Chemical Engineering Science 309 (2025) 121425

Contents lists available at ScienceDirect

Chemical Engineering Science

journal homepage: www.elsevier.com/locate/ces

A comprehensive approach to incorporating intermolecular dispersion into the openCOSMO-RS model. Part 1. Halocarbons

Daria Grigorash [a,b], Simon Müller [c], Patrice Paricaud [d], Erling H. Stenby [a,b], Irina Smirnova [c], Wei Yan [a,b,*]

[a] Department of Chemistry, Technical University of Denmark, Kgs. Lyngby, 2800, Denmark
[b] Center for Energy Resources Engineering, Technical University of Denmark, Kgs. Lyngby, 2800, Denmark
[c] Institute of Thermal Separation Processes, Hamburg University of Technology, Hamburg, 21073, Germany
[d] UCP, ENSTA Paris Institut Polytechnique de Paris, Palaiseau, 91762, France

ARTICLE INFO

Keywords:
COSMO-RS
Dispersion
Parameterization
Refrigerants

ABSTRACT

The COSMO-RS (Conductor-like Screening Model for Real Solvents) is a predictive thermodynamic model that has found diverse applications in various domains like chemical engineering, environmental chemistry, nanotechnology, material science, and biotechnology. Its core concept involves calculating the screening charge density on the surface of each molecule and letting these surface patches interact with each other to calculate thermodynamic properties. In this study, we aim to enhance the performance of the open-source implementation openCOSMO-RS by incorporating dispersive interactions between the paired segments. Several parametrizations were systematically evaluated through the extensive regression analysis using a comprehensive database of Vapor-Liquid Equilibrium (VLE), Liquid-Liquid Equilibrium (LLE) and Infinite Dilution Activity Coefficients (IDACs). Furthermore, the influence of different combinatorial terms on the model performance was investigated. Our findings indicate that incorporating dispersive interactions significantly improves the accuracy of phase equilibrium predictions for halocarbons and refrigerant mixtures.

1. Introduction

Over the last two decades, the developments in computational chemistry methods have laid the foundation for a new class of predictive thermodynamic models originally based on the COSMO (COnductor-like Screening Model) (Klamt and Schüürmann, 1993) quantum chemical (QC) method. It is one of the dielectric continuum solvation models that is used to determine the dielectric screening effects on molecules in solvents. The pioneering work by Klamt (1995) on integrating COSMO into a thermodynamic model created COSMO-RS. Since then, several implementations based on the COSMO-RS model, including COSMO-SAC (Lin and Sandler, 2002), COSMO-RS(ol) (Grensemann and Gmehling, 2005) and JCOSMO Soares (2023) have been proposed. Additionally, the concept of segment-segment interactions appeared in the group-contribution method F-SAC (Soares and Gerber, 2013). However, in F-SAC, regression of experimental thermodynamic data replaced QC calculations.

COSMO-RS is a predictive model that allows the estimation of thermodynamic properties even when no experimental data is available. Since its creation, COSMO-RS and its implementations have been assessed and applied for the prediction of solvation Gibbs free energies (Saidi et al., 2020), VLE and IDACs across diverse chemical families of compounds (Fingerhut et al., 2017), and various thermodynamic properties in systems such as ionic liquids (Jiřiště and Klajmon, 2022; Lee and Lin, 2017; Diedenhofen and Klamt, 2010), pharmaceuticals (Klajmon, 2022; Rodríguez-Llorente et al., 2023; Loschen and Klamt, 2016; Klamt et al., 2002), polymer solutions (Silva et al., 2023; Loschen and Klamt, 2014), electrolytes (Müller et al., 2019; Kröger et al., 2020; Müller et al., 2020; González de Castilla et al., 2021; Ritter et al., 2016; Arrad et al., 2024; González de Castilla et al., 2022) and refrigerants (Mambo-Lomba and Paricaud, 2021). This study focuses on refrigerant systems, which are vital for industrial processes, especially in optimizing refrigerant production and operating energy systems like heat pumps. Understanding the phase behavior of refrigerant blends, including azeotropes and VLE, is critical for selecting their optimal combinations.

Within the COSMO-RS framework, molecules are fragmented into interacting segments, primarily considering electrostatic interactions and hydrogen bonding. These interactions are characterized by a set of descriptors or properties of these segments, with the surface charge density being the most important among them. In the early versions of

* Corresponding author at: Department of Chemistry, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark.
E-mail address: weya@kemi.dtu.dk (W. Yan).

https://doi.org/10.1016/j.ces.2025.121425
Received 4 June 2024; Received in revised form 19 September 2024; Accepted 23 February 2025
Available online 2 March 2025

COSMO-RS, dispersive intermolecular interactions were not accounted for in fluid mixtures. These interactions are caused by attractive forces between instantaneous dipoles (Prausnitz et al., 1999). Since these dipoles are not permanent, their interactions could not be captured with classic electrostatics (Stone, 2013). Nevertheless, dispersive interactions have already been considered in several implementations of COSMO-RS. For instance, COSMO-SAC-dsp by Hsieh et al. (2014) introduces a one-constant Margules term to activity coefficients. The constant is determined from atomic dispersion parameters which are general for the model. As previously mentioned by Krooshof et al. (2019b), this treatment greatly simplifies the dispersion phenomenon, since using the Margules model implies that dispersion comes only from similar-sized interactions. Moreover, in mixtures of species that belong to one aliphatic hydrocarbon homologous series, dispersion is neglected, since the contribution of hydrogen (H-atom) in such compounds equals to zero. Therefore, other equations for the dispersive contribution to the activity coefficient were proposed in works by Krooshof et al. (2019a,b). However, to our knowledge, none of those is implemented in the open-source software. Mambo-Lomba and Paricaud (2021) modified the atomic dispersion parameter of fluorine (F-atom) in the COSMO-SAC-dsp model, which significantly improved VLE predictions of the studied refrigerants. For high-pressure systems, the MHV1 EoS/G^E mixing rule approach was used to combine the modified COSMO-SAC-dsp with the Peng-Robinson equation of state. In the F-SAC model, dispersion is addressed via general group interaction parameters (Flôres et al., 2016). The group–group interactions in F-SAC are analogous to segment–segment interactions in COSMO-RS. Therefore, the integration of dispersive interactions in F-SAC is conceptually different from COSMO-SAC-dsp. Furthermore, in the recent versions of COSMOtherm software (BIOVIA, 2020), a dispersion term is also introduced for fluid mixtures, using the matrix of pairwise element-element specific van der Waals parameters (Klamt and Eckert, 2007).

The major challenge in modeling the phase equilibrium of halocarbons as part of refrigerants is the substantial difference in polarizabilities of halogens, e.g., fluorine having an extremely low polarizability in contrast to iodine having a very high one. Therefore, there is a clear need for special consideration of the dispersive contribution regarding this type of systems.

In the present study, we aim to develop the dispersion contribution to the segment-segment interactions, currently absent in the open-source implementation of COSMO-RS: openCOSMO-RS by Gerlach et al. (2022). To outline the main concepts of COSMO-RS, its major equations are presented in the Methods 2 section: residual contribution, including electrostatic misfit and hydrogen bonding in Section 2.1.1, the newly-implemented dispersive contribution in Section 2.1.2, and various combinatorial terms in Section 2.1.3. The workflow used for the QC calculations and regression procedure are explained in Sections 2.2 and 2.3, respectively. The database introduced in Section 3.1 was used for the regressions discussed in Section 3.2. Subsequently, we evaluated the impact of the combinatorial term on the IDAC of hydrocarbons in Section 3.3. Finally, in Section 3.4, we examined the VLE and LLE predictions by the modified models and compared them with existing open-source COSMO-RS implementations.

2. Methods

2.1. COSMO-RS

As most of the existing activity coefficient models, COSMO-RS includes both residual (Sections 2.1.1, 2.1.2 and combinatorial (Section 2.1.3) terms:

$$\ln(\gamma_i) = \ln(\gamma_i^{\text{res}}) + \ln(\gamma_i^{\text{comb}}). \tag{1}$$

2.1.1. Residual contribution: electrostatic misfit and hydrogen bonding

The detailed equations used in openCOSMO-RS are reported in Klamt et al. (1998), Klamt and Eckert (2000) and Gerlach et al. (2022). Following the COSMO-RS approach, molecules are represented by a set of surface patches or segments, each characterized by a group of descriptors. One of the primary descriptors is the surface charge density (σ). It is determined using the COSMO or similar C-PCM (Conductor-like Polarizable Continuum Model) QC method. Continuum solvation models represent each solvent as a dielectric continuum with a dielectric permittivity ε. The solute molecule is then embedded into this dielectric continuum, which, in the case of COSMO, is modeled as a perfect conductor with $\varepsilon = \infty$ (Klamt, 2005a). In response, screening charge is generated on the surface of the molecular cavity, constructed using element-specific radii. While σ serves as a major descriptor, the openCOSMO-RS implementation used in this study also incorporated additional descriptors and simplified the procedure for integrating new ones into the model if needed.

In COSMO-RS, thermodynamic properties are related to the interactions between molecular segments:

$$\mu_s(\sigma_i) = -kT \ln\left(\int d\sigma_j p_s(\sigma_j) \exp\left[(-E_{ij} + \mu_s(\sigma_j))/kT\right]\right), \tag{2}$$

where σ_i and σ_j, $\mu_s(\sigma_i)$ and $\mu_s(\sigma_j)$ are the surface screening charge densities and chemical potentials of segments i and j, respectively. k is Boltzmann's constant, and T is the temperature. $p_s(\sigma_j)$ or the σ-profile defines the probability of finding the segment with charge density σ_j on the surface of the molecule. It is formed by clustering segments onto a predetermined grid of selected values of σ_j and projecting them onto a histogram. Before clustering segments into the σ-profile, they are averaged over a region of radius r_{av}. The areas of the segments are an output from the currently used COSMO/C-PCM implementation of the underlying QC package. However, these areas might significantly differ from each other. Therefore, averaging over a region of radius r_{av} is performed to spread the charge and to reduce numerical noise. Finally, E_{ij} represents the interaction free energy between segments i and j.

In the early versions of COSMO-RS (Klamt et al., 1998; Klamt and Eckert, 2000), two contributions to the interaction free energy were considered for modeling the condensed phase equilibrium: the repulsive misfit free energy $E_{\text{mf}}(\sigma_i, \sigma_j)$ and the attractive hydrogen bonding energy $E_{\text{hb}}(\sigma_i, \sigma_j)$. The former captures electrostatic interactions between the segments, while the latter represents an additional attractive contribution arising from hydrogen bond formation. The electrostatic misfit free energy is calculated as:

$$E_{\text{mf}}(\sigma_i, \sigma_j) = 0.5 a_{\text{eff}} \alpha_{\text{mf}} [(\sigma_i + \sigma_j)^2 + f_{\text{corr}}(\sigma_i + \sigma_j)(\sigma_i^{\perp} + \sigma_j^{\perp})], \tag{3}$$

involving both the screening charge density σ and the correlation screening charge density $\sigma^{\perp}$ to account for the effect of the segment surroundings. f_{corr} is the correlation correction factor adjusted to the dielectric energy data (Klamt et al., 1998). The general COSMO-RS parameters α_{mf} and a_{eff} are the misfit prefactor and the effective contact area of a segment, respectively. The hydrogen bond contribution, assuming that segment i is a donor and j is an acceptor ($\sigma_i < \sigma_j$), is expressed as:

$$E_{\text{hb}}(\sigma_i, \sigma_j) = a_{\text{eff}} c_{\text{hb}}(T) [\min(0; \sigma_i + \sigma_{\text{hb}}) \max(0; \sigma_j - \sigma_{\text{hb}})]. \tag{4}$$

Due to this functional form, a non-zero hydrogen bond contribution would be observed only for segments with opposing signs of σ, the absolute value of which exceeds the threshold value of σ_{hb}. This ensures that this effect is considered only for highly polar segments. To account for the temperature dependence of hydrogen bond formation, the following temperature-dependent coefficient $c_{\text{hb}}(T)$ is introduced:

$$c_{\text{hb}}(T) = c_{\text{hb}} \max(0; 1 - c_{\text{hb}}^T + c_{\text{hb}}^T \cdot 298.15/T) \tag{5}$$

At this stage, seven general COSMO-RS parameters are considered: r_{av}, α_{mf}, a_{eff}, f_{corr}, c_{hb}, c_{hb}^T and σ_{hb}. Since in this work we focus only on halocarbons, the hydrogen bonding contribution is irrelevant.

2.1.2. Residual contribution: dispersive interactions

The interaction energy between two segments E_{ij} does not initially include the dispersion contribution, which is arguably negligible for

highly polar mixtures but crucial for systems of low polarity such as halocarbons. In this study, we incorporated the dispersion energy of interactions between segments i and j using a method analogous to the Lorentz-Berthelot combining rule:

$$E_{\mathrm{vdW}} = (1 - k_{ij}) a_{\mathrm{eff}} \tau_i^{\mathrm{vdW}} \tau_j^{\mathrm{vdW}}. \tag{6}$$

Here τ_i^{vdW} and τ_j^{vdW} are the dispersion coefficients of segments i and j, respectively. The classical combining rule for dispersion energies in mixtures, which takes the square root of the product of individual dispersion energies, is derived under several assumptions, one of which is the equality of ionization potentials. However, for hydrocarbons and fluorocarbons, ionization potentials can differ significantly (Hudson and McCoubrey, 1960), leading to deviations from the classical combining rule in their mixtures. To account for this discrepancy, we introduce additional k_{ij} parameters, which allow for adjustments based on the specific characteristics of the interacting components. However, this significantly increases the number of adjustable parameters.

Finally, we obtain the following equation for the segment-segment interaction free energy:

$$E_{ij} = E_{\mathrm{mf}}(\sigma_i, \sigma_j) + E_{\mathrm{hb}}(\sigma_i, \sigma_j) - E_{\mathrm{vdW}}. \tag{7}$$

The functional form in Eq. (6) was inspired by the London attractive potential (London, 1937):

$$\phi_{ij}^{\mathrm{London}} = -\frac{3}{2} \frac{\alpha_i \alpha_j}{r_{ij}^6} \cdot \frac{h\nu_i h\nu_j}{h(\nu_i + \nu_j)}, \tag{8}$$

which quantifies the attractive potential between two spherically symmetrical systems with polarizabilities α_i and α_j separated by the distance r_{ij} from each other. Additionally, Eq. (8) demonstrates the influence of zero-point vibrations with their frequencies ν_i and ν_j.

Within the COSMO-RS framework, direct consideration of the London potential is not feasible due to central assumptions and the two-dimensional nature of the statistical mechanics involved (Klamt, 2005b). However, the functional form of the dispersion parameters in Eq. (6) resembles the dependence of the London potential on atomic polarizabilities in Eq. (8).

2.1.3. Combinatorial contribution

For the combinatorial contribution, we consider the following widely used terms: the Staverman–Guggenheim term (Staverman, 1950) Eq. (9), the Flory–Huggins term (Flory, 1942) Eq. (12) and the Flory–Huggins term modified by Elbro et al. (1990). The equation for the Staverman–Guggenheim (SG) term is:

$$\ln(\gamma_i^{\mathrm{comb}}) = \ln\left(\frac{\phi_i}{x_i}\right) + 1 - \frac{\phi_i}{x_i} - 0.5 z \frac{A_i}{A_{\mathrm{std}}} \left(\ln\left(\frac{\phi_i}{\theta_i}\right) + 1 - \frac{\phi_i}{\theta_i}\right), \tag{9}$$

with the volume fractions ϕ_i and the surface fractions θ_i calculated as:

$$\frac{\phi_i}{x_i} = \frac{V_i}{\sum_j x_j V_j}, \tag{10}$$

$$\frac{\theta_i}{x_i} = \frac{A_i}{\sum_j x_j A_j}. \tag{11}$$

In the equations above, V_i represents the hard-core volume, A_i is the molecular cavity surface area determined using the QC calculations. For the SG term, the coordination number z was set to 10, and A_{std} was treated as an adjustable parameter. The Flory–Huggins (FH) term is expressed as:

$$\ln(\gamma_i^{\mathrm{comb}}) = \ln\left(\frac{\phi_i}{x_i}\right) + 1 - \frac{\phi_i}{x_i}. \tag{12}$$

Elbro et al. (1990) proposed to use 'free' volume V_i' instead of V_i in Eq. (12), calculated as:

$$V_i' = V_i - V_i^{\mathrm{m}}, \tag{13}$$

where V_i^{m} is the molar volume determined from the experimental densities of pure compounds.

2.2. QC calculations

For QC calculations, the open-source workflow described in detail in Gerlach et al. (2022) was used. Following this methodology, all calculations were performed in ORCA 5.0.3 (Neese, 2012; Neese et al., 2020). Initially, a set of conformers was generated using RDKit Python library (RDKit, 2023; Ebejer et al., 2012), followed by ALPB (Bannwarth et al., 2021b; Ehlert et al., 2021) geometry optimizations in water using GFN2-xTB (Bannwarth et al., 2021a) calculations from within ORCA. Then the conformers were filtered and their geometry was optimized with C-PCM at DFT/BP86/def2-TZVP(-f) level. Subsequently, the conformer with the lowest energy was selected and its geometry was reoptimized at DFT/BP86/def2-TZVP. Finally, C-PCM calculations were performed at DFT/BP86/def2-TZVPD level for the most stable conformer. It is worth mentioning that although C-PCM is not identical to the originally used COSMO method, these two methods are essentially the same for a solvent with a dielectric constant $\varepsilon = \infty$. An archive with examples of ORCA files can be found in Supplementary Material. The entire procedure is automated in the openCOSMO-RS conformer generator pipeline code, and the tools involved in all these steps are freely accessible for *academia*.

2.3. Regression procedure

To optimize COSMO-RS parameters, various types of phase equilibrium data were used: IDAC, partition coefficients of component i between water and organic liquid ($K_i^{\mathrm{org/w}}$), VLE and LLE. For binary VLE, we calculated experimental activity coefficients using saturated vapor pressures of pure components either directly from experimental data, or if unavailable, calculated from the correlations provided by DIPPR (DIPPR, 2023), NIST (Burgess Donald, 2024) or Yaws and Satyro (2015). These experimental activity coefficients $\ln(\gamma_i^{\mathrm{exp}})$ were then used to minimize the following objective function:

$$OF = \sum_s^{Ns} w_s \sum_i^{Nc} (\ln(\gamma_i^{\mathrm{calc}}) - \ln(\gamma_i^{\mathrm{exp}}))^2, \tag{14}$$

where N_s is the number of datasets, and N_c is the number of components for which data is available. All weights w_s were set to unity. The same objective function as Eq. (14) was used to regress IDAC data.

The distribution coefficient K_i of component i between phases α and β is calculated as:

$$K_i \equiv \frac{x_i^{\beta,\mathrm{exp}}}{x_i^{\alpha,\mathrm{exp}}} = \frac{\gamma_i^{\alpha,\mathrm{exp}}}{\gamma_i^{\beta,\mathrm{exp}}}. \tag{15}$$

Here, the ratio of mole fractions in separated phases on the left-hand side of this equation was determined from the LLE experimental data, while the ratio of activity coefficients was calculated with openCOSMO-RS at experimental concentration. Although this type of simplified regression does not guarantee the equality of fugacities, previous studies (Gerlach et al., 2018; Müller et al., 2020; Marques et al., 2023), as well as our own results in Section 3.4, have demonstrated that this approach can accurately describe LLE. For the regression of LLE data, we employed the following objective function:

$$OF = \sum_s^{Ns} w_s \sum_i^{Nc} (\ln(K_i^{\mathrm{calc}}) - \ln(K_i^{\mathrm{exp}}))^2, \tag{16}$$

which was also used for partition coefficients $K_i^{\mathrm{org/w}}$.

3. Results and discussion

3.1. Database

To optimize the dispersion parameters for the openCOSMO-RS model, we focused on three types of hydrocarbon and halocarbon thermodynamic data that are widely used for parameterizing thermodynamic models. We believe that these data types effectively capture the

Table 1
An overview of the collected experimental data on hydrocarbons and halocarbons.

Type	Data points	Datasets
IDAC	310	224
LLE	1534	85
VLE	10,104	397

intermolecular interactions occurring in mixtures and cover a broad range of industrial processes where accurate phase equilibrium modeling is essential. For example, precise modeling of VLE data is crucial for distillation processes and is necessary for the design and optimization of refrigeration cycles. Binary VLE data illustrate situations where molecules are surrounded by both similar and different molecules, leading to non-idealities arising from these intermolecular interactions. For IDAC data, the solute molecules are highly diluted in the solvent, allowing us to observe interactions primarily between the solute and solvent molecules. Including LLE data makes the thermodynamic model more versatile, ensuring it can handle this type of phase equilibrium as well. Additionally, LLE data are particularly sensitive to changes in the activity coefficients. In contrast, while VLE predictions also rely on activity coefficients, their accuracy is heavily influenced by the vapor pressure of the pure components, making them somewhat less sensitive to the activity coefficient model compared to LLE data. Therefore, LLE data can then serve as a stringent test for the accuracy of thermodynamic models.

Table 1 provides an overview of the collected datasets, including their types, the number of datasets, and the number of data points used for regressions. More details on the sources of the datasets are reported in Table S1 of the Supplementary Material.

The majority of the VLE datasets are from the Korean Database (KDB, 2024) and from the database by Jaubert et al. (2020). While the former did not report any thermodynamic consistency analysis of the data, the latter was specifically developed for benchmarking thermodynamic models, ensuring a quality assessment of the data. To further assure data quality, we performed the van Ness thermodynamic consistency test (Van Ness, 1995) on the VLE data. Additionally, all datasets were plotted in $\ln(\frac{\gamma_1}{\gamma_2}) - x_1$ coordinates to visually verify a reasonable trend. Approximately half of the initially collected datasets were eliminated due to inconsistencies.

For LLE datasets, a significant fraction of experimental studies carried out the cloud point measurements, which provide only one end of the tie-line. Thus, we correlated these data with the UNIQUAC thermodynamic model (Abrams and Prausnitz, 1975; Anderson and Prausnitz, 1978) and extrapolated the missing tie-lines.

As shown in Fig. 1, the VLE data exhibit a prevalence of both fluorinated and chlorinated compounds, while the LLE data primarily feature florinated compounds. Although experimental data for these compounds are more abundant than those for brominated and iodinated hydrocarbons, it is crucial to include all species to ensure coverage of all possible halogen-halogen combinations. While the dataset used for regression lacks iodinated and brominated data, we expect the model to perform well in predicting these cases due to its underlying physical principles. Fig. S1 compares the total AAD (Average Absolute Deviation) of predictions from experimental data classified by the type of halogen present in the systems. We observe a decrease in AAD across all categories, including those involving brominated and iodinated compounds, which confirms that despite their limited representation in the objective function, the model can accurately predict these species. The frequency of atom-atom pair occurrence in the collected VLE and LLE data is illustrated in Fig. 2. For instance, the binary mixture of *n*-pentane and dichloromethane comprises various atom-atom contacts, including H–H, H–C, H–Cl, C–C, C–Cl, Cl–Cl. Each data point related to this mixture is assigned all the listed pairs to cover the possible segment-segment interactions with atomic number as a descriptor. To illustrate this, we calculated a percentage of data points containing each atom-atom pair relative to the total number of data points. As shown in Fig. 2, data for mixtures containing both Br and I atoms, or Cl and I atoms, is particularly scarce.

3.2. Parametrization procedures

To incorporate the dispersion term into openCOSMO-RS, we introduced several atom-specific and cross-atom parameters (Eq. (6)). The parametrization approaches considered in this chapter, along with their corresponding abbreviations for further reference, are presented in Table 2.

At this stage, our focus is on hydrocarbons and halocarbons, however, the approach will be extended to other types of compounds in the future. Therefore, currently, the initial parameter set corresponds to C, H, F, Cl, Br, and I atoms, resulting in a total of 6 parameters. It is worth mentioning that the polarizability of C-atom vary with its hybridization. Hence, we also used two parameters to describe C(sp^3) and C(sp^2) atoms, leading to 7 parameters in total. The inclusion of cross-interaction parameters k_{ij} (Eq. (6)) expands the total number of parameters to 28 when accounting for differences in C-atom hybridization, and 21 if not. Furthermore, all these parametrization schemes were incorporated with three combinatorial terms (SG, FH, Elbro), resulting in 12 possible combinations.

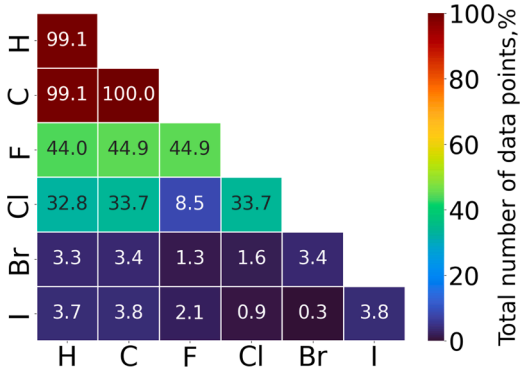

Fig. 2. Frequency of occurrence of atom-atom combinations in each pair of molecules across the VLE and LLE data.

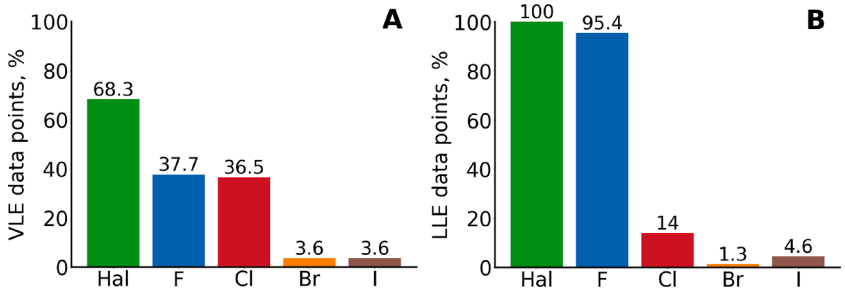

Fig. 1. The distribution of halogen-atoms (Hal) in the VLE (A) and LLE (B) data.

Table 2
Characteristics of the performed parametrizations, including the type of combinatorial term used, the number of atom-specific parameters, and the inclusion of cross-atom k_{ij} parameters, along with their abbreviations.

Combinatorial term	Atom-specific parameters		Atom-specific parameters + cross-atom coefficients	
	6 parameters	7 parameters	6 parameters	7 parameters
Staverman–Guggenheim	SG_6	SG_7	SG_6_cross	SG_7_cross
Flory–Huggins	FH_6	FH_7	FH_6_cross	FH_7_cross
Elbro	Elbro_6	Elbro_7	Elbro_6_cross	Elbro_7_cross

Tables 3 and 4 summarize all the parameters regressed in this study. Table 3 covers parametrizations with only 6 or 7 atom-specific dispersion parameters. For SG_6 and SG_7, the general model parameters (a_{eff}, r_{av}, α_{mf}, f_{corr}, c_{hb}, c_{hb}^{T}, σ_{hb}, and A_{std}) were fixed to the corresponding values of ORCA-B parametrization with the SG term from Gerlach et al. (2022) (the benchmark model without the dispersion term, denoted as 'no dispersion' in subsequent Figures). For the FH term, this set of parameters was regressed to the $K_i^{\text{org/w}}$ and IDAC data using the same database as for SG_6 and SG_7 reported in Gerlach et al. (2022). With the Elbro term, we followed a different approach by regressing both general and dispersion parameters simultaneously only to the data in Table 1 due to poor performance of the Ebro term for the highly polar and aqueous systems, which are abundant in IDAC and $K_i^{\text{org/w}}$ data collection. Consequently, the general parameters for the Elbro term are biased towards the hydrocarbon and halocarbon data, diverging from those regressed for other combinatorial terms using a dataset that is more evenly distributed across various types of chemicals. While all sets of general and atom-specific dispersion parameters in Table 3 are implemented across the considered models that share the same combinatorial term and number of atom-specific parameters, Table 4 further extends these sets with respective parameters for cross-interactions.

For regressions, we utilized the differential evolution algorithm provided in the SciPy Python package (Virtanen et al., 2020). While this algorithm does not require initial estimates and avoids using gradient methods, it explores a large sample parameter space, resulting in longer optimization times. For example, the optimization of atomic dispersion parameters took approximately 10 hours, while the regression of cross-interaction parameters required up to 5 days on HPC cluster using 1 to 2 cores. Regression times can be significantly reduced by using Nedler-Mead algorithms. However, we opted for differential evolution as its strength lies in its ability to identify the global minimum within the parameter space more effectively than conventional gradient methods. Furthermore, the minimizations can be constrained using bounds and dependencies between parameters. To impose physically meaningful constraints, we used the values of atomic polarizabilities listed in Table 5. Consequently, the atom-specific parameters align well with these polarizability values, as demonstrated in Fig. 3, with parametrizations involving the FH combinatorial term being the closest match.

Additionally, we estimated the local sensitivities S_x of the τ^{vdW} parameters using a backward difference method:

$$S_x = \frac{1}{N_P}\frac{OF(x-\epsilon) - OF(x)}{\epsilon}, \tag{17}$$

where x is the evaluated parameter, with the others being fixed, OF is the value of the objective function at x, and ϵ is the increment, which was set to 5 % of the parameter value. Due to a scarcity of data with Br and I atoms compared to the data with F and Cl atoms, we weighted the sensitivity with the number of data points N_P across the data containing these atoms. An example of such sensitivity analysis is demonstrated in Fig. 4 for models without cross-atom parameters. Notably, the highest sensitivities are mostly attributed to the F-atom parameter, while H and C atoms parameters show low sensitivities. Interestingly, all sensitivities for Elbro_6 and Elbro_7 models are much higher than those estimated with other combinatorial terms.

To assess the performance of the modified models, Tables 3 and 4 also provide the AAD estimated using the following equations:

$$\text{AAD}_{\text{LLE}} = \frac{1}{N_p}\sum^{N_s}\sum_{i}^{Nc}\frac{\left|\ln(K_i^{\text{calc}}) - \ln(K_i^{\text{exp}})\right|}{N_c}, \tag{18}$$

$$\text{AAD}_{\text{VLE}} = \frac{1}{N_p}\sum^{N_s}\sum_{i}^{Nc}\frac{\left|\ln(\gamma_i^{\text{calc}}) - \ln(\gamma_i^{\text{exp}})\right|}{N_c}, \tag{19}$$

Table 3
Parameter values and deviations for openCOSMO-RS dispersion parametrizations without cross-interaction parameters. Parameters that were fixed during optimization are marked with asterisk.

Parameters and deviations (AAD)	Parametrizations					
	SG_6	SG_7	FH_6	FH_7	Elbro_6	Elbro_7
a_{eff} [Å^2]	6.115		5.034		2.745619	2.773896
r_{av}^{*} [Å]	0.5					
α_{mf} [J/(mol Å^2)/e^2]	7.584E+06		7.592E+06		1.210E+07	1.198E+07
f_{corr}^{*}	2.4					
c_{hb} [J/(mol Å^2)/e^2]	3.093E+07		3.094E+07		3.093E+07*	
c_{hb}^{T*}	1.5					
σ_{hb} [e/Å^2]	0.007876		0.007276		0.007876*	
Astd [Å^2]	41.89					
$\tau^{\text{vdW}}_{\text{C(sp}^3\text{)}}$ [J$^{0.5}$/Å]	11.193	9.425	14.123	10.577	17.675	18.221
$\tau^{\text{vdW}}_{\text{C(sp}^2\text{)}}$ [J$^{0.5}$/Å]		10.235		11.480		19.054
$\tau^{\text{vdW}}_{\text{H}}$ [J$^{0.5}$/Å]	10.041	9.021	12.581	10.300	15.636	17.506
$\tau^{\text{vdW}}_{\text{F}}$ [J$^{0.5}$/Å]	3.240	1.977	5.319	2.522	8.914	10.260
$\tau^{\text{vdW}}_{\text{Cl}}$ [J$^{0.5}$/Å]	11.865	10.647	14.575	11.735	18.289	19.669
$\tau^{\text{vdW}}_{\text{Br}}$ [J$^{0.5}$/Å]	17.602	16.414	20.796	17.545	25.965	27.218
$\tau^{\text{vdW}}_{\text{I}}$ [J$^{0.5}$/Å]	19.578	18.236	22.823	20.031	26.716	28.244
AAD_{TOT}	0.0473	0.0469	0.0466	0.0460	0.0491	0.0489
AAD_{IDAC}	0.1949	0.1929	0.2528	0.2521	0.1007	0.0998
AAD_{LLE}	0.1344	0.1328	0.0964	0.0937	0.1055	0.1045
AAD_{VLE}	0.0296	0.0294	0.0327	0.0324	0.0390	0.0389

Table 4
Cross-interaction parameter values and deviations for the respective openCOSMO-RS dispersion parametrizations.

Parameters and Deviations (AAD)	Parametrizations					
	SG_6_cross	SG_7_cross	FH_6_cross	FH_7_cross	Elbro_6_cross	Elbro_7_cross
$k_{H\text{-}C(sp^3)}$	0.18159	0.01833	0.24871	0.00159	0.15927	−0.81875
$k_{H\text{-}C(sp^2)}$		−0.11976		0.15274		−0.34582
$k_{H\text{-}F}$	0.06897	0.66191	0.09735	0.74790	0.07372	−0.17848
$k_{H\text{-}Cl}$	0.22346	0.25723	0.16732	0.14724	0.14310	−0.04194
$k_{H\text{-}Br}$	−0.10650	−0.24431	0.16923	0.16498	0.10780	−0.25552
$k_{H\text{-}I}$	0.56261	0.29858	0.29177	0.22584	0.36641	0.11027
$k_{C(sp^3)\text{-}C(sp^2)}$		0.13688		0.01047		−0.07518
$k_{C(sp^3)\text{-}F}$	0.25838	−0.73798	0.24347	−0.94343	0.07947	−0.63680
$k_{C(sp^2)\text{-}F}$		−0.38252		0.01254		−0.14392
$k_{C(sp^3)\text{-}Cl}$	−0.09500	−0.19281	−0.01799	−0.10530	−0.03320	−0.33904
$k_{C(sp^2)\text{-}Cl}$		−0.19363		−0.03107		−0.11235
$k_{C(sp^3)\text{-}Br}$	−0.13948	0.07020	−0.08836	−0.20372	−0.09358	−0.04681
$k_{C(sp^2)\text{-}Br}$		−0.15348		0.00222		−0.04118
$k_{C(sp^3)\text{-}I}$	−0.47451	−0.25369	−0.25303	−0.21984	−0.30310	−0.45880
$k_{C(sp^2)\text{-}I}$		−0.42763		−0.37158		−0.28746
$k_{F\text{-}Cl}$	−0.02230	−0.14832	−0.03033	−0.10208	−0.01986	−0.02789
$k_{F\text{-}Br}$	−0.40404	−0.33715	−0.01564	−0.31896	−0.11410	0.02610
$k_{F\text{-}I}$	−0.12331	0.02270	−0.19483	0.05881	−0.12197	−0.11127
$k_{Cl\text{-}Br}$	0.11674	0.00928	−0.00186	0.06092	0.04101	−0.02772
$k_{Cl\text{-}I}$	−0.03422	−0.03802	−0.01921	−0.02616	−0.03001	−0.02558
$k_{Br\text{-}I}$	0.08028	−0.04237	0.01017	0.08432	0.05495	0.02172
AAD_{TOT}	0.0431	0.0424	0.04015	0.03878	0.04004	0.03822
AAD_{IDAC}	0.1838	0.1898	0.22506	0.22706	0.07917	0.08143
AAD_{LLE}	0.1326	0.1268	0.09632	0.08888	0.10071	0.09645
AAD_{VLE}	0.0252	0.0250	0.02595	0.02539	0.02962	0.02806

Table 5
Atomic polarizabilities from Schwerdtfeger and Nagle (2019).

Atomic symbol	Polarizability [a.u.^{3}]
H	4.5
C	11.3
F	3.74
Cl	14.6
Br	21
I	32.9

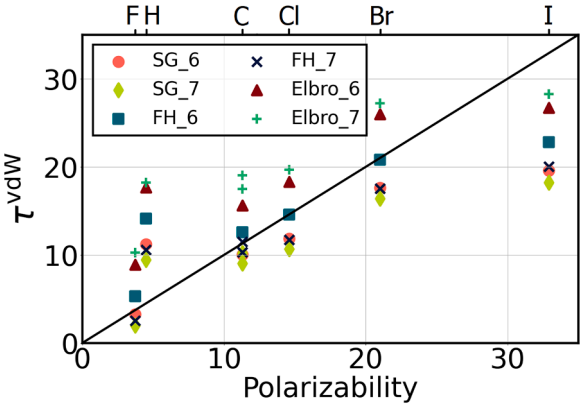

Fig. 3. Comparison between atomic polarizabilities ([a.u.^{3}]) from Schwerdtfeger and Nagle (2019) and atom-specific dispersion parameters (τ^{vdW}).

$$AAD_{IDAC} = \frac{1}{N_p}\sum^{N_s}\left|\ln(\gamma_\infty^{calc}) - \ln(\gamma_\infty^{exp})\right|, \quad (20)$$

and AAD_{TOT} calculated as the sum of all the absolute deviations divided by the total number of the data points across all types of data.

3.3. Evaluation of combinatorial terms for IDAC predictions

For the IDAC data, the dispersive contribution does not significantly improve openCOSMO-RS predictions as shown in Fig. 5, since most of the collected data consists of asymmetric hydrocarbon mixtures with only a few halocarbon datasets. Typically, both models tend to underestimate the activity coefficients. There are, however, a few outliers to this trend, specifically in the benzene/toluene-hexadecane systems. As shown in Fig. S2, the IDACs of benzene and toluene sharply decrease with increasing temperature, forming an S-shaped curve. This peculiar trend may raise questions about the quality of the data.

For highly asymmetric alkane mixtures, the combinatorial term plays a crucial role. Thus, we evaluated three combinatorial terms: SG, FH and Elbro on the IDAC data of alkane mixtures collected by Soares (2011). It should be noted that the evaluated IDAC data consists of short-chain solutes in long-chain solvents only. As shown in Figs. 6 and 7, the Elbro term outperforms both SG and FH, which show comparable performance, with FH being the most inferior. Remarkably, the evidence from molecular simulations studies (Iwai et al., 2010) and polymers research (Silva et al., 2023) indicates that the Elbro term is in better agreement with Monte Carlo simulations and leads to more accurate predictions for polymer solutions, when coupled with COSMO-RS implemented in COSMOtherm. It should be noted that several combinatorial terms are implemented in COSMOtherm software and the Elbro term is one of them. The performance of the FH and Elbro terms and their modifications for asymmetric alkanes and polymer mixtures has been previously investigated by Kontogeorgis et al. (1994, 1997) and Kouskoumvekaki et al. (2002). It was found that the Elbro term performs well for short-chain solutes in long-chain solvents but not for long-chain solutes in short-chain solvents. Kouskoumvekaki et al. (2002) evaluated the performance of the Elbro term and suggested that the inaccessible hard-core volume might be higher than the van der Waals volume currently used. They increased its value by a factor of 1.2, which led to improved results. Interestingly, the volumes of COSMO cavities are related to the van der Waals volumes by the same factor (Klamt et al., 1998). Since these COSMO cavity volumes were used in our calculations, this could account for the superior performance of the Elbro term for the evaluated set. As stated by Kontogeorgis et al. (1994), it is challenging to predict both the activity coefficients of short-chain solutes in long-chain solvents and long-chain solutes in short-chain solvents with a single model. Generally speaking, to select the optimal combinatorial term for openCOSMO-RS, a broader experimental dataset

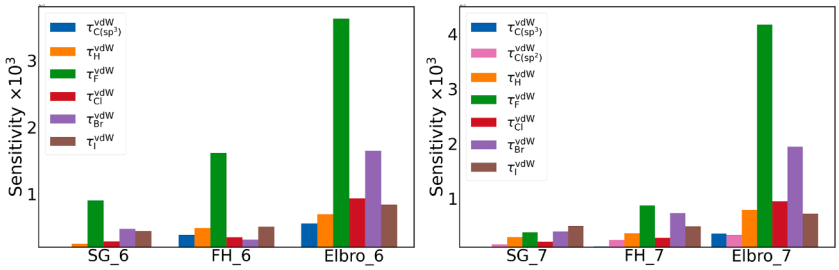

Fig. 4. Local sensitivities for atom-specific parameters estimated for all types of phase equilibrium data collectively.

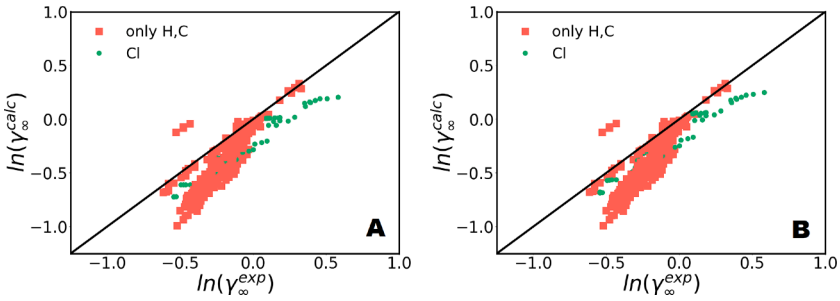

Fig. 5. Parity plots comparing predicted values (calc) with experimental data (exp) for IDACs without the dispersion term (A) or using SG_6 parametrization (B). The symbols are colored according to the type of atoms in the data.

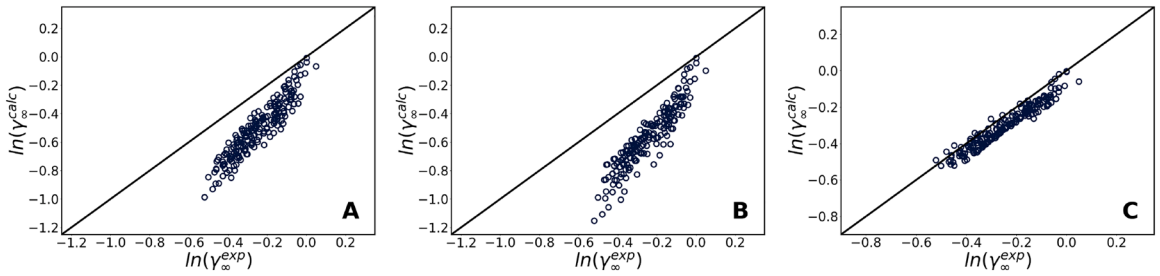

Fig. 6. Parity plots for predicted IDAC (calc) versus experimental data (exp) using one of the following combinatorial terms solely: SG (A), FH (B) or Elbro (C).

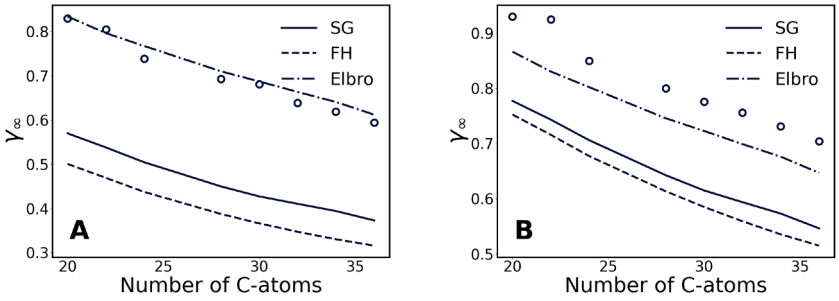

Fig. 7. The experimental (**o**) and predicted by SG, FH and Elbro terms IDAC values of *n*-butane (A) and 4-methylheptane (B) diluted in long-chain alkanes with carbon number indicated on the *x-axis*.

should be considered in the evaluation, which we plan to address in future research. It is also important to emphasize that the Elbro term comes with practical limitations, one of which is the necessity of additional experimental data on the density of pure components to calculate the temperature-dependent molar volume. This dependency on experimental data can restrict the model's predictive capabilities, particularly when such data are unavailable or difficult to obtain.

3.4. Evaluation of the parametrization approaches for VLE and LLE predictions

Disregarding the dispersion contribution leads to substantial deviations from the experimental data for both VLE and LLE, as one can observe in Fig. 8A and B. Conversely, Fig. 8C and D show that incorporating the dispersion term with a minimal number of adjustable parameters results in significant improvements in prediction accuracy. For comparison, we calculated LLE, VLE, and IDAC using the COSMO-SAC-dsp implementation by Bell et al. (2020) and COSMOtherm v.24, with the results presented in Fig. 8E–H and Table 6, respectively. The COSMO-SAC-dsp model was evaluated using the same dataset, but it lacks predictions for brominated and iodinated compounds due to the absence of corresponding dispersion parameters in its parameterization. While COSMO-SAC-dsp performs slightly worse than our model for VLE data, it shows a more pronounced inferiority in predicting LLE data. When compared with COSMOtherm (Table 6), which provided predictions for all the datasets considered, we analyzed two parameterizations: "BP-TZVP"

Table 6
Comparison of AAD between openCOSMO-RS SG_6 and COSMOtherm 24.

		openCOSMO-RS SG_6	COSMOtherm 24 BP-TZVP		COSMOtherm 24 BP-TZVPD-FINE	
	Data points	Single conformer	Single conformer	Multiple conformers	Single conformer	Multiple conformers
IDAC	310	0.1949	0.2194	0.21936	0.18488	0.18488
LLE	1534	0.1344	0.2528	0.25242	0.1037	0.10480
VLE	10,104	0.0296	0.0500	0.04483	0.0332	0.02762
Total	11,948	0.0473	0.0804	0.07601	0.0462	0.04161

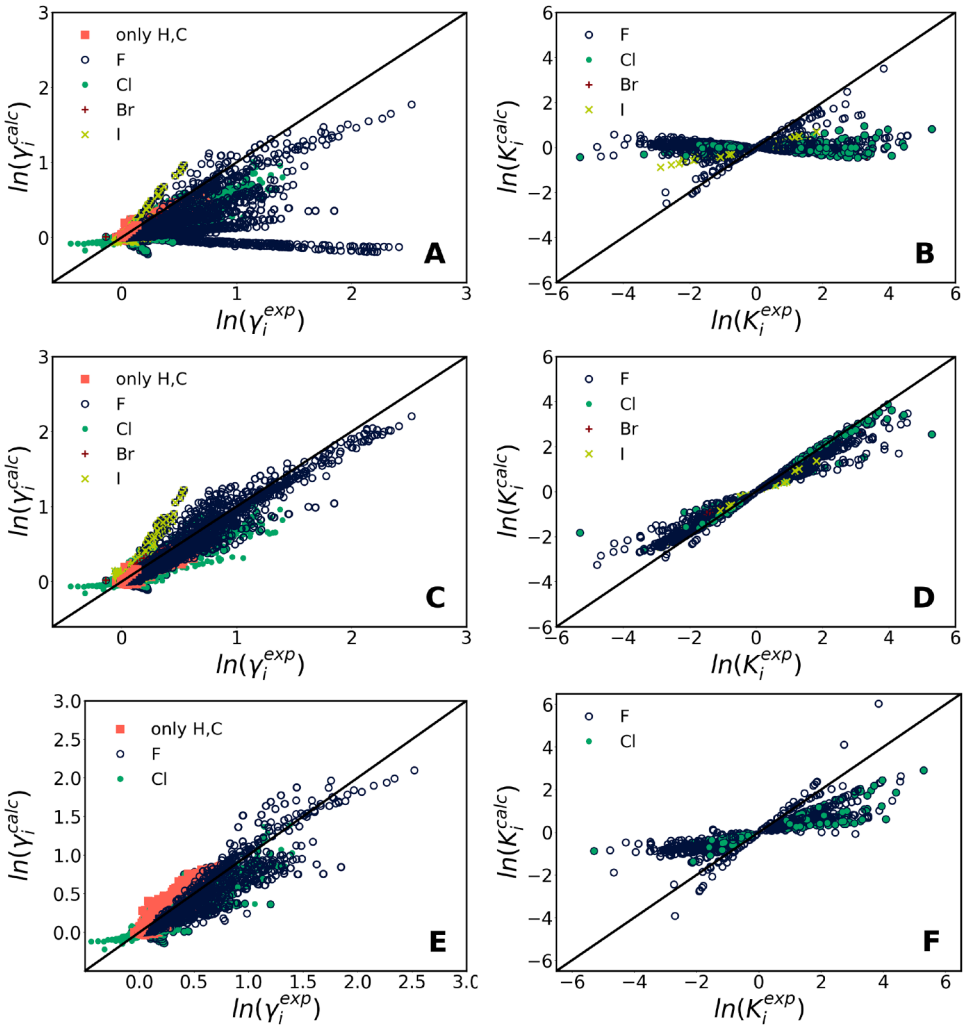

Fig. 8. Parity plots comparing predicted values (calc) with experimental data (exp) for VLE activity coefficients without the dispersion term (A), using SG_6 parametrization (C) or COSMO-SAC-dsp (E), and for LLE partition coefficients without the dispersion term (B), using SG_6 parametrization (D) or COSMO-SAC-dsp (F). The symbols are colored according to the type of atoms in the data.

and "BP-TZVPD-FINE". The "BP-TZVPD-FINE" parameterization outperforms "BP-TZVP" due to its more advanced basis set (TZVPD), a refined cavity construction algorithm (FINE) and the underlying COSMO-RS model. Additionally, we assessed whether the inclusion of multiple conformers improves prediction accuracy. The comparison indicates that our approach is comparable to the "BP-TZVPD-FINE" parameterization and superior to the "BP-TZVP" one. Including multiple conformers appears to enhance the predictions, likely due to the presence of flexible molecules where *cis-trans* isomerism is relevant.

The AAD values (Tables 3 and 4) for all the regressed models indicate that differentiating between sp^3 and sp^2 hybridizations of C-atoms does not significantly affect the predictions of phase equilibrium. While improvements can be seen for some systems, as shown in Fig. 9, overall, not much difference is observed on a large scale. Besides, we did not consider sp hybridization due to the absence of corresponding experimental data in our database.

To demonstrate the impact of the dispersion term on VLE predictions, Fig. 10A and B present two P-*xy* phase diagrams of fluorinated refrigerant mixtures. In the absence of the dispersion contribution, the model fails to capture azeotropic behavior and even predicts nearly ideal behavior for *n*-propane-pentafluoroethane (Fig. 10B). However, incorporating the dispersion term, even via the simplest models not considering C-atom hybridization or cross-interactions, leads to more physically meaningful predictions. All illustrated models successfully capture the azeotrope. Moreover, the choice of combinatorial term does not have a substantial effect in those cases.

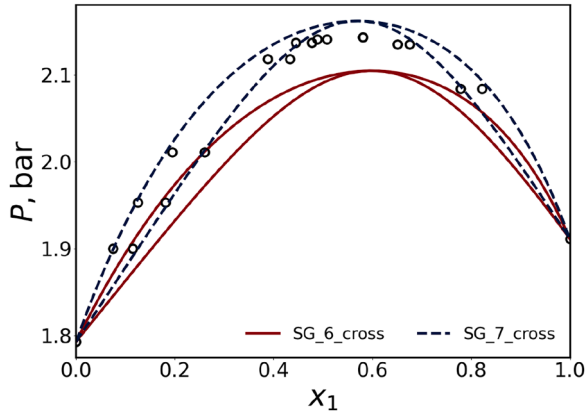

Fig. 9. P-*xy* phase diagram of trifluoroiodomethane-*trans*-1,3,3,3-tetrafluoropropene at 268.15 K. Experimental data (Guo et al., 2012) is represented by symbols (**o**), while corresponding lines depict predictions made using various openCOSMO-RS dispersion parametrizations. These predictions illustrate the effect of differentiating between C-atoms with sp^3 and sp^2 hybridizations (SG_7_cross) compared to using a single dispersion parameter for C-atom (SG_6_cross).

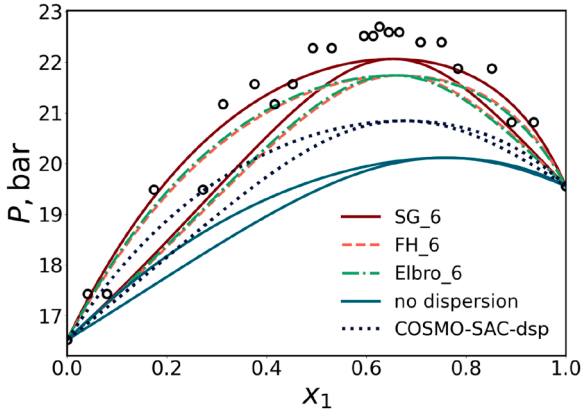

Fig. 11. P-*xy* phase diagram of chlorodifluoromethane-octafluoropropane at 323.15 K. Experimental data (KDB, 2024) is represented by symbols (**o**), while corresponding lines are the predictions made using various openCOSMO-RS dispersion parametrizations and COSMO-SAC-dsp.

The maximum saturation pressure for the selected VLE systems is approximately 20 bar. At higher pressures, one should consider the non-idealities in the gas phase. Moreover, for certain mixtures, even at 20 bar, the non-ideal gas behaviour could not be ignored. However, as depicted in Fig. 11, all dispersion models without cross-interactions are in a good agreement with the experimental data, although some points are at pressures higher than 20 bar.

As reported in Tables 3 and 4, the inclusion of additional cross-interaction parameters improves predictions for various types of phase equilibrium but not significantly. In the case of VLE, most systems are modeled similarly with and without additional cross-interaction parameters k_{ij}. Nevertheless, certain mixtures containing multiple halogen atoms benefit from the additional k_{ij}. For instance, as shown in Fig. 12A and B, the models lacking cross-interaction parameters tend to overestimate saturation pressures. To better understand the impact of the cross-interaction parameters, we estimated the local sensitivities of the k_{ij} using the backward difference method mentioned previously (Eq. (17)). One example is Fig. 13 that demonstrates the local sensitivity analysis for the SG_6_cross parametrization. Notably, the highest sensitivities are mostly attributed to the halogen-halogen parameters (e.g. Cl-I, Br-I, Cl-Br, F-I), which are responsible for the improved predictions in the discussed examples. Although there is no clearly defined trend of the k_{ij} parameters, one could notice some correlation between the sign and value of k_{ij} and the size differences of the interacting atoms. For instance, accounting for k_{ij} typically increases E_{vdW} compared to $k_{ij} = 0$ for interactions involving the I-atom, while the opposite is generally true for interactions with H-atom.

The workflow utilized for the QC calculations had some difficulty in identifying the most stable conformer for two molecules: 1,2-dichloroethane (Fig. 15) and 1,2-dibromoethane. For instance, we observed unreasonably high deviations between predictions made using openCOSMO-RS and COSMO-SAC-dsp for the 1,2-dichloroethane-tetrachloromethane mixture, irrespective of the dispersion term used. These significant discrepancies arise due to the utilization of different conformers in calculations. Initially (Fig. 14A), the workflow determined the *cis*-conformer of 1,2-dichloroethane as the most stable. However, it is widely acknowledged that *trans*-conformers are typically more stable. Consequently, we repeated the calculations using the more stable *trans*-conformer (Fig. 14B), resulting in a remarkable difference. Although the assessment of conformers is beyond the scope of the present study, it is essential to emphasize their critical role in COSMO-RS predictions (Klamt, 2005c). Therefore, an improved conformer search algorithm is currently under development.

Despite the significant improvements in VLE predictions due to incorporation of the dispersion term into openCOSMO-RS, several challenges remain. One such challenge is illustrated in Fig. 16, where poor predictions may arise from neglecting the steric hindrance and the equilibrium distribution of conformers. The model without the dispersion contribution predicts a nearly ideal mixture, while other dispersion models overestimate the saturation pressures.

In LLE predictions, both openCOSMO-RS without the dispersion term and COSMO-SAC-dsp fail to identify any phase split for more than half of the collected systems. However, in certain cases, such as Fig. 17A and C,

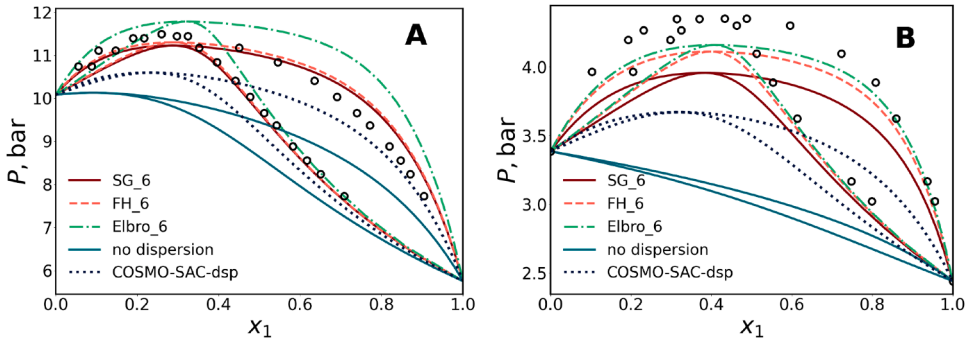

Fig. 10. P-*xy* phase diagram of isopentane-1,1,1,3,3-pentafluoropropane at 362.94 K (A) and of *n*-propane-pentafluoroethane at 253.15 K (B). Experimental data A - (El Ahmar et al., 2012) and B - (Kim et al., 2003) is represented by symbols (**o**), while corresponding lines are the predictions made using various openCOSMO-RS dispersion parametrizations and COSMO-SAC-dsp.

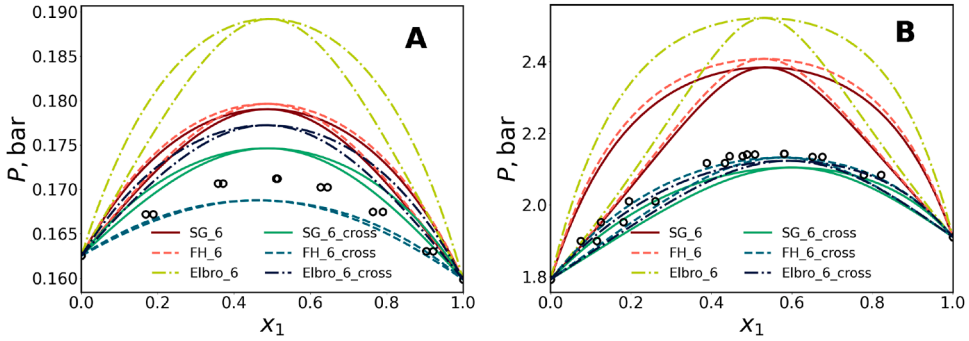

Fig. 12. P-*xy* phase diagrams of 1,2-dibromoethane - chlorobenzene at 348.12 K (A) and of trifluoroiodomethane-*trans*-1,3,3,3-tetrafluoropropene at 268.15 K (B). Experimental data A - (Lacher and Hunt, 1941) and B - (Guo et al., 2012) is represented by symbols (**o**), while corresponding lines depict predictions made using various openCOSMO-RS dispersion parametrizations.

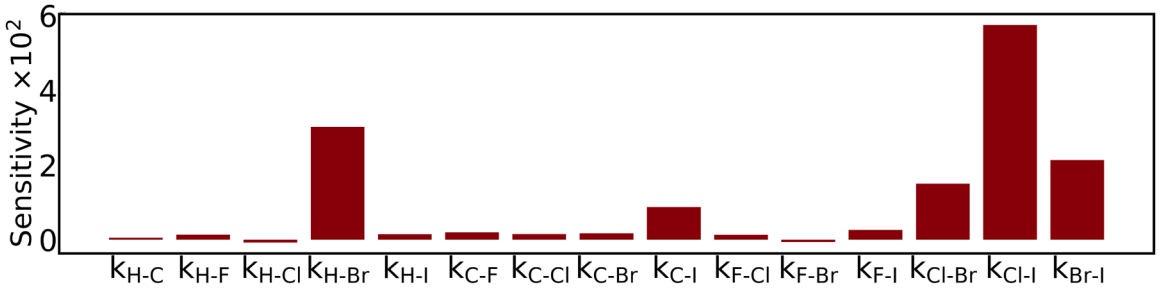

Fig. 13. Local sensitivities for cross-atom parameters of SG_6_cross model estimated for VLE data.

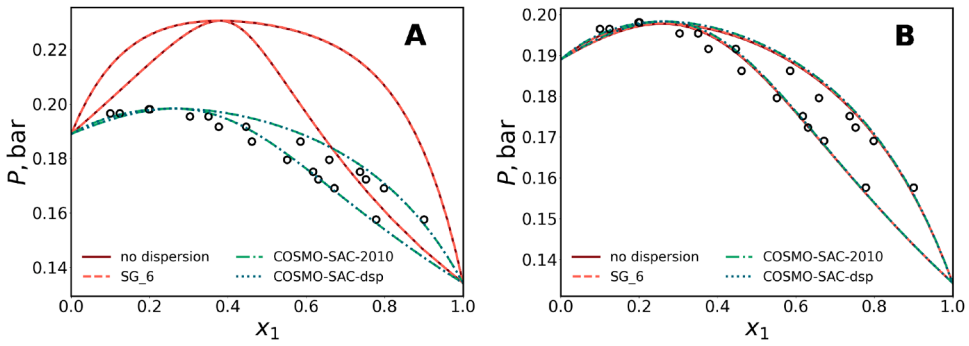

Fig. 14. P-*xy* phase diagrams of 1,2-dichloroethane-tetrachloromethane at 303.15 K. Experimental data (Jaubert et al., 2020) is represented by symbols (**o**), while corresponding lines depict predictions made using various openCOSMO-RS and COSMO-SAC models. For openCOSMO-RS, the *cis*-conformer of 1,2-dichloroethane is used in (A) and *trans*-conformer is used in (B).

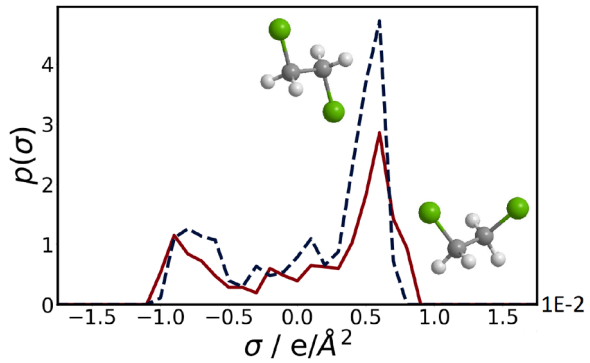

Fig. 15. σ-profiles of *cis*- and *trans*- conformers of 1,2-dichloroethane. The 3D structure of the conformers is visualized using Chem3D software.

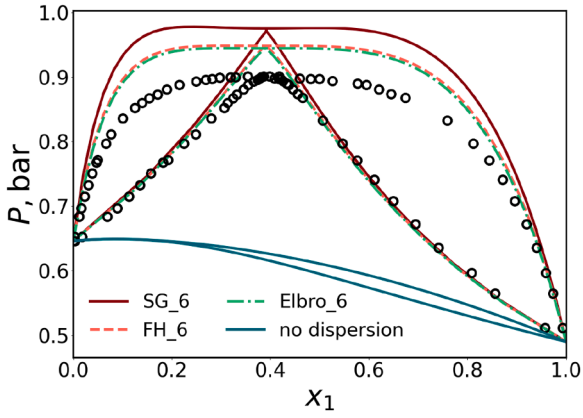

Fig. 16. P-*xy* phase diagram of perfluoromethylcyclohexane-n-hexane at 328.15 K. Experimental data (KDB, 2024) is represented by symbols (**o**), while corresponding lines depict predictions made using various openCOSMO-RS models.

reasonable predictions were achieved regardless of the absence of the dispersion term. As demonstrated in Tables 3 and 4, models utilizing the SG combinatorial term are slightly inferior to those using the FH or Elbro terms. Specifically, models involving the FH combinatorial term

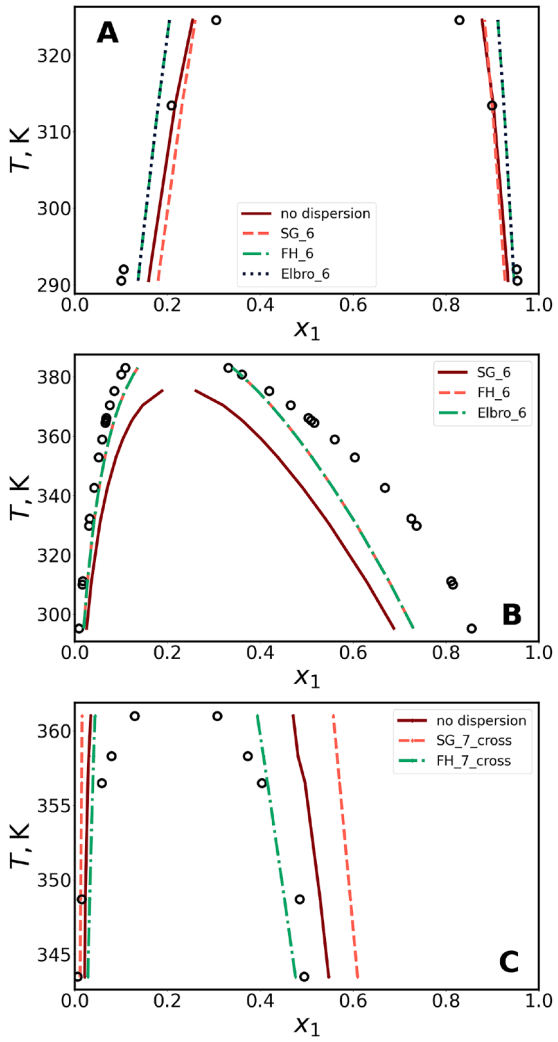

Fig. 17. LLE of the octane-perfluoroheptane (A), the perfluoroheptane-benzene (B), and the octadecafluorooctane-dichloromethane (C) systems. Experimental data (Lo Nostro et al., 2005) is represented by symbols (**o**), while corresponding lines depict predictions made using various openCOSMO-RS models.

lead to the lowest deviations between experimental data and predictions for LLE. For instance, in Fig. 17B, the SG_6 parametrization captures the curvature of the phase split, however it underestimates the critical solution temperature, which is a notoriously challenging property to predict. Conversely, predictions by FH_6 and Elbro_6 are in better agreement with the experimental data. It is noteworthy that capturing the dome-shape of the binodal curve is a general challenge for all parametrizations across the majority of systems, regardless of the number of parameters. One can observe in Fig. 17C that even the most accurate parametrization for LLE FH_7_cross does not entirtely follow the dome-shape of the binodal curve.

The impact of the combinatorial term on VLE and LLE is not as pronounced as on IDAC data, although still substantial. Despite our expectations after analyzing IDAC predictions, the Elbro term did not demonstrate an outstanding performance for other types of phase equilibrium evaluated in this study. AAD values for VLE data listed in Tables 3 and 4 indicate its less satisfactory performance compared to other models. However, for LLE data, the Elbro term performs almost as well as the FH term. One could notice from the list of systems in Tables S1 and S2 of the Supplementary Material that LLE datasets comprise more asymmetrical systems than VLE datasets. The advantage of the Elbro term in modeling asymmetric systems agrees with previous findings. The SG term models tend to produce VLE predictions the most accurately. Although the FH term leads to the poorest predictions of IDACs, it seems to be the best choice for LLE. Therefore, across all the parametrizations obtained in this study, there are no universal solutions leading to the superior predictions for all types of phase equilibrium. Nevertheless, when modeling halocarbons/refrigerants phase equilibrium, we consider the FH term to be the optimal choice. Models incorporating the FH term predict all types of the phase equilibrium reasonably well. Unlike the SG term, it does not require an additional general parameter to the model, and unlike Elbro, it does not require density experimental data. Moreover, in a recent work by Krooshof and de With (2024), it was demonstrated that many of the modifications to the FH term are inconsistent in their application. Consequently, the authors strongly recommend the use of the original FH term for more reliable and consistent modeling. That said, in general, the performance of the combinatorial terms varies for different phase equilibrium types, however in the light of developing a general predictive model, finding a satisfactory compromise is important. In the future work, we plan to extend our analysis on a larger dataset with more chemically diverse compounds.

4. Conclusions

In this study, we improved openCOSMO-RS by incorporating the dispersion contribution and applied the modified version to predict phase equilibrium of halocarbons and refrigerants. We collected and evaluated an extensive database of VLE, LLE, and IDAC to parameterize and assess the model's performance.

Furthermore, our analysis showed that differentiating between carbon atom hybridizations did not yield significant improvements in predictions. While extending the parameter set with k_{ij} cross-atom parameters proved beneficial for certain systems with multiple halogens, the overall improvements were not substantial compared to parametrizations without them. Therefore, a relatively simple parameter set without additional k_{ij} and without a second parameter for carbon atom proved sufficient for satisfactory phase equilibrium predictions of the considered mixtures.

Additionally, we evaluated the performance of several combinatorial terms. The theoretically well-supported Elbro term demonstrated remarkable accuracy in predicting IDAC data of alkanes. However, for VLE, its accuracy was inferior to the SG and FH terms. Specifically, the FH term performed better for LLE, while the SG term excelled for VLE.

Overall, for modeling VLE and LLE of halocarbons and refrigerants, we recommend using the FH term coupled with the dispersion parametrization consisting of six atom-specific parameters. In future research, we aim to implement the dispersive contribution in the improved version of openCOSMO-RS (Müller et al., 2024) for a broader range of systems.

Data availability

Data are available on request from authors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Daria Grigorash: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization; **Simon Müller:** Writing – review & editing, Validation, Software, Methodology, Investigation, Conceptualization; **Patrice Paricaud:** Writing – review & editing,

Supervision, Investigation, Conceptualization; **Erling H. Stenby:** Supervision, Project administration; **Irina Smirnova:** Supervision, Project administration; **Wei Yan:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Supplementary material

Supplementary material associated with this article can be found in the online version at 10.1016/j.ces.2025.121425.

References

Abrams, D.S., Prausnitz, J.M., 1975. Statistical thermodynamics of liquid mixtures: a new expression for the excess Gibbs energy of partly or completely miscible systems. AIChE J. 21 (1), 116–128. https://doi.org/10.1002/aic.690210115

Anderson, T.F., Prausnitz, J.M., 1978. Application of the UNIQUAC equation to calculation of multicomponent phase equilibria. 1. Vapor-liquid equilibria. Ind. Eng. Chem. Process Des. Dev. 17 (4), 552–561. https://doi.org/10.1021/i260068a028

Arrad, M., Thomsen, K., Müller, S., Smirnova, I., et al.. Thermodynamic modeling using extended UNIQUAC and COSMO-RS-ES models: case study of the cesium nitrate-water system over a large range of temperatures 580, 114037. https://www.sciencedirect.com/science/article/pii/S0378381224000141. https://doi.org/10.1016/j.fluid.2024.114037

Bannwarth, C., Caldeweyher, E., Ehlert, S., Hansen, A., Pracht, P., Seibert, J., Spicher, S., Grimme, S., 2021a. Extended tight-binding quantum chemistry methods. WIREs Comput. Mol. Sci. 11 (2), e1493. https://doi.org/10.1002/wcms.1493

Bannwarth, C., Caldeweyher, E., Ehlert, S., Hansen, A., Pracht, P., Seibert, J., Spicher, S., Grimme, S., 2021b. Extended tightbinding quantum chemistry methods. WIREs Comput. Mol. Sci. 11 (2). https://doi.org/10.1002/wcms.1493

Bell, I.H., Mickoleit, E., Hsieh, C.M., Lin, S.T., Vrabec, J., Breitkopf, C., Jäger, A., 2020. A benchmark open-source implementation of COSMO-SAC. J. Chem. Theory Comput. 16 (4), 2635–2646. https://doi.org/10.1021/acs.jctc.9b01016

BIOVIA. Cosmotherm. https://www.3ds.com/products/biovia/cosmo-rs/cosmotherm.

Burgess Donald, Jr., R., 2024. NIST Chemistry WebBook, NIST Standard Reference Database Number 69. Vol. Thermochem. National Institute of Standards and Technology, Gaithersburg MD, 20899.

Diedenhofen, M., Klamt, A., 2010. COSMO-RS as a tool for property prediction of IL mixtures—A review. Fluid Phase Equilib. 294 (1–2), 31–38. https://doi.org/10.1016/j.fluid.2010.02.002

DIPPR, 2023. https://www.aiche.org/dippr/projects/801.

Ebejer, J.-P., Morris, G.M., Deane, C.M., 2012. Freely available conformer generation methods: how good are they? J. Chem. Inf. Model. 52 (5), 1146–1158. https://doi.org/10.1021/ci2004658

Ehlert, S., Stahn, M., Spicher, S., Grimme, S., 2021. Robust and efficient implicit solvation model for fast semiempirical methods. J. Chem. Theory Comput. 17 (7), 4250–4261. https://doi.org/10.1021/acs.jctc.1c00471

El Ahmar, E., Valtz, A., Paricaud, P., Coquelet, C., Abbas, L., Rached, W., 2012. Vapour–liquid equilibrium of binary systems containing pentafluorochemicals from 363 to 413 K: measurement and modelling with Peng–Robinson and three SAFT-like equations of states. Int. J. Refrig 35 (8). https://doi.org/10.1016/j.ijrefrig.2012.05.016

Elbro, H.S., Fredenslund, A., Rasmussen, P., 1990. A New Simple Equation for the Prediction of Solvent Activities in Polymer Solutions. Technical report. https://www.pubs.acs.org/sharingguidelines.

Fingerhut, R., Chen, W.L., Schedemann, A., Cordes, W., Rarey, J., Hsieh, C.M., Vrabec, J., Lin, S.T., 2017. Comprehensive assessment of COSMO-SAC models for predictions of fluid-phase equilibria. Ind. Eng. Chem. Res. 56 (35), 9868–9884. https://doi.org/10.1021/acs.iecr.7b01360

Flôres, G.B., Staudt, P.B., Soares, R.de P., 2016. Including dispersive interactions in the F–SAC model. Fluid Phase Equilib. 426. https://doi.org/10.1016/j.fluid.2016.02.043

Flory, P.J., 1942. Themodynamics of high polymer solutions. J. Chem. Phys. 10 (1), 51–61. https://doi.org/10.1063/1.1723621

Gerlach, T., Müller, S., de Castilla, A.G., Smirnova, I., 2022. An open source COSMO-RS implementation and parameterization supporting the efficient implementation of multiple segment descriptors. Fluid Phase Equilib. 560. https://doi.org/10.1016/j.fluid.2022.113472

Gerlach, T., Müller, S., Smirnova, I., 2018. Development of a COSMO-RS based model for the calculation of phase equilibria in electrolyte systems. AlChE J. 64. https://doi.org/10.1002/aic.15875

Grensemann, H., Gmehling, J., 2005. Performance of a conductor-like screening model for real solvents model in comparison to classical group contribution methods. Ind. Eng. Chem. Res. 44 (5), 1610–1624. https://doi.org/10.1021/ie049139z

González de Castilla, A., Müller, S., Smirnova, I., 2021. On the analogy between the restricted primitive model and capacitor circuits: semi-empirical alternatives for over- and underscreening in the calculation of mean ionic activity coefficients. J. Mol. Liq. 326. https://doi.org/10.1016/j.molliq.2020.115204

González de Castilla, A., Müller, S., Smirnova, I., et al. On the analogy between the restricted primitive model and capacitor circuits. Part II: a generalized Gibbs–Duhem consistent extension of the Pitzer–Debye-hückel term with corrections for low and variable relative permittivity 360, 119398. https://www.sciencedirect.com/science/article/pii/S0167732222009369. https://doi.org/10.1016/j.molliq.2022.119398

Guo, H., Gong, M., Dong, X., Wu, J., 2012. (Vapour + liquid) equilibrium data for the binary system of {trifluoroiodomethane (R13I1) + *trans*-1, 3, 3, 3-tetrafluoropropene (R1234ze (E))} at various temperatures from (258.150 to 298.150)K. J. Chem. Thermodyn. 47. https://doi.org/10.1016/j.jct.2011.11.024

Hsieh, C.M., Lin, S.T., Vrabec, J., 2014. Considering the dispersive interactions in the COSMO-SAC model for more accurate predictions of fluid phase behavior. Fluid Phase Equilib. 367, 109–116. https://doi.org/10.1016/j.fluid.2014.01.032

Hudson, G.H., McCoubrey, J.C., 1960. Intermolecular forces between unlike molecules. A more complete form of the combining rules. Trans. Faraday Soc. 56. https://doi.org/10.1039/tf9605600761

Iwai, Y., Imamura, Y., Shimoyama, Y., 2010. Test of athermal terms of activity coefficient models by Monte Carlo simulation with hard-core models. Fluid Phase Equilib. 297 (2), 221–226. https://doi.org/10.1016/j.fluid.2010.05.008

Jaubert, J.-N., Le Guennec, Y., Piña-Martinez, A., Ramirez-Velez, N., Lasala, S., Schmid, B., Nikolaidis, I.K., Economou, I.G., Privat, R., 2020. Benchmark database containing binary-system-high-quality-certified data for cross-comparing thermodynamic models and assessing their accuracy. Ind. Eng. Chem. Res. 59 (33). https://doi.org/10.1021/acs.iecr.0c01734

Jiřiště, L., Klajmon, M., 2022. Predicting the thermodynamics of ionic liquids: what to expect from PC-SAFT and COSMO-RS? J. Phys. Chem. B 126 (20), 3717–3736. https://doi.org/10.1021/acs.jpcb.2c00685

KDB, 2024. CHERIC, Chemical Engineering and Materials Research Information Center. https://www.cheric.org/research/kdb.

Kim, J.H., Kim, M.S., Kim, Y., 2003. Vaporliquid equilibria for pentafluoroethane + propane and difluoromethane + propane systems over a temperature range from 253.15 to 323.15 K. Fluid Phase Equilib. 211 (2). https://doi.org/10.1016/S0378-3812(03)00237-1

Klajmon, M., 2022. Purely predicting the pharmaceutical solubility: what to expect from PC-SAFT and COSMO-RS? Mol. Pharm. 19 (11), 4212–4232. https://doi.org/10.1021/acs.molpharmaceut.2c00573

Klamt, A., 1995. Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena starting from the question of why dielectric continuum models give a fairly good description of molecules. Technical report. https://pubs.acs.org/sharingguidelines.

Klamt, A., 2005a. Chapter 2—Dielectric continuum solvation models and COSMO. In: Klamt, A. (Ed.), COSMO-RS. Elsevier, Amsterdam, pp. 11–41. https://doi.org/10.1016/B978-044451994-8/50002-X

Klamt, A., 2005b. Chapter 5—Statistical thermodynamics of interacting surfaces. In: Klamt, A. (Ed.), COSMO-RS. Elsevier, Amsterdam, pp. 59–81. https://doi.org/10.1016/B978-044451994-8/50005-5

Klamt, A., 2005c. Chapter 7—Refinements, parameterization, and the complete COSMO-RS. In: Klamt, A. (Ed.), COSMO-RS. Elsevier, Amsterdam, pp. 109–125. https://doi.org/10.1016/B978-044451994-8/50007-9

Klamt, A., Eckert, F., 2000. COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids. Technical report. https://www.elsevier.nlrlocaterfluid.

Klamt, A., Eckert, F., 2007. Prediction, fine tuning, and temperature extrapolation of a vapor liquid equilibrium using COSMOtherm. Fluid Phase Equilib. 260 (2), 183–189. https://doi.org/10.1016/j.fluid.2007.07.055

Klamt, A., Eckert, F., Hornig, M., Beck, M.E., Brger, T., 2002. Prediction of aqueous solubility of drugs and pesticides with COSMO-RS. J. Comput. Chem. 23 (2), 275–281. https://doi.org/10.1002/jcc.1168

Klamt, A., Jonas, V., Bu, T., Lohrenz, J. C.W., 1998. Refinement and parametrization of COSMO-RS. Technical report. https://pubs.acs.org/sharingguidelines.

Klamt, A., Schüürmann, G., 1993. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. J. Chem. Soc., Perkin Trans. 2 (5). https://doi.org/10.1039/P29930000799

Kontogeorgis, G.M., Coutsikos, P., Tassios, D., Fredenslund, A., 1994. Improved models for the prediction of activity coefficients in nearly atherrnal mixtures Part I. Empirical modifications of free-volume models. Technical Report.

Kontogeorgis, G.M., Nikolopoulos, G.I., Fredenslund, A., Tassios, D.P., 1997. Improved models for the prediction of activity coefficients in nearly athermal mixtures 1 Part II. A theoretically-based GE-model based on the van der Waals partition function. Technical report.

Kouskoumvekaki, I.A., Michelsen, M.L., Kontogeorgis, G.M., 2002. An improved entropic expression for polymer solutions. Technical report.

Krooshof, G. J.P., de With, G., 2024. Gibbs probability entropy and its implication to combinatorial entropy models. Fluid Phase Equilib. 584, 114146. https://www.sciencedirect.com/science/article/pii/S0378381224001237. https://doi.org/10.1016/j.fluid.2024.114146

Krooshof, G. J.P., Tuinier, R., de With, G., 2019a. Dispersion activity coefficient models. Part 1: cubic equations of state. Fluid Phase Equilib. 501. https://doi.org/10.1016/j.fluid.2019.112275

Krooshof, G. J.P., Tuinier, R., de With, G., 2019b. Dispersion activity coefficient models. Part 2: perturbed chain equations of state. Fluid Phase Equilib. 502. https://doi.org/10.1016/j.fluid.2019.112286

Kröger, L.C., Müller, S., Smirnova, I., Leonhard, K., et al. Prediction of solvation free energies of ionic solutes in neutral solvents 124 (20), 4171–4181. Publisher: American Chemical Society. https://doi.org/10.1021/acs.jpca.0c01606

Lacher, J.R., Hunt, R.E., 1941. Vapor pressures of binary liquid mixtures *. J. Am. Chem. Soc. 63 (6), 1752–1756. https://doi.org/10.1021/ja01851a070

Lee, B.S., Lin, S.T., 2017. Prediction and screening of solubility of pharmaceuticals in single- and mixed-ionic liquids using COSMO-SAC model. AlChE J. 63 (7), 3096–3104. https://doi.org/10.1002/aic.15595

Lin, S.T., Sandler, S.I., 2002. A priori phase equilibrium prediction from a segment contribution solvation model. Ind. Eng. Chem. Res. 41 (5), 899–913. https://doi.org/10.1021/ie001047w

Lo Nostro, P., Scalise, L., Baglioni, P., 2005. Phase separation in binary mixtures containing linear perfluoroalkanes. J. Chem. Eng. Data 50. https://doi.org/10.1021/je049620w

London, F., 1937. The general theory of molecular forces. Trans. Faraday Soc. 33. https://doi.org/10.1039/tf937330008b

Loschen, C., Klamt, A., 2014. Prediction of solubilities and partition coefficients in polymers using COSMO-RS. Ind. Eng. Chem. Res. 53 (28). https://doi.org/10.1021/ie501669z

Loschen, C., Klamt, A., 2016. Computational screening of drug solvates. Pharm. Res. 33 (11), 2794–2804. https://doi.org/10.1007/s11095-016-2005-2

Mambo-Lomba, D., Paricaud, P., 2021. Predictions of thermodynamic properties and phase equilibria of refrigerant systems with COSMO approaches. Int. J. Refrig. 124, 52–63. https://doi.org/10.1016/j.ijrefrig.2020.11.005

Marques, H., González de Castilla, A., Müller, S., Smirnova, I., 2023. Impact of extended long-range electrostatics on the correlation of liquid-liquid equilibria in aqueous ionic liquid systems. Fluid Phase Equilib. 569, 113765. https://www.sciencedirect.com/science/article/pii/S037838122300047X. https://doi.org/10.1016/j.fluid.2023.113765

Müller, S., González de Castilla, A., Taeschler, C., Klein, A., Smirnova, I., 2019. Evaluation and refinement of the novel predictive electrolyte model COSMO-RS-ES based on solid-liquid equilibria of salts and Gibbs free energies of transfer of ions. Fluid Phase Equilib. 483, 165–174. https://doi.org/10.1016/j.fluid.2018.10.023

Müller, S., González de Castilla, A., Taeschler, C., Klein, A., Smirnova, I., 2020. Calculation of thermodynamic equilibria with the predictive electrolyte model COSMO-RS-ES: improvements for low permittivity systems. Fluid Phase Equilib. 506. https://doi.org/10.1016/j.fluid.2019.112368

Müller, S., González de Castilla, A., Taeschler, C., Klein, A., Smirnova, I., 2020. Calculation of thermodynamic equilibria with the predictive electrolyte model COSMO-RS-ES: improvements for low permittivity systems. Fluid Phase Equilib. 506, 112368. https://www.sciencedirect.com/science/article/pii/S0378381219304297. https://doi.org/10.1016/j.fluid.2019.112368

Müller, S., Nevolianis, T., Garcia-Ratés, M., Riplinger, C., Leonhard, K., Smirnova, I., 2024. Predicting solvation free energies for neutral molecules in any solvent with openCOSMO-RS. https://arxiv.org/abs/2407.03434.

Neese, F., 2012. The ORCA program system. WIREs Comput. Mol. Sci. 2 (1), 73–78. https://doi.org/10.1002/wcms.81

Neese, F., Wennmohs, F., Becker, U., Riplinger, C., 2020. The ORCA quantum chemistry program package. J. Chem. Phys. 152 (22). https://pubs.aip.org/jcp/article/152/22/224108/1061982/The-ORCA-quantum-chemistry-program-package. https://doi.org/10.1063/5.0004608

Prausnitz, J.M., Lichtenthaler, R.N., de Azevedo, E.G., 1999. Molecular Thermodynamics of Fluid-Phase Equilibria. Prentice-Hall PTR. third ed..

RDKit2023.03.3, 2023. RDKit, version 2023.03.3. https://www.rdkit.org/.

Ritter, E., Racheva, R., Storm, S., Müller, S., Ingram, T., Smirnova, I., et al.. Influence of inorganic salts on the phase equilibrium of triton x-114 aqueous two-phase systems 61 (4), 1496–1501. https://doi.org/10.1021/acs.jced.5b00821

Rodríguez-Llorente, D., Hernández, E., Gutiérrez-Sánchez, P., Navarro, P., Ismael Águeda, V., Álvarez-Torrellas, S., García, J., Larriba, M., 2023. Extraction of pharmaceuticals from hospital wastewater with eutectic solvents and terpenoids: computational, experimental, and simulation studies. Chem. Eng. J. 451. https://doi.org/10.1016/j.cej.2022.138544

Saidi, C.N., Mielczarek, D.C., Paricaud, P., 2020. Predictions of solvation Gibbs free energies with COSMO-SAC approaches. Fluid Phase Equilib. 517. https://doi.org/10.1016/j.fluid.2020.112614

Schwerdtfeger, P., Nagle, J.K., 2019. 2018 table of static dipole polarizabilities of the neutral elements in the periodic table. Mol. Phys. 117 (9–12). https://doi.org/10.1080/00268976.2018.1535143

Silva, G.M.C., Pantano, D.A., Loehlé, S., Coutinho, J.A.P., 2023. The challenges of using COSMO-RS to describe polymer solution behavior. Ind. Eng. Chem. Res. 62 (48), 20936–20944. https://doi.org/10.1021/acs.iecr.3c03310

Soares, R.de P., 2011. The combinatorial term for COSMO-based activity coefficient models. In: Industrial and Engineering Chemistry Research. 50, pp. 3060–3063. https://doi.org/10.1021/ie102087p

Soares, R.de P., 2023. Jcosmo. https://www.ufrgs.br/lvpp/2023/09/22/jcosmo-download/.

Soares, R.de P., Gerber, R.P., 2013. Functional-segment activity coefficient model. 1. Model formulation. Ind. Eng. Chem. Res. 52 (32), 11159–11171. https://doi.org/10.1021/ie400170a

Staverman, A.J., 1950. The entropy of high polymer solutions. Generalization of formulae. Recl. Trav. Chim. Pays-Bas 69 (2), 163–174. https://doi.org/10.1002/recl.19500690203

Stone, A., 2013. The Theory of Intermolecular Forces. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199672394.001.0001

Van Ness, H.C., 1995. Thermodynamics in the treatment of vapor/liquid equilibrium (VLE) data. Pure Appl. Chem. 67 (6). https://doi.org/10.1351/pac199567060859

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A. R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Yaws, C.L., Satyro, M.A., 2015. Vapor pressure—Organic compounds. In: The Yaws Handbook of Vapor Pressure. Elsevier. https://doi.org/10.1016/B978-0-12-802999-2.00001-5