



HAL
open science

Estimation of Solvation Quantities from Experimental Thermodynamic Data: Development of the Comprehensive CompSol Databank for Pure and Mixed Solutes

Edouard Moine, Romain Privat, Baptiste Sirjean, Jean-Noël Jaubert

► To cite this version:

Edouard Moine, Romain Privat, Baptiste Sirjean, Jean-Noël Jaubert. Estimation of Solvation Quantities from Experimental Thermodynamic Data: Development of the Comprehensive CompSol Databank for Pure and Mixed Solutes. *Journal of Physical and Chemical Reference Data*, 2017, 46, pp.033102. 10.1063/1.5000910 . hal-01703358

HAL Id: hal-01703358

<https://hal.univ-lorraine.fr/hal-01703358>

Submitted on 23 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Estimation of Solvation Quantities from Experimental Thermodynamic Data: Development of the Comprehensive CompSol Databank for Pure and Mixed Solutes

Edouard Moine, Romain Privat,^{a)} Baptiste Sirjean, and Jean-Noël Jaubert

Laboratoire Réactions et Génie des Procédés (UMR CNRS 7274), Ecole Nationale Supérieure des Industries Chimiques, Université de Lorraine, 1 Rue Grandville, 54000 Nancy, France

(Received 9 May 2017; accepted 18 August 2017; published online 25 September 2017)

The Gibbs energy of solvation measures the affinity of a solute for its solvent and is thus a key property for the selection of an appropriate solvent for a chemical synthesis or a separation process. More fundamentally, Gibbs energies of solvation are choice data for developing and benchmarking molecular models predicting solvation effects. The Comprehensive Solvation—CompSol—database was developed with the ambition to propose very large sets of new experimental solvation chemical-potential, solvation entropy, and solvation enthalpy data of pure and mixed components, covering extended temperature ranges. For mixed compounds, the solvation quantities were generated in infinite-dilution conditions by combining experimental values of pure-component and binary-mixture thermodynamic properties. Three types of binary-mixture properties were considered: partition coefficients, activity coefficients at infinite dilution, and Henry's-law constants. A rigorous methodology was implemented with the aim to select data at appropriate conditions of temperature, pressure, and concentration for the estimation of solvation data. Finally, our comprehensive CompSol database contains 21 671 data associated with 1969 pure species and 70 062 data associated with 14 102 binary mixtures (including 760 solvation data related to the ionic-liquid class of solvents). On the basis of the very large amount of experimental data contained in the CompSol database, it is finally discussed how solvation energies are influenced by hydrogen-bonding association effects. © 2017 AIP Publishing LLC for the National Institute of Standards and Technology. <https://doi.org/10.1063/1.5000910>

Key words: activity coefficient; chemical potential of solvation; databank; database; Henry's law; infinite dilution; solvation data; solvation process; thermodynamics.

CONTENTS

		solubility data) measured in a composition range where Henry's law can be used as a sound approximation of solute behavior	9
1.	Introduction		3
2.	Expressions of Solvation Energies for Their Practical Estimation from Experimental Data . .		6
	2.1. Expression of the Gibbs energy of solvation derived from Ben-Naim's definition		6
	2.2. General expressions of chemical potentials of solvation		7
	2.3. Practical estimation of chemical potentials of solvation for pure components		8
	2.4. Expressions of chemical potentials of solvation of infinitely dilute solutes		8
	2.4.1. Estimation of $\Delta_{\text{solv}}\bar{g}_i^\infty(T, P)$ from Henry's-law constant measurements		8
	2.4.2. Estimation of $\Delta_{\text{solv}}\bar{g}_i^\infty(T, P)$ from partition coefficients (i.e., VLE		
		solubility data) measured in a composition range where Henry's law can be used as a sound approximation of solute behavior	9
		2.4.3. Estimation of $\Delta_{\text{solv}}\bar{g}_i^\infty(T, P)$ from Raoult's-law infinite-dilution activity-coefficient measurements	9
		2.5. Estimation of partial molar entropy and enthalpy of solvation	10
		2.6. Summary of the expressions relating the chemical potential of solvation to experimentally accessible properties	10
3.	Methodology Used for Developing the CompSol Database		11
	3.1. Generation of pure-component solvation chemical potential data		11
	3.2. Generation of binary solvation chemical potential data at infinite dilution		12
	3.2.1 Estimation of $\Delta_{\text{solv}}\bar{g}_i^\infty(T, P)$ from complete VLE measurements		12
	3.2.2. Estimation of $\Delta_{\text{solv}}\bar{g}_i^\infty(T, P)$ from Raoult's-law infinite-dilution activity-coefficient measurements		13

^{a)}Author to whom correspondence should be addressed; electronic mail: romain.privat@univ-lorraine.fr.
 © 2017 AIP Publishing LLC.

3.2.3. Estimation of $\Delta_{\text{solv}}\bar{g}_i^\infty(T, P)$ from Henry's-law constant measurements	14
3.2.4. Methodology for systems using an ionic liquid as solvent	14
4. Global Analysis of the Obtained Results: Presentation of the CompSol Database	15
4.1. On the behavior of pure-component solvation properties	15
4.1.1. Description of the CompSol pure-component database	15
4.1.2. Mutual agreement between the CompSol and MNSol databases	15
4.1.3. What can be learnt from self-solvation data?	15
4.2. On the behavior of binary-system solvation properties	16
4.2.1. Description of the CompSol binary-system database	16
4.2.2. Data consistency	16
4.2.3. Mutual agreement between the CompSol and MNSol databases	16
4.2.4. Focus on the generation of infinite-dilution solvation enthalpy and entropy data	17
4.2.5. Effect of hydrogen bonding on binary-system solvation properties	17
5. Conclusion	19
Acknowledgments	20
6. Appendix A: On the Pressure Independence of the Pseudo-chemical Potential of a Molecule in a Perfect-gas Mixture	20
7. Appendix B: On the Pressure Independence of the Solvation Chemical Potential	21
8. References	21

List of Tables

1. List of expressions and related assumptions potentially usable for the estimation of the chemical potential of solvation (CPS) of species i	11
------------------------------------------------------------------------------------------------------------------------------------------------------------	----

2. Description of the specifications and assumptions mentioned in Table 1	11
3. Description of the four association codes for pure components	18

List of Figures

1. Occurrences of the keywords "solvation energy" and "solvation process" in published scientific journals (bar chart) Number of citations of Ben-Naim's book <i>Solvation Thermodynamics</i> (line).	4
2. Illustration of Ben-Naim's solvation process	7
3. Summary of the various assumptions used to derive Eq. (21) in the Pxy isothermal VLE phase diagram of an arbitrary binary system	9
4. Saturation fugacity coefficients of pure components versus reduced pressure	13
5. Parity plot comparing $\Delta_{\text{solv}}g_{\text{pure } i}^{\text{sat}}$ data at 298.15 K contained in the MNSol and CompSol databases, on the basis of 72 pure components.	15
6. (a) $\Delta_{\text{solv}}g_{\text{pure } i}^{\text{sat}}$ vs. T and (b) $(\Delta_{\text{solv}}g_{\text{pure } i}^{\text{sat}})_r$ vs. $T_{r,i}$ for different linear alkanes, benzene, cyclohexane, and water	15
7. (a) $(\Delta_{\text{solv}}g_{\text{pure } i}^{\text{sat}})_r = \Delta_{\text{solv}}g_{\text{pure } i}^{\text{sat}}/R$ and (b) $(\Delta_{\text{solv}}h_{\text{pure } i}^{\text{sat}})_r = \Delta_{\text{solv}}h_{\text{pure } i}^{\text{sat}}/(RT_{c,i})$ vs. $T_{r,i} = T/T_{c,i}$ for different linear alkanes, benzene, cyclohexane, and water	16
8. Parity plot comparing $\Delta_{\text{solv}}\bar{g}_i^\infty$ data at 298.15 K contained in the MNSol and CompSol databases, based on 601 binary systems	17
9. ST = solute; SV = solvent. (a) Examples of binary systems showing a linear $\Delta_{\text{solv}}\bar{g}_i^\infty(T)$ function. (b) Example of binary systems such that it is not possible to conclude about the linearity of the $\Delta_{\text{solv}}\bar{g}_i^\infty(T)$ function.	17
10. Classification of binary systems into 16 categories deduced from pure-component association codes presented in Table 3	18
11. Distribution of $\Delta_{\text{solv}}\bar{g}_i^\infty$ values at 298.15 K following the seven classes of binary systems defined in Fig. 10	19

NOMENCLATURE

List of symbols

B_i	Second virial coefficient of component i , $\text{m}^3 \text{mol}^{-1}$
c_i	Molar concentration of component i , mol m^{-3}
f_i	Fugacity of component i , Pa
\bar{g}_i	Chemical potential of component i (also called <i>partial molar Gibbs energy</i> of i), J mol^{-1}

\bar{g}_i^0	Pseudo chemical potential of component i , J mol^{-1}
h	Planck constant, $\text{m}^2 \text{kg s}^{-1}$
$H_{i,j}$	Henry's-law constant of solute i in solvent j , Pa
k	Boltzmann constant, J K^{-1}
K_i	Partition coefficient of component i , dimensionless
m_i	Molecular mass of component i , kg
\mathbf{n}	Vector containing the mole numbers n_i of all components i
n_i	Mole number of component i , mol
N_A	Avogadro constant, mol^{-1}

P	Pressure, Pa
R	Gas constant, $\text{J mol}^{-1} \text{K}^{-1}$
T	Temperature, K
V	Volume, m^3
v	Molar volume, $\text{m}^3 \text{mol}^{-1}$
\mathbf{x}	Vector containing the molar fractions x_i of all components i of a liquid system
x_i	Molar fraction of component i in a liquid system, dimensionless
\mathbf{y}	Vector containing the molar fractions y_i of all components i of a gaseous system
y_i	Molar fraction of component i in a gaseous system, dimensionless
z	Molar compressibility factor, dimensionless

Greek letters

γ_i	Activity coefficient of component i in a liquid system, dimensionless
Λ	Thermal de Broglie wavelength, m
$\Delta_{\text{solv}}\bar{g}_i$	Chemical potential of solvation of mixed i , J mol^{-1} (also called <i>partial molar Gibbs energy of solvation</i> or <i>partial molar solvation Gibbs free energy</i>)
$\Delta_{\text{solv}}g_{\text{pure } i}$	Chemical potential of solvation of pure i , J mol^{-1} (also called <i>pure-component molar Gibbs energy of solvation</i> or <i>pure-component molar solvation Gibbs free energy</i>)
$\Delta_{\text{solv}}\bar{h}_i$	Partial molar enthalpy of solvation of mixed i , J mol^{-1}
$\Delta_{\text{solv}}h_{\text{pure } i}$	Molar enthalpy of solvation of pure i , J mol^{-1}
$\Delta_{\text{solv}}\bar{s}_i$	Partial molar entropy of solvation of mixed i , $\text{J mol}^{-1} \text{K}^{-1}$
$\Delta_{\text{solv}}s_{\text{pure } i}$	Molar entropy of solvation of pure i , $\text{J mol}^{-1} \text{K}^{-1}$
ρ	Molar density, mol m^{-3}
φ_i	Fugacity coefficient of i , dimensionless

Subscripts

c	Critical property
gas	Gas phase
i, j, k	Component (i : solute, j : solvent, k : arbitrary component)
liq	Liquid phase
pure	Pure-component property
r	Reduced property (pure-component property divided by critical property)

Superscripts

●	Perfect-gas property
(x)	For activity coefficients defined in the molar-fraction scale
∞	Property at infinite dilution
sat	Property at saturation

Abbreviations

DDB	Dortmund databank
DIPPR	Design Institute for Physical Properties databank
EoS	Equation of state
HWR	Hougen-Watson-Ragatz (law of corresponding states)
PCP	Pseudo-chemical potential
VLE	Vapor-liquid equilibrium conditions

1. Introduction

The term *solvation* commonly defines a physico-chemical process in which a small quantity of a solid or gaseous substance called *solute* is spread out in large quantities of a liquid compound or mixture called *solvent*. Similar general definitions of *solvation* can be found in the mainstream scientific literature with more or less emphasis on the nature of the physical interactions involved in the solvation process; for instance, while *Encyclopædia Britannica* solely mentions that the solvent and solute molecules “link together with weak bonds,”¹ the International Union of Pure and Applied Chemistry (IUPAC) provides more details: “Such interactions generally involve electrostatic forces and van der Waals forces, as well as chemically more specific effects such as hydrogen bond formation.”² The quantification of solvation effects is also a recurrent issue for scientists, which has given birth to a variety of roughly similar definitions. For example, for IUPAC, the so-called *solvation energy* is defined as “the change in Gibbs energy when an ion or molecule is transferred from a vacuum (or the gas phase) to a solvent.”³ In the late 1970s, Ben-Naim asserts that a measure of solvation energy⁴

- (i) is “presumed to describe the ‘free energy of interaction with the solvent’ [and] is expected to decrease to zero when the solvent density tends to zero;”
- (ii) should be “devoid of the liberation free energy” (i.e., “the free energy change due to the release of the new particle from a fixed position;” this term accounts for the contribution to the free energy due to translational freedom).

This leads him to conclude that the “conventional” measures of solvation quantity are not the most appropriate ones and “cannot serve as *bona fide* measures of the free energy of interaction with solvent.”⁴ In addition, Ben-Naim observes that the multiplicity of solvation-process definitions does not induce a univocal way of quantifying the energies governing the process and questions the classical assumption that the solvation phenomenon implies binary or pseudo-binary systems in which it is possible to identify a solvent (that is, a molecule present in large amounts) and a solute (that is, a molecule present in small amounts) and excludes the case of binary solutions of arbitrary compositions.⁵

Starting from the facts that “thermodynamics alone cannot discriminate between the relative significance of the various standard quantities”⁴ and that the solvation process is “inherently a molecular process, dependent upon local rather than macroscopic properties,”⁵ he introduces new definitions based

on statistical mechanical considerations, which in his opinion “deserve to be referred to as *bona fide* measures of solvation.”⁵ “We now define the solvation process of a molecule s in a fluid L ” (which can be a pure phase or a mixture and which may possibly contain a large amount of s) “as the process of transferring the molecule s from a fixed position in an ideal gas phase g into a fixed position in the fluid or liquid phase L . The process is carried out at constant temperature T and pressure P . Also, the composition of the system is unchanged.”⁵ It is here meant that the ideal-gas phase g and the fluid or liquid phase L have the same temperature and pressure and that the composition of the liquid phase is not affected by the transfer. Note that the composition of the ideal-gas phase is not fixed by the definition for the simple reason that solvation quantities (e.g., Gibbs energy, enthalpy, and entropy of solvation) are independent of the ideal-gas composition, as discussed below and proved in Appendix A.

Among other advantages, his new definition makes it possible to “describe the thermodynamics of solvation without reference to any standard state. This is a considerable improvement relative to the present situation in the field.”⁵ Following this definition, the Gibbs energy of solvation can be pragmatically seen as the Gibbs energy change that would be measured in an isothermal and isobaric device in which one molecule of fixed s (i.e., without kinetic energy) would be transferred from a perfect-gas mixture to a liquid L . As a straightforward consequence, a negative molar Gibbs energy of solvation indicates that “with respect to the adopted standard states, the (liquid) solution state is thermodynamically more stable than the gas phase.”⁶

The rate of adherence of the scientific community to these definitions can be evaluated through the growing occurrence of Ben-Naim’s book citations in peer-reviewed publications during the past four decades, as shown in Fig. 1. In this study, the definition of the solvation process proposed by Ben-Naim is thus definitively adopted. Figure 1 also illustrates the growing interest of scientists for solvation thermodynamics

issues, as revealed by the occurrences of the keywords “solvation energy” and “solvation process” in the same period.

The usefulness of these concepts can be simply appreciated through the variety of practical and theoretical applications referring to solvation quantities.

- (i) Since these energies basically describe and quantify interactions, mainly size and energetic effects, resulting from the insertion of a static molecule (solute) in a “medium” (solvent), they can be used to measure and compare affinities of solutes for solvents. In this respect, different methodologies in the same spirit have been proposed: for a given solvent, a general chart representing the solvation energies of reference solutes as a function of a solute-specific or mixture property is built in order to highlight reference behaviors. Once this general chart is available, points specific to new solutes are placed on the chart so that their affinities can be quantified through the distance between points and reference-solute behaviors.

For illustration, some authors have proposed to use as a general chart the *molar Gibbs energy of solvation* versus the *van der Waals surface area of the solute*. Doing so, for a given solvent, linear trends are observed in a homologous series of solutes.^{6–8} Among others, this tool has been used to study how the branching of alkanes modifies their affinity for a given solvent.

As another illustration, working on solvophobic effects (i.e., low affinity of a solute for a solvent), Sedov *et al.* have used as a general chart the plane *solvation Gibbs energy* versus *solvation enthalpy*. They showed that points related to hydrocarbons in non-associating solvents are all located on the same universal straight line.^{9–12} Considering this straight line as reference behavior for systems made up of solutes showing high affinity for solvents, they proposed a way to quantify the solvophobic effect through

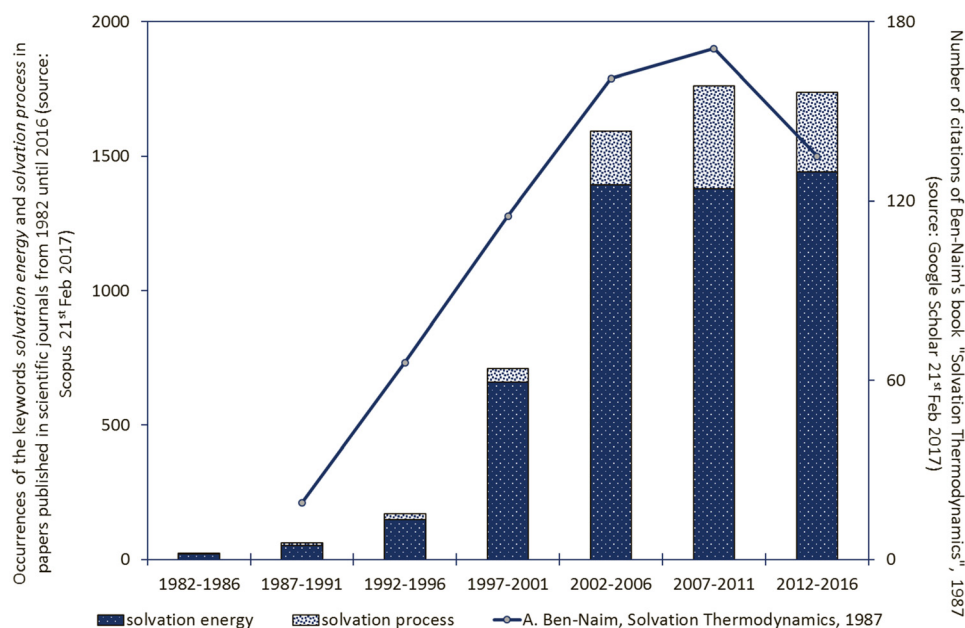


FIG. 1. Occurrences of the keywords “solvation energy” and “solvation process” in published scientific journals (bar chart). Number of citations of Ben-Naim’s book *Solvation Thermodynamics* (line).

the deviation between points related to a given (solute + solvent) pair and the reference straight line.

Such methods make it possible to simply quantify the degree of affinity between two compounds and can be obviously used to select an appropriate solvent for a given application.

- (ii) Solvation quantities are also extensively considered in biological applications where solvent effects play a key role. In pharmacology, the properties of active molecules are affected by their matrix/solvent/environment, and therefore special attention needs to be paid to the solvation process. For instance, remembering that the conformations of proteins depend upon their solvent and determine their biological function, solvation quantities appear as key properties for the study of protein-structure stability and fluctuation in solution.^{13–18}

Solvation properties are also of particular importance for “drug design” applications, the aim of which is to produce a molecule capable of binding to a target molecule.^{14,18,19}

A particular drug is indeed expected “to have a favorable interaction with both water (for transport in the blood) and lipophilic media (for transport through cell membranes)” (chapter 4 of Ref. 20). In the same way, solvation quantities can be used to study the interaction of nanoparticles (e.g., fullerenes) with biological systems. Functionalized fullerenes are considered as potential HIV protease inhibitors or immunosensors for the detection of cancers and viruses. Their toxicity may be linked to their partitioning between aqueous and organic phases. In this respect, Redmill *et al.* have computed values of molar Gibbs solvation energy of fullerene derivatives in octanol and water from molecular simulation with the aim of providing “molecular insight into the solubility and partitioning of fullerenes.”²¹

Note that an interesting review highlighting—among others—the usefulness of solvation energies for drug-design and environmental applications has been recently published.¹⁹

- (iii) Solvation quantities were also used for the selection of appropriate stationary phases for gas-chromatography applications. To do so, a *solvation parameter model* was used to predict the distribution of a solute between a gas phase and various potential stationary phases.^{22–24}

More anecdotally, some authors have recently reworked macroscopic thermodynamic models in order to use solvation energies as input parameters. As a first illustration, Pliego has shown how Gibbs energies of solvation can be incorporated into the Scatchard-Hildebrand activity-coefficient model to predict liquid-liquid phase equilibria.²⁵ As another illustration, Hsieh and Lin have introduced Gibbs energies of solvation in the attractive-parameter expression of the Peng-Robinson cubic equation of state (EoS).^{26–28} The COSMO-SAC solvation model was used to generate Gibbs energies of solvation.

To conclude the list of potential applications related to solvation quantities, we mention the generation of detailed kinetic models (involving potentially thousands of species and reactions) for liquid-phase reactions.^{29,30} To estimate rate constants of reactions taking place in liquid phases, a theory

allowing to switch from gas-phase-reaction rates to liquid-phase-reaction rates is considered. The implementation of this theory requires the preliminary knowledge of Gibbs energies of solvation, which makes it possible to build a bridge between liquid-phase and gas-phase properties.

Considering that each of the applications mentioned above requires the preliminary knowledge of Gibbs energies of solvation, it becomes necessary to dispose of accurate estimation methods such as experimental determinations [which, as explained below, are all based on vapor-liquid equilibrium (VLE) property measurements] or model predictions.

The estimation of molar Gibbs energies of solvation of a given solute requires the combined knowledge of the solvent density and the fugacity of the solute in the liquid phase. Under certain conditions, this latter quantity can be estimated from partition-coefficient data (accessible from direct vapor-liquid equilibrium measurements) or from alternative sets of properties such as {pure-solute and pure-solvent vapor pressures + pure-solute density + pure-solute fugacity coefficient at saturation + infinite-dilution activity coefficients} or Henry’s-law constants. A large number of papers detail experimental determinations of Gibbs energies of solvation for various pairs of solute + solvent systems.^{6–8,19,31} To the best of our knowledge, only a few databanks reporting such data or data estimated from molecular-simulation techniques exist. The main ones are as follows:

- (i) the Minnesota Solvation (MNSol) database developed by Truhlar, Cramer, and co-workers, updated in 2012, which compiles experimental values of molar Gibbs energy of solvation at 298 K for 3037 binary systems (involving 790 solutes and 92 solvents), estimated from experimentally measured properties (solute densities, solute vapor pressures, solute solubilities, partition coefficients, and Henry’s-law constants),³²
- (ii) another database developed in 2006 by the same research group reporting 2364 experimental hydration Gibbs energy values (which is the solvation Gibbs energy in the solvent water) for 182 solutes at different temperatures (estimated from solute vapor pressures, solute solubilities, infinite-dilution activity coefficients, and Henry’s-law constants),³³
- (iii) the FreeSolv database proposed in 2014 by Mobley and Guthrie reporting hydration Gibbs energy values at 298 K for more than 600 solutes.³⁴

As mentioned above, predictive methods can be used alternatively for estimating Gibbs energies of solvation. A comprehensive review of these methods can be found in a recent article by Jalan *et al.*²⁹ The main ones are as follows:

- (i) The Linear Solvation Energy Relationship (LSER) model is of the QSAR (Quantitative Structure-Activity Relationship) type and can be used to express the solvation Gibbs energy with respect to solvent coefficients and solute molecular descriptors (excess molar refractivity, solute dipolarity/polarizability, hydrogen bond acidity and basicity, and McGowan characteristic volume). Solvent coefficients can be obtained from solubility

or chromatographic data, from databases, or can be estimated from group-contribution methods.^{23,31,35,36}

(ii) Various more theoretical *a priori* calculation methods can be used to predict Gibbs energies of solvation. They can be divided into two categories:

(a) Solvation models considering the solute molecule surrounded by a discrete number of solvent molecules: in this approach, the solvation quantities can be deduced from molecular simulation techniques (molecular dynamics or Monte Carlo) or integral equation theory.^{14,17,19,21,37–39}

(b) Continuum solvation models (also called *implicit solvation models*): in this case, the solute is represented by a discrete molecule while the solvent is considered as a continuum surrounding the solute.^{33,40–43} For instance, in COSMO-based models, the solvent is a dielectric continuum of given permittivity surrounding the solute molecule lying in a molecular cavity. These models process the screening charge density on the surface of solute molecules to calculate chemical potentials, from which solvation Gibbs energies are deduced. Screening charge densities being accessible from classical *ab initio* calculations, COSMO-based models can thus be considered totally predictive.^{44–46}

Note that some other continuum solvation models exist. A comparison of the performances of different implicit solvation models is proposed in Guthrie's study.⁴⁷

As a final remark regarding prediction methods, let us mention that although not addressed in the open literature, solvation Gibbs energies could be estimated from any existing predictive EoS, with the advantage that these models account naturally for the temperature dependence of these quantities. The relationships connecting solvation Gibbs energies and EoS can be found in papers by Hsieh and Lin.^{26–28}

The above presentation of the application fields and estimation methods of Gibbs solvation energies shows the great need of reliable data for practical and theoretical studies but also for a proper parameterization of models and assessment of their performances. Observing that current databases contain a rather limited number of data, that most of these data are provided at a fixed temperature, and that data usable for solvation Gibbs energy estimation (partition coefficients, infinite-dilution activity coefficients, Henry's-law constants, and so on) are today extensively accessible in the open literature or in commercial databanks for a very large number of systems, a comprehensive databank of solvation Gibbs energies, solvation entropies, and solvation enthalpies of pure components (solvation of a species by itself) and {infinitely dilute solute + solvent} binary systems has been developed and is proposed in the present article (the database is given in the [supplementary material](#)). For pure components, the database contains solvation quantities at saturated-liquid conditions for the sole reason that solvation Gibbs energy data can be simply generated under such conditions. For binary mixtures, it has been decided to only consider species that are infinitely dilute in a solvent because the corresponding solvation data are the focus of interest of most studies dealing with solvation issues.

To create a comprehensive databank, the highest number of experimental measurements describing the largest possible ranges of temperature were considered:

- (i) For pure species: vapor-pressure and saturated-liquid density data have been used.
- (ii) For binary systems: pure-component properties have been combined with either experimental vapor-liquid equilibrium measurements (temperature, pressure, liquid-phase, and vapor-phase compositions) or infinite-dilution activity coefficients or Henry's-law constants. Note that we considered data published up to and including 2015.

Surprisingly, while the aforementioned types of thermodynamic data are profuse, they have almost never been used for the estimation of solvation Gibbs energies.

Eventually, it has been possible to estimate molar Gibbs energies of solvation for around 2000 pure components and 14 100 binary mixtures (non-electrolytic solutes were exclusively considered). In particular, we note that 136 binary systems use ionic liquids as solvents. For all pure components and the majority of binary systems, Gibbs energies of solvation were estimated at different temperatures, making it possible to estimate enthalpies and entropies of solvation.

2. Expressions of Solvation Energies for Their Practical Estimation from Experimental Data

2.1. Expression of the Gibbs energy of solvation derived from Ben-Naim's definition

Ben-Naim's definition of a solvation process has been previously introduced. As an important feature, during the solvation process, a molecule devoid of liberation energy, stemming from a p -component perfect-gas mixture at a fixed temperature T , pressure P , and composition $\mathbf{n}^\bullet = (n_1^\bullet, n_2^\bullet, \dots, n_p^\bullet)$ (vector of mole numbers), is transferred to a 1-phase real-fluid mixture (assumed to be liquid in the remainder of the paper) of composition $\mathbf{n} = (n_1, n_2, \dots, n_p)$ having the same temperature and pressure as the perfect-gas mixture. Note that the term *perfect gas* is preferred to *ideal gas* following a recent discussion.⁴⁸

The transfer of a molecule i through Ben-Naim's solvation process is illustrated in Fig. 2. Notations used thereafter are introduced in this figure. In particular, note that superscript “ \bullet ” designates perfect-gas properties.

It should be noted that Ben-Naim's definition does not necessitate any assumption about the concentration of the solute in the liquid mixture, this latter being either infinitely dilute or concentrated. Regarding vocabulary issues, the use of the word *solute*—which generally refers to infinite-dilution conditions—could be avoided by using the more general term *solvaton* instead, as advised by Ben-Naim. However, in the present paper, following general practices, it has been decided to consider the word “solute” as a synonym of “solvaton.”

The partial molar **Gibbs energy of solvation of a solute i** (or equivalently, its *chemical potential of solvation*) is the chemical potential change accompanying the solvation process. The chemical potential of a solute i free of liberation

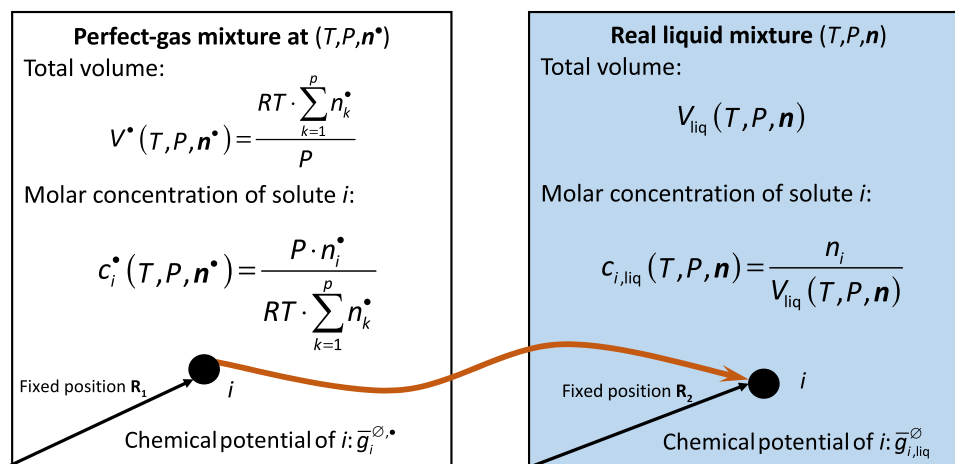


FIG. 2. Illustration of Ben-Naim's solvation process.

energy (i.e., without kinetic energy) in the initial state is denoted by $\bar{g}_i^{\ominus,*}$; the chemical potential of a fixed solute i in the final-state liquid mixture is denoted by $\bar{g}_{i,\text{liq}}^{\ominus}$. The expression of the solvation chemical potential $\Delta_{\text{solv}}\bar{g}_i$ is thus

$$\Delta_{\text{solv}}\bar{g}_i(T, P, \mathbf{n}) := \bar{g}_{i,\text{liq}}^{\ominus}(T, P, \mathbf{n}) - \bar{g}_i^{\ominus,*}(T, P). \quad (1)$$

The quantity \bar{g}_i^{\ominus} was christened **pseudo chemical potential** (PCP) by Ben-Naim;⁵ in Eq. (1), $\bar{g}_{i,\text{liq}}^{\ominus}$ is the molar PCP of i in the real liquid phase and $\bar{g}_i^{\ominus,*}$ denotes the molar PCP of i in the corresponding fictitious perfect-gas mixture at the same T and P as the liquid mixture. Note that $\bar{g}_i^{\ominus,*}$ has the feature to be independent of the composition \mathbf{n}^* (a proof of this assertion is given in Appendix A).

Following statistical mechanical derivations⁵ (assuming that the gas-phase internal partition function is not affected by the solvent), the molar PCP of a rigid molecule i can be expressed as a function of the (classical) molar chemical potential of i (denoted \bar{g}_i) following

$$\bar{g}_i^{\ominus}(T, P, \mathbf{n}) = \bar{g}_i(T, P, \mathbf{n}) - RT \ln [c_i(T, P, \mathbf{n}) \cdot N_A \cdot \Lambda_i^3(T)], \quad (2)$$

where R is the gas constant, c_i is the molar concentration of i , N_A is the Avogadro constant, and Λ_i is the thermal

De Broglie wavelength of i which is temperature-dependent,

$$\Lambda_i(T) = \frac{h}{\sqrt{2\pi m_i kT}}, \quad (3)$$

where h , k , and m_i are, respectively, the Planck constant, the Boltzmann constant, and the molecular mass of i .

Sections 2.2–2.6 explain how $\Delta_{\text{solv}}\bar{g}_i$ can be linked to thermodynamic properties accessible from experiments.

2.2. General expressions of chemical potentials of solvation

By introducing x_i and x_i^* , the mole fractions of i in the liquid and perfect-gas systems, respectively, \mathbf{x} and \mathbf{x}^* , the corresponding mole fraction vectors, and by combining Eqs. (1) and (2), one obtains

$$\begin{aligned} \Delta_{\text{solv}}\bar{g}_i(T, P, \mathbf{x}) &= \bar{g}_{i,\text{liq}}(T, P, \mathbf{x}) - \bar{g}_i^*(T, P, \mathbf{x}^*) \\ &\quad - RT \ln \left[\frac{c_{i,\text{liq}}(T, P, \mathbf{x})}{c_i^*(T, P, \mathbf{x}^*)} \right]. \end{aligned} \quad (4)$$

The following substitutions are made:

$$\left\{ \begin{aligned} \bar{g}_{i,\text{liq}}(T, P, \mathbf{x}) - \bar{g}_i^*(T, P, \mathbf{x}^*) &= \left[\bar{g}_{i,\text{liq}}(T, P, \mathbf{x}) - \bar{g}_i(T, P, \mathbf{x}) \right] + \left[\bar{g}_i(T, P, \mathbf{x}) - \bar{g}_i^*(T, P, \mathbf{x}^*) \right], \\ &= RT \ln \varphi_{i,\text{liq}}(T, P, \mathbf{x}) + RT \ln (x_i/x_i^*) \\ c_{i,\text{liq}}(T, P, \mathbf{x}) &= \frac{n_i}{V_{\text{liq}}(T, P, \mathbf{n})} = \frac{n_i}{\sum_{k=1}^p n_k} \cdot \frac{\sum_{k=1}^p n_k}{V_{\text{liq}}(T, P, \mathbf{n})} = x_i \cdot \rho_{\text{liq}}(T, P, \mathbf{n}), \\ c_i^*(T, P, \mathbf{x}^*) &= \frac{n_i^*}{V^*(T, P, \mathbf{n}^*)} = \frac{x_i^* \sum_{k=1}^p n_k^*}{V^*(T, P, \mathbf{n}^*)} = \frac{x_i^* P}{RT}, \end{aligned} \right. \quad (5)$$

where $\varphi_{i,\text{liq}}$ is the *fugacity coefficient* of species i . Hence,

$$\Delta_{\text{solv}}\bar{g}_i(T, P, \mathbf{x}) = RT \ln \left[\frac{P \cdot \varphi_{i,\text{liq}}(T, P, \mathbf{x})}{RT \cdot \rho_{\text{liq}}(T, P, \mathbf{x})} \right]. \quad (6)$$

Note that Eq. (6) was derived straightforwardly from the definition of the solvation process, without making any assumptions, and applies for any mixture containing an arbitrary number of components. It can be shown (see Appendix B) that $\Delta_{\text{solv}}\bar{g}_i$ is a typical liquid-phase property and is thus not very sensitive to pressure.

The aim of the present study being to develop a database of solvation data, the first concern that comes in mind is the availability of the property data required for the estimation of $\Delta_{\text{solv}}\bar{g}_i$. While density data are rather plentiful, fugacity coefficients cannot be simply estimated at any temperature and pressure from properties accessible from experiments. More information on how they can be estimated is given in Secs. 2.3 and 2.4.

2.3. Practical estimation of chemical potentials of solvation for pure components

When dealing with pure species, the solvation process is called *self-solvation*. The expression of the solvation chemical potential of i is deduced from Eq. (6),

$$\Delta_{\text{solv}}g_{\text{pure } i}(T, P) = RT \ln \left[\frac{P \cdot \varphi_{\text{pure } i,\text{liq}}(T, P)}{RT \cdot \rho_{\text{pure } i,\text{liq}}(T, P)} \right], \quad (7)$$

with $\varphi_{\text{pure } i,\text{liq}}$ the fugacity coefficient of pure i and $\rho_{\text{pure } i,\text{liq}}$ the molar density of pure liquid i at T and P . This molar density is reported for a large number of compounds but only at saturation conditions (sat), i.e., for a pure liquid phase in equilibrium with its vapor. By applying Eq. (7) in the case where the pure component i is in VLE at temperature T , one obtains

$$\Delta_{\text{solv}}g_{\text{pure } i}^{\text{sat}}(T) = RT \ln \left[\frac{P_i^{\text{sat}}(T) \cdot \varphi_i^{\text{sat}}(T)}{RT \cdot \rho_{i,\text{liq}}^{\text{sat}}(T)} \right], \quad (8)$$

where P_i^{sat} , the vapor pressure of i , is an extensively measured property. $\varphi_i^{\text{sat}}(T)$ is the fugacity coefficient of pure i in VLE that is common to the saturated-liquid and saturated-vapor phases. Although not measurable and thus not reported in usual databanks, φ_i^{sat} can be estimated from simple though efficient models, e.g., models derived from the law of corresponding states or the truncated virial EoS involving the second virial coefficient (denoted B_i thereafter), which is a property frequently tabulated in pure-component databanks.

Note that the pure-component VLE specification is subsequently denoted as S1. The term *specification* must be understood as “particular application conditions” of a general formula (in other words, the temperature, pressure, or composition domains of application are restricted but the equation

remains exact); on the contrary, the term *assumption* must be understood as “approximation” and the formula is no longer rigorously exact.

2.4. Expressions of chemical potentials of solvation of infinitely dilute solutes

Binary mixtures contain two components denoted as i and j thereafter. In the case of infinitely dilute solutions, indices i and j designate the solute and the solvent, respectively. In order to derive Henry's-law constants, partition coefficients, or infinite-dilution activity coefficients in the mathematical expression of $\Delta_{\text{solv}}\bar{g}_i(T, P, \mathbf{x})$ for an infinitely dilute solute, it is convenient to introduce in Eq. (6) the fugacity $f_{i,\text{liq}}(T, P, \mathbf{x})$ of solute i . By definition,

$$f_{i,\text{liq}}(T, P, \mathbf{x}) = P \cdot x_i \cdot \varphi_{i,\text{liq}}(T, P, \mathbf{x}). \quad (9)$$

By combining Eqs. (6) and (9), one has

$$\Delta_{\text{solv}}\bar{g}_i(T, P, \mathbf{x}) = RT \ln \left[\frac{f_{i,\text{liq}}(T, P, \mathbf{x})}{x_i} \times \frac{1}{RT \cdot \rho_{\text{liq}}(T, P, \mathbf{x})} \right]. \quad (10)$$

In the case where solute i is infinitely dilute in solvent j , the mixture density tends to the pure solvent density: $\rho_{\text{liq}}(T, P, \mathbf{x}) \rightarrow \rho_{\text{pure } j,\text{liq}}(T, P)$, and Eq. (10) becomes

$$\Delta_{\text{solv}}\bar{g}_i^{\infty}(T, P) = RT \ln \left[\left(\frac{f_{i,\text{liq}}(T, P, \mathbf{x})}{x_i} \right)^{\infty} \times \frac{1}{RT \cdot \rho_{\text{pure } j,\text{liq}}(T, P)} \right]. \quad (11)$$

Note that the superscript “ ∞ ” is used to indicate infinite-dilution conditions.

2.4.1. Estimation of $\Delta_{\text{solv}}\bar{g}_i^{\infty}(T, P)$ from Henry's-law constant measurements

Classical thermodynamics defines the Henry's-law constant of solute i in solvent j , denoted as $H_{ij}(T, P)$, as

$$H_{ij}(T, P) = \lim_{x_i \rightarrow 0} \left[\frac{f_{i,\text{liq}}(T, P, \mathbf{x})}{x_i} \right]. \quad (12)$$

By incorporating Eq. (12) in Eq. (11), the following expression of $\Delta_{\text{solv}}\bar{g}_i^{\infty}$ is obtained:

$$\Delta_{\text{solv}}\bar{g}_i^{\infty}(T, P) = RT \ln \left[\frac{H_{ij}(T, P)}{RT \cdot \rho_{\text{pure } j,\text{liq}}(T, P)} \right]. \quad (13)$$

Equation (13) is rigorous and can be applied if experimental values of both Henry's-law constant and pure-solvent density

are available at a given (T, P) . In practice, several cases may be encountered:

Case 1: $H_{ij}(T, P)$ is not very sensitive to pressure, and thus the pressure at which it is measured is frequently not reported. In such a case, it is assumed that Henry's-law constant is measured using an ebulliometer so that the total pressure of the VLE system (involving a solute nearly infinitely dilute in the liquid phase) is close to the solvent vapor pressure: $P_j^{\text{sat}}(T)$. In such a case, $\rho_{\text{pure } j, \text{liq}}(T, P) \rightarrow \rho_{j, \text{liq}}^{\text{sat}}(T)$ (assumption A1) and $\Delta_{\text{sol}v} \bar{g}_i^\infty$ becomes pressure-independent. It can be calculated through

$$\Delta_{\text{sol}v} \bar{g}_i^\infty(T, P \approx P_j^{\text{sat}}(T)) = RT \ln \left[\frac{H_{ij}(T, P)}{RT \cdot \rho_{j, \text{liq}}^{\text{sat}}(T)} \right]. \quad (14)$$

Case 2: the pressure at which $H_{ij}(T, P)$ was measured is reported. In the particular case where P does not deviate much from $P_j^{\text{sat}}(T)$, let us say $|P - P_j^{\text{sat}}(T)| \leq 10$ bar, it can be assumed that the liquid solvent j is incompressible in the pressure range $[P_j^{\text{sat}}(T); P]$ and consequently $\rho_{\text{pure } j, \text{liq}}(T, P)$ can be replaced by $\rho_{j, \text{liq}}^{\text{sat}}(T)$ (assumption A1). As explained above, this latter property is reported in a number of pure-component property databanks. If an experimental or estimated value of $\rho_{j, \text{liq}}^{\text{sat}}(T)$ is available, Eq. (14) can be used. On the other hand, if $|P - P_j^{\text{sat}}(T)| > 10$ bar, $\rho_{\text{pure } j, \text{liq}}(T, P)$ cannot be replaced by $\rho_{j, \text{liq}}^{\text{sat}}(T)$ and in such a case, $\Delta_{\text{sol}v} \bar{g}_i^\infty(T, P)$ cannot be calculated.

2.4.2. Estimation of $\Delta_{\text{sol}v} \bar{g}_i^\infty(T, P)$ from partition coefficients (i.e., VLE solubility data) measured in a composition range where Henry's law can be used as a sound approximation of solute behavior

In the composition range where Henry's law can be used as a sound approximation of solute i fugacity (assumption A2), Eq. (12) can be simplified as

$$f_{i, \text{liq}}(T, P, \mathbf{x}) \underset{x_i \text{ low enough}}{\approx} x_i \cdot H_{ij}(T, P). \quad (15)$$

If such a liquid phase is in equilibrium with a gas phase (specification S2), the equilibrium condition

$$f_{i, \text{liq}}(T, P, \mathbf{x}) = f_{i, \text{gas}}(T, P, \mathbf{y}) \quad (16)$$

makes it possible to write

$$x_i \cdot H_{ij}(T, P) = P \cdot y_i \cdot \varphi_{i, \text{gas}}(T, P, \mathbf{y}). \quad (17)$$

By introducing the partition coefficient

$$K_i = \frac{y_i}{x_i} \quad (18)$$

in Eq. (17), one has

$$H_{ij}(T, P) = P \cdot K_i \cdot \varphi_{i, \text{gas}}(T, P, \mathbf{y}). \quad (19)$$

Finally, by combining Eqs. (13) and (19), one obtains

$$\Delta_{\text{sol}v} \bar{g}_i^\infty = RT \ln \left[\frac{P \cdot K_i \cdot \varphi_{i, \text{gas}}(T, P, \mathbf{y})}{RT \cdot \rho_{\text{pure } j, \text{liq}}(T, P)} \right]. \quad (20)$$

In Eq. (20), for a zeotropic binary system, the total pressure P at which a VLE is measured lies between $P_i^{\text{sat}}(T)$ and $P_j^{\text{sat}}(T)$. Since Henry's law applies to dilute solute i , the total pressure is certainly close to $P_j^{\text{sat}}(T)$. By assuming that the liquid solvent is incompressible between $P_j^{\text{sat}}(T)$ and P , $\rho_{\text{pure } j, \text{liq}}(T, P)$ can be approximated by $\rho_{j, \text{liq}}^{\text{sat}}(T)$ (assumption A1). Moreover, in the case where the total pressure is low to moderate, the gas phase may be considered as perfect and the fugacity coefficient of any component in the gas phase is equal to one: $\varphi_{i, \text{gas}}(T, P, \mathbf{y}) = 1$ (assumption A3).

To conclude

- if the compositions of the liquid (x_i) and gas (y_i) phases in equilibrium are experimentally determined at low to moderate pressure,
- if Henry's law is a sound approximation of solute i behavior in the liquid equilibrium phase (x_i),
- if the solvent can be considered as incompressible [and thus $\rho_{\text{pure } j, \text{liq}}(T, P) \approx \rho_{j, \text{liq}}^{\text{sat}}(T)$],

therefore, Eq. (20) simplifies as follows:

$$\Delta_{\text{sol}v} \bar{g}_i^\infty(T, P) \approx RT \ln \left[\frac{P \cdot K_i(T, P)}{RT \cdot \rho_{j, \text{liq}}^{\text{sat}}(T)} \right]. \quad (21)$$

For better understanding the various assumptions, these latter are summarized in Fig. 3, representing an isothermal projection of the VLE behavior of an arbitrary binary system.

2.4.3. Estimation of $\Delta_{\text{sol}v} \bar{g}_i^\infty(T, P)$ from Raoult's-law infinite-dilution activity-coefficient measurements

According to classical thermodynamics, a straightforward relation exists between Henry's-law constants and Raoult's-law

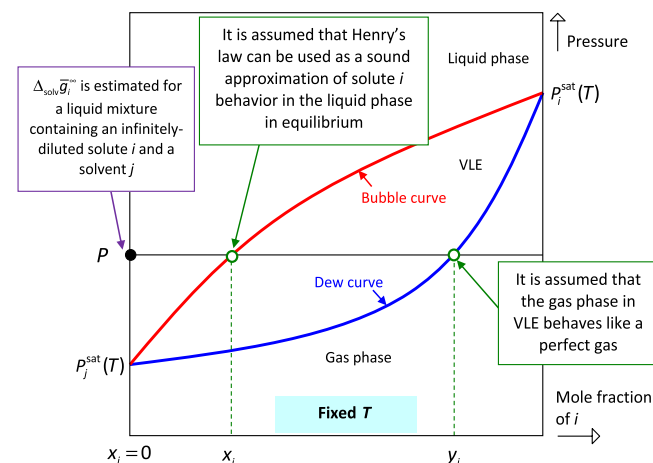


FIG. 3. Summary of the various assumptions used to derive Eq. (21) in the Pxy isothermal VLE phase diagram of an arbitrary binary system.

infinite-dilution activity coefficients (denoted as γ_i^∞). Consequently, γ_i^∞ can also be used to determine $\Delta_{\text{solv}}\bar{g}_i^\infty(T, P)$. To start, it is recalled that the fugacity of species i ($f_{i,\text{liq}}(T, P, \mathbf{x})$) is related to the Raoult's-law activity coefficient through

$$f_{i,\text{liq}}(T, P, \mathbf{x}) = x_i \cdot \gamma_{i,\text{liq}}(T, P, \mathbf{x}) \cdot f_{\text{pure } i,\text{liq}}(T, P). \quad (22)$$

By combining Eqs. (12) and (22), one has

$$\begin{aligned} H_{ij}(T, P) &= \lim_{x_i \rightarrow 0} \left[\gamma_{i,\text{liq}}(T, P, \mathbf{x}) \cdot f_{\text{pure } i,\text{liq}}(T, P) \right] \\ &= \gamma_i^\infty(T, P) \times f_{\text{pure } i,\text{liq}}(T, P). \end{aligned} \quad (23)$$

Equations (13) and (23) lead to

$$\Delta_{\text{solv}}\bar{g}_i^\infty(T, P) = RT \ln \left[\frac{\gamma_i^\infty(T, P) \cdot f_{\text{pure } i,\text{liq}}(T, P)}{RT \cdot \rho_{\text{pure } j,\text{liq}}(T, P)} \right]. \quad (24)$$

Equation (24) is rigorous but not easily usable in practice and thus must be adapted. Several cases are now considered.

Case 1: $\gamma_i^\infty(T, P)$ is not very sensitive to pressure so that the pressure at which it is measured is frequently not reported. In such a case, it is assumed that $\gamma_i^\infty(T, P)$ is measured using an ebulliometer, so that at infinite dilution the total pressure is close to the solvent vapor pressure $P_j^{\text{sat}}(T)$. In such a case,

$$\rho_{\text{pure } j,\text{liq}}(T, P) \rightarrow \rho_{j,\text{liq}}^{\text{sat}}(T) \quad (25)$$

(assumption A1) and

$$f_{\text{pure } i,\text{liq}}(T, P) \rightarrow f_{\text{pure } i,\text{liq}}(T, P_j^{\text{sat}}(T)). \quad (26)$$

The fugacity of pure solute i can be estimated through the following exact relation:

$$\begin{aligned} f_{\text{pure } i,\text{liq}}(T, P_j^{\text{sat}}(T)) \\ = P_i^{\text{sat}}(T) \cdot \varphi_i^{\text{sat}}(T) \cdot \exp \left[\frac{1}{RT} \int_{P_i^{\text{sat}}(T)}^{P_j^{\text{sat}}(T)} \frac{dP}{\rho_{\text{pure } i,\text{liq}}(T, P)} \right]. \end{aligned} \quad (27)$$

Under the assumption (A4) that pure liquid i is incompressible (its density does not depend on pressure), one has

$$f_{\text{pure } i,\text{liq}}(T, P_j^{\text{sat}}(T)) = \frac{P_i^{\text{sat}}(T) \cdot \varphi_i^{\text{sat}}(T)}{\exp \left[\frac{P_i^{\text{sat}}(T) - P_j^{\text{sat}}(T)}{RT \cdot \rho_{i,\text{liq}}^{\text{sat}}(T)} \right]}. \quad (28)$$

Combining Eqs. (24), (25), and (28), the following expression of the solvation chemical potential of i is obtained:

$$\begin{aligned} \Delta_{\text{solv}}\bar{g}_i^\infty(T, P = P_j^{\text{sat}}(T)) \\ \approx RT \ln \left[\frac{P_i^{\text{sat}}(T) \cdot \gamma_i^\infty(T) \cdot \varphi_i^{\text{sat}}(T)}{RT \cdot \rho_{j,\text{liq}}^{\text{sat}}(T) \cdot \exp \left[\frac{P_i^{\text{sat}}(T) - P_j^{\text{sat}}(T)}{RT \cdot \rho_{i,\text{liq}}^{\text{sat}}(T)} \right]} \right]. \end{aligned} \quad (29)$$

Case 2: the pressure at which $\gamma_i^\infty(T, P)$ was measured is reported. In the particular case where P does not deviate much from $P_j^{\text{sat}}(T)$, let us say $|P - P_j^{\text{sat}}(T)| \leq 10$ bar, it can be assumed that the liquid solvent j is incompressible in the pressure range $[P_j^{\text{sat}}(T); P]$ and consequently $\rho_{\text{pure } j,\text{liq}}(T, P)$ can be replaced by $\rho_{j,\text{liq}}^{\text{sat}}(T)$ (assumption A1). If, simultaneously, P does not deviate much from $P_i^{\text{sat}}(T)$, let us say $|P - P_i^{\text{sat}}(T)| \leq 10$ bar, it can be reasonably assumed that pure i is incompressible in the pressure range $[P; P_i^{\text{sat}}(T)]$ (assumption A4). Then Eq. (28) becomes

$$f_{\text{pure } i,\text{liq}}(T, P) = \frac{P_i^{\text{sat}}(T) \cdot \varphi_i^{\text{sat}}(T)}{\exp \left[\frac{P_i^{\text{sat}}(T) - P}{RT \cdot \rho_{i,\text{liq}}^{\text{sat}}(T)} \right]}. \quad (30)$$

The final expression of $\Delta_{\text{solv}}\bar{g}_i^\infty$ with respect to γ_i^∞ is given by

$$\Delta_{\text{solv}}\bar{g}_i^\infty(T, P) \approx RT \ln \left[\frac{P_i^{\text{sat}}(T) \cdot \gamma_i^\infty(T, P) \cdot \varphi_i^{\text{sat}}(T)}{RT \cdot \rho_{j,\text{liq}}^{\text{sat}}(T) \cdot \exp \left[\frac{P_i^{\text{sat}}(T) - P}{RT \cdot \rho_{i,\text{liq}}^{\text{sat}}(T)} \right]} \right]. \quad (31)$$

On the other hand, if $|P - P_j^{\text{sat}}(T)| > 10$ bar or $|P - P_i^{\text{sat}}(T)| > 10$ bar, $\Delta_{\text{solv}}\bar{g}_i^\infty(T, P)$ cannot be calculated.

2.5. Estimation of partial molar entropy and enthalpy of solvation

These quantities are classically derived from the chemical potential of solvation,

$$\begin{cases} \Delta_{\text{solv}}\bar{s}_i = - \left(\frac{\partial \Delta_{\text{solv}}\bar{g}_i}{\partial T} \right)_{P, \text{composition}}, \\ \Delta_{\text{solv}}\bar{h}_i = \Delta_{\text{solv}}\bar{g}_i + T \Delta_{\text{solv}}\bar{s}_i. \end{cases} \quad (32)$$

2.6. Summary of the expressions relating the chemical potential of solvation to experimentally accessible properties

This summary appears in Tables 1 and 2.

TABLE 1. List of expressions and related assumptions potentially usable for the estimation of the chemical potential of solvation (CPS) of species i . Specifications S1, S2 and assumptions A1 to A4 are explicated in Table 2

Applicability	Assumptions/ specifications	Expressions of the chemical potential of solvation (CPS) of species i	Equation
Pure component	None	$\Delta_{\text{solv}} g_{\text{pure } i}(T, P) = RT \ln \left[\frac{P \cdot \phi_{\text{pure } i, \text{liq}}(T, P)}{RT \cdot \rho_{\text{pure } i, \text{liq}}(T, P)} \right]$	(7)
	S1	$\Delta_{\text{solv}} g_{\text{pure } i}^{\text{sat}}(T) = RT \ln \left[\frac{P_i^{\text{sat}}(T) \cdot \phi_i^{\text{sat}}(T)}{RT \cdot \rho_{i, \text{liq}}^{\text{sat}}(T)} \right]$	(8)
Binary mixture	A1	$\Delta_{\text{solv}} \bar{g}_i^{\infty}(T, P \approx P_j^{\text{sat}}(T)) = RT \ln \left[\frac{H_{ij}(T)}{RT \cdot \rho_{j, \text{liq}}^{\text{sat}}(T)} \right]$	(14)
	S2 + A1 + A2 + A3	$\Delta_{\text{solv}} \bar{g}_i^{\infty}(T, P) \approx RT \ln \left[\frac{P}{RT} \cdot \frac{K_i}{\rho_{j, \text{liq}}^{\text{sat}}(T)} \right]$ with $K_i = \frac{y_i}{x_i}$	(21)
	A1 + A4	$\Delta_{\text{solv}} \bar{g}_i^{\infty}(T, P = P_j^{\text{sat}}(T)) \approx RT \ln \left[\frac{P_i^{\text{sat}}(T) \cdot \gamma_i^{\infty}(T, P) \cdot \phi_i^{\text{sat}}(T)}{RT \cdot \rho_{j, \text{liq}}^{\text{sat}}(T) \cdot \exp \left[\frac{P_i^{\text{sat}}(T) - P_j^{\text{sat}}(T)}{RT \cdot \rho_{i, \text{liq}}^{\text{sat}}(T)} \right]} \right]$ or	(29) or (31)
		$\Delta_{\text{solv}} \bar{g}_i^{\infty}(T, P) \approx RT \ln \left[\frac{P_i^{\text{sat}}(T) \cdot \gamma_i^{\infty}(T, P) \cdot \phi_i^{\text{sat}}(T)}{RT \cdot \rho_{j, \text{liq}}^{\text{sat}}(T) \cdot \exp \left[\frac{P_i^{\text{sat}}(T) - P}{RT \cdot \rho_{i, \text{liq}}^{\text{sat}}(T)} \right]} \right]$	

3. Methodology Used for Developing the CompSol Database

Experimental data contained in databanks of pure-component and binary-mixture thermodynamic properties cover large ranges of temperature and pressure (and composition also, in the case of binary mixtures). Equations connecting chemical potentials of solvation to thermodynamic properties (introduced in Sec. 2) were established provided a series of assumptions can be made, thus limiting the number of experimental data likely to be integrated within the solvation database. In this section, the algorithms used to select appropriate data and to derive solvation quantities are detailed.

3.1. Generation of pure-component solvation chemical potential data

The estimation of experimentally derived values of chemical potentials of solvation $\Delta_{\text{solv}} g_{\text{pure } i}^{\text{sat}}(T)$ for pure

TABLE 2. Description of the specifications and assumptions mentioned in Table 1

List of specifications:	
S1	Pure i is in VLE at given T
S2	A VLE exists for the binary system at fixed (T, P)
List of assumptions:	
A1	The liquid-phase density of pure j is independent of P [and thus $\rho_{\text{pure } j, \text{liq}}(T, P) \approx \rho_{j, \text{liq}}^{\text{sat}}(T)$]
A2	Henry's law assumption for solute i at (T, P, x)
A3	The solute i in the saturated gas phase at fixed (T, P) exhibits a perfect gas behavior, and thus $\phi_{i, \text{gas}} = 1$
A4	The liquid-phase density of pure i is independent of P [and thus $\rho_{\text{pure } i, \text{liq}}(T, P) \approx \rho_{i, \text{liq}}^{\text{sat}}(T)$]

components is based on Eq. (8). Chemical potentials of solvation of pure i can thus be evaluated at T provided the saturation pressure $P_i^{\text{sat}}(T)$, saturated-liquid density $\rho_{i, \text{liq}}^{\text{sat}}(T)$, and saturation fugacity coefficient $\phi_i^{\text{sat}}(T)$ can be estimated. In this study, the DIPPR (Design Institute for Physical Properties) database has been used to generate *pseudo-experimental* data of $P_i^{\text{sat}}(T)$ and $\rho_{i, \text{liq}}^{\text{sat}}(T)$.⁴⁹ These data are considered as pseudo-experimental, as for each pure component, the DIPPR database provides $P_i^{\text{sat}}(T)$ and $\rho_{i, \text{liq}}^{\text{sat}}(T)$ data in the form of temperature-dependent correlations fitted to real experimental data. These correlations can only be used in limited ranges of temperatures (denoted as $[T_{\text{min}}^{i, \text{psat}}, T_{\text{max}}^{i, \text{psat}}]$ and $[T_{\text{min}}^{i, \text{psat}}, T_{\text{max}}^{i, \text{psat}}]$). In most cases, one has $T_{\text{max}}^{i, \text{psat}} = T_{\text{max}}^{i, \text{psat}} = T_{c, i}$, where $T_{c, i}$ is the critical temperature of pure i . Saturation fugacity coefficients $\phi_i^{\text{sat}}(T)$ cannot be directly measured and are thus absent from pure-component property databases. The $\phi_i^{\text{sat}}(T)$ quantities have been estimated using the corresponding-states law as proposed by Hougen, Watson, and Ragatz (HWR).⁵⁰ The HWR method has been preferred over other corresponding-states methods as it provides correlations for $\phi_i^{\text{sat}}(T)$ (i.e., fugacity coefficients of pure components in VLE). Note however that the HWR model only applies for molecules of critical compressibility factor ranging between 0.25 and 0.29 and for water as well. For other molecules, the truncated virial EoS has been used after having verified the applicability of this model. Eventually, the following rules were applied to generate $\Delta_{\text{solv}} g_{\text{pure } i}^{\text{sat}}(T)$ data from $P_i^{\text{sat}}(T)$, $\rho_{i, \text{liq}}^{\text{sat}}(T)$, $B_i(T)$, and $\phi_i^{\text{sat}}(T)$ data in the largest possible range of temperatures:

- (1) Only components having a critical temperature $T_{c, i}$ lower than 950 K are considered for the next steps. Species showing higher critical temperatures are actually pure metals which are out of the scope of this study.

- (2) For each component i , a temperature range $[T_{\min,i}, T_{\max,i}]$, such that $P_i^{\text{sat}}(T)$ and $\rho_{i,\text{liq}}^{\text{sat}}(T)$ correlations can be simultaneously applied, is defined. The lower and upper limits are calculated as $T_{\min,i} = \max[T_{\min,i}^{\text{Psat}}, T_{\min,i}^{\rho_{i,\text{liq}}^{\text{sat}}}]$ and $T_{\max,i} = \min[T_{\max,i}^{\text{Psat}}, T_{\max,i}^{\rho_{i,\text{liq}}^{\text{sat}}}]$. If no correlation is available either for $P_i^{\text{sat}}(T)$ or $\rho_{i,\text{liq}}^{\text{sat}}(T)$ or if $[T_{\min,i}, T_{\max,i}] = \emptyset$, then the pure component is not considered further.
- (3) The estimation of saturation fugacity coefficients $\varphi_i^{\text{sat}}(T)$ was mainly carried out by using the 3-parameter HWR corresponding-states law which has the advantage to be both accurate and applicable from low to critical pressures. In the rare cases where this law could not be applied, the virial EoS, which can only be used to model weakly imperfect gases, was used instead. The following procedure has finally been followed:
- CASE 1: When the reduced pressure $P_{r,i}^{\text{sat}}(T) = P_i^{\text{sat}}(T)/P_{c,i} < 0.05$, then pure gaseous component i in VLE exhibits a perfect-gas behavior and the saturation fugacity coefficient of i is nearly equal to 1.
 - CASE 2: Otherwise, the HWR⁵⁰ corresponding-states law can be used for water and for any other pure compound i having its critical compressibility factor $z_{c,i}$ within the range $0.25 \leq z_{c,i} \leq 0.29$.

- (iii) CASE 3: Otherwise, the truncated virial EoS is used on condition that the reduced saturated pressure $P_{r,i}^{\text{sat}}(T)$ does not exceed 0.5. This model expresses the saturation fugacity coefficient as

$$\varphi_i^{\text{sat}}(T) = \exp\left[\frac{B_i(T) \cdot P_i^{\text{sat}}(T)}{RT}\right]. \quad (33)$$

The *second virial coefficient* $B_i(T)$ is obtained from DIPPR correlations (provided temperature T is included in the second-virial coefficient application range).

- (iv) Finally, if at the considered temperature conditions for applying cases 1, 2, or 3 are not met, then the pure component is not considered further.

For illustration, the three methods used to estimate $\varphi_i^{\text{sat}}(T)$ are compared in Fig. 4. It is observed that the HWR corresponding-states law and the virial EoS lead to similar trends and perfectly coincide for reduced pressures around 0.4.

- (4) $\Delta_{\text{sol}v}g_{\text{pure } i}^{\text{sat}}(T)$ data are then generated at different temperatures in the range $[T_{\min,i}, T_{\max,i}]$. For each pure component, the number of regularly spaced temperatures (N_{temp}) has been determined as follows:

$$\left\{ \begin{array}{ll} N_{\text{temp}} = 1 & \text{if } (T_{\max,i} - T_{\min,i}) < 5 \text{ K} \\ N_{\text{temp}} = 3 & \text{if } 5 \text{ K} \leq (T_{\max,i} - T_{\min,i}) < 50 \text{ K} \\ N_{\text{temp}} = 5 & \text{if } 50 \text{ K} \leq (T_{\max,i} - T_{\min,i}) < 100 \text{ K} \\ N_{\text{temp}} = 5(i+1) & \text{if } (i \times 100 \text{ K}) < (T_{\max,i} - T_{\min,i}) < [(i+1) \times 100 \text{ K}] \text{ with } i = 1, 2, \dots \end{array} \right.$$

When possible, $\Delta_{\text{sol}v}g_{\text{pure } i}^{\text{sat}}$ at 298.15 K is also estimated.

- (5) For each pure component and at a given temperature, $\Delta_{\text{sol}v}g_{\text{pure } i}^{\text{sat}}(T)$ is calculated according to Eq. (8).
- (6) Molar enthalpies $\Delta_{\text{sol}v}h_{\text{pure } i}^{\text{sat}}(T)$ and entropies $\Delta_{\text{sol}v}s_{\text{pure } i}^{\text{sat}}(T)$ of solvation have been numerically estimated from Eq. (32). Uncertainties of $\Delta_{\text{sol}v}g_{\text{pure } i}^{\text{sat}}(T)$ data estimations have been also estimated from uncertainties in $P_i^{\text{sat}}(T)$ and $\rho_{i,\text{liq}}^{\text{sat}}(T)$ data reported in the DIPPR database and by using the propagation-error formula. Due to a lack of reliable information on the accuracy of $\varphi_i^{\text{sat}}(T)$ estimation methods, the corresponding uncertainty has been neglected. The calculated relative uncertainty estimation for $\Delta_{\text{sol}v}g_{\text{pure } i}^{\text{sat}}(T)$ is thus certainly optimistic,

$$\Delta\left(\Delta_{\text{sol}v}g_{\text{pure } i}^{\text{sat}}(T)\right)_{\%} = \frac{\sqrt{\Delta(P_i^{\text{sat}})_{\%}^2 + \Delta(\rho_{i,\text{liq}}^{\text{sat}})_{\%}^2}}{\left| \ln\left[\frac{P_i^{\text{sat}}(T) \cdot \varphi_i^{\text{sat}}(T)}{RT \cdot \rho_{i,\text{liq}}^{\text{sat}}(T)}\right] \right|}. \quad (34)$$

3.2. Generation of binary solvation chemical potential data at infinite dilution

3.2.1 Estimation of $\Delta_{\text{sol}v}\bar{g}_i^{\infty}(T, P)$ from complete VLE measurements

Following Eq. (21), the solvation chemical potential $\Delta_{\text{sol}v}\bar{g}_i^{\infty}(T, P)$ of species i at infinite dilution at a given (T, P) can be estimated from simultaneous knowledge of a complete VLE datapoint (T, P, x_i, y_i) and the saturated-liquid density $\rho_{j,\text{liq}}^{\text{sat}}(T)$ of solvent j . As previously detailed, Eq. (21) was derived under certain assumptions (A1–A3), which must be carefully verified before applying it. To do so, and as explained below, additional properties are required such as the critical pressure of pure solute i ($P_{c,i}$) or the saturation pressure of pure solvent j (P_j^{sat}). It is recalled that pure-component properties were taken from the DIPPR databank. Experimental binary VLE data were extracted from the Dortmund databank (DDB) which has the feature to be nearly exhaustive. The following steps have been followed to estimate $\Delta_{\text{sol}v}\bar{g}_i^{\infty}(T, P)$ in keeping with assumptions A1–A3:

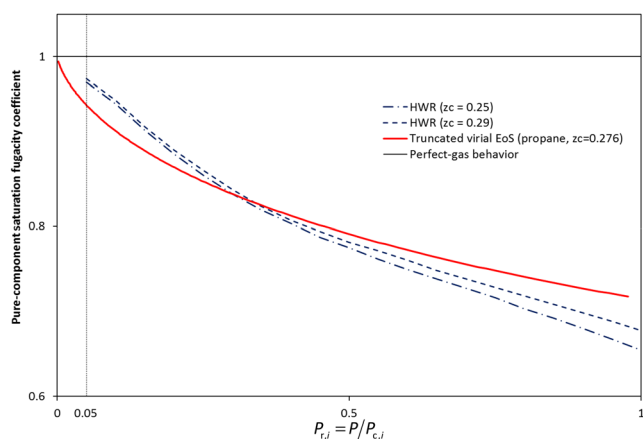


Fig. 4. Saturation fugacity coefficients of pure components versus reduced pressure. Dashed and dashed-dotted lines: predictions from the HWR corresponding-states law for a pure component with $z_c = 0.25$ and $z_c = 0.29$, respectively. Continuous bold line: prediction from the truncated virial EoS for pure propane ($z_c = 0.279$). Horizontal line: perfect-gas behavior.

(1) For a given pair of solute + solvent, it is checked first whether the DDB and DIPPR databanks provide a consistent property set for the estimation of $\Delta_{\text{sol}}\bar{g}_i^\infty(T, P)$:

(1.1) Binary VLE data extracted from the DDB database must be complete, i.e., the temperature, pressure, and compositions of the two phases ($T, P, \mathbf{x}, \mathbf{y}$) must be known. Otherwise, VLE data are not considered for our study. Pure-component VLE data (such that $x_i = 0$ or $y_i = 0$) are also disregarded.

(1.2) For a given VLE datapoint ($T, P, \mathbf{x}, \mathbf{y}$), the words solute and “solvent” are assigned following the criterion $x_{\text{solute } i} < x_{\text{solvent } j}$.

(1.3) For each binary system, the presence of pure-component properties [$\rho_{j,\text{liq}}^{\text{sat}}(T)$, $P_j^{\text{sat}}(T)$, and $P_{c,i}$] in the DIPPR database is checked. In case of failure, the binary mixture is not considered further.

(1.4) Is the VLE temperature T of the binary-system data within the temperature ranges of DIPPR correlations for $\rho_{j,\text{liq}}^{\text{sat}}(T)$ and $P_j^{\text{sat}}(T)$? If not, corresponding VLE datapoints are removed.

(2) Validation of the assumptions associated with Eq. (21):

(2.1) To consider that the liquid-phase density of pure j at (T, P) is pressure-independent (A1) and thus can be approximated by its saturated-liquid density at T , the pressure P must be close enough to the saturation pressure $P_j^{\text{sat}}(T)$ of pure j at T . The following criterion has been used to quantify the closeness of the VLE pressure and the vapor pressure of pure j at the same temperature:

$$\left| P - P_j^{\text{sat}}(T) \right| \leq 10 \text{ bar}. \quad (35)$$

(2.2) According to assumption (A2), the Henry’s-law activity coefficient of solute i at (T, P, \mathbf{x}) must be equal to one (in other words, Henry’s law must be applicable to component i). The Gibbs-Duhem equation makes it possible to deduce that if Henry’s law is valid for the solute, then the solvent obeys

Raoult’s law and reciprocally. By combining Raoult’s law for solvent j and the binary VLE relation, one obtains

$$P \cdot y_j(T, P) = P_j^{\text{sat}}(T) \cdot x_j(T, P). \quad (36)$$

Finally, assumption (A2) will be considered as valid provided

$$\left| 1 - \frac{P \cdot y_j(T, P)}{P_j^{\text{sat}}(T) \cdot x_j(T, P)} \right| \leq 0.01. \quad (37)$$

(2.3) To ensure that the gas phase in VLE at (T, P) can be considered as a perfect-gas mixture (A3) (and thus that the fugacity coefficient of i in the gas phase is equal to 1), the VLE pressure P must be both low enough and far from the critical pressure of i . The following criterion is used:

$$P < \min \left\{ 0.05 P_{c,i}; 5 \text{ bar} \right\}. \quad (38)$$

(3) Due to the observation of inconsistent solvation-chemical-potential data in the CompSol database, the following two steps have been added *a posteriori*:

(3.1) High deviations from general trends in the ($\Delta_{\text{sol}}\bar{g}_i^\infty, T$) plane were observed in the cases where VLE data associated with very low molar fraction of solute were used. Such deviations result from the inaccuracy accompanying the measurement of mole fractions lower than 0.01. To overcome this issue, only points for which $y_i > 0.01$ and $x_i > 0.01$ were considered.

(3.2) Some series of experimental VLE datasets contain many datapoints measured at the same (T, P), by the same author, for the same binary mixture but leading to scattered values of the partition coefficient. To identify these inconsistent experimental behaviors, the following test was implemented:

(i) For each considered dataset, the maximum and minimum experimental temperatures T_{min} and T_{max} , the maximum and minimum experimental pressures P_{min} and P_{max} , and the maximum and minimum partition coefficients of the solute $K_{i,\text{min}}$ and $K_{i,\text{max}}$ are calculated.

(ii) If $T_{\text{max}} - T_{\text{min}} < 1 \text{ K}$, $P_{\text{max}} - P_{\text{min}} < 0.05 \text{ bar}$, or $\frac{K_{i,\text{max}} - K_{i,\text{min}}}{K_{i,\text{min}}} > 0.07$, then the dataset is considered as thermodynamically inconsistent and is not considered further.

(4) Once steps (1)–(3) are validated, the chemical potential of solvation of i can be estimated from Eq. (21).

3.2.2. Estimation of $\Delta_{\text{sol}}\bar{g}_i^\infty(T, P)$ from Raoult’s-law infinite-dilution activity-coefficient measurements

Equation (29) or (31) is used to relate $\Delta_{\text{sol}}\bar{g}_i^\infty(T, P)$ to $\gamma_i^\infty(T, P)$ depending on whether an experimental pressure is

explicitly associated with the γ_i^∞ data. These equations also involve pure-component saturated-liquid properties: the saturation pressures $P_i^{\text{sat}}(T)$ and $P_j^{\text{sat}}(T)$, the saturated-liquid densities $\rho_{i,\text{liq}}^{\text{sat}}(T)$ and $\rho_{j,\text{liq}}^{\text{sat}}(T)$, and the fugacity coefficient of i at saturation $\varphi_i^{\text{sat}}(T)$. The critical pressure of pure i ($P_{c,i}$) is indirectly required to estimate this latter quantity. As previous, the binary data [$\gamma_i^\infty(T)$ in the present case] are extracted from the DDB while saturated-liquid properties of pure species stem from the DIPPR databank.

- (i) For only 0.34% of the experimental γ_i^∞ data reported in the DDB, the temperature T and pressure P are both indicated. In such a case, $\Delta_{\text{sol}}\bar{g}_i^\infty(T, P)$ can be evaluated using Eq. (31).
- (ii) For all other experimental datapoints, the experimental pressure value is not mentioned. As explained below, Eq. (29) is used in such cases.

The following algorithm is proposed to estimate $\Delta_{\text{sol}}\bar{g}_i^\infty(T, P)$ or $\Delta_{\text{sol}}\bar{g}_i^\infty(T, P = P_j^{\text{sat}}(T))$ from γ_i^∞ data:

- (1) This step is similar to step (1) implemented in Sec. 3.2.1.
 - (i) Exploitable and compatible experimental data are selected both from the Dortmund and DIPPR databanks [regarding the compatibility of the data, it is checked whether the temperature of the $\gamma_i^\infty(T)$ data is within the common temperature range of the $P_i^{\text{sat}}(T)$, $P_j^{\text{sat}}(T)$, $\rho_{i,\text{liq}}^{\text{sat}}(T)$, and $\rho_{j,\text{liq}}^{\text{sat}}(T)$ DIPPR correlations].
 - (ii) For each datapoint, the names of the solute and solvent are extracted from the DDB.
- (2) Validation of the assumptions associated with Eqs. (29) and (31):
 - (i) Assumptions (A1) and (A4) are considered as valid provided the following criteria are simultaneously verified:

$$\begin{cases} |P - P_i^{\text{sat}}(T)| \leq 10 \text{ bar} \\ |P - P_j^{\text{sat}}(T)| \leq 10 \text{ bar} \end{cases} \quad (39)$$

- (ii) Note that if not indicated, the experimental pressure is approximated by the vapor pressure of pure j at T .
- (3) The fugacity coefficient of pure i at saturation is estimated as previous.
- (4) Calculation of $\Delta_{\text{sol}}\bar{g}_i^\infty(T, P)$ or $\Delta_{\text{sol}}\bar{g}_i^\infty(T, P = P_j^{\text{sat}}(T))$ using Eq. (29) or (31).

3.2.3. Estimation of $\Delta_{\text{sol}}\bar{g}_i^\infty(T, P)$ from Henry's-law constant measurements

Equation (14) is used for the estimation of $\Delta_{\text{sol}}\bar{g}_i^\infty(T, P)$ from Henry's-law constant data $H_{i,j}(T, P)$ (extracted from the DDB). As for activity coefficients at infinite dilution, although Henry's-law constant theoretically depends on pressure, experimental pressure values are not systematically indicated since Henry's-law constant is not very sensitive to pressure (like most incompressible liquid properties). In the DDB, it was observed that for 84% of the data, the pressure P was not reported. In such cases, the

pressure was approximated by $P = P_j^{\text{sat}}(T)$. The following saturated-liquid properties involved in Eq. (14) or used to verify assumption (A1) were taken from the DIPPR database: saturated-liquid density of the solvent $\rho_{j,\text{liq}}^{\text{sat}}(T)$ and vapor pressure of the solvent $P_j^{\text{sat}}(T)$. The following algorithm was used:

- (1) This step is similar to step (1) implemented in Sec. 3.2.2.
- (2) Equation (14) is based on one assumption to be checked:
 - (i) The following criterion is used to check assumption (A1):

$$|P - P_j^{\text{sat}}(T)| \leq 10 \text{ bar}. \quad (40)$$
 - (ii) If not indicated, the experimental pressure is approximated by the vapor pressure of pure j at T .
- (3) Calculation of $\Delta_{\text{sol}}\bar{g}_i^\infty(T, P)$ or $\Delta_{\text{sol}}\bar{g}_i^\infty(T, P = P_j^{\text{sat}}(T))$ using Eq. (14).

3.2.4. Methodology for systems using an ionic liquid as solvent

Systems involving an ionic liquid as solvent require a specific treatment. Indeed, while Henry's-law constant data are available in the literature for such systems, the database used for the extraction of pure-component properties (DIPPR) does not contain any ionic liquids. However, pure-ionic-liquid density data can also be found in the literature.

To include ionic-liquid systems in our database, we have thus decided to compile density data of pure ionic liquids from the literature and, for each ionic liquid, to estimate the four coefficients (C_1 – C_4) and the corresponding validity range [$T_{\text{min}}^{\rho_{i,\text{liq}}^{\text{sat}}}$, $T_{\text{max}}^{\rho_{i,\text{liq}}^{\text{sat}}}$] associated with the following equation, generally used in the DIPPR database for correlating liquid density data:

$$\rho_{i,\text{liq}}^{\text{sat}}(T) / (\text{kmol m}^{-3}) = \frac{C_1}{C_2 [1 + [1 - (T/K)/C_3]^{C_4}]},$$

for $T \in [T_{\text{min}}^{\rho_{i,\text{liq}}^{\text{sat}}}, T_{\text{max}}^{\rho_{i,\text{liq}}^{\text{sat}}}]$. (41)

Then, by knowing at a given temperature a Henry's-law constant datum and a solvent (ionic liquid) density, Eq. (14) can be used to estimate solvation energy data.

It is worth noting that the density of pure solvent is the only additional property required to generate solvation Gibbs energies from Henry's-law constants. To generate $\Delta_{\text{sol}}\bar{g}_i^\infty(T, P)$ from other mixture properties such as infinite-dilution activity coefficient and partition coefficient data, some other pure-component properties must be known, in particular the vapor pressures of pure ionic liquids which are very low at moderate temperature and therefore very tricky to measure. For this reason, infinite-dilution activity coefficient and partition coefficient data were not considered for generating solvation data for ionic-liquid systems.

4. Global Analysis of the Obtained Results: Presentation of the CompSol Database

4.1. On the behavior of pure-component solvation properties

4.1.1. Description of the CompSol pure-component database

The CompSol databank (provided as the [supplementary material](#)) contains 1969 pure components and reports for each of them solvation chemical potentials, enthalpies, and entropy data at different temperatures. In total, CompSol reports 21 671 values of self-solvation chemical potential data (expressed in kcal mol^{-1} ; $1 \text{ kcal} = 4.184 \text{ kJ}$), 21 671 values of self-solvation entropy data (in $\text{kcal mol}^{-1} \text{ K}^{-1}$), and 21 671 values of self-solvation enthalpy data (in kcal mol^{-1}). For each chemical-potential data point, the database indicates an estimate of the relative error (%) obtained from Eq. (34).

4.1.2. Mutual agreement between the CompSol and MNSol databases

As a validation step, pure-component data contained in the CompSol database have been compared to data extracted from the MNSol database (mentioned in Sec. 1) on the basis of 72 pure compounds common to the two databases. It is recalled that the MNSol database contains exclusively solvation chemical-potential data at 298.15 K. An average deviation of 1.46% between both databases has been obtained, indicating their good reciprocal agreement, as highlighted by the parity plot shown in Fig. 5 (note that only one point, associated with pure pentanol, deviates significantly from the linear trend).

4.1.3. What can be learnt from self-solvation data?

Taking advantage of the number and variety of the compounds contained in the CompSol database, and for each of them, of the number of data at different temperatures as well,

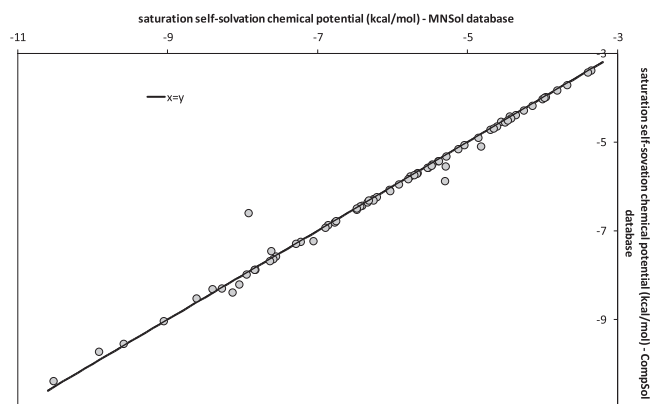


FIG. 5. Parity plot comparing $\Delta_{\text{solv}}g_{\text{pure } i}^{\text{sat}}$ data at 298.15 K contained in the MNSol and CompSol databases, on the basis of 72 pure components.

we examined a complete series of $\Delta_{\text{solv}}g_{\text{pure } i}^{\text{sat}}(T)$ curves generated from CompSol to draw some general observations. Figure 6(a) shows a set of $\Delta_{\text{solv}}g_{\text{pure } i}^{\text{sat}}(T)$ curves related to compounds of various critical compressibility factors (denoted $z_{c,i}$ for species i) and chemical families, considered as well representative of the whole database content. It appears that self-solvation chemical potentials are all negative and increasing functions of temperature. An inflection point seems to systematically arise at an intermediate temperature separating the low-temperature concave behavior from the high-temperature convex behavior.

To better understand the nature of the factors that influence the shapes of the $\Delta_{\text{solv}}g_{\text{pure } i}^{\text{sat}}(T)$ curves, the law of corresponding states can be invoked. According to this theory initially developed by van der Waals, two pure compounds of similar critical compressibility factors exhibit the same (dimensionless) VLE reduced properties at the same reduced temperature $T_{r,i} = T/T_{c,i}$. Note that the critical compressibility factor is sometimes replaced by alternative properties, e.g., the acentric factor. In the present case, the reduced self-solvation chemical potential is defined as $(\Delta_{\text{solv}}g_{\text{pure } i}^{\text{sat}})_r = \Delta_{\text{solv}}g_{\text{pure } i}^{\text{sat}}/(RT_{c,i})$. As highlighted in Fig. 6(b), the $(\Delta_{\text{solv}}g_{\text{pure } i}^{\text{sat}})_r$ versus T_r curves of different pure compounds of similar $z_{c,i}$ values are nearly superimposed, which seems to indicate that $\Delta_{\text{solv}}g_{\text{pure } i}^{\text{sat}}$ follows the

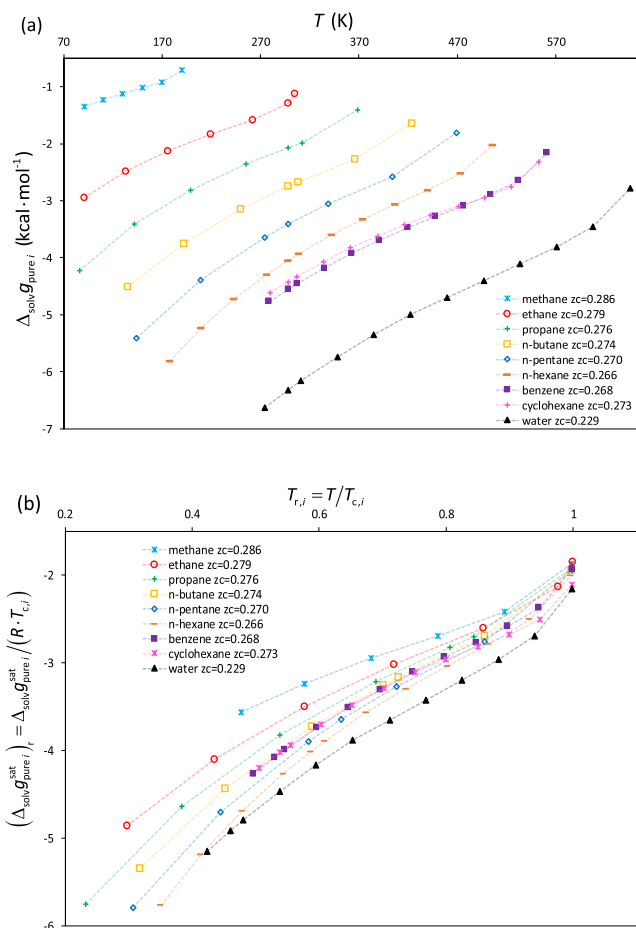


FIG. 6. (a) $\Delta_{\text{solv}}g_{\text{pure } i}^{\text{sat}}$ vs. T and (b) $(\Delta_{\text{solv}}g_{\text{pure } i}^{\text{sat}})_r$ vs. $T_{r,i}$ for different linear alkanes, benzene, cyclohexane, and water.

corresponding-states law. As a consequence, it can be claimed that $(\Delta_{\text{solv}} g_{\text{pure } i}^{\text{sat}})_r$ only depends on the reduced temperature and the critical compressibility factor. Figure 6(b) shows that for a given reduced temperature, the lower the critical compressibility factor, the lower the property $\Delta_{\text{solv}} g_{\text{pure } i}^{\text{sat}}$.

In the same way, Figs. 7(a) and 7(b) show that the reduced solvation entropies $(\Delta_{\text{solv}} s_{\text{pure } i}^{\text{sat}})_r = \Delta_{\text{solv}} s_{\text{pure } i}^{\text{sat}}/R$ and reduced solvation enthalpies $(\Delta_{\text{solv}} h_{\text{pure } i}^{\text{sat}})_r = \Delta_{\text{solv}} h_{\text{pure } i}^{\text{sat}}/(RT_{c,i})$ follow the corresponding-states law but, in addition, are nearly independent of temperature over a large temperature range. Consequently and as confirmed by Fig. 6(a), $\Delta_{\text{solv}} g_{\text{pure } i}^{\text{sat}}(T)$ can be approximated by a linear function in this temperature range,

$$\Delta_{\text{solv}} g_{\text{pure } i}^{\text{sat}}(T) \approx \Delta_{\text{solv}} h_{\text{pure } i}^{\text{sat}}(T_0) - T \cdot \Delta_{\text{solv}} s_{\text{pure } i}^{\text{sat}}(T_0), \quad (42)$$

where T_0 is a reference temperature for the estimation of $\Delta_{\text{solv}} h_{\text{pure } i}^{\text{sat}}(T)$ and $\Delta_{\text{solv}} s_{\text{pure } i}^{\text{sat}}(T)$.

4.2. On the behavior of binary-system solvation properties

4.2.1. Description of the CompSol binary-system database

The CompSol databank (available as the [supplementary material](#)) contains 14 102 binary mixtures involving 865

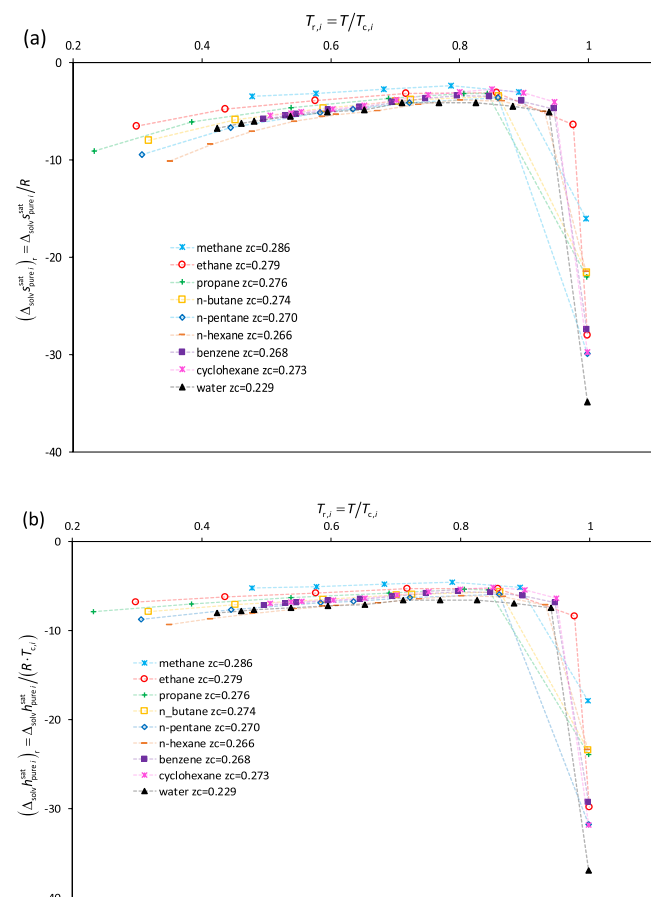


FIG. 7. (a) $(\Delta_{\text{solv}} s_{\text{pure } i}^{\text{sat}})_r = \Delta_{\text{solv}} s_{\text{pure } i}^{\text{sat}}/R$ and (b) $(\Delta_{\text{solv}} h_{\text{pure } i}^{\text{sat}})_r = \Delta_{\text{solv}} h_{\text{pure } i}^{\text{sat}}/(RT_{c,i})$ vs. $T_{r,i} = T/T_{c,i}$ for different linear alkanes, benzene, cyclohexane, and water.

different solutes and 775 different solvents (including ionic liquids). In total, 70 062 values of infinite-dilution solvation chemical potential data (expressed in kcal mol^{-1}) at different temperatures (K) are reported in CompSol as well as 29 344 values of infinite-dilution solvation entropy data (in $\text{kcal mol}^{-1} \text{K}^{-1}$) and 29 344 values of infinite-dilution solvation enthalpy data (in kcal mol^{-1}).

Note that systems involving ionic liquids are reported separately from other systems. For each solvation datum, additional information is also provided:

- (i) The pressure of the experimental data used for the estimation of $\Delta_{\text{solv}} \bar{g}_i^\infty$ is indicated (as explained above, in the many cases where $\Delta_{\text{solv}} \bar{g}_i^\infty$ was deduced from infinite-dilution activity coefficient or Henry's-law constant data, this pressure is either the experimental value or if not known, the saturation pressure of the solvent j at the temperature of interest). It is, however, recalled that due to the incompressible character of liquid phases under low to moderate pressures, solvation energies can be considered as pressure-independent (a proof of this statement is given in Appendix B).
- (ii) The nature of the experimental data used for estimating $\Delta_{\text{solv}} \bar{g}_i^\infty$ is mentioned ("VLE" for VLE solubility data, "ACO" for infinite-dilution activity-coefficient data, and "HLC" for Henry's-law constant data).
- (iii) The complete reference of the paper in which the VLE, ACO, or HLC data were published.
- (iv) When a sufficient number of infinite-dilution solvation chemical potential data exist in a large enough temperature range, infinite-dilution solvation entropy ($\Delta_{\text{solv}} \bar{s}_i^\infty$) and enthalpy ($\Delta_{\text{solv}} \bar{h}_i^\infty$) data were estimated by using Eq. (32).

As mentioned in Sec. 3.2.4, systems containing ionic liquids require special treatment for the estimation of liquid density data. For each ionic liquid contained in the database, coefficients C_1 – C_4 of Eq. (41), their corresponding validity range $[T_{\text{min}}^{\text{sat}}, T_{\text{max}}^{\text{sat}}]$, and the references of the data used for estimating these quantities are specifically reported in the database.

4.2.2. Data consistency

Although many consistency tests are proposed in the literature to analyze the reliability of binary-system VLE data, none of them can be applied to infinite-dilution data. For this reason, it was not possible (except in some rare cases) to analyze the consistency of the solvation data we generated. Note that, in the final CompSol database, some solute + water mixtures that can be considered as reacting systems (especially when the solute is an acid or a base) have been specifically highlighted to draw the reader's attention.

4.2.3. Mutual agreement between the CompSol and MNSol databases

As for pure components, a comparison between the CompSol and MNSol databases has been performed to check their mutual agreement (by considering all the data common to the two

databases). To do so, 601 binary-system experimental $\Delta_{\text{solv}}\bar{g}_i^\infty$ data at 298.15 K were extracted from both databases and compared. The parity plot in Fig. 8 shows a satisfactory agreement between the CompSol and MNSol databanks. An average deviation of 4.9% and an absolute deviation of 0.11 kcal mol⁻¹ have been obtained. Note that an average value of $\Delta_{\text{solv}}\bar{g}_i^\infty$ was calculated when the CompSol database provided different values for the same binary mixture at 298.15 K.

4.2.4. Focus on the generation of infinite-dilution solvation enthalpy and entropy data

As previously stated, it is worth noting that the CompSol databank also includes infinite-dilution solvation entropy ($\Delta_{\text{solv}}\bar{s}_i^\infty$) and enthalpy ($\Delta_{\text{solv}}\bar{h}_i^\infty$) data for 3684 binary mixtures. These properties were only calculated by applying Eq. (32) to binary systems showing a linear $\Delta_{\text{solv}}\bar{g}_i^\infty(T)$ function (other behaviors have been considered as unverifiable or uncertain and have thus been disregarded). When $\Delta_{\text{solv}}\bar{g}_i^\infty(T)$ data describe a linear trend, the quantities $\Delta_{\text{solv}}\bar{g}_i^\infty$, $\Delta_{\text{solv}}\bar{h}_i^\infty$, and $\Delta_{\text{solv}}\bar{s}_i^\infty$ are related through the following equation:

$$\Delta_{\text{solv}}\bar{g}_i^\infty(T) \approx \Delta_{\text{solv}}\bar{h}_i^\infty(T_0) - T \cdot \Delta_{\text{solv}}\bar{s}_i^\infty(T_0) \quad (43)$$

with $T \in [T_{\min}; T_{\max}]$,

where $[T_{\min}; T_{\max}]$ is the temperature range covered by the $\Delta_{\text{solv}}\bar{g}_i^\infty$ data; $\Delta_{\text{solv}}\bar{h}_i^\infty(T_0)$ and $\Delta_{\text{solv}}\bar{s}_i^\infty(T_0)$ are the infinite-dilution solvation enthalpy and entropy, respectively, which are assumed constant on $[T_{\min}; T_{\max}]$. For illustration, four binary systems showing a linear $\Delta_{\text{solv}}\bar{g}_i^\infty(T)$ function are represented in Fig. 9(a); reciprocally, four other binary systems exhibiting non-linear or unclassifiable behaviors are presented in Fig. 9(b). To sum up, for a given binary system, if the series of $\Delta_{\text{solv}}\bar{g}_i^\infty(T)$ data describe a straight line, constant $\Delta_{\text{solv}}\bar{h}_i^\infty$ and $\Delta_{\text{solv}}\bar{s}_i^\infty$ values are reported for all of these data except those deviating significantly from the linear behavior. On the contrary, if the series of $\Delta_{\text{solv}}\bar{g}_i^\infty(T)$ data does not contain enough data to conclude about their linearity or if the data series do not describe a linear behavior, then no $\Delta_{\text{solv}}\bar{h}_i^\infty$ or $\Delta_{\text{solv}}\bar{s}_i^\infty$ values are reported.

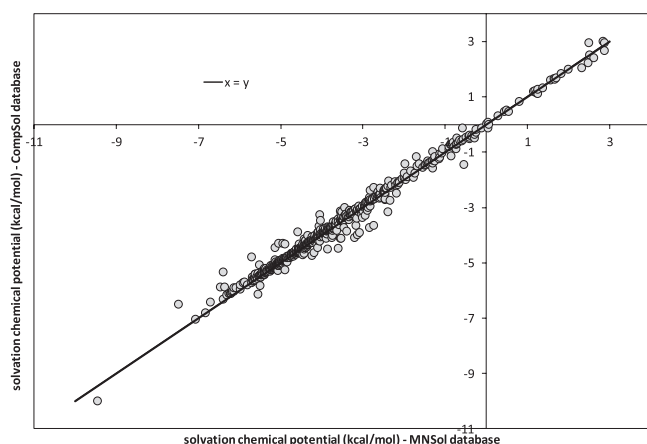


FIG. 8. Parity plot comparing $\Delta_{\text{solv}}\bar{g}_i^\infty$ data at 298.15 K contained in the MNSol and CompSol databases, based on 601 binary systems.

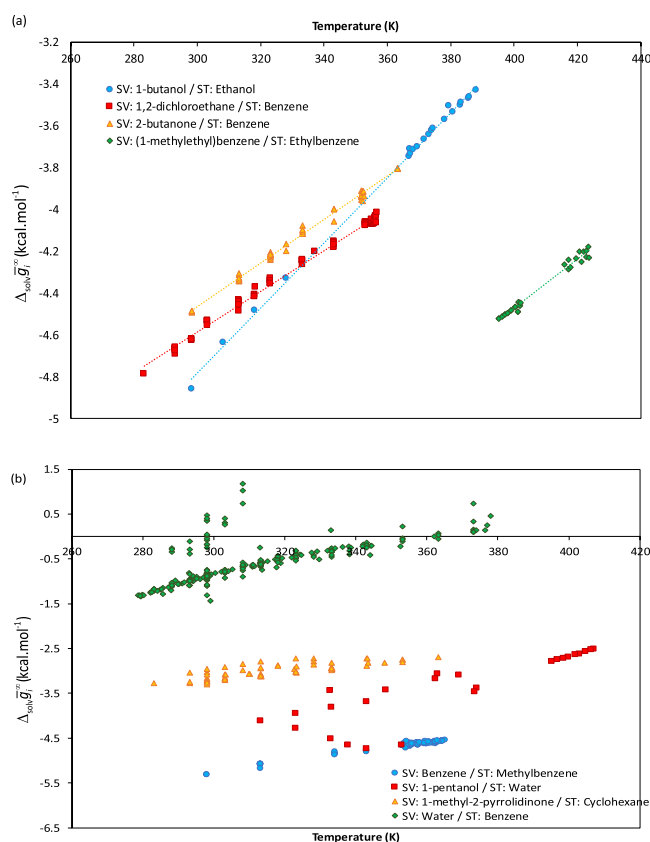


FIG. 9. ST = solute; SV = solvent. (a) Examples of binary systems showing a linear $\Delta_{\text{solv}}\bar{g}_i^\infty(T)$ function. (b) Example of binary systems such that it is not possible to conclude about the linearity of the $\Delta_{\text{solv}}\bar{g}_i^\infty(T)$ function.

4.2.5. Effect of hydrogen bonding on binary-system solvation properties

By using the corresponding-states principle, it has been previously shown that reduced pure-component solvation data values are essentially governed by two factors:

- (i) the reduced temperature;
- (ii) the critical compressibility factor.

When two non-reacting pure compounds i and j are mixed together, self-interactions i - i and j - j between pure species still exist but there are also cross interactions i - j between unlike species. Following the classical solution theory, these latter mainly account for combinatorial (or size, or entropic) effects and residual (or energetic, or enthalpic) effects. Eventually, binary-mixture solvation data under infinite-dilution conditions are influenced by the following:

- (i) the temperature;
- (ii) the molecular sizes and shapes of the two compounds;
- (iii) the i - j and j - j energetic effects (i being infinitely dilute in j , interactions i - i are absent).

How solvation properties are influenced by combinatorial effects (i.e., change in solvation property induced by differences of size and shape between the two components of a binary system) is an issue frequently addressed in literature studies, e.g., these effects are typically discussed when

comparing the solvation of a given solute in linear alkanes and in branched alkanes.^{7,8} Working on the quantification of such effects, some authors have highlighted that solvation is favored with solvents of high Bondi's surface area parameters.⁶⁻⁸

Energetic effects are revealed by the presence of intermolecular weak bonds due to, e.g., van der Waals forces or association through hydrogen bonding. Because of their high intensity, hydrogen bonds may significantly influence solution properties. As they may favor or perturb the solvation process, solvation properties are specifically affected by these effects.

The variety of chemicals present in the CompSol database, and for each of them the number of solvation data, makes it possible to evidence relationships between chemical structures and solvation energies. Consequently, this allows us to validate or disconfirm intuitions on solvation effects. In the present case, it is proposed to discuss how energetic effects induced by association phenomena (i.e., ability of a solute and/or a solvent to create hydrogen bonds) may influence the solvation of a solute in a solvent.

Following the spirit of the classification proposed by Danner and Gess⁵¹ in the early 1990s, it has been decided to order pure components according to their ability to generate association phenomena in solution (i.e., ability of a solute and/or a solvent to create hydrogen bonds). To observe chemical association between two molecules, one of them must possess a labile hydrogen atom (a hydrogen atom linked to a hetero-atom conferring it a high tendency to switch on a H⁺ form), as in carboxylic acid or water, while the second one must show a lone pair of electrons able to link with a labile hydrogen, as in ammonia or water.

Starting from this basic analysis, four classes of pure components can be identified (numbered with pure association codes 1–4) and are presented in Table 3.

A classification of binary systems can be immediately deduced by simple combination of pure-component association codes. Because there are four possible pure-component classes for a solute and four possible pure-component classes for a solvent, there are thus 16 possible combinations for the corresponding binary system, as shown in Fig. 10.

By gathering binary-system classes leading to the same kind of association phenomena in solution, the 16 classes can be reduced to seven classes that are as follows:

- (i) NA-NA: both pure compounds are non-associating and cannot create hydrogen bonds.
- (ii) SA-NA: self-associating solute mixed with a non-associating solvent. Hydrogen bonds are only possible between two molecules of solute.

		Solute			
		1	2	3	4
Solvent	1	1-1 NA-NA	1-2 NA-NA	1-3 NA-NA	1-4 SA-NA
	2	2-1 NA-NA	2-2 NA-NA	2-3 E/H-H/E	2-4 SA-E/H
	3	3-1 NA-NA	3-2 E/H-H/E	3-3 NA-NA	3-4 SA-E/H
	4	4-1 NA-SA	4-2 E/H-SA	4-3 E/H-SA	4-4 SA-SA

FIG. 10. Classification of binary systems into 16 categories deduced from pure-component association codes presented in Table 3 (1-1, 1-2, 1-3, ...). Classification of binary systems into seven categories called "association classes" (NA-NA, SA-NA, SA-E/H, ...) according to their ability to generate hydrogen bonding in solution.

- (iii) NA-SA: self-associating solvent mixed with a non-associating solute. Hydrogen bonds are only possible between two molecules of solvent.
- (iv) E/H-H/E: the solute presents a lone pair of electrons or a labile hydrogen while the solvent presents a labile hydrogen or a lone pair of electrons, respectively. Only cross association between unlike molecules is possible.
- (v) SA-E/H: self-associating solute mixed with a solvent showing either a lone pair of electrons or labile hydrogen. Hydrogen bonds are possible either between two molecules of solute (self-association) or between two unlike molecules (cross association).
- (vi) E/H-SA: self-associating solvent mixed with a solute showing either a lone pair of electrons or labile hydrogen. Hydrogen bonds are possible either between two molecules of solvent (self-association) or between two unlike molecules (cross association).
- (vii) SA-SA: self-associating solute mixed with a self-associating solvent. Hydrogen bonds are possible either between two molecules of solute (self-association), two molecules of solvent (self-association), or two unlike molecules (cross association).

Figure 11 shows how 6409 infinite-dilution solvation energy data estimated at a reference temperature of 298.15 K [$\Delta_{\text{solv}}\bar{g}_i^\infty(298.15 \text{ K})$] fall into the seven association classes of binary systems defined above. For a given binary-system class, binary systems are distributed into eight ranges of $\Delta_{\text{solv}}\bar{g}_i^\infty(298.15 \text{ K})$ values (the darkest color is associated with highly negative solvation-energy values while the lightest color

TABLE 3. Description of the four association codes for pure components

Pure-component association code (abbreviation)	Description	Example
1 (NA for "non associating")	Molecules do not present any association site (neither labile hydrogen nor lone pair of electrons)	Alkanes, non-substituted aromatics
2 (H)	Molecules present a labile hydrogen atom only	Organochlorides, organobromides
3 (E)	Molecules present a lone pair of electrons only	Ketones, amides, esters
4 (SA for "self-associating")	Molecules present both a labile hydrogen atom and a lone pair of electrons and give birth to self-association (SA)	Alcohols, water, amines, carboxylic acids

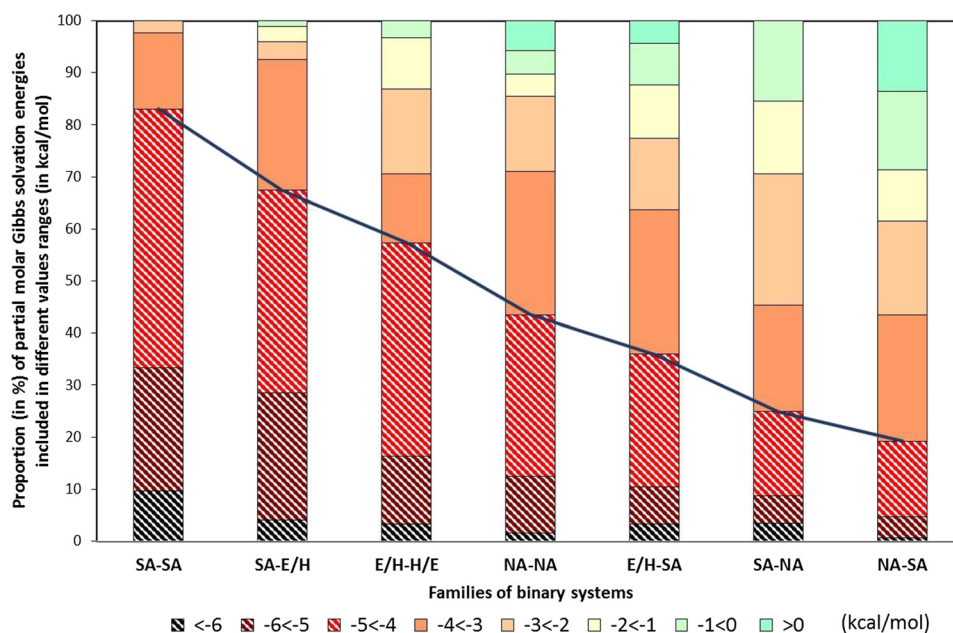


Fig. 11. Distribution of $\Delta_{\text{solv}}\bar{g}_i^\infty$ values at 298.15 K following the seven classes of binary systems defined in Fig. 10.

indicates positive values); the area associated with a given color is proportional to the ratio of binary systems falling into the corresponding range of solvation energies.

It is recalled that the more $\Delta_{\text{solv}}\bar{g}_i^\infty$ is negative, the more solute i is stable in the solute + solvent liquid mixture with respect to a hypothetical perfect-gas mixture and consequently solute i is considered as well solvated by its solvent as $\Delta_{\text{solv}}\bar{g}_i^\infty$ is low. Figure 11 shows that the seven binary-system association-classes can be ordered following the crosshatched area, thus revealing relationships between association phenomena and solvation.

- (i) The NA-NA class, which is characterized by the absence of hydrogen bonds in solution, is found exactly in the middle of the seven classes (see Fig. 11) and can be considered as reference behavior in the sense that association does not play any role in the solvation process (neither contributing nor perturbing). When no association effect is present in the mixture (no self-association and no cross association), only temperature, size, and weak-interaction (e.g., van der Waals forces) effects determine $\Delta_{\text{solv}}\bar{g}_i^\infty$ values. Note that around 50% of the considered experimental data fall into the NA-NA association class.
- (ii) Binary classes located on the left-hand side of NA-NA systems (in Fig. 11) exhibit more negative $\Delta_{\text{solv}}\bar{g}_i^\infty$ values, and consequently it can be claimed that, in these cases, association phenomena have a positive effect on the solvation process. In the SA-E/H class, although the solute has the capacity to self-associate, the probability for two solute molecules to meet tends to zero as solutes are infinitely diluted in their solvents. Consequently, cross association dominates. Therefore, in the SA-E/H and E/H-H/E classes of systems, it can be claimed that only cross association interactions take place, leading to higher stability of the solute-solvent pair. SA-SA type

- systems show the more favorable distribution of solvation energies to the solvation process. It must be concluded that in the families of chemical compounds likely to self-associate, cross associations i - j are generally more stable than solvent-solvent (j - j) associations.
- (iii) On the right-hand side of the NA-NA class of systems (in Fig. 11), association phenomena tend to hinder the solvation process. In the E/H-SA and NA-SA binary-system classes, the solvent self-associates, thus generating aggregates not conducive to the insertion of a solute molecule. In addition, solute and solvent molecules involved in binary systems of the NA-SA or SA-NA classes are characterized by the absence of stabilizing cross associations and are generally very asymmetric in terms of physical properties (e.g., polarity) and chemical functions, thus contributing to their low reciprocal affinity. The SA-NA class is found to be the less favorable one to the solvation process.

5. Conclusion

By starting from a rigorous definition of the solvation process, exact relations between thermodynamic properties accessible from experiments and solvation properties have been evidenced. As a second step, pure-component and binary-system thermodynamic-property data have been extracted from comprehensive commercial databases. By applying the evidenced relations to the experimental property data, experimental solvation data have been generated, thus giving birth to the CompSol database (given in the [supplementary material](#)). CompSol stands for “Comprehensive Solvation” (database), as we endeavored to consider the largest possible sets of thermodynamic data to generate this database: CompSol contains 21 671 data associated with 1969 pure species and 70 062 data

associated with 14 102 binary mixtures. Beyond the number of the data and the variety of pure compounds and binary systems considered, the two key strengths of this database are certainly the presence of solvation data at different temperatures and of solvation enthalpy and entropy data.

The CompSol database offers today the advantage to be the most extensive databank of pure and mixed solvation-energy data and could be used in many respects, e.g., search for an appropriate solvent for a given solute or mixture of solutes, development of solvation models, qualitative or quantitative studies of the factors influencing the solvation process.

What now remains to be done to complete our work? Regarding binary-system data, we only focused on solvation properties under infinite-dilution conditions, which are certainly the most useful ones when one wants to compare affinities of solutes for solvents. However, for some other applications, e.g., estimation of kinetic rate constants of reactions taking place in liquid phases or development of solvation models, solvation data at higher concentrations \mathbf{x} of solute could be very useful. To reach this goal, estimates of activity coefficients and liquid-mixture densities are necessary, as highlighted by Eq. (44),

$$\Delta_{\text{solv}}\bar{g}_i(T, P, \mathbf{x}) = RT \ln \left[\frac{P}{RT \cdot \rho_{\text{liq}}(T, P, \mathbf{x})} \cdot K_i(T, P) \cdot \frac{\gamma_{i,\text{liq}}^{(x)}(T, P, \mathbf{x})}{\gamma_{i,\text{liq}}^{(x)}(T, P, \mathbf{x}^{\text{VLE}})} \right], \quad (44)$$

which can be applied regardless of the solute composition domain considered [provided VLE exists at the given (T, P) and the gas phase can be considered as perfect]; \mathbf{x}^{VLE} denotes the composition of the liquid phase involved in the VLE. The estimation of activity coefficients at any composition is difficult in the sense that it requires the use of an appropriate activity-coefficient model (also called *excess Gibbs energy model*). However, it should be noted that the estimation of solvation chemical potentials at a bubble-point condition becomes easier, since in such a case one has $\mathbf{x} = \mathbf{x}^{\text{VLE}}$ and consequently $\gamma_{i,\text{liq}}^{(x)}(T, P, \mathbf{x}) = \gamma_{i,\text{liq}}^{(x)}(T, P, \mathbf{x}^{\text{VLE}})$. Doing so, Eq. (44) is

$$\Delta_{\text{solv}}\bar{g}_i(T, P, \mathbf{x} = \mathbf{x}^{\text{VLE}}) = RT \ln \left[\frac{P \cdot K_i(T, P)}{RT} \times \frac{x_i}{\rho_{\text{pure } i,\text{liq}}(T, P)} + \frac{x_j}{\rho_{\text{pure } j,\text{liq}}(T, P)} + v_{\text{liq}}^{\text{E}}(T, P, \mathbf{x}) \right], \quad (45)$$

where $v_{\text{liq}}^{\text{E}}$ denotes the excess molar volume of the system. By neglecting this quantity (which is a reasonable assumption for many binary systems), $\Delta_{\text{solv}}\bar{g}_i(T, P, \mathbf{x} = \mathbf{x}^{\text{VLE}})$ could be estimated for any composition \mathbf{x}^{VLE} . Eventually, any value of K_i makes it possible to generate immediately a $\Delta_{\text{solv}}\bar{g}_i$ datum. By considering the very extensive number of VLE data present in the literature on one hand, and the simplicity of the relation between K_i and $\Delta_{\text{solv}}\bar{g}_i$ on the other hand, we think that developing such a database would be both fastidious and useless.

Another possible extension of our work could be the use of calorimetric data to generate infinite-dilution solvation enthalpies. It can indeed be shown that under the assumption that the gas phase is perfect and that the liquid phase is incompressible, one has⁶

$$\Delta_{\text{solv}}\bar{h}_i^{\infty}(T) \approx \bar{h}_i^{\text{E},\infty}(T) - \Delta_{\text{vap}}h_{\text{pure } i}(T), \quad (46)$$

where $\bar{h}_i^{\text{E},\infty}$ is the partial molar excess enthalpy of solute i at infinite dilution (estimable from calorimetric data) and $\Delta_{\text{vap}}h_{\text{pure } i}$ is its molar enthalpy of vaporization (e.g., available in the DIPPR databank).

Supplementary Material

See [supplementary material](#) at [[EPAPS]] for the complete CompSol Database.

Acknowledgments

The authors would like to thank Dr. Andreas Klamt (COSMOlogic GmbH&CoKG & Inst. of Phys. and Theor. Chemistry, Regensburg) warmly for helpful discussions and suggestions regarding the generation of solvation Gibbs energy data.

6. Appendix A: On the Pressure Independence of the Pseudo-chemical Potential of a Molecule in a Perfect-gas Mixture

Classically, the chemical potential of a molecule i in a perfect-gas mixture at fixed T, P and of molar fraction y_i is expressed as

$$\bar{g}_i^{\circ}(T, P, y_i) = g_i^{\circ}(T, P) + RT \ln y_i. \quad (A1)$$

The molar fraction can be related to the molar concentration through

$$y_i = c_i \cdot \frac{RT}{P}. \quad (A2)$$

By incorporating Eq. (A2) to Eq. (A1), one has

$$\bar{g}_i^{\circ}(T, P, c_i) = g_i^{\circ}(T, P) + RT \ln \left(c_i \cdot \frac{RT}{P} \right). \quad (A3)$$

By combining Eqs. (2) and (A3), one obtains

$$\bar{g}_i^{\circ,\bullet}(T, P) = g_i^{\circ}(T, P) + RT \ln \left[\frac{RT}{P \cdot N_A \cdot \Lambda_i^3(T)} \right]. \quad (A4)$$

It can be observed that the pseudo-chemical potential of a molecule i in a perfect gas mixture depends on the temperature and the pressure but not on the mixture composition.

7. Appendix B: On the Pressure Independence of the Solvation Chemical Potential

Equation (6) relates the solvation chemical potential of species i in a liquid phase [$\Delta_{\text{solv}}\bar{g}_i(T, P, \mathbf{x})$] to the product $P \cdot \varphi_{i,\text{liq}}(T, P, \mathbf{x})$ (fugacity coefficient of i times pressure) and to the liquid-phase density $\rho_{\text{liq}}(T, P, \mathbf{x})$.

It is well known that the density of a liquid phase depends only slightly on the pressure (because of the incompressible character of a liquid phase under moderate pressure). The same conclusion can be reached for $P \cdot \varphi_{i,\text{liq}}(T, P, \mathbf{x})$. Here is a short proof: the fugacity coefficient of i is related to the activity coefficient of i through the relation: $\varphi_{i,\text{liq}}(T, P, \mathbf{x}) = \gamma_{i,\text{liq}}^{(x)}(T, P, \mathbf{x}) \cdot \varphi_{\text{pure } i,\text{liq}}(T, P)$. The product $P \cdot \varphi_{\text{pure } i,\text{liq}}(T, P)$ defines the *fugacity* of pure liquid i at T ; under moderate pressure, this quantity can be approximated by the vapor pressure of pure i at T (this is a classical assumption). Consequently, one can claim that under moderate pressure,

$$\begin{aligned} \Delta_{\text{solv}}\bar{g}_i(T, P, \mathbf{x}) &= RT \ln \left[\frac{P \cdot \varphi_{i,\text{liq}}(T, P, \mathbf{x})}{RT \cdot \rho_{\text{liq}}(T, P, \mathbf{x})} \right] \\ &\approx RT \ln \left[\frac{\gamma_{i,\text{liq}}^{(x)}(T, P, \mathbf{x}) \cdot P_i^{\text{sat}}(T)}{RT \cdot \rho_{\text{liq}}(T, P, \mathbf{x})} \right]. \end{aligned} \quad (\text{B1})$$

By considering the activity coefficient $\gamma_{i,\text{liq}}^{(x)}(T, P, \mathbf{x})$ as pressure-independent (once again, because of the incompressible character of the liquid phase, this is a classical assumption), it appears that $\Delta_{\text{solv}}\bar{g}_i(T, P, \mathbf{x})$ can be considered independent of pressure.

8. References

- ¹ *Encyclopaedia Britannica*, edited by M.J. Adler, W. Benton, and C. E. Swanson, 15th ed. (Encyclopædia Britannica, Inc., Chicago, 1974).
- ² P. Muller, *Pure Appl. Chem.* **66**, 1164 (1994).
- ³ V. I. Minkin, *Pure Appl. Chem.* **71**, 1919 (1999).
- ⁴ A. Ben-Naim, *J. Phys. Chem.* **82**, 792 (1978).
- ⁵ A. Ben-Naim, *Solvation Thermodynamics* (Springer US, Boston, MA, 1987).
- ⁶ L. Bernazzani, M. R. Carosi, N. Ceccanti, G. Conti, P. Gianni, V. Mollica, M. R. Tinè, L. Lepori, and E. Matteoli, *Phys. Chem. Chem. Phys.* **2**, 4829 (2000).
- ⁷ L. Lepori, P. Gianni, and E. Matteoli, *J. Therm. Anal. Calorim.* **124**, 1497 (2016).
- ⁸ L. Lepori, E. Matteoli, P. Gianni, and M. C. Righetti, *Fluid Phase Equilib.* **387**, 198 (2015).
- ⁹ I. A. Sedov, M. A. Stolov, and B. N. Solomonov, *J. Phys. Org. Chem.* **24**, 1088 (2011).
- ¹⁰ I. A. Sedov, M. A. Stolov, and B. N. Solomonov, *Fluid Phase Equilib.* **354**, 95 (2013).
- ¹¹ I. A. Sedov, M. A. Stolov, and B. N. Solomonov, *Russ. J. Phys. Chem. A* **85**, 621 (2011).
- ¹² I. A. Sedov, T. I. Magsumov, M. A. Stolov, and B. N. Solomonov, *Thermochim. Acta* **623**, 9 (2016).
- ¹³ N. Matubayasi, *Curr. Opin. Struct. Biol.* **43**, 45 (2017).
- ¹⁴ S. A. Martins, S. F. Sousa, M. J. Ramos, and P. A. Fernandes, *J. Chem. Theory Comput.* **10**, 3570 (2014).
- ¹⁵ B. Dominy, *Curr. Pharm. Biotechnol.* **9**, 87 (2008).
- ¹⁶ K. Raha and K. M. Merz, *Annu. Rep. Comput. Chem.* **1**, 113 (2005).
- ¹⁷ M. R. Shirts, J. W. Pitera, W. C. Swope, and V. S. Pande, *J. Chem. Phys.* **119**, 5740 (2003).
- ¹⁸ W. L. Jorgensen, *Science* **303**, 1813 (2004).
- ¹⁹ E. L. Ratkova, D. S. Palmer, and M. V. Fedorov, *Chem. Rev.* **115**, 6312 (2015).
- ²⁰ *Free Energy Calculations in Rational Drug Design*, edited by M. Rami Reddy and M. D. Erion (Kluwer Academic/Plenum Publishers, New York, 2001).
- ²¹ P. S. Redmill, S. L. Capps, P. T. Cummings, and C. McCabe, *Carbon* **47**, 2865 (2009).
- ²² C. West, J. Ogden, and E. Lesellier, *J. Chromatogr. A* **1216**, 5600 (2009).
- ²³ M. H. Abraham, C. F. Poole, and S. K. Poole, *J. Chromatogr. A* **842**, 79 (1999).
- ²⁴ J. V. Seeley, E. M. Libby, K. A. H. Edwards, and S. K. Seeley, *J. Chromatogr. A* **1216**, 1650 (2009).
- ²⁵ J. R. Pliego, Jr., *J. Braz. Chem. Soc.* **26**, 1737 (2015).
- ²⁶ C. Hsieh and S. Lin, *AIChE J.* **54**, 2174 (2008).
- ²⁷ C.-M. Hsieh and S.-T. Lin, *Ind. Eng. Chem. Res.* **48**, 3197 (2009).
- ²⁸ S.-T. Lin, *Fluid Phase Equilib.* **245**, 185 (2006).
- ²⁹ A. Jalan, R. W. Ashcraft, R. H. West, and W. H. Green, *Annu. Rep., Sect. C: Phys. Chem.* **106**, 211 (2010).
- ³⁰ A. Jalan, R. H. West, and W. H. Green, *J. Phys. Chem. B* **117**, 2955 (2013).
- ³¹ M. H. Abraham, G. S. Whiting, R. Fuchs, and E. J. Chambers, *J. Chem. Soc., Perkin Trans. 2*, 291 (1990).
- ³² A. V. Marenich, C. P. Kelly, J. D. Thompson, G. D. Hawkins, C. C. Chambers, D. J. Giesen, P. Winget, C. J. Cramer, and D. G. Truhlar, *Minnesota Solvation Database—Version 2012*, University of Minnesota, Minneapolis, 2012.
- ³³ A. C. Chamberlin, C. J. Cramer, and D. G. Truhlar, *J. Phys. Chem. B* **110**, 5665 (2006).
- ³⁴ D. L. Mobley and J. P. Guthrie, *J. Comput.-Aided Mol. Des.* **28**, 711 (2014).
- ³⁵ J.-C. Bradley, M. H. Abraham, W. E. Acree, and A. Lang, *Chem. Cent. J.* **9**, 12 (2015).
- ³⁶ M. H. Abraham, A. Ibrahim, and A. M. Zissimos, *J. Chromatogr. A* **1037**, 29 (2004).
- ³⁷ J. Zhang, B. Tuguldur, and D. van der Spoel, *J. Chem. Inf. Model.* **55**, 1192 (2015).
- ³⁸ T. P. Straatsma, H. J. C. Berendsen, and J. P. M. Postma, *J. Chem. Phys.* **85**, 6720 (1986).
- ³⁹ B. Wang, Z. Zhao, and G.-W. Wei, *J. Chem. Phys.* **145**, 124110 (2016).
- ⁴⁰ J. D. Thompson, C. J. Cramer, and D. G. Truhlar, *J. Chem. Phys.* **119**, 1661 (2003).
- ⁴¹ J. Wang, W. Wang, S. Huo, M. Lee, and P. A. Kollman, *J. Phys. Chem. B* **105**, 5055 (2001).
- ⁴² T. Sulea, D. Wanapun, S. Dennis, and E. O. Purisima, *J. Phys. Chem. B* **113**, 4511 (2009).
- ⁴³ A. Nicholls, S. Wlodek, and J. A. Grant, *J. Phys. Chem. B* **113**, 4521 (2009).
- ⁴⁴ A. Klamt and M. Diedenhofen, *J. Phys. Chem. A* **119**, 5439 (2015).
- ⁴⁵ A. Klamt, *J. Phys. Chem.* **99**, 2224 (1995).
- ⁴⁶ A. Klamt, V. Jonas, T. Bürger, and J. C. W. Lohrenz, *J. Phys. Chem. A* **102**, 5074 (1998).
- ⁴⁷ J. P. Guthrie, *J. Phys. Chem. B* **113**, 4501 (2009).
- ⁴⁸ R. Privat, J.-N. Jaubert, and E. Moine, *J. Chem. Educ.* **93**, 2040 (2016).
- ⁴⁹ R. Rowley, W. Wilding, J. Oscarson, Y. Yang, and N. Giles, *DIPPR Data Compilation of Pure Chemical Properties* (Design Institute for Physical Properties, AIChE, New York, 2010).
- ⁵⁰ O. A. Hougen, K. M. Watson, and R. A. Ragatz, *Chemical Process Principles. 2: Thermodynamics*, 2nd ed., [NACH-DR] (Wiley, New York, 1959).
- ⁵¹ R. P. Danner and M. A. Gess, *Fluid Phase Equilib.* **56**, 285 (1990).