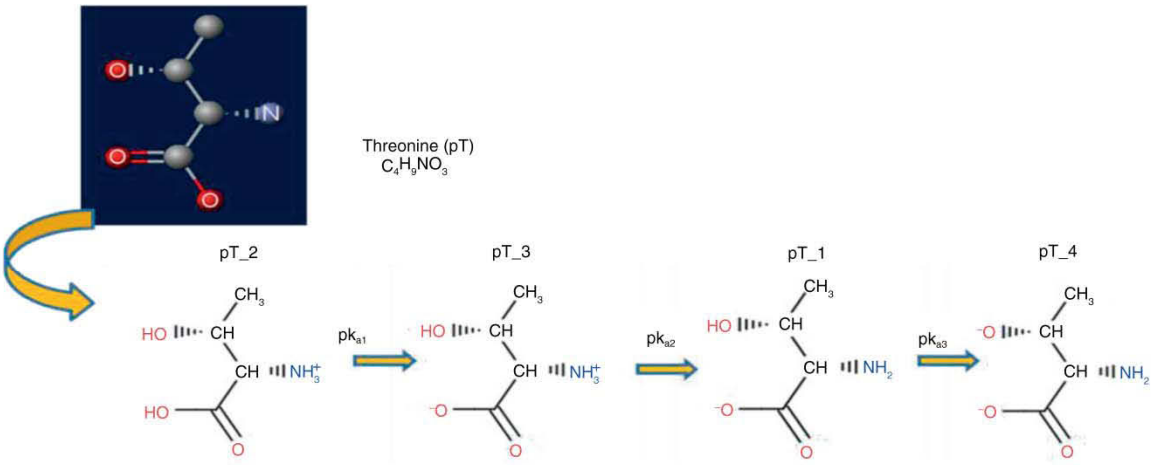


Figure 1

Neutral and ionic forms of threonine (pT). The names of the cationic form (pT_2), the neutral form (pT_3), the anion (pT_1) and the di-anion (pT_4) are evidenced. The two experimental values of this molecule calculated in the present study are $pK_{a1} = 2.09$ and $pK_{a2} = 8.81$.

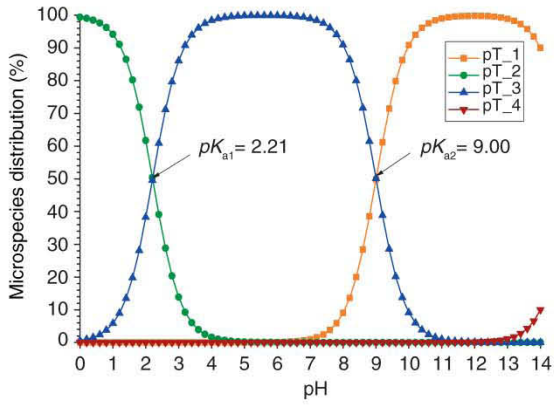


Figure 2

Illustration of the pK_a values of threonine (pT) predicted using the ChemAxon method. This graph shows the evolution of microspecies distribution *vs* pH values, and enlightens the ChemAxon predicted pK_a values that are respectively pK_{a1} = 2.21 and pK_{a2} = 9.00.

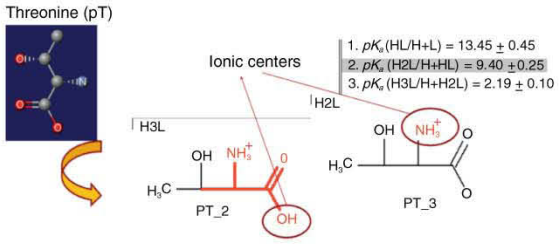


Figure 3

Illustration of the pK_a values of threonine (pT) predicted using the ACD/Labs method. This image shows the different ionic centers in the structures of pT_2 (H3L form) and pT_3 (H2L form) for which ACD predicted pK_a values are respectively pK_{a1} = 2.19 and pK_{a2} = 9.40.

Regarding Figure 2, one can see that the pK_a value pK_{a1} = 2.21 corresponds to the intercept between the microspecies distribution plots of ionic species 2 (pT_2) and 3 (pT_3). Likewise, the pK_a value pK_{a2} = 9.00 corresponds to the intercept between the microspecies distribution plots of ionic species 3 (pT_3) and 1 (pT_1). These values are in very good agreement with the experimental values (respectively pK_{a1} = 2.09 and pK_{a2} = 8.81).

1.3 The ACD/Labs Method

The ACD/Labs pK_a prediction method [11] enables to calculate single pK_a values for all possible dissociation centers when the rest of the molecule is considered neutral, using an internal database containing chemical structures and experimental data.

The algorithm of this calculation mimics the experimental order of protonation of the drawn molecule and determines the pK_a values which can be experimentally measured in aqueous solution.

This model is based on Linear Free Energy Relationships (LFER), applying the Hammett equation [2, 7]:

$$pK_a = pK_a^0 + \Delta(pK_a) \qquad (7)$$

where pK_a^0 is the ionization constant for the parent molecule, and $\Delta(pK_a) = \sum \Delta(pK_a)^i$ is the sum over the influence of all other functional groups on the respective pK_a of the $\Delta(pK_a)^i$ values of each reaction center;

$$\Delta(pK_a)^i = -\sum \rho^j \sigma^j \qquad (8)$$

where ρ^j is the constant for a particular class j of molecules, and σ^j is the electronic effect of the jth substituent on the ionization constant of the parent molecule.

For this purpose, every ionizable group is characterized by several Hammet-type equations that have been parameterized to cover the most popular ionizable functional groups. The ACD/Labs internal training set contains more than 2 000 derived experimental electronic constants (σ^j). When the required substituent constant is not available from the experimental database; the ACD/Labs method uses another algorithm to describe electronic effect transmissions through the molecular system. The flaw in this method is that the parent molecules inherently carry the majority of the chemical information and without training on a particular parent, predictions for such compounds are impossible. That's why an Internal Reaction Centers Database is used for pK_a prediction using this method. The ACD/Labs internal training set contains more than 31 000 experimental values for 15 932 structures. These data are taken from various articles published in peer-reviewed scientific journals [18].

As an illustration case, the pK_a calculation for threonine (pT) is described in Figure 3.

The ACD/Labs predicted pK_a values corresponding to the two ionic centers studied earlier (in *Sect. 1.4*, using the ChemAxon method) are respectively pK_{a1} = 2.19 and pK_{a2} = 9.40 (when the corresponding experimental values were respectively pK_{a1} = 2.09 and pK_{a2} = 8.81).

1.4 The COSMO-RS Method

COSMO-RS [13, 19, 20] is a predictive method for thermodynamic equilibrium of fluids and liquid mixtures that is a combination of the quantum chemical dielectric continuum solvation model COSMO [12] (acronym of Conductor-like Screening Model) with a statistical thermodynamics treatment for more Realistic Solvation (RS) simulations.

The equilibrium thermodynamic properties derived from the COSMO-RS theory are computed in COSMOtherm, a command line/file driven program which can be run directly from a UNIX or DOS shell. In the present study, we used the C21_0111 version of COSMOthermX [8], a Graphical User Interface to the COSMOtherm [9] command line program.

In this software, the pK_a of a solute j can be estimated from the Linear Free Energy Relationship (LFER) [21, 22]:

$$pK_a = A\left(\frac{\Delta G^j_{neutral} - \Delta G^j_{ion}}{RT \ln 10}\right) + B \qquad (9)$$

where $\Delta G^j_{neutral}$ and ΔG^j_{ion} are respectively the free energies of the neutral and ionic compounds, in the solvent (water in our case) at infinite dilution; A and B denote LFER parameters that were determined for example by correlating calculated free energies of dissociation with the experimental aqueous pK_a for a set of 64 organic and inorganic acids (not including any peptide) [21].

Equation (9) should also be rewritten as:

$$pK_a = c_0 + c_1\left(\Delta G^j_{neutral} - \Delta G^j_{ion}\right) \qquad (10)$$

Thus to obtain a pK_a value it is necessary to do quantum COSMO calculations of a molecule in its neutral and in its ionic state. All the pK_a calculations (of the present study) were done at the large TZVP basis set, in the following denoted BP-TZVP, using a full Turbomole BP-RI-DFT COSMO optimization of the molecular structure.

The LFER parameters $c_0 = B$ and $c_1 = \frac{A}{RT \ln 10}$ used to predict the pK_a values of interest in this study were read from the COSMOtherm parameter file. At ambient temperature, their values are respectively $c_0 = -120.29804$ and $c_1 = 0.10927$ mol/kJ.

pK_a prediction by COSMOtherm is not restricted to aqueous acid pK_a. However, both aqueous base pK_a prediction and pK_a in non-aqueous solvents require reparametrization of the pK_a LFER parameters. Likewise, to compute pK_a at non-ambient temperature, a reparametrization of the LFER parameters is required.

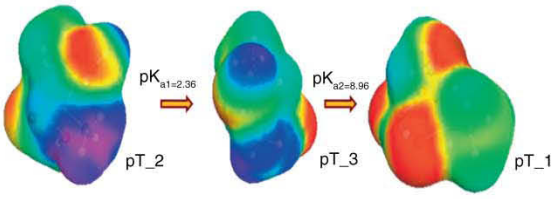


Figure 4

Illustration of the pK_a values of threonine (pT) predicted using the COSMO-RS method. This image shows the different ionic structures involved in the dissociation reactions for which COSMO-RS predicted pK_a values are respectively $pK_{a1} = 2.36$ and $pK_{a2} = 8.96$.

TABLE 1

Summary of the observed (exp.) and predicted pK_a values for threonine

Method	pK_{a1}	pK_{a2}
ChemAxon	2.21	9.00
ACD/Labs	2.19	9.40
COSMO-RS	2.36	8.96
Exp.	2.09	8.81

As an illustration case the pK_a calculation for threonine (pT) at ambient temperature is described in Figure 4.

Regarding this image, we can see that the predicted pK_a values of threonine, using the COSMO-RS method, are respectively $pK_{a1} = 2.36$ and $pK_{a2} = 8.96$ (while the experimental values were respectively $pK_{a1} = 2.09$ and $pK_{a2} = 8.81$). All the predicted pK_a values of the illustrative case are shown in Table 1.

One has to note that all the 3 methods tested (ChemAxon, ACD/Labs, and COSMO-RS) enable the user to input (in their software package) structures in SMILES format [23], and optionally to incorporate "local" data in order to bias predictions. However these optional facilities were not studied, since we are looking for a predictive tool for complex structures.

1.5 Data Analysis

To compare the predicted values *versus* observed ones, we first perform a graphical analysis of pK_a results. Then statistical tests to compare the three prediction methods (ChemAxon, ACD/Labs, and COSMO-RS) are performed.

1.5.1 Indices of Performance of Models

The bias factor (B_f) and the accuracy factor (A_f) are two indices of performance that enable to compare the goodness-of-fit of competing models [24, 25].

The bias factor provides an indication of the average deviation between the model predictions and the average deviation between the model predictions and observed results, it is defined as:

$$B_f = 10^{\frac{\sum \log\left(\frac{pK_a^{cal}}{pK_a^{exp}}\right)}{n}} \tag{11}$$

where pK_a^{cal} is the predicted pK_a value, pK_a^{exp} and is the experimental pK_a value and n is the number of observations. A bias factor of 1 indicates perfect agreement between observed and predicted pK_a values. Because over- and under-predictions may cancel out, the bias factor provides no indication of the range of the deviation between predictions and observations. A bias factor greater (resp. lower) than 1 indicates that the model predicts, on average, pK_a values higher (resp. lower) than experimental ones [24, 25].

The accuracy factor (A_f) seeks to provide an estimate of the average deviation between prediction and observation, and is defined as:

$$A_f = 10^{\frac{\sum \left|\log\left(\frac{pK_a^{cal}}{pK_a^{exp}}\right)\right|}{n}} \tag{12}$$

1.5.2 Statistical Analysis of Prediction Errors

We calculate the RMSE of the predicted pK_a values for each method:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (\Delta pK_a)^2}{n}} \tag{13}$$

Then, a normalization of the residuals (noticing that 95% of these points must be located between −2 and +2) is performed on the predicted values using the following equations:

$$\text{Normalized residual} = \frac{\Delta pK_a - \overline{\Delta pK_a}}{\sigma_{\Delta pK_a}} \tag{14}$$

where $\Delta pK_a = pK_a^{calc} - pK_a^{exp}$ is the error on predicted value; $\overline{\Delta pK_a}$ and $\sigma_{\Delta pKa}$ are respectively the average and the standard deviation of the prediction errors on pK_a values; and are given by:

$$\begin{cases} \overline{\Delta pK_a} = \dfrac{\sum_{i=1}^{n} \Delta pK_a}{n} \\ \sigma_{\Delta pK_a} = \dfrac{\sum_{i=1}^{n} \left(\Delta pK_a - \overline{\Delta pK_a}\right)^2}{n} \end{cases} \tag{15}$$

To have a more general comparison of the errors on predicted pK_a, we perform several normality tests on error bars using the statistical tools [26-27] available in the R software [28]. For this purpose, we mainly compare the cumulative distribution functions, the Quantile *versus* Quantile (Q-Q) plots, and the Percentile *versus* Percentile (P-P) plots, of the prediction errors distributions of the competing methods. The CFD describes the probability of "hitting" a value x or less in a given distribution (a normal or Gaussian distribution in our test). The Q-Q plot represents the quantiles of the theoretical fitted distribution (x-axis) against the empirical quantiles of the sample data (y-axis). Likewise, for each value of the data set the P-P plot represents the cumulative density function of the fitted distribution (x-axis) against the empirical cumulative density function of the sample data (y-axis).

1.5.3 Factors that Can Affect the Comparison Between Predicted and Observed pK_a Values

The almost of experimental pK_a values [14, 15] used in the present study, were fitted to zero ionic strength unless otherwise indicated under "Remarks", in which cases, an ionic strength correction of the experimental values should be necessary to perform a better comparison to predicted values. Indeed, if the solutions were no more concentrated than 0.01 M, the corrections of experimental data would be small, and the author may choose to neglect them for his purposes [1]. However there are circumstances in which they must not be neglected because ionization constants change with dilution (although there is seldom a detectable change below 0.001 M) [1]. Since for experimental pK_a measurements methods like potentiometric titrations or spectrometric determinations, the corrections for diluter solutions involve the ionic strength, written as I and define as:

$$\mathrm{I} = \frac{1}{2}\sum_i \left(C_i \eta_i^2\right) \tag{16}$$

where, C_i is the molar concentration of an ion, and η_i is the charge of the ion.

The ionization constant yielded directly by potentiometric titration is appropriately denoted as K_a' (or K_a^M) because it is a mixed constant [1], partly thermodynamic and yet partly concentration-dependent. This mixed character arises from the fact that a pH set is calibrated in terms of hydrogen ion activity (not hydrogen ion concentration), whereas the ionic term is a concentration (not an activity). Thus,

$$K_a'(\text{or } K_a^M) = \frac{(\mathrm{H^+ \text{ or } H_3O^+})[\mathrm{B}]}{[\mathrm{A}]} \tag{17}$$

A relation between pK_a and pK'_a can be derived beginning from:

$$a_i = c_i \gamma_i \quad (18)$$

where γ_i is the activity coefficient (molar scale) of an ion of activity a_i and molar concentration c_i.

For an ion of charge η_i, the activity coefficient is given for dilute solutions by activity coefficients models [1, 2, 29]. In the present study, we use the correction proposed by Ould-Moulaye (in his PhD Thesis [29]) to take into account the influence of ionic strength on pK_a values, using a simplified Goldberg model [30], and assuming that the acidic form is not very concentrated ($\gamma^m_{AH} = 1$). It was demonstrated that (at 25°C):

$$pK_a = pK'_a + 0.51065 \sum_i \left(v_i \eta_i^2\right) \frac{\sqrt{I_m}}{1 + 1.6\sqrt{I_m}} \quad (19)$$

where v_i represents the stoechiometric coefficient of the species in dissociation reaction, $v_i > 0$ for products and $v_i < 0$ for reactants; I_m represents the ionic strength in molality scale.

Due to the fact that ChemAxon and ACD/Labs use 2D structures to predict pK_a values, while COSMO-RS performs a geometrical optimization of the 3D structure in its Quantum Chemistry (QC) calculations, it would be interesting to have a look on the influence of conformations set treatment in the COSMO-RS pK_a predictions. For this purpose, the influence of conformations treatment for 2 molecules namely histidine (pH) and histidylglutamic acid (pHE) in COSMO-RS calculations is studied. Sometimes, there are several different reported pK_a values for the same dissociation reaction (ideally, a reliable prediction method would be able to choose the right experimental value). So by using all these points with only one predicted value and 2 or more experimental values, one can influence the global RMSE of the prediction methods. This point is discussed later.

2 RESULTS AND DISCUSSION

2.1 Graphical Comparisons

The predicted pK_a values are plotted *versus* the experimental values in Figure 5. A linear fit of the predicted pK_a data enables to determine the coefficient of determination (r^2) of each competed model. As results, we got respectively $r^2 = 0.98$ for ACD/Labs, $r^2 = 0.98$ for ChemAxon and $r^2 = 0.96$ for COSMO-RS. However, it is very difficult to compare the three prediction methods from this graph, as there are lots of points for which one method gives better predicted pK_a values compared to the others. However, one can already see that the maximum error on predicted pK_a values is observed for the 60th data-point ($pK_a^{exp} = 10.98$) corresponding to one of the dissociation constants of hystidylglutamic acid (pHE). For this data-point, ACD/Labs and ChemAxon methods give erroneous prediction results ($pK_a = 12.86$ and 14.03 respectively) while COSMO-RS predicts a $pK_a = 11.44$ which is significantly closer to the experimental value. One justification of these differences on predicted pK_a should be the fact that the pK_a calculated for this structure (on a dissociation center which is a cyclic nitrogen atom) that was probably not used in the internal database of ACD/Labs and ChemAxon, while COSMO-RS overcome this problem when performing a geometrical optimization during the QC calculations preceding its pK_a prediction.

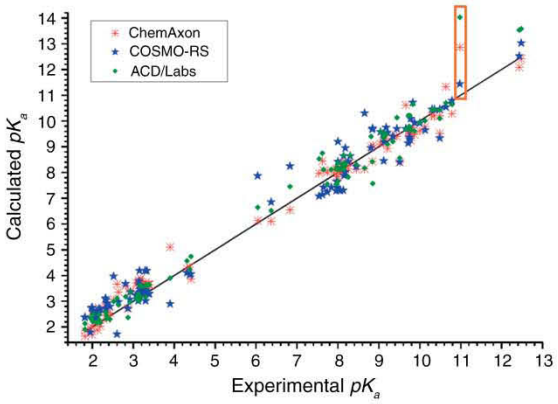


Figure 5

pK_a results for several amino-acids, dipeptides and tripeptides using the ChemAxon (red stars), the COSMO-RS (blue filled stars) and the ACD/Labs (green diamonds) methods. The line represents the equality between predicted values and experimental data and the orange box shows the predicted pK_a data corresponding to the maximum error (data-point number 60).

2.2 Bias Factor and Accuracy Factor

The calculated values of the bias and accuracy factors (using respectively *Eq. 11* and *Eq. 12*) are shown in Tables 2 and 3 for each family, and on the overall data. These values were very close to 1 that means each of the 3 studied methods performs a good prediction of the pK_a values.

Using the bias factor criterion, ACD/Labs and ChemAxon give slightly better pK_a predictions compared to COSMO-RS. For amino-acids the bias factor of COSMO-RS (1.09) is slightly greater than those of ChemAxon (1.02) and ADC/Labs (1.04), while for

TABLE 2
Values of the bias factor (B_f) for each of the 3 methods used for pK_a prediction

	B_f ChemAxon	B_f COSMO-RS	B_f ACD/Labs	Number of points
Amino-acids (AA)	1.02	1.09	1.04	45
Dipeptides	1.07	1.03	1.04	37
Tripeptides	1.04	1.01	1.05	25
Overall data	1.04	1.05	1.04	107

TABLE 3
Values of the accuracy factor (A_f) for each of the 3 methods used for pK_a prediction

	A_f ChemAxon	A_f COSMO-RS	A_f ACD/Labs	Number of points
Amino-acids (AA)	1.06	1.13	1.06	45
Dipeptides	1.11	1.11	1.06	37
Tripeptides	1.05	1.07	1.05	25
Overall data	1.07	1.11	1.06	107

dipeptides and tripeptides COSMO-RS has the smallest bias factor (resp. 1.03 and 1.01), followed by ACD/Labs (1.04 and 1.05) and ChemAxon (1.07 and 1.04). Using the accuracy factor criterion, ACD/Labs and ChemAxon give better pK_a predictions compared to COSMO-RS. For each family, the accuracy factor of COSMO-RS is slightly greater than those of ChemAxon and ADC/Labs. Since ACD/Labs and ChemAxon are parameterized using experimental data of molecules, the slightly higher values in the A_f and B_f values for COSMO-RS is quite normal as discussed later in Section 2.5.

2.3 Statistical Analysis of Prediction Errors

The RMSE of the overall predicted pK_a values for each method compared to experimental results were respectively 0.596 for COSMO-RS, 0.445 for ChemAxon and 0.490 for ACD/Labs (*Tab. 4*).

These RMSE results are in good agreement with literature values. Indeed, for complex molecular structures a RMSE of 0.50 is expected for ACD/Labs [11] (version 10). In a recent study, on 211 drug like-compounds, Manchester *et al.* (2010) [7] got a RMSE of 0.6 for ACD/Labs method (version 10), and 0.8 for the ChemAxon method (Marvin, version 5.2). Their results agree well with the RMSE values given in the present study since we used a more recent version of Marvin (version 5.4.1), that should explain the difference in RMSE.

Then one can plot Figure 6 which compares the normalized residuals (that should ideally be equal to zero) of the 3 prediction methods, using Equation (13). One has to note that the maximum normalized error is got for data-point number 60 (pHE) discussed earlier (6.27 for ACD/Labs; 4.10 for ChemAxon, and 0.54 for COSMO-RS). But, when analyzing this plot, it becomes also clearer that there is not a distinguishable difference between the predictions methods using this comparison criterion, since for data-point number 19 (pH) COSMO-RS has the largest normalized residual (2.90 *versus* −0.09 for ChemAxon and 0.91 for ACD/Labs), and for data-point number 9 (pD) ChemAxon has the largest normalized error (2.53 *versus* −1.97 for COSMO-RS and −0.40 for ACD/Labs). Furthermore, 6 points for COSMO-RS, 5 points for ChemAxon and 6 points for ACD/Labs are located outside the range of [−2, +2] for normalized residual values. That are in good agreement with the 95% points expected to be in the same range, since 107 data-points were used for each of the studied models. These examples illustrated that all the studied methods are undistinguishable in that each can sometimes give large errors. This statement is confirmed in another study [7], in the case of ACD/Labs and ChemAxon.

Figure 7 represents the comparison between the theoretical and observed cumulative distribution functions (CFD) [27] of the prediction errors for each pK_a prediction method. One can notice that the COSMO-RS method's sample CFD is the closest to its theoretical CFD, followed by ChemAxon, while the ACD/Labs method's

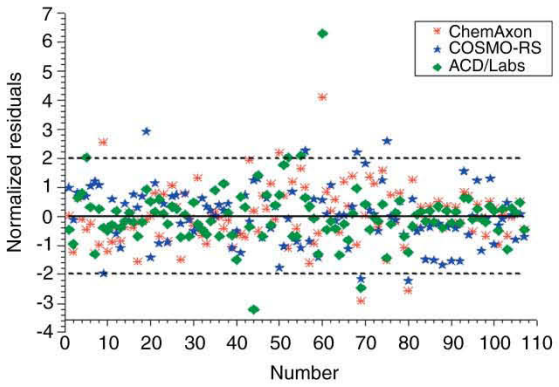


Figure 6

Normalized residuals of the pK_a predicted values for several amino-acids, dipeptides and tripeptides using the ChemAxon, the COSMO-RS and the ACD/Labs methods.

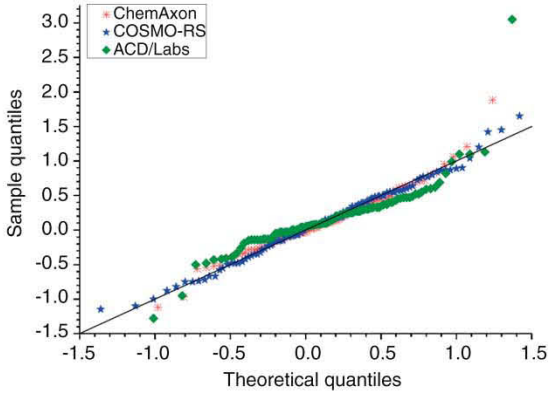


Figure 8

Quantile-Quantile (Q-Q) plots of errors distribution for each of the 3 pK_a prediction methods.

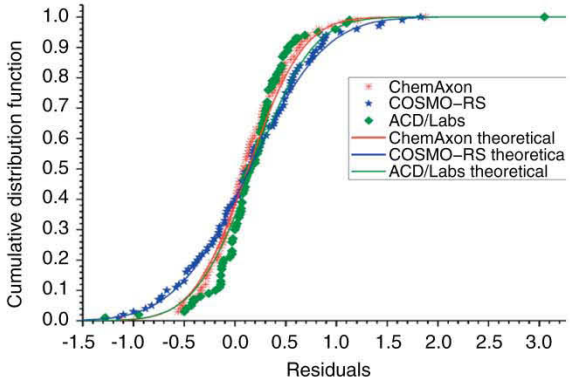


Figure 7

CFD of errors distribution for each of the 3 pK_a prediction methods.

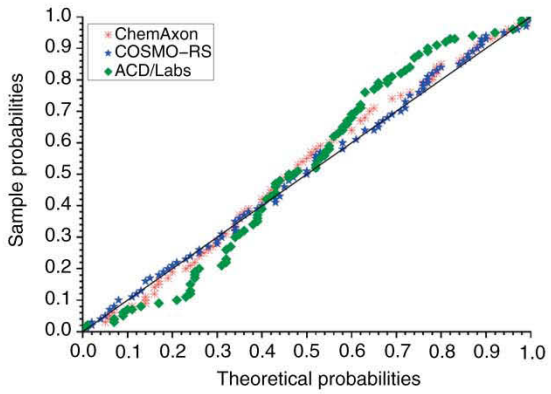


Figure 9

Percentile-Percentile (P-P) plots of errors distribution for each of the 3 pK_a prediction methods.

sample CFD is relatively different compared to its theoretical CFD.

Figures 8 and 9 show respectively the Q-Q and the P-P plots [27] of the prediction errors distribution for each method.

These graphical normality tests (CFD, Q-Q and P-P plots) show that COSMO-RS prediction errors have the closest distribution to the Gaussian distribution. The ChemAxon prediction errors have a slightly less normal distribution, while ACD/Labs method's errors distribution is significantly different from the normal distribution. Several normality tests (*Tab. 5*) confirmed these results.

For each of the studied pK_a prediction methods, several statistical tests on the distribution shown were performed to get the Cullen and Frey graph that enables to have an idea about the position of the observed distribution (red circle point) compared to different theoretical distributions (normal, log-normal, exponential, etc.) and the results of a bootstrap (performed on sample points by randomly taking off one or more points) done 100 times (hollow blue circles) (*Fig. 10-12*).

These confirmed that the nearest theoretical distribution should be the normal distribution, and that

TABLE 4

Values of the average RMSE (Root Mean Square Errors of the differences ΔpK_a between the predicted values and the experimental ones, $\Delta pK_a = pK_a^{calc} - pK_a^{exp}$) for each of the 3 methods used for pK_a prediction (ChemAxon, COSMO-RS and ACD/Labs)

	RMSE ChemAxon	RMSE COSMO-RS	RMSE ACD/Labs	Number of points
Amino-acids (AA)	0.361	0.577	0.407	45
Dipeptides	0.612	0.668	0.669	37
Tripeptides	0.239	0.510	0.259	25
Overall data	0.445 (r^2 = 0.98)	0.596 (r^2 = 0.96)	0.490 (r^2 = 0.98)	107

TABLE 5

Summary of the results of several normality tests on errors distribution for each pK_a prediction method

	Kolmogorov-Smirnov statistic	Cramer-von Mises statistic	Anderson-Darling statistic	Conclusion
ChemAxon	0.06307036 (Not rejected)	0.09646038 (Not rejected)	0.6444582 (Not rejected)	Gaussian distribution
COSMO-RS	0.03971827 (Not rejected)	0.02223085 (Not rejected)	0.2031585 (Not rejected)	Gaussian distribution
ACD/Labs	0.1329101 (Rejected)	0.5802089 (Rejected)	3.523414 (Rejected)	Non-Gaussian distribution

TABLE 6

Values of the average RMSE (Root Mean Square Errors of the differences ΔpK_a between the predicted values and the experimental ones: $\Delta pK_a = pK_a^{calc} - pK_a^{exp}$) for each of the 3 methods used for pK_a prediction (ChemAxon, COSMO-RS and ACD/Labs), after taking into account the influence of ionic strength. The values in parenthesis are those found in literature without any correction to get zero ionic strength pK_a values (see *Tab. 4*)

	RMSE ChemAxon	RMSE COSMO-RS	RMSE ACD/Labs	Number of points
Amino-acids	0.355 (0.361)	0.560 (0.577)	0.424 (0.407)	45
Dipeptides	0.708 (0.612)	0.691 (0.668)	0.830 (0.669)	37
Tripeptides	0.178 (0.239)	0.456 (0.510)	0.208 (0.259)	25
Overall data	0.484 (0.445)	0.588 (0.596)	0.569 (0.490)	107

COSMO-RS errors distribution are the closest to this theoretical distribution.

2.4 Analysis of Several Factors that Can Influence Predicted pK_a Values

2.4.1 Influence of Ionic Strength on pK_a Values

The influence of ionic strength on the experimental pK_a values used has been taken into account using (*Eq. 19*). This correction changes the values of the RMSE of the different prediction methods as shown in Table 6.

Table 6 shows that the correction of ionic strength influence on pK_a decreased the RMSE of the COSMO-RS pK_a values by 0.01 while the ChemAxon and ACD/Labs RMSE increased respectively by about 0.04 and 0.08. That should probably due to the fact that the *ab-initio* calculation performed in the COSMO-RS method enables it to reach a better treatment of electrostatic interactions when compared to other methods using purely empirical parameters. Indeed, the ionic strength influence should only affect the experimental pK_a value, not the performance of a given model. But because ACD/Labs and ChemAxon models are parameterized on a large set of experimental pK_a (probably including non zero ionic strength pK_a values), a small effect of ionic strength should bias their respective RMSE values as illustrated in Table 6.

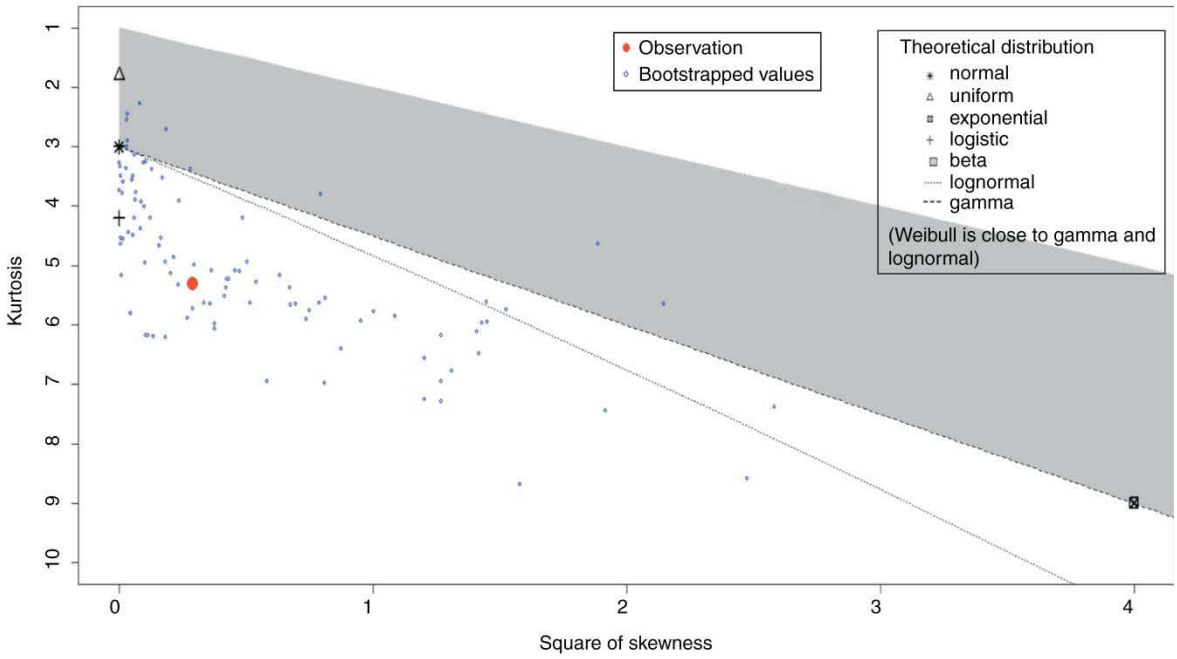


Figure 10

Cullen and Frey graph for the errors distribution of the ChemAxon method.

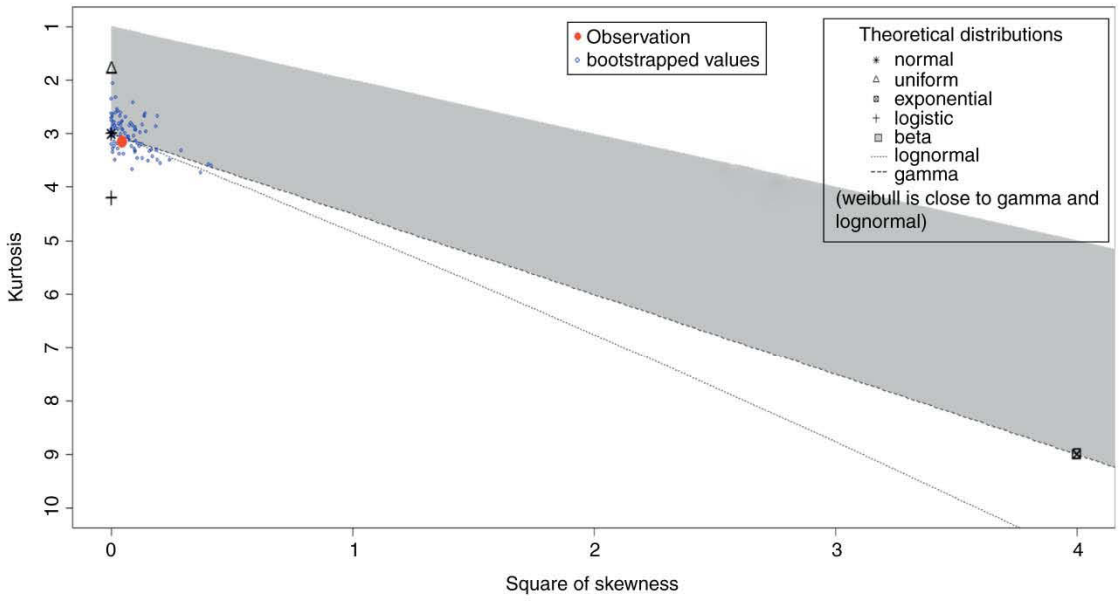


Figure 11

Cullen and Frey graph for the errors distribution of the COSMO-RS method.

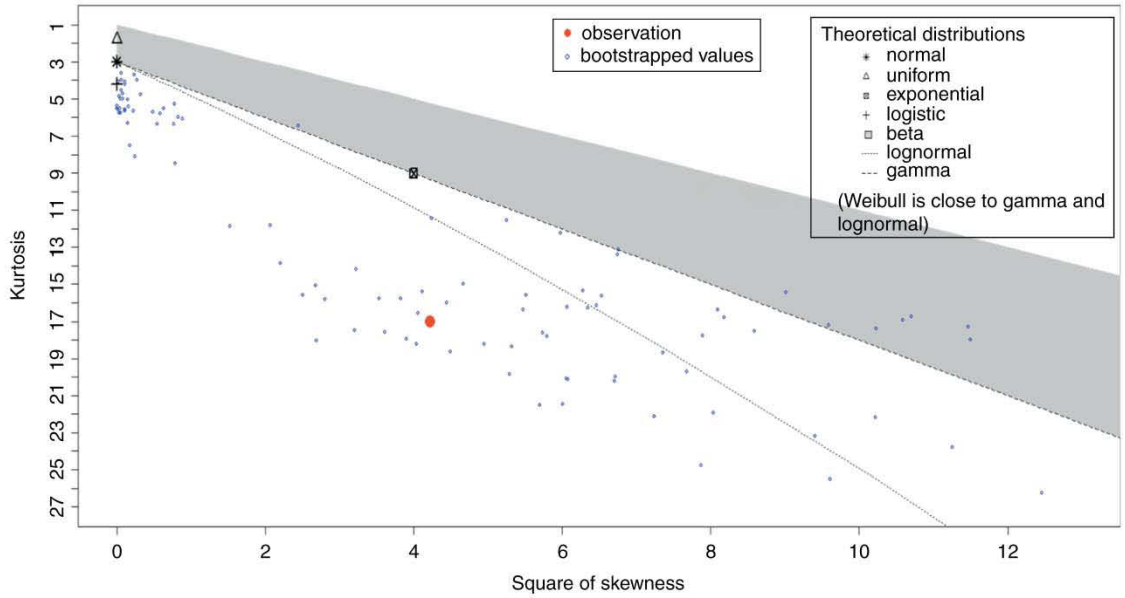


Figure 12

Cullen and Frey graph for the errors distribution of the ACD/Labs method.

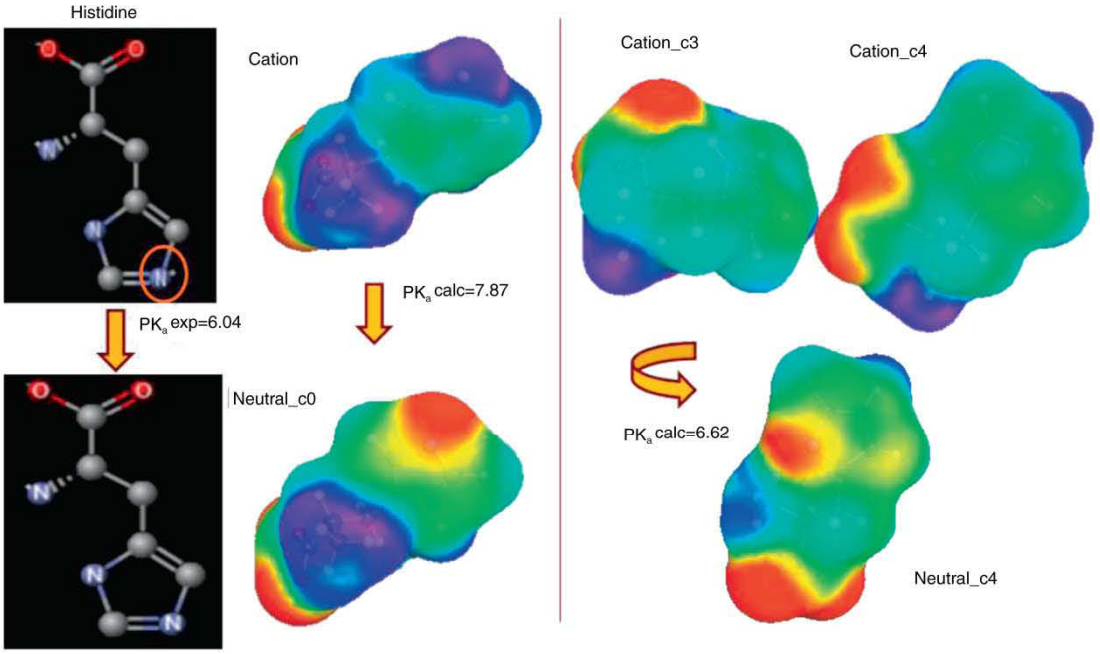


Figure 13

Illustration of the conformers set treatment in the case of histidine molecule. Only the most stable conformations in water are shown on the right (the COSMO-geometries used in the first prediction are shown on the left).

TABLE 7

Values of the average RMSE (Root Mean Square Errors of the differences ΔpK_a between the predicted values and the experimental ones: $\Delta pK_a = pK_a^{calc} - pK_a^{exp}$) for each of the 3 methods used for pK_a prediction (ChemAxon, COSMO-RS and ACD/Labs), after taking into account the conformations of histidine and histidyl-glutamic acid in COSMO-RS calculations. The values in parenthesis show the COSMO-RS pK_a results before this conformers set treatment

	RMSE ChemAxon	RMSE COSMO-RS	RMSE ACD/Labs	Number of points
Amino-acids (AA)	0.361	0.536 (0.577)	0.407	45
Dipeptides	0.612	0.633 (0.668)	0.669	37
Tripeptides	0.239	0.510 (0.510)	0.259	25
Overall data	0.445	0.566 (0.596)	0.490	107

TABLE 8

Values of the average RMSE (Root Mean Square Errors of the differences ΔpK_a between the predicted values and the experimental ones: $\Delta pK_a = pK_a^{calc} - pK_a^{exp}$) for each of the 3 methods used for pK_a prediction (ChemAxon, COSMO-RS and ACD/Labs), after taking into account the conformations of histidine and histidyl-glutamic acid in COSMO-RS calculations and averaging the multiple experimental pK_a data for the same dissociation reaction. The values in parenthesis are those given in Table 4

	RMSE ChemAxon	RMSE COSMO-RS	RMSE ACD/Labs	Number of points
Amino-acids (AA)	0.361 (0.361)	0.536 (0.577)	0.407 (0.407)	45 (45)
Dipeptides	0.622 (0.612)	0.646 (0.668)	0.684 (0.669)	35 (37)
Tripeptides	0.219 (0.239)	0.430 (0.510)	0.273 (0.259)	11 (25)
Overall data	0.468 (0.445)	0.570 (0.596)	0.520 (0.490)	91 (107)

2.4.2 Conformations Treatment Influence on COSMO pK_a Results

The averaged RMSE of the predicted pK_a values for each method compared to experimental results were respectively 0.596 for COSMO-RS, 0.445 for ChemAxon and 0.490 for ACD/Labs.

But, while the ChemAxon and ACD/Labs use a 2D structure to predict pK_a values, the COSMO-RS approach uses a 3D geometry. So the conformations treatment can have a non-negligible influence on COSMO-RS pK_a predicted values. For instance, using the most stable (in water) conformation sets of two of the studied molecules (histidine (*see Fig. 13*) and histidyl-glutamic acid), the RMSE of the COSMO-RS is improved by 0.02 as shown in Table 7.

2.4.3 Multiple Experimental Data Points Influence on pK_a Results

Furthermore, in the experimental pK_a values used, there were some cases where we got several different experimental values listed for the same dissociation center. By averaging these "multiple" experimental data to compare the prediction methods, we got 91 different data points (instead of 107) and the RMSE were slightly different than those got in Table 1.

The combination of a conformations set treatment for 2 molecules (histidine and histidylglutamic acid) and the averaging of multiple pK_a values (for the same dissociation center) decreased the RMSE of the COSMO-RS pK_a values by 0.02 while the ChemAxon and ACD/Labs RMSE increased by approximately 0.03 for each method (*Tab. 8*).

2.5 Discussion

It is well established that Group Contributions Methods (GCM) like ACD/Labs and ChemAxon method are in general more accurate in pK_a calculations. However, they are limited to some chemical families and their respective accuracy depends on the availability of experimental data. Compared to these methods, the main advantage of the COSMO-RS prediction method is that it is fully predictive. Indeed the LFER parameters of the COSMO-RS method were determined on a training set which does not include any of the molecules studied in this paper and the pK_a predicted are quite accurate.

Moreover, ACD/Labs and ChemAxon methods both use a 2D description of molecular structure in their respective algorithms for pK_a calculation. Thus these methods are not able to distinguish conformation treatment effect (which is a 3D effect) while COSMO-RS is able to perform this task (as shown in *Sect. 2.4.2*).

As mentioned earlier, each of the studied methods includes optionally parameterization tools to incorporate "local" data in order to bias predictions. This will increase for sure the prediction results; however these optional facilities were not studied in this paper since we are looking for a predictive tool that is able to predict pK_a of complex structures (that are ubiquitous in foods and biological systems). The COSMO-RS method seems very promising to determine the pK_a of a given molecule in a fully predictive way (especially when the conformations are well treated in calculations) with no available experimental data.

Ideally, a benchmarking of pK_a prediction models would require a universal training set to train all models, and a universal disjoint and similarly the use of diverse test set to compare their prediction. However it is very difficult to perform such task on all of the commercial pK_a prediction utilities used in the present study. With no true benchmarks for pK_a prediction utilities, the only way to identify a superior model is by trusting statistics. The statistics for both training and test data should be separate; unfortunately this is not the case in the present study since we are not able to distinguish the molecules that were used to parameterized ChemAxon and ACD/Labs methods.

All empirically based models should have r^2 closed to 1.0 and RMSE as close to 0.0 as possible over a wide range of compounds. Regarding the r^2 values, each of the studied models should be considered as accurate (r^2 = 0.98 for ChemAxon and ACD/Labs and r^2 = 0.96 for COSMO-RS). The averaged RMSE of the predicted pK_a values for each method compared to experimental results were respectively 0.596 for COSMO-RS, 0.445 for ChemAxon and 0.490 for ACD/Labs. Since, all of these RMSE values are close to 0.0; one can conclude that each model is suitable for pK_a prediction. This statement is confirmed by other statistical analysis (normality tests, bias factor and accuracy factor that are ubiquitous in comparing models in food science).

Moreover, it has been reported that a successful evaluation of seafood spoilage models present a bias factor (B_f) in the range 0.75-1.25. More drastically, a bias factor in the range 0.90-1.05 is considered as good for models dealing with pathogens growth (no 'fail-dangerous' predictions) [25]. The B_f values obtained (on the overall data) in this study are not greater than 1.05 which indicates a very good prediction of the pK_a values by all the models.

Likewise, it is also reported that the best performance that might be expected from a kinetic model encompassing the effect of temperature, pH and aw on growth rate, is $\sim$30%, or an accuracy factor A_f of 1.3. This value is greater than all the A_f values determined within this study, confirming again that all the studied models are reliable.

However, it is difficult to determine the best model because we were not able to distinguish the data used to train each model (especially ChemAxon and ACD/Labs). As suggested by Lee and Crippen [10], it should be interesting to study the performance of a consensus model based on these 3 methods. Since the statistics obtained from a consensus model may not reflect its performance on new data, this kind of study is out of the scope of this paper in which we are looking for a fully predictive model able to treat other products of interest in foods and biological systems.

CONCLUSION

The results presented in this study indicate that all the 3 methods (ChemAxon, ACD/Labs and COSMO-RS) are effective in predicting pK_a values (with a RMSE of about 0.5 pK_a-unit) for compounds of interest in food sciences like amino-acids, dipeptides and tripeptides. Furthermore, it appears that ACD/Labs, ChemAxon and COSMO-RS each can sometimes give large errors. A statistical study of the prediction errors (for this training set) showed that COSMO-RS method has the closest distribution to a normal (or Gaussian) distribution followed by ChemAxon, and that ACD/Labs pK_a errors do not follow a normal law. This was confirmed when analyzing the influence of ionic strength. Since COSMO-RS performs a Quantum Chemistry (QC) calculation on a 3D geometry while the two other methods are using a 2D structure (generated directly from SMILES file), one can expect that conformations treatment has to be taken into account to have a better pK_a prediction. This effect was studied for the case of 2 molecules of our training set and reduced the RMSE of the COSMO-RS method by about 0.02. All these packages include the ability to bias predictions using "local" data but these facilities were not evaluated. ChemAxon's Marvin and ACD/Labs are the fastest tools in term of computer-time. Due to the time-consuming QC calculations preceding its thermodynamics calculations, the COSMO-RS is less fast. But when this calculation is done once for each molecule and ion of interest, the COSMO-RS thermodynamics algorithm takes the same amount of time as ChemAxon and ACD/Labs to perform pK_a predictions.

Regarding these results, COSMO-RS appears as a promising method to predict the pK_a values of molecules of interest in food science with scarce available pK_a values such as peptides.

The final goal of this study is to use the pK_a values in a predictive thermodynamics model for products of interest in food industry.

ACKNOWLEDGMENTS

This work was funded by the Na⁻ integrated programme (ANR-09-ALIA-013-01) financed by the French National Research Agency. Thanks to Dr. Andreas Klamt and Dr. Fabrice Audonnet for helpful discussions.

REFERENCES

1 Katritzky A.R. (1963) *Physical Methods in Heterocyclic Chemistry*, Academic Press, New York, USA.

2 Perrin D.D. (1981) *pKa Prediction for Organic Acids and Bases*, Dempsey B., Serjeant E.P. (Eds), London, Great Britain, ISBN 0 412 22190 X.

3 Marvin (2011): "Marvin was used for drawing, displaying and characterizing chemical structures, substructures and reactions, *Marvin 5.4.1.1, 2011*, ChemAxon (http://www.chemaxon.com)".

4 Calculator Plugins (2011): "Calculator Plugins were used for structure property prediction and calculation, *Marvin 5.4.1.1, 2011*, ChemAxon (http://www.chemaxon.com)".

5 Harding A.P., Wedge D.C., Popelier P.L.A. (2009) pK_a Prediction from "Quantum Chemical Topology" Descriptors, *J. Chem. Inf. Model* **49**, 1914-1924.

6 Lebert I., Lebert A. (2006) Quantitative prediction of microbial behaviour during food processing using an integrated modelling approach: a review, *Int. J. Refrig.* **29**, 968-984.

7 Manchester J., Walkup G., Rivin O., You Z. (2010) Evaluation of *pKa* Estimation Methods on 211 Druglike Compounds, *J. Chem. Inf. Model.* **50**, 565-571.

8 COSMOthermX (7 December 2011) A Graphical User Interface to the COSMOtherm Program, Tutorial for version C21_0111, *COSMOlogic GmbH & Co. KG*, Leverkusen, Germany.

9 Eckert F., Klamt A. (2010) COSMOtherm, Version C2.1. Release 01.11, *COSMOlogic GmbH & Co. KG*, Leverkusen, Germany.

10 Lee A.C., Crippen G.M. (2009) Predicting pK_a, *J. Chem. Inf. Model* **49**, 2013-2033.

11 ACD/ChemSketch (2006) version 10.01 (Release 10.00), *Advanced Chemistry Development, Inc.*, Toronto, ON, Canada, www.acdlabs.com.

12 Klamt A., Schüürmann G. (1993) COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expression for the Screening Energy and its Gradients, *J. Chem. Soc. Perkin Trans.* **2**, 799.

13 Klamt A., Jonas V., Bürger T., Lohrenz J.W.C. (1998) Refinement and Parameterization of COSMO-RS, *J. Phys. Chem. A* **102**, 5074.

14 Perrin D.D. (1965) *Dissociation constants of organic bases in aqueous solution*, International Union of Pure and Applied Chemistry, Butterworths, London, England.

15 Perrin D.D. (1972) *Dissociation constants of organic bases in aqueous solution*, supplement 1972, International Union of Pure and Applied Chemistry, Butterworths, London, England, ISBN 0 408 70408 X.

16 Szegezdi J., Csizmadia F. (2004) Prediction of dissociation constant using microconstants, *27th ACS (American Chemical Society) National Meeting*, Anaheim, California, 28 March-1 April.

17 Szegezdi J., Csizmadia F. (2007) Method for calculating *pKa* values of small and large molecules, *233rd ACS (American Chemical Society) National Meeting*, IL, Chicago, 25-29 March.

18 ACD/pK_a DB (2006) version 10.0 for Microsoft Windows, Reference Manual, Comprehensive Interface Description, *Advanced Chemistry Development, Inc.*, Toronto, ON, Canada, www.acdlabs.com.

19 Eckert F., Klamt A. (2002) Fast Solvent Screening *via* Quantum Chemistry: COSMO-RS Approach, *AIChE J.* **48**, 369.

20 Klamt A., Eckert F. (2000) COSMO-RS: a novel and efficient method for the *a priori* prediction of thermophysical data of liquids, *Fluid Phase Equilib.* **172**, 43.

21 Klamt A., Eckert F., Diedenhofen M. (2003) First principles calculations of aqueous pK_a values for organic and inorganic acids using COSMO-RS reveal an inconsistency in the slope of the pK_a scale, *J. Phys. Chem. A* **107**, 9380-9386.

22 Eckert F., Diedenhofen M., Klamt A. (2010) Towards a first principles prediction of *pKa*: COSMO-RS and the cluster-continuum approach, *Molec. Phys.* **108**, 3-4, 229-241.

23 Weininger D. (1988) SMILES, A Chemical Language and Information System, 1. Introduction to Methodolody and Encoding Rules, *J. Chem. Inf. Model.* **28**, 31-36.

24 Neumeyer K., Ross T., Thomson G., McMeekin T.A. (1997) Validation of a model describing the effects of temperature and water activity on the growth of psychrotrophic pseudomonads, *Int. J. Food Microbiol.* **38**, 55-63.

25 Mellefont L.A., McMeekin T.A., Ross T. (2003) Performance evaluation of a model describing the effects of temperature, water activity, pH and lactic acid concentration on the growth of *Escherichia coli*, *Int. J. Food Microbiol.* **82**, 45-58.

26 Delignette-Muller M.L., Pouillot R., Denis J.-B., Dutang C. (2010) Fitdistrplus: help to fit of a parametric distribution to non-censored or censored data, R package version 0.1-3, http://CRAN.R-project.org/package=fitdistrplus.

27 Dalgaard P. (2002) *Introductory Statistics with R (Statistics and Computing)*, Springer-Verlag, New York, USA, ISBN 0-387-95475-9.

28 The R software: R Development Core Team (2011) R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, URL http://www.R-project.org/.

29 Ould-Moulaye C.B. (1998) Calcul des propriétés de formation en solution aqueuse des composés impliqués dans les procédés microbiologiques et alimentaires - prédiction et réconciliation de données - modélisation des équilibres chimiques et des équilibres entre phases, *PhD Thesis/Thèse*, Université Blaise Pascal.

30 Goldberg R.N. (1981) Evaluated activity and osmotic coefficients for aqueous solutions: Thirty-six uni-bivalent electrolytes, *J. Phys. Chem. Ref. Data* **10**, 671.

Final manuscript received in December 2012
Published online in May 2013

This study was later extended to the determination of the pK_a of several other organic acids (mainly carboxylic acids and diacids). The latter represent generally the structural units of fatty acids which are the main components of lipids and polyamides of amino-carboxylic acids also included proteins. Carboxylic acids are used in the production of polymers, pharmaceuticals, solvents, and food additives. Industrially important, carboxylic acids include acetic acid (component of vinegar, precursor to solvents and coatings), adipic acid (polymers), citric acid (beverages), ethylenediaminetetraacetic acid (chelating agent), fatty acids (coatings), maleic acid (polymers), propionic acid (food preservative), terephthalic acid (polymers).

This extended study leads to the same conclusions as Touré et al. (2013). Indeed, as illustrated in Table IV-1, the RMSE of the predicted results (shown in Figure IV-2 and Figure IV-3) are slightly greater for COSMO-RS because the available experimental data were not used to train this model, while in almost all of the other predictive tools experimental data concerning these acids are provided and ubiquitously used to bias the predicted results. Therefore, it must be outlined that COSMO-RS results are *a priori* predictions revealing the underlying power of the model.

Table IV-1: Values of the average RMSE of the difference ΔpK_a between the predicted values and the experimental ones, $\Delta pK_a = pK_a^{calc} - pK_a^{exp}$) for each of the 3 methods used for pK_a prediction (ChemAxon, COSMO-RS and ACD/labs) on an extended set of molecules.

	RMSE ChemAxon	RMSE COSMO-RS	RMSE ACD/Labs	Number of data points
Organic acids	0.060	0.270	0.030	33
Amino acids (AA)	0.361	0.577	0.407	45
Dipeptides	0.612	0.668	0.669	37
Tripeptides	0.239	0.510	0.259	25
Overall data	**0.402**	**0.537**	**0.429**	140

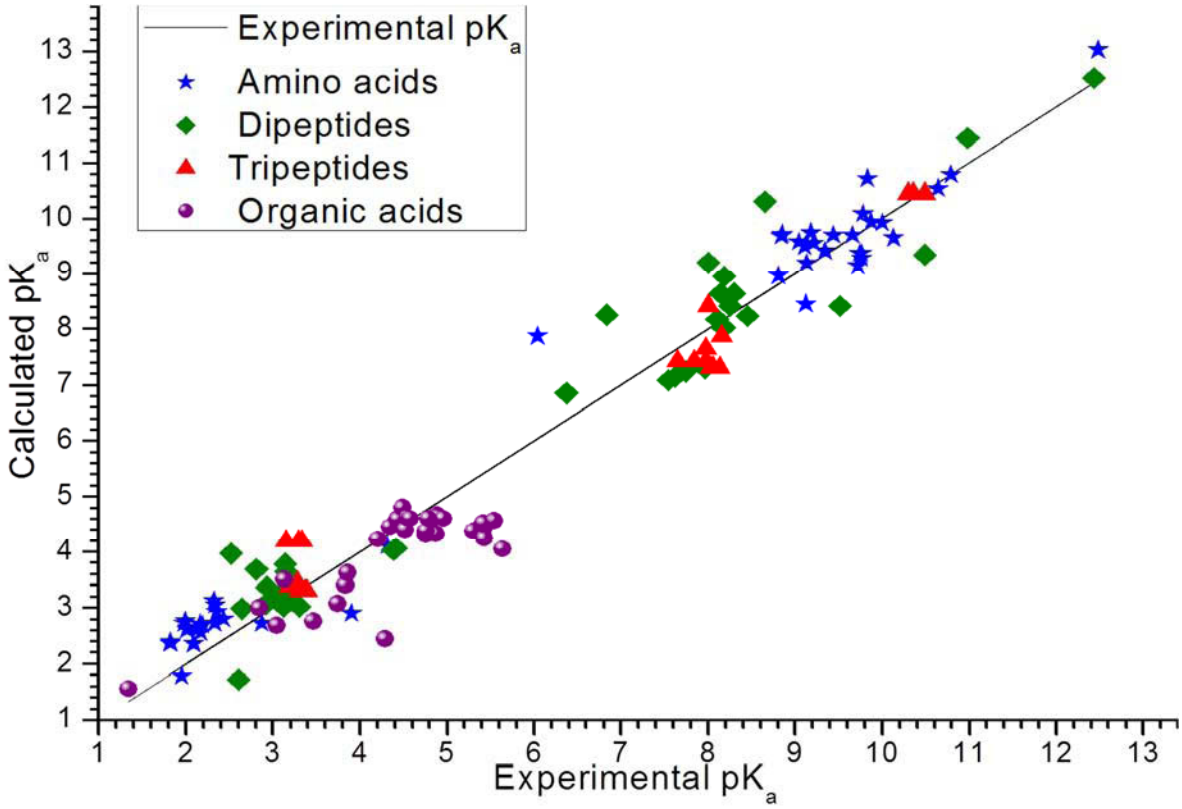


***Figure IV-2** : Comparison of the predicted pK_a values using COSMO-RS to the experimental data for amino acids, dipeptides, tripeptides and several organic acids.*

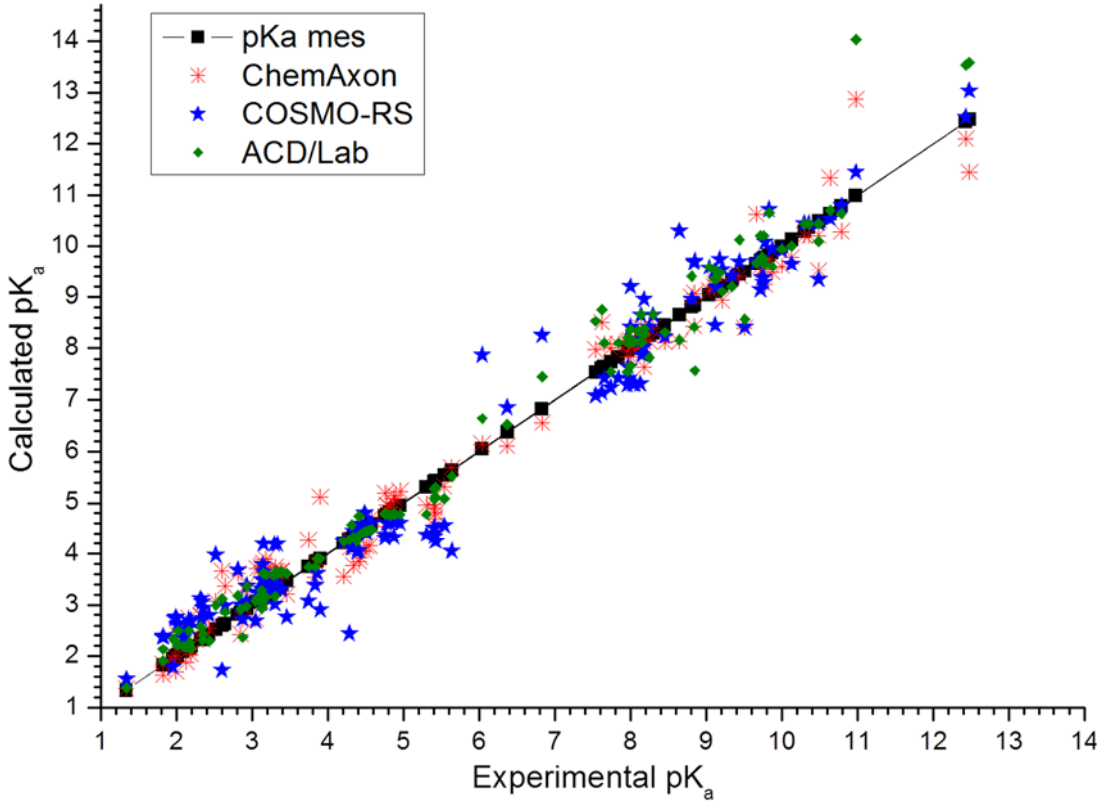


***Figure IV-3** : Comparison of the predicted pK_a values using COSMO-RS, ACD/Labs and ChemAxon methods for amino acids, peptides and several organic acids.*

As discussed above, the COSMO-RS model appears as a promising tool for *pKa* predictions for a wide variety of molecules. In its current *pKa* prediction algorithm, a linear free energy relationship (LFER) is used in the COSMO-RS method.

$$pK_a = A\left(\frac{\Delta G^j_{ion} - \Delta G^j_{neutral}}{RT\ln 10}\right) + B = c_0 + c_1\left(\Delta G^j_{ion} - \Delta G^j_{neutral}\right) \qquad \text{(IV-1)}$$

where $\Delta G^j_{neutral}$ and ΔG^j_{ion} are respectively the free enthalpies of the neutral and ionic compounds in the solvent (water in our case) at infinite dilution. This introduces two LFER parameters (*A* and *B*) or (c_0 and c_1) that have theoretical physical meanings, though they are generally determined from a linear regression of experimental data. Indeed, since the *pKa* values characterize a proton exchange between the protonated and deprotonated form of a given acid, the LFER parameters has to be identified only once.

For instance, Klamt et al. (2003) have determined these parameters by correlating calculated free energies of dissociation with the experimental aqueous *pKa* for a set of 64 organic and inorganic acids (including any peptide). The LFER parameters $c_0 = B$ and $c_1 = \frac{A}{RT\ln 10}$ used to predict the *pKa* values of interest in this study were read from the COSMOtherm parameter file. At ambient temperature, their values are respectively c_0=-120.29804 and c_1=0.10927 mol/kJ. Thus, $A = c_1 RT\ln 10 = 0.6$ and $B = c_0$ =-120.29804 (using the BP_TZVP_C21_0111 parameterization file in the COSMOtherm software).

To compute the values of $\Delta G^j_{neutral}$ and ΔG^j_{ion} in order to calculate $\Delta G^j_{diss} = \Delta G^j_{ion} - \Delta G^j_{neutral}$ and then to deduce a *pKa* value, it is necessary to do quantum COSMO calculations of a molecule in its neutral and in its ionic state.

If the formation properties were used to calculate the chemical potential (at infinite dilution in water) of each species and then deduce the value of the pKa from the following reaction,

$$AH \rightleftharpoons A^- + H^+ \qquad \text{(IV-2)}$$

one would have (by definition in reference conditions, $\mu^0_{H^+} = 0$) :

$$c_0 = B = \frac{\mu^0_{H^+}}{RT\ln 10} = 0 \qquad \text{(IV-3)}$$

$$c_1 = \frac{A}{RT\ln 10} = \frac{1}{RT\ln 10} \qquad \text{(IV-4)}$$

However, because the proton H^+ cannot be simulated in DFT calculations (it has no electron), it is replaced by its hydrated form H_3O^+.Then, we are studying the reaction,

$$AH + H_2O \rightleftharpoons A^- + H_3O^+ \qquad \text{(IV-5)}$$

and the values of c_0 and c_1 become:

$$c_0 = B = \frac{\mu^0_{H_2O} - \mu^0_{H_3O^+}}{RT\ln 10} \neq 0 \quad \text{(IV-6)}$$

$$c_1 = \frac{A}{RT\ln 10} \quad \text{(IV-7)}$$

In theory, c_0 can be determined from the analysis of the pK_a value of H_3O^+ determined from the study of the equation:

$$H_3O^+ \rightleftharpoons H_2O + H^+ \quad \text{(IV-8)}$$

It has been published that the experimental pK_a value of H_3O^+ is equal to -1.74 (Starkey et al. 1986) that would correspond to the absolute value of c_0^{th}, *i.e.* the theoretical value of the LFER constant c_0. However, because in the COSMO-RS algorithm quantum calculated free energies are used for both H_3O^+ and H_2O, then c_0 becomes:

$$c_0 = c_0^{th} + c_1\left(\Delta G^{CRS}_{H_2O} - \Delta G^{CRS}_{H_3O^+}\right) = c_1\left(\Delta G^{CRS}_{H_2O} - \Delta G^{CRS}_{H_3O^+}\right) - pK_a^{\exp}\left(H_3O^+\right) \quad \text{(IV-9)}$$

After calculations, we get c_0=-120.4936, using c_1=0.10927 mol/kJ, $c_0^{th} = -pK_a^{\exp}\left(H_3O^+\right)$=1.74 and $\Delta G^{CRS}_{H_3O^+} - \Delta G^{CRS}_{H_2O}$=-1118.6378 kJ/mol. This value of c_0 perfectly corresponds to the value c_0=-120.29804 determined by Klamt (2003) directly from the linear regression of experimental data.

Likewise, theoretically the LFER parameter A must be equal to unity; however since the quantum DFT energy is used to determine the total Gibbs free energy from quantum calculations in the COSMO-RS method, due to the discrepancies between quantum and macroscopic states, the value of A becomes lower than unit. This lower slope is described as having a "mysterious reason outside the COSMO-RS calculation methods" in Klamt (2005).

As it will be discussed later, this lower slope (*i.e.* value of A) comes from the enthalpic contribution which is overestimated in any quantum calculation tools compared to the enthalpy of formation. Indeed, the LFER coefficient A is about 0.6 in almost all of quantum DFT based pK_a prediction algorithms (Klamt et al. 2003, Klamt 2005, Eckert et al. 2010). Using an atom-based contribution method to correlate DFT level calculated energies (expressed in hartrees, *i.e.* the atomic unit of energy ;1 hartree = 2625.5 kJ/mol = 627.51 kcal/mol) to enthalpy of formation data, it can be proved that the slope A corresponds to the correlation coefficient of hydrogen atom $\varepsilon(H)$=0.59, as discussed in section IV.2. To our knowledge, this is the first plausible explanation of this lower slope value of A in quantum based pK_a prediction tools.

Ould Moulaye (1998) has demonstrated that the consistent determination of the pK_a values requires the knowledge of the formation properties in the appropriate reference state. Quantum based

calculations are performed generally in the ideal gas reference state while the reactions do happen in liquid solutions (as treated by the RS part of the COSMO-RS method to determine excess properties). In the present case, instead of choosing an arbitrary reference as usually done in excess free enthalpies models, the reference chemical potential must be equal to the chemical potential of formation in order to predict the pKa data. To do so, a new prediction method based on quantum calculations has been developed during this thesis. This tool is also extendable to the prediction of the formation properties at infinite dilution in water, *i.e.* the reference state chosen for molecules in solution within this study, as detailed below.

IV.3. Prediction of formation properties in gas phase and at infinite dilution in water

The basis of the representation of thermodynamic properties of solutions depends on the choice of the reference state of each component included in the solution. Regarding the choice of reference temperature and pressure, there is no particular difficulty (in general, it is widely accepted that $T_r = 25°C = 298.15K$ and $p_r = 1atm = 1.01325*10^5 Pa$ even if several sources used $p_r = 1bar = 10^5 Pa$). However the choice of the physical state of reference of a given compound is more challenging since the latter can be simulated in the pure solid, liquid or gas reference states.

The formation properties of a given molecule are defined as the thermodynamic equilibrium properties of the reaction of formation of this molecule from a stoichiometric mixture of reference elements (taken in their respective physical reference states). Classically, the organic molecules of the $C_xH_yO_zN_t$ type are treated by using as reference elements: carbon (C solid, graphite), hydrogen (H_2 gas), oxygen (O_2 gas) and nitrogen (N_2 gas). In this case the formation reaction can be written as:

$$xC(\text{gr}) + \frac{y}{2}H_2(gas) + \frac{z}{2}O_2(gas) + \frac{t}{2}N_2(gas) \rightleftharpoons C_xH_yO_zN_t \qquad \text{(IV-10)}$$

The formation enthalpy ($h_i^0 = \Delta H_{f,i}^0$) and Gibbs free energy ($\mu_i^0 = \Delta G_{f,i}^0$) represent the change in the appropriate thermodynamic quantity when one mole of the substance in its standard state is formed, isothermally at the indicated temperature, from the elements, each in its appropriate standard reference state. By convention, the formation properties of the reference elements are exactly equal to zero. In thermodynamic tables, the standard reference state at 25 °C for each element has been chosen to be the physical state that is thermodynamically stable at 25°C and 1 atm pressure. In reference conditions, the compound can also be simulated at the infinite dilution in

water, as ubiquitously done for ionic species. The latter choice is adopted for all species within the present study.

In this context, since this kind of data is not systematically given in existing tables, there are two main thermodynamic modelling tasks: the prediction and/or the collection of formation properties data in gas phase, and the conversion of gas phase data to calculate the corresponding properties at infinite dilution in water, as discussed below.

IV.3.1. Prediction of formation properties in gas phase

For a gas, the standard state is the hypothetical ideal gas at unit fugacity, in which the state of the enthalpy is that of the real gas at the same temperature and at zero pressure.

In the hypothetical gas standard state, some computational methods (particularly *ab initio* techniques) produce detailed molecular information but no thermodynamic information directly. Further calculations are needed to generate familiar, ideal-gas quantities such as the standard molar entropy ($S°$), heat capacity ($C_p°$), and enthalpy change [$H°$ (T)-$H°$ (0)]. Thermo-chemical calculations are extended to transition states of chemical reactions, as detailed in Irikura et al. 1998.

In this context, statistical thermodynamic calculations are necessary to compute properties as functions of the temperature. Moreover, in some computations, such as *ab initio* electronic calculations of the molecular energy, the raw results do not even correspond to properties at absolute zero temperature and must always be corrected. All the corrections are based upon molecular spectroscopy, with temperature-dependence implicit in the molecular partition function. The partition function is used not only for theoretical predictions, but also to generate most of the published thermochemical tables. Many data compilations include descriptions of calculational procedures (Irikura et al. 1998).

By convention, energies from *ab initio* calculations are reported in hartrees. These energies are negative, with the defined zero of energy being the fully-dissociated limit (free electrons and bare nuclei). *Ab initio* models also invoke the approximation that the atomic nuclei are stationary, with the electrons swarming around them. This is a good approximation because nuclei are much heavier than electrons. Consequently, the resulting energies are for a hypothetical, non-vibrating molecule. Although oscillators may be at rest in classical mechanics, real (quantum-mechanical) oscillators are always in motion. The small residual motion at absolute zero temperature is the zero-point vibrational energy, abbreviated as ZPVE or ZPE. For a simple harmonic oscillator, the ZPE equals half of the vibrational frequency. Although all the real molecular vibrations are at least

slightly anharmonic, they are usually approximated as harmonic. Thus, the molecule's ZPE may be taken as one-half of the sum of the vibrational frequencies (Irikura et al. 1998).

However, it is very time consuming to compute vibrationnal frequencies in quantum calculation tools, especially for the most important molecules of interest in food and biological systems. To overcome this difficulty, Rice et al. (1999) have used the method of atom equivalents to determine the gas phase heats of formation $h_i^{0,gas}$ from a linear regression of the quantum DFT calculated energies ($E_i^{DFT} = E_i^{gas,CRS}$) in standard conditions. They obtained a root mean square error RMSE of 3.1 kcal/mol (*i.e.* 12.96 kJ/mol) on a set of 35 molecules studied, suggesting that this kind of approximation (that has been slightly improved by Byrd and Rice (2006)) seems to be very promising in order to determine reasonable enthalpies of formation data, especially for molecules with high molecular weight that cannot be simulated in a fully quantum *ab initio* framework.

The formation properties prediction method developed in this study is based on the use of several quantum results which are also used as input data in the COSMO-RS method (mainly in the vapour pressure calculation algorithm), in order to calculate the gas phase formation properties of the pure compounds. The main advantage of this approach is that quantum calculated energies include structural information that cannot be taken into account in a fully predictive way in any empirical approach (especially when there are not enough experimental data like the case of complex structures which are ubiquitous in foods and biological systems), as well as the opportunity to take into account conformation effects.

The enthalpy of formation prediction method is inspired from the work of Rice et al. (1999) and Byrd and Rice (2006) who have earlier published several data about $h_i^{0,gas}$ prediction from quantum DFT calculated energies. In comparison to the original work of Rice and co-workers who used an atom-based contribution approach, the enthalpy of formation prediction method developed herein is based on group contributions since several studies (Benson and Buss 1958; Benson et al. 1969) have proved that a group-contribution based method is the most successful way to compute thermodynamic properties of pure components.

The main originality of this study is that it introduces a similar method to predict the Gibbs free energy of formation $\mu_i^{0,gas}$. For this purpose, $\mu_i^{0,gas}$ is correlated to the gas phase chemical potential $\mu_i^{gas,CRS}$ which is computed in the COSMO-RS algorithm in the vapour pressure prediction module using the following equation:

$$\mu_i^{gas,CRS} = E_i^{gas,CRS} - E_i^{COSMO} + \eta^{gas} - \omega_{Ring} n_{Ring}^{X_i} + \tau_\eta - \tau_\eta \frac{T}{T_{room}} - \omega_{Ring} n_{Ring}^{X_i} \left(-\tau_{\omega_{Ring}} + \tau_{\omega_{Ring}} \frac{T}{T_{room}} \right) \quad \text{(IV-11)}$$

where the energies $E_i^{gas,CRS}$ and E_i^{COSMO} are calculated from quantum calculations (as discussed in chapter II). The remaining contributions in equation IV-11 consist of a correction term for ring shaped molecules with $n_{Ring}^{X_i}$ being the number of ring atoms in the molecule and ω_{Ring} an adjustable parameter, the temperature dependency parameters τ_η and $\tau_{\omega_{Ring}}$ (that will not contribute if T= T_r = 25°C=298.15 K) and the parameter η^{gas} which provides the link between the reference states of the system's free energy in the gas phase and in the liquid (Klamt 2005). The latter (η^{gas}) is one of the seven parameters of the COSMO-RS model. It is important to notice that within the COSMO-RS algorithm the parameter η^{gas} has a constant value for all compounds and that its value has been adjusted from a fitting of the model in order to predict vapour pressures data. Since several authors (Klamt 1995; Klamt 2005; Nakajoh et al. 2006, Eckert and Klamt 2010) have already proven that COSMO-RS is very powerful in the prediction of vapour pressures, this new prediction method would give accurate values of the gas phase Gibbs free energy of formation $\mu_i^{0,gas}$.

Because this study is performed in reference conditions (*i.e.* at room temperature and atmospheric pressure), the gas phase chemical potential $\mu_i^{gas,CRS}$, which is computed in the COSMO-RS algorithm becomes (if we assume that there is no ring correction to take into account):

$$\mu_i^{gas,CRS} = E_i^{gas,CRS} - E_i^{COSMO} + \eta^{gas} \qquad \text{(IV-12)}$$

Since η^{gas} has a constant value for all molecules (Klamt 2005), it will be neglected within the formation properties regression procedure developed herein.

Thus, in summary, $h_i^{0,gas}$ and $\mu_i^{0,gas}$ are calculated as:

$$\begin{cases} h_i^{0,gas} = \Delta H_{f,i}^{0,gas} = E_i^{gas,CRS} + \sum_k \alpha_k \varepsilon_k \; ; \\ \mu_i^{0,gas} = \Delta G_{f,i}^{0,gas} = \mu_i^{gas,CRS} + \sum_k \alpha_k \mu_k \end{cases} \qquad \text{(IV-13)}$$

where α_k denotes either the number of atoms $k = \{C; H; O; N; ....\}$ contained in molecule *i* (if the atom equivalents method is used to predict formation properties) or the number of groups $k = \{CH_3; CH_2; OH; ....\}$ in the UNIFAC decomposition of the molecule *i* (if a group-equivalents method is used to predict formation properties); ε_k and μ_k are determined from a multi-linear regression of enthalpies and Gibbs free energies data. It is important to notice that all the UNIFAC decomposition were determined using the Simulis Thermodynamics software which also includes several DIPPR experimental gas phase enthalpies and Gibbs free energies data.

IV.3.1.1. Formation properties results predicted using both atom-equivalents and group-equivalents methods on a small set of molecules

Before going further, it is interesting to compare the two kinds of approaches (atom equivalents and group equivalents methods) in a small set of molecules including alkanes, alcohols and ketones.

To do so, the regression coefficients have been determined on a training set including 16 linear alkanes (from methane to n-hexadecane), 14 linear primary alcohols (from methanol to 1-tetradecanol), and 9 ketones (acetone, methylethylketone, 3-pentanone, 3-heptanone, 4-heptanone, 3-hexanone, 2-pentanone, 2-hexanone and 2-heptanone). Likewise, the performance of the prediction method has been tested on a test set including 16 linear alkanes (from n-heptadecane to n-triacontane, n-dotriacontane and n-hexatriacontane), 6 linear primary alcohols (from 1-pentadecanol to 1-eicosanol), and 3 ketones (5-nonanone, 2-nonanone and 2-octanone).

Figure IV-4 and Figure IV-6 show the results obtained respectively on heat of formation and Gibbs free energy of formation for a set of molecules selected in the (training set) to determine the regression parameters of equation (IV-9). Likewise, Figure IV-5 and Figure IV-7 represent the results obtained on the external set of molecules (*i.e.* the test set).

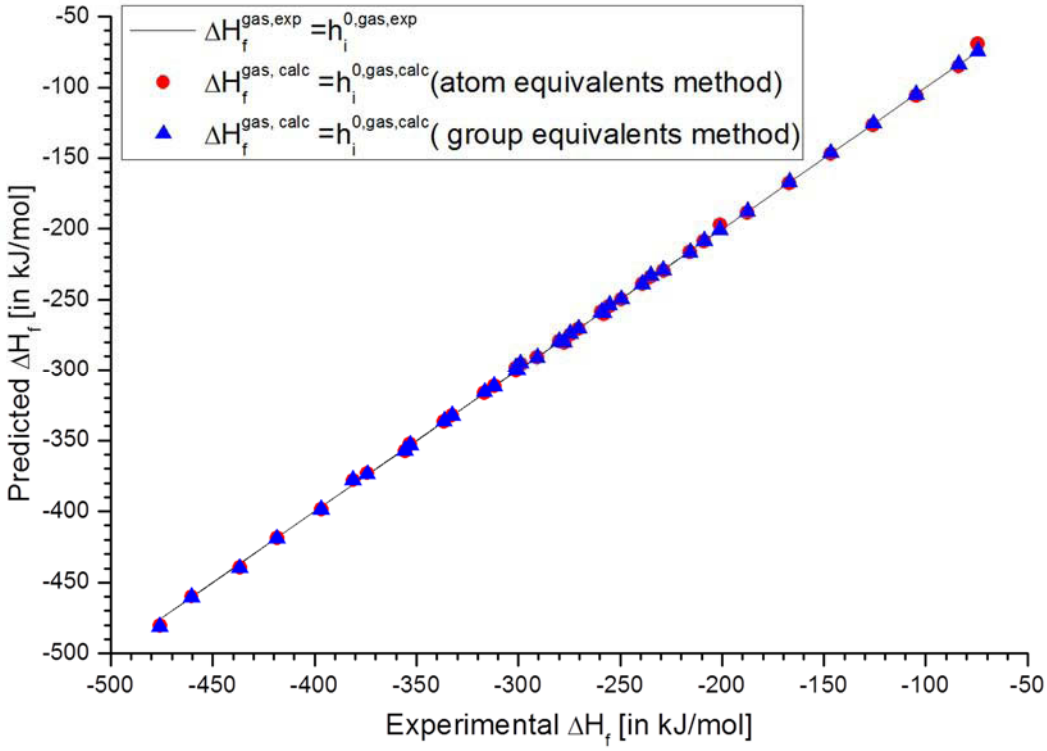


Figure IV-4 : Results of the prediction of enthalpies of formation for molecules included in the training set (including alkanes, alcohols and ketones).

The line represents the equality between calculated and predicted data, the circles and triangles denote the respective results of the atom and group equivalents methods.

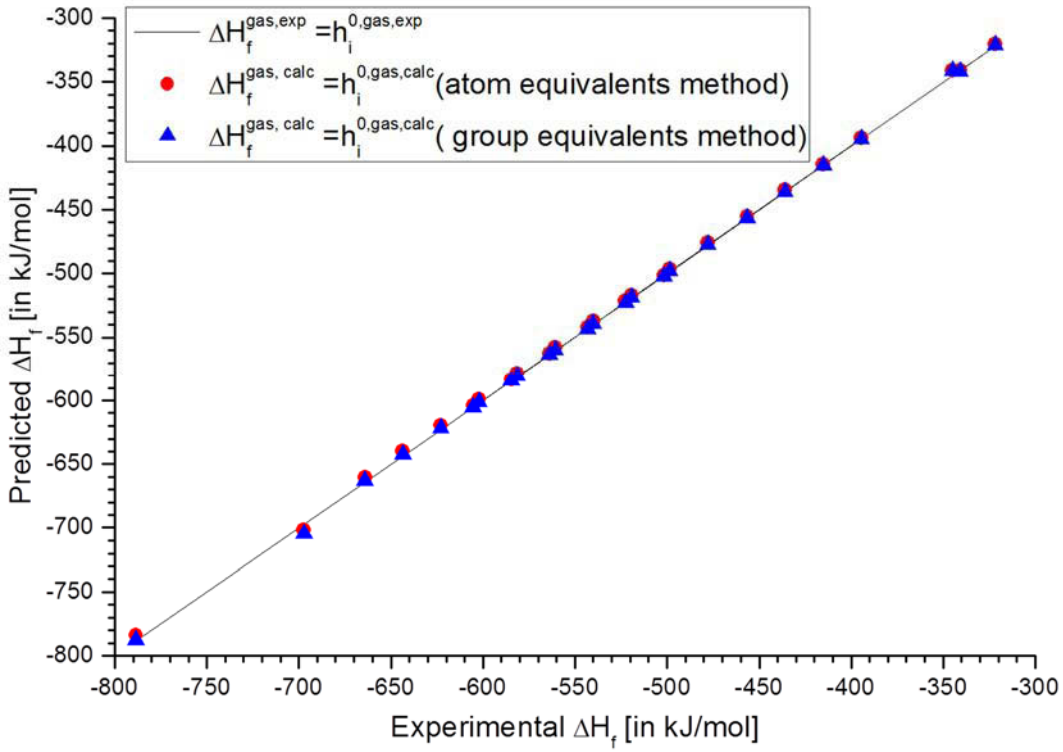


Figure IV-5 : Results of the prediction of enthalpies of formation for molecules included in the validation set (including alkanes, alcohols and ketones).

Same legend as in Figure IV-4.

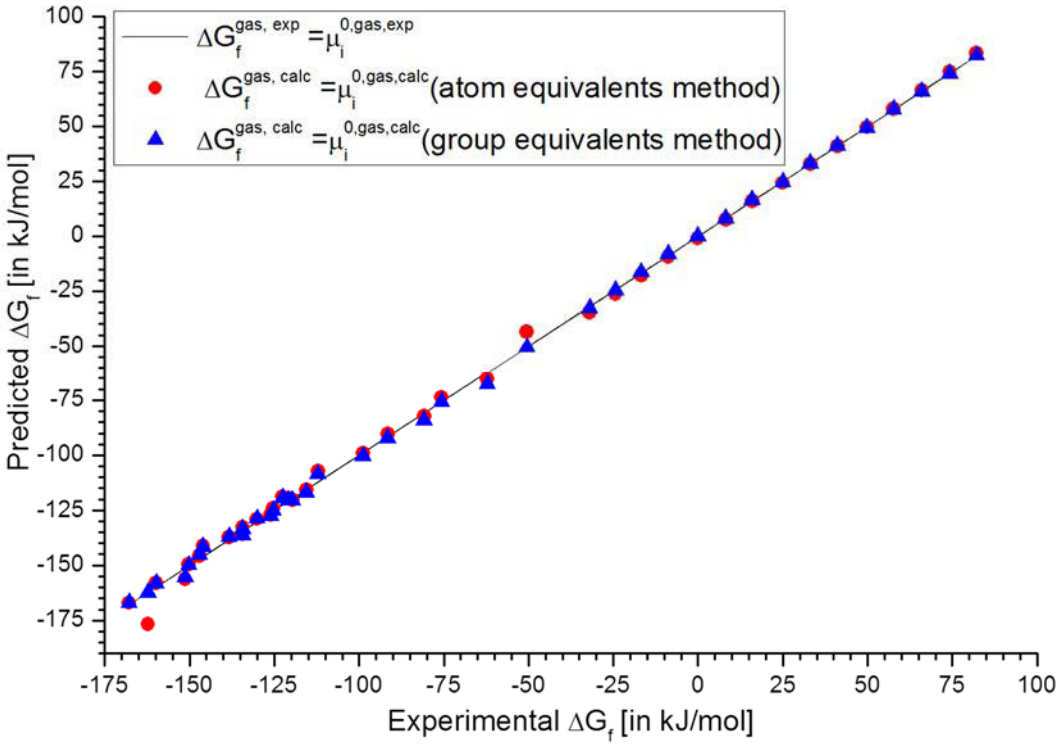


Figure IV-6 : Results of the prediction of chemical potentials of formation for molecules included in the training set (including alkanes, alcohols and ketones).
The line represents the equality between calculated and predicted data, the circles and triangles denote the respective results of the atom and group equivalents methods.

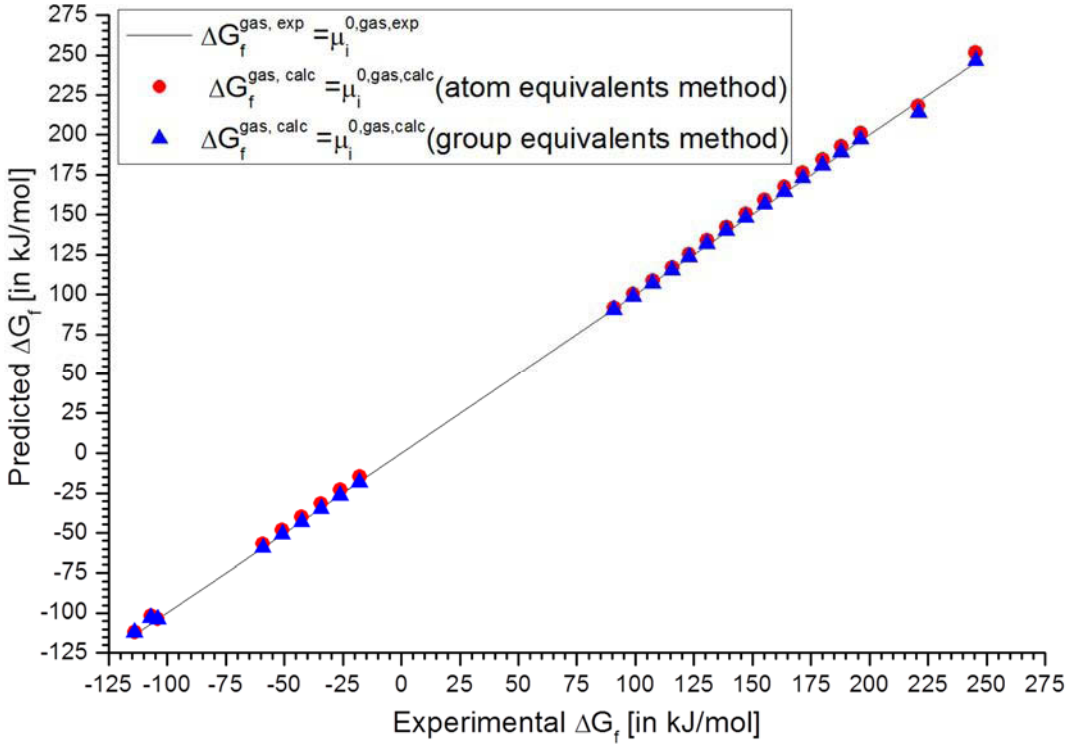


Figure IV-7 : Results of the prediction of chemical potentials of formation for molecules included in the validation set (including alkanes, alcohols and ketones).
Same legend as in Figure IV-6.

Tables IV-2 and IV-3 summarize the RMSE obtained respectively for the enthalpies of formation and the chemical potentials of formation predictions using in each case both atom equivalents and group equivalent methods.

Table IV-2: RMSE of the prediction method for the enthalpy of formation data for a set of alkanes, alcohols and ketones.

[a]Number of points in the training set (same legends for the other occurrences)
[b] Number of points in the validation set.

	Atom equivalents method Enthalpy of formation ΔH_f^{gas} (kJ.mol^{-1})		Group equivalents method Enthalpy of formation ΔH_f^{gas} (kJ.mol^{-1})	
Compound groups	RMSE (training set)	RMSE (validation set)	RMSE (training set)	RMSE (validation set)
Alkanes (16[a] +15[b])	1.48	7.71	0.23	0.57
Alcohols(14[a] + 6[b])	2.27	1.12	2.36	0.03
Ketones(9[a] + 3[b])	1.58	2.30	1.43	2.29
Overall data(39[a] + 24[b])	0.95	2.19	0.14	0.60

Table IV-3: RMSE of the prediction method for the chemical potential of formation data for a set of alkanes, alcohols and ketones.

[a]Number of points in the training set (same legends for the other occurrences)
[b] Number of points in the validation set.

	Atom equivalents method Chemical potential of formation ΔG_f^{gas} (kJ.mol^{-1})		Group equivalents method Chemical potential of formation ΔG_f^{gas} (kJ.mol^{-1})	
Compound groups	RMSE (training set)	RMSE (validation set)	RMSE (training set)	RMSE (validation set)
Alkanes (16[a] +15[b])	2.04	12.97	0.36	0.79
Alcohols(14[a] + 6[b])	4.44	2.72	2.42	0.32
Ketones(9[a] + 3[b])	2.29	3.11	2.09	2.45
Overall data(39[a] + 24[b])	1.31	2.85	0.23	0.70

The analysis of these RMSE data suggests that the predicted results are in good agreement with literature data, and that a group equivalents method is more suitable for formation properties

prediction than the atom-equivalent one. This confirms why Byrd and Rice (2006) have used a group-equivalents method to improve their previous results (Rice et al. 1999).

Table IV-4: ***Regression coefficients obtained for the training set containing alkanes, alcohols and ketones, using either the atom equivalents method or the group equivalents method.***

In the atom equivalent method, O' denotes oxygen atom that are involved in multiple-bonded environments, while C, H and O denote atoms that are involved in single-bonded environments.

[a] This value corresponds to the value of the LFER parameter A used in the pK_a prediction algorithm of the COSMO-RS method (as discussed in section IV.1). That explains the lower slope observed in the LFER relation when using quantum calculated energies for pK_a predictions.

		Enthalpy of formation regression coefficients		Chemical potential of formation Regression coefficients
	Elements (k)	ε_k (in kJ/mol)	ε_k (in hartrees)	μ_k (in kJ/mol)
Atom equivalents method	C	100096.464	38.125	59.331
	H	**1558.426**	**0.594[a]**	**-25.322**
	O	197387.454	75.181	-111.373
	O'	197398.346	75.185	-157.283
Group equivalents method (UNIFAC groups)	[CH_3OH]	303713.909	115.679	-139.054
	[CH_4]	106325.032	40.497	-49.027
	[CH_3]	104772.426	39.906	-15.691
	[CH_2]	103213.169	39.312	8.486
	[OH]	198946.508	75.775	-137.321
	[CH_3CO]	402265.820	153.215	-114.647
	[CH_2CO]	400707.512	152.621	-91.293

Since the previous study has proved that a group-contribution method is more accurate than an atom-based one, only the group-equivalent method will be used in the rest of this work for formation properties prediction.

IV.3.1.2. Formation properties results predicted using the group-equivalents method on an extended set of molecules

This extended study has been performed on a larger set of molecules including alkanes, alcohols, ketones, alkenes, alkynes, aromatics, phenols and aldehydes.

Figure III-8 and Figure IV-10 show the results obtained respectively on heat of formation and Gibbs free energy of formation for a set of molecules selected to determine the regression parameters of equation (IV-9). Likewise, Figure III-9 and Figure IV-11 represent the results obtained on an external set of molecules.

Table IV-5 summarizes the RMSE obtained in each case, demonstrating that the predicted results are in very good agreement with literature data, and the RMSE obtained here are lower than that observed by other authors (Rice et al. 1999; Byrd and Rice 2006; Reid et al. 1977).

Table IV-5: RMSE of the prediction method on different compound groups.

[a]Number of points in the training set (same legends for the other occurrences)

[b] Number of points in the validation set.

	Enthalpy of formation ΔH_f^{gas} (kJ.mol^{-1})		Chemical potential of formation ΔG_f^{gas} (kJ.mol^{-1})	
Compound groups	RMSE (training set)	RMSE (validation set)	RMSE (training set)	RMSE (validation set)
Alkanes (16[a] +15[b])	0.23	0.43	0.37	0.64
Alcohols(14[a] + 6[b])	2.45	0.03	2.51	0.10
Ketones(9[a] + 3[b])	1.52	2.47	2.22	1.93
Alkenes(14[a] + 5[b])	2.14	0.40	-3.49	0.64
Alkynes(12[a] + 3[b])	2.80	2.59	1.49	1.14
Aromatics(41[a] + 26[b])	2.31	5.28	2.43	4.94
Phenols(12[a] + 10[b])	1.51	7.64	3.56	5.63
Aldehydes(10[a] + 3[b])	1.13	0.15	1.91	1.76
Overall data(128[a] + 71[b])	1.97	4.34	2.11	3.95

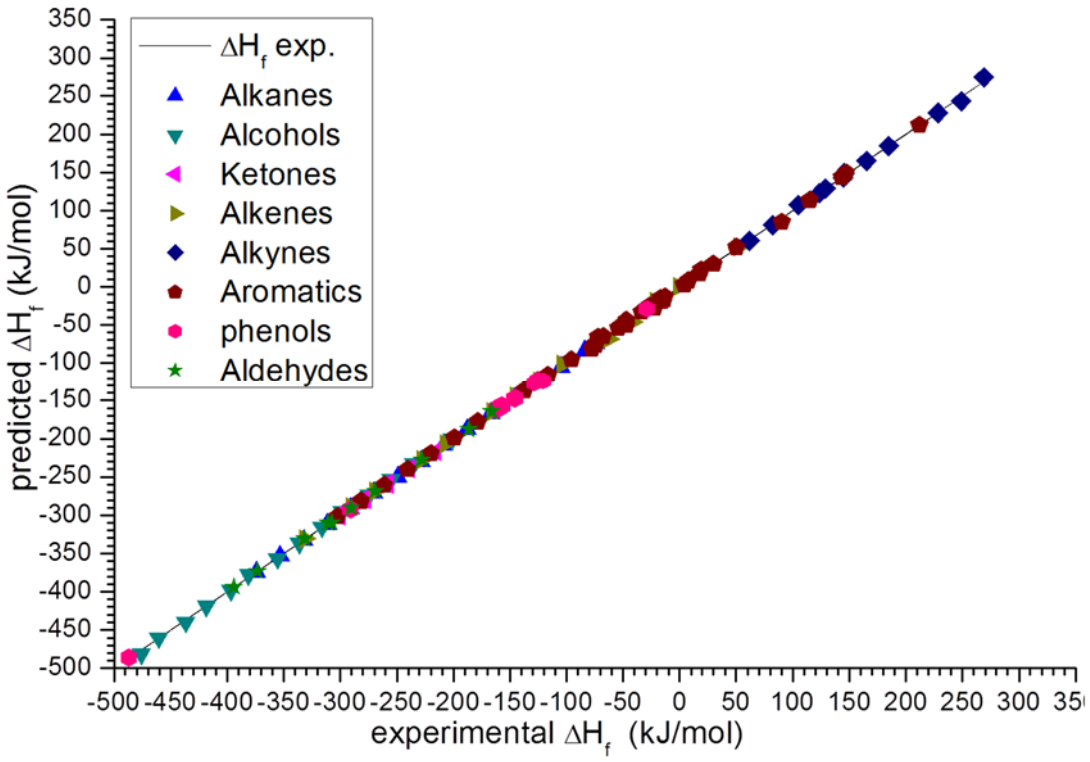


Figure IV-8 : Results of the prediction of the enthalpies of formation for molecules included in the training set (using the group equivalents method).
The line represents the equality between calculated and predicted data.

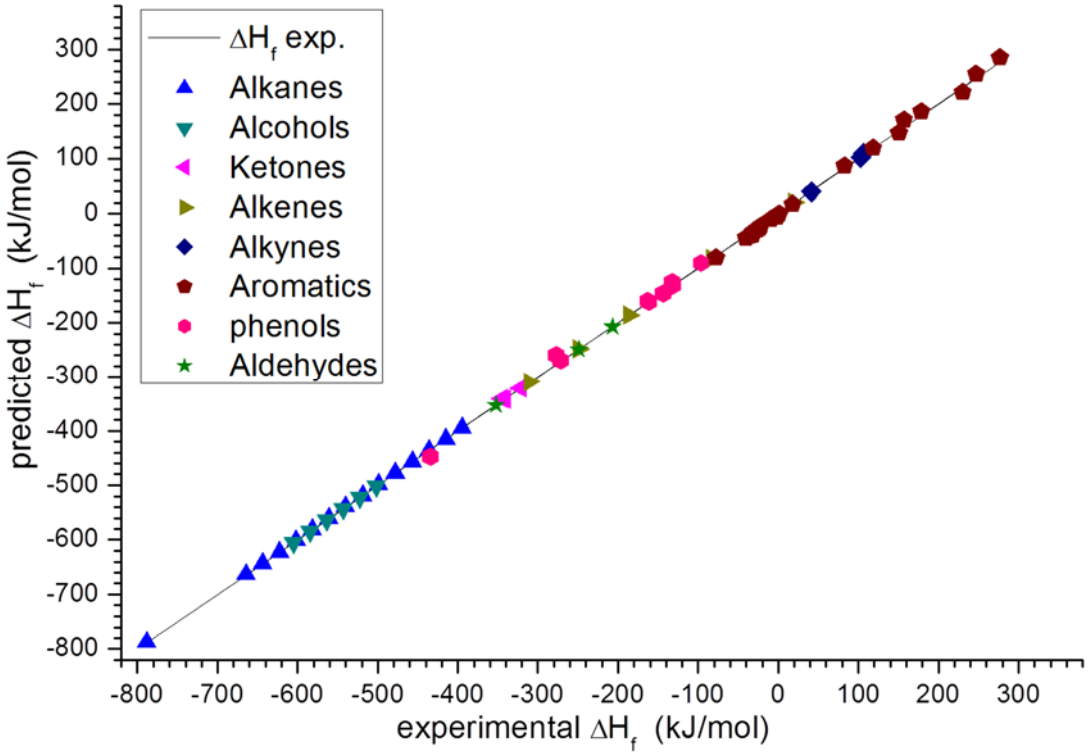


Figure IV-9 : Results of the prediction of the enthalpies of formation for molecules included in the validation test (using the group equivalents method).
The line represents the equality between calculated and predicted data.

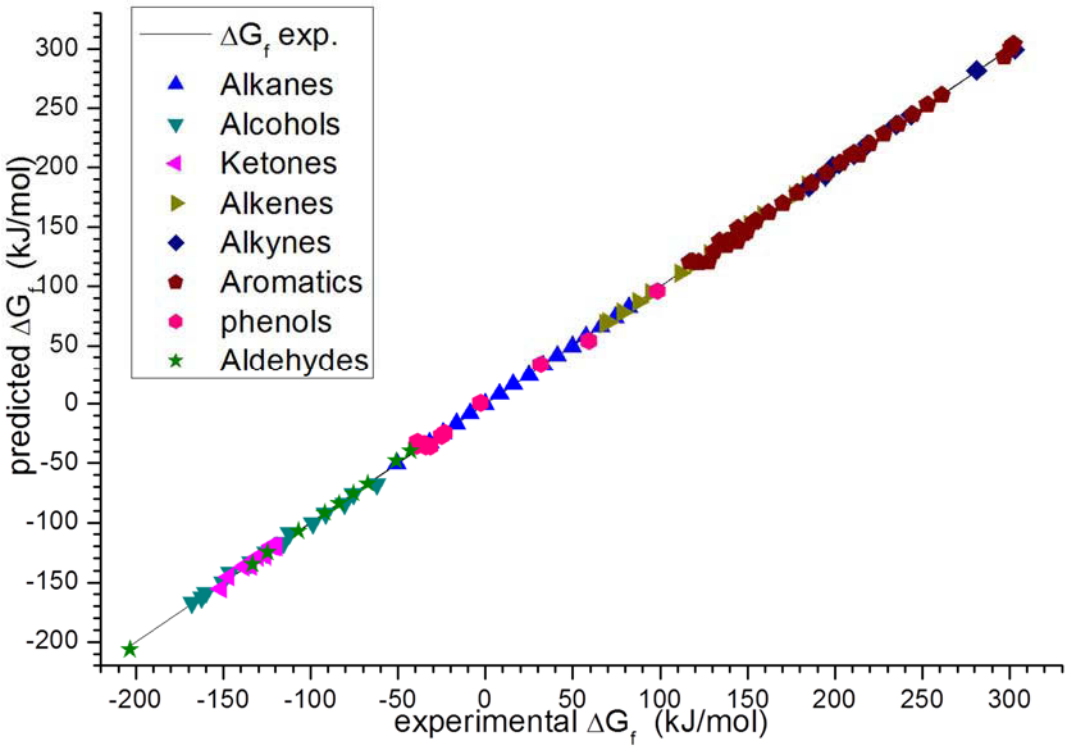


Figure IV-10 : Results of the prediction of the chemical potentials of formation for molecules included in the training set (using the group equivalents method).
The line represents the equality between calculated and predicted data.

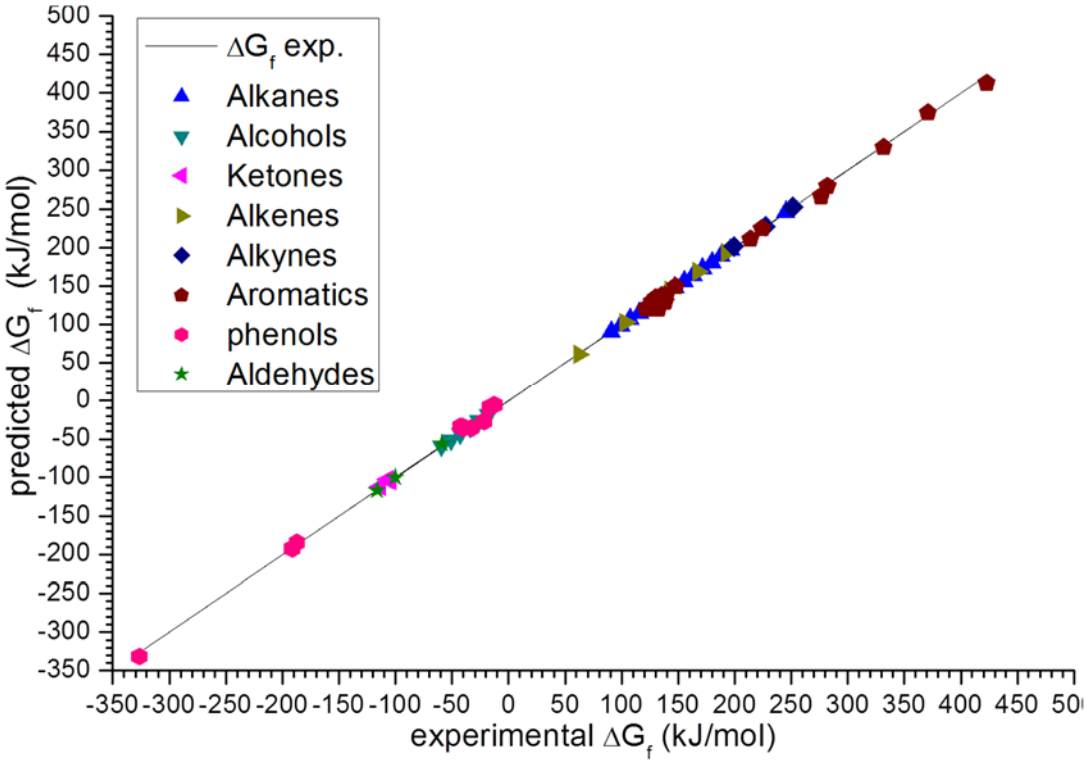


Figure IV-11 : Results of the prediction of the chemical potentials of formation for molecules included in the validation test (using the group equivalents method).
The line represents the equality between calculated and predicted data.

The regression coefficients obtained during the multi-linear regression of the enthalpy of formation and the chemical potential of formation (on molecules included in the training set) are given in Table IV-6. These coefficients will be used to predict the gas phase formation properties (especially the chemical potential of formation) of several compounds like free radicals for which there are not enough experimental data, in order to understand the redox mechanism and to predict standard redox potentials values, as discussed in section IV.3.3. However, because these redox reactions have to be treated at infinite dilution in water, it is necessary to convert the gas phase chemical potentials of formation into the corresponding values at infinite dilution in water. That is the purpose of section IV.3.2.

***Table IV-6*: Regression coefficients obtained on the training set containing alkanes, alcohols, ketones, alkenes, alkynes, aromatics, phenols and aldehydes ; using the group equivalents method.**
*In the UNIFAC groups, the symbols = and * denote respectively a double and a triple liaison between carbon atoms.*

		Enthalpy of formation regression coefficients		Chemical potential of formation regression coefficients
	Elements (*k*) *i.e.* UNIFAC groups	ε_k (in kJ/mol)	ε_k (in hartrees)	μ_k (in kJ/mol)
Group equivalents method	[CH_3OH]	303713.909	115.679	-139.054
	[CH_4]	106325.032	40.497	-49.027
	[CH_3]	104772.426	39.906	-15.691
	[CH_2]	103213.169	39.312	8.486
	[C_2H_4]	206426.985	78.624	74.633
	[CH_2=CH]	204872.639	78.032	83.470
	[OH]	198946.508	75.775	-137.321
	[CH_3CO]	402265.820	153.215	-114.647
	[CH_2CO]	400707.512	152.621	-91.293
	[C_2H_2]	203299.791	77.433	225.778
	[CH*C]	201751.067	76.843	224.284
	[C*C]	200204.725	76.254	228.631
	[ACH]	101661.511	38.721	22.176
	[$ACCH_3$]	204869.935	78.031	22.345
	[$ACCH_2$]	203308.160	77.436	46.787
	[ACCH]	201741.123	76.839	68.106
	[AC]	100099.377	38.126	32.475
	[ACOH]	299049.830	113.902	-118.495

IV.3.2. Determination of chemical potentials of formation at infinite dilution in water from gas phase data

IV.3.2.1. Use of correspondence relations between reference properties

Classically, to calculate the chemical potential of formation obtained at infinite dilution in water from that obtained in the gas phase, it is necessary to use the combination of several relations established in Chapter I and the liquid-vapour equilibrium properties in order to study the thermodynamic equilibrium between the liquid and gas phases.

In this context, the chemical potential of a given compound in a liquid mixture is calculated by,

$$\mu_i^L = \mu_i^{0L} + RT \ln \gamma_i x_i \tag{IV-14}$$

When the compound is taken in reference conditions at infinite dilution in water, Ould-Moulaye (1997) has used that the formation properties of a given compound at infinite dilution in water are calculated as,

$$\mu_i^{0,aq,\infty} = \mu_i^{0L} + RT \ln \gamma_i^{\infty} \tag{IV-15}$$

Likewise, at equilibrium conditions between the gas and liquid phases (we are on the vaporization curve where the pressure is equal to vapour pressure P_i^0), the respective chemical potentials of the liquid and gas phases are:

$$\begin{cases} \mu_i^L(T, P_i^0) = \mu_i^{0L}(T, p_{\mathrm{r}}) + v_i^{0L}(P_i^0 - p_{\mathrm{r}}) \\ \mu_i^{gas}(T, P_i^0) = \mu_i^{0,gas}(T, p_{\mathrm{r}}) + RT \ln \dfrac{P_i^0}{p_{\mathrm{r}}} \end{cases} \tag{IV-16}$$

Because the two phases are taken at equilibrium, the gas phase and liquid phases chemical potentials are equal,

$$\mu_i^L(T, P_i^0) = \mu_i^{gas}(T, P_i^0) \tag{IV-17}$$

Assuming that the term $v_i^{0L}(P_i^0 - p_{\mathrm{r}})$ can be neglected in chemical potentials calculations Ould-Moulaye (1997) has demonstrated that:

$$\mu_i^{0L}(T, p_r) = \mu_i^{0,gas}(T, p_r) + RT \ln \frac{P_i^0}{p_r} \tag{IV-18}$$

Thus, $\mu_i^{0,aq,\infty}(T, p_r)$ can be calculated from $\mu_i^{0,gas}(T, p_r)$ as,

$$\mu_i^{0,aq,\infty}(T,p_r) = \mu_i^{0L}(T,p_r) + RT\ln\gamma_i^{\infty}$$
$$= \mu_i^{0,gas}(T,p_r) + RT\ln\frac{P_i^0}{p_r} + RT\ln\gamma_i^{\infty} \qquad \text{(IV-19)}$$

The latter equation states that in order to convert the gas phase chemical potential to determine the corresponding value at infinite dilution in water for a given compound X_i, it is necessary to use two kinds of data: its vapour pressure (P_i^0) and its activity coefficient obtained at infinite dilution (γ_i^{∞}).

Regarding the determination of γ_i^{∞} that is an excess property, one can either use experimental data (if there are available) or use a predictive model like COSMO-RS. The results obtained in the two cases are compared in Figure IV-12 for a small set of compounds.

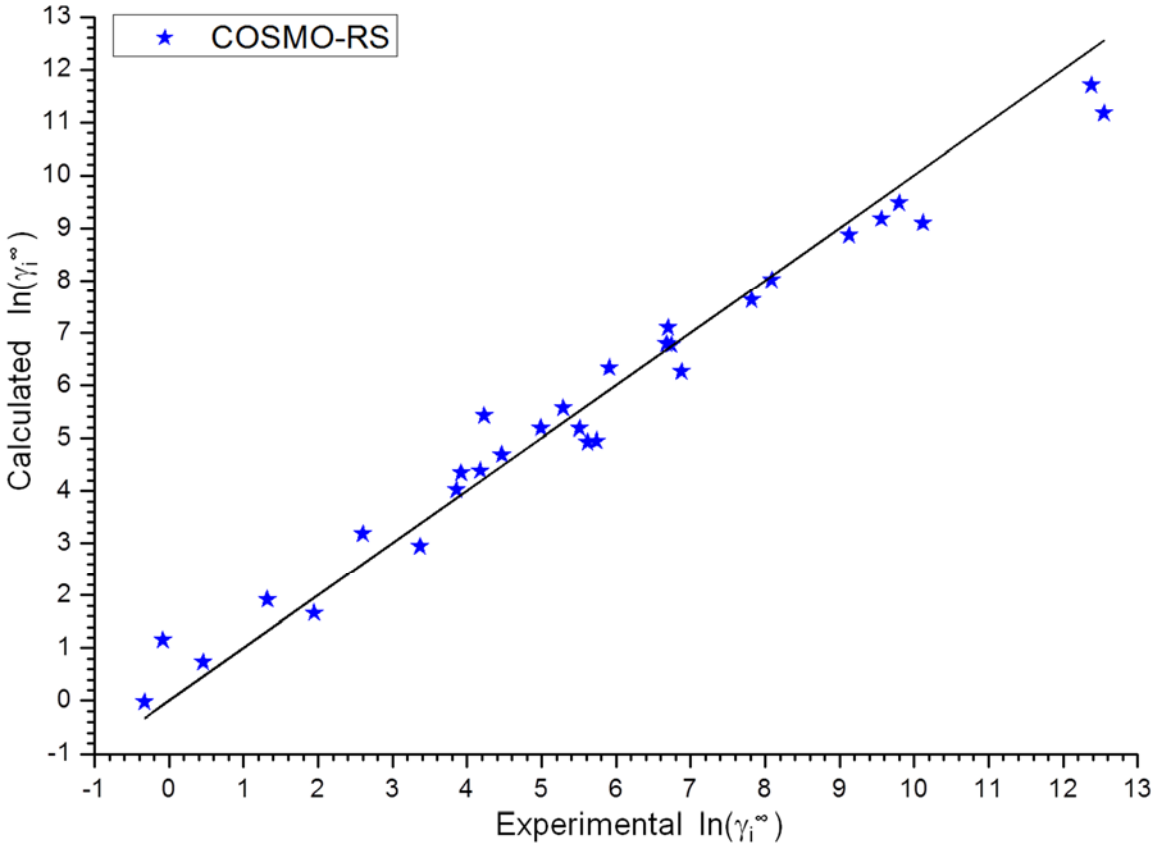


Figure IV-12 : Comparison between the experimental activity coefficients obtained at infinite dilution (γ_i^{∞}) and those calculated using the COSMO-RS method.
The line represents the equality between experimental and predicted data.

This figure confirms that the COSMO-RS model is a very efficient activity coefficient model for non-electrolyte mixtures as discussed in chapters II and III.

Because COSMO-RS seems to be very promising as a prediction tool to predict activity coefficients, it remains to determine the values of the vapour pressure, which was not the primary purpose of excess free energies models like COSMO-RS though this facility is included in its

algorithm. It is commonly accepted that vapour pressures data can be calculated from the Antoine law or several correlations if the coefficients are available in literature (Reid et al. 1977).

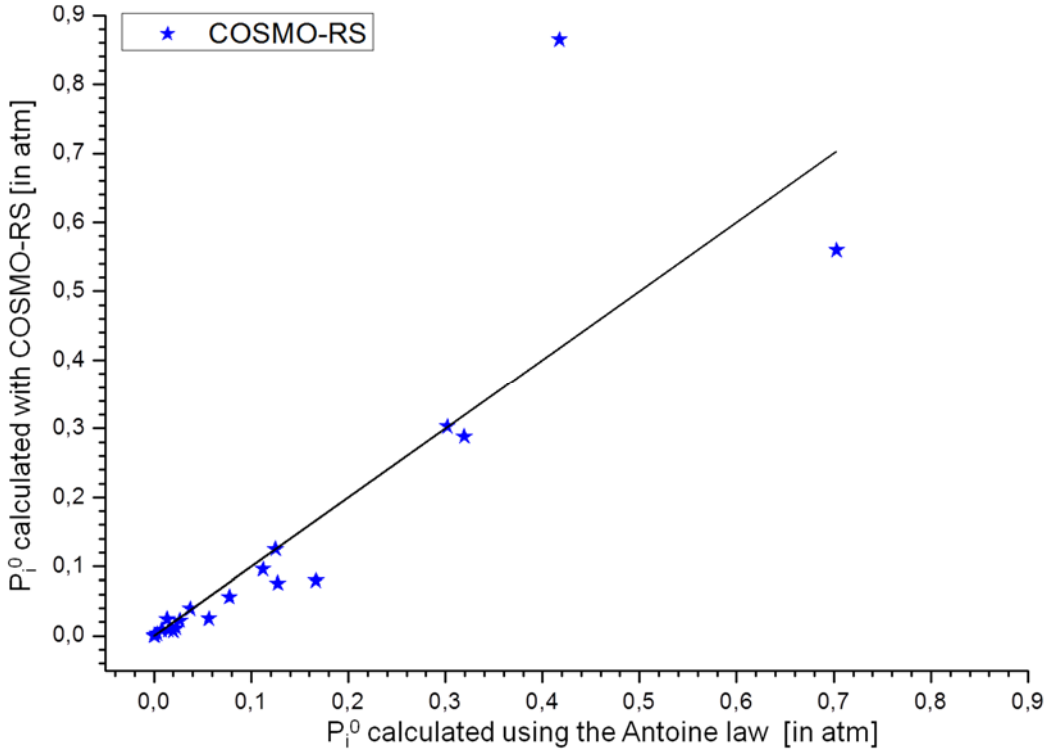


Figure IV-13 : Comparison between the vapour pressure data calculated using the Antoine law (coefficients from Reid et al. 1977) and those calculated using the COSMO-RS method.
The line represents the equality between the two kinds of data.

This figure points out that the COSMO-RS algorithm is not yet enough powerful for predicting vapour pressure data even if its predicted results can be used when there are not enough experimental data available for a given compound. For this reason, in order to use equation (IV-19) only the vapour pressure data calculated using the Antoine law will be used in the rest of this work.

IV.3.2.2. Use of the solvation properties

On the other hand, it is also possible to use the thermodynamic properties of the "solvation" process in order to determine infinite dilution chemical potentials from gas phase data (Ben Naim 1987 and 2001). The latter is only viable in terms of Gibbs energies, since it is defined as the chemical potential required for transferring a given molecule from a fixed position in the gas phase to the same position in solution (especially at infinite dilution in water, in which case the solvation process is called hydration).

IV.3.2.3. Comparison of the infinite dilution chemical potentials determined using the correspondence relations between reference properties and the solvation properties

To compare the $\mu_i^{0,aq,\infty}$ values computed using the solvation data to those computed using Equation IV-19, a set of 30 molecules (including alcohols, ketones, cycloalkanes, aldehydes, aromatics and phenols) has been selected. For all these compounds, the gas phase formation properties are already available in existing databanks like DIPPR that is included in the Simulis Thermodynamics software, as discussed in section IV.2.1. Likewise the Antoine law coefficients are available in Reid et al. 1977. The chemical potentials obtained at infinite dilution using the two processes schematized in Figure IV-14 are shown in Figure IV-15.

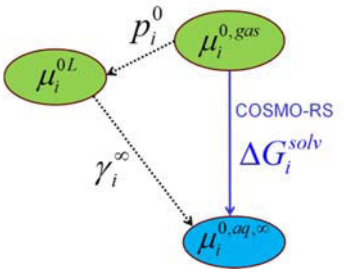


Figure IV-14 : Summary of the different ways to calculate using equation IV-19 and those calculated using the solvation data predicted with the COSMO-RS method.
The straight line represents the solvation process in the framework of Ben Naim (1987 and 2001) and the dotted lines represent the classical process described for using Equation IV-19 in which case two kinds of data are required (vapour pressure data P_i^0 and activity coefficient data γ_i^∞).

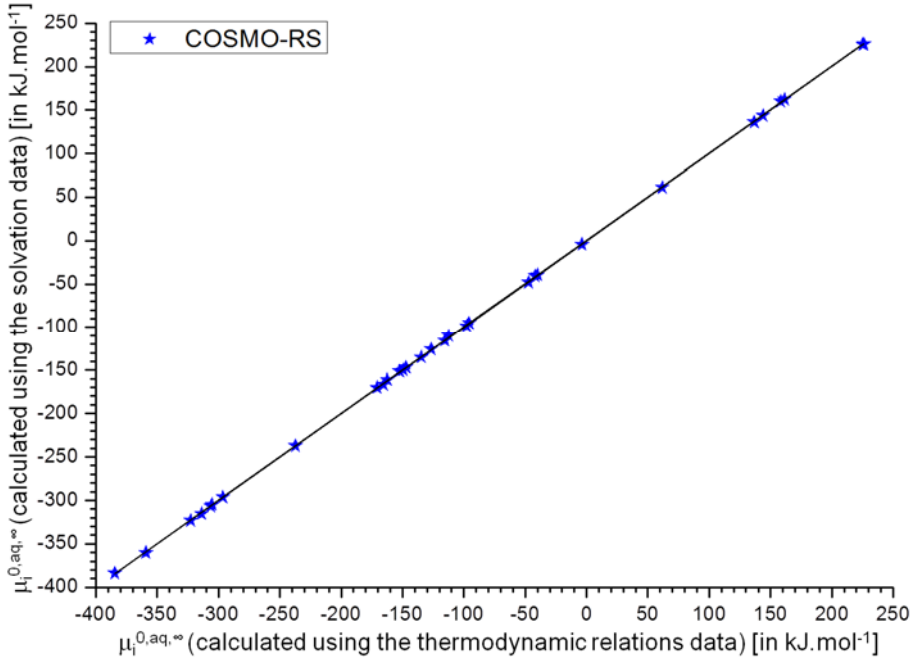


Figure IV-15 : Comparison between the infinite dilution chemical potentials data calculated using equation IV-19 and those calculated using the solvation data predicted with the COSMO-RS method.
The line represents the equality between the two kinds of data.

Figure IV-15 shows that there is an excellent agreement between the $\mu_i^{0,aq,\infty}$ data obtained using the solvation properties and those that would be obtained by the use of equation IV-19 (in which case the knowledge of activity data at infinite dilution in water, and vapour pressure data are required, as illustrated in Figure IV-14).

Furthermore, Klamt (2011) has described the COSMO-RS method as the most accurate quantum based tool for estimating solvation (especially hydration free energies). In this context, this method seems to be enough reliable and accurate to ensure the transition from gas phase to infinite dilution properties of formation in water. For this reason, the COSMO-RS method is used in the rest of this chapter to compute the solvation free energies that are then summed with the gas phase chemical potentials of formation to calculate the chemical properties of formation at infinite dilution in water.

In summary, the work described within this section is a decisive step for the development of a tool able to predict at infinite dilution (in water) the formation properties $\mu_i^{0,aq,\infty}$ of complex molecules for which experimental data are scarce like purines and pyrimidines (Ould-Moulaye et al. 2001). Hence, these formation properties data should be used to calculate (in a consistent way) the reaction properties in aqueous phase (like dissociation constants pK_a and standard redox potentials E^0). In this regards, Touré et al. (2013) have already demonstrated that COSMO-RS is efficient to predict the pK_a of complex molecules like amino-acids and peptides, and behaves like the ChemAxon and ACD/Labs pK_a prediction methods. It remains to check the performance of the COSMO-RS method as a prediction tool for redox properties data. This is the purpose of the next section.

IV.3.3. Application to the prediction of standard redox potential E^0

The standard redox potential E^0 is related to the formation properties (as discussed in chapter I), especially the chemical potentials of formation obtained at infinite dilution in water ($\mu_i^{0,aq,\infty}$). As illustrated in Table IV-7, Ould Moulaye (1998) has tested the validity of this statement by comparing the E^0 values computed from $\mu_i^{0,aq,\infty}$ data to experimental E^0 values given by Bratsch (1989).

Table IV-7: Standard redox potentials of several common half-reactions.

The values of E^0 reported by Ould-Moulaye (1997) from the chemical potentials of formation of ions at infinite dilution in water are compared here with the experimental E^0 data given by Bratsch (1989).

Reaction Numbers	Half-reaction	E^0_{calc} (V)	E^0_{table} (V)	$\Delta E^0 = E^0_{calc} - E^0_{table}$ (V)
(1)	$2H^+ + 2e^- \rightleftharpoons H_2$	0.000	0.000	0.000
(2)	$2H_2O + 2e^- \rightleftharpoons H_2 + 2OH^-$	-0.828	-0.828	0.000
(3)	$O_2 + 4H^+ + 4e^- \rightleftharpoons 2H_2O$	1.229	1.229	0.000
(4)	$O_2 + 2H_2O + 2e^- \rightleftharpoons H_2O_2(liq) + 2OH^-$	-0.20	-0.146	-0.054
(5)	$O_2 + 2H_2O + 4e^- \rightleftharpoons 4OH^-$	0.401	0.401	0.000
(6)	$O_3 + 2H^+ + 2e^- \rightleftharpoons O_2 + 2H_2O$	2.07	2.07	0.000
(7)	$O_3 + H_2O + 2e^- \rightleftharpoons O_2 + 2OH^-$	1.247	1.24	0.007
(8)	$Ca^{2+} + 2e^- \rightleftharpoons Ca$	-2.871	-2.868	-0.003
(9)	$Fe^{2+} + 2e^- \rightleftharpoons Fe$	-0.409	-0.44	0.031
(10)	$Fe^{3+} + 3e^- \rightleftharpoons Fe$	-0.016	-0.037	0.021
(11)	$Fe^{3+} + e^- \rightleftharpoons Fe^{2+}$	0.770	0.771	-0.001
(12)	$Cl_2(g) + 2e^- \rightleftharpoons 2Cl^-$	1.359	1.360	-0.001
(13)	$K^+ + e^- \rightleftharpoons K$	-2.933	-2.936	0.003
(14)	$Mg^{2+} + 2e^- \rightleftharpoons Mg$	-2.356	-2.36	0.004
(15)	$Mn^{2+} + 2e^- \rightleftharpoons Mn$	-1.182	-1.182	0.000
(16)	$Na^+ + e^- \rightleftharpoons Na$	-2.716	-2.714	-0.002

Since it can be assumed that the knowledge of the formation properties is the most important information to model standard redox potentials, to validate a E^0 prediction model using a tool based on a molecular approach, it is necessary to identify the structure of the different molecules appearing during the redox process. This kind of information is generally given on the Pourbaix diagram also called potential/pH diagram or E_h-pH diagram, as illustrated in Figure IV-16. The latter is generally determined from electrochemistry experiments and maps out the possible stable (equilibrium) phases of an aqueous electrochemical system. Predominant ion boundaries are represented by lines. As indicated in its name, this diagram was first proposed by Marcel Pourbaix (1904-1998) and can be read much like a standard phase diagram with a different set of axes. But as for phase diagrams, they do not allow for reaction rate or kinetic effects.

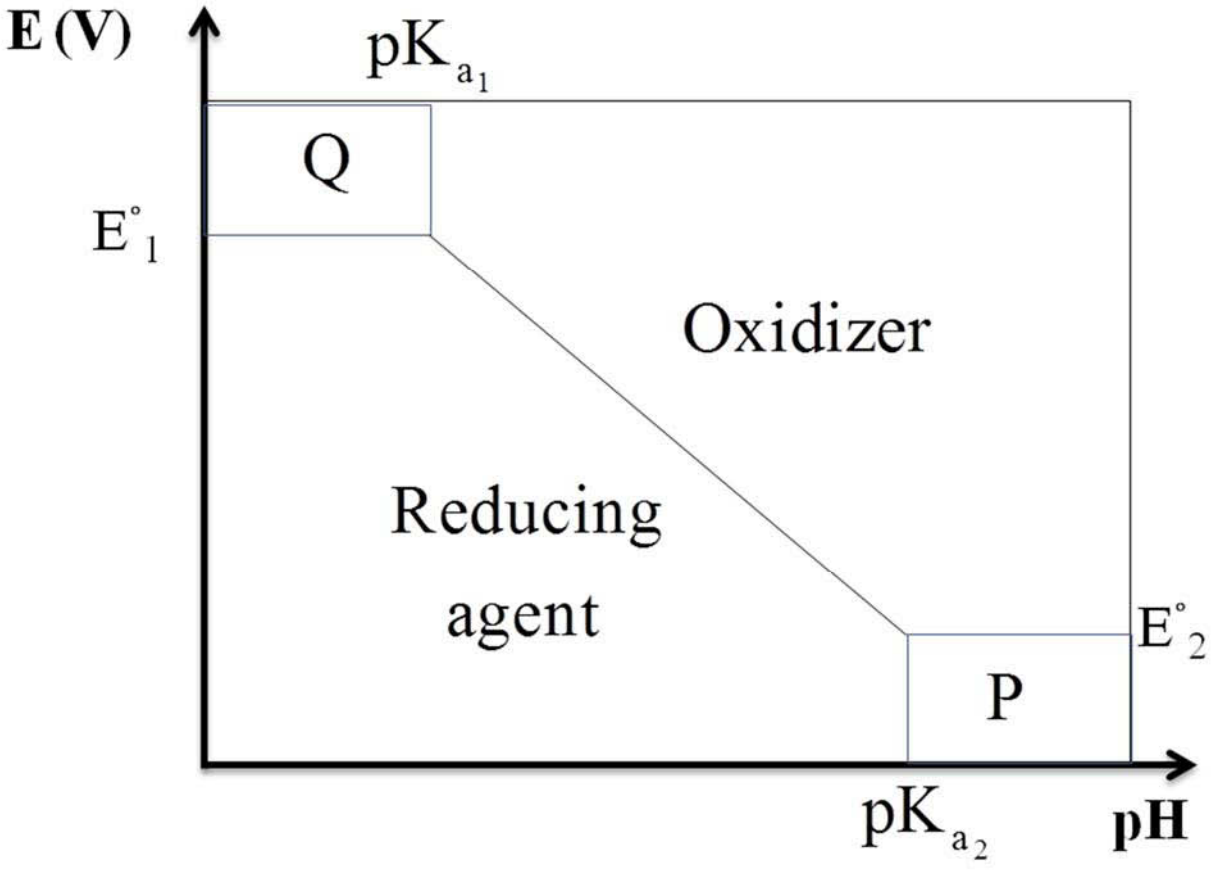


Figure IV-16 : General form of the Pourbaix diagram.

Q and P denote respectively the oxidized protonated form and the reduced deprotonated form. Vertical lines separate species that are in acid-base equilibrium. Non vertical lines separate species related by redox equilibria. Horizontal lines separate species in redox equilibria not involving hydrogen or hydroxide ions. Diagonal boundaries separate species in redox equilibria in which hydroxide or hydrogen ions are involved.
Any point on the diagram will give the thermodynamically most stable (and theoretically most abundant) form of that element at a given potential and pH condition. Strong oxidizing agents and oxidizing conditions are found only at the top of Pourbaix diagrams. Strong oxidizing agents have lower boundaries that are also high on the diagram. Reducing agents and reducing conditions are found at the bottom of a diagram and not elsewhere. Strong reducing agents have low upper boundaries on the diagram. A species that ranges from the top to the bottom of the diagram at a given pH will have no oxidizing or reducing properties at that pH.

Due to their redox properties, antioxidants are commonly used in food science as conservators, for instance in order to prevent the rancidity of fatty un-saturated acids which usually induce the degradation of foods. In biological terms, antioxidants protect the living organisms from oxidant molecules which cause the aging cell. Natural antioxidants (like polyphenols, vitamins C and E, etc.) are present in almost all of fruits and legumes. In order to optimize the use of the antioxidants in food industries, it is necessary to better understand their behaviour, *i.e.* their redox potentials in solution and semi-crystals like bubbles (E_H) even for single antioxidant or for mixtures of antioxidants. For this purpose, in the context of the Na^- project (especially, during the Master 2

internship of Trebla (2012)), it was decided to study the redox potentials of several antioxidants (including phenol, the ortho, meta and para isomers of cresol, which are the main elementary components of polyphenols) in aqueous solutions, as a first step for understanding what will happen in a complex food system like meat.

To do so, several experimental redox potential measurements were made (electrochemically) on the selected compounds in order to deduce their respective Pourbaix diagram, *i.e.* the redox mechanism and the pK_a and E^0 values. Likewise, the (difference of) E^0 values were predicted using the reverse way, from the thermodynamic analysis of the redox reactions properties at infinite dilution in water, using the COSMO-RS based prediction tools discussed earlier.

IV.3.3.1. Experimental determination of Pourbaix diagrams of phenol and cresols

The experimental measurements are done using a potentiostat (of type Solartron SI 1287) connected to a computer. 50 mL of Britton-Robinson solution in presence of KNO_3, containing 2.10^{-4} $mol.L^{-1}$ of the studied subtract (phenol or cresols in this study), is placed in the electrochemical cell which is thermostatically controlled to 24.9± 0.1°C. Three-electrode cells that are the most commonly used setup in electrochemical studies, as illustrated in Figure IV-17, are used in this study. The reference electrode (Saturated Calomel Electrode) denoted RE and the separate auxiliary (counter) electrode (platinum) denoted CE are constantly immerged in solution while the working electrode (Glassy Carbon of 3 mm of diameter imbedded in a 10 mm Teflon cylinder) denoted WE is not in solution during the launching of the acquisition process. The WE is a rotating electrode (200 rotations per minute). Because of the reactions of the solvent due to the presence of H^+/H_2 and O_2/H_2O redox couples, the baselines of solutions were measured under the same conditions without the presence of the substrate in order to subtract them from those obtained in presence of substrate. Prior to the first measurement, the oxygen dissolved in the solution is eliminated by bubbling gaseous nitrogen in the solution during about ten minutes, and 5 more minutes between each pH measurement. The voltamograms are obtained at scan speeds of 200 $mV.s^{-1}$.

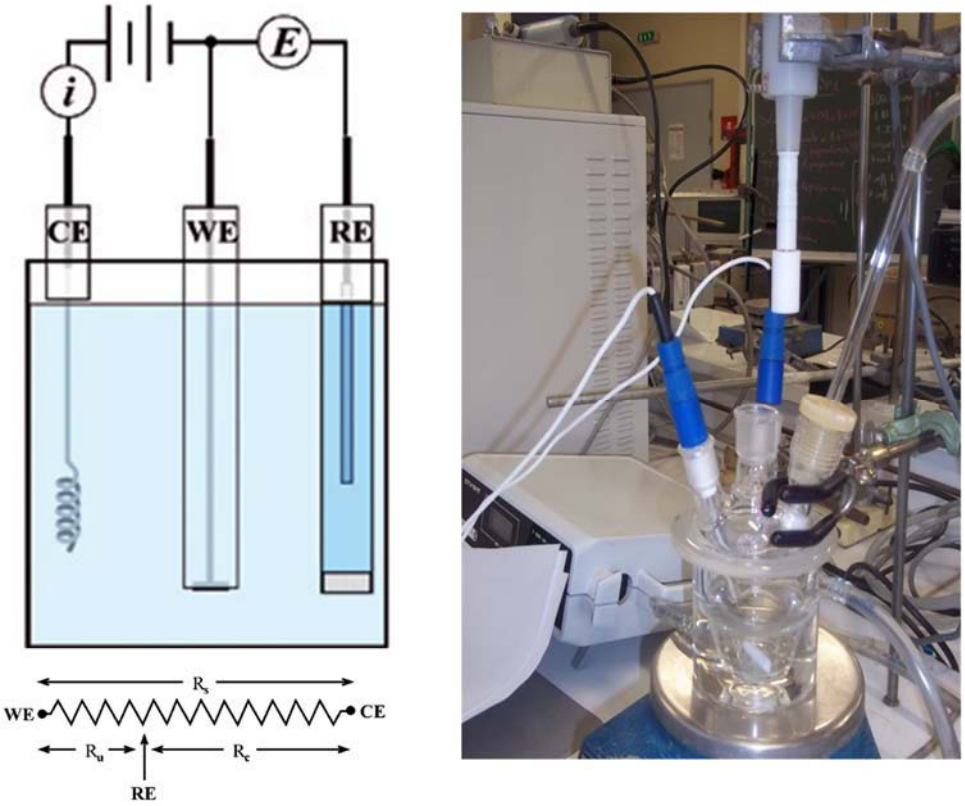


Figure IV-17 : Experimental setup for redox potentials measurements (from Zoski 2007).

Three-electrode cells are the most commonly used setup in electrochemical studies, especially when the cell solution resistance (R_s) is relatively high. In this configuration, the potential of the working electrode (WE) is still monitored relative to the reference potential of the reference electrode (RE); however, the current (i) passes between the working electrode and a separate auxiliary (counter) electrode (CE). Since no (or little) current passes to the reference electrode (RE), it approaches ideal nonpolarizability and is hence a reliable reference for potential control (E). Experimentally, the tip of the reference electrode is placed as close as possible to the working electrode in order to minimize (uncompensated) solution resistance (R_u), whereas the (compensated) solution resistance (R_c) is taken in account by the use of the counter electrode (CE).

This work is focused on the study of the redox properties of the phenol and some derivatives like the three isomers of cresol. The phenol is a well-know compound as well for its physico-chemical properties as for its redox ones (potential as a function of the pH). However, the mechanisms of redox, even for a so "simple" compound (in electrochemical viewpoint) is not still completely clear, even if some reaction pathways are proposed in the literature (Costentin et al. 2010, Bonin et al. 2010), as illustrated in Figure IV-18.

One interesting thing is the fact that using the Pourbaix diagram obtained by cyclic voltammetry, it is possible to extract the pK_a of the acid-basic couple ($ArOH/ArO^-$), but above all, the different standard redox potentials (E^0) associated to each supposed mechanism. By definition, the standard redox potential is defined at pH = 0.

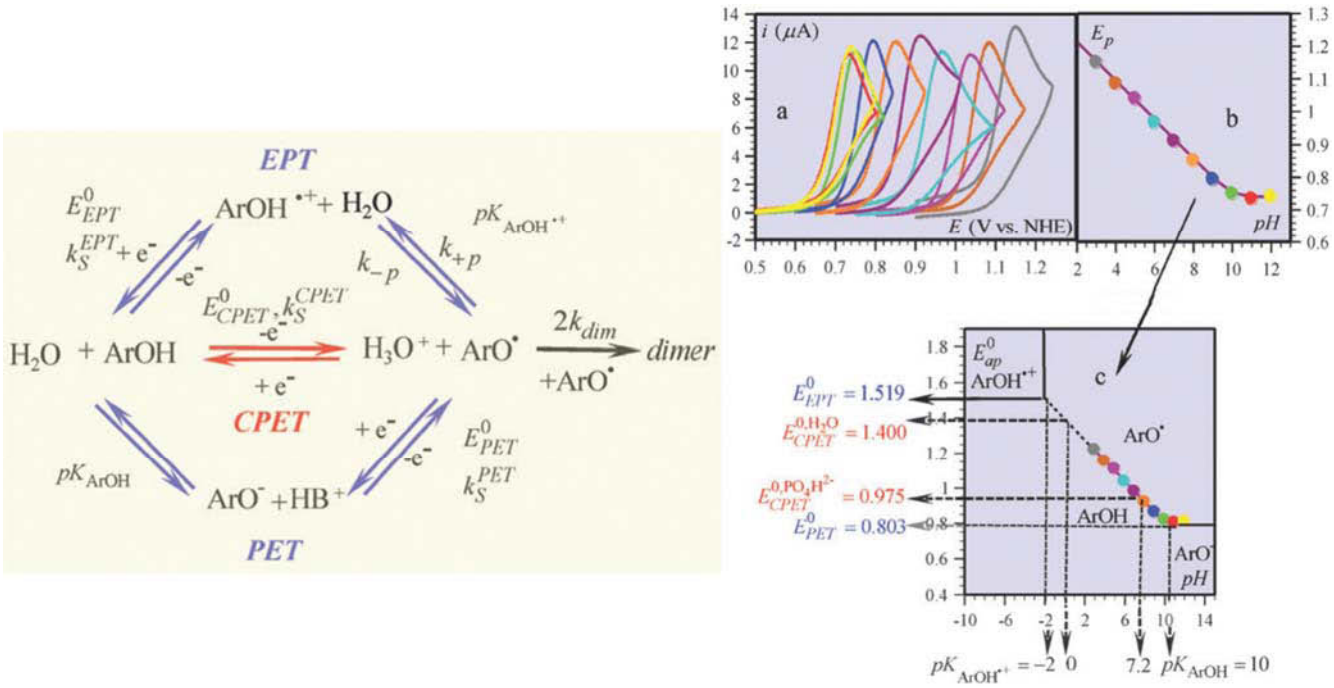


Figure IV-18 : Some reaction pathways proposed in the literature for the redox mechanism of phenol and its derivatives like cresols (adapted from Costentin et al 2010).
Here, ArOH denotes the neutral form of the compound, ArO^- the depronoted one and ArO° are the free radicals. EPT, PET and CPET denote respectively Electron-Proton Transfer, Proton-Electron Transfer and Concerted Proton-Electron Transfer.

As experimental results, the pK_a of the acid-basic couple ($ArOH/ArO^-$) and the different standard redox potentials (E^0) associated to each supposed mechanism are given in Table IV-8.

Table IV-8: Experimental pK_a and E^0 results for phenol and the cresols.
The values of $E^0_{1,mes}$ and $E^0_{2,mes}$ denote respectively the standard redox potentials of the half-reactions $ArOH + H_2O \rightleftharpoons ArO° + e^- + H_3O^+$ and $ArO^- \rightleftharpoons ArO° + e^-$. This table points out that even if they have the same formula, the three isomers of cresol do have different pK_a and E^0 values.

Compounds	$pK_{a_{2,mes}}$ (ArOH/ArO⁻) [-]	$E^0_{1,mes}$ (ArO°/ArO⁻) [in mV]	$E^0_{2,mes}$ (ArO°/ArOH) [in mV]	$E^0_{1,mes} - E^0_{2,mes}$ [in mV]
Phenol	9.81	803	1415	612
p-cresol	10.26	690	1302	612
m-cresol	10.05	762	1364	602
o-cresol	10.54	707	1343	636

IV.3.3.2. Predictive modelling of Pourbaix diagrams of phenol and cresols

The modelling tasks for determining the Pourbaix diagram is summarized in Figure IV-19 which states that it is necessary to know the molecular structures of the species appearing during the redox reactions, in order to predict their corresponding pK_a and E^0 values.

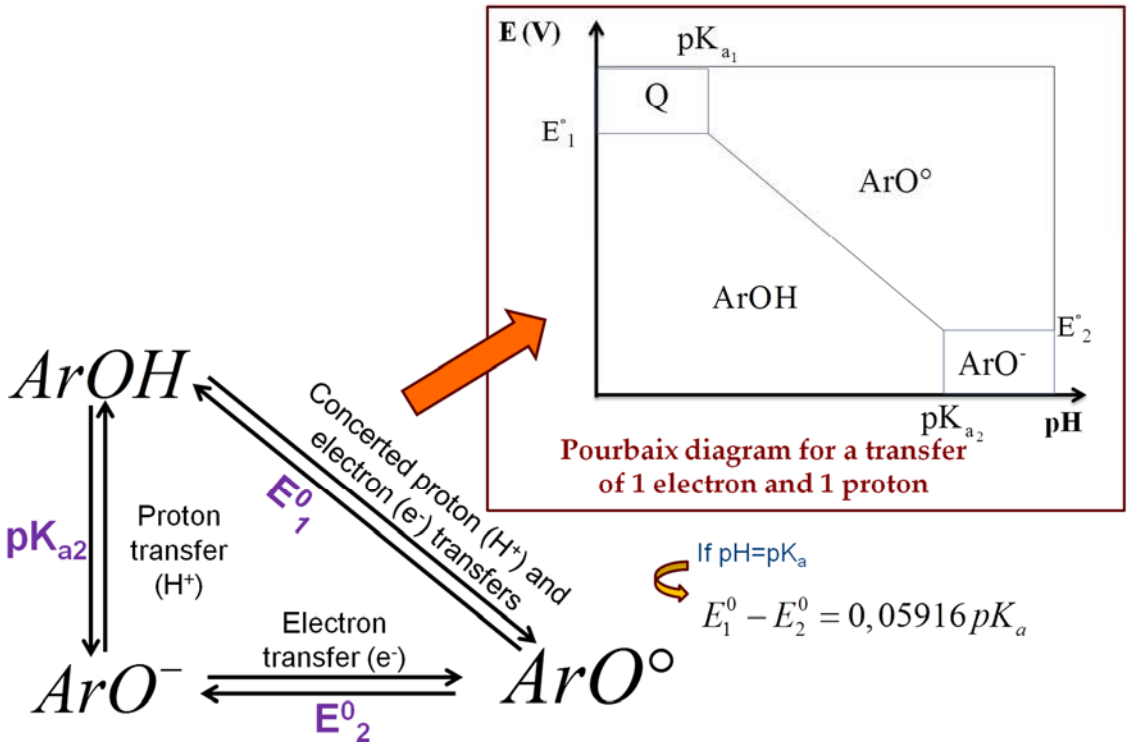


Figure IV-19 : Illustration of the modelling requirements to predict Pourbaix diagram for the studied cases of phenol and cresols (at room temperature).
To determine the Pourbaix diagram, it is necessary to be able to predict both pK_a and E^0 at the same time.

To do so, it has been demonstrated in the earlier sections that it is necessary to be able to predict (at infinite dilution in water) the chemical potentials of each species, as illustrated in Figure IV-20. For this purpose, using the prediction tools developed during this thesis, it is necessary to know the molecular structures of each species in solution. Regarding the modelling of species appearing during the dissociation reaction, which is only a proton exchange, the molecular structures of the different species are well known in the literature. However, this is not the case for radical species. Indeed, the abbreviation ArO° ubiquitously used in electrochemistry does not give any indication about the position of the radical atom. For instance in the illustrative case of phenol, the radical species denoted as PhO° can either be an alcohol or a ketone, as illustrated in Figure IV-21.

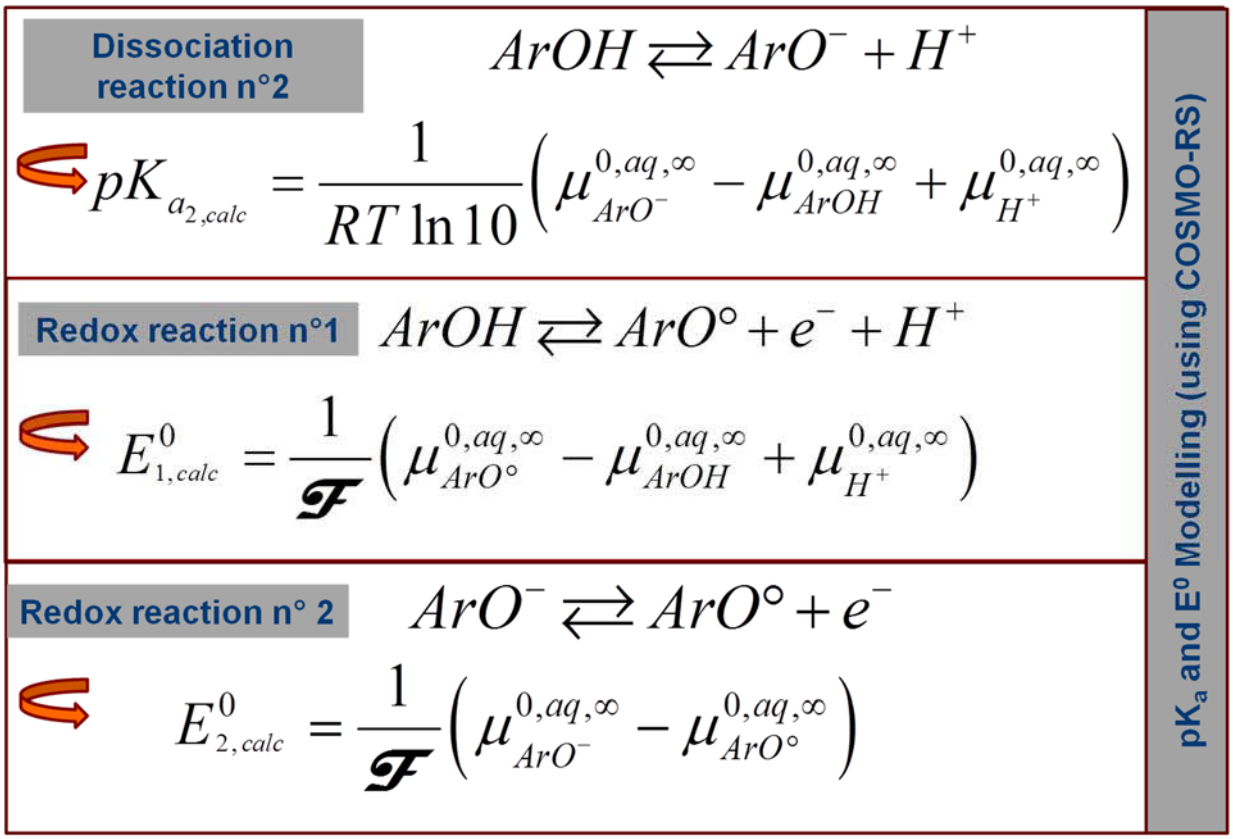


Figure IV-20 : Illustration of the E^0 prediction requirements for the studied cases of phenol and cresols.
To predict E^0 values, it is necessary to know the chemical potential of each species.

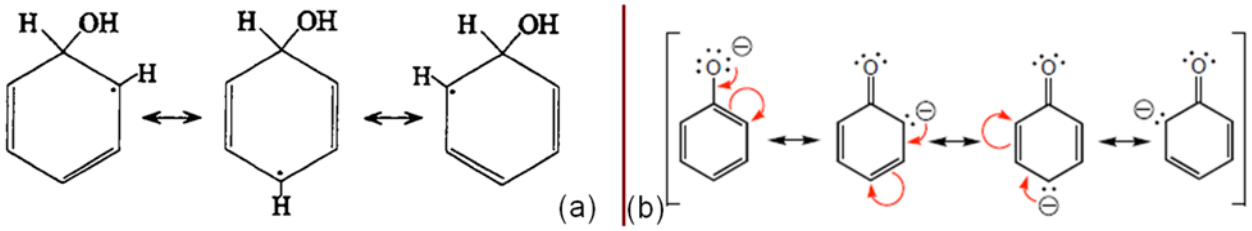


Figure IV-21 : Overview of the different forms of the radical (pHO°) that can appear during the redox mechanism of phenol.
The radical form (pHO°) of phenol can either be (a) an alcohol or (b) a ketone.

The quantum calculations (DFT energy optimizations in the COSMO framework) did not converge at all if the aromatic radicals are assumed to be ketones, contrarily to the case of alcohols where the convergence was easily reached, one can deduce that the ArO° remains an alcohol (like the reducing agent ArOH). In other words, the radical atom is one of the carbon atoms (*i.e.* not the only oxygen atom present in the molecule, as implied by the notation used).

In this context, the COSMO-RS method allows the determination of the population of each form of the radical, as discussed in chapter II. Furthermore, the formation properties prediction tool can be used to determine the chemical potential of formation of each of the different isomers of the radical PhO° illustrated in Figure IV-22.

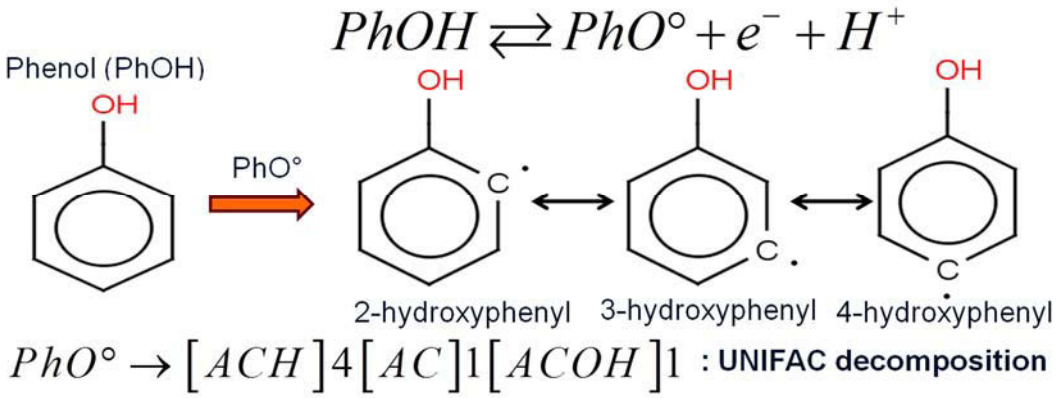


Figure IV-22 : Illustration of the three isomers of PhO° that could participate to the redox reaction.

The UNIFAC decomposition is the same for all the three isomers of the radical (PhO°). However, these isomers have different quantum calculated energies and chemical potentials values. Their respective population can also be estimated using COSMO-RS. In the gas phase for instance, there are 41.23% of 2-hydroxyphenyl, 53.72% of 3-hydroxyphenyl and 5.05% of 4-hydroxyphenyl. At infinite dilution in water, these populations become 10.85%, 81.20% and 7.95% respectively for 2-hydroxyphenyl, 3-hydroxyphenyl and 4-hydroxyphenyl.

It is possible to predict using the COSMO-RS algorithm the pK_a values of each acid/base couple, even those of free radicals that cannot be determined experimentally due to their short living times in solution, as illustrated in Figure IV-23.

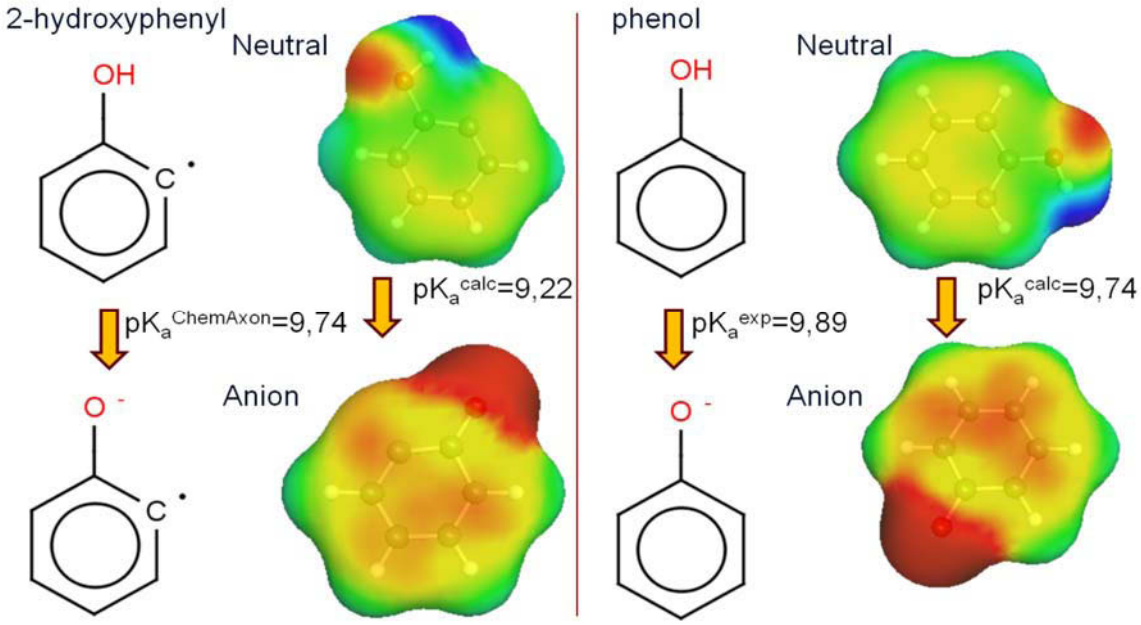


Figure IV-23 : Illustration of the pK_a calculation for phenol (right side) and for 2-hydroxyphenyl (left side) that is a radical form (PhO°) of phenol.

The COSMO-RS method can be used to determine the pK_a of free radicals, even if the experimental measurement of such data is a real challenge due to their short life time in solution. That means that the COSMO-RS method can be used (in its current state of art) to predict the vertical lines of the Pourbaix diagram.

Since COSMO-RS is an efficient pK_a prediction tool (*i.e.* vertical lines in the Pourbaix diagram), to predict the Pourbaix diagram it remains to be able to determine either horizontal lines (*i.e.* E^0 values) and/or diagonal lines. In this context, the prediction tool described in section IV.2.1 can be used to predict the gas phase chemical potentials, that will be combined with the solvation properties (at least for neutral compounds) to get their corresponding chemical potentials values at infinite dilution in water. For the free radicals and neutral compounds included in the redox mechanisms of phenol, *o*-cresol, *m*-cresol and *p*-cresol, the chemical potential results are summarized in Table IV.9.

Table IV-9: Chemical potentials results for the neutral and free radical compounds participating to the redox mechanisms of phenol and the cresols.

The.N.A. denotes not available in the used DIPPR database.

Compounds	Formula	$\mu_i^{0,gas,DIPPR}$ [in kJ.mol^{-1}]	$\mu_i^{0,gas,calc}$ [in kJ.mol^{-1}]	$\left(\Delta\mu_i^{0,gas}\right)_{calc\text{-}exp}$ [in kJ.mol^{-1}]	ΔG_i^{solv} [in kJ.mol^{-1}]	$\mu_i^{0,aq,\infty,calc}$ [in kJ.mol^{-1}]
PHENOL	C_6H_6O	-32.64	**-36.00**	-3.36	-8.18	-44.18
m-CRESOL	C_7H_8O	-40.19	**-35.88**	4.31	-6.51	-42.39
o-CRESOL	C_7H_8O	-35.43	**-34.06**	1.37	-1.40	-35.46
p-CRESOL	C_7H_8O	-31.66	**-35.64**	-3.98	-6.25	-41.89
2-hydroxyphenyl	C_6H_5O	N.A.	**-20.47**	N.A.	-3.09	-23.57
3-hydroxyphenyl	C_6H_5O	N.A.	**-24.51**	N.A.	-7.43	-31.93
4-hydroxyphenyl	C_6H_5O	N.A.	**-24.97**	N.A.	-7.53	-32.50
Salicyl	C_7H_7O	N.A.	**-7.71**	N.A.	-2.07	-9.78
3-hydroxybenzyl	C_7H_7O	N.A.	**-24.80**	N.A.	-6.70	-31.50
4-hydroxybenzyl	C_7H_7O	N.A.	**-25.57**	N.A.	-7.70	-33.27

However, since the solvation process is not yet able to accurately predict the hydration free energies of charged species (*i.e.* ions A^-) in the current state of art of the COSMO-RS algorithm, it can be assumed that these formation properties can be derived from the simultaneous knowledge of the pK_a value and the chemical potential of formation at infinite dilution of the protonated form AH. Thus, our prediction tool is not yet able to determine in a fully predictive (and thermodynamically

consistent) way E^0 values directly since redox reactions contains at the same time neutral form, free radical and ionic species.

However, as illustrated in Table IV.10 it is efficient to predict the pK_a and the differences between two E^0 values of the same species.

Table IV-10: Comparison between predicted and experimental pK_a and E^0 results for phenol and the cresols.

The values of E^0_1 and E^0_2 denote respectively the standard redox potentials of the redox couples (ArO°/ArO-) and (ArO°/ArOH), corresponding respectively to the half-reactions $ArOH + H_2O \rightleftharpoons ArO^\circ + e^- + H_3O^+$ and $ArO^- \rightleftharpoons ArO^\circ + e^-$. This table points out that even if they have the same formula, the three isomers of the cresol do have different pK_a and E^0 values.

Compounds	$pK_{a_{2,mes}}$ [-]	$pK_{a_{2,calc}}$ [-]	$\Delta\left(pK_{a_2}\right)_{calc-mes}$ [-]	$E^0_{1,mes} - E^0_{2,mes}$ [in mV]	$E^0_{1,calc} - E^0_{2,calc}$ [in mV]	$\Delta\left(E^0_1 - E^0_2\right)_{calc-mes}$ [in mV]
Phenol	9.81	9.74	-0.07	612	576	-20
p-cresol	10.26	10.06	-0.20	612	595	-17
m-cresol	10.05	9.87	-0.18	602	584	-18
o-cresol	10.54	10.24	-0.30	636	606	-30

The model is also powerful in its current development stage, to distinguish between the 3 isomers of cresol (the methyl group being in ortho-, meta- or para- position). Furthermore, in a way similar to COSMO-RS pK_a prediction LFER, by adding a constant factor that is equal about 1.2V (approximately $c'_0 = -\frac{c_0 * 1000}{\mathcal{F}} = -\frac{-120.298 * 1000}{96485} \cong 1.2$), the predicted E^0 values of phenol and cresol can be corrected to get reliable data even if the physical meaning of this correction remains to be elucidated. Even if this kind of correction is not uncommon in prediction tools (based generally on empirical parameters) this kind of correction remains out of the scope of the present study, where we are looking for fully predictive tools that are based on consistent thermodynamic description of phenomena.

The fact that the differences of E^0 values are accurately predicted instead of direct E^0 values is already far above the current expected accuracy of DFT calculations in thermochemistry application.

Now that we have proved that COSMO-RS is a reliable promising tool for predicting physico-chemical properties (like pK_a, activities, etc.) that are of great interest in complex processes like those encountered in foods science, it will be interesting to evaluate its performance as a

qualitative understanding tool for phenomena appearing in complex mixtures, especially in the context of the Na- project, as discussed in section IV.3.

IV.4. Use of COSMO-RS as a qualitative understanding tool of phenomena appearing during food processes

The σ-profiles of a given mixture can be used to provide a good qualitative description of its physico-chemical properties, or to choose the best suitable solvent for a given reaction (Klamt 2008).

Below, several applications of the σ-profiles used to provide a qualitative understanding of salting effect on a given meat (beef. or pork) in the context of the Na- project, are discussed.

A raw muscle contains water, proteins, lipids, and several minerals (less than 1%). The protein composition can be approximated as the sum of amino acids contents. Similarly the lipids composition can be approximated as the sum of fatty acid contents. An approximate composition of a muscle is given in Table IV-11.

Table IV-11: Approximate composition of a semi-membranous(SM) muscle in 100 g of raw meat for beef and pork.

These data are taken from Jensen et al. 2014.

	Compounds	S.M (Beef)	S.M (Pork)
Approximate composition of a semi-membranous muscle (S.M) / 100g of raw meat	Moisture (water content)	73.9 ± 0.4	73.6 ± 0.2
	Proteins (sum of amino acids)	18.9 ± 0.5	18.3 ± 0.4
	Lipids (sum of fatty acids)	3.2 ± 0.4	4.5 ± 0.3

Since the σ-profiles of all these elementary compounds are already available in our databases of COSMO-files, the aim of the present study is to simulate the σ -profile of the entire muscle. Indeed, because the σ-profile is an additive property, the σ-profile of the muscle is expected to be the sum of the σ-profiles of all the molecules included in its composition, pondered by their respective ratios. To do so, it is necessary to simulate in a first step the protein composition (*i.e.* the sum of amino acids) and the lipid composition (*i.e.* the sum of fatty acids).

In this context, though not recent, a detailed composition of the amino acids contained in 100 g of protein as well as the fatty acids contained in 100 g of lipids are given by Anderson in

1988. These data are rescaled to get the appropriate value of protein and lipid contents given in a more recent work (*i.e.* the data given in Table IV-11). The resulting amino acids content in 100 g of raw SM are given respectively in Table IV-12 and the σ-profile of the main amino acids are given in Figure IV-24. The approximated σ-profiles of the protein contents of the SM muscles (for both beef and pork) are given in Figure IV-25.

Table IV-12: Approximate amino-acids contents in a semi-membranous(SM) muscle in 100 g of raw meat for beef and pork.

These data are taken from Anderson (1988) after a rescaling to get the appropriate protein composition given by Jensen et al.(2014).

Compounds (Amino acids)	S.M (Beef)	S.M (Pork)
tryptophan	0.00	0.00
threonine	0.61	0.81
isoleucine	0.75	0.90
leucine	1.53	1.44
lysine	1.48	1.55
dl-cystine	0.29	0.22
methionine	0.54	0.41
phenylalanine	0.82	0.85
tyrosine	0.49	0.68
valine	1.17	1.20
arginine	1.48	1.41
histidine	0.57	0.71
alanine	1.29	1.15
asparticacid	1.68	1.51
glutamicacid	2.71	2.60
glycine	1.58	1.13
proline	1.18	0.94
serine	0.73	0.79
Total	18.90	18.30

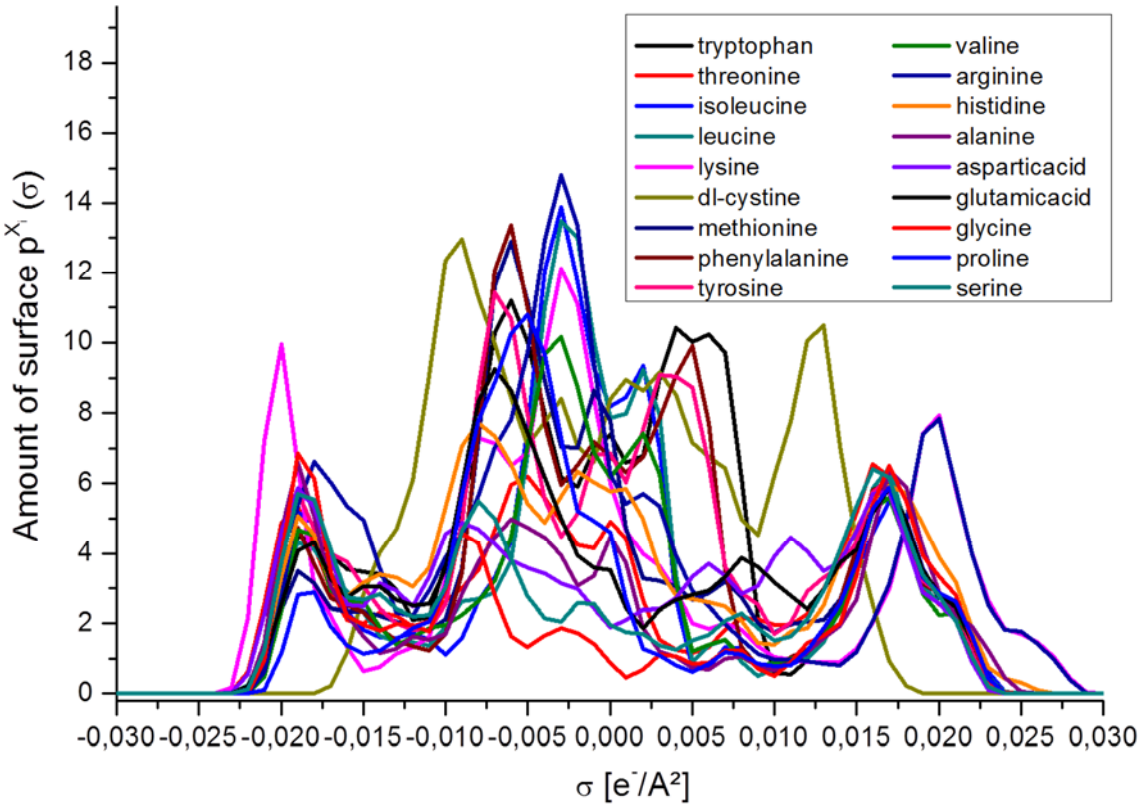


Figure IV-24 : σ-profiles of the main amino acids contained in meat.

The protein content of a meat is approximated as the sum of the amino acids.

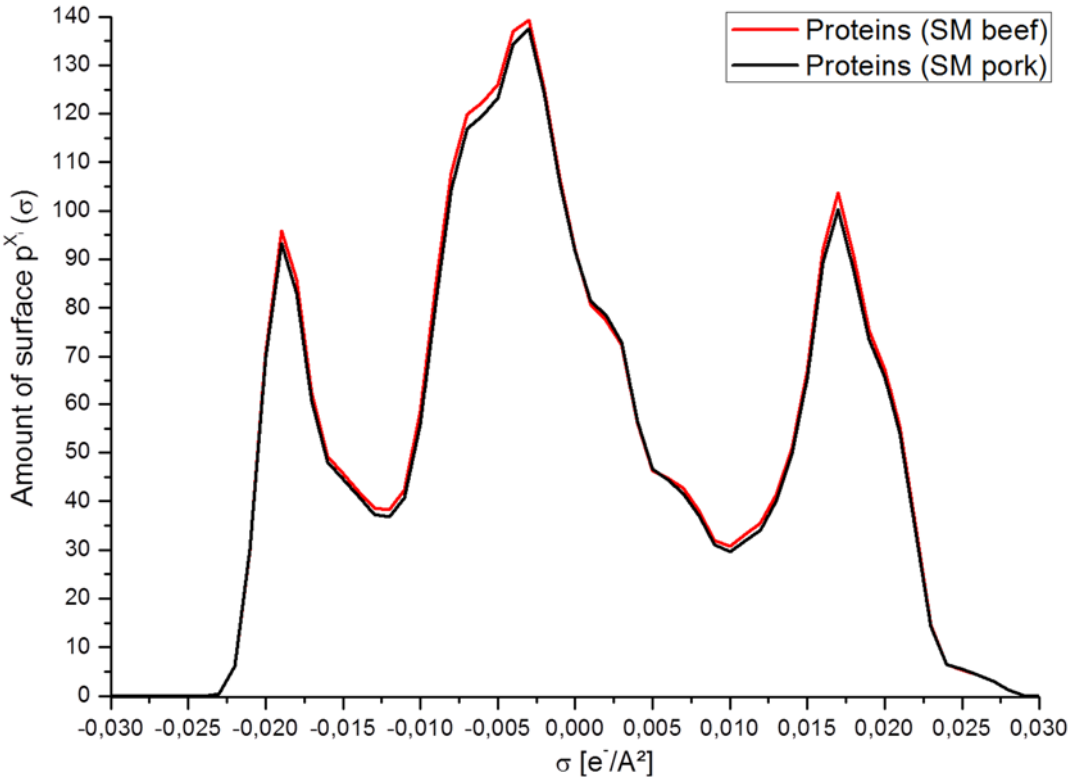


Figure IV-25 : Averaged σ-profiles of the protein content of a Semi membranous (SM) muscle of beef and pork meats.

The black line represents the beef while the red line represents the pork. This clearly confirms that the proteins content of beef and meat are identical.

Likewise, the resulting fatty acids content in 100 g of raw SM is given respectively in Table IV-13.

Table IV-13: Approximate fatty acids contents in a semi-membranous(SM) muscle in 100 g of raw meat for beef and pork.

These data are taken from Anderson (1988) after a rescaling to get the appropriate lipid composition given by Jensen et al.(2014).

Compounds (Fatty acids)	S.M (Beef)	S.M (Pork)
Dodecanoicacid	0.00	0.00
Tetradecanoic acid	0.10	0.04
Hexadecanoic acid	0.90	1.14
Octadecanoic acid	0.77	0.63
Palmitoleic acid	0.09	0.08
cis-9-octadecenoic acid	1.20	2.12
11-eicosenoic acid	0.01	0.06
(z,z)-9,12-octadecadienoic acid	0.09	0.40
Gammalinoleic acid	0.03	0.03
Arachidonic acid	0.00	0.01
Total	3.20	4.50

The σ-profiles of the main fatty acids are illustrated in Figure IV-26. The approximated σ-profiles of the lipid content of the SM muscles (for both beef and pork) are given in Figure IV-27. Finally, the resulting σ-profiles of 100g of the SM muscles (for both beef and pork) are respectively given in Figure IV-28 and Figure IV-29. The latter are compared in Figure IV-30.
Likewise, the σ-profile of a mixture containing 100 g of SM muscles (for both beef and pork) and 6.9 g of NaCl is given in Figure IV-31.

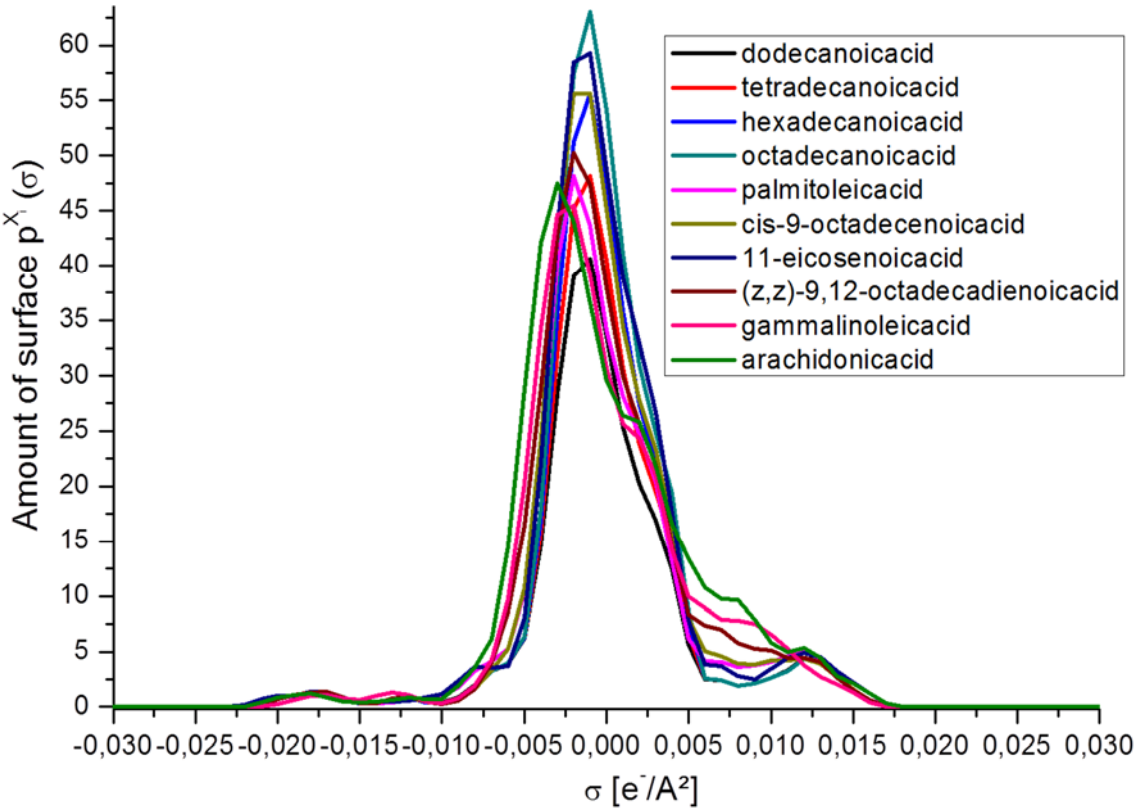


Figure IV-26 : σ-profiles of the main fatty acids contained in meat.
The lipid content of a meat is approximated as the sum of the fatty acids.

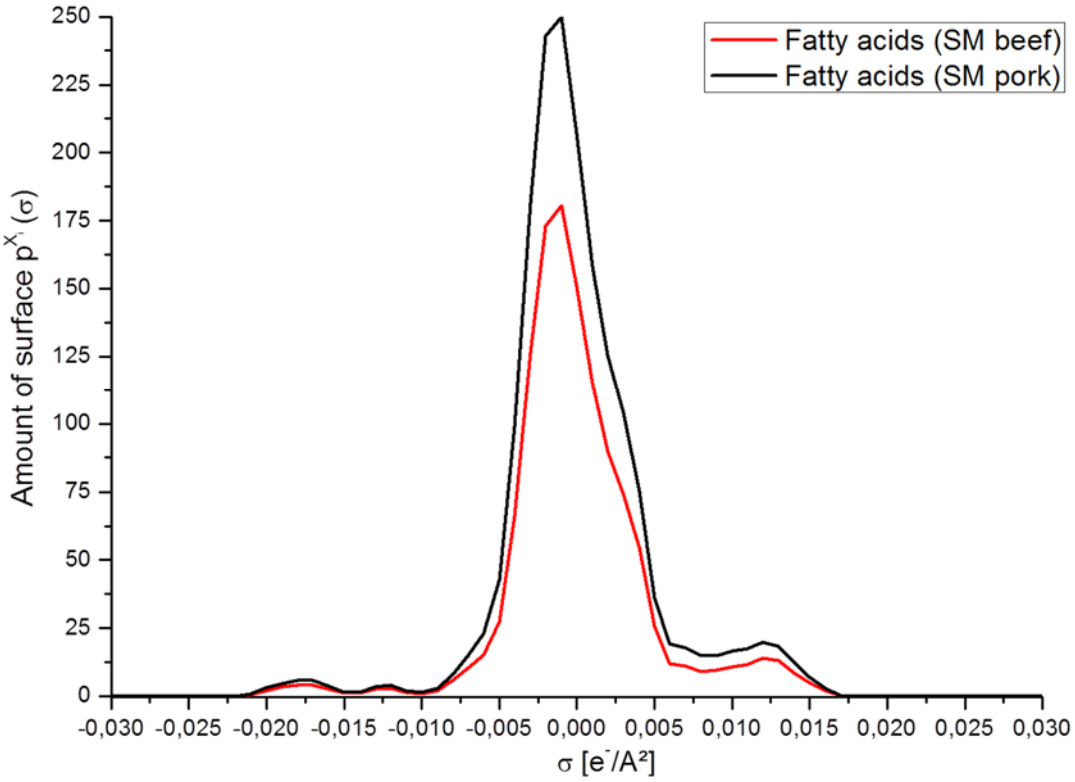


Figure IV-27 :Averaged σ-profiles of the lipid content of a Semi membranous (SM) muscle of beef and pork meats.
The black line represents the beef while the red line represents the pork.

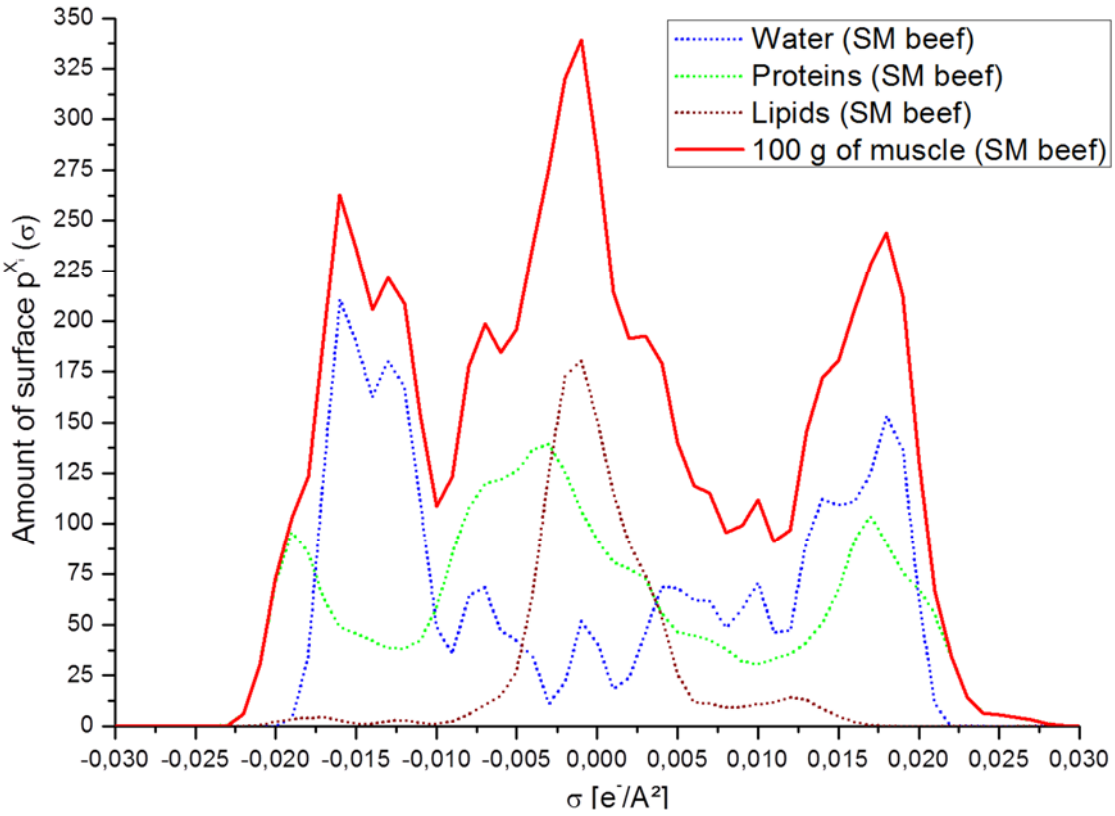


***Figure IV-28** : Averaged σ-profile of the semi membranous (SM) muscle of a beef meat.*
The water, protein and lipid contents are also given.

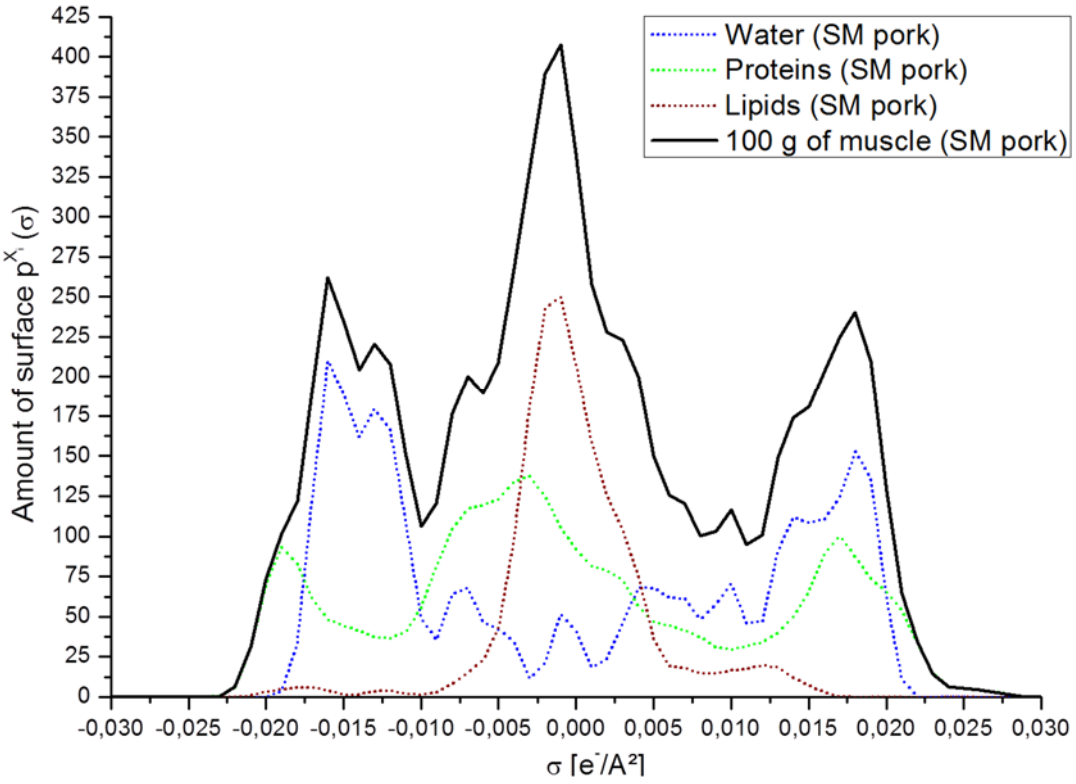


***Figure IV-29** :Averaged σ-profile of the semi membranous (SM) muscle of a pork meat.*
The water, protein and lipid contents are also given.

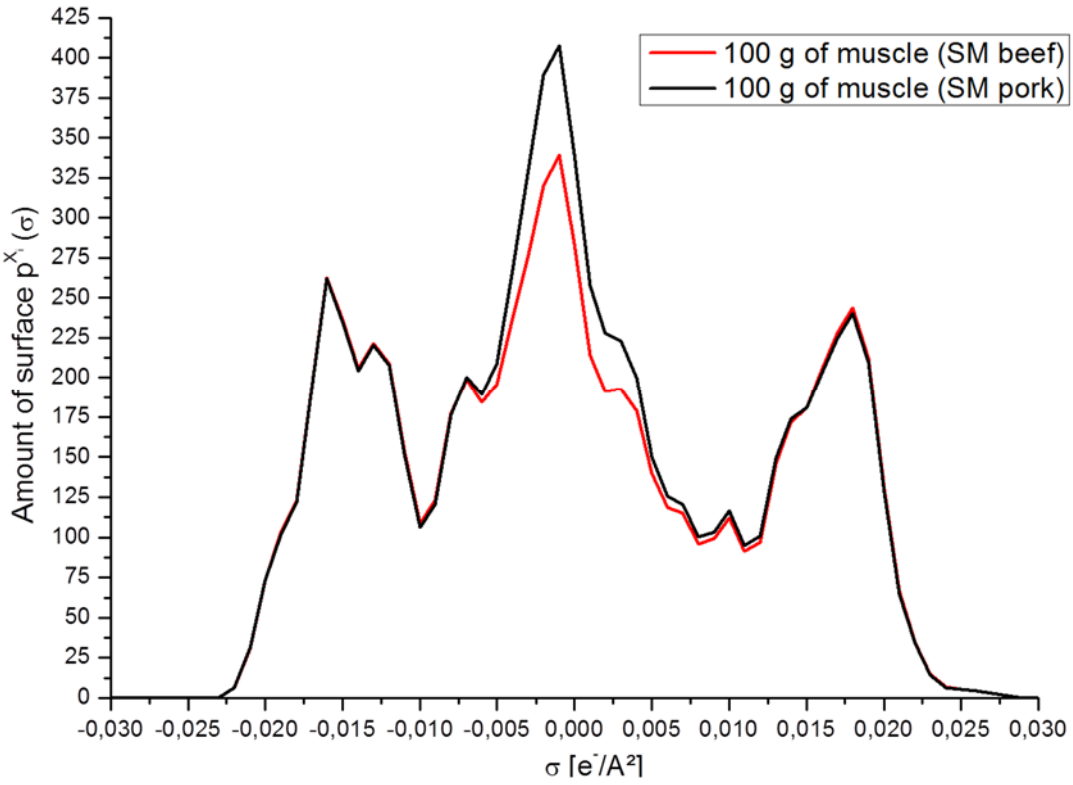


Figure IV-30 :Comparison of the approximate σ-profiles of the semi membranous (SM) muscle of beef and pork meats.

Except the lipid content, the two muscles are practically identical in terms of σ-profiles.

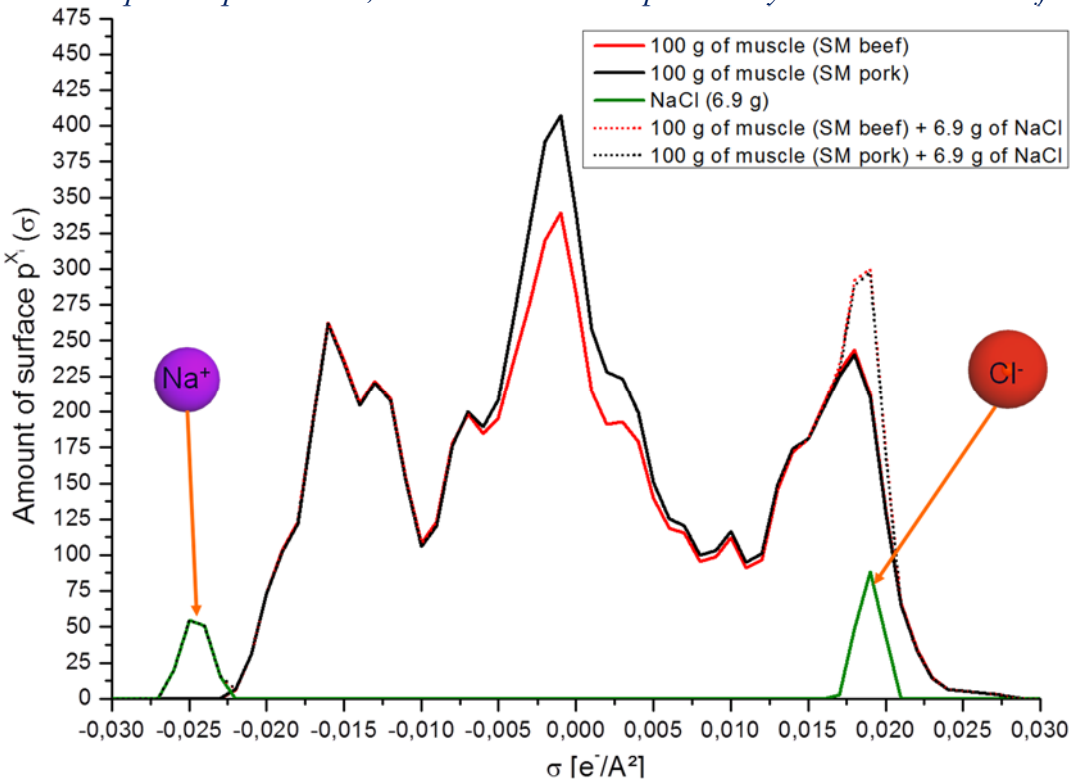


Figure IV-31 : Comparison of the approximate σ-profiles of the semi membranous (SM) muscle of beef and pork meats, before and after adding 6.9 g of NaCl.

The addition of salt slightly changes only the most polar regions of the averaged σ-profile of the meat.

Finally, the calculated σ-profiles of the raw and salted meats are used to explain qualitatively the behaviour of the SM muscle of beef and pork. In this context, Figure IV-32 shows that the SM muscles of beef and pork have the same titration curves (*i.e.* pH *vs.* volume of acid or base curves), while Figure IV-33 illustrates the fact that the addition of salt does not change the titration curve of meat.

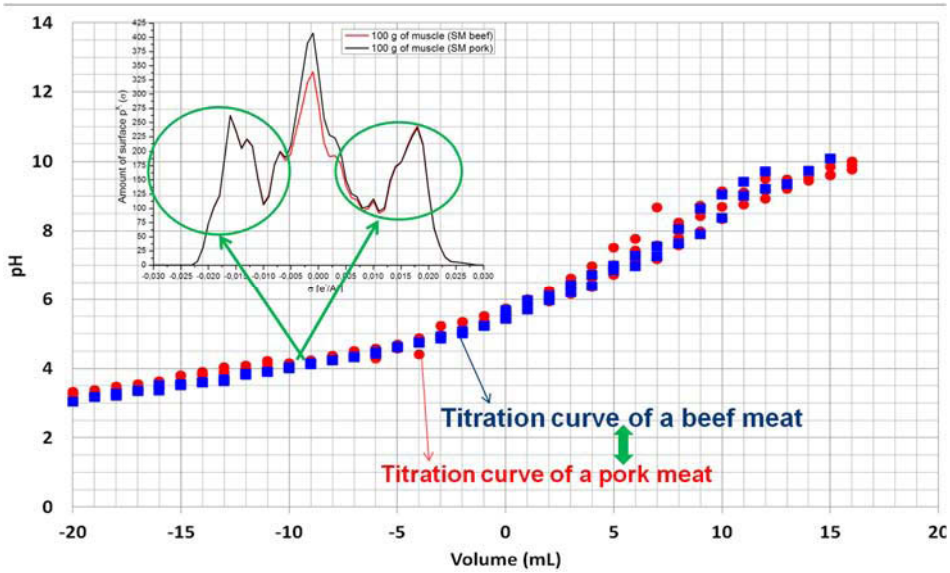


Figure IV-32 : σ-profiles as a qualitative tool for comparing the titration curves of raw meat in beef and pork SM muscles.
Because the titration curve would vary when there are more charged species in solution, the circled regions suggest that the respective σ-profiles of SM in beef and pork are identical in this case, that is why their resulting titration curves are identical.

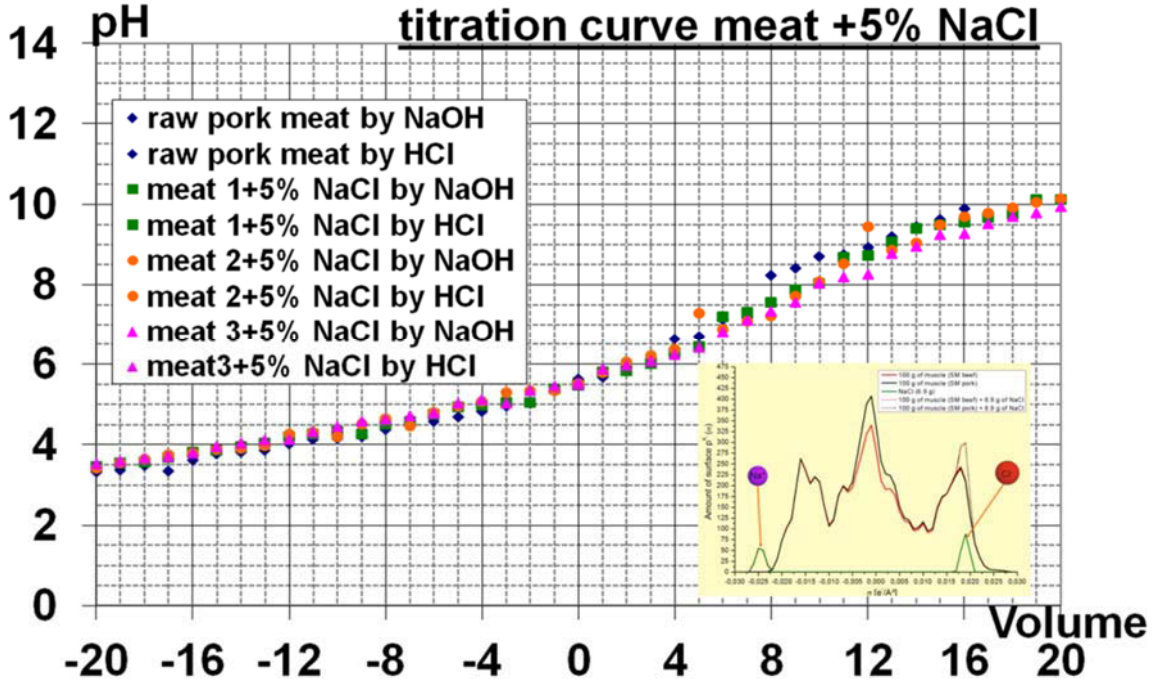


Figure IV-33 : σ-profiles as a qualitative tool for comparing the titration curves of raw meat in beef and pork SM muscles, after adding 6.9 g of NaCl.
Because the addition of salt does not change the regions of the σ-profile that are the more important for pH calculations (i.e. between -0.03 and 0.03 $e^-/Å^2$), the titration curve of meat does not change after the addition of a salt.

The sorption curve represents the evolution of the water content as function of the water activity of the meat. In this study, the experimental protocol defined by Baucour and Daudin (2000) is used to measure water sorption curves data in meats. In this context, Figure IV-34 shows that the sorption curves of a given meat do change after the addition of a salt.

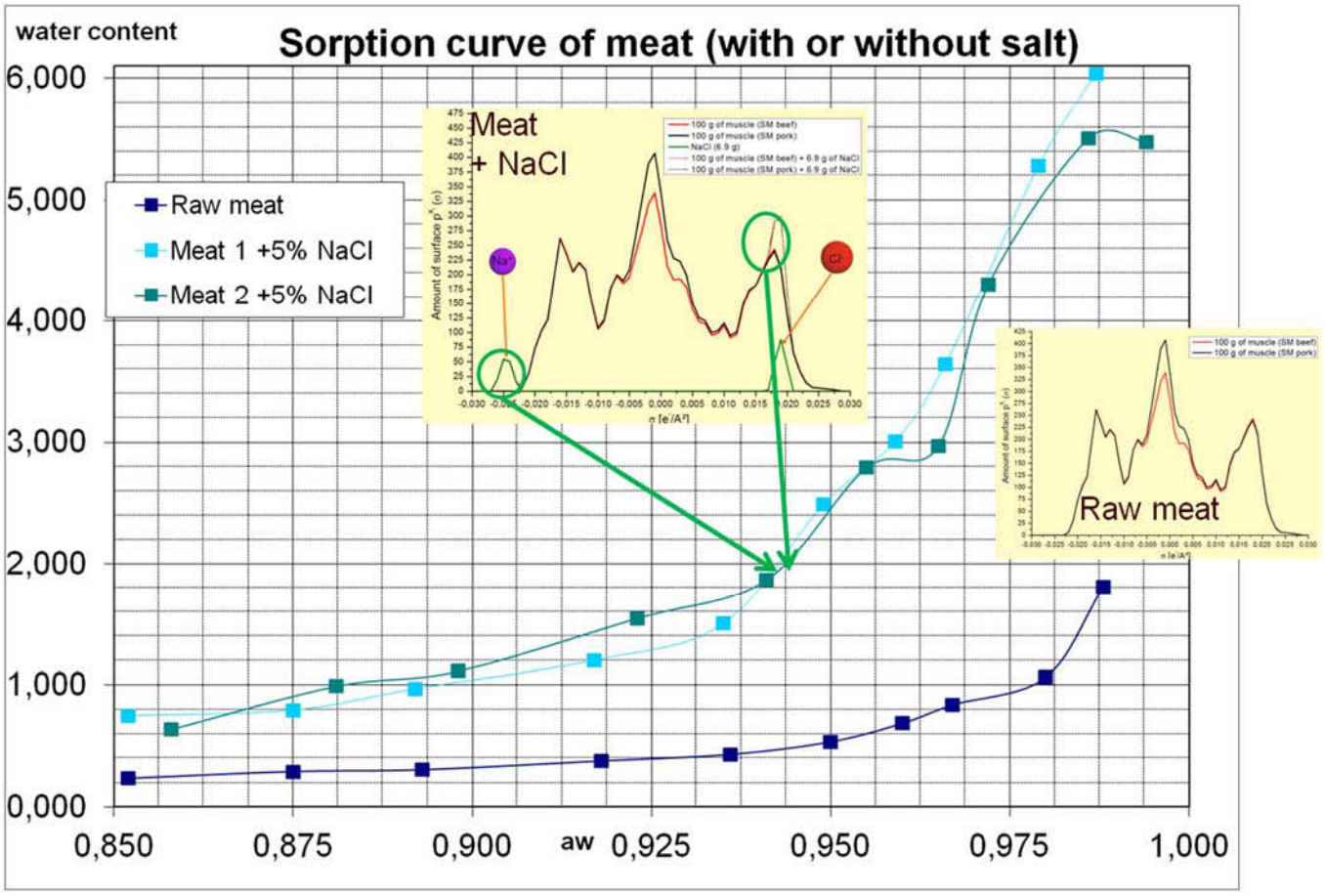


Figure IV-34 : : σ-profiles as a qualitative tool for comparing the sorption curves of raw meat in pork SM muscles, after adding 6.9 g of NaCl.

Contrarily to the previous case, the addition of a salt changes the value of the water activity of the meat. As result, the addition of a salt would change the sorption curve of the meat. This phenomenon is validated experimentally.

It has been demonstrated in chapter III that it is necessary to take into account the hydration of the ions in order to predict accurately activity coefficients and water activities data in aqueous binary mixtures. This finding remains evident in the case of complex system containing simultaneously water and salts. As illustrated in Figure IV-35, this translates to the question of the necessity to use either solid salts or dissolved salts during the salting process. This figure points out that by hydrating the salt, the latter will form a better "electrostatic" mixture with the muscle (since the σ-profile of an ideal mixture is highly symmetric). Furthermore, the hydrated ions spread a lower range of σ values than the non-hydrated ones. Similarly, the σ-profiles of several edible ions that could be used as substitute salt in meat processes are given in Figure IV-36. In this context, to

improve the symmetry of the σ-profile of the resulting mixture, the most interesting candidate that should be able to replace sodium is the potassium ion, similarly a mixture of fluorine and iodine can be used to replace the chlorine anion. In the latter case, it would be interesting to add a little amount of calcium to compensate the peak of the fluorine in the opposite σ region.

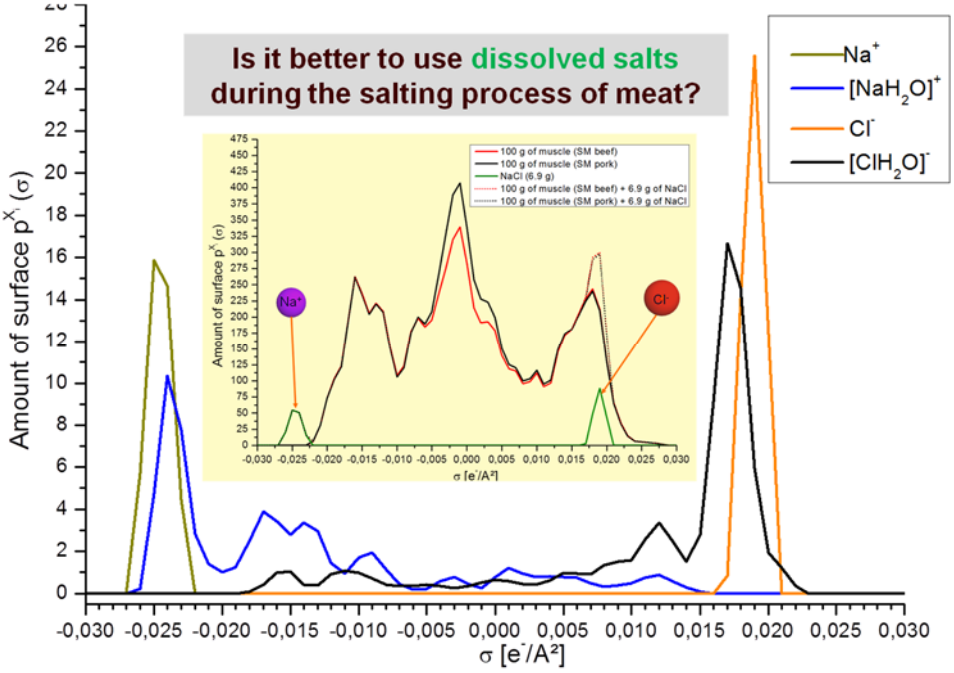


Figure IV-35 : σ-profiles as a qualitative tool for selecting the most suitable form of the salt to add during the SM muscle processing.

The black line represents the beef while the red line represents the pork.

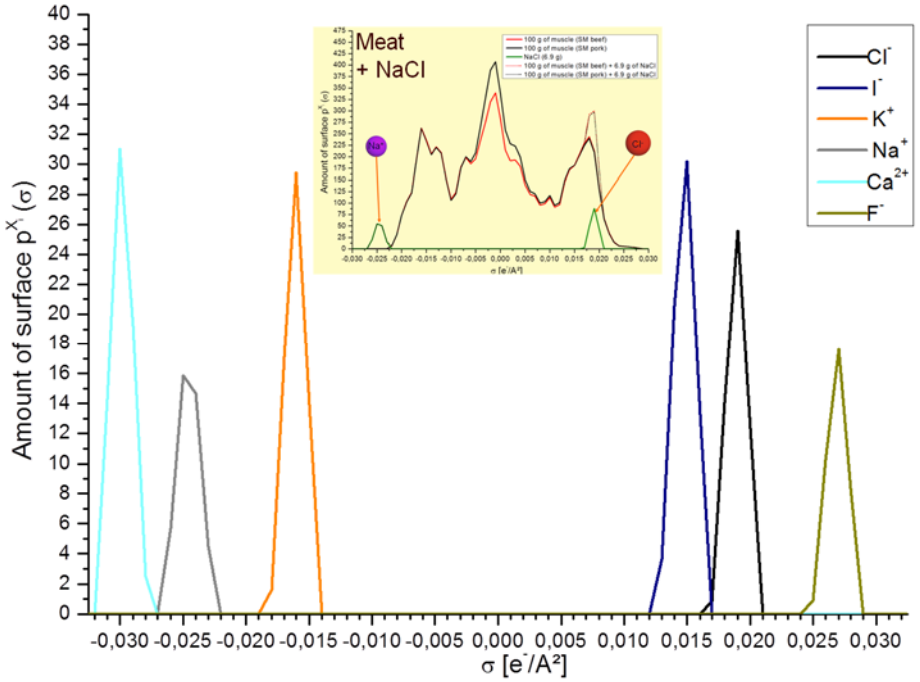


Figure IV-36 : σ-profiles of several ions that can be used to determine substitute salts in meat process.

The most suitable ionic candidates are: potassium, iodine, fluorine.

IV.5. Concluding remarks about the prediction of formation properties

In this chapter, we have seen that the COSMO-RS method is very efficient as a pK_a prediction tool. Likewise, a new prediction method was introduced and used to determine the formation properties of several compounds in the gas phase and at infinite dilution in water (using solvation properties predicted using COSMO-RS for neutral compounds).

It has been shown that to predict the standard redox potentials (required to determine the entire Pourbaix diagram) of several compounds of interest like phenol and cresols, it is necessary to identify the molecular species participating to the redox mechanisms and to be able to predict simultaneously their gas phase chemical potentials and solvation properties even if the studied molecule is a free radical or an ionic species.

Furthermore, it is possible to simulate the σ-profile of a meat to qualitatively explain several phenomena observed experimentally during food processes and determine several substitute salts that should be used in processes to replace NaCl in the context of the Na^- project.

Now that we are familiar with the most common applications of the COSMO-RS model (*i.e.* excess properties modelling) and that we have developed several extensions of this model to the prediction of excess properties of electrolytes systems and to the prediction of the formation properties (both in gas phase and at infinite dilution in water), it is interesting to summarize and revisit the different steps of the roadmap of this study. This is the purpose of chapter V.

Chapter V:Revisit of the pathway from gas phase to condensed phase using the COSMO-RS model

V.1. Introduction

The COSMO-RS method is a very efficient model for the prediction of activity coefficients, *i.e.* an excellent model for excess properties. Indeed, by considering a liquid as an ensemble of almost closely packed ideally screened molecules, the interactions of the molecules (of short range nature like electrostatic misfit interactions and hydrogen bonding) are expressed as pairwise interactions of the screening charges that are computed using (COSMO) quantum calculations tools. With this reduction of molecular interactions to surface contacts, the COSMO-RS method uses basic concepts of statistical physics to derive the chemical potential of a given molecule resulting from these SR interactions. The latter is combined with a van der Waals term and a combinatorial term to determine the final value of the chemical potential of the compound in the mixture that is used to compute activity coefficients data. In chapter III, it has been shown that when combined with the PDH model and a chemical treatment of the hydration of electrolyte ions, the performance of the COSMO-RS model can be extended to the prediction of activity coefficients of aqueous-electrolytes binary mixtures. The resulting model, denoted as COSMO-RS-PDHS, assumes at this stage a constant hydration of ionic species. This description can be improved if it is possible to take into account the variable hydration of ions that would require the availability of thermochemistry data like reaction and formation properties data. In chapter IV it has been shown that these kinds of data are not always available in the literature. To overcome this lack of data, we have introduced a new method for predicting formation properties in the gas phase (using quantum calculations tools) and at infinite dilution in water (using the solvation process introduced by Ben Naim and available in the COSMO-RS algorithm). This justifies why COSMO-RS is an excellent pK_a prediction tool, and a promising tool for redox potential predictions. In the latter case, because the oxidizer and the reducer do not have necessarily to be the same molecule, it is necessary to be able to predict simultaneously and consistently the formation properties of neutral species (including free radicals) and charged species that are not so far taken into account efficiently in the COSMO-RS algorithm.

In other words, COSMO-RS which is a model originally developed for liquid mixtures, is a very efficient predictive model for excess properties (excess enthalpies and entropies and consequently excess chemical potentials and activity coefficients in liquid mixtures). The tentative work developed in this chapter aims at extending the COSMO-RS approach to consistently predict formation properties in the gas phase and the changes of

enthalpies and entropies between gas phase and condensed (liquid) phase. This has been shown to be feasible in chapter IV.

To our knowledge, all the statistical thermodynamic equations currently used in the COSMO-RS model assume a molecule to behave like a liquid or a pseudo-liquid. It is also assumed that the translational motions cancel out during the solvation process (since the molecule is also assumed to be immobilized). This simplified assumption is justified to determine excess properties; however it must be reviewed when the target is to determine absolute thermodynamic properties to describe the transition from gas phase to condensed phase. This is the goal of the present chapter, which aims at revisiting the roadmap of the whole study and to propose a complete and thermodynamically consistent pathway from gas phase to real solutions. For this purpose, we propose in Figure V-1 a study scheme that includes 6 steps, which are detailed throughout this chapter.

All these steps involve different scientific areas, including thermodynamics, statistical physics, quantum physics and electrostatics. A brief description of the main goals of each discipline is shortly given in section V.2.

Before starting a detailed analysis of Figure V-1, it is mandatory to have in mind the primary goals of each discipline. As discussed below, quantum calculations provide energy data while statistical physics enables to determine "statistical" entropy data. Taken alone, any of these data (energy or entropy) is not enough to fully describe the macroscopic behaviour of a system that is provided by thermodynamic quantities. Indeed, to explain the physical significance of the statistical physics parameter β that is an intensive property (whose dimensions are those of a reciprocal energy), it was necessary to establish the link between this parameter and the thermodynamic temperature T using a simple proportionality, *i.e.* $T=1/(k\beta)$ where k is the universal Boltzmann's constant ($k=1.38066\times10^{23}$ J.K^{-1}) (Prausnitz et al. 1999). Furthermore, the laws of thermodynamics state that Absolute Zero (where the entropy S reaches its minimum value) cannot be reached using only thermodynamic means. A system at absolute zero still possesses quantum mechanical zero-point energy, the energy of its ground state. The kinetic energy of the ground state cannot be removed although, in the classical interpretation, it is zero and the thermal energy of matter vanishes (Irikura et al. 1998).

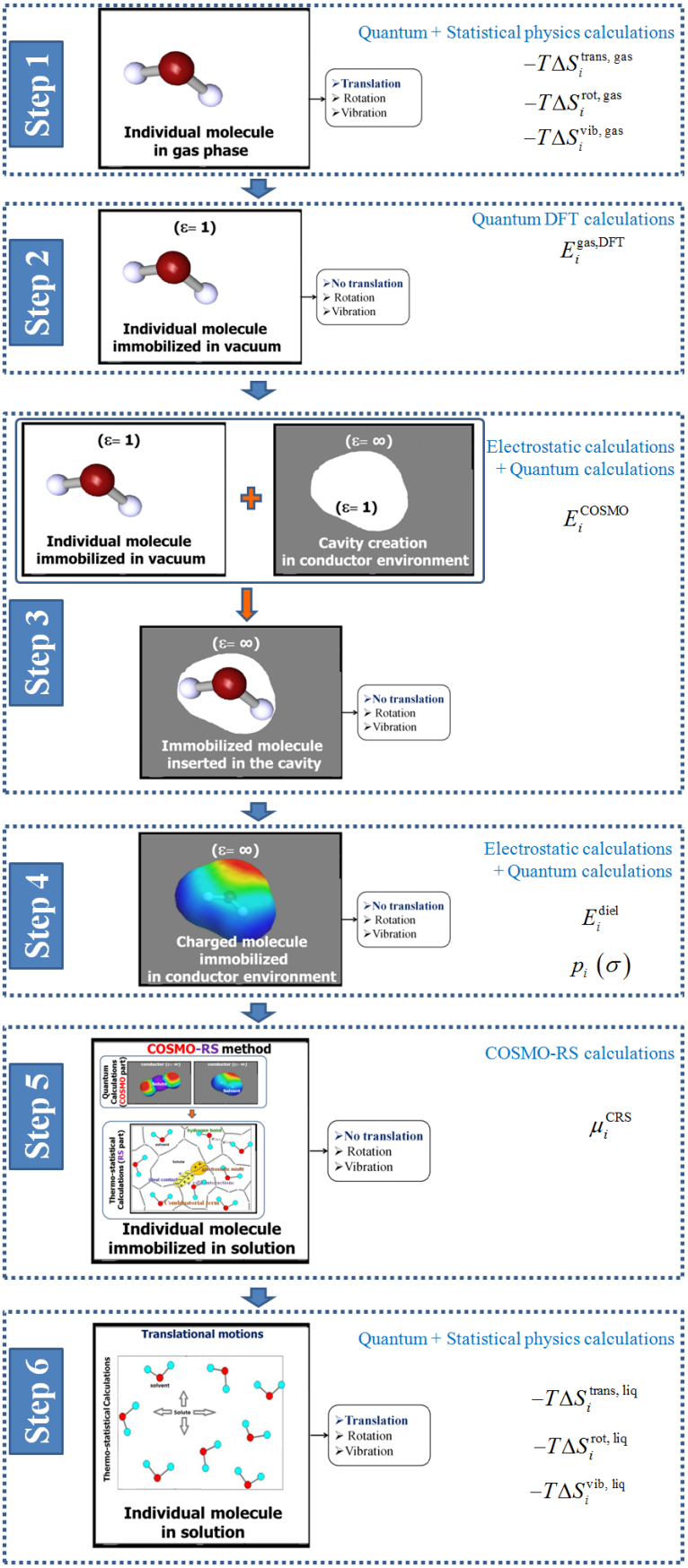


***Figure V-1**: The pathway to go from gas phase to condensed phase.*

The main outlines of the pathway described in Figure V-1 are shortly summarized below.

In Step 1, all the thermodynamic relations allowing the calculation of the entropies are already available in the literature; the reference equations established by Irikura et al. (1998) have been used in the present work. These relations are summarized in section V.3.

The quantum calculations tools (that are used in Steps 2 and 3) are developed to compute energy data. To get chemical potentials from quantum calculations, it is necessary to perform external vibrational frequencies calculations, and then to deduce the value of the ZPE, before using the relations established in Step 1 to determine the enthalpy of formation and then the chemical potential of formation in the gas phase. It must be kept in mind that, the accuracy of the calculated chemical potential will strongly depend on the quantum level used and it has been shown that this kind of calculations is challenging for large molecules, like those involved in foods and biological systems.

The chemical potential of a molecule in a mixture is computed from the information resulting from quantum COSMO-calculations, namely from the screening charge density σ that is converted to a histogram, *i.e.* the σ-profile, as pointed out in step 4. In Step 5, this histogram, that is a distribution, is used in the statistical thermodynamics calculations to determine the chemical potential of each segment and then to deduce the corresponding value for the entire individual molecule. Thus, Steps 3, 4 and 5 have already been introduced in this study (especially in Chapter II).

Finally, the analogue of the relations used in the gas phase is used to compute the entropic contributions (especially the translational one) in step 6. However, even if we have the chemical potential of the individual molecule in the COSMO-RS approach, its analogue in the gas phase still remains to be determined before performing the transition between the gas and the condensed phases. This is the main originality of the present work.

Indeed, we have decided to introduce a virtual state (called COSMO-vac) to determine this contribution resulting from the statistical thermodynamics relations for interacting segments, as discussed in details in Section V.5. For this purpose, it is necessary to earlier introduce (in section V.4) several concepts of statistical thermodynamics for interacting particles (*i.e.* surface segments in the present case).

V.2. About thermodynamics, statistical physics and quantum physics

The following discussion is essentially inspired from the reference work of Prausnitz et al. (1999).

The goal of classical ***thermodynamics*** is to describe macroscopic properties in terms of specific physical quantities as well as to establish relations between these quantities. A ***thermodynamic system*** is an arbitrary collection of objects, which is described uniquely and completely by state functions (volume V, energy E, number of particles N, electric charge Q, etc.) (Prausnitz et al. 1999).

According to ***quantum physics***, the most complete description that is possible to obtain about a system is a statement of its wave function, *i.e.* the quantity Ψ that appears in Schrodinger's equation. When Ψ is known as a function of the coordinates of the elementary particles, this provides a specification of the ***quantum state*** of the system. For a macroscopic system (including about 10^{24} electrons and nuclei), these quantum states may all be compatible with the same total energy, volume and composition. When measuring a macroscopic property (*e.g.*, pressure, density, etc.), the value obtained results from the chaotic motions and collisions of a large number of molecules; when viewed over a short time scale (*e.g.* 10^{-8} s), this macroscopic property is a fluctuating quantity. In practice, the time required for a macroscopic measurement is usually much longer than 10^{-8} s so that fluctuations are not observed. In other words, the macroscopic properties are time averages over the very large number of possible quantum states that a system may assume, even though each quantum state is compatible with the macroscopically observed values (Prausnitz et al. 1999).

Statistical physics in general, especially statistical mechanics, describes the behaviour of macroscopic systems in terms of microscopic properties, *i.e.* those of particles such as atoms, molecules, ions, etc. The part of statistical mechanics that deals with equilibrium states is called ***statistical thermodynamics***. It applies probability theory, which contains mathematical tools for dealing with large populations, to the study of the thermodynamic behaviour of systems composed of a large number of particles. Statistical mechanics provides a framework for explaining thermodynamics as a result of the classical- and quantum-mechanical descriptions of statistics and mechanics at the microscopic level. The aim of statistical thermodynamics is to calculate the ***time average*** macroscopic properties as a function of molecular properties (Prausnitz et al. 1999).

To calculate time averages over all possible quantum states, some postulates are necessary. The exact formulation of basic postulates uses the definition of an ***ensemble***, which is a large number of imagined systems. In the ensemble, each system has the same macroscopic properties as those chosen to describe the thermodynamic state of some real systems of interest in this study. Although the single systems of ensemble all have the same macroscopic properties, they may have different quantum states (Prausnitz et al. 1999).

A thermodynamic system is generally confined in an enclosure (of volume V) having boundaries (which can be material or immaterial) that separate it from the outer. In this context, the outer can mainly exchange energies (work and heat) and sometimes matter with the macroscopic system (Prausnitz et al. 1999).

If the system cannot exchange either energy or matter, with the outer, it is called an ***isolated system***. In such a system the number of particles N and the total energy E are the conserved variables and the system is called ***microcanonical ensemble***. If for example, the total energy of a real system is E, the volume is V, and the number of molecules is N, then every system in the ensemble also has energy E, volume V and N molecules (Prausnitz et al. 1999).

If the system can exchange energy but not any matter, with the outer, it is called a ***closed system***. In such a system, the total number of particles N is conserved while the total energy E is not conserved. The system is called ***canonical ensemble***. Such an ensemble is characterized by its ***temperature*** T. For instance, if a real system having N molecules in volume V with heat-conducting walls is immersed in a large heat bath, each system of the ensemble containing N molecules in volume V with heat-conducting walls is also immerged in the same large heat bath (Prausnitz et al. 1999).

Finally, if the system can exchange energy and matter with the outer, it is called an ***open system***. When an open system is in equilibrium with its environment, only the averaged values of the total energy E and the total number of particles N are known and have to be taken into account and the system is called ***macro canonical*** or ***grand canonical ensemble***. Such an ensemble is characterized by its temperature T and its ***chemical potential*** μ. Although the chemical potential plays a role analogous to temperature and pressure, understanding the nature of the chemical potential is more difficult. Indeed, if two systems are at different temperatures and are then placed in thermal contact, there will be a net transfer of energy from one system to the other until the temperatures of the two systems become equal. Likewise, if there is a movable wall between two systems at different pressures, then the wall will move so as to change the volume of each system to make the pressures equal. Similarly,

if two systems are initially at different chemical potentials and are then allowed to exchange particles, there will be a net transfer of particles from the system at the higher chemical potential to the one at the lower chemical potential until the chemical potentials become equal (Prausnitz et al. 1999).

Having briefly explained what we mean by ensemble, it is now possible to formulate the first postulate of statistical physics: "The time average of a dynamic property (*i.e.* a property that fluctuates in time like the pressure, the temperature, number of molecules, etc.) of a real system is equal to the ensemble average of that property" (Prausnitz et al. 1999).

To calculate the ensemble average, it is necessary to know the probabilities of the different quantum states of the systems of the ensemble. These probabilities are given by the second postulate of statistical mechanics: "All accessible and distinguishable quantum states of a closed system of fixed energy (microcanonical ensemble) are equally probable" (Prausnitz et al. 1999).

These postulates are expressed by the following equations:

$$X = \sum_i p_i X_i \tag{V-1}$$

In particular,

$$\begin{cases} E = \sum_i p_i E_i \\ N = \sum_i p_i N_i \end{cases} \tag{V-2}$$

and

$$p_1 = p_2 = p_3 = \ldots = p_i = \ldots. \tag{V-3}$$

where X denotes the measured macroscopic dynamic property (E, N, etc.) of the real system and X_i is the value of this property in a system of the ensemble that is in quantum state i.

p_i denotes the probability of quantum state i of the systems of the ensemble, normalized such that $\sum_i p_i = 1$ (the summation runs over all possible quantum states).

To facilitate understanding of the terms "ensemble" and "ensemble average", consider the following example. In a box (which is the real system), there are six spheres of equal size, one is white, two are red, and three are blue. Without looking into the box, let us take out one sphere, note its colour, and put it back again. If one repeats this experience often enough, *e.g.*, 1000 times, the total numbers of white, red and blue spheres taken out are in the ratio 1:2:3, respectively. Now, let us imagine that there are 1000 of these boxes, each containing six

spheres as described above; this set of 1000 boxes is the ensemble. If we take one sphere out of each box and note the colour of the sphere, the ratio of total numbers of white, red and blue spheres is again respectively 1:2:3 (Prausnitz et al. 1999).

This example illustrates that the result of an experiment often repeated with one box is identical to that obtained with only one experiment with an ensemble of many boxes (systems). In other words, it is assumed that "the time average is equal to the ensemble average". This belief is called the ***ergodic hypothesis*** (Prausnitz et al. 1999).

Because a uniform probability distribution reflects the largest randomness, a system with n allowed states (n=3 in the previous example since a given sphere of the system can be white, red or blue, but nothing else) will have the greatest entropy when each state is equally likely. In this situation, the probabilities become

$$p_i = \frac{1}{\Omega} \tag{V-4}$$

where Ω is the total number of microstates (*i.e.* the total number of spheres in the previous example). The entropy is thus:

$$\begin{aligned} S &= -k\sum_{i=1}^{\Omega} p_i \ln\left(p_i\right) = -k\sum_{i=1}^{\Omega} \frac{1}{\Omega}\ln\left(\frac{1}{\Omega}\right) \\ &= -k\Omega\left[\frac{1}{\Omega}\ln\left(\frac{1}{\Omega}\right)\right] = -k\ln\left(\frac{1}{\Omega}\right) = k\ln\left(\Omega\right) \end{aligned} \tag{V-5}$$

k =1.38066 × 10^{23} J/K is the Boltzmann constant, *i.e.* a constant introduced to give to the statistical entropy the same units as the thermodynamic entropy. This relation results in the expression (S = k ln(Ω)) proposed in 1872 by Ludwig Boltzmann (1844-1906) that is a particular case of the general expression given by Gibbs (1910). This expression states that to determine the value of the entropy, it is necessary to know the total number of accessible microstates and the behaviour of the system in the microscopic level. Furthermore, the entropy increases when the number of accessible microstates increases. But when the latter grows, the disorder of the system also increases: the statistical entropy is thus a measure of the disorder of the system (Prausnitz et al. 1999).

The essential problem in statistical thermodynamics is to calculate the distribution of a given amount of energy E over N identical systems. The goal of statistical thermodynamics is to understand and to interpret the measurable macroscopic properties of materials in terms of the properties of their constituent particles and the interactions between them. This is done by

connecting thermodynamic functions to quantum-mechanical equations (Prausnitz et al. 1999).

The total energy E of a thermodynamic system is also called the internal energy, denoted as U. The latter is defined as the energy needed to create the system but excludes the energy to displace the system's surroundings, any energy associated with a move as a whole, or due to external force fields. Internal energy has two major components, kinetic energy and potential energy. The kinetic energy is related to the motion of the system's particles (translations, rotations, vibrations); the potential energy is associated with the static rest mass energy of the constituents of matter, static electric energy of atoms within molecules or crystals, and the static energy of chemical bonds (Prausnitz et al. 1999).

The fundamental equation of thermodynamics for the internal energy U includes terms for various types of work and involves differentials of extensive variables. The fundamental equation for U yields intensive variables as partial derivatives of the internal energy with respect to other extensive properties. In addition to the terms from the combined first and second laws for a system involving pV work, the fundamental equation for the internal energy involve terms for chemical work, gravitational work, work of electric transport, elongation work, surface work, work of electric and magnetic polarization, and other kinds of work.

Thermodynamic properties such as the internal energy U, entropy S, temperature T, pressure p, and volume V behave like mathematical functions, and many relations between thermodynamic properties are obtained by simply using the operations of calculus. When adding other types of work than pV, the number of thermodynamic variables is increased. Thermodynamic potentials are extensive properties that, like the potential energy in mechanics, give information about the most stable state of the system. When terms for non-pV work are introduced, the number of possible thermodynamic potentials also increases. These thermodynamic potentials are obtained by use of Legendre transforms of U and defining intensive properties as the derivative of internal energy over the considered non-pV term. The thermodynamic potential obtained is U minus the conjugate pair of intensive and extensive variables involved in the work term (Alberty 2001,Prausnitz et al. 1999).

The independent variables represented by differentials in a fundamental equation are referred to as natural variables. If a thermodynamic potential is determined as a function of its natural variables, all of the thermodynamic properties of the system are obtained by taking partial derivatives of the thermodynamic potential with respect to the natural variables. The natural variables are also held constant in the criterion for spontaneous change and equilibrium based on that thermodynamic potential. By use of Legendre transforms any

desired set of natural variables are obtained. The most classical example concerns the definition of enthalpy H, Helmholtz energy A, and Gibbs energy G that are defined by Legendre transforms that introduce p, V, and S and T together or separately as natural variables (Alberty 2001, Prausnitz et al. 1999).

$$H = U - \left(-pV\right) = U + pV \tag{V-6}$$

$$A = U - TS \tag{V-7}$$

$$G = U - TS - \left(-pV\right) = H - TS \tag{V-8}$$

Legendre transforms are used to introduce the chemical potential of any species, the gravitational potential, electric potentials of phases, surface tension, the force of elongation, electric field strength, magnetic field strength, and other intensive variables as natural variables. The large number of transformed thermodynamic potentials that can be defined raises serious nomenclature problems (Alberty 2001). In the present study, the definition of thermodynamic potentials will include the electric potential in the variables in order to account for a consistent derivation of COSMO thermodynamic functions.

When there are several distinguishable species or particles in a system, the chemical potential μ_i of species i is especially important. This intensive property is built when building the Legendre transform of the number of moles of species i. This intensive property uniform throughout a multiphase system at equilibrium even though the phases may be different states of matter or be at different pressures, gravitational potentials, or electric potentials. When the chemical potential of a species is held constant, a Legendre transform can be used to define a transformed Gibbs energy, which is minimized at equilibrium at a specified chemical potential of that species (Alberty 2001). Importantly in statistical physics, the concept of chemical potential is only introduced as a descriptor in the grand canonical ensemble (*i.e.* not in the canonical, nor in the microcanonical ensembles) for the reason that the number of moles must not be fixed for defining the chemical potential.

In the following discussion, the goal is to determine at each step of the working scheme (Figure V-1) the most relevant information that enables the computation of thermodynamic data, or data needed for this calculation. The ultimate goal of all this study is to provide thermodynamic data H, S and G (or μ_i) in the gas phase and in liquid state with a common reference state in any solvent (so at infinite dilution in water when the solvent is water).

V.3. Individual molecule in the hypothetical ideal gas phase (step 1)

Since the descriptions given by Irikura et al. (1998) well represent the needs of the following discussion, we have decided to use the same notations convention in the present work and almost all of this discussion is taken from their reference work.

When an individual molecule is in the hypothetical ideal gas phase, it is free to translate, rotate and vibrate. All these molecular motions are described by a combination of quantum calculations and statistical physics tools (Irikura et al. 1998).

Indeed, statistical thermodynamics is meant to include the methods used to convert molecular energy levels into macroscopic properties, especially enthalpies, entropies, and heat capacities. Molecular energy levels arise from molecular translation (*i.e.*, motion through space), rotation, vibration, and electronic excitation. This information constitutes the spectroscopy of the molecule of interest and can be obtained experimentally or from calculations (Irikura et al. 1998).

The molecular energy levels ε_i are used to compute the molecular partition function, usually denoted by the symbol Z, as shown in equation below:

$$Z(T) = \sum_i \exp\left(\frac{-\varepsilon_i}{kT}\right) \tag{V-9}$$

The sum extends over all energy levels, sometimes this sum is written only over all unique energy levels, in which case a level degeneracy g_i must be included in the sum. However, for very high temperatures at which the molecule becomes unstable, the extent of the sum may be ambiguous. Tabulated thermo chemical data must be used with caution under such conditions; the values (1) may depend strongly upon the high-energy cut-off procedure adopted and (2) may deviate implicitly from the ideal-gas model (Irikura et al. 1998).

One typically chooses the lowest energy level to be the zero of energy, so that no levels lie at negative energies. From equation (V-9) it follows that the largest contributions to Z are from the lowest energy levels. Conversely, levels that lie far above kT have only a minor effect on Z and its derivative thermodynamic quantities (Irikura et al. 1998).

Given the partition function, the usual molar thermodynamic functions are calculated based upon the following general equations (Irikura et al. 1998).

$$S = Nk\left[\frac{\partial}{\partial T}\left(T \ln Z\right) - \ln N + 1\right] \tag{V-10}$$

$$C_v = NkT\frac{\partial^2}{\partial T^2}\left(T\ln Z\right) \tag{V-11}$$

$$C_p = C_v + R \tag{V-12}$$

$$H\left(T\right) - H\left(0\right) = \int_0^T C_p dT = \frac{RT^2}{Q}\frac{\partial Z}{\partial T} + RT \tag{V-13}$$

$$\frac{\partial}{\partial T}\left(T\ln Z\right) = \ln Z + \frac{T}{Z}\frac{\partial Z}{\partial T} \tag{V-14}$$

$$\frac{\partial^2}{\partial T^2}\left(T\ln Z\right) = \frac{2}{Z}\frac{\partial Z}{\partial T} + \frac{T}{Z}\frac{\partial^2 Z}{\partial T^2} - \frac{T}{Z^2}\left(\frac{\partial Z}{\partial T}\right)^2 \tag{V-15}$$

$$\frac{\partial Z}{\partial T} = \frac{1}{kT^2}\sum_i \varepsilon_i \exp\left(\frac{-\varepsilon_i}{kT}\right) \tag{V-16}$$

$$\frac{\partial^2 Z}{\partial T^2} = \frac{-2}{T}\frac{\partial Z}{\partial T} + \frac{1}{k^2T^4}\sum_i \varepsilon_i^2 \exp\left(\frac{-\varepsilon_i}{kT}\right) \tag{V-17}$$

Equation (V-10) is for the entropy, equation (V-11) for the heat capacity at constant volume, equation (V-12) for the heat capacity at constant pressure, and equation (V-13) for the enthalpy difference relative to absolute zero temperature. N is Avogadro's number (6.022137 $\times$ 10^{23}), k is Boltzmann's constant (1.38066 $\times$ 10^{23} J/K), and the ideal-gas constant is $R \equiv Nk$ (Cohen and Taylor 1988). The last two terms inside the brackets in equation (V-10) arise from the indistinguishability of identical molecules, which requires a factor of (1/N!) in the partition function for the ensemble. Expressions V-10 to V-13 may more easily be evaluated using equations V-14 to V-17 for the various derivatives (Irikura et al. 1998).

In practice, a complete set of molecular energy levels is almost never available. To simplify the problem, one usually adopts a model in which translation, rotation, vibration, and electronic excitation are uncoupled. In other words, one makes the approximation that the different types of motion are unaffected by each other and do not mix together. This leads to a separability of Z into four factors that correspond to separate partition functions for translation, rotation, vibration, and electronic excitation. This is shown in equation (V-18), where the explicit dependence upon temperature has been dropped for simplicity (Irikura et al. 1998).

$$Z = Z_{trans}Z_{rot}Z_{vib}Z_{elec} \tag{V-18}$$

When electronically excited states are considered, one often assumes that the translational, rotational, and vibrationnal spectra of the excited state are the same as those of the ground electronic state. This is crude, but is convenient when no other information is available. Moreover, if the excited state lies far above *kT*, the final results will not be sensitive to such details (Irikura et al. 1998).

V.3.1. Translational Partition Function

Rigorously, Z_{trans} must be calculated from a sum over all the translational energy levels that are available to a molecule confined to a cubic box of volume $V = RT/p$ (molar volume of an ideal gas at temperature T and pressure p). This is seldom done. Instead, the sum is approximated as an integral to obtain equations V-19 to V-22 (Irikura et al. 1998).

$$Z_{trans} = \left(2\pi mkT\right)^{3/2} h^{-3} V \tag{V-19}$$

$$S^{\circ}_{trans} = R\left[\frac{3}{2}\ln\left(\frac{2\pi m}{h^2}\right) + \frac{5}{2}\ln\left(kT\right) - \ln\left(p\right) + \frac{5}{2}\right] \tag{V-20}$$

$$C^{\circ}_{p,trans} = \frac{5}{2}R \tag{V-21}$$

$$\left[H^{\circ}(T) - H^{\circ}(0)\right]_{trans} = \frac{5}{2}RT \tag{V-22}$$

This approximation is good as long as $m^{3/2}T^{5/2}p^{-1} >> h^3\,(2\pi)^{-3/2}k^{-5/2}$. At the standard pressure p = 1 bar = 10^5 Pa, this condition is met for sufficiently heavy molecules, m (in amu) >> 11.4 $T^{5/3}$, and for sufficiently high temperatures, $T >> 4.31\ m^{-3/5}$ (m expressed in amu). Fortunately, this covers the conditions of common chemical interest. For an atomic ideal gas, there is no vibrationnal or rotational motion (Irikura et al. 1998).

V.3.2. Rotational Partition Function

The free rotation of a rigid molecule is also quantized (the angular momentum and its projection are integer multiples of $h/2\pi$), so that the rotational energy is restricted to certain discrete levels. Rotational spectra are characterized by the constants A, B, and C, where $A \equiv h/(8\pi^2 I_A)$ and likewise for B and C. The quantities $I_{A,B,C}$ are the principal moments of inertia of the molecule, with the convention $I_A \le I_B \le I_C$ (or $A \ge B \ge C$). Many computer programs, including *ab initio* packages, report the rotational constants when provided with a molecular geometry. The moments can also be calculated manually as the eigenvalues of the inertial tensor, which has elements like $I_{xy} = -\sum m_i x_i y_i$ and $I_{xx} = \sum m_i(y_i^2 + z_i^2)$, where the index i runs over

all atoms in the molecule and the coordinate origin is at the centre of mass. The product of moments of inertia is all that is needed for the partition function, not the individual eigenvalues of the inertial tensor, and the product can be obtained directly from the elements:

$$I_A I_B I_C = I_{xx} I_{yy} I_{zz} + 2 I_{xy} I_{xz} I_{yz} - I_{xx} {I_{yz}}^2 - I_{yy} {I_{xz}}^2 - I_{zz} {I_{xy}}^2 \quad \text{(V-23)}$$

Linear molecules (I_A = 0) are described by a single rotational constant, B, and a single moment of inertia, I. Details may be found in textbooks of molecular spectroscopy.

Fortunately, at high enough temperatures ($kT >> hA$), the sum can be replaced by an integral as it is for translation. In the general case, the rotational partition function is given by:

$$Z_{rot} = \frac{8\pi^2}{\sigma h^3} \left(2\pi m k T\right)^{3/2} \left(I_A I_B I_C\right)^{3/2} = \left(\frac{kT}{h}\right)^{3/2} \left(ABC\right)^{-1/2} \pi^{1/2} \sigma^{-1} \quad \text{(V-24)}$$

For linear molecules, equation V-25 should be used instead of equation V-24:

$$Z_{rot}^{linear} = \frac{8\pi^2 I k T}{\sigma h^2} = \frac{kT}{\sigma h B} \quad \text{(V-25)}$$

In these equations, the symbol σ denotes the "rotational symmetry number" or "external symmetry number" for the molecule. This is the number of unique orientations of the rigid molecule that only interchange identical atoms. It preserves parity restrictions on the interchange of identical nuclei when summation is replaced by integration. Identifying the correct symmetry number is a common point of difficulty; it is discussed further below.

For the typical case (equation (V-24)), the thermodynamic functions are given by equations (V-26 to V-28) (Irikura et al. 1998).

$$\begin{aligned} S_{rot}^{\circ} &= R\left[\ln\left(\frac{8\pi^2}{\sigma}\right) + \frac{3}{2}\ln\left(\frac{2\pi k T}{h^2}\right) + \frac{1}{2}\ln\left(I_A I_B I_C\right) + \frac{3}{2}\right] \\ &= R\left[\frac{3}{2}\ln\left(\frac{kT}{h}\right) - \frac{1}{2}\ln\left(\frac{ABC}{\pi}\right) - \ln\left(\sigma\right) + \frac{3}{2}\right] \end{aligned} \quad \text{(V-26)}$$

$$C_{p,rot}^{\circ} = \frac{3}{2} R \quad \text{(V-27)}$$

$$\left[H^{\circ}(T) - H^{\circ}(0)\right]_{rot} = \frac{3}{2} RT \quad \text{(V-28)}$$

For linear molecules (equation (V-25)), equations (V-29 to V-31) are used instead.

$$S_{rot}^{\circ,linear} = R\left[\ln\left(\frac{8\pi^2 I k T}{\sigma h^2}\right) + 1\right] = R\left[\ln\left(\frac{kT}{\sigma h B}\right) + 1\right] \quad \text{(V-29)}$$

$$C_{p,rot}^{\circ,linear} = R \tag{V-30}$$

$$\left[H^{\circ}(T) - H^{\circ}(0)\right]_{rot}^{linear} = RT \tag{V-31}$$

V.3.3. External symmetry number

Some computer programs, such as many *ab initio* packages, determine the molecular symmetry and external symmetry number (σ) automatically. If such a program is unavailable, σ may be determined by hand. With practice, this becomes very fast. If one is familiar enough with group theory to identify the molecule's point group (Cotton 1971), σ can be determined from Table V-1 (Herzberg 1945). Without identifying the point group, one can count manually the number of orientations of the rigid molecule that interchange only identical atoms (Irikura et al. 1998).

Table V-1: Symmetry numbers corresponding to symmetry point groups.

These data are taken from Herzberg 1945.

Group	Symmetry number (σ)	Group	Symmetry number (σ)	Group	Symmetry number (σ)	Group	Symmetry number (σ)
C_1, C_i, C_s, $C_{\infty v}$	1	$D_{\infty h}$	2	T, T_d	12	O_h	24
C_n, C_{nv}, C_{nh}	n	D_n, D_{nh}, D_{nd}	2n	S_n	n/2	I_h	60

For example, the benzene molecule (C_6H_6) belongs to the D_{6h} point group. From Table V-1, σ=12. Alternatively, one can draw the molecule as a hexagon with numbered vertices. Rotating the drawing by $n \times 60°$, where n runs from 0 to 5, generates six different orientations that are distinguished only by the artificial numbering of the vertices. Each of these six orientations can be flipped over to generate another orientation, for a total of 12 unique orientations, σ = 12. Another example is methyl chloride, CH_3Cl. This belongs to the C_{3v} point group, so σ = 3. Alternatively, one can artificially number the hydrogen atoms and see that there are three unique orientations, related by rotations of n × 120° (n = 0-2) around the C-Cl bond axis. Chlorobenzene, C_6H_5Cl, belongs to the C_{2v} point group, so σ = 2. Alternatively, one can again number the hydrogen atoms and see that there are two unique orientations, related by rotations of n × 180° (n = 0-1) around the C-Cl bond axis. In contrast, toluene ($C_6H_5CH_3$) belongs to the C_s point group, so σ = 1. There are no ways to rotate or flip the molecule rigidly that will leave it unchanged (Irikura et al. 1998).

V.3.4. Vibrationnal Partition Function

To complete the simple rigid-rotator/harmonic oscillator (RRHO) model, one must consider the molecular vibrations. A molecule that contains N atoms has $3N$-6 vibrational frequencies ($3N$-5 for linear molecules). The partition function is given in equation (V-32), where the product runs over all vibrational frequencies ν_i. The corresponding thermodynamic functions are given by equations (V-33 to V-35) (Irikura et al. 1998).

$$Z_{vib} = \prod_i \left(1 - e^{-h\nu_i/kT}\right)^{-1} \tag{V-32}$$

$$S^\circ_{vib} = -R\sum_i \ln\left(1 - e^{-h\nu_i/kT}\right) + R\sum_i \frac{h\nu_i}{kT}\left(\frac{e^{-h\nu_i/kT}}{1 - e^{-h\nu_i/kT}}\right) \tag{V-33}$$

$$C^\circ_{p,vib} = R\sum_i \left(\frac{h\nu_i}{kT}\right)^2 \frac{e^{-h\nu_i/kT}}{\left(1 - e^{-h\nu_i/kT}\right)^2} \tag{V-34}$$

$$\left[H^\circ(T) - H^\circ(0)\right]_{vib} = RT\sum_i \left(\frac{h\nu_i}{kT}\right)\frac{e^{-h\nu_i/kT}}{\left(1 - e^{-h\nu_i/kT}\right)} \tag{V-35}$$

V.3.5. Electronic Partition Function

Although they may not have low-lying electronic excited states, some molecules have degenerate electronic ground states. Free radicals are a common example. They may have unpaired electrons in their electronic ground states and a net electron spin of $S = n_{unpaired}/2$, where $n_{unpaired}$ is the number of unpaired electrons (beware not to confuse the spin quantum number S with the entropy). The multiplicity, or degeneracy g, of such a state is $g = (2S+1)$. Using degeneracy numbers is equivalent to an explicit count of all states, including degenerate ones. Thus, $Q_{elec} = g$ is a constant and only affects the entropy: $S_{elec} = R \ln(g)$ and $C_{p,\,elec} = [H(T)-H(0)]_{elec} = 0$. Since most free radicals have only a single unpaired electron, the usual effect is to increase the entropy by $R \ln(2)$. In addition to spin degeneracies, some states have spatial degeneracies. This situation is most common for diatomic molecules. Linear molecules with a spatial symmetry other than Σ (*e.g.*, Π or Δ) have a spatial degeneracy of 2. For example, the OH radical has a $^2\Pi$ ground state, so its degeneracy is g = 2 (spin) × 2 (spatial) = 4. If there are both spin and spatial degeneracies, spin-orbit coupling lifts the

degeneracy, often significantly. In the example of OH, the 4-fold degenerate ground state is split into two doubly-degenerate levels separated by 139.2 cm^{-1} (Huber and Herzberg 1979). In such a case the low-lying excited states should be included in the calculation of thermodynamic quantities. The partition function is given by equation (V-36), where ε_i and g_i are the excitation energies (spectroscopic T_0) and degeneracies of the excited states, g_0 and $\varepsilon_0 \equiv 0$ are for the ground state, and the sum runs over all the electronic states being considered, including the ground state. The contributions to the thermal functions are given by equations (V-37 and V-39). This treatment assumes, rather crudely, that the rotations and vibrations are unaffected by electronic excitation (Irikura et al. 1998).

$$Z_{elec} = \sum_i g_i \exp\left(-\varepsilon_i / kT\right) \tag{V-36}$$

$$S^\circ_{elec} = R \ln\left(\sum g_i \exp\left(-\varepsilon_i / kT\right)\right) + R \frac{\sum g_i \left(\varepsilon_i / kT\right) \exp\left(-\varepsilon_i / kT\right)}{\sum g_i \exp\left(-\varepsilon_i / kT\right)} \tag{V-37}$$

$$C^\circ_{p,elec} = R\left(\frac{\sum g_i \left(\varepsilon_i / kT\right)^2 \exp\left(-\varepsilon_i / kT\right)}{\sum g_i \exp\left(-\varepsilon_i / kT\right)}\right) - R\left(\frac{\sum g_i \left(\varepsilon_i / kT\right) \exp\left(-\varepsilon_i / kT\right)}{\sum g_i \exp\left(-\varepsilon_i / kT\right)}\right)^2 \tag{V-38}$$

$$\left[H^\circ(T) - H^\circ(0)\right]_{elec} = RT \frac{\sum g_i \left(\varepsilon_i / kT\right) \exp\left(-\varepsilon_i / kT\right)}{\sum g_i \exp\left(-\varepsilon_i / kT\right)} \tag{V-39}$$

V.3.6. Concluding remarks about step 1

At this stage, it becomes possible to compute, at a given temperature T, ideal gas thermodynamic quantities such as the standard molar entropy $S°(T)$, heat capacity $C_p°(T)$, and enthalpy change $[H°(T)-H°(0)]$ which are the sum of their corresponding translational, rotational, vibrationnal and electronic components (Irikura et al. 1998).

$$S^\circ(T) = S^\circ_{trans}(T) + S^\circ_{rot}(T) + S^\circ_{vib}(T) + S^\circ_{elec}(T) \tag{V-40}$$

$$C^\circ_p(T) = C^\circ_{p,trans}(T) + C^\circ_{p,rot}(T) + C^\circ_{p,vib}(T) + C^\circ_{p,elec}(T) \tag{V-41}$$

$$\left[H^\circ(T) - H^\circ(0)\right] = \left[H^\circ(T) - H^\circ(0)\right]_{trans+rot+vib+elec} \tag{V-42}$$

To generate the enthalpy of formation at a given temperature, especially at room temperature, it remains to predict $\Delta H_f°(0)$ for the reference elements as well (using the Hess law).

$$\Delta H_f^{0,gas}(T) = \Delta H_f^{\circ,gas}(0) + \Delta\left\{\left[H°(T) - H°(0)\right]\right\}_f^{gas} \qquad \text{(V-43)}$$

However, as shortly discussed in Chapter IV, in almost all of quantum computational methods which are *ab initio* electronic calculations of molecular energy, the raw results do not even correspond to properties at absolute zero temperature and must always be corrected. All the corrections are based upon molecular spectroscopy, with temperature-dependence implicit in the molecular partition function. By convention, energies from *ab initio* calculations are reported in hartrees, the atomic unit of energy (1 hartree = 2625.5 kJ/mol = 627.51 kcal/mol). These energies are negative, with the defined zero of energy being the fully-dissociated limit (free electrons and bare nuclei). *Ab initio* models also invoke the approximation that the atomic nuclei are stationary, with the electrons swarming about them. This is a good approximation because nuclei are much heavier than electrons. Consequently, the resulting energies are for a hypothetical, non-vibrating molecule. Although oscillators may be at rest in classical mechanics, real (quantum-mechanical) oscillators are always in motion (Irikura et al. 1998).

The small residual motion at absolute zero temperature is the zero-point vibrationnal energy, abbreviated ZPVE or ZPE. For a simple harmonic oscillator, the ZPE equals one-half the vibrationnal frequency. Although all real molecular vibrations are at least slightly anharmonic, they are usually approximated as harmonic. Thus, the molecule's ZPE may be taken as one-half the sum of the vibrationnal frequencies (Irikura et al. 1998).

$$ZPE = \frac{1}{2}\sum_{i=1}^{3N-6} \nu_i \qquad \text{(V-44)}$$

In equation (V-44), N is the number of atoms in the molecule and the ν_i are the fundamental vibrational frequencies. There are $3N$-6 vibrations in a non-linear molecule and $3N$-5 in a linear molecule; the above equation is for the more common non-linear case. The ZPE must be added to the raw *ab initio* energy ($E^{ab\ initio}$) to obtain an energy corresponding to absolute zero temperature, T = 0 K, thus the corresponding enthalpy (Irikura et al. 1998).

$$H°(0) = E°(0) = E^{ab\,initio} + ZPE \qquad \text{(V-45)}$$

In practice, the ZPE correction is slightly complicated by the observation that *ab initio* vibrational frequencies are often in error by +5% to +10%. To compensate for this error, the

computed frequencies are usually multiplied by empirical scaling factors. The most recent recommendations are those of Scott and Radom (1996) (Irikura et al. 1998).

When the enthalpy of formation is known, the absolute entropy $S°$ can be used to calculate the entropy of formation $\Delta S_f^{0,gas}$ that is summed with $\Delta H_f^{0,gas}$ to determine the gas phase chemical potential of formation.

$$\begin{cases} \mu_i^{0,gas} = \Delta G_f^{0,gas} = \Delta H_f^{0,gas} - T\Delta S_f^{\circ,gas} \\ \text{and } \Delta S_f^{\circ,gas} = S^{\circ,gas} + \sum_{\text{ref.elements}} \lambda_i S_i^{\circ,gas} \end{cases} \quad \text{(V-46)}$$

where λ_i denotes the elementary composition of the pure component and $S_i^{\circ,gas}$ the standard entropies (in the gas phase) of the reference elements.

The above equations demonstrate that by combining a high level quantum calculation tool with a statistical physics treatment of molecular motions, it is possible to determine all the thermodynamic quantities required for thermochemistry calculations in the hypothetical ideal gas state.

It is also possible to use this kind of method to approach a realistic description of the liquid state where intermolecular interactions must be taken into account efficiently. This is done in molecular dynamics or Monte Carlo simulations. However, such a simulation is computer time consuming and hardly doable in complex systems like those encountered in biological and food systems. Thus, it is necessary to look for another way to compute relevant thermochemistry data for such compounds. To do so, it was decided within this work to develop a pathway that is similar to those used in classical thermodynamics to go from gas phase to the thermodynamic properties of liquids. In this context, Ould-Moulaye (1997) has proven that the classical relations allowing the transition (in Figure V-1) from Step 1 (gas phase reference state) to Step 5 (infinite dilution in water reference state) provide reasonable accuracy for all relevant properties in biological media, even if it is necessary to provide activity coefficients data, pure component molar volume and vapour pressures.

As discussed in chapter IV, regarding the changes of chemical potentials between the gas phase and the infinite dilution, it is also possible to use the solvation process described by Ben Naim (1987 and 2001) that consists to move the molecule from a fixed position in the gas phase to the same position at infinite dilution in water (in this case, in the COSMO-RS algorithm). COSMO-RS combines concepts of electrostatics and quantum physics (Steps 2 to 4) as well as statistical physics to provide activity coefficients of liquids systems (Step 5 in Figure V-1). However, instead of a fully statistical thermodynamic description of the 3D

molecule (with all their complicated geometry constraints), the COSMO-RS model treats only local pairwise interactions of charged surfaces. Because this method was described in details in chapter II, we will only give further details about the statistical treatment leading to the expression of the chemical potential of a given compound in the following paragraph. Since the concept of chemical potential has a real meaning only in the grand canonical ensemble, to check the thermodynamic consistency of the COSMO-RS model regarding the pathway proposed in this work, it is necessary to establish the expression of the chemical potential of a given compound in this ensemble of pairwise interacting charged surfaces.

V.4. Thermodynamic properties of interacting surfaces in the grand canonical ensemble

V.4.1. For a mono-constituent system

Let us consider an ensemble of N_t particles in pairwise interactions. The aim is here to derive the partition function in the grand canonical ensemble, *i.e.* in an ensemble where V is fixed.

The partition function of this ensemble (also called grand partition function) is given by:

$$\xi = \sum_{E,N} \Omega_{EN} e^{\beta\mu N} e^{-\beta E} \tag{V-47}$$

E denotes the total energy of N particles. Ω_{EN} denotes for the number of microstates such as the total energy is E for a number N of particles.

The variables E, N, and Ω are functions of total volume V. This has not been reported in the previous equation for simplifying the notations. The summation extends on all possible values of E and N.

The double summation in (V-47) is expressed as:

$$\xi = \sum_{E,N} \Omega_{EN} e^{\beta\mu N} e^{-\beta E} = \sum_{N} \sum_{E(N)} \Omega_{E(N)} e^{\beta\mu N} e^{-\beta E(N)} \tag{V-48}$$

Let ε denotes the energy of a single interaction: for n identical interactions (interaction of surfaces in our case), the total energy is calculated as:

$$E = n\varepsilon \tag{V-49}$$

In the same way, for n_j interactions of energy ε_j the total energy is given by:

$$E = \sum_j n_j \varepsilon_j \quad \text{and} \quad n = \sum_j n_j \tag{V-50}$$

The average energy per interaction here becomes: $\varepsilon = \frac{E}{n}$.

For N particles in pairwise interactions, there are $n = \frac{N}{2}$ interactions.

Now, let us consider that the n interactions have ω available microstates. For the total set of N_t particles, the total number of available microstates ω is:

$$\omega = \frac{N_t}{2} \tag{V-51}$$

In order to obtain the Ω parameter, we must now calculate the number of possible states for the interactions of N particles. Following a classical demonstration in statistical physics, three cases must be analysed:

- Distinguishable interactions: $\Omega_D = \omega^n$
- Non-distinguishable interactions; only one by microstate: $\Omega_F = \frac{\omega!}{n!(\omega-1)!} \approx \frac{\omega^n}{n!}$
- Non-distinguishable interactions; variable number by microstate: $\Omega_B = \frac{(n+\omega-1)!}{n!(\omega-1)!} \approx \frac{\omega^n}{n!}$

The approximations in previous equation stand for $\omega >> n$ that is satisfied.

In quantum physics, interactions (and particles) are considered as non-distinguishable, so that we obtain:

$$\Omega = \frac{\omega^n}{n!} = \frac{\omega^{N/2}}{\left(N/2\right)!} \tag{V-52}$$

The sum in (V-48) is therefore given by:

$$\sum_N \sum_{E(N)} \Omega_{E(N)} e^{\beta\mu N} e^{-\beta E} = \sum_N \sum_{E(N)} \Omega_{E(N)} e^{\beta\mu N} e^{-\beta\frac{N}{2}\varepsilon} = \sum_N \frac{\omega^{N/2}}{\left(N/2\right)!} e^{\beta\mu N} e^{-\beta\frac{N}{2}\varepsilon} = \sum_N \frac{\left(\omega e^{\beta(2\mu-\varepsilon)}\right)^{N/2}}{\left(N/2\right)!} \tag{V-53}$$

where $E(N) = n\varepsilon = \frac{N}{2}\varepsilon$ and the summation over all even values of N (we are dealing with pairwise interactions). Thus the grand partition function ξ becomes:

$$\xi = \sum_N \frac{\left(\omega e^{\beta(2\mu-\varepsilon)}\right)^{N/2}}{\left(N/2\right)!} \tag{V-54}$$

Considering that $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$, the partition function becomes:

$$\xi \approx e^{\left(\omega e^{\beta(2\mu-\varepsilon)}\right)} \tag{V-55}$$

Thus,

$$\ln \xi = \ln\left(\omega e^{\beta(2\mu-\varepsilon)}\right) \tag{V-56}$$

So that:

$$\ln \xi = \omega e^{\beta(2\mu-\varepsilon)} \tag{V-57}$$

ω is an extensive constant that is proportional to the total number of particles by definition as given by Eq. (V-51). Now it is necessary to relate the grand partition function to the thermodynamic variables.

The fundamental relations to determine thermodynamic variables (especially the Gibbs energy) in the grand canonical ensemble are (Prausnitz et al. 1999):

$$G = \frac{1}{\beta}\left(\frac{\partial \ln \xi}{\partial \ln \mu}\right)_{T,V} \tag{V-58}$$

$$pV = kT \ln \xi \left(\text{if } p \text{ and } V \text{ are the variables of the system}\right) \tag{V-59}$$

So that:

$$G = 2\mu\omega e^{\beta(2\mu-\varepsilon)} \tag{V-60}$$

By definition, the total Gibbs energy of a given compound is also given by:

$$G = N_t \mu \tag{V-61}$$

where μ denotes the chemical potential.

Using Eq. V-51, it follows that:

$$e^{\beta(2\mu-\varepsilon)} = 1 \tag{V-62}$$

So that:

$$\mu = \frac{\varepsilon}{2} \tag{V-63}$$

It appears then that this demonstration defines the origin of the chemical potentials values (thus of the entropies).

V.4.2. For a multi-constituent system

Now, let us consider a mixture containing respectively N_{1t} and N_{2t} particles in pairwise interactions in the grand canonical ensemble.

Before going further, let us establish the balance equations for describing the contacts. If N_{11}, N_{22} and N_{12} denote respectively the number of contacts of type 1-1 (energy ε_{11}), 2-2 (energy ε_{22}) and 1-2 (energy ε_{12}); the balance of contacts is written as:

$$\begin{cases} N_{1t} = 2N_{11t} + N_{12t} \\ N_{2t} = 2N_{22t} + N_{21t} \end{cases} \quad \text{(V-64)}$$

If we assume that the system is a completely random mixture, using Guggenheim's hypothesis (Prausnitz et al. 1999), it can be written that:

$$N_{12t} = \frac{N_{1t} N_{2t}}{N_{1t} + N_{2t}} \quad \text{(V-65)}$$

By coupling these sets of equations, it appears that:

$$\begin{cases} \omega_{11} = N_{11t} = \frac{1}{2} x_1 N_{1t} \\ \omega_{22} = N_{22t} = \frac{1}{2} x_2 N_{2t} \\ \omega_{12} = N_{12t} = x_2 N_{1t} = x_1 N_{2t} = \frac{1}{2}\left(x_2 N_{1t} + x_1 N_{2t} \right) \end{cases} \quad \text{(V-66)}$$

where:

$x_1 = \frac{N_1}{N_1 + N_2}$ and $x_2 = \frac{N_2}{N_1 + N_2}$ denote the molar fractions.

In order to fit with the notations used in previous paragraph, we have also defined ω_{ij} that are the numbers of microstates available for interaction of type *i-j*

.

The grand partition function of this ensemble is given by:

$$\xi = \sum_{E, N_1, N_2} \Omega_{E N_1 N_2} e^{\beta \mu_1 N_1} e^{\beta \mu_2 N_2} e^{-\beta E(N_1 N_2)} \quad \text{(V-67)}$$

As previously the grand partition function is also given by:

$$\xi = \sum_{N_1} \sum_{N_2(N_1)} \sum_{E(N_1, N_2)} \Omega_{E(N_1, N_2)} e^{\beta \mu_1 N_1} e^{\beta \mu_2 N_2} e^{-\beta E(N_1, N_2)} \quad \text{(V-68)}$$

The total interaction energy *E* for N_1 and N_2 pairwise interacting particles is given by:

$$E = N_{11}\varepsilon_{11} + N_{12}\varepsilon_{12} + N_{22}\varepsilon_{22} \quad \text{(V-69)}$$

We now use the same balance equations as Eqs. (V-64) and (V-65) keeping in mind that here the hypothesis of random mixing (Guggenheim hypothesis) is applied to fluctuating variables N_1 and N_2:

$$E = \frac{1}{2}\left(x_1\varepsilon_{11} + x_2\varepsilon_{12} \right) N_1 + \frac{1}{2}\left(x_1\varepsilon_{12} + x_2\varepsilon_{22} \right) N_2 \quad \text{(V-70)}$$

From the above, the grand partition function is rewritten as:

$$\xi = \sum_{N_1} \sum_{N_2(N_1)} \sum_{E(N_1, N_2)} \Omega_{E(N_1, N_2)} e^{\beta(2\mu_1 - x_1\varepsilon_{11} - x_2\varepsilon_{12})\frac{N_1}{2}} e^{\beta(2\mu_2 - x_1\varepsilon_{12} - x_2\varepsilon_{22})\frac{N_2}{2}} \quad \text{(V-71)}$$

In the form of Eq. (V-71), the sequence of these 3 sums is impossible to express directly. The combined probabilities law enables to express directly $\ln \xi$ as

$$\ln \xi = \ln \xi_{11} + \ln \xi_{12} + \ln \xi_{22} \tag{V-72}$$

$\ln\xi_{11}$ and $\ln\xi_{22}$ are expressed using Eq. (V-57) established earlier for mono-constituent system:

$$\ln \xi_{11} = N_{11t} e^{\beta(2\mu_1 - \varepsilon_{11})} \tag{V-73}$$

$$\ln \xi_{22} = N_{22t} e^{\beta(2\mu_2 - \varepsilon_{22})} \tag{V-74}$$

$\ln \xi_{12}$ is determined in a similar way. Considering that there are N_{12} particles of type N_1 that interact with N_{12} particles of type N_2, ξ_{12} is given similarly to Eq. (V-53):

$$\sum_{E(N_{12})} \Omega_{E(N_{12})} e^{\beta(\mu_1+\mu_2)N_{12}} e^{-\beta\varepsilon_{12}N_{12}} = \frac{\omega_{12}^{N_{12}}}{N_{12}!} e^{\beta(\mu_1+\mu_2)N_{12}} e^{-\beta\varepsilon_{12}N_{12}} = \frac{\left(\omega_{12} e^{\beta(\mu_1+\mu_2-\varepsilon_{12})}\right)^{N_{12}}}{N_{12}!} \tag{V-75}$$

As previously ξ_{12} is calculated after the sum over all N_{12} interactions:

$$\xi_{12} = \sum_{N_{12}} \frac{\left(\omega_{12} e^{\beta(\mu_1+\mu_2-\varepsilon_{12})}\right)^{N_{12}}}{N_{12}!} \tag{V-76}$$

So that:

$$\ln \xi_{12} = N_{12t} e^{\beta(\mu_1+\mu_2-\varepsilon_{12})} \tag{V-77}$$

Finally, the grand partition function $\xi = (\xi_{11})(\xi_{12})(\xi_{22})$ is calculated:

$$\ln \xi = N_{11t} e^{\beta(2\mu_1-\varepsilon_{11})} + N_{12t} e^{\beta(\mu_1+\mu_2-\varepsilon_{12})} + N_{22t} e^{\beta(2\mu_2-\varepsilon_{22})} \tag{V-78}$$

Using the definition of the total partial Gibbs energy of component 1, we obtain:

$$G_1 = N_{1t}\mu_1 = \frac{1}{\beta}\left(\frac{\partial \ln \xi}{\partial \ln \mu_1}\right)_{T,V,\mu_2} = \mu_1\left[2N_{11t} e^{\beta(2\mu_1-\varepsilon_{11})} + 2N_{12t} e^{\beta(\mu_1+\mu_2-\varepsilon_{12})}\right] \tag{V-79}$$

With the relations of definition of N_{11t} and N_{12t} given in Eqs (V-66), we obtain:

$$x_1 e^{\beta(2\mu_1-\varepsilon_{11})} + x_2 e^{\beta(\mu_1+\mu_2-\varepsilon_{12})} = 1 \tag{V-80}$$

This is the same as:

$$e^{-\beta\mu_1} = x_1 e^{\beta(\mu_1-\varepsilon_{11})} + x_2 e^{\beta(\mu_2-\varepsilon_{12})} \tag{V-81}$$

Likewise,

$$e^{-\beta\mu_2} = x_1 e^{\beta(\mu_1-\varepsilon_{12})} + x_2 e^{\beta(\mu_2-\varepsilon_{22})} \tag{V-82}$$

These relations can be extended easily to an ensemble containing p systems:

$$-\beta\mu_i = \ln\left(\sum_j x_j e^{\beta(\mu_j - \varepsilon_{ij})}\right) \quad \text{(V-83)}$$

It should be noticed that equation (V-83) is a more general form of equation (II-44) that results from the demonstration given in details by Klamt in the COSMOSPACE thermo-statistical approach (Klamt et al. 2002) and used in all COSMO-RS relevant papers even if it is presented differently.

In this context, the interaction energies ε_{ij} are denoted (like in chapter II) as E_{int} that includes the electrostatic misfit interactions and the hydrogen bonding. Note that E_{vdW} is not included in Eq. (V-83) *i.e.* in the statistical averaging because it is not a function of individual surface contacts (Eckert and Klamt 2002). Instead, E_{vdW} is added to the reference energy in solution *a posteriori*.

The fact that the above equations are all determined in the grand canonical ensemble where the chemical potentials are absolute chemical potentials is an additional proof of why COSMO-RS is an excellent activity coefficients model for liquids (since the latter is a difference of chemical potentials in the liquid state). It is worth mentioning that this view of liquid interactions does not include any concept of translational motions.

However, the chemical potential of the gas phase is not yet provided in this surfaces pairwise view of molecules. This is done in the following paragraph.

V.5. Surfaces pairwise view of molecules in the gas phase (steps 2, 3, 4 and 5)

The starting point of the COSMO-RS calculation is the screened molecule in the conductor state (COSMO state, *i.e.* Steps 3 and 4). In order to make a correspondence between this state and the immobilized molecule in the ideal gas phase (Step 2), we introduce the following "pseudo-chemical" reaction and more exactly the following equilibrium that accounts for the transformation between Steps 2 and 3 and Step 4:

$$\text{gas phase (Step 2)} + \text{cavity (Step 3)} \rightleftharpoons \text{COSMO state (Step 4)} \quad \text{(V-84)}$$

This equation provides several important information regarding the thermodynamic pathway given in Figure V-1:

- In the COSMO state (Step 4) the molecule is embedded by its surface elements having charge σ_i and probability p_i. The quantum calculations tools provide directly in this step energy data $\tilde{E}^{COSMO} = E^{COSMO} + E^{diel}$ and E^{diel}. The latter is calculated from the screening charge (q_i) and the local electrostatic potential ϕ_i

on each piece of surface. q_i is the extensive property and ϕ_i is the intensive associated property. Even if this step is directly chosen as the starting point to predict liquid properties in the COSMO-RS framework, it is interesting to decompose step 4 into 2 preliminary steps (namely Step 2 and Step 3 in the thermodynamic pathway) for the reasons discussed below.

- In the cavity (step 3) the molecule is located in a cavity created in an ideal conductor environment (where the total electrostatic potential is null). Let us denote the chemical potential expressed in this reference state as μ_i^{cav}. Furthermore, it is important to notice that each piece of surface has a screening charge (q_i) and a local electrostatic potential ϕ_i. Similarly as it is classically done with the couple of variables p and V to take into account the pressures forces ($-pV$) term, the electrostatic energy term must be taken into account in the grand canonical ensemble. This electrostatic energy is called total screening dielectric energy in the COSMO-RS framework and is given as:

$$E_{diel} = \sum_i q_i \phi_i \tag{V-85}$$

Thus, the relation $pV = kT \ln \xi$ (equation V-59) becomes:

$$-E_{diel} = kT \ln \xi \left(\text{if } q \text{ and } \phi \text{ are the variables of the system}\right) \tag{V-86}$$

- Finally, in the gas state (step 2) the molecule is immobilized with no translational motions allowed. Let us denote the chemical potential expressed in this reference state as $\mu_i^{gas,immo}$.

In order to use the COSMO-RS prediction capacities of chemical potentials of the pairwise interactions surfaces in the gas phase, we now use a pseudo molecule called “vacuum” which does not contain any atom as solvent in step 5. By using the same view of molecular interactions, it is necessary to take into account the different types of interactions between this vacuum molecule and any given solute. Since there is no atom in the vacuum molecule, only the electrostatic misfit interactions energies have to be considered. The question is now to calculate the chemical potential of the cavity using the same absolute reference, in order to obtain the value of the chemical potential of the immobilized molecule in the gas phase (step 2) using equation (V-87). Importantly it must be kept in mind that E_{diel} corresponds to an energy but not to a chemical potential; exactly as for an ideal gas the pV term is different from the chemical potential. This is a direct consequence of considering successive Legendre transforms of the energy. For solving this problem, we consider the infinite dilution in the vacuum, the COSMO-vac reference state and denote the chemical potential as $\mu_i^{COSMO-vac}$.

From the reaction equation (V-85), it can be written that

$$\mu_i^{gas,immo} + \mu_i^{cav} = \mu_i^{COSMO-vac} \tag{V-87}$$

The latter equation should be rewritten as:

$$\mu_i^{gas,immo} = \mu_i^{COSMO-vac} - \mu_i^{cav} \tag{V-88}$$

$\mu_i^{COSMO-vac}$ is computed directly in the COSMO-RS algorithm using the same equations as those used in the liquid states, except that in this case there is no combinational term to add since in reality the molecule is isolated in an hypothetical ideal gas phase , *i.e.* the vacuum (there is no geometrical constraints) like in the liquid phase.

It remains to determine the expression of μ_i^{cav} in order to predict the value of $\mu_i^{gas,immo}$. To do so, we calculate as previously (in section V.4) the grand partition function of the interactions between surfaces elements *i* and *j* forming the cavity, given by:

$$\ln \xi = \frac{1}{2} \sum_{i,j} x_i x_j e^{\beta\left(\mu_i + \mu_j - \varepsilon_{ij}\right)} \tag{V-89}$$

$$\ln \xi_{cav} = \alpha \ln \xi \tag{V-90}$$

In order to obtain a value of the grand partition function we consider two different contributions, which are respectively external and internal to the cavity:

$$\ln \xi_{cav} = \ln \xi_{cav,ext} + \ln \xi_{cav,int} \tag{V-91}$$

- The external contribution to the grand partition function of the cavity is calculated considering that the surface elements forming the cavity are immobile and belong to a same perfect conductor, so that $\varepsilon_{ij} = 0$. This gives:

$$\ln \xi_{cav,ext} = \frac{1}{2} \sum_{i,j} x_i x_j e^{\beta\left(\mu_{i,ext} + \mu_{j,ext}\right)} \tag{V-92}$$

  As demonstrated in the previous section, the normalization condition of this relation is:

$$-\beta\mu_{i,ext} = \ln\left(\sum_j x_j e^{\beta\mu_j} \right) \quad , \forall i \tag{V-93}$$

  The trivial solution of the latter equation is $\mu_{i,ext} = 0$, $\forall i$, so that:

$$\ln \xi_{cav,ext} = \frac{1}{2} \tag{V-94}$$

- The internal contribution to the grand partition function of the cavity is given from equation V-87 (which is an analogue of the fundamental relations of thermodynamic

quantities determination in the grand canonical ensemble in the case of electrostatic interactions) as:

$$\ln \xi_{cav,\text{int}} = \frac{-E_{diel}}{kT} \tag{V-95}$$

This relation allows the definition of the origin of the chemical potentials μ_i.

Finally, we obtain:

$$\ln \xi_{cav} = \ln \xi_{cav,ext} + \ln \xi_{cav,\text{int}} = \frac{1}{2} - \frac{E_{diel}}{kT} \tag{V-96}$$

The latter relation can be rewritten as:

$$\ln \xi_{cav} = \frac{1}{2} e^{\frac{2a}{RT}} = \frac{1}{2} \sum_{i,j} x_i x_j e^{\beta\left(\mu_{i,cav} + \mu_{j,cav}\right)} \tag{V-97}$$

This leads directly to

$$\mu_{i,cav} = \mu_{j,cav} = a \tag{V-98}$$

where a is calculated from the relation

$$e^{\frac{2a}{RT}} = 1 - \frac{2E_{diel}}{RT} \tag{V-99}$$

so that:

$$a = \frac{RT}{2} \ln\left(1 - \frac{2E_{diel}}{RT}\right) \tag{V-100}$$

Finally, the chemical potential of the entire cavity becomes:

$$\mu_{cav} = \sum_i x_i \mu_{i,cav} = a = \frac{RT}{2} \ln\left(1 - \frac{2E_{diel}}{RT}\right) \tag{V-101}$$

It follows that the chemical potential of the immobilized molecule in the gas phase is given by

$$\mu_i^{gas,immo} = \mu_i^{COSMO-vac} - \mu_i^{cav} = \mu_i^{COSMO-vac} - \frac{RT}{2} \ln\left(1 - \frac{2E_{diel}}{RT}\right) \tag{V-102}$$

The previous new relation permits to link in a consistent way the calculation of the chemical potentials in the gas phase (immobile molecule) and in the perfect conductor.

Now it has been established that, in the grand canonical ensemble, the chemical potentials of a given molecule *i* viewed as an ensemble of charged surfaces is calculated both in the immobilized liquid state as well as in the immobilized gas phase using the COSMO-RS method in a consistent thermodynamic pathway. It remains to take into account the translational, rotational and vibrationnal motions of the molecule in order to give a more

realistic description of the pathway from gas phase to condensed phase. It can be assumed that for a given molecule, the rotational and vibrationnal motions do not depend on the physical state of the fluid. This simplifies the calculation tasks to the treatment of the translational motions. This is the purpose of the next section.

V.6. Treatment of translational motions in the pathway from the gas phase to condensed phase

Regarding the treatment of the translational motions, the main contribution to the chemical potential will come from the entropic effects (-TΔS).

As demonstrated earlier in step 1, in the gas phase the absolute entropy is given as

$$S_{trans}^{\circ,gas} = R\left[\frac{3}{2}\ln\left(\frac{2\pi m}{h^2}\right)+\frac{5}{2}\ln\left(kT\right)-\ln\left(p\right)+\frac{5}{2}\right] \qquad \text{(V-103)}$$

By using a similar reasoning (noting that the pV term is not equal to RT for a liquid) it can be demonstrated that, in the liquid phase, the absolute entropy becomes

$$S_{trans}^{\circ,liq} = R\left[\frac{3}{2}\ln\left(\frac{2\pi mkT}{h^2}\right)+\ln\left(v_{liq}\right)+\frac{5}{2}\right] \qquad \text{(V-104)}$$

V.7. Summary of the presentation of the pathway and application to the prediction of ΔG_{vap} and ΔH_{vap} at the normal boiling point T_b

In summary, in order to go from the hypothetical ideal gas phase to condensed phase using a consistent thermodynamic pathway in which any molecule are viewed as an ensemble of electrostatic charged surfaces, it is necessary to:

- Immobilize the molecule in the gas phase and compute its corresponding chemical potential (this passes through two intermediary steps, the COSMO-vac reference state where the molecule is taken at infinite dilution in the vacuum molecule and a robust treatment of the cavity creations)
- Immobilize the molecule in the liquid state and compute its corresponding chemical potential (this is a direct application of the COSMO-RS method that has been developed for this purpose)
- Take into account translational motions in both phases

Thus, the respective chemical potentials of a given molecule in the liquid and gas phases are

$$\begin{cases} \mu_i^{liq} = \mu_i^{liq,COSMO-RS} - TS_i^{\circ,liq} \\ \mu_i^{gas} = \mu_i^{COSMO-vac} - \dfrac{RT}{2}\ln\left(1 - \dfrac{2E_{diel}}{RT}\right) - TS_i^{\circ,gas} \end{cases} \quad (V\text{-}105)$$

It can be deduced that:

$$\Delta G_{vap} = \mu_i^{gas} - \mu_i^{liq}$$
$$= \left[\mu_i^{COSMO-vac} - \frac{RT}{2}\ln\left(1 - \frac{2E_{diel}}{RT}\right) - TS_i^{\circ,gas}\right] - \left(\mu_i^{liq,COSMO-RS} - TS_i^{\circ,liq}\right) \quad (V\text{-}106)$$

Since the COSMO-RS model is able to take into account the temperature dependency of the predicted chemical potentials, it is possible to check the performance and the consistency of our proposed thermodynamic pathway to go from gas phase to condensed phase, at the normal boiling point for several compounds. The corresponding values of the heat of vaporization are determined from the knowledge of the chemical potentials data, using a quite simple numerical derivation scheme.

The results obtained are summarized in Table V-2 and illustrated in Figure V-2 and Figure V-3. These results are in very good agreement with experimental data.

This study demonstrates that it is possible to use a robust and consistent thermodynamic pathway in order to predict relevant thermochemistry data required for the transition between gas phase and condensed phase. Combined with the gas phase formation properties method introduced in chapter IV, due to its thermodynamical consistency this pathway opens the door of thermochemistry to liquid state models like COSMO-RS. This will be very helpful especially to handle the prediction of excess properties (like activity coefficients) of systems containing electrolytes that passes through a robust and consistent treatment of physical interactions (SR and LR) as well as chemical solvation.

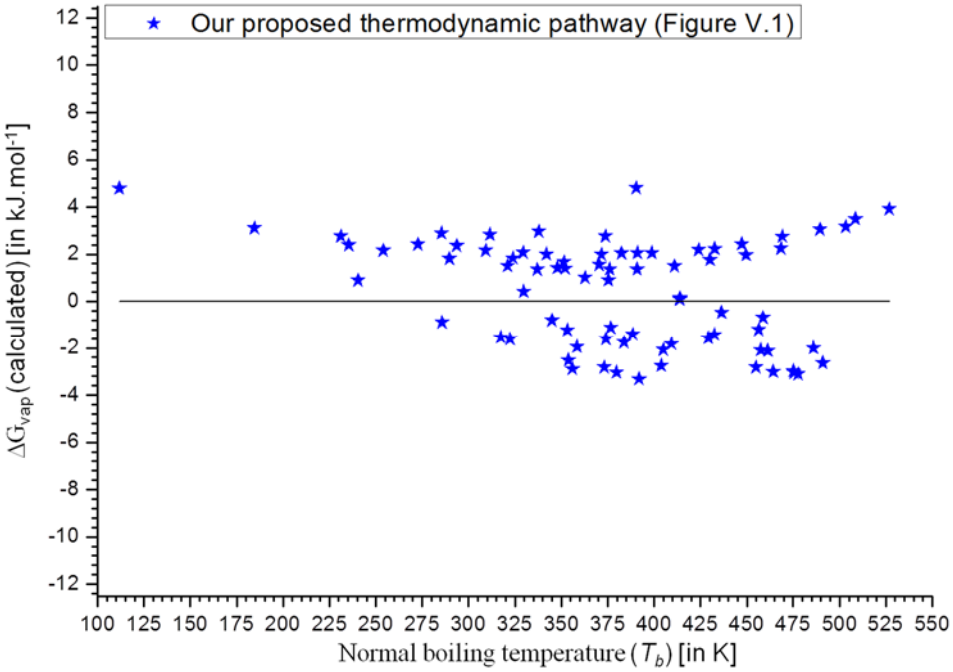


***Figure V-2*: Vaporization Gibbs free energies computed at the normal boiling temperature using our new thermodynamic pathway.**

This figure demonstrates that without identifying any parameter, it is possible to predict the vaporization Gibbs free energy of several compounds on a large range of temperature (from 100 to 550 K) with an accuracy lower than 1kcal/mol.

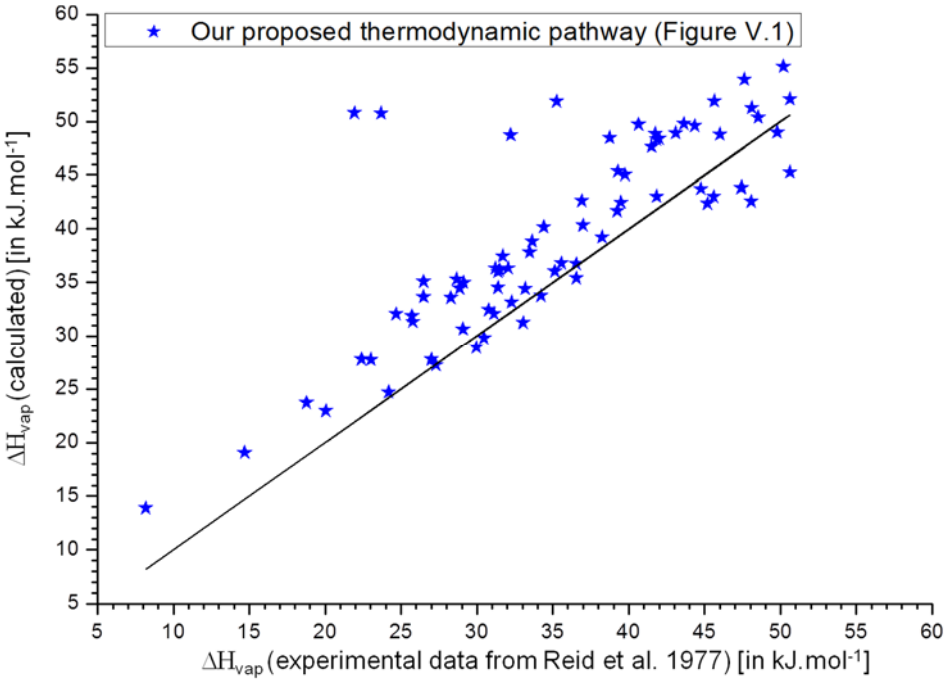


***Figure V-3*: Vaporization enthalpies computed at the normal boiling temperature using our new thermodynamic pathway.**

The computed ΔH_{vap} data are derived from a numerical derivation of the corresponding Gibbs free energies of vaporization, using the classical relation of thermodynamics. This figure demonstrates that without identifying any parameter, it is possible to also predict the vaporization enthalpies ΔH_{vap} of several compounds on a large range of temperature (from 100 to 550 K) with an accuracy lower than 4 kJ.mol^{-1} (i.e. 1 kcal mol^{-1}).

Table V-2 : Summary of the vaporization data computed at the normal boiling temperature using our new thermodynamic pathway.

Molecules	T_b [K]	E_{diel} [kJ.mol^{-1}]	a [kJ.mol^{-1}]	$\mu_i^{COSMO-vac}$ [kJ.mol^{-1}]	$\mu_i^{liq,COSMO-RS}$ [kJ.mol^{-1}]	$-T\left(S_i^{\circ,gas}-S_i^{\circ,liq}\right)$ [kJ.mol^{-1}]	ΔG_{vap} [kJ.mol^{-1}]	ΔH_{vap}(calc) [kJ.mol^{-1}]	ΔH_{vap}(exp.) [kJ.mol^{-1}]
methane	111.70	-1.85	0.75	0.63	-10.49	5.58	4.79	13.91	8.18
ethane	184.50	-1.85	0.94	0.45	-12.92	9.32	3.11	19.06	14.70
propane	231.10	-2.23	1.15	0.56	-14.85	11.49	2.77	23.71	18.77
butane	272.70	-2.57	1.34	0.67	-16.48	13.39	2.42	27.77	22.39
pentane	309.20	-2.94	1.53	0.78	-17.91	15.01	2.15	31.34	25.77
hexane	341.90	-3.31	1.71	0.89	-19.25	16.43	2.00	34.47	28.85
n-heptane	371.60	-3.68	1.88	1.01	-20.54	17.68	1.98	37.49	31.69
octane	398.80	-4.04	2.05	1.12	-21.79	18.81	2.06	40.20	34.41
n-nonane	424.00	-4.42	2.21	1.23	-23.00	19.83	2.19	42.61	36.91
n-decane	447.30	-4.78	2.37	1.34	-24.20	20.75	2.43	45.37	39.27
n-undecane	469.10	-5.16	2.52	1.45	-25.41	21.58	2.76	47.64	41.50
dodecane	489.50	-5.53	2.67	1.57	-26.53	22.38	3.05	49.77	43.63
tridecane	508.60	-5.90	2.82	1.68	-27.72	23.08	3.49	51.90	45.64
tetradecane	526.70	-6.28	2.96	1.79	-28.84	23.76	3.91	53.93	47.61
methanol	337.80	-28.96	4.32	23.99	-2.59	19.30	2.97	51.87	35.25
ethanol	351.50	-27.94	4.39	22.34	-2.82	19.12	1.66	48.49	38.74
propanol	370.40	-27.87	4.54	21.98	-3.58	19.46	1.55	48.85	41.75
1-butanol	390.90	-27.94	4.71	21.86	-4.24	20.02	1.36	48.92	43.09
1-pentanol	411.00	-28.31	4.90	22.01	-5.02	20.63	1.50	49.59	44.34
1-hexanol	430.20	-28.66	5.07	22.12	-5.93	21.21	1.77	50.38	48.53
1-heptanol	449.50	-28.96	5.24	22.18	-6.87	21.85	1.96	51.26	48.11
1-octanol	468.40	-29.29	5.40	22.24	-7.88	22.48	2.24	52.10	50.62
1-decanol	503.40	-30.10	5.72	22.55	-9.97	23.64	3.16	55.14	50.20
benzene	353.30	-15.14	3.56	12.69	-7.55	17.90	-1.23	32.44	30.76
toluene	383.80	-15.53	3.79	12.27	-8.90	19.14	-1.76	34.42	33.17
ethylbenzene	409.30	-15.68	3.95	12.10	-10.14	20.10	-1.82	36.79	35.56
n-propylbenzene	432.40	-16.11	4.13	12.23	-11.43	20.94	-1.42	39.26	38.24
butylbenzene	456.40	-16.53	4.31	12.36	-12.62	21.87	-1.21	41.67	39.24
propanone	329.40	-34.43	4.47	24.65	0.96	17.16	2.07	35.00	29.12
butanone	352.08	-32.52	4.60	22.78	-1.13	17.91	1.40	36.35	31.21
3-pentanone	375.50	-31.02	4.74	21.04	-3.34	18.73	0.90	37.84	33.47
formaldehyde	254.00	-24.44	3.36	19.77	0.12	14.14	2.15	27.74	23.01
acetaldehyde	293.60	-30.66	3.98	22.92	0.85	15.74	2.35	31.86	25.73
propanal	321.00	-29.53	4.19	21.64	-0.75	16.69	1.51	33.57	28.28
butanal	348.00	-29.40	4.43	21.40	-2.10	17.66	1.42	36.16	31.50
isobutanal	337.00	-27.71	4.25	19.75	-2.87	17.01	1.35	34.53	31.37
valeraldehyde	376.00	-29.88	4.69	21.60	-3.21	18.75	1.36	38.85	33.63
phenol	455.00	-35.27	5.63	28.63	2.15	23.66	-2.82	43.02	45.60
2-methylphenol	464.20	-33.59	5.62	26.25	0.06	23.58	-3.01	42.39	45.18
3-hydroxytoluene	475.40	-35.76	5.83	28.10	1.08	24.25	-3.06	43.84	47.40

Molecules	T_b [K]	E_{diel} [kJ.mol^{-1}]	a [kJ.mol^{-1}]	$\mu_i^{COSMO-vac}$ [kJ.mol^{-1}]	$\mu_i^{liq,COSMO-RS}$ [kJ.mol^{-1}]	$-T\left(S_i^{\circ,gas}-S_i^{\circ,liq}\right)$ [kJ.mol^{-1}]	ΔG_{vap} [kJ.mol^{-1}]	ΔH_{vap}(calc) [kJ.mol^{-1}]	ΔH_{vap}(exp.) [kJ.mol^{-1}]
4-methylphenol	475.10	-35.23	5.80	27.98	0.95	24.21	-2.98	43.84	47.44
2-ethylphenol	477.70	-32.67	5.68	24.49	-1.87	23.80	-3.11	42.56	48.07
4-ethylphenol	491.00	-35.36	5.94	27.76	-0.11	24.57	-2.63	45.24	50.62
formicacid	373.80	-38.54	5.05	32.45	3.26	21.36	2.77	50.80	21.92
aceticacid	391.10	-39.09	5.24	31.46	2.74	21.42	2.06	50.75	23.68
acrylicacid	414.00	-38.69	5.43	30.31	2.55	22.19	0.14	48.81	46.02
propionicacid	414.00	-37.53	5.38	29.59	2.12	22.00	0.09	48.72	32.21
butyricacid	436.40	-37.39	5.58	29.30	1.57	22.63	-0.48	48.39	42.00
n-bucooh	458.70	-37.69	5.78	29.35	0.94	23.32	-0.70	48.98	49.78
pyridine	388.50	-29.01	4.75	22.52	-0.97	20.15	-1.41	36.07	35.14
1,2-diaminoethane	390.40	-47.52	5.54	40.32	9.27	20.72	4.80	48.34	41.83
dimethylamine	289.70	-19.79	3.44	15.74	-4.61	15.10	1.81	33.65	26.48
acetylchloride	323.90	-26.49	4.08	18.53	-4.19	16.81	1.83	35.31	28.66
bromoethane	311.50	-17.96	3.50	12.41	-9.96	16.04	2.83	35.12	26.48
chloroethane	285.40	-17.41	3.26	12.04	-8.78	14.66	2.88	32.06	24.68
fluoroethane	235.40	-16.47	2.82	12.20	-5.10	12.10	2.38	25.16	NaN
aziridine	329.80	-28.51	4.22	23.63	1.08	17.91	0.42	36.35	32.04
bromobenzene	429.20	-17.69	4.26	13.42	-10.88	21.61	-1.57	40.20	NaN
chlorobenzene	404.90	-17.14	4.06	12.95	-9.39	20.35	-2.07	36.73	36.54
fluorobenzene	358.50	-16.91	3.75	13.27	-6.53	17.99	-1.94	31.80	NaN
piperidine	379.70	-18.23	3.99	13.33	-6.76	19.14	-3.04	33.76	34.22
iodobenzene	461.40	-18.30	4.52	14.13	-11.50	23.24	-2.12	42.50	39.49
aniline	457.50	-36.77	5.73	29.31	1.98	23.68	-2.08	43.05	41.83
dipropylamine	382.40	-18.51	4.03	12.58	-11.90	18.40	2.05	40.34	36.98
triethylamine	362.70	-10.40	3.12	5.60	-15.84	17.33	1.00	36.02	31.37
dibutylamine	432.80	-19.27	4.43	12.82	-14.24	20.41	2.22	45.03	39.74
ethylbenzoate	485.90	-33.93	5.82	23.76	-3.77	23.72	-2.00	43.75	44.76
water	373.20	-40.13	5.11	23.84	-2.06	23.61	-2.81	49.72	40.65
cyclopropane	240.40	-6.67	2.04	4.49	-10.71	12.28	0.89	22.94	20.04
cyclobutane	285.70	-2.67	1.40	1.10	-13.87	14.45	-0.88	24.67	24.18
cyclopentene	317.40	-6.95	2.42	4.77	-12.09	15.98	-1.54	27.75	26.98
cyclopentanone	403.90	-32.89	5.08	24.20	1.06	20.81	-2.76	35.40	36.56
cyclopentane	322.40	-2.52	1.42	0.77	-15.15	16.12	-1.62	27.23	27.29
cyclohexene	356.10	-7.48	2.67	4.82	-12.73	17.77	-2.88	29.78	30.45
cyclohexane	353.90	-2.35	1.40	0.35	-16.07	17.53	-2.52	28.89	29.95
methylcyclopentane	345.00	-2.91	1.59	0.82	-16.86	16.89	-0.80	30.63	29.07
cycloheptane	391.90	-2.59	1.55	0.50	-16.93	19.19	-3.31	31.25	33.05
ethylcyclopentane	376.60	-3.17	1.73	0.78	-18.07	18.24	-1.13	33.17	32.27
methylcyclohexane	374.10	-2.91	1.64	0.60	-17.60	18.17	-1.61	32.08	31.12

General conclusion

The goal of our study was to develop a predictive thermodynamic model that enables to characterize the equilibrium properties of aqueous mixtures encountered in food and biological systems (the kinetic aspect is not treated herein) and to compensate for the lack of experimental data or more exactly for the huge number of variables that appear in a complex biological solution. For this purpose, it is necessary to extend the performance of the existing thermodynamic models up to the prediction of the physico-chemical properties of systems containing simultaneously electrolytes and non-electrolytes, solvated and un-solvated substances, etc..

To do this, it was necessary to be familiar with several important concepts; this has been the purpose of Chapter I. The main outlines are summarized below.

The description of the evolution of a food during processing uses a combined modelling approach. The global model should include a water transfer model, which determines the water content that is further used to predict physico-chemical properties (pH, pK_a, a_w, E_h) with a thermodynamic model. It is outlined in chapter I that all these properties are deduced from the knowledge of the chemical potential μ_i.

Several fundamental notions are required to describe the chemical potential, which characterizes the equilibrium state of the studied system. The chemical potential of a given compound in solution is calculated as the sum of two terms: *i)* the reference chemical potential of the species (which should correspond to one of its formation properties, the Gibbs free energy of formation in order to perform thermochemistry calculations) and *ii)* the activity coefficient (which is an excess property).

It is shown in chapter I that the thermodynamic properties of a mixture depend on the interaction forces that exist between the species in the mixture. The pieces of the puzzle of molecular behaviour are reviewed. Food and biological systems are treated as multicomponent aqueous mixtures that contain simultaneously water, salts or electrolytes, sugars, etc. Thus the system is characterized by the presence of both molecular species and ionic species, resulting in three different types of interactions: ion-ion, molecule-molecule and ion-molecule. Ion-ion interactions are governed by electrostatic forces between ions that have a much longer range than other intermolecular forces. Molecule-molecule and ion-molecule interaction forces are known to be short-range in nature. The excess Gibbs energy of systems containing electrolytes is considered as the sum of two terms, one related to long-range (LR) forces between ions and the other to short-range (SR) forces between all the species. The UNIFAC model has been adopted by the IP-GePEB team and extended to the treatment of food and biological systems. In the resulting model (ULPDHS), the term of Pitzer based on the Debye-Hückel theory was added to the UNIFAC-Larsen model in previous studies (Achard 1992, Catté 1994, Ben Gaïda 2007). Ions were considered as UNIFAC independent groups. Considering the solvation of charged species giving clusters by means of a hydration number for each ion at infinite dilution, two interaction parameters (water-

anion and water-cation interactions) were sufficient to characterize water-salt systems. However, some UNIFAC group interactions parameters are not available in the literature and the extension to highly complex solutions is not achievable. In the present study, it was decided to study another alternative that should overcome these limitations.

The COSMO-RS method (Klamt 1995, 1998, 2005 and 2011) is an *a priori* fully predictive method widely used in chemical engineering. Our goal was to apply it to biological systems. COSMO-RS is an excess Gibbs energy method like UNIFAC, with the advantage that it does not require identification of interactions parameters between chemical groups.

A detailed description of the COSMO-RS method is given in Chapter II. This method combines concepts of electrostatics, quantum chemistry and statistical thermodynamics to predict the chemical potential of liquid systems. Liquid in COSMO-RS is considered to be an ensemble of almost closely packed ideally screened molecules, and the interactions of the molecules are expressed as pairwise interactions of the screening charges. This includes electrostatic interactions as well as hydrogen bonding (and the van der Waals interactions). By this reduction of molecular interactions to surface contacts, the statistical thermodynamics is reduced to a simple set of equations, which are similar to, but even somewhat more accurate than, the UNIQUAC type of equations used in the UNIFAC models. Several atom-based internal constants are necessary using this method (*e.g.* the COSMO-radii of the elements which is used for cavity construction and corresponds to the Bondi radii multiplied by a scaling factor of about 1.17; the ring correction coefficient ω, the coefficient λ used in the combinatorial part of the chemical potential; the transfer constant η which connects gas and condensed reference states. These constants have been determined once for several atomic elements (H, C, N, O, F, P, S, Cl, Br ou I) and have since been improved in several revisions of the COSMO-RS model that was originally developed to treat aqueous mixtures containing mainly neutral compounds.

The applicability of COSMO-RS model to predict the activity coefficients of systems containing electrolytes is at its beginning. Our contribution was to reinforce these new applications. In chapter III, we developed a new tool able to extend the performance of the COSMO-RS method towards the prediction of activity coefficients in binary aqueous electrolyte systems (water-salt). This prediction tool has the same structure as the ULPDHS model (Achard 1992). For this purpose, the molar fractions of each compound of mixtures containing water, one anion and one cation (or its corresponding cluster) are calculated using the equations developed in the chemical part of the ULPDHS model. Then, for each mixture composition the corresponding activity coefficients are predicted using the default parameters of the COSMOthermX software (if not stated otherwise), and mean activity coefficients of salts are deduced. Long-range interactions between ions of opposite

charges are dominated by electrostatic forces and are accounted for by a PDH term. The resulting model called "COSMO-RS-PDHS" is fully predictive with no identified parameter.

To test the predictive power of the COSMO-RS-PDHS model, it was necessary to investigate the determination of the hydration number of several cations (namely H^+; Li^+, Na^+, K^+, Mg^{2+} and Ca^{2+}, Rb^+, Cs^+, Be^{2+}, Sr^{2+} and Ba^{2+} to complete the IA and IIA columns) and to determine a new set of COSMO-radii suitable for these elements. A similar study can be performed for anions which are generally less hydrated than cations. Indeed the main elements of the IA and IIA group (except H, Li, Na, Be and Mg) are not yet introduced in the current COSMO-RS parameterization. In this context, a new prediction method of the maximum value of the hydration number (that is the key parameter for taking into account the hydration of the ion by water molecules) was introduced in Chapter III. For this purpose, the hydration number (n) is defined as the number of water molecules forming a stoichiometric complex with the studied ion. Four complementary procedures (three procedures at the gas phase and the last one at infinite dilution in water) have been described and used to predict the corresponding hydration number values. In order to perform COSMO calculations, the COSMO-radii had to be determined. Because this is not the case in the litterature regarding almost all of the studied ionic elements, a strategy of estimation of radii has also been introduced in Chapter III. This strategy starts with the gas phase calculations where DFT results (mainly energies and optimized geometries) are independent of ionic radii. Thus, by analyzing the calculated mean distances between the oxygen atom of one water molecule and a given ion (inside the cluster), it becomes possible to choose a set of radii that are in agreement with these values. Importantly, the values of the distance ion-water ($d_{ion\text{-}water}$) calculated in this study are consistent with the sum $R_{ion}+R_{water}$ given by Marcus (1988, 1997 and 2012) where R_{water} = 1.38 nm. This justifies the choice of the ionic radii values given by Marcus as reference values for calculating COSMO-radii of cations and applying the scaling factor of 1.17 times R_{ion}. This study of hydration process has also demonstrated that the DFT/QM level used in COSMO-RS is enough accurate to determine the hydration number of ions.

Then, the predictive power of the COSMO-RS-PDHS model has been investigated. For the elements (like H, Li, Na, Cl, Br, I) that are partially parameterized in the current COSMO-RS algorithm, the predicted water activity and mean activity coefficients of salts, were very satisfying. Likewise, even for systems containing unparameterized elements such as K^+, the predicted activity coefficients data were also in very good agreement with literature data when we use the new set of COSMO-radii proposed within this study. Thus, similarly to the ULPDHS model that has been validated on a large set of systems containing electrolytes, the COSMO-RS-PDHS method appears as a promising and fully predictive activity coefficient model, *i.e.* an excellent model for excess properties. In order to enlarge the domain of validity of the COSMO-RS-PDHS model, it will be

necessary to use a variable hydration of ionic species. It has been demonstrated that this can be performed if one has an accurate treatment of the different chemical equilibria between the different hydrated clusters formed by water molecules and a given ion. To do so, it is necessary to compute the equilibrium constant of each hydration reaction (Ben Gaïda 2007). The latter is calculated in reference conditions, from the formation properties of the species in solution. Thus, for consistency purpose, we have decided to develop a COSMO-RS based method to test the feasibility of such thermochemistry calculations, on a set of neutral compounds.

As mentioned earlier, food and biological systems also contain a large number of species that are involved in chemical equilibria (*e.g.* dissociation, complexation and redox). To handle the prediction of these properties, it is necessary to know the chemical potential of formation at infinite dilution in water. In Chapter IV, the performance of the COSMO-RS method in predicting the pK_a of amino-acids and peptides was tested by comparing the COSMO-RS results with those obtained by other pK_a prediction tools like ChemAxon and ACD/Labs methods, as well as several experimental data. The predicted results were in very good agreement with literature data, and COSMO-RS appeared to perform as well as the other methods. However, COSMO-RS uses quantum calculated "free" energies to predict pK_a using a linear free energy relationship (LFER) instead of chemical potential of formation data. This is a strong limitation for extrapolating COSMO-RS predictions to larger set of reaction data (complexation, solvation, redox).

A new method of prediction of formation properties of several molecules of interest using the same input data as those of the COSMO-RS model (*i.e.* results of quantum DFT/COSMO calculations), was introduced and validated in Chapter IV. The development of this model, that is fully predictive and allows the treatment of conformations, permits to explain the physical origins of the LFER parameters (especially the slope that is lower than the theoretically expected value which equals 1), as detailed in chapter IV.

Since the processes studied within this work do occur in solution, it was necessary to convert gas phase data to their corresponding values at infinite dilution in water. This transition is done by using several facilities included in the COSMO-RS algorithm like the calculation tool of solvation free energies data. In chapter IV, a comparison between the results obtained using the solvation data predicted by the COSMO-RS model and those obtained using the results of the thermodynamic relations of phase equilibria has shown a very good agreement between all data. However, to use these relations derived from phase equilibria, it is necessary to use both vapour pressure and activity coefficient data. Thus the solvation process has been combined with the gas phase formation properties to predict the chemical potential obtained at infinite dilution in water for a large set of neutral compounds including free radicals. The chemical potential values were successfully used to predict several portions of the Pourbaix diagram of phenol and the isomers of the cresol (including

both pK_a and standard redox potentials E^0 data) that are in very good agreement with experimental data. This study also provides several important explanations about the redox mechanism and the 3D structure of the free radical, especially in the case of phenol.

Since the approximate composition of a meat muscle is known, a section was devoted to the evaluation of the performance of the COSMO-RS model as a qualitative understanding tool for phenomena appearing in complex mixtures, especially in the context of the Na^- project. To do so, the σ-profile of meat has been approximated. This permits to argue that (in meat processes) the most interesting candidate for replacing sodium is the potassium ion; similarly a mixture of fluorine and iodine can be used to replace the chlorine anion. Likewise, the calculated σ-profiles of the raw and salted meats have been used to explain why the SM muscles of beef and pork have the same titration curves (*i.e.* pH *vs.* volume of acid or base curves) and to illustrate the fact that the addition of salt does not change the titration curve of meat. But it is also proved that the addition of salt changes the sorption curve of meat, since the latter includes water activity data.

This work demonstrates that the COSMO-RS method is an excellent activity coefficient model, *i.e.* an excellent model for excess properties that can be extended to the representation of the equilibrium properties of complex electrolytes solutions and consequently of foods and biological systems.

However, the previous results have provided a thorough experience showing that the solvation concept used in the COSMO-RS algorithm, even if it is very efficient for neutral compounds, allows only a partial description of the process of transition between gas and condensed phases. Thus the whole study was revisited in order to develop a complementary and physically consistent view of a solvation process that passes through the virtual step of charged surfaces molecules in both phases. This pathway introduced in Chapter V contains several steps that combine simultaneously concepts of classical thermodynamics, quantum physics, electrostatics and statistical physics. At each step the scientific knowledge and data are pointed out. This study demonstrates that it is possible to define a robust and consistent thermodynamic pathway in order to predict relevant thermochemistry data required for the transition between gas phase and condensed phase with the simultaneous determination of energies and chemical potentials. The results are validated with the determination of Gibbs free energies of vaporization (so vapour pressures) and enthalpies of vaporization. This new pathway would enable to reduce the number of parameters of the COSMO-RS model trough the suppression of the transfer constant η, which is replaced by the use of the electrostatic energy of the cavity (E_{diel}) in a suitable expression for determining the chemical potential. No regression of new parameters is necessary, showing that: *i-)* the COSMO

model is intrinsically very robust; *ii-)* the inaccuracy of quantum mechanics calculations cancels when it is used as differences between different states, namely ideal gas and COSMO state.

Combined with the gas phase formation properties method, this pathway opens the door to a completely predictive thermochemistry tool suited for complex systems containing simultaneously neutral liquid and gaseous compounds, and then for charged species. This will be very helpful to handle the prediction of excess properties of systems containing electrolytes that passes through consistent treatment of physical interactions (SR and LR) as well as chemical solvation.

The results obtained with this new description of the pathway from gas to condensed phase could be improved if a re-identification of the COSMO-RS parameters is performed. This would lead to a reduction of the number of parameters since the transfer constant (η that connects reference states in gas and solution), the ring correction coefficient (ω) and the combinatorial part of the chemical potential (coefficient λ) would be removed, at least for the transition from gas phase to condensed phase using this new pathway.

If such identification was performed, it would probably be interesting to restart the whole study using the reverse way that is more familiar to chemical engineers. Even if the COSMO-RS model is one of the most powerful of the existing predictive thermodynamic models, we are convinced that this new study pathway would improve the interest of chemical engineers in the use of such a model.

This work can be used in its current state to investigate the almost of the thermodynamic tasks required in the framework of the "Na^-" project and far beyond. Indeed, this study provides:

- For the final user (especially people involved in meat processing)
  - a prediction tool of physico-chemical properties (a_w, pK_a, pH and E_h) of complex systems like meat in presence of different additives (salts, organic acids, sugars, amino-acids, etc.) enabling thus to test for instance the mixtures allowing a partial or total substitution of sodium chloride;
  - a more physical understanding of the phenomena linked to the discolouration problems of meat linked to the presence of micro-organisms since the growth of the latter depends on the above mentioned physico-chemical properties;
- For the modellers (especially people involved in meat processing)
  - A physico-chemical tool that can be directly combined with heat and substances transfers models even if NaCl is substituted;
  - Explanations about the behaviour of micro-organisms during a given process by the prediction of the evolution of the redox potential and more generally this model can be integrated in microbiological models.

On the long term scale, this work can be used to:

- Investigate several other complex systems like cheese matrices, cereals in which the redox problem of the components is partially involved ;
- Explore new predictive tools of water, salts, sugars, or acids diffusivities in different matrices using the thermodynamic model that would enable to avoid long and expensive experimental measurements of these diffusivities.

Finally, the thermodynamic model developed in this study can be improved trough:

- A more detailed investigation of the variable hydration of ions ;
- The treatment of the complexation reactions which are a more general extension of the hydration (*i.e.* complexation by water molecules).

References

Aaqvist, J. (1990), 'Ion-water interaction potentials derived from free energy perturbation simulations', *The Journal of Physical Chemistry* **94**(21), 8021-8024.

Abraham, S., Cachon R., Colas B., Feron G., DeConinck J.R. (2007), 'E_h and pH gradients in camembert cheese during ripening: Measurements using microelectrodes and correlations with texture', *Int. Dairy J.* **17**, 954-960.

Achard, C. (1992), Modélisation des propriétés d'équilibre de milieux biologiques et alimentaires à l'aide de modèles prédictifs. Contribution à la mise en place d'un logiciel de simulation des procédés biotechnologiques, *PhD thesis*, Université Blaise Pascal-Clermont-Ferrand II.

Achard, C., Dussap, C. and Gros, J. (1994), 'Representation of vapour-liquid equilibria in water-alcohol-electrolyte mixtures with a modified UNIFAC group-contribution method', *Fluid Phase Equilibria* **98**, 71-89.

Alberty, R. A. (2001), 'Use of legendre transforms in chemical thermodynamics (IUPAC technical report)', *Pure and Applied Chemistry* **73**(8), 1349-1380.

Archer, D. G. and Wang, P. (1990), 'The dielectric constant of water and Debye-Hückel limiting low slopes', *J. Phys. Chem. Ref. data* **19** (2), 371-411.

Asghar A., Gray J.I., Booren A.M., Gomaa E.A., Abouzied M.M., Miller E.R. (1991), 'Effects of superanutritional dietary vitamin E levels on sub cellular deposition of -tocopherol in the muscle and on pork quality', *J. Sci. Food Agric.* **57**, 31.

Astruc T., Labas R., Vendeuvre J. L., Martin J. L., Taylor R. G. (2008), 'Beef sausage structure affected by sodium chloride and potassium lactate', *Meat Science.* **80**, 1092-1099.

Asthagiri, D. and Pratt, L. R. (2003), 'Quasi-chemical study of Be^{2+}(aq) speciation', *Chemical physics letters* **371**(5), 613-619.

Barrière, C., Brückner, R., Centeno, D. and Talon, R. (2002), 'Characterisation of the kata gene encoding a catalase and evidence for at least a second catalase activity in staphylococcus xylosus, bacteria used in food fermentation', *FEMS microbiology letters* **216**(2), 277-283.

Barrière, C., Brückner, R. and Talon, R. (2001a), 'Characterization of the single superoxide dismutase of staphylococcus xylosus', *Applied and environmental microbiology* **67**(9), 4096-4104.

Barriere, C., Centeno, D., Lebert, A., Leroy-Sétrin, S., Berdague, J. and Talon, R. (2001b), 'Roles of superoxide dismutase and catalase of staphylococcus xylosus in the inhibition of linoleic acid oxidation', *FEMS Microbiology Letters* **201**(2), 181-185.

Baucour, P. and Daudin, J. (2000), 'Development of a new method for fast measurement of water sorption isotherms in the high humidity range validation on gelatine gel', *Journal of Food Engineering* **44**(2), 97-107.

Becke, A. D. (1988), 'Density-functional exchange-energy approximation with correct asymptotic behavior', *Physical Review A* **38**(6), 3098.

Beglov, D. and Roux, B. (1994), 'Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations', *The Journal of Chemical Physics* **100**(12), 9050-9063.

Belitz, H. D., Grosch, W. and Schieberle, P. (2009), *Food Chemistry, 4th revised and extended Edition*,. Springer Verlag, Berlin, Germany.

Ben-Naim, A. (1987), *Solvation thermodynamics*, Plenum Press, New York.

Ben-Naim, A. (2001), 'On the evolution of the concept of solvation thermodynamics', *Journal of Solution Chemistry* **30**, 475-487.

Ben Gaïda L., Dussap C.G., Gros J.B. (2006), 'Variable hydration of small carbohydrates for predicting equilibrium properties in diluted and concentrated solutions', *Food Chem.* **96**, 387-401.

Ben Gaïda, L. (2007), Méthode de contribution de groupes pour la représentation des propriétés d'équilibre dans les solutions aqueuses. Prise en compte de l'hydratation variable des espèces, *PhD thesis*, Université Blaise Pascal-Clermont-Ferrand II.

Ben Gaïda L., Gros J.B., Dussap C.G. (2010), 'Activity coefficients of concentrated strong and weak electrolytes by a hydration equilibrium and group contribution model', *Fluid Phase Equilibria.* **289**, 40-48.

Bendall, J. and Swatland, H. (1988), 'A review of the relationships of ph with physical aspects of pork quality', *Meat science* **24**(2), 85-126.

Benson, S. W. and Buss, J. H. (1958), 'Additivity rules for the estimation of molecular properties. Thermodynamic properties', *The Journal of Chemical Physics* **29**(3), 546-572.

Benson, S. W., Cruickshank, F. R., Golden, D. M., Haugen, G. R., O'neal, H. E., Rodgers, A. S., Shaw, R. and Walsh, R. (1969), 'Additivity rules for the estimation of thermochemical properties', *Chemical Reviews* **69**(3), 279-324.

Blas, F. J. and Vega, L. F. (1998), 'Prediction of binary and ternary diagrams using the statistical associating fluid theory (saft) equation of state', *Industrial & engineering chemistry research* **37**(2), 660-674.

Bockris, J. O. and Saluja, P. (1972), 'Ionic solvation numbers from compressibilities and ionic vibration potentials measurements', *The Journal of Physical Chemistry* **76**(15), 2140-2151.

Bockris, J. O. and Reddy, A. K. N. (1970), *Modern Electrochemistry*, 3rd edition, Plenum Press, New York.

Boda, A., De, S., Ali, S. M., Tulishetti, S., Khan, S. and Singh, J. K. (2012), 'From microhydration to bulk hydration of Sr^{2+} metal ion: DFT, MP2 and molecular dynamics study', *Journal of Molecular Liquids* **172**, 110-118.

Bondi, A. (1964), 'van der waals volumes and radii', *The Journal of Physical Chemistry* **68**(3), 441-451.

Bonin, J., Costentin, C., Louault, C., Robert, M., Routier, M. and Savéant, J.-M. (2010), 'Intrinsic reactivity and driving force dependence in concerted proton-electron transfers to water illustrated by phenol oxidation', *Proceedings of the National Academy of Sciences* **107**(8), 3367-3372.

Brocklehurst, T. F., Parker, M. L., Gunning, P. A. and Robins, M. M. (1993), Microbiology of food emulsions: physicochemical aspects, *Lipid Technology*, **5**(4), 83-88.

Brooijmans, R., Smit, B., Santos, F., van Riel, J., de Vos, W. M., Hugenholtz, J. et al. (2009), 'Heme and menaquinone induced electron transport in lactic acid bacteria', *Microb Cell Fact* **8**(1), 28.

Buchner, R., Hefter, G. T. and May, P. M. (1999), 'Dielectric relaxation of aqueous nacl solutions', *The Journal of Physical Chemistry A* **103**(1), 1-9.

Buscailhon S., Monin G., Cornet M. (1994), 'Time-related changes in nitrogen fraction and free amino acids of lean tissue of French dry cured ham', *Meat Science* **37**, 449-456.

Byrd, E. F. and Rice, B. M. (2006), 'Improved prediction of heats of formation of energetic materials using quantum mechanical calculations', *The Journal of Physical Chemistry A* **110**(3), 1005-1013.

Car, R. and Parrinello, M. (1985), 'Unified approach for molecular dynamics and density-functional theory', *Physical review letters* **55**(22), 2471.

Catté, M. (1994), Représentation des propriétés d'équilibre des systèmes biologiques par des méthodes de contribution de groupes: solubilité des gaz et comportement des solutions aqueuses de sucres, *PhD thesis*, Université Blaise Pascal-Clermont-Ferrand II.

Catté, M., Dussap, C. G., and Gros, J. B. (1995), A physical chemical UNIFAC model for aqueous solutions of sugars, *Fluid Phase Equilibria* **105**, 1-25.

Chaillou, S., Champomier-Vergès, M.-C., Cornet, M., Crutz-Le Coq, A.M., Dudez, A.-M., Martin, V., Beaufils, S., Darbon-Rongère, E., Bossy, R., Loux, V. et al. (2005), 'The complete genome sequence of the meatborne lactic acid bacterium lactobacillus sakei 23k', *Nature biotechnology* **23**(12), 1527-1533.

Chapman, W. G., Gubbins, K. E., Jackson, G. and Radosz, M. (1989), 'SAFT: Equation-of-state solution model for associating fluids', *Fluid Phase Equilibria* **52**, 31-38.

Chapman, W. G., Gubbins, K. E., Jackson, G. and Radosz, M. (1990), 'New reference equation of state for associating liquids', *Industrial & Engineering Chemistry Research* **29**(8), 1709-1721.

Chipot, C. and Pohorille, A. (1998), 'Folding and translocation of the undecamer of poly-l-leucine across the waterhexane interface. a molecular dynamics study', *Journal of the American Chemical Society* **120**(46), 11912-11924.

Clark, W. M. and Clark, W. M. (1960), '*Oxidation-reduction potentials of organic systems*'.

Cohen, E. R., Taylor, B. N. et al. (1988), *The 1986 CODATA recommended values of the fundamental physical constants*, American Chemical Society and the American Institute of Physics for the National Bureau of Standards.

Coroller, L., Leguérinel, Y., and Mafart, P. (2001), 'Effect of water activities of heating and recovery media on apparent heat resistence of Bacillus cereus spores, *App. Environ. Microbiol.* **67**, 317-322.

COSMOthermX (2011). **COSMOthermX** A Graphical User Interface to the COSMOtherm Program, Tutorial for version C21 0111, COSMOlogic GmbH & Co. KG, Leverkusen, Germany, 07 December 2011.

Costentin, C., Robert, M., Savéant, J.-M. and Tard, C. (2011), 'H-bond relays in proton-coupled electron transfers. oxidation of a phenol concerted with proton transport to a distal base through an OH relay', *Phys. Chem. Chem. Phys.* **13**(12), 5353-5358.

Cotton, F, A. (1971), *Chemical Applications of Group Theory*; 2nd ed., Wiley-Interscience : New York.

Cramer, C. J. and Truhlar, D. G. (1999), Implicit solvation models: equilibria, structure, spectra, and dynamics, *Chemical Reviews* **99**(8), 2161-2200.

Derr, E. L. and Deal, C. H. (1969), 'Analytical solution of groups: correlation of activity coefficients through structural group parameters', *Inst. Chem. Eng. Symp. Ser* **32**(3), 40.

Derr, E. L. and Deal, C. H. (1973), 'Predicted compositions during mixed solvent evaporation from resin solutions using the analytical solutions of groups method', *Adv. Chem. Ser., Am. Chem. Soc.* **124**, 11.

Desnier-Lebert, I. (2004), Prédiction de la croissance de Listeria innocua par une approche phénoménologique: modélisations complémentaires des propriétés du milieu, des transferts d'eau et des cinétiques, *PhD thesis*, Université Blaise Pascal-Clermont-Ferrand II.

Driesner, T., Ha, T.-K. and Seward, T. (2000), 'Oxygen and hydrogen isotope fractionation by hydration complexes of Li^+, Na^+, K^+, Mg^{2+}, F^-, Cl^-, and Br^-: a theoretical study', *Geochimica et COSMOchimica Acta* **64**(17), 3007-3033.

Dzidic, I. and Kebarle, P. (1970), 'Hydration of the alkali ions in the gas phase. enthalpies and entropies of reactions M+(h2o)n-1 + h2o = M+(h2o)n', *The Journal of Physical Chemistry* **74**(7), 1466-1474.

Eckert, F. and Klamt, A. (2002), 'Fast solvent screening via quantum chemistry: COSMO-RS approach', *AIChE Journal* **48**(2), 369-385.

Eckert, F. and Klamt, A. (2006), 'Accurate prediction of basicity in aqueous solution with COSMO-RS', *Journal of computational chemistry* **27**(1), 11-19.

Eckert F., Diedenhofen M., and Klamt A. (2010), 'Towards a first principles prediction of pK_a : COSMO-RS and the cluster-continuum approach', *Molecular Physics* **108**(3-4), 229-241.

Eckert, F. and Klamt, A. (2010). COSMOtherm, Version C2.1. Release 01.11, COSMOlogic GmbH & Co. KG, Leverkusen, Germany.

Eichkorn, K., Weigend, F., Treutler, O. and Ahlrichs, R. (1997), 'Auxiliary basis sets for main row atoms and transition metals and their use to approximate coulomb potentials', *Theoretical Chemistry Accounts* **97**, 119-124.

Errington, J. R. and Panagiotopoulos, A. Z. (1998), 'A fixed point charge model for water optimized to the vapor-liquid coexistence properties', *The Journal of Physical Chemistry B* **102**(38), 7470-7475.

Fock, V. (1930), 'Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems' *Zeitschrift für Physik* **61**(1-2), 126-148.

Fredenslund, A., Jones, R.L., Prausnitz, J.M. (1975), 'Group contribution estimation of activity coefficients in nonideal liquid mixtures', *AIChE J.* **21**, 1086-1099.

Galster, H. (2000), *Technique of measurement, electrode processes and electrode treatment*, Redox. Springer Berlin Heidelberg, 13-23.

Gatellier, P., Hamelin, C., Durand, Y. and Renerre, M. (2001), 'Effect of a dietary vitamin E supplementation on colour stability and lipid oxidation of air-and modified atmosphere-packaged beef', *Meat Science* **59**(2), 133- 140.

Gatellier, P., Kondjoyan, A., Portanguen, S., Grève, E., Yoon, K. and Santé-Lhoutellier, V. (2009), 'Determination of aromatic amino acid content in cooked meat by derivative spectrophotometry: Implications for nutritional quality of meat', *Food Chemistry* **114**(3), 1074-1078.

Geay, Y., Bauchart, D., Hocquette, J.-F. and Culioli, J. (2001), 'Effect of nutritional factors on biochemical, structural and metabolic characteristics of muscles in ruminants, consequences on dietetic value and sensorial qualities of meat', *Reproduction Nutrition Development* **41**(1), 1-26.

Gil-Villegas, A., Galindo, A., Whitehead, P. J., Mills, S. J., Jackson, G. and Burgess, A. N. (1997), 'Statistical associating fluid theory for chain molecules with attractive potentials of variable range', *The Journal of chemical physics* **106**, 4168.

Gmehling, J., Li, J. and Schiller, M. (1993), 'A modified UNIFAC model. 2. present parameter matrix and results for different thermodynamic properties', *Industrial & Engineering Chemistry Research* **32**(1), 178-193.

Gray, J. and Pearson, A. (1987), 'Rancidity and warmed-over flavor', *Advances in meat research* **3**, 221-269.

Green, J. and Paget, M. S. (2004), 'Bacterial redox sensors', *Nature Reviews Microbiology* **2**(12), 954-966.

Grensemann, H. and Gmehling, J. (2005), 'Performance of a conductorlike screening model for real solvents model in comparison to classical group contribution methods', *Industrial & engineering chemistry research* **44**(5), 1610-1624.

Gros, J.-B. and Dussap, C. (2003), 'Estimation of equilibrium properties in formulation or processing of liquid foods', *Food Chemistry* **82**(1), 41-49.

Gross, J. and Sadowski, G. (2002), 'Application of the perturbed-chain saft equation of state to associating systems', *Industrial & engineering chemistry research* **41**(22), 5510-5515.

Guggenheim, E. A. (1966), *Applications of statistical mechanics*, Oxford : Oxford University Press

Guilbaud M, Zagorec M, Chaillou S, Champomier-Vergès MC. (2012), 'Intraspecies diversity of *Lactobacillus sakei* response to oxidative stress and variability of strain performance in mixed strains challenges', *Food Microbiol.* **29**, 2, 197-204

Haar, L., Gallagher, J. S. and Kell, G. S. (1984), *NBS/NRC steam tables: Thermodynamic and transport properties and computer programs for vapor and liquid states of water in SI units, 320 p*, Hemisphere, Washington, DC.

Haigh R. (1986), 'Safety and stressity of antioxidants: EEC approval', *Food Chem. Technol.* **24**, 1031

Hamer, W. J. and Wu, Y.-C. (1972), 'Osmotic coefficients and mean activity coefficients of uni-univalent electrolytes in water at 25 c', *Journal of Physical and Chemical Reference Data* **1**, 1047.

Hammett, L. P. (1937), 'The effect of structure upon the reactions of organic compounds. benzene derivatives', *Journal of the American Chemical Society* **59**(1), 96-103.

Hancock, R. D., Siddons, C. J., Oscarson, K. A. and Reibenspies, J. M. (2004), 'The structure of the 11-coordinate barium complex of the pendant-donor macrocycle 1,4,7,10-tetrakis(carbamoylmethyl)-1,4,7,10-tetraazacyclododecane: an analysis of the coordination numbers of barium(ii) in its complexes', *Inorganica Chimica Acta* **357**(3), 723 - 727.

Hansch, C., Leo, A. and Taft, R. (1991), 'A survey of hammett substituent constants and resonance and field parameters', *Chemical Reviews* **91**(2), 165-195.

Harding A. P., Wedge D. C., and Popelier P. L. A. (2009), 'pK_a Prediction from "Quantum Chemical Topology" Descriptors', *J. Chem. Inf. Model* **49**, 1914–1924

Hartree, D. R. (1928), The wave mechanics of an atom with a non-Coulomb central field. Part I. Theory and methods, *Mathematical Proceedings of the Cambridge Philosophical Society*, **24**(01), 89-110, Cambridge University Press.

Hashimoto, K. and Kamimoto, T. (1998), 'Theoretical study of microscopic solvation of lithium in water clusters: Neutral and cationic li (h2o) n (n= 1-6 and 8)', *Journal of the American Chemical Society* **120**(15), 3560- 3570.

Herzberg, G. (1945), *Molecular Spectra and Molecular Structure: II. Infrared and Raman Spectra of Polyatomic Molecules*, van Nostrand Reinhold: New York.

Hoge, B. and Bader, J. (2007), 'A qualitative scale for the electron withdrawing effect of substituted phenyl groups and heterocycles', *Journal of fluorine chemistry* **128**(7), 857-861.

Hohenberg, P. and Kohn, W. (1964), 'Inhomogeneous electron gas', *Physical review* **136**(3B), B864.

Huang, S. H. and Radosz, M. (1990), 'Equation of state for small, large, polydisperse, and associating molecules', *Industrial & Engineering Chemistry Research* **29**(11), 2284-2294.

Huang, S. H. and Radosz, M. (1991), 'Equation of state for small, large, polydisperse, and associating molecules: extension to fluid mixtures', *Industrial & Engineering Chemistry Research* **30**(8), 1994-2005.

Huber, K, P. and Herzberg, G. (1979), *Molecular Spectra and Molecular Structure: IV. Constants of Diatomic Molecules*, van Nostrand Reinhold: New York.

Ingram, T., Gerlach, T., Mehling, T. and Smirnova, I. (2012), 'Extension of COSMO-RS for monoatomic electrolytes: Modeling of liquid-liquid equilibria in presence of salts', *Fluid Phase Equilibria* **314**, 29-37.

Irikura, K. K. and Frurip, D. J. (1998), *Computational thermochemistry*, in 'ACS symposium series', Vol. 677, ACS Publications, 2-19.

-Israelachvili, J. N. (2011), *Intermolecular and Surface Forces, 3rd edition*, Academic Press, London.

Jensen, I.-J., Dort, J. and Eilertsen, K.-E. (2014), 'Proximate composition, antihypertensive and antioxidative properties of the semimembranosus muscle from pork and beef after cooking and *in vitro* digestion', *Meat science* **96**(2), 916-921.

Joo, S., Kauffman, R., Kim, B. and Park, G. (1999), 'The relationship of sarcoplasmic and myofibrillar protein solubility to colour and water holding capacity in porcine longissimus muscle', *Meat Science* **52**(3), 291- 297.

Jorgensen, W. (1991), 'Computational insights on intermolecular interactions and binding in solution', *Chemtracts Org. Chem* **4**, 91-119.

Karel, M. and Lund, D. B. (2003),*Physical principles of food preservation (Vol. 129),* CRC Press.

Kebarle, P. (1977), 'Ion thermochemistry and solvation from gas phase ion equilibria', *Annual Review of Physical Chemistry* **28**(1), 445-476.

Kiriukhin, M. Y. and Collins, K. D. (2002), 'Dynamic hydration numbers for biologically important ions', *Biophysical Chemistry* **99**(2), 155-168.

Klamt, A. (1995), 'Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena', *The Journal of Physical Chemistry* **99**(7), 2224-2235.

Klamt, A. (2005), *COSMO-RS from quantum chemistry to fluid phase thermodynamics and drug design*, Elsevier.

Klamt, A. (2011), 'The COSMO and COSMO-RS solvation models', *Wiley Interdisciplinary Reviews: Computational Molecular Science* **1**(5), 699-709.

Klamt, A. and Eckert, F. (2000), 'COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids', *Fluid Phase Equilibria* **172**(1), 43-72.

Klamt, A., Eckert, F., Diedenhofen, M. and Beck, M. E. (2003), 'First principles calculations of aqueous pK_a values for organic and inorganic acids using COSMOrs reveal an inconsistency in the slope of the pK_a scale', *The Journal of Physical Chemistry A* **107**(44), 9380-9386.

Klamt, A., Krooshof, G. J. and Taylor, R. (2002), 'COSMOSPACE: Alternative to conventional activity-coefficient models', *AIChE journal* **48**(10), 2332-2349.

Klamt, A., Jonas, V., Bürger, T. and Lohrenz, J. C. W. (1998), 'Refinement and parametrization of COSMO-RS', *The Journal of Physical Chemistry A* **102**(26), 5074-5085.

Klamt, A. and Schüürmann, G. (1993), 'COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient', *J. Chem. Soc., Perkin Trans.* **2**, 799-805.

Kojima, K. and Tochigi, K. (1979), *Prediction of vapor-liquid equilibria by the ASOG method*, Tokyo: Kodansha.

Kontogeorgis, G. M. and Folas, G. K. (2009), *Thermodynamic models for industrial applications: from classical and advanced mixing rules to association theories*, John Wiley & Sons.

Kontogeorgis, G. M., Michelsen, M. L., Folas, G. K., Derawi, S., von Solms, N. and Stenby, E. H. (2006), 'Ten years with the cpa (cubic-plus-association) equation of state. part 1. pure compounds and self-associating systems', *Industrial & engineering chemistry research* **45**(14), 4855-4868.

Larrea V., Pérez-Munuera I., Hernando I., Quiles A., Llorca E., Lluch M. A. (2007), 'Microstructural changes in teruel dry-cured ham during processing', *Meat Science* **76**, 574-582.

Larsen, B. L., Rasmussen, P. and Fredenslund, A. (1987), 'A modified UNIFAC group-contribution model for prediction of phase equilibria and heats of mixing', *Industrial & engineering chemistry research* **26**(11), 2274-2286.

Lawrie R.A. (1985), *Meat Science, 4th ed.*, Oxford: Pergamon Press, 173-175.

Le Maguer, M. (1992), *Thermodynamics and vapor–liquid equilibria*, H.G. Schwartzberg, R.W. Martel (Eds.), Physical chemistry of foods, Marcel Dekker, New York.

Le Page, J. F., Dachraoui, O., Daudin, J. D., and Mirade, P. S. (2012), Time-course estimation of evaporated water fluxes together with mean surface water activity in uncooked pressed cheeses during ripening using a purpose-built micro-bioreactor, *International Dairy Journal*, **22**(1), 66-72.

Le Page, J. F., Mirade, P. S. and Daudin, J. D. (2010), Development of a device and method for the time-course estimation of low water fluxes and mean surface water activity of food products during ripening and storage, *Food research international*, **43**(4), 1180-1186.

Lebert, I. and Lebert, A. (2006), 'Quantitative prediction of microbial behaviour during food processing using an integrated modelling approach: a review', *International Journal of Refrigeration* **29**(6), 968-984.

Lebert, I., Dussap, C. and Lebert, A. (2005), 'Combined physico-chemical and water transfer modelling to predict bacterial growth during food processes', *International journal of food microbiology* **102**(3), 305-322.

Lebert, A. and Richon, D. (1984), 'Infinite dilution activity coefficients of n-alcohols as a function of dextrin concentration in water-dextrin systems', *Journal of Agricultural and Food Chemistry* **32**(5), 1156-1161.

Lee A.C., and Crippen G.M. (2009), 'Predicting pK_a', *J. Chem. Inf. Model* **49**, 2013–2033

Lei, Z., Chen, B., Li, C., & Liu, H. (2008), 'Predictive molecular thermodynamic models for liquid solvents, solid salts, polymers, and ionic liquids', *Chemical reviews* **108**(4), 1419-1455.

Lin, S.-T., Chang, J., Wang, S., Goddard, W. A. and Sandler, S. I. (2004), 'Prediction of vapor pressures and enthalpies of vaporization using a COSMO solvation model', *The Journal of Physical Chemistry A* **108**(36), 7429- 7439.

Lin, S.-T. and Sandler, S. I. (2002), 'A priori phase equilibrium prediction from a segment contribution solvation model', *Industrial & engineering chemistry research* **41**(5), 899-913.

Lindahl, G., Lundström, K. and Tornberg, E. (2001), 'Contribution of pigment content, myoglobin forms and internal reflectance to the colour of pork loin and ham from pure breed pigs', *Meat Science* **59**(2), 141-151.

Magnussen, T., Rasmussen, P. and Fredenslund, A. (1981), 'UNIFAC parameter table for prediction of liquid-liquid equilibriums', *Industrial & Engineering Chemistry Process Design and Development* **20**(2), 331-339.

Manchester J., Walkup G., Rivin O., and You Z. (2010), 'Evaluation of pK_a Estimation Methods on 211 Druglike Compounds', *J. Chem. Inf. Model* **50**, 565–571

Mantina, M., Chamberlin, A. C., Valero, R., Cramer, C. J. and Truhlar, D. G. (2009), 'Consistent van der waals radii for the whole main group', *The Journal of Physical Chemistry A* **113**(19), 5806-5812.

Marcus, Y. (1988), 'Ionic radii in aqueous solutions', *Chemical Reviews* **88**(8), 1475-1498.

Marcus, Y. (1997), *Ion Propierties, Vol. 1*, CRC Press.

Marcus, Y. (2002), *Solvent mixtures: properties and selective solvation*, CRC Press.

Marcus, Y. (2012), *Ions in Water and Biophysical Implications: From Chaos to COSMOs*, Springer.

Martin, F., Cayot, N., Marin, A., Journaux, L., Cayot, P., Gervais, P. and Cachon, R. (2009), 'Effect of oxidoreduction potential and of gas bubbling on rheological properties and microstructure of acid skim milk gels acidified with glucono-δ-lactone', *Journal of dairy science* **92**(12), 5898-5906.

Martin, F., Cayot, N., Vergoignan, C., Journaux, L., Gervais, P. and Cachon, R. (2010), 'Impact of oxidoreduction potential and of gas bubbling on rheological properties of non-fat yoghurt', *Food Research International* **43**(1), 218-223.

Martin F., Ebel B., Gervais P., Cayot N., Cachon R.. (2012), 'Redox potential: monitoring and role in development of aroma compounds, rheological properties and survival of oxygen sensitive strains during the manufacture of fermented dairy products', *Lactic acid bacteria. J.*, Marcelino Kongo, InTech.

Martin, M. G. and Siepmann, J. I. (1998), 'Transferable potentials for phase equilibria. 1. united-atom description of n-alkanes', *The Journal of Physical Chemistry B* **102**(14), 2569-2577.

Mercier Y., Gatellier P., Renerre M. (2004), 'Lipid and protein oxidation in vitro, and antioxidant potential in meat from charolais cows finished on pasture or mixed diet.', *Meat Sci.* **66**, 4

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953), 'Equation of state calculations by fast computing machines', *The journal of chemical physics* **21**, 1087.

Michelon, D., Abraham, S., Ebel, B., De Coninck, J., Husson, F., Feron, G., Gervais, P. and Cachon, R. (2010), 'Contribution of exofacial thiol groups in the reducing activity of lactococcus lactis', *FEBS Journal* **277**(10), 2282-2290.

Michelon, D., Tachon, S., Ebel, B., De Coninck, J., Feron, G., Gervais, P., Yvon, M. and Cachon, R. (2012), 'Screening of lactic acid bacteria for reducing power using a tetrazolium salt reduction method on milk agar', *Journal of bioscience and bioengineering* .

Moss, G. P. (1996), 'Basic terminology of stereochemistry (IUPAC Recommendations 1996)', *Pure and applied chemistry* **68**(12), 2193-2222.

Nakajoh, K., Grabda, M., Oleszek-Kudlak, S., Shibata, E., Eckert, F. and Nakamura, T. (2009), 'Prediction of vapour pressures of chlorobenzenes and selected polychlorinated biphenyls using the COSMO-RS model', *Journal of Molecular Structure: THEOCHEM* **895**(1-3), 9-17.

Nath, S. K., Escobedo, F. A. and de Pablo, J. J. (1998), 'On the simulation of vapor-liquid equilibria for alkanes', *The Journal of chemical physics* **108**, 9905.

Nielsen, S. B., Masella, M. and Kebarle, P. (1999), 'Competitive gas-phase solvation of alkali metal ions by water and methanol', *The Journal of Physical Chemistry A* **103**(48), 9891-9898.

Offer, G. and Knight, P. (1988), 'Structural basis of water-holding in meat. 2. Drip losses', *Developments in meat science*.

Offer, G. and Trinick, J. (1983), 'On the mechanism of water holding in meat: the swelling and shrinking of myofibrils', *Meat Science* **8**, 245-281.

Ohtaki, H. and Radnai, T. (1993), 'Structure and dynamics of hydrated ions', *Chemical Reviews* **93**(3), 1157-1204.

Ould Moulaye, C. B. (1998), Calcul des propriétés de formation en solution aqueuse des composés impliqués dans les procédés microbiologiques et alimentaires Prédiction et réconciliation de données Modélisation des équilibres chimiques et des équilibres entre phases, *PhD thesis*, Université Blaise Pascal-Clermont-Ferrand II.

Parsons, N. and Knight, P. (1990), Origin of variable extraction of myosin from myofibrils treated with salt and pyrophosphate, *Journal of the Science of Food and Agriculture*, **51**(1), 71-90.

Partanen, J. A. and Minkkinen, P. O. (1993), 'Thermodynamic activity quantities in aqueous sodium and potassium chloride solutions at 298.15 K up to a molality of 2.0 mol kg^{-1}.', *Acta Chemica Scandinavica* **47**, 768-776.

Pedersen, M. B., Gaudu, P., Lechardeur, D., Petit, M.-A. and Gruss, A. (2012), 'Aerobic respiration metabolism in lactic acid bacteria and uses in biotechnology', *Annual Review of Food Science and Technology* **3**, 37-58.

Perrin D.D. (1965), *Dissociation constants of organic bases in aqueous solution*, International Union of Pure and Applied Chemistry, BUTTERWORTHS, London, England

Perrin D.D. (1972), *Dissociation constants of organic bases in aqueous solution, supplement 1972*, International Union of Pure and Applied Chemistry, BUTTERWORTHS, London, England

Perrin, D. D., Dempsey, B. and Serjeant, E. P. (1981), *pK_a prediction for organic acids and bases, Vol. 1*, Chapman and Hall London; New York.

Pitzer, K. S. (1973), 'Thermodynamics of electrolytes. i. theoretical basis and general equations', *The Journal of Physical Chemistry* **77**(2), 268- 277.

Pitzer, K. S. (1980), 'Electrolytes. from dilute solutions to fused salts', *Journal of the American Chemical Society* **102**(9), 2902-2906.

Planchon, S., Chambon, C., Desvaux, M., Chafsey, I., Leroy, S., Talon, R. and Hebraud, M. (2007), 'Proteomic analysis of cell envelope from staphylococcus xylosus c2a, a coagulase-negative staphylococcus', *Journal of proteome research* **6**(9), 3566-3580.

Pratt, L. R. and Rempe, S. B. (1999), 'Quasi-chemical theory and implicit solvent models for simulations', *AIP Conference Proceedings* **492**(1), 172- 201.

Prausnitz, J., Lichtenthaler, R. and de Azevedo, E. (1999), *Molecular thermodynamics of fluid-phase equilibria, Prentice-Hall international series in the physical and chemical engineering sciences*, Prentice-Hall PTR.

Pye, C. C., Rudolph, W. and Poirier, R. A. (1996), 'An *ab initio* investigation of lithium ion hydration', *The Journal of Physical Chemistry* **100**(2), 601- 605.

Rahman, M. S. (2007), *Handbook of food preservation*, CRC press.

Reid, R, C., Prausnitz, J, M. and Sherwood, T. (1977), *Properties of Gases and Liquids, 3rd ed.,* McGraw-Hill, New York.

Rempe, S. B., Pratt, L. R., Hummer, G., Kress, J. D., Martin, R. L. and Redondo, A. (2000), 'The hydration number of Li^+ in liquid water', *Journal of the American Chemical Society* **122**(5), 966-967.

Renerre, M., Dumont, F. and Gatellier, P. (1996), 'Antioxidant enzyme activities in beef in relation to oxidation of lipid and myoglobin', *Meat Science* **43**(2), 111-121.

Renerre, M., Poncet, K., Mercier, Y., Gatellier, P. and Métro, B. (1999), 'Influence of dietary fat and vitamin e on antioxidant status of muscles of turkey', *Journal of Agricultural and Food Chemistry* **47**(1), 237-244.

Renerre, M. (1990), 'Factors involved in the discoloration of beef meat', *International Journal of Food Science & Technology* **25**(6), 613-630.

Rice, B. M., Pai, S. V. and Hare, J. (1999), 'Predicting heats of formation of energetic materials using quantum mechanical calculations', *Combustion and Flame* **118**(3), 445-458.

Robinson, R. A. and Stokes, R. H. (1959), *Electrolyte solutions*, Butter-worths, London.

Rosso, L., Zuber, E., Pichat, C. and Flandrois, J. P. (1997), Simple relationship between acid dissociation constant and minimal pH for microbial growth in laboratory medium, *International journal of food microbiology*, **35**(1), 75-81.

Rowlinson, J. S. and Swinton, F. L. (1982), *Liquids and liquid mixtures*, Butter worths, London.

Sandler, S. I. (1994), *Models for thermodynamic and phase equilibria calculations*, Marcel Dekker, Inc., New-York.

Sancho, M. F., Rao, M. A. and Downing, D. L. (1997), 'Infinite dilution activity coefficients of apple juice aroma compounds', *Journal of food engineering* **34**(2), 145-158.

Scott, A. P., Platz, M. S. and Radom, L. (2001), 'Singlet-triplet splittings and barriers to wolff rearrangement for carbonyl carbenes', *Journal of the American Chemical Society* **123**(25), 6069-6076.

Sengers, J. V. and Watson, J. T. R. (1986), 'Improved international formulations for the viscosity and thermal conductivity of water substance', *J. Phys. Chem. Ref. Data* **15**, 1291.

Simatos, D. (2002), 'Propriétés de l'eau dans les produits alimentaires: activité de l'eau, diagrammes de phase et d'états. L'eau dans les aliments', Paris: Lavoisier, 49-83.

Simulis Thermodynamics, (http://www.prosim.net/en/index.php).

Steffen, C., Thomas, K., Huniar, U., Hellweg, A., Rubner, O. and Schroer, A. (2010), 'Tmolex, a graphical user interface for turbomole', *Journal of Computational Chemistry* **31**(16), 2967-2970.

Sutton, L. E. (1965), *Tables of Interatomic Distances and Configuration in Molecules and Ions: Supplement 1956-59, number 18*, Chemical Society.

Szegezdi, J. and Csizmadia, F. (2004), 'Prediction of dissociation constant using microconstants', *27th ACS National Meeting*, Anaheim, California · March 28-April 1, 2004

Szegezdi, J. and Csizmadia, F. (2007), 'Calculating pK_a values of small and large molecules', *American Chemical Society Spring meeting*, March 25-29th, 2007

Tabilo G., Flores M., Fiszman S. M., Toldra F. (1999), 'Postmortem meat quality and sex affect textural properties and protein breakdown od dry cured ham', *Meat Science* **69**, 451-457.

Tachon, S., Michelon, D., Chambellon, E., Cantonnet, M., Mezange, C., Henno, L., Cachon, R. and Yvon, M. (2009), 'Experimental conditions affect the site of tetrazolium violet reduction in the electron transport chain of lactococcus lactis', *Microbiology* **155**(9), 2941-2948.

Talon, R. and Leroy, S. (2011), 'Diversity and safety hazards of bacteria involved in meat fermentations', *Meat science* **89**(3), 303-309.

Terada K. and Breene W.M. (1989), 'Migration of BHT through airspace in food package system', *Packag. Technol. Sci.* **2**, 165

Toldra F. (1998), 'Proteolysis and lipolysis in flavour development of dry-cured meat products', *Meat Science* **49**, 101-110.

Tomasi, J. and Persico, M. (1994), 'Molecular interactions in solution: an overview of methods based on continuous distributions of the solvent', *Chemical Reviews* **94**(7), 2027-2094.

Tomasi, J., Mennucci, B. and Cammi, R. (2005), 'Quantum mechanical continuum solvation models', *Chemical Reviews* **105**(8), 2999-3094.

Topcu, A., McKinnon, I. and McSweeney, P. (2008), 'Measurement of the oxidation-reduction potential of cheddar cheese', *Journal of food science* **73**(3), C198-C203.

Toure, O., Dussap, C.-G. and Lebert, A. (2013), 'Comparison of predicted pK_a values for some amino-acids, dipeptides and tripeptides, using COSMO-RS, ChemAxon and ACD/LABS methods', *Oil Gas Sci. Technol. -Rev. IFP Energies nouvelles* **68**(2), 281-297.

Trebla, M. (2012), Mesure du potential d'oxydoréduction d'antioxydants en solution aqueuse par voie électrochimique, *Master of science thesis*, Université Blaise Pascal-Clermont-Ferrand II.

Tunell, I. and Lim, C. (2006), 'Factors governing the metal coordination number in isolated group IA and IIA metal hydrates', *Inorganic Chemistry* **45**(12), 4811-4819.

TURBOMOLE (2007). TURBOMOLE, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com.

Van den Berg, C. and Bruin, S. (1981), 'Water activity and its estimation in food systems: theoretical aspects, Rockland LB, Stewart GF, Water Activity: Influences on Food Quality, 1-61', Academic Press, New York.

Van Gunsteren, W. F. and Berendsen, H. J. (1990), 'Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry', *Angewandte Chemie International Edition in English* **29**(9), 992-1023.

Varma, S. and Rempe, S. B. (2006), 'Coordination numbers of alkali metal ions in aqueous solutions', *Biophysical Chemistry* **124**(3), 192-199.

Vega, L. F., and Jackson, G. (2011), '20 Years of the SAFT equation of state - Recent advances and challenges: Symposium held in Bellaterra, Barcelona, 19–21 September 2010', *Fluid Phase Equilibria* **306**(1), 1-3.

Weidlich, U. and Gmehling, J. (1987), 'A modified UNIFAC model. 1. prediction of VLE, H^E, and. gamma.. infin.', *Industrial & engineering chemistry research* **26**(7), 1372-1381.

White, S. H. and Wimley, W. C. (1999), 'Membrane protein folding and stability: physical principles', *Annual review of biophysics and biomolecular structure* **28**(1), 319-365.

Whitfield, T. W., Varma, S., Harder, E., Lamoureux, G., Rempe, S. B. and Roux, B. (2007), 'Theoretical study of aqueous solvation of K^+ comparing *ab initio*, polarizable, and fixed-charge models', *Journal of Chemical Theory and Computation* **3**(6), 2068-2082.

Wolbach, J. P. and Sandler, S. I. (1997), 'Using molecular orbital calculations to describe the phase behavior of hydrogen-bonding fluids', *Industrial & engineering chemistry research* **36**(10), 4041-4051.

Wu, H. S. and Sandler, S. I. (1991), 'Use of ab initio quantum mechanics calculations in group contribution methods. 1. Theory and the basis for group identifications', *Industrial & engineering chemistry research* **30**(5), 881-889.

Zavitsas, A. A. (2001), 'Properties of water solutions of electrolytes and nonelectrolytes', *The Journal of Physical Chemistry B* **105**(32), 7805-7817.

Zemaitis, J. F., Clark, D. M., Rafal, M. and Scrivner, N. C. (1986), *Handbook of aqueous electrolyte thermodynamics*, AIChE, New York: Deisgn Institute for Physical Property Data,.Zoski, C. G. (2007), *Handbook of electrochemistry*, Access Online via Elsevier.

Résumé: Les milieux biologiques et alimentaires sont généralement des mélanges contenant un nombre élevé de constituants (eau, solvants organiques, solides dissous, gaz dissous, espèces ioniques, macromolécules). La prédiction des propriétés d'équilibre de tels milieux requiert l'utilisation d'un modèle thermodynamique entièrement prédictif. Ce modèle doit également permettre d'assurer la cohérence entre des données expérimentales et garantir la robustesse de la représentation simultanée des équilibres physiques (liquide-vapeur, solubilité, etc.) et chimiques (dissociation, oxydo-réduction, complexation, etc.). Le potentiel chimique est une donnée indispensable pour modéliser ces équilibres. Sa connaissance dépend de la prédiction de deux variables : l'enthalpie libre de formation dans un état de référence choisi, et le coefficient d'activité qui dépend aussi de l'état de référence choisi.

Le modèle COSMO-RS est un excellent modèle de prédiction des coefficients d'activité très largement utilisé dans le domaine du génie chimique où on s'intéresse essentiellement à des molécules neutres. Ce travail de thèse a permis d'étendre les performances du modèle COSMO-RS au traitement des milieux biologiques et alimentaires dans lesquels on trouve systématiquement des électrolytes en solution (en plus des molécules neutres). Un nouvel outil utilisant les récentes avancées de la mécanique quantique a été développé pour prédire les propriétés de formation à l'état gaz. En combinant des concepts de la thermodynamique, de la mécanique quantique, de l'électrostatique, et de la physique statistique, il a été démontré qu'il est possible d'utiliser le modèle COSMO-RS pour faire la transition entre l'état gaz et la phase condensée. Partant de là, ce travail démontre qu'il est maintenant possible de traiter simultanément les équilibres physiques et chimiques et donc de prédire les propriétés physico-chimiques (a_W, pH, E_h) dans les milieux biologiques et alimentaires par le modèle COSMO-RS.

Mots-clés: Propriétés d'équilibre, COSMO-RS, modèle thermodynamique, prediction, hydratation.

Abstract: Food and biological systems are generally multicomponent mixtures (including water, organic solvents, dissolved solids, dissolved gases, ionic species, macromolecules). The prediction of the equilibrium properties of such environments requires the use of a fully predictive thermodynamic model. This model must be able to ensure the consistency between experimental data and to ensure the robustness of the simultaneous representation of physical equilibria (liquid-vapour, solubility, etc.) and chemical equilibria (dissociation, redox, complexation, etc.). The chemical potential is an essential property for modelling such equilibria. Its determination depends on two variables: the Gibbs free energy of formation in a chosen reference state, and the prediction of the activity coefficient which also depends on the chosen reference state.

The COSMO-RS model is an excellent model for predicting activity coefficients that is widely used in chemical engineering where the studied molecules are generally neutral. This PhD study enabled to extend the performance of the COSMO-RS model toward the treatment of food and biological systems where there are systematically electrolytes in solution (in addition to neutral molecules). A new tool based on the recent advances of quantum mechanics has been developed in order to predict gas phase formation properties. By combining concepts of thermodynamics, quantum physics, electrostatics and statistical physics, it has been demonstrated that it is possible to use the COSMO-RS model to ensure the transition between the gas phase and the condensed phase. In this context, this work demonstrates that it is possible to treat simultaneously physical and chemical equilibria and thus to predict physico-chemical properties (a_W, pH, E_h) in food and biological systems using the COSMO-RS model.

Key words: Equilibrium properties, COSMO-RS, thermodynamic model, prediction, hydration.