

This is a postprint of *J. Phys. Chem. A*, 2016, 120 (12), pp 2049–2056. The original article can be found under <http://pubs.acs.org/doi/abs/10.1021/acs.jpca.6b00757>

COSMO-RSC: 2nd Order Quasi-Chemical Theory

Recovering Local Surface Correlation Effects

A. Klamt^{1,2}

¹ *COSMOlogic GmbH&CoKG, Imbacher Weg 46, D-51379 Leverkusen, Germany*

² *Institute of Physical and Theoretical Chemistry, University of Regensburg, Germany*

* Tel.: +49 2171 731681, e-mail: klamt@cosmologic.de

Abstract

The conductor-like screening model for realistic solvation (COSMO-RS) was introduced 20 years ago and meanwhile has become an important tool for the prediction of fluid phase equilibrium properties. Starting from quantum chemical information about the surface polarity of solutes and solvents, it solves the statistical thermodynamics of molecules in liquid phases by the very efficient approximation of independently pair-wise interacting surfaces, which meanwhile was shown to be equivalent to Guggenheim's quasi-chemical theory. One of the basic limitations of COSMO-RS, as of any quasi-chemical model, is the neglect of neighbor information, i.e. of local correlations of surface types on the molecular surface. In this paper we present the completely novel concept of using the first order COSMO-RS contact probabilities for the construction of local surface correlation functions. These are fed as an entropic correction for the pair interactions into a second COSMO-RS self-consistency loop, which yields new contact probabilities, enthalpies, free energies and activity coefficients recovering much of the originally lost neighbor effects. By a novel analytic correction for concentration dependent interactions the resulting activity coefficients remain exactly Gibbs-Duhem consistent. The theory is demonstrated on the example of a lattice Monte-Carlo fluid of dimerizing pseudo-molecules. In this showcase the strong deviations of the lattice Monte-Carlo fluid from quasi-chemical theory are almost perfectly reproduced by COSMO-RSC.

Introduction

The conductor-like screening model for realistic solvation (COSMO-RS)¹⁻³ originally presented in 1995 starts from quantum-chemical calculations of molecules embedded in a virtual conductor, employing the conductor-like screening model (COSMO).⁴ In a second step these polarization charge densities are used for the quantification of the intermolecular electrostatic and hydrogen-bond interactions. In a final step these local surface interaction energies are turned into free energies and enthalpies of fluids, using a specially derived thermodynamic self-consistency equation, which later was published in a slightly more general form under the name COSMOSPACE (COSMO surface pair activity coefficient equations).⁵

COSMO-RS meanwhile has become one of the most important and accurate tools for the prediction of free energies of solvation and important fluid phase equilibrium properties as solubilities, vapor pressures, activity and partition coefficients.⁶

A few years after the publication of COSMO-RS it was shown⁷ that the thermodynamic model of interacting surface segments derived in COSMO-RS directly from statistical thermodynamics, i.e. COSMOSPACE, is equivalent to an efficient and exact solution of Guggenheim's quasi-chemical theory (QUAC)⁸, directly yielding Gibbs-Duhem consistent activity coefficients. Usually the QUAC equations are either solved by the less efficient variational optimization of all contact probabilities, as in GEQUAC⁹, or by additional approximations as in UNIQUAC¹⁰ and UNIFAC¹¹. In 1986 Larsen and Rasmussen¹² had published an algorithm for solving the quasi-chemical equation systems, which is completely equivalent to the COSMOSPACE equations, but they did not recognize that these equations directly lead to free energies and activity coefficients. In 2001 Lin and Sandler re-derived the COSMO-RS algorithm under the name COSMO-SAC (COSMO segment activity coefficients)¹³.

Despite its considerable success, one of the fundamental limitations of COSMO-RS, as of all quasi-chemical models, is the almost complete neglect of effects resulting from the spatial neighborhood of different segments on the molecular surface. For example, the dimerization of acetic acid diluted in non-polar solvents is out of reach for such models, because in the quasi-chemical approximation the hydrogen bond donors and acceptors of the carboxylic acid group would form individual strong

contacts, leading to a 4th power dependence of the dimer concentration with respect to the concentration of the acid, while in reality the first contact between two acid molecules will induce a high probability for the formation of the second hydrogen bond, which then leads to a 2nd order concentration dependence of the dimer concentration. Indeed, recently an extension of COSMO-RS for dimerization, association, and reaction equilibria (COSMO-RS-DARE) has been presented¹⁴, which is able to describe such phenomena if the dimers or association species are introduced as special species into the COSMO-RS ensemble. But this extension requires considerable special parameterization and it is not extendable to less specific surface correlation phenomena. Other cases in which the lack of surface correlation may cause inaccuracies in COSMO-RS are micro phase separations, as they are observed in amphiphilic liquids as octanol, in which the alkyl chains and the polar groups tend to preferentially cluster, even before micellization occurs.

In this paper we present a novel theory for taking into account local correlations of surface polarities in a much more systematic and almost parameter-free way. This 2nd order COSMO-RS, which we will denote as COSMO-RSC further on, uses the contact probabilities of surface segments derived in a first COSMO-RS cycle for the quantification of the induced local concentrations of surface types or polarization charge densities due to the neighbor segments of contacting neighbor segments. These induced local concentrations then are converted into an entropic correction for the pair-wise interaction of segments. Then a second COSMO-RS loop is performed, now utilizing the local concentration corrected interaction (free) energies, resulting in modified contact probabilities, which reflect the local correlations. This easily leads to corrected interaction enthalpies of the liquid ensemble, while more effort is required in order to get thermodynamically consistent free energies. This is because the use of concentration dependent interaction operators, as the concentration dependent correlation corrections, in the COSMOSPACE algorithm lead to Gibbs-Duhem inconsistent free energies and activity coefficients. A theoretical analysis and a resulting correction formula for this problem has already been developed some time ago. Using this correction, we end up in COSMO-RSC with exactly Gibbs-Duhem consistent activity coefficients.

We demonstrate and test the new theory using an artificial lattice Monte-Carlo (LMC) liquid, which allows for a clearer analysis of the local surface correlations than real fluids would do. Applications to real liquids will be described in forthcoming papers.

Theory

Using the notations introduced in the COSMOSPACE paper, we consider an ensemble of molecules that represents an essentially incompressible liquid. We neglect the combinatorial free energy, since it is not of relevance for the aspects considered here, and only analyze the so-called "residual" contribution, which arises from the interactions between molecules. The basic assumption of surface pair interaction models is that residual - i.e., non-steric - interactions can be described by pairs of geometrically independent surface segments.

We consider a system having ν different types of surface. The total number of surface segments n_i on a molecule i is given by

$$n_i = \sum_{\nu} n_i^{\nu} = \frac{q_i}{a_{eff}} \quad (1)$$

where q_i is the total surface area of molecule i and n_i^{ν} denotes the numbers of segments of type ν on molecule i . a_{eff} is the size of a thermodynamically independent segment, which in all COSMO-RS parameterizations is in the range of 7 \AA^2 .^{2,3,13}

If $N_i = N x_i$ is the number of molecules of species i in the system, then the total number, n , of segments in the system is given by

$$n = \sum_i N_i n_i \quad (2)$$

In the same way, the number of segments of type ν is given by

$$n^{\nu} = \sum_i N_i n_i^{\nu} \quad (3)$$

The relative number or surface fraction of segments of type ν is

$$\Theta^{\nu} = \frac{n^{\nu}}{n} \quad (4)$$

From elementary statistical thermodynamics we know that for an incompressible system the Gibbs free energy of the system is

$$G = -k_B T \ln Z \quad (5)$$

where Z is the total partition function of this system. The chemical potential of species i in the mixture is given by

$$\mu_i = \frac{\partial G}{\partial N_i} = -k_B T \frac{\partial \ln Z}{\partial N_i} \quad (6)$$

In view of our earlier assumption that the partition function Z depends only on the segment composition we have

$$\mu_i = -k_B T \frac{\partial \ln Z}{\partial N_i} = -k_B T \sum_{\nu} \frac{\partial \ln Z}{\partial n^{\nu}} \frac{\partial n^{\nu}}{\partial N_i} = -k_B T \sum_{\nu} \frac{\partial \ln Z}{\partial n^{\nu}} n_i^{\nu} = \sum_{\nu} n_i^{\nu} \mu^{\nu} \quad (7)$$

where μ^{ν} is the pseudo-chemical potential of a segment of type ν in the ensemble of interacting surface segments. The activity coefficient, γ_i , of species i is defined by

$$k_B T \ln \gamma_i = \mu_i - \mu_{ii} \quad (8)$$

where μ_{ii} denotes the chemical potential of compound i in a system with $x_i = 1$. In the same way we have

$$k_B T \ln \gamma^{\nu} = \mu^{\nu} - \mu^{\nu\nu} \quad (9)$$

for the activity coefficient of a segment of kind ν , with $\mu^{\nu\nu}$ being the segment chemical potential at $\Theta^{\nu} = 1$. If we define μ_i^{ν} and γ_i^{ν} to be the chemical potential and activity coefficient of a segment ν in an ensemble of pure compound i , respectively, and apply these definitions to Eq. 7, we obtain

$$\begin{aligned} \ln \gamma_i &= \frac{\mu_i - \mu_{ii}}{k_B T} = \sum_{\nu} n_i^{\nu} \frac{\mu^{\nu} - \mu_i^{\nu}}{k_B T} \\ &= \sum_{\nu} n_i^{\nu} \frac{(\mu^{\nu} - \mu^{\nu\nu}) - (\mu_i^{\nu} - \mu_i^{\nu\nu})}{k_B T} = \sum_{\nu} n_i^{\nu} (\ln \gamma^{\nu} - \ln \gamma_i^{\nu}) \end{aligned} \quad (10)$$

for the activity coefficient of compound i . By Eq. 10 we have reduced the thermodynamics of the system of chemical compounds to the thermodynamics of an ensemble of pair-wise interacting surface segments.

Let μ and ν denote different kinds of surface segment and let

$$g_{\mu\nu} = h_{\mu\nu} - T s_{\mu\nu} \quad (11)$$

be the interaction energy of a pair $\mu\nu$. $g_{\mu\nu}$ in general may be a free energy of interaction of the two pairs, i.e. it may include enthalpic contributions $h_{\mu\nu}$ and entropic contributions $-Ts_{\mu\nu}$, while mostly in literature only enthalpic interactions are considered. We always assume symmetric interactions of the segments, i.e. $g_{\mu\nu} = g_{\nu\mu}$. The Boltzmann factor for the interaction energy of the segments is usually introduced as

$$\tau_{\mu\nu} = \exp\left\{-\frac{g_{\mu\nu}}{k_B T}\right\} \quad (12)$$

In the COSMOSPACE paper it was shown that a general and very efficient solution for the complete thermodynamics of such system can be achieved by solving the set of equation

$$\frac{1}{\gamma^\nu} = \sum_{\mu} \tau_{\mu\nu} \Theta^\mu \gamma^\mu \quad (13)$$

for the activity coefficients of segment types ν in the ensemble. In general this has to be solved numerically by iteration, usually starting from $\gamma^\mu = 1$ on the right hand side. In some cases, especially if only two different segment types are considered, analytic solutions can be derived. The efficiency of the COSMOSPACE equations results from the fact that via eq. 10 the segment activity coefficients do directly lead to the activity coefficients and thus to the free energies of the compounds. The contact probabilities of segments are given by

$$p_{\mu\nu} = \Theta^\mu \gamma^\mu \tau_{\mu\nu} \gamma^\nu \Theta^\nu \equiv \Theta^\mu \Psi_{\mu\nu} \Theta^\nu \quad (14)$$

Trivially the enthalpy of the system is then given as expectation value of the interaction enthalpy, i.e.

$$H = \frac{1}{2} \sum_{\mu\nu} h_{\mu\nu} p_{\mu\nu} \quad (15)$$

As proven in appendix B of the COSMOSPACE paper, Gibbs-Duhem consistency of the resulting activity coefficients is automatically guaranteed within COSMOSPACE. Nevertheless in the derivation of the COSMOSPACE equations it was implicitly assumed that the segment interaction energies do not depend on the composition of the liquid system. But for mole fraction dependent segment interaction energies $g_{\mu\nu}(x)$ the activity coefficients of the compounds calculated by COSMOSPACE (eq.10) are no longer Gibbs-Duhem consistent and hence thermodynamically wrong. This results

from the fact that the additional composition dependence of the total free energy, which goes beyond the composition dependence resulting from the contact probabilities, and which is caused by the composition dependence of the interactions, is not taken into account in the variational solution of the statistical thermodynamics. Nevertheless, at a fixed value of x , the contact probabilities calculated within the COSMOSPACE framework must still be correct, because at a fixed value of x we have fixed interaction energies and hence no deviation from the assumptions made in COSMOSPACE.

An exact correction for the originally thermodynamic inconsistent activity coefficients of COSMOSPACE in the presence of composition dependent interaction energies can be derived, if we write the total free energy of a quasi-chemical system in the fundamental way

$$G = \frac{N}{2} \sum_{\mu\nu} \Theta^\mu \Theta^\nu \Psi_{\mu\nu} [g_{\mu\nu} + \ln \Psi_{\mu\nu}] \quad (15)$$

as nicely described within the GEQUAC model.⁹ Applying the definition of the compound chemical potentials according to eq. 6

$$\begin{aligned} \mu_i &= \frac{\partial G}{\partial N_i} = \frac{\partial}{\partial N_i} \left(\frac{N}{2} \sum_{\mu\nu} \Theta^\mu \Theta^\nu \Psi_{\mu\nu} [g_{\mu\nu}(x) + \ln \Psi_{\mu\nu}] \right) \\ &= \mu_i^0 + \frac{N}{2} \sum_{\mu\nu} \Theta^\mu \Theta^\nu \Psi_{\mu\nu} \left[\frac{\partial g_{\mu\nu}(x)}{\partial N_i} \right] = \mu_i^0 + \frac{N}{2} \sum_{\mu\nu} P_{\mu\nu} \left[\frac{\partial g_{\mu\nu}(x)}{\partial N_i} \right] \equiv \mu_i^0 + \mu_i^{corr} \end{aligned} \quad (16)$$

where we have separated the contribution of G , which arises from the composition dependence of the interactions, into μ_i^{corr} and subsumed all other parts into μ_i^0 . Therefore μ_i^0 can be calculated as usual from COSMOSPACE, and μ_i^{corr} can be easily calculated as the expectation value of the derivatives of the interaction energies with respect to the particle number. Eq. 16 is by itself a considerable extension of COSMOSPACE models, because it allows for taking into account general composition dependencies of the segment interactions, as long as the derivatives of the segment interaction energies on the composition can be calculated.

With that we have the framework for working out the 2nd order COSMO-RS theory. Let us assume that our segments μ and ν have direct neighbor segments μ' and ν' . Due to the neighborhood of the segments a contact between μ' and ν' induces an

increased local concentration of v in the environment of μ and vice versa, leading to an increased contact probability of μ and v . Any effects and energy corrections resulting from such neighbor induced local concentration are not taken into account in COSMO-RS and COSMOSPACE, nor in any quasi-chemical model so far. If the segments on the molecular surface would be in a planar region and of quadratic shape, then geometrically $1/4$ of the $(\mu'v')$ -contacts would lead to a (μv) -contact. If we would have a regular hexagonal lattice, the geometric correlation coefficient would reduce to $1/6$. In the realistic case of segments on molecular surfaces, which have non-vanishing and irregular curvature, an exact evaluation of the induced local concentration of v in the surrounding of μ due to contacts of neighbor segments μ' and v' will not be possible. If we count each segment on each compound as a separate type of segments, approximate expressions of the form

$$\tilde{\Theta}_{\mu}^v = \sum_{\substack{\mu' \in i(\mu) \\ \mu' \neq \mu}} \sum_{\substack{v' \in j(v) \\ v' \neq v}} \Theta^{v'} \gamma^{v'} \tau_{\mu'v'} \gamma^{\mu'} f_{\mu'v'}^{corr} \quad (17)$$

can be constructed for the induced local concentration of type v in the vicinity of segment type μ , in which $i(\mu)$ and $j(v)$ denote the molecules i and j to which segments μ and v belong. The summations are over all ‘neighbor’ segments μ' and v' of μ and v , respectively, i.e. over all segments on the same molecules. According to COSMOSPACE, $\Theta^{v'} \gamma^{v'} \tau_{\mu'v'} \gamma^{\mu}$ is the probability that segment v' is partner of μ' . The local correlation function $f_{\mu'v'}^{corr}$ should decay rapidly with the distances $d(\mu'\mu)$ and $d(v'v)$ of the segments on the two molecules, as well as with the angles $\alpha(\mu'\mu)$ and $\alpha(v'v)$ between the normal vectors of the segment. Furthermore, it should be a kind of δ -function of the two distances, because only neighbor segments of μ and v at about the same distance can cause additional (μv) contacts. Thus a plausible functional form for $f_{\mu'v'}^{corr}$ may be

$$f_{\mu'v'}^{corr} \cong f_{corr}^0 \exp\left\{-\frac{d(\mu'\mu)^2}{d_{corr}^2}\right\} \exp\left\{-\frac{d(v'v)^2}{d_{corr}^2}\right\} \\ * \exp\left\{-\frac{\alpha(\mu'\mu)^2}{\alpha_{corr}^2}\right\} \exp\left\{-\frac{\alpha(v'v)^2}{\alpha_{corr}^2}\right\} \exp\left\{-\frac{|d(\mu'\mu) - d(v'v)|^2}{d_{tol}^2}\right\} \quad (18)$$

but the detailed and optimal expression for the correlation function is not relevant for the purpose of this paper. It will have to be worked out based on plausible

assumptions and fitting of a few empirical parameters. The relevant fact is that the induced local concentration $\tilde{\Theta}_\mu^v$ of v in the surrounding of μ is given by a sum of neighbor contact probabilities, which are known from the initial COSMOSPACE calculation, and a geometric function, which can be easily calculated from the segment coordinates stored in the COSMO files. Indeed, the intramolecular segment distances and normal vector angles can be precalculated and stored in each COSMO file, making the evaluation of the correlation functions more efficient. If, as usual in COSMO-RS, several segments, e.g. segments of similar polarization charge densities σ , are comprised to one segment type, the appropriate summations and averages of the induced local concentrations have to be calculated. But still the induced contact probabilities will be summations of COSMOSPACE contact probabilities multiplied by geometric correlation functions.

If we now introduce the sum of all induced local concentrations at μ as

$$\tilde{\Theta}_\mu^{tot} = \sum_v \tilde{\Theta}_\mu^v \quad (19)$$

we can define a locally corrected concentration of v in the vicinity of μ as

$$\hat{\Theta}_\mu^v = \frac{\Theta^v + \tilde{\Theta}_\mu^v}{1 + \tilde{\Theta}_\mu^{tot}} \quad (20)$$

assuming that the global concentration Θ^v of segment type v adds to the induced local concentrations. The denominator in eq. 20 results from normalization. Thus the ratio of the locally corrected concentration and the initially used global concentration of v in the vicinity of μ is given by the ratio of $\hat{\Theta}_\mu^v$ and Θ^v . If we convert this into an entropic correction for the contact probability of μ and v , we get

$$g_{\mu\nu}^{lcc} = -\frac{k_B T}{2} \left[\ln \left\{ \frac{\Theta^v + \tilde{\Theta}_\mu^v}{(1 + \tilde{\Theta}_\mu^{tot}) \Theta^v} \right\} + \ln \left\{ \frac{\Theta^\mu + \tilde{\Theta}_\nu^\mu}{(1 + \tilde{\Theta}_\nu^{tot}) \Theta^\mu} \right\} \right] \quad (21)$$

where the factor $1/2$ avoids double counting, because we are simultaneously correcting for both induced concentrations in order to get a symmetric free energy correction for the contacts of μ and ν .

Now we can introduce 2nd order interaction free energies

$$g_{\mu\nu}^* = g_{\mu\nu} + g_{\mu\nu}^{lcc} \quad (22)$$

and solve the COSMOSPACE equations again, yielding 2nd order segment activity coefficients γ_i^* and new contact probabilities $p_{\mu\nu}^*$ as well as a new total enthalpy, which will reflect the geometry induced local correlation of segment types. Since the $g_{\mu\nu}^{lcc}$ include the first order contact probabilities and by that depend on the composition of the system, the calculation of free energies and activity coefficients of the compounds in the system requires the correction term as given in eq. 16. Hence we finally have for the free energies of compound i in the system

$$\mu_i^* = \mu_i^0 + \mu_i^{*corr} = k_B T \sum_{v \in i} \ln \gamma_v^* + \frac{N}{2} \sum_{\mu\nu} p_{\mu\nu}^* \left[\frac{\partial g_{\mu\nu}^{lcc}}{\partial N_i} \right] \quad (23)$$

The $g_{\mu\nu}^{lcc}$ depend on the composition and by that on the particle number N_i only via the surface fractions and 1st order contact probabilities, and thus on the 1st order segment activity coefficients γ^μ . The analytic derivatives of the latter with respect to the particle numbers can be easily calculated, as described in detail in appendix D of the COSMOSPACE paper.

Thus we now have the complete 2nd order fluid phase thermodynamics of the system in our hands, which recovers local correlations of the surface segments on the molecular surfaces, which was ignored in the 1st order COSMO-RS and COSMOSPACE equations, as well as in any kind of quasi-chemical approach.

Model example

For the validation of the COSMO-RSC theory it is useful to consider a case in which the correlation effects can be clearly identified and quantified, and which is simple enough in order to construct the surface correlation function analytically. A rather realistic model systems can be provided by lattice fluids and lattice Monte-Carlo simulations. Such systems have already been considered during the original development of COSMO-RS, and later LMC calculations with cubic pseudo-molecules have been used for validation purposes in the COSMOSPACE paper.

In order to study substantial neighbor correlation effects the cubic molecules are not sufficient. Therefore we will consider in the following a lattice fluid consisting of N_c cubic pseudo-molecules, and N_t straight trimers molecules, which are just 3 cubes

stuck together linearly. Thus the cube molecules have 6 surface segments, and the trimers have 14 external surface segments. In general each of the 20 segments may be of different type ν . A lattice Monte-Carlo program trimer-LMC was written, which allows the specification of a polarization charge density σ_ν on each segment type ν , and the interaction energies are given by the generic functional form

$$g_{\mu\nu}^{LMC} = h_{\mu\nu}^{LMC} = \sigma_\nu \sigma_\mu \quad (24)$$

We consider a cubic simulation box of size L^3 (default $L = 20$) and apply periodic boundary condition. Since we want to consider a densely filled box, the number of cube molecules is given by

$$N_c = L^3 - 3N_t \quad (25)$$

The four types of LMC moves required in such system are quite simple:

- 1) random rotation of a cube into one of its 24 possible orientations.
- 2) random rotation by 90° or 180° of a trimer either around its long axis or rotation by 180° around one of the perpendicular axes.
- 3) selection of a trimer which has a cube at one of its head faces and exchange of the two molecules.
- 4) selection of a trimer and of three linearly neighboring cubes and exchange of the trimer with and the three cubes.

These LMC moves guarantee that all arrangements of the LMC mixture can be achieved. We performed $2 \cdot 10^7$ LMC steps for each composition and temperature and used the last half of the LMC steps for the evaluation of the average interaction enthalpy H_{LMC} . By the proper choice of polar and less polar surface segments, i.e. by the proper choice of the 20 types of polarization charge densities σ_ν on the faces of the cube and trimer molecules, many different systems can be mimicked, in which different types of neighbor correlation effects can be studied.

For the analytic validation of COSMO-RSC we consider a very simple but quite realistic case, mimicking a mixture of methane and acetic acid, where the pseudo-methane is represented by the cubes, with all surface segments being neutral (or green, if we use the standard color coding of COSMO-RS), and the acetic acid molecules are represented by trimer molecules, which have a strongly negatively polar surface segment representing a hydrogen bond acceptor (i.e. a red segment with positive σ) in the middle of one of the long sides, and a positively polar segment

representing a hydrogen bond donor (i.e. a blue segment with negative σ) directly adjacent to the red acceptor segment. The cubic and trimer molecules are schematically visualized in figure 1. LMC simulation have been performed for this cube-trimer mixture over a wide range of compositions expressed as mole fractions of trimers x_t and temperatures, expressed in units of σ^2/k_B , where σ is the polarization charge density of the positively polar segment on the trimer. The resulting interaction enthalpies $H_{LMC}(x_t, T)$ are shown in figure 2.

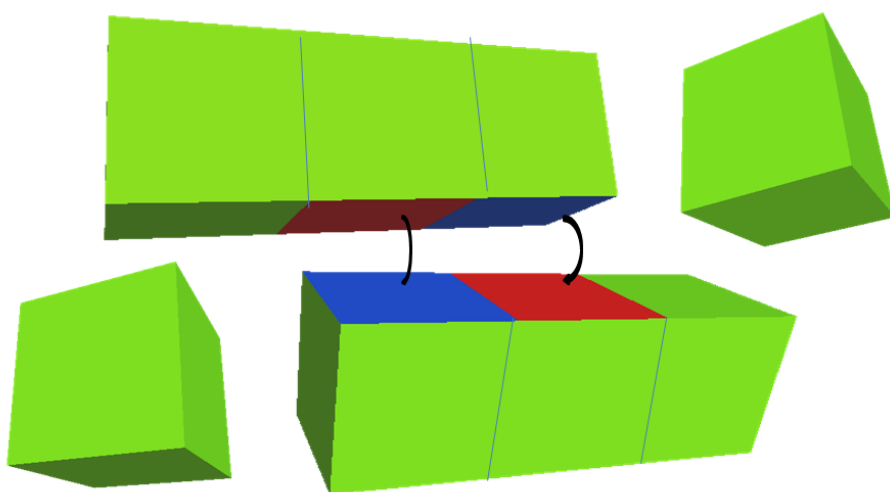


Figure 1: Non-polar green cube molecules and trimer molecules with adjacent oppositely polar surface segments. Due to their direct neighborhood the polar segments on the trimers can form a very favorable double interaction.

Next we perform the 2nd order COSMO-RS calculation for this system. First we have to solve the 1st order COSMOSPACE equations. Let us denote all green, non-polar surface segments as type A, and the donor and acceptor segments as types B and B'. Since the polarities of the donor and acceptor segments are symmetric, and since their surface fractions are always identical, their segment activity coefficients γ^B and $\gamma^{B'}$ must be identical. Because the green segments have $\sigma_A = 0$, the Boltzmann factors τ_{AA} , τ_{AB} and $\tau_{AB'} = 1$ are unity. Denoting the Boltzmann factor of a BB' contact as $\tau = \tau_{BB'}$, the Boltzmann factors for the like contacts are $\tau_{BB} = \tau_{B'B'} = 1/\tau$. Then the COSMOSPACE equations become:

$$1 = \Theta^A \gamma^{A^2} + \Theta^B \gamma^B \gamma^A + \Theta^{B'} \gamma^{B'} \gamma^A = \Theta^A \gamma^{A^2} + 2\Theta^B \gamma^B \gamma^A \quad (26)$$

and

$$1 = \Theta^{B'} \gamma^{B'} \gamma^B \tau^{-1} + \Theta^B \gamma^{B^2} \tau + \Theta^A \gamma^A \gamma^B = \Theta^B \gamma^{B^2} \tau^{-1} + \Theta^B \gamma^{B^2} \tau + \Theta^A \gamma^A \gamma^B \quad (27)$$

with $\theta^B = \theta^{B'} = (1-\theta^A)/2$. Solving eq. 26 for γ^B and inserting this into eq. 27 leads to a quadratic equation with respect to γ^{A^2} which can be solved by standard algebraic methods. Thus we yield the segment activity coefficients and we calculate the enthalpy $H_{\text{CRS1}}(x_t, T)$ of the system according to standard, 1st order COSMO-RS. As can be seen in figure 2, these first order results deviate substantially from the LMC results. The reason is the strong local correlation of the neighboring donor and acceptor segments, B and B'.

While the geometry of the trimer molecules also induces some correlation between A type segments (green), this will be very small, because type A segments are plentiful at all compositions of the system. The induced correlation of BB contact due to B'B' contacts can also be neglected, since like contacts of the polar segments have an extremely low probability. Therefore the effects of the correlation between B and B' due to neighboring B' and B contacts will clearly be dominating. For the sake of clarity and simplicity we will only consider the induced local concentration of this type. According to eq. 17, the induced local concentration of B' in the vicinity of B is

$$\tilde{\Theta}_B^{B'} = \frac{1}{4} \Theta^B \gamma^B \tau \gamma^{B'} = \frac{1}{4} \Theta^B \gamma^{B^2} \tau \quad (28)$$

Using eq. 20 and 21 we yield the local concentrations free energy corrections

$$\begin{aligned} g_{BB'}^{lcc} &= -\frac{k_B T}{2} \left[\ln \left\{ \frac{\Theta^{B'} + \tilde{\Theta}_B^{B'}}{(1 + \tilde{\Theta}_B^{B'}) \Theta^{B'}} \right\} + \ln \left\{ \frac{\Theta^B + \tilde{\Theta}_B^B}{(1 + \tilde{\Theta}_B^B) \Theta^B} \right\} \right] \\ &= -k_B T \left[\ln \left\{ \Theta^B + \tilde{\Theta}_B^{B'} \right\} - \ln \left\{ 1 + \tilde{\Theta}_B^{B'} \right\} - \ln \left\{ \Theta^B \right\} \right] \equiv -k_B T \left[\ln \{s\} \right] \end{aligned} \quad (29)$$

$$g_{BA}^{lcc} = -\frac{k_B T}{2} \left[\ln \left\{ \frac{\Theta^A}{(1 + \tilde{\Theta}_B^{B'}) \Theta^A} \right\} \right] = \frac{k_B T}{2} \ln \left\{ 1 + \tilde{\Theta}_B^{B'} \right\} \equiv -k_B T \ln \{r\} \quad (30)$$

and

$$g_{BB}^{lcc} = -\frac{k_B T}{2} \left[\ln \left\{ \frac{\Theta^B}{(1 + \tilde{\Theta}_B^{B'}) \Theta^B} \right\} + \ln \left\{ \frac{\Theta^{B'}}{(1 + \tilde{\Theta}_B^B) \Theta^{B'}} \right\} \right] = k_B T \ln \left\{ 1 + \tilde{\Theta}_B^{B'} \right\} = -k_B T \ln \{r^2\} \quad (31)$$

Using the shortcuts s and r defined in eqs. 29 and 30, the COSMOSPACE equations for the 2nd order COSMO-RS correction become:

$$1 = \Theta^A \gamma_*^{A^2} + 2r \Theta^B \gamma_*^B \gamma_*^A \quad (32)$$

$$1 = \Theta^B \gamma_*^{B^2} r^2 \tau^{-1} + \Theta^B \gamma_*^{B^2} s \tau + \Theta^A r \gamma_*^A \gamma_*^B \quad (33)$$

The segment activity coefficients γ_*^A and γ_*^B can now again be calculated analytically from eqs. 32 and 33 by standard algebra. From this we can easily calculate the 2nd order enthalpy $H_{\text{CRS2}}(x_t, T)$. This is shown in figure 2 together with the LMC enthalpy and the 1st order COSMO-RS result. As can be seen, $H_{\text{CRS2}}(x_t, T)$ is in good agreement with the LMC results. Especially at low temperatures, where correlation effects are most important, deviations between 1st order COSMO-RS are largest, the 2nd order COSMO-RS results agree very well with the LMC simulations. At higher temperatures the LMC simulations are almost half way between the 1st and 2nd order COSMO-RS results and the agreement appears to be less good on a logarithmic scale, but on an absolute scale the deviations here are very small.

We have added in figure 2 the results achieved by numerical COSMO-RSC calculations which take into account all correlations, and not just the correlation between the polar surface segments. These results agree perfectly with the analytical results at lower temperatures, while at high temperature ($T = 3$) these results slightly deviate from the simplified analytical model and are even closer to description of the LMC results.

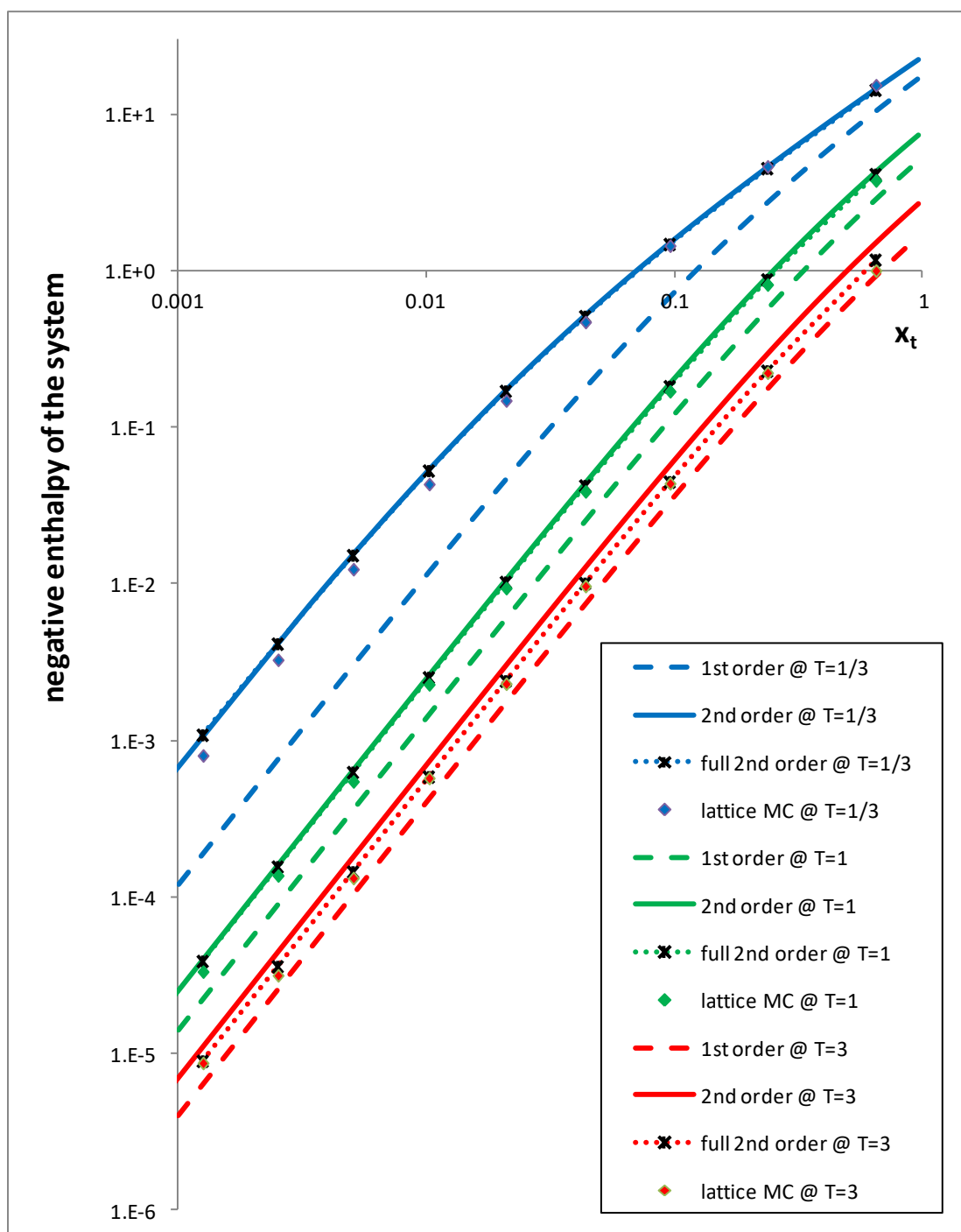


Figure 2: Comparison of molar enthalpies of the cube-trimer mixture calculated by lattice Monte Carlo simulations (solid symbols) and different level COSMO-RS simulations. The logarithmic horizontal axis denotes the negative of the molar enthalpy. The black crosses and dotted lines give the results of numerical COSMO-RSC calculations, taking into account all segment correlations.

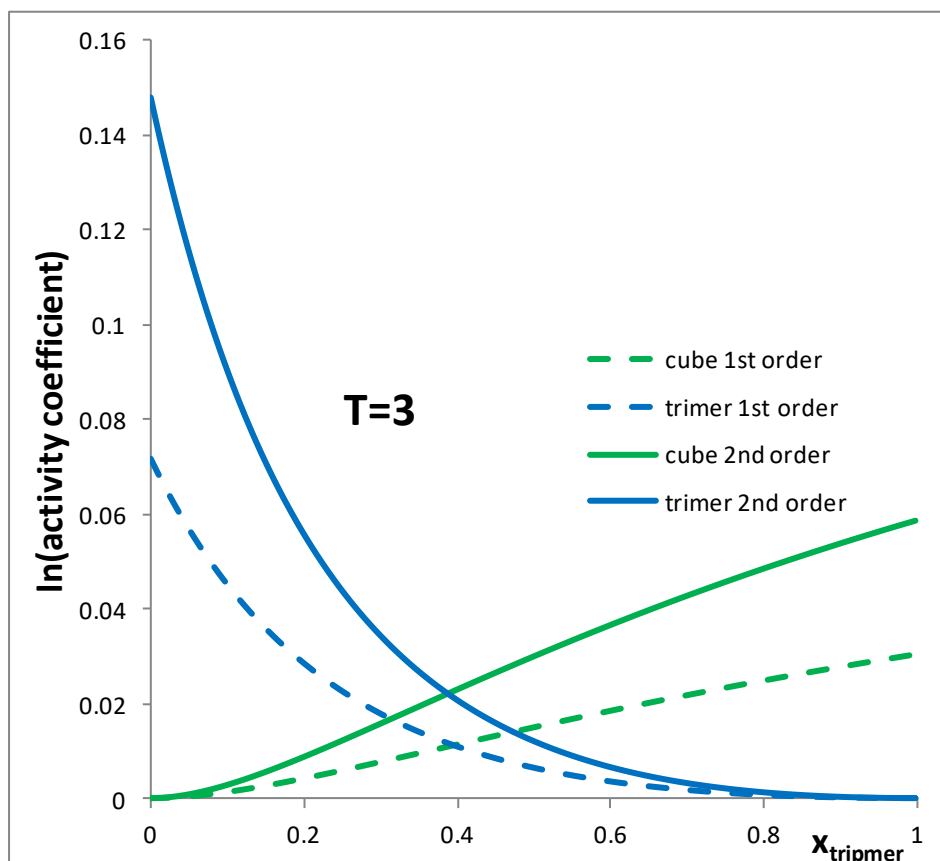
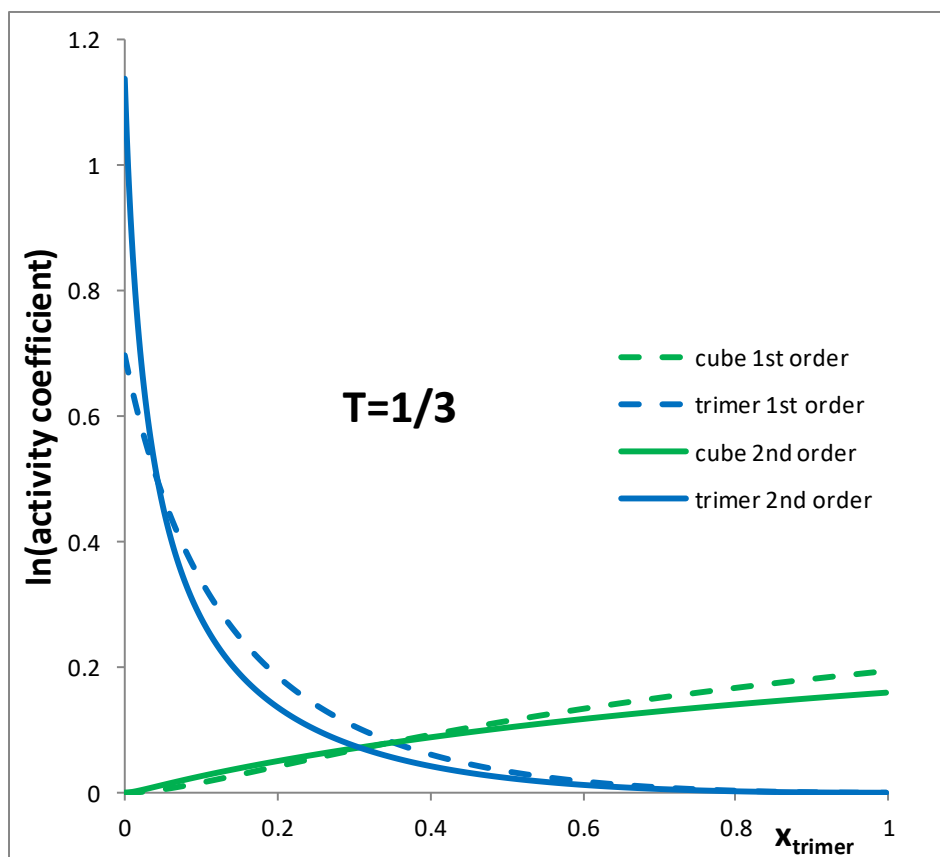


Figure 3: 1st order and 2nd order COSMO-RS activity coefficients for the cubes and trimers molecules, a) at $T = 1/3$, b) at $T = 3$.

Finally we need to calculate the free energies μ_c and μ_t of the cube and trimer molecules according to eq. 23. The uncorrected contributions μ_c^0 and μ_t^0 can be derived by summing up the logarithmic 2nd order segment activity coefficients. Since we have explicit analytic expressions for the 1st order segment activity coefficients, we can calculate the analytic derivatives of the γ^A and γ^B with respect to the surface fraction θ^A of non-polar segments. From that we easily get the derivatives of the entropic local correlation corrections. Combining these with the derivatives of the θ^A with respect to the particle numbers N_c and N_t , respectively, we yield the derivatives of the interaction energy corrections with respect to the particle numbers, and combining these with the second order contact probabilities we get the desired chemical potential corrections μ_c^{*corr} and μ_t^{*corr} and the total chemical potentials of both species. Figure 3 shows the chemical potentials resulting from first order COSMO-RS and the uncorrected and corrected chemical potentials resulting from 2nd order COSMO-RS of the compounds at three different temperatures. Figure 4 shows the corresponding Gibbs-Duhem consistency tests

$$\Delta^{GD} = \frac{x_c}{k_B T} \frac{\partial \mu_c}{\partial x_c} + \frac{x_t}{k_B T} \frac{\partial \mu_t}{\partial x_c} \quad (34)$$

for all three types of chemical potentials displayed in figure 3. Consistency is given if Δ^{GD} is zero. The derivatives of the chemical potentials are calculated numerically. As can be seen in figure 4, the 1st order chemical potentials and the corrected 2nd order chemical potentials are perfectly Gibbs-Duhem consistent at all temperatures, while the uncorrected 2nd order chemical potential are massively inconsistent.

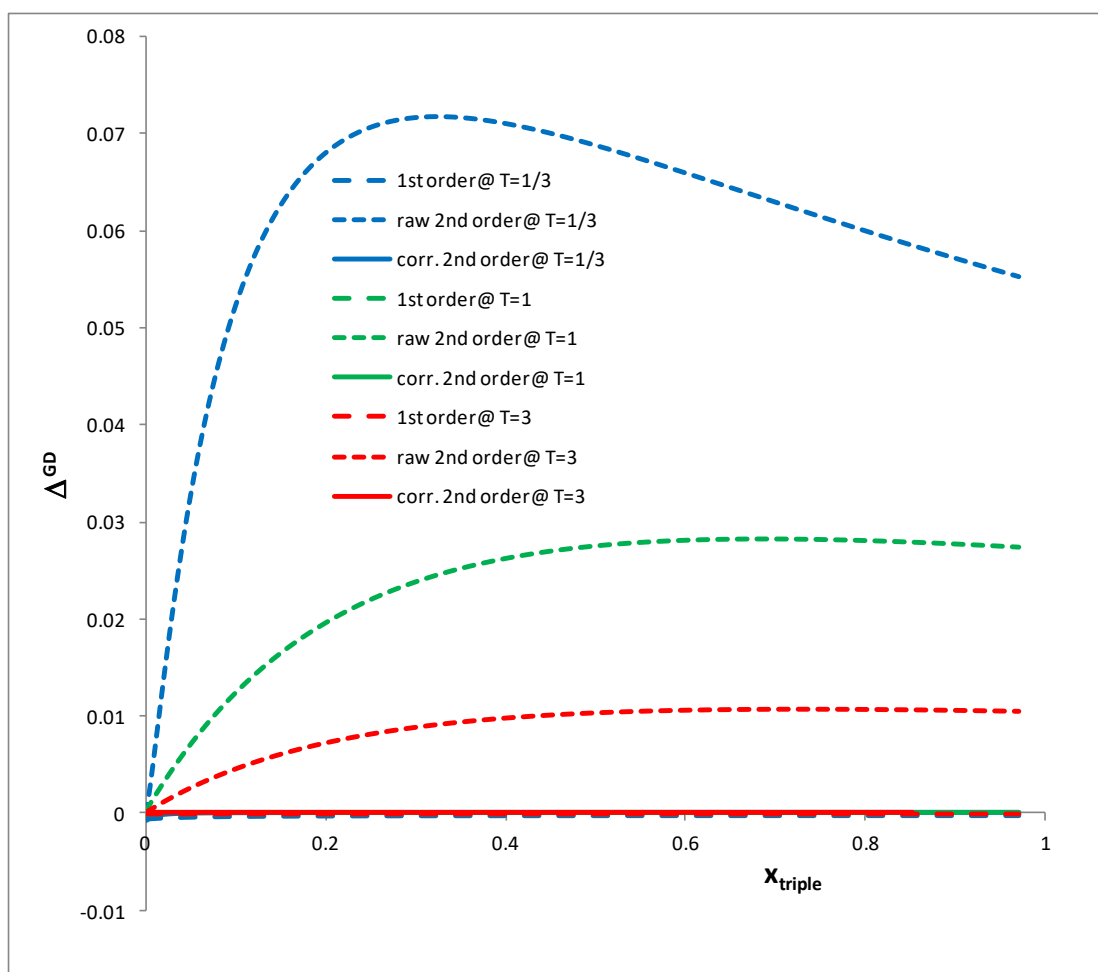


Figure 4: Gibbs-Duhem consistency test for the 1st and 2nd order chemical potentials as well as for the raw 2nd order chemical potentials before correction. The latter show a massive deviation from Gibbs-Duhem consistency, while the 1st order and the corrected 2nd order chemical potentials are perfectly Gibbs-Duhem consistent at all temperatures.

Summary, discussion and outlook

A 2nd order surface polarity correlation correction, COSMO-RSC, has been developed, which makes use of the contact probabilities derived from standard 1st order COSMO-RS calculations, derives from these local concentrations of surface polarities and feeds them back as entropic interaction free energy corrections into a second COSMO-RS loop. Utilizing a novel, thermodynamically exact correction method, the resulting 2nd order free energies and activity coefficients are completely Gibbs-Duhem consistent. As shown on the example of a lattice Monte-Carlo fluid

mimicking the dimerization of acetic acid in alkane solvents, the 2nd order COSMO-RS thermodynamics recovers the initially neglected effects resulting from coordinated interactions of neighboring surface segments almost perfectly.

In the model example considered in this paper the computational costs of COSMO-RSC are just about 3 times the costs of simple COSMO-RS, because the COSMOSPACE equations have to be solved a second time and derivatives of the contact probabilities have to be evaluated. In COSMO-RS calculations for real molecules most likely the evaluation of the geometrical segment correlation functions will be the major computational task in COSMO-RSC, because they involve the evaluation of many distances and angles between segments and segment normal vectors, respectively. Nevertheless, much of that can be pre-calculated and stored as an appendix to the COSMO file, which holds the polarity and geometric information of each molecule. Compared to the quantum chemical calculation required for the generation of the COSMO file the calculation of the segment correlation functions will be only a small additional effort. By such techniques it can be expected that the increase in computational costs caused by COSMO-RSC will be in the order of a factor of 3 or less.

Thus we can expect, while keeping the thermodynamic consistency and much of the efficiency of the COSMO-RS approach, COSMO-RSC will lead to improved overall accuracy for the prediction of thermodynamic properties, because the so far neglected effects of surface polarity correlation will be taken into account to a large degree. Furthermore, many extension of the application range can be expected, e.g. aggregation phenomena, hydrotropes, micro phase separation or even hydrophobic collapse of long alkyl or polymer chains. Improvements can also be expected for smaller ions, especially in their interaction with chelating solvents, because such ions always have strong and similar polarization charge density on neighboring surface segments, and especially strong interactions will arise, if a solvent molecule offers neighboring suitably arranged, i.e. correlated partners segments of the opposite polarity.

If necessary, higher order COSMO-RS, i.e. using the COSMO-RSC information for correlation effects in a 3rd COSMO-RS loop, should be feasible. Since the

composition derivatives of all quantities entering into the 2nd order COSMO-RS are available, the derivatives of the resulting contact probabilities should be available as well and thus even 3rd order and higher order COSMO-RS should yield thermodynamically consistent chemical potentials. Potentially this may be iterated to complete self-consistency. Nevertheless, before working on higher orders, the next step needs to be the development of optimal neighbor segment correlation functions for realistic, irregularly shaped molecular surfaces.

References

- (1) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995**, *99*, 2224–2235.
- (2) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. Refinement and Parametrization of COSMO-RS. *J. Phys. Chem. A* **1998**, *102*, 5074–5085.
- (3) Klamt, A. *COSMO-RS From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*; Elsevier: Amsterdam, The Netherlands; Boston, MA, USA, 2005.
- (4) Klamt, A.; Schüürmann, G. 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* **1993**, *1993*, 799–805.
- (5) Klamt, A.; Krooshof, G. J. P.; Taylor, R. COSMOSPACE: Alternative to Conventional Activity-Coefficient Models. *AIChE J.* **2002**, *48*, 2332–2349.
- (6) Klamt, A.; Eckert, F.; Arlt, W. COSMO-RS: An Alternative to Simulation for Calculating Thermodynamic Properties of Liquid Mixtures. *Annu. Rev. Chem. Biomol. Eng.* **2010**, *1*, 101–122.
- (7) K. Lucas, RWTH Aachen, personal communication 2001
- (8) Guggenheim, E. A., *Mixtures*, Oxford University Press, Oxford, 1952.
- (9) Egner, K.; Gaube, J.; Pfennig, A. GEQUAC, an Excess Gibbs Energy Model Describing Associating and Nonassociating Liquid Mixtures by a New Model Concept for Functional Groups. *Fluid Phase Equilibria* **1999**, *158*, 381–389.
- (10) Abrams, D. S.; Prausnitz, J. M. Statistical Thermodynamics of Liquid Mixtures: A New Expression for the Excess Gibbs Energy of Partly or Completely Miscible Systems. *AIChE J.* **1975**, *21*, 116–128.

- (11) Fredenslund, A.; Jones, R. L.; Prausnitz, J. M. Group-Contribution Estimation of Activity Coefficients in Nonideal Liquid Mixtures. *AIChE J.* **1975**, *21*, 1086–1099.
- (12) Larsen, B. L.; Rasmussen, P. A Comparison between the Quasichemical Model and Two-Fluid Local-Composition Models. *Fluid Phase Equilibria* **1986**, *28*, 1–11.
- (13) Lin, S.-T.; Sandler, S. I. Reply to Comments on “A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model.” *Ind. Eng. Chem. Res.* **2002**, *41*, 2332–2334.
- (14) Sachsenhauser, T.; Rehfeldt, S.; Klamt, A.; Eckert, F.; Klein, H. Consideration of Dimerization for Property Prediction with COSMO-RS-DARE. *Fluid Phase Equilibria* **2014**, *382*, 89–99.

TOC-graphic

