The Challenges of Using COSMO-RS To Describe Polymer Solution Behavior

Gonçalo M. C. Silva, Diego A. Pantano, Sophie Loehlé, and João A. P. Coutinho*

ABSTRACT: This work discusses the ability and limitations of the conductor-like screening model for realistic solvation (COSMO-RS) to predict the thermodynamic properties of polymer solutions over the entire concentration range. For that purpose, aqueous mixtures of polyethylene glycol (PEG) are used as a case study. The activity coefficients of PEG + water mixtures were calculated as functions of molecular weight and temperature. The influence of the conformations of the solvated polymer molecules, the inclusion of a combinatorial term, and the description of the volume dependence were assessed. Molecular dynamics simulations of water-solvated polymer molecules were performed to obtain representative polymer conformers. COSMO-RS is shown to be able to quantitatively predict the activity coefficients of these systems with the inclusion of a combinatorial term when the conformational and density dependence are well described. It is also shown that the scaleup of the polymer to large molecular weights can be achieved from the sigma profiles of oligomers with low molecular weight.

1. INTRODUCTION

The conceptualization of the conductor-like screening model for realistic solvation (COSMO-RS)\textsuperscript{1,2} has revolutionized the way chemical compounds can be screened by offering a fast calculation method that requires simple inputs obtained from quantum calculations without the need for parametrization on a case-by-case basis, such as typical equations of state (EoS) or activity coefficient models. While conventional EoS can provide accurate thermodynamic data with a good description of the temperature/pressure dependence, they mostly rely on the availability of experimental data to fit the models, even in the case of group-contribution methods, which is a significant disadvantage, compared to COSMO-RS, in terms of predictive capability. COSMO-RS has been extensively used in recent years to predict thermodynamic properties of several types of systems, from simple organic compounds\textsuperscript{3} to more-complex ionic liquids\textsuperscript{4} or deep eutectic solvents.\textsuperscript{5} Nonetheless, COSMO-RS has not gained considerable traction with regard to the study of polymeric systems due to their complex nature; the high molecular weight and high polydispersity, the number of possible conformers, and the complexity of the solid amorphous/crystalline phases make them difficult to model. Moreover, due to these properties, obtaining reference experimental values for model validation can also be a significant hurdle when treating polymeric systems with theoretical models. Thus, COSMO-RS could be a valuable asset to obtain thermodynamic predictions, even in cases where experimental data are not available. The use of COSMO-RS for polymeric systems, however, does not come without its caveats; due to the high molecular weight of polymers, quantum chemical calculations required to obtain the σ-profiles can be computationally demanding, and, for this reason, some authors have employed methods to circumvent this limitation. For instance, methodologies where the σ-profile of large polymers can be obtained by scaling up smaller oligomers with the same repeating units have previously been used.\textsuperscript{6,7} The importance of using representative polymer conformations for the COSMO calculations cannot be overlooked. To address this issue, molecular dynamics (MD) simulations have been proposed to source realistic conformers as input for the quantum calculations, based on the radius of gyration and solvent-accessible surface area (SASA) distributions calculated from the trajectories.\textsuperscript{8,9}

Polyethers are a class of compounds with a wide range of industrially relevant applications such as the formulation of cosmetics, pharmaceuticals, and lubricants,\textsuperscript{10−12} among others. In particular, aqueous-based lubricants are an interesting area of research that aims to replace the common oil-based lubricants with more environmentally friendly ones. As such, the ability to predict the solubility of polyethers in water under...
different conditions is of utmost importance (operation conditions can vary widely in temperature and pressure), which allows one to narrow down the number of candidates to be tested in the laboratory to assess the systems with more-suitable properties. For their importance in industrial applications and also due to their relatively simple structure, we use polyethers as the subject of this study to demonstrate the importance of different aspects when performing COSMO-RS calculations for polymeric systems, including the inclusion of a combinatorial term, the description of the polymer volumetric data, the effect of conformations, and the scaling of molecular weight from the $\sigma$-profile of small oligomers. In this work, we investigate the use of MD simulations to obtain realistic conformers for COSMO-RS calculations and discuss the obtained activity coefficient results in light of the aforementioned topics.

2. METHODS

2.1. MD Simulation Details. Atomistic simulations were performed for polyether molecules (PEG400, PEG1000, and PEG4000) in water, using the Gromacs 2020.6 software. The SPC/E model was used for water and the OPLS-AA force field was used to model the polymer molecules. The equations of motion were calculated with the leapfrog algorithm, using a time step of 2 fs. The forces were calculated exactly within a cutoff distance of 1.4 nm. Beyond the cutoff, the PME algorithm was used to correct the electrostatic interactions and the missing pressure and energy components were added to the dispersive term. Bonds that contain H atoms were constrained using the LINCS algorithm. Random initial cubic simulation boxes were generated with PACKMOL. An energy minimization step was performed before an equilibration stage in the NPT ensemble, which was run at a temperature of 298.15 K for 2 ns. Following this step, production runs of 100 ns were carried out in the NpT ensemble, where the pressure (1 bar) and temperature (298.15 K) were kept constant with the Parrinello–Rahman barostat and the Nose–Hoover thermostat, using coupling constants of 10 and 0.5 ps, respectively. Visualization of the simulation trajectories was made with VMD 1.9.4.

2.2. COSMO-RS Calculations. COSMO-RS calculations were made with COSMOTHERM version 21.0 software, using the “BP_TZVP_21.crd” parametrization. The DFT quantum mechanics calculations to obtain the $\sigma$-profiles were made with Turbomole version 17 at the BP86/def-TZVP level of theory. The calculations were made with polymer geometries obtained from MD simulations. A fine mesh was used to more accurately sample the charge density inside cavities. Before the single-point calculation, a relaxation step using the BP86/def-SV(P) method was necessary to avoid unrealistic highly polarized regions. The calculations for higher-molecular-weight polymers were made by scaling the $\sigma$-profile of the repeating units (excluding the end-groups). In addition to the residual term, the Elbro combinatorial term was used to account for the asymmetry of molecular volumes to the entropy of mixing. The molar volumes used to calculate the Elbro combinatorial term were derived from experimental data and, in some cases, extrapolated to the desired temperatures when no data were available.

3. RESULTS AND DISCUSSION

3.1. The Choice of Conformers. 3.1.1. MD Simulations. The adequate choice of conformers used for COSMO calculations is of importance, as they can have a visible impact on the quality of the inputs in the form of the $\sigma$-profiles. For polymeric systems, the conformations are inherently linked to the affinity toward the solution media and can deviate considerably from gas-phase optimized geometries; thus, the obtention of representative conformers of the polymer chains is of particular importance. Due to the size and flexibility of the polymer chains, the number of possible conformers is too large to be assessed in a deterministic way. Moreover, the use of optimized geometries or heuristic generation of conformers, such as with COSMOconf, might not be suitable to obtain realistic backbone structures of large and flexible molecules. However, atomistic MD simulations for the targeted mixtures can be a viable option to extract conformers that resemble the real molecules, as has been recently demonstrated by Zhou et al. Nevertheless, this approach has been used with extensive sampling of several geometries to construct the input for COSMO calculations. In this work, the effect of molecular geometries is assessed for a system of poly(ethylene glycol) and water. First, the effect of polymer concentration was investigated by performing MD simulations of PEG 1000 at two different concentrations: at infinite dilution ($W_{\text{PEG1000}} \approx 0$) and $W_{\text{PEG1000}} = 0.2$. The radius of gyration of PEG 1000 at the two concentrations and the corresponding simulation snapshots are presented in Figure 1. The calculated average radii of gyration were 0.9 nm at both infinite dilution and $W_{\text{polymer}} = 0.2$, suggesting that the radius of gyration is independent of the polymer concentration. The most obvious distinction in amplitude of the fluctuations between the two curves is simply related to the differing number of molecules within the simulation boxes (1 vs 300 molecules). Given this evidence, it seems that simulations at infinite dilution are sufficient to obtain realistic conformers for COSMO-RS calculations, which is convenient in order to reduce the computational time required for the calculations.

![Figure 1. Radius of gyration ($R_g$) of PEG 1000 at two different concentrations (blue denotes infinite dilution; orange denotes $W_{\text{polymer}} = 0.2$) calculated throughout the simulation length. The final MD configurations are also represented at the bottom: at infinite dilution (left) and for $W_{\text{polymer}} = 0.2$ (right). The water molecules have been omitted for added clarity.](https://doi.org/10.1021/acs.iecr.3c03310)
The analysis of conformers, presented in Figure 2, is based on the methodology described by Zhou et al., where the combined probability distributions of the $R_g$ and SASA values were plotted to assess the most representative conformers. This plot is represented for the case of PEG400, where the values of $R_g$ and SASA have been normalized. In conjunction, a few chosen conformers, corresponding to extreme cases for the calculated radius of gyration or SASA and also a conformer that corresponds to a state of highest probability, are indicated in a $\sigma$-surface representation. The choice of conformers, including some extreme cases, are meant to assess the influence of the geometry with the properties calculated using COSMO-RS, namely the activity coefficients. This analysis is discussed in detail in section 3.1.2. As expected, there is a clear correlation between the calculated $R_g$ and SASA, where low values correspond to more-spherical conformers and high values correspond to more-extended molecules.

### 3.1.2. COSMO-RS Calculations

The impact of the conformers chosen on the COSMO-RS calculations is assessed in this section for the case of PEG400. The activity coefficients obtained for five different conformers, four of them obtained from the MD simulation trajectories (and depicted in Figure 2) and one corresponding to the optimized geometry, are presented in Figure 3. The results show a considerable dependence on the conformation for the activity coefficients of both the solvent and polymer. By comparison with experimental results, the choice of an all-trans optimized geometry seems to be the worst approximation, providing a description of the polymer and solvent that are worse than any of the conformers extracted from the simulation. Both water

---

Figure 2. Solvent-accessible surface area (SASA) vs radius of gyration ($R_g$) for PEG400 at 298.15 K and infinite dilution in water. The $\sigma$-surface of the chosen conformers for this work is represented on the right side.

Figure 3. Activity coefficients of water (left) and PEG400 (right) for different conformers at 298.15 K. The color and numbering scheme for the conformers follows what is shown in Figure 2. Full lines correspond to the total COSMO-RS activity coefficients (residual + combinatorial terms), and dashed lines represent the residual contribution. Experimental data (open circles) was sourced from the work of Ninni et al.
and polymer activity coefficients show that the affinity for water is decreased in the all-trans optimized conformation, displaying values above unity in considerable concentration ranges, indicating that phase separation occurs at room temperature, which does not correspond to the experimental evidence. The all-trans conformer thus seems like a coarser approximation to the real molecules, when compared to geometries extracted from MD trajectories, and is not suitable to obtain accurate predictions from the COSMO-RS calculations.

 Depending on the morphology of the MD simulation conformers, some differences are also observed regarding their calculated activity coefficients. In particular, conformers 1 and 2, which correspond to high/low values of SASA and $R_H$, respectively, display very similar activity coefficients for both the polymer and solvent. This is not surprising as the two conformers have very similar morphologies in an extended geometry. Certainly, more surprising is the fact that, of all of the tested conformers, these are the ones that further deviate from the results obtained for the all-trans optimized geometry. Nonetheless, the extended arrangement of the PEG molecules is likely to contribute to more accessible association centers along the molecule, which explains the low activity coefficients. The remaining conformers 3 and 4, corresponding to the maximum probability in the SASA versus $R_H$ and the lowest value of $R_H$/SASA, respectively, also display similar results between them. These two conformers are significantly more curled when compared to conformers 1 and 2 and, following the same reasoning, the lower accessibility of the charged ethoxy or hydroxyl groups likely contributes to slightly larger activity coefficients. Nonetheless, the differences observed between the two sets of MD conformers are not very significant, especially when compared to the optimized geometry. The small differences can be justified by the available polar regions at the surface of the molecules, which is visible in the $\sigma$-profiles (see the Supporting Information) and, in turn, in the residual contribution to the activity coefficients, but a small contribution can also be attributed to the combinatorial term, where very small differences in the calculated COSMO-RS volume can be observed. When compared with experimental data (only found for water activity coefficients), the agreement for all MD conformers is very satisfactory overall. As such, it seems like the use of a single conformer from MD simulations might be enough to get satisfactory COSMO-RS predictions and should be a better approximation than using optimized geometries.

### 3.2. The Combinatorial Term

A key contribution to the description of the mixing properties of polymer solutions lies in the combinatorial term, which should account for the asymmetry in space occupation between the solvent and solute. The Flory–Huggins (F-H) equation tackles the size asymmetry issue by taking into account the volume fractions of the compounds. However, due to the difference in intrinsic hard-core volume between polymer and solvent, this description can be insufficient in some situations where the free volumes of the compounds differ significantly. In order to solve this issue, Elbro et al. modified the F-H equation by accounting for the inaccessible volume occupied by the hard core of the molecules, which resulted in a significant improvement over the F-H model. The combinatorial term proposed by Elbro et al., written for the chemical potential of compound $i$, is shown below:

$$\mu_i^{FV} = RT \left[ \ln \frac{\phi_i^{FV}}{x_i} + 1 - \phi_i^{FV} \right]$$

where $x_i$ is the molar fraction of compound $i$ and $\phi_i^{FV}$ is its free volume fraction, which can be written as follows:

$$\phi_i^{FV} = \frac{x_i \bar{V}_i - r_i}{\sum_j x_j \bar{V}_j - r_j}$$

where $\bar{V}_i$ is the molar volume of compound $i$ and $r_i$ is the hard-core volume. Further modifications have been introduced, such as the one proposed by Kouskoumvekaki et al., where the hard-core volume is empirically fitted with a scaling factor in order to optimize the agreement with a dataset of activity coefficients for polymer solutions. This approach, besides not being as elegant as the ones proposed by Flory and Huggins or Elbro et al., can lead to free volumes that are too sensitive to the value of the molar volume and can even result in...
unphysical outcomes such as “negative” free volumes. In general, the Elbro equation seems to be a balanced and robust solution to approximate the combinatorial contribution to the free energy, as previously mentioned by Loschen and Klamt. The estimation of the molar volume, $V_i$, or the density of the pure compounds, is more important than it might seem at first glance, because it can have a significant impact on the calculated free volume in the Elbro equation. The use of predictive methods such as the QSPR model, while convenient, can bring significant issues into the calculation. One issue with the QSPR estimated densities is its dependence on the molecule geometry, which can lead to deviations of $\sim 5\%$ between the highest and lowest values for the conformers of PEG400 tested in this work. An even more significant problem arises from scaling up the polymer molecular weight, which can result in densities that are unrealistically low. After some studies not detailed in this work, we realized that, whenever possible, the use of experimental data, even if extrapolated, is always preferred over any estimation method. Alternatively, the use of EoS to predict the molar volumes can be also a viable alternative. As we will show in the upcoming sections, an accurate description of the densities of both solvent and solute is essential to obtain correct trends as a function of polymer molecular weight. For now, we will focus on showcasing the impact of the density value at a constant temperature and MW.

The activity coefficients of PEG400 using the QSPR method to estimate the densities are listed in Figure 4. The curves presented also compare the impact of using (solid lines) or not using (dashed lines) the combinatorial term in the predictions of the activity coefficients, as previously shown in Figure 3. The impact is clear: when the combinatorial entropic term, which always contributes to lowering the activity coefficients, is absent, the results are unrealistic and incompatibility between the compounds (i.e., positive deviations to ideality with values that could result in phase separation) is observed. As such, the entropic contribution associated with mixing compounds must be included, in addition to the enthalpic residual term for all mixtures, in order to obtain reasonable and physically sound results.

In comparison with using experimental densities, a change in density that goes up to 6% in the worst case in relation to the experimental data will have an impact on the obtained results. The predicted densities are lower than the experimental densities, leading to an increase in the activity coefficients. The predictions deviate further from the experimental data (by excess for all conformers), and only two of the tested conformers would not display phase separation at room temperature. The results show that (i) the importance of the combinatorial term cannot be overstated and (ii) an accurate description of the densities is required.

3.3. The Residual Term. The residual term, responsible for the enthalpic contribution to the Gibbs free energy, can be
densities from the methodology is crucial to eliminate unreasonably high charge fine mesh were used or absent. As can be observed, this conformer (and activity coefficients are represented for the same quantum calculations is given in Figure 5, where the profiles correspond to reasonable screening charge densities as important as the combinatorial term to describe polymer solutions. In COSMO-RS, the residual term is calculated as the sum of energetic contributions that account for van der Waals, hydrogen bonding, and misfit interactions. The three contributions can be calculated as a function of the $\sigma$-profile, which contains the charge density information for each molecule, meaning that a correct description of the surface charge density is key to correctly model polymer solutions within COSMO-RS. We established in section section 3.1 that MD simulations can be a source of realistic conformers as input for the quantum calculations. However, certain precautions must be taken to ensure that the obtained $\sigma$-profiles correspond to reasonable screening charge densities and avoid overly polarized regions. As such, for each conformer extracted from the MD trajectory, a low-level DFT calculation on the conformers extracted from atomistic simulations: highly charged cavities are observed near ether compounds. The presence of an additional methyl group per monomer significantly impacts the structure of the polymeric chain. That is the case for PEG400 and PPG400, where, despite the structural similarity between the two polyether compounds, they behave very differently in the presence of water. Experimental observations at room temperature show that while PEG400 is fully miscible with water, PPG400 undergoes phase separation in the same solvent.

The calculated and experimental activity coefficients of PEG400 and PPG400 in water solutions are illustrated in Figure 6. It is interesting to see how COSMO-RS can differentiate between the two types of polymers, which behave very differently, despite the seemingly similar structure. The higher than unity activity coefficients for PPG400 indicate a tendency to phase separate, while the contrary is true for PEG400. The predictions are also in overall good agreement with the experimental data for the case of the water activity coefficients, following very similar trends. It is important to note that the distinction between the two polymers is seen if we isolate the residual term as well, which is indicative of the model being able to differentiate the polarities of the two compounds. The presence of an additional methyl group per monomer significantly impacts the $\sigma$-profile (see the Supporting Information), decreasing the availability of polar regions along the molecular surface, as indicated by the decrease in probability at $\sigma > 0.01$ e/Å$^2$. This example demonstrates that the COSMO-RS model can be very sensitive to small changes in the structure, making it a viable option to screen polymers with various constitutions/structures.

3.4. Molecular Weight Scaling. Perhaps the most challenging task when modeling polymeric systems with the COSMO-RS model is capturing the effect of the molecular weight. For larger molecules, the quantum mechanics calculations required to obtain the $\sigma$-profiles become
increasingly more demanding and, at some point, impossible to perform with readily available hardware. In order to circumvent these limitations, the σ-profiles can be upscaling from smaller oligomers by applying a weighting factor to the atoms of the repeating core of the polymer chain. This approach allows performing calculations for high-molecular-weight polymers without the need to perform computationally expensive quantum calculations. This effect can be visualized in Figure 7, where the σ-profiles for different PEG molecular weights is presented.

![Figure 7. σ-profiles showing the effect of molecular weight scaling for PEG400, PEG1500, PEG4000, and PEG20000. Inset shows a closer look at the negatively charged region.](image)

As a noteworthy observation, while the lower polarity sections of the σ-profile are proportionally increasing with the molecular weight, the most positively charged regions (negative σ), corresponding mostly to the hydroxyl hydrogen of the terminal groups, are mostly independent of the molecular weight scaling but are, nonetheless, still present, as seen in the inset in Figure 7. The approach that we followed in this work does not neglect the influence of the terminal groups, which can have an impact especially on the calculation for small molecules but whose effect is negligible for long polymers. While the σ-profile is the only required input to calculate the residual part of the chemical potential, the combinatorial term requires the molar volume. As shown above in section 3.2, an accurate volumetric description is important to accurately describe the experimental trends in the activity coefficients. Since the longer PEG compounds are solid at room temperature, densities of the liquid phase were predicted by extrapolation from the data obtained for shorter polymers.

The COSMO-RS calculations are presented in Figure 8 (lines), together with the experimental datasets (circles), as a function of the polymer mass fraction. Experimental data show that the water activity coefficients are mostly independent of the polymer molecular weight. COSMO-RS is able to quantitatively predict this observation, as the activity coefficients show no dependence of the activity coefficients in relation to the polymer molecular weight. Conceptually, for the same mass fraction of the polymer, it is logical that the effect on the water interactions is mostly irrelevant. However, the same is not true for the polymer activity coefficients, as can be seen in Figure 8 (right). An increase in molecular weight leads to significantly lower PEG activity coefficients as a function of the mass fraction. Once again, the COSMO-RS is able to correctly predict the trends that arise from changing the polymer molecular weight and, to some extent, quantitatively predict the correct PEG activity coefficients. Still, some significant deviations are observed in the case of PEG35000, where the COSMO-RS consistently underpredicts the activity coefficients. It should be noted, however, that the data points presented were obtained via indirect calculations (eq S1 in the Supporting Information) from SLE diagrams and that each experimental point corresponds to a different melting temperature. Nonetheless, the theoretical calculations obtained from scaling the σ-profiles produce very satisfactory results, following the experimental observations with the correct trends and even in quantitative agreement, demonstrating the

![Figure 8. – Activity coefficients of water (left) and PEG (right) as a function of the polymer molecular weight. COSMO-RS data (lines) was calculated at 298.15K. The experimental γwater data was sourced from the work of Ninni et al. at 298.15K and the γPEG data was calculated from the SLE diagrams found in the works of Kuttich et al. (PEG15000 and PEG8000), Hager et al. (PEG4000) and Gosh et al. (PEG35000). The molar enthalpies of melting used to calculate (eq S1 in the Supporting Information) the activity coefficients from SLE data were taken from the works of Kou et al.](image)
potential of COSMO-RS to be used in polymer—solvent mixture calculations.

As a final note, despite the potential that we recognize in the COSMO-RS approach, we were unable to calculate the LLE diagrams with COSMOtherm, due to numerical limitations that result from the high-molecular-weight asymmetry between polymer and solvent, thus focusing the bulk of the discussion on the activity coefficients. Hopefully, this limitation will be addressed in future software revisions, which could allow the proliferation of COSMO-RS as a method for the prediction of thermodynamic data for polymeric systems.

4. CONCLUSIONS

In this work, we discussed the details concerning polymer solution calculations with the COSMO-RS theory using the PEG family as a case study. Several factors were evaluated, such as the influence of conformations, the description of the combinatorial/residual terms, and the upscaling with molecular weight.

The use of realistic polymer conformers, sourced from atomistic MD simulations, resulted in significant improvements over the quantum mechanics optimized geometries. Nonetheless, due to the total or partial absence of intramolecular interactions between neighboring atoms in most force fields, certain precautions must be addressed in order to avoid highly polarized regions. A low-level energy minimization optimization step followed by a single-point calculation using the fine-cavity option led to realistic σ-profiles and good agreement between the COSMO-RS predictions and experimental data. Moreover, a good description of the combinatorial term was shown to be essential to predict the activity coefficients with accuracy. In this regard, the use of experimental data is preferred over predictive methods, such as QSPR, even if some data extrapolation is necessary.

The molecular weight scaling method, applied to the repeating units of an oligomer, allows the use of the COSMO-RS model without performing computationally intensive calculations for very large molecules. This methodology proved to be accurate in predicting the activity coefficients of water + polymer mixtures, with the right trends and, to a certain extent, in quantitative agreement when compared with experimental data. Overall, COSMO-RS can be a very good predictive tool for polymeric systems as long as some details inherent to the calculations discussed in this work are correctly addressed, opening perspectives for future work to scan the phase behavior of different polymer families.

ASSOCIATED CONTENT

Supporting Information
The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.iecr.3c03310.

Additional σ-profiles (PDF)

AUTHOR INFORMATION

Corresponding Author
João A. P. Coutinho — CICECO, Departamento de Química, Universidade de Aveiro, 3810-193 Aveiro, Portugal; orcid.org/0000-0002-3841-743X; Phone: +351 234 370 200; Email: jcoutinho@ua.pt

Authors
Gonçalo M. C. Silva — CICECO, Departamento de Química, Universidade de Aveiro, 3810-193 Aveiro, Portugal
Diego A. Pantano — TotalEnergies OneTech, 92400 Courbevoie, France
Sophie Loché — TotalEnergies OneTech, 92400 Courbevoie, France

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.iecr.3c03310

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors acknowledge the funding by TotalEnergies OneTech Solaize Research Center through a research partnership. This work was partly developed within the scope of the project CICECO-Aveiro Institute of Materials (Nos. UIDB/50011/2020, UIDP/50011/2020 & LA/P/0006/2020), financed by national funds through the FCT/MCTES (PIDDAC).

REFERENCES