MOQUAC, a New Expression for the Excess Gibbs Free Energy Based on Molecular Orientations

Von der Fakultät für Maschinenwesen der Rheinisch-Westfälischen Technischen Hochschule Aachen zur Erlangung des akademischen Grades eines Doktors der Ingenieurwissenschaften genehmigte Dissertation

vorgelegt von

Rob Anna Hubertus Bronneberg

Berichter: Universitätsprofessor Dr.-Ing. Andreas Pfennig
Universitätsprofessor Dr.-Ing. André Bardow

Tag der mündlichen Prüfung: 6.7.2012

Preface and Acknowledgement

This doctoral thesis resulted from my work at the chair of Thermal Process Engineering of the RWTH Aachen University, Germany, under the supervision of Prof. Dr.-Ing. Andreas Pfennig from 10/2007 until 11/2011.

I wish to thank Prof. Dr.-Ing. Andreas Pfennig for his academic guidance and the freedom of development he granted me. I am grateful for the many fruitful discussions and his support and encouragement throughout my time at his chair.

I wish to thank Prof. Dr.-Ing. André Bardow for reviewing my doctoral thesis and Prof. Dr.-Ing. Reinhold Kneer for taking the position of chairperson during the examination.

I am grateful to all colleagues and students, who contributed with their work or with ideas and discussions to my thesis. I also wish to thank my colleagues for the pleasant four years at AVT.TVT.

I wish to thank Rebecca Buchbender for helping me with the English spelling.

I wish to thank my family, especially my parents, brother and sisters, for their support. Above all, I wish to thank my wife Lizbeth for her support, encouragement and patience at all times.

Ludwigshafen, August 2012 Rob Bronneberg

a mi amada esposa

Contents

Abstract

The three-dimensional structure of molecules determines if effects like steric hindrance or multiple contact points upon a molecular contact occur. These effects are especially important for molecules with several strongly interacting functional groups, since contacts between such groups have a strong influence on the behavior of liquid systems. The three-dimensional structure of molecules also needs to be considered if stereoisomers are to be distinguished. Bio-based molecules often have several strongly interacting functional groups and show enantiomerism. Because of the gradual shift towards greater use of renewable resources in the chemical industry, the consideration of the molecular geometry in G^{E}-models will become increasingly important. However, for all state-of-the-art G^{E}-models, assumptions are made that lead to the loss of information about the three-dimensional molecular structure. Therefore, in this work a new model is derived that does not need such assumptions.

The new model MOQUAC described in this thesis is based on a quasi-chemical approach. For this model the orientation of molecules upon a molecular contact is considered, which enables consideration of the three-dimensional structure of molecules. By comparison to results from lattice simulations, it is shown that MOQUAC can describe systems with coupled interactions. Additionally, a model to predictively describe the interaction energy between real molecules as a function of their orientation is derived. It is shown that MOQUAC can be used together with this interaction-energy model to predict the behavior of systems of real components.

In addition to a new expression for the residual contribution to the Gibbs energy, MOQUAC consists of a physically-founded improvement of the UNIQUAC combinatorial term. For the combinatorial contribution, a standard segment is used to determine the structural parameters r and q. It is generally assumed that the choice of the standard segment and the influence of the absolute value of q are negligible. The standard segment area, however, does not cancel out in the model equation and is therefore a model parameter. The improvement of the combinatorial term consists of a fit of the size of the standard segment to carefully selected experimental data. It is shown that the new standard segment significantly improves the performance of the UNIQUAC combinatorial term and that the physically founded improved model performs at least as well as empirical modifications of the original term.

Kurzfassung

Die dreidimensionale Molekülstruktur bestimmt ob Effekte wie sterische Hinderungen oder gekoppelte Wechselwirkungen bei einem Molekülkontakt vorkommen. Diese Effekte sind insbesondere wichtig für Moleküle mit verschiedenen stark wechselwirkenden funktionellen Gruppen, weil Kontakte zwischen solchen Gruppen eine starke Auswirkung auf das Verhalten von flüssigen Systemen haben. Die dreidimensionale Molekülstruktur muss auch zur Unterscheidung von Stereoisomeren berücksichtigt werden. Biobasierte Moleküle haben häufig verschiedene stark wechselwirkende funktionelle Gruppen und kommen als Enantiomeren vor. Wegen des zu erwartenden Rohstoffwandels hin zu nachwachsenden Rohstoffen in der chemischen Industrie wird die Berücksichtigung der Molekülstruktur in G^{E}-Modelle zunehmend an Bedeutung gewinnen. Jedoch werden derzeit für alle G^{E}-Modelle Annahmen gemacht, die zum Verlust der Information über die dreidimensionale Molekülstruktur führen. Deshalb wird in dieser Arbeit ein Modell hergeleitet, welches ohne solche Annahmen auskommt.

Das neue Modell MOQUAC, welches in dieser Dissertation beschrieben wird, basiert auf einem quasichemischen Ansatz. Für das Modell wird die Ausrichtung eines Moleküls beim Molekülkontakt berücksichtigt, was es ermöglicht, die dreidimensionale Molekülstruktur zu berücksichtigen. Durch einen Vergleich mit Ergebnissen aus einer Gittersimulation wird gezeigt, dass MOQUAC Systeme mit gekoppelten Wechselwirkungen beschreiben kann. Zusätzlich wird ein Modell hergeleitet, welches prädiktiv die Wechselwirkungsenergie zwischen realen Molekülen in Abhängigkeit von ihrer Ausrichtung beschreiben kann. Es wird gezeigt, dass MOQUAC zusammen mit dem Wechselwirkungsenergiemodell zur Vorhersage des Verhaltens von Stoffsystemen aus realen Komponenten angewendet werden kann.

Zusätzlich zu einem neuen Ausdruck für den residuellen Beitrag zur freien Enthalpie besteht MOQUAC aus einer physikalisch fundierten Verbesserung des kombinatorischen Terms des UNIQUAC-Modells. Für den kombinatorischen Beitrag wird ein Standardsegment zur Bestimmung der strukturellen Parameter r und q benutzt. Es wird allgemein angenommen, dass die Wahl des Standardsegmentes und der Einfluss des absoluten Wertes von q vernachlässigbar sind. Die Oberfläche des Standardsegmentes jedoch kürzt sich nicht aus der Modellgleichung heraus und ist deshalb ein Modellparameter. Die Verbesserung des kombinatorischen Terms besteht aus einer Anpassung der Größe des Standardsegmentes an sorgfältig ausgewählte experimentelle Daten. Es wird gezeigt, dass das neue Standardsegment die Leistung des UNIQUAC kombinatorischen Term signifikant verbessert und dass das physikalisch fundiert verbesserte Modell mindestens genauso leistungsfähig wie die empirischen Modifikationen des ursprünglichen Terms ist.

1 Introduction

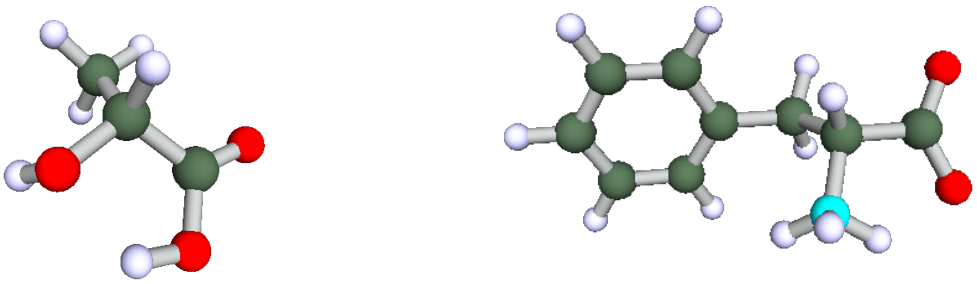
For almost every product of the chemical industry, specifications concerning purity and maximal allowed concentrations of contaminants are defined. To reach these specifications, thermal separation processes such as crystallization, absorption, distillation or liquid-liquid extraction are normally applied in downstream processes. According to Gmehling [1], these processes can account for up to 80 % of the required investment costs of a chemical process. Distillation processes in particular can be very energy intensive and often account for the majority of the process' energy demand. Klemm and Emig [2] estimate that energy costs make up for 15 % of the total production cost of chemical products. Consequently, an exact design and the determination of optimal operating conditions for thermal separation processes are of great importance.

Modeling of the excess Gibbs energy, G^{E}, allows description of the activity of components in liquid mixtures. These activities are required in many fields of chemical engineering such as description of vapor-liquid and liquid-liquid equilibria. Phase equilibria are needed for the design and optimization of various thermal separation processes. Several G^{E}-models are available nowadays, of which the most relevant models for the chemical industry are described in chapter 2. Many of these models require experimental data to determine their parameters for the description of a specific system. Only few models are capable of a purely predictive description of systems. A predictive description is required if no adequate experimental data of the system under consideration are available, as is often the case in the phase of conceptual process design.

All state-of-the-art G^{E}-models have in common that they were designed to serve in a chemical industry that is based upon crude oil. Many substances derived from crude oil only have a small number of strongly interacting functional groups. A molecule with several strongly interacting functional groups is challenging to model because the three-dimensional structure of the molecule needs to be taken into account. It is the structure that determines, for example, whether multiple contact points between strongly interacting groups upon a molecular contact occur. Such contacts with multiple contact points typically have high interaction energies and can therefore significantly influence the behavior of a system. For example, in non hydrogen-bonding solvents and at low concentrations, acetic acid forms cyclic dimers that are characterized by two simultaneously occurring hydrogen bonds [3].

The three-dimensional molecular structure also needs to be taken into account to describe effects like steric hindrance or to enable a model to distinguish between enantiomers. Typical bio-based molecules in particular often shown enantiomerism and have several strongly interacting functional groups, for example, lactic acid and phenylalanine, which are shown in Fig. 1.1. Because of the limited availability of crude oil and an increasing awareness of the need for sustainable production, it is to be expected that the chemical industry will face a shift towards greater use of renewable resources. With this change in raw materials it will become more important for G^{E}-models to take molecular structure into account. However, as demonstrates chapter 2, no state-of-the-art G^{E}-model is able to adequately account for the three-dimensional molecular structure.

The objective of this work is to improve the predictive modeling of the excess Gibbs energy by taking account of the full three-dimensional structure of molecules. To do this, an existing



(a) r-lactic acid

(b) r-phenylalanine

Figure 1.1: Examples of bio-based molecules.

model for the combinatorial contribution to G^E will be improved and a new model for the residual contribution to G^E will be derived. The combinatorial contribution to G^E describes size and shape effects and the residual contribution to G^E describes all energetic effects in liquid mixtures.

There are physically founded models that describe the combinatorial contribution to G^E, for example, the Guggenheim-Staverman model. The combinatorial term in UNIQUAC is based on the Guggenheim-Staverman model. However, in the application of the combinatorial term of UNIQUAC to real mixtures, large deviations were found. Several authors have thus proposed modifications to this combinatorial term. All of these modifications significantly improved the performance of the term, but they are all empirical. In chapter 3, a physically-founded modification to the UNIQUAC combinatorial term is presented that improves the original term in a similar or even better way than the empirical modifications.

In chapter 4, MOQUAC will be presented. This is the new model for the residual contribution to G^E. MOQUAC is based on an explicit consideration of the orientations of molecules upon molecular contacts and therefore allows consideration of the full three-dimensional structure of molecules. Because of this, the model is capable of describing effects like multiple contact points and steric hindrance. MOQUAC also allows enantiomers to be distinguished. Thus, compared to the state-of-the-art G^E-models MOQUAC is better suited for the description of more complex molecules that have several strongly interacting functional groups.

For MOQUAC, the interaction energy between two molecules as a function of the orientation of both molecules is a model parameter. In chapter 5, a first simple, empirical interaction-energy model that allows a-priori description of this interaction energy will be presented. To produce this a-priori description, the model uses information from a quantum-chemical calculation and can thus, in principle, be applied to any kind of molecule. The combination of the interaction-energy model with MOQUAC allows for a predictive description of the excess Gibbs energy. The few parameters in the interaction-energy model are first fitted to experimental data. Using these parameters the combination of both models is then used to predict the behavior of real systems. These results are then discussed and compared to results with the modified UNIFAC (Dortmund) model and COSMO-RS, which are currently two of the most widely used, predictive G^E-models in the chemical industry.

2 State-of-the-art models to describe the excess Gibbs energy

For the design and optimization of thermal-separation units, it is important to describe the vapor-liquid and liquid-liquid equilibrium. For the modeling of strongly non-ideal systems, models that describe the excess Gibbs energy, G^{E}, are often applied. Frequently, a combinatorial and a residual contribution for such models are distinguished:

$$G^{\mathrm{E}} = G^{\mathrm{E}}_{\mathrm{comb}} + G^{\mathrm{E}}_{\mathrm{res}} \tag{2.1}$$

The combinatorial contribution $G^{\mathrm{E}}_{\mathrm{comb}}$ describes size and sometimes shape effects of the components in the mixture, whereas the residual contribution $G^{\mathrm{E}}_{\mathrm{res}}$ describes all energetic effects. The denotation "residual" is misleading however, since it suggests that this term is of less importance. This only applies for athermal or almost athermal mixtures. Most common mixtures in chemical engineering can, however, not be treated as athermal mixtures.

This chapter has two main sections. First, some of the most common models for the combinatorial term alone are described. This is followed by a description of some of the most popular state-of-the-art G^{E}-models that also include an expression for the residual term. Special attention is paid to the reason why none of these models can take the full three-dimensional molecular structure into account. These models are therefore unable to describe effects like multiple contact points, something which is important for the description of components with several strongly interacting functional groups. In chapter 4, the new G^{E}-model MOQUAC is presented. MOQUAC overcomes this shortcoming. In this work molar units are generally used for all quantities.

2.1 The combinatorial contribution

For a mixture of molecules of different size and shape, there are different ways of arranging the molecules in space for the mixture and for the pure components. This results in a contribution to the entropy of mixing that is independent of molecular interaction energies. This contribution is called the combinatorial contribution and plays an important role in the modeling of athermal solutions of molecules of very different size [4]. The combinatorial contribution to the excess Gibbs energy is only determined by the combinatorial contribution to the excess entropy $S^{\mathrm{E}}_{\mathrm{comb}}$:

$$G^{\mathrm{E}}_{\mathrm{comb}} = -TS^{\mathrm{E}}_{\mathrm{comb}} \tag{2.2}$$

The vast majority of models for the combinatorial contribution are based on a lattice picture of fluids. Here, a molecule i consists of r_i segments that each occupies one lattice site. The first model to be presented is the Flory-Huggins model. This model only takes molecular size into account. Next the Staverman-Guggenheim model is presented. This model takes both molecular size and shape into account. Finally, some empirical modifications of these physically founded lattice models are discussed.

2.1.1 The Flory-Huggins model

Flory [5] and Huggins [6] considered the number of possible ways of arranging N_c different components on a lattice, where each component consists of r_i segments each occupying exactly one lattice site. The lattice has a coordination number of z and it is assumed that there are no empty lattice sites. Flory and Huggins assumed for their derivation that the probability of finding an empty lattice site while placing a molecule on the lattice is determined by the ratio of the number of free lattice sites to the total number of lattice sites. Their derivation leads to

$$G^{E}_{comb} = RT \sum_{i=1}^{N_c} x_i \ln \frac{\phi_i}{x_i} \tag{2.3}$$

for the excess part of the combinatorial Gibbs energy. R is the universal gas constant, x_i is the mole fraction and ϕ_i is the volume fraction of component i in the mixture

$$\phi_i = \frac{x_i r_i}{\sum_{j=1}^{N_c} x_j r_j} \tag{2.4}$$

A detailed derivation of Eq. 2.3 can, for example, be found in Pfennig [7].

2.1.2 The Staverman-Guggenheim model

Flory and Huggins' assumption concerning the probability of finding an empty lattice site is only a crude approximation, since no notice is taken of the fact that the empty lattice sites are not distributed randomly. In fact, this distribution is structured in relation to the structured molecules. This non-random distribution of empty lattice sites is due to the fact that the r_i segments of a molecule are connected, leading to larger coherent regions of empty lattice sites than considered in the Flory-Huggins model. These regions increase the number of different possible positions for the molecules. Staverman and Guggenheim distinguish between internal and external contacts between the segments of molecules in order to account for this effect. They define zq_i as the total number of external segment contacts of molecule i, where $q_i < r_i$ except in the trivial case of $r_i = 1$.

Guggenheim [8] derived an expression for the combinatorial entropy for linear and branched molecules with the assumption that molecules show no ring formation. In this case, the number of segments and the number of external contact sites are related by

$$\frac{z}{2} q_i = \frac{z}{2} r_i - r_i + 1 \tag{2.5}$$

Guggenheim's equation for the combinatorial entropy reads

$$S_{comb} = R \sum_{i=1}^{N_c} x_i \left(\ln \rho_i + \frac{z}{2} q_i \ln \frac{\bar{q}}{\bar{r}} + \ln \frac{\bar{r}}{x_i} \right) \tag{2.6}$$

where ρ_i is the number of possible positions of a molecule of component i on an empty lattice when the first segment of the component is kept fixed. According to Sanchez [9], for rigid molecules ρ_i is less or equal to $z^2 - z$. $\bar{r}$ and $\bar{q}$ are average structural parameters and are defined as

$$\bar{r} = \sum_{i=1}^{N_c} x_i r_i \tag{2.7}$$

and

$$\bar{q} = \sum_{i=1}^{N_c} x_i q_i \tag{2.8}$$

However, Staverman [10] showed that inconsistencies arise when Guggenheim's equation for the combinatorial entropy is generalized to any value of q_i different from Eq. 2.5. He removed these inconsistencies, which resulted in a different equation for the combinatorial entropy

$$S_{\text{comb}} = R \sum_{i=1}^{N_c} x_i \left(\ln \rho_i + \frac{z}{2} q_i \ln \frac{\bar{q}}{\bar{r}} + \ln \frac{\bar{r}}{x_i} + 1 - r_i + \frac{z}{2} r_i - \frac{z}{2} q_i \right) \tag{2.9}$$

A comparison of Eqs. 2.6 and 2.9 shows that both equations only yield identical results when Eq. 2.5 applies.

The excess entropy of a mixture is defined by

$$S^{\text{E}}_{\text{comb}} = S_{\text{comb}} - \sum_{i=1}^{N_c} x_i S^0_{\text{comb},i} + R \sum_{i=1}^{N_c} x_i \ln x_i \tag{2.10}$$

where S_{comb} is the combinatorial entropy of the mixture and $S^0_{\text{comb},i}$ the combinatorial entropy of the pure component i. The combinatorial entropy of the mixture and of the pure components can both be determined either with Eq. 2.6 or with Eq. 2.9. For the excess part of the combinatorial entropy, both equations yield the same result

$$S^{\text{E}}_{\text{comb}} = -R \sum_{i=1}^{N_c} x_i \ln \frac{\phi_i}{x_i} - \frac{1}{2} R \sum_{i=1}^{N_c} z q_i x_i \ln \frac{\psi_i}{\phi_i} \tag{2.11}$$

where ψ_i is the surface fraction

$$\psi_i = \frac{x_i q_i}{\sum_j^{N_c} x_j q_j} \tag{2.12}$$

of component i in the mixture. With Eq. 2.2, the combinatorial contribution to the excess Gibbs energy equals

$$G^{\text{E}}_{\text{comb}} = RT \sum_{i=1}^{N_c} x_i \ln \frac{\phi_i}{x_i} + \frac{1}{2} RT \sum_{i=1}^{N_c} z q_i x_i \ln \frac{\psi_i}{\phi_i} \tag{2.13}$$

and is referred to as the Staverman-Guggenheim equation. Since Staverman did not assume Eq. 2.5, the Staverman-Guggenheim equation can also be applied to molecules with ring formation.

2.1.3 Empirical modifications of existing models

Abrams and Prausnitz proposed using the Staverman-Guggenheim model, Eq. 2.13, to describe the combinatorial term of UNIQUAC [11]. To determine the structural parameters r_i and q_i from the point of view of lattice models, a standard segment needs to be defined. With the surface area A_{ref} and volume V_{ref} of this standard segment, the structural parameters are then determined from

$$q_i = \frac{A_i}{A_{\text{ref}}} \tag{2.14}$$

and

$$r_i = \frac{V_i}{V_{\text{ref}}} \tag{2.15}$$

where A_i and V_i are the molecular surface area and volume of component i.

For the molecular surface area and volume, Abrams and Prausnitz proposed using the values determined with Bondi's [12] group-increment method. They regard the standard segment as a sphere with radius r_{ref} so its surface area is calculated by

$$A_{\mathrm{ref}} = 4N_{\mathrm{A}}\pi r_{\mathrm{ref}}^2 \tag{2.16}$$

and volume by

$$V_{\mathrm{ref}} = \frac{4}{3}N_{\mathrm{A}}\pi r_{\mathrm{ref}}^3 \tag{2.17}$$

N_{A} is the Avogadro constant. Abrams and Prausnitz chose the standard segment for the UNIQUAC combinatorial term such that it satisfies Eq. 2.5 for a linear polymethylene molecule of infinite length [11]. This choice was arbitrary and leads to

$$r_{\mathrm{ref}} = 1.818 \times 10^{-10}\,\mathrm{m} \tag{2.18}$$

Kikic et al. [13] investigated the UNIQUAC combinatorial term and found that the degree of non-ideality predicted by the model is greatly exaggerated for some binary alkane mixtures. A similar comparison to that of Kikic is shown in Fig. 2.1, where experimental data on the activity coeffient of *n*-hexane at infinite dilution in other *n*-alkanes is compared to model results. Mixtures of *n*-alkanes behave nearly athermally and are therefore adequate for such comparison. Kikic et al. proposed a modification of the UNIQUAC combinatorial term analogous to a modification proposed by Donohue and Prausnitz [14] for the Flory-Huggins model. This was:

$$G^{\mathrm{E}}_{\mathrm{comb}} = RT\sum_{i=1}^{N_{\mathrm{c}}} x_i \ln\frac{\phi_i^{\mathrm{Kikic}}}{x_i} + \frac{1}{2}RT\sum_{i=1}^{N_{\mathrm{c}}} zq_i x_i \ln\frac{\psi_i}{\phi_i} \tag{2.19}$$

with

$$\phi_i^{\mathrm{Kikic}} = \frac{x_i r_i^{\frac{2}{3}}}{\sum_{j=1}^{N_{\mathrm{c}}} x_j r_j^{\frac{2}{3}}} \tag{2.20}$$

The exponent $2/3$ was determined by comparison with experimental data. This modification greatly improves the performance of the model, which now only slightly overestimates the non-idealities of shorter *n*-alkanes, as can be seen in Fig. 2.1.

Huyskens and Haulait-Pirson [16] state that many authors are aware that the Flory-Huggins term with the reference segment of UNIQUAC does not describe reality quite correctly, as can also be seen in Fig. 2.1. Like the UNIQUAC combinatorial term, the Flory-Huggins term greatly overestimates the non-idealities of the mixture. Huyskens and Haulait-Pirson realized that reality lies almost exactly in the middle between the ideal solution and the Flory-Huggins equation and thus write

$$G^{\mathrm{E}}_{\mathrm{comb}} = RT\frac{1}{2}\sum_{i=1}^{N_{\mathrm{c}}} x_i \ln\frac{\phi_i}{x_i} \tag{2.21}$$

They show that this equation yields better results for solubilities of solid *n*-alkanes in liquid alkanes than the ideal solution term or the Flory-Huggins term alone. Their equation is simple and yields clearly improved results, as shown in Fig. 2.1, although most non-idealities are overestimated.

Weidlich and Gmehling proposed a modification similar to that of Kikic et al., but instead of an exponent of $2/3$, they used $3/4$ for their modified UNIFAC (Dortmund) model [17]. This

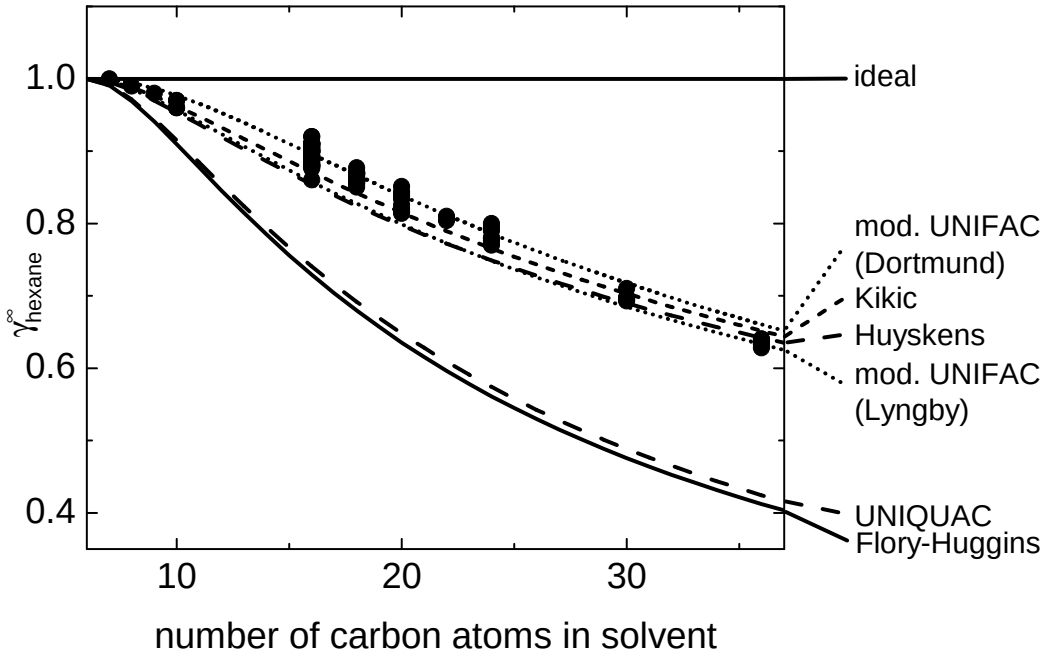


Figure 2.1: Comparison of the different combinatorial terms with experimental data on the activity coefficient of *n*-hexane at infinite dilution in other *n*-alkanes. All experimental data were taken from the DECHEMA data series [15].

value resulted from a fit of the activity coefficient at infinite dilution to experimental data of mixtures of alkane + alkane, alkane + alcohol, and alcohol + alcohol. The parameters q and r which they use are determined from different group contributions. These group contributions for the parameters q and r were also obtained from a fit to experimental data and are no longer calculated from molecular parameters as in UNIQUAC. Thus, pure-component parameters have in effect been fitted to mixture data, which may be causing inconsistencies. Fig. 2.1 shows that Weidlich and Gmehling's model reproduces the experimental data very well. It should be stressed, however, that for the modified UNIFAC (Dortmund) model, the modifications to the structure of the model are merely empirical.

For their modified UNIFAC (Lyngby) model, Larsen et al. proposed using an altered Flory-Huggins combinatorial term with modified volume fractions following Kikic et al. [18]

$$G^{\mathrm{E}}_{\mathrm{comb}} = RT \sum_{i=1}^{N_{\mathrm{c}}} x_i \ln \frac{\phi_i^{\mathrm{Kikic}}}{x_i} \tag{2.22}$$

instead of using the Staverman-Guggenheim combinatorial term. The volume parameters r are calculated as in UNIQUAC [18]. They argue that the Staverman-Guggenheim term represents a corrected Flory-Huggins term, which they correct in an alternative way, as specified in Eq. 2.22. The correcting contribution of the last term of Eq. 2.13 is said to frequently be quite small, but it may, however, in some cases give large corrections, even leading to negative values of the combinatorial excess entropy, which are not considered realistic. Fig. 2.1 shows that the

modified UNIFAC (Lyngby) model performs similarly to the Huyskens model, overestimating almost all non-idealities. In chapter 3 it will be shown that this correction term is not negligible when the standard segment is determined meaningfully.

2.2 The residual contribution

The combinatorial contribution only describes entropic effects due to differences in the size and shape of the molecules, whereas the residual contribution considers all energetic effects in the mixture. If the interaction energies between molecules are different in the mixture and in the pure components, heat can be liberated or absorbed when the pure components are mixed. This leads to a contribution to the residual part of the excess enthalpy $H^{\mathrm{E}}_{\mathrm{res}}$. In addition, the interaction energies can cause distribution of the molecules in the mixture and in the pure components that differs from random distribution. This leads to a contribution to the residual part of the excess entropy $S^{\mathrm{E}}_{\mathrm{res}}$. The residual Gibbs energy is defined as

$$G^{\mathrm{E}}_{\mathrm{res}} = H^{\mathrm{E}}_{\mathrm{res}} - TS^{\mathrm{E}}_{\mathrm{res}} \tag{2.23}$$

As with modeling of the combinatorial term, a lattice view is often applied in modeling the energetic effects in liquid mixtures. It is assumed that molecules linger on lattice positions, around which they oscillate. Interchanges of positions are so rare in liquids that they can be disregarded for thermodynamic equilibrium. Guggenheim [19] in particular but also Barker [20, 21] have significantly influenced the development of these lattice theories. Because their perception of liquid mixtures is the basis for many known G^{E}-models, the lattice-based model of Guggenheim is presented first in the following section. The concept of local composition is also introduced. This is followed by the presentation of some semi-empirical models that are based on Guggenheim's ideas and the concept of local compositions and that are still in use today. After that, the most relevant group-contribution models and surface-segment models are presented. Special attention is paid to the reason why all of these models cannot take the full three-dimensional molecular structure into account correctly.

2.2.1 Guggenheim's lattice-based model for liquids

For Guggenheim's lattice-based model, a mixture of N_{c} components is regarded. All molecules are considered to be approximately of equal size, such that it can be assumed that each molecule occupies exactly one lattice site. Exchanging the position of two different molecules does thus not lead to a steric contribution. It is also assumed that no lattice site is empty, that the coordination number z of the lattice is constant and that the interaction energy between two molecules does not depend on the orientation of the molecules in space. Further assumptions are that the interaction energy between two molecules is not influenced by other molecules and that for the system only the interaction energy between directly neighboring molecules needs to be considered.

Wilson introduced the concept of local compositions [22], where $x_{j,i}$ is the local composition of component j in the direct vicinity of component i. $x_{j,i}$ is the normalized fraction of all contacts, originating from component i with component j. Based on this definition,

$$\sum_{j=1}^{N_{\mathrm{c}}} x_{j,i} = 1 \quad \text{for all} \quad i = 1,\ldots N_{\mathrm{c}} \tag{2.24}$$

applies. Since the number of contacts originating from molecule i with molecule j must be equal to the number of contacts originating from molecule j with molecule i, the symmetry condition

$$x_i x_{j,i} = x_j x_{i,j} \tag{2.25}$$

applies.

According to the assumptions made above, the system's energy U equals

$$U = \sum_{i=1}^{N_c} \sum_{j=1}^{N_c} \frac{1}{2} z x_i x_{j,i} u_{j,i} \tag{2.26}$$

where the factor $1/2$ accounts for the fact that in the sum each contact is considered twice. $u_{j,i}$ is the interaction energy between molecules i and j. The interaction energy is symmetrical, so

$$u_{j,i} = u_{i,j} \tag{2.27}$$

applies. For lattice systems no volume dependence is assumed, so the system's energy is equal to the system's enthalpy

$$H = U \tag{2.28}$$

The excess enthalpy of a mixture is defined as

$$H^E = H - \sum_{i=1}^{N_c} x_i H_i^0 \tag{2.29}$$

where H is the enthalpy of the mixture and H_i^0 is the enthalpy of pure component i. Applying this definition and Eqs. 2.28 and 2.26 yields

$$H_{res}^E = \sum_{i=1}^{N_c} \sum_{j=1}^{N_c} \frac{1}{2} z x_i x_{j,i} u_{j,i} - \sum_{i=1}^{N_c} \frac{1}{2} z u_{i,i} \tag{2.30}$$

for the excess enthalpy.

Hu et al. [23] give an expression for the molar excess entropy with use of the local composition:

$$S_{res}^E = -\frac{z}{2} R \sum_{i=1}^{N_c} \sum_{j=1}^{N_c} x_i x_{j,i} \ln \frac{x_{j,i}}{x_j} \tag{2.31}$$

Eq. 2.31 is based on the assumptions of Guggenheim. The derivation of Eq. 2.31 is not given in the paper of Hu et al., but is given, for example, by Lucas [24].

Unknown so far are the local compositions $x_{j,i}$ that describe a representative system state at equilibrium. Within his framework of the quasi-chemical theory, Guggenheim derived the relation

$$\frac{x_{j,i} x_{i,j}}{x_{i,i} x_{j,j}} = \exp\left(-\frac{\omega_{j,i}}{RT}\right) \tag{2.32}$$

where $\omega_{j,i}$ is defined as:

$$\omega_{j,i} = u_{j,i} + u_{i,j} - u_{i,i} - u_{j,j} \tag{2.33}$$

$z\omega_{j,i}$ is called the interchange energy and equals the energy difference of interchanging one molecule of component i and of component j from their pure solutions respectively. The denomination "quasi-chemical" has its origin in the fact that Eq. 2.32 resembles the formulation of a

chemical equilibrium reaction. The reaction in this case is that of one $i-i$ and one $j-j$ contact reacting to one $i-j$ and one $j-i$ contact. This is illustrated in Fig. 2.2.

Eq. 2.32 is equivalent to the law of mass action. The term on the right-hand side gives the equilibrium constant of the quasi-chemical reaction in the form of the Boltzmann factor. The Boltzmann factor describes the temperature dependence of the equilibrium.

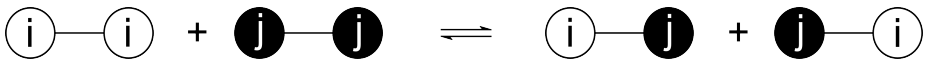


Figure 2.2: Quasi-chemical reaction of one $i-i$ and one $j-j$ contact reacting to one $i-j$ and one $j-i$ contact.

Guggenheim's model was of great significance for advances in the modeling of liquids and the basic idea of this model is still part of state-of-the-art G^{E}-models. However, since the model was derived for spherical molecules whose interaction energy with other molecules does not depend on their molecular orientation, only higher order upon mixing can be predicted. Higher order means a negative excess entropy. For many nonpolar mixtures, however, positive excess entropies have been observed experimentally [4]. These positive excess entropies cannot be described by the model. Guggenheim [19] recognized this shortcoming and proposed considering $u_{j,i}$ as a free energy instead of only an energy parameter. For $u_{j,i}$ Guggenheim always assumed a linear dependence of temperature

$$u_{j,i} = h_{j,i} - T s_{j,i} \tag{2.34}$$

where $h_{j,i}$ and $s_{j,i}$ are considered constant in a certain temperature range. With this modification, positive excess entropies can also be described. Guggenheim did not give an interpretation for this approach, but according to Prausnitz [4], positive excess entropies are a result of other effects (neglected by the lattice theory), such as changes in volume and changes in excitation of internal degrees of freedom (rotation, vibration) resulting from the mixing process [25].

The necessary transfer from an energy to a Gibbs energy can also be caused by the fact that formally the direction dependent molecular potentials of real molecules were replaced by direction independent molecular potentials [24]. It is clear that direction dependent molecular potentials have to be considered in order to take the full three-dimensional molecular structure into account.

2.2.2 Semi-empirical models for the excess Gibbs energy on the basis of local compositions

Numerous models have been developed based on Guggenheim's idea that because of an energetic preference, some contacts between molecules are more frequent than others. Models that have been established in industrial applications are the Wilson model [22], the NRTL model [26] and the UNIQUAC model [11].

To describe the preference of certain contacts, the Wilson model uses

$$\frac{G^{\mathrm{E}}}{RT} = -\sum_{i=1}^{N_{\mathrm{c}}} x_i \ln\left(\sum_{j=1}^{N_{\mathrm{c}}} \Lambda_{i,j} x_j\right) \tag{2.35}$$

with

$$\Lambda_{i,j} = \frac{V_j^{\mathrm{L}}}{V_i^{\mathrm{L}}} \exp\left(\frac{u_{i,i} - u_{i,j}}{RT}\right) \tag{2.36}$$

For each binary component combination, the model shows the two parameters $u_{i,i} - u_{i,j}$ and $u_{j,j} - u_{j,i}$. $u_{i,j}$ is the interaction energy between components i and j.

Wilson recognized that size differences between molecules have to be accounted for. To do this, he used the quotient of the molar volume of the pure liquid V_i^{L}. These volumes can be found for many substances in the DIPPR database [27], for example. Eq. 2.35 is a complete G^{E}-model that in the case when all parameters $u_{i,i} - u_{i,j}$ equal zero reduces to the Flory-Huggins equation Eq. 2.3.

The NRTL (Non-Random Two Liquid) model by Renon and Prausnitz [26] is also based on the concept of local compositions as well as on the two-liquid theory proposed by Scott [28]. The two-liquid theory states that a binary mixture can be described as a combination of two liquids. While one liquid consists of cells that contain a molecule of type 1 as the central molecule that is surrounded by molecules of type 1 and 2, the other liquid consists of cells that contain a molecule of type 2 as the central molecule that is also surrounded by molecules of type 1 and 2. The NRTL model describes the excess Gibbs energy as

$$\frac{G^{\mathrm{E}}}{RT} = \sum_{i=1}^{N_{\mathrm{c}}} x_i \frac{\sum_{j=1}^{N_{\mathrm{c}}} \tau_{j,i} G_{j,i} x_j}{\sum_{k=1}^{N_{\mathrm{c}}} G_{k,i} x_k} \tag{2.37}$$

with

$$\tau_{j,i} = \frac{u_{j,i} - u_{i,i}}{RT} \tag{2.38}$$

$$G_{j,i} = \exp\left(-\alpha_{j,i}\tau_{j,i}\right) \tag{2.39}$$

$$u_{j,i} = u_{i,j} \tag{2.40}$$

Here too $u_{j,i}$ is the interaction energy between the components i and j. In contrast to the Wilson model, the NRTL model possesses a further parameter $\alpha_{j,i}$. Typically

$$\alpha_{j,i} = \alpha_{i,j} \tag{2.41}$$

is applied, so three parameters are needed to describe a binary mixture. Since, however, two parameters often already offer the model sufficient flexibility, usually all $\alpha_{j,i}$ are set to a fixed value, e.g. 0.2 or 0.3 [7].

The UNIQUAC (UNIversal QUAsi-Chemical theory) model by Abrams and Prausnitz [11] and Maurer and Prausnitz [29] distinguishes a combinatorial and a residual contribution to the excess Gibbs energy using Eq. 2.1. The UNIQUAC combinatorial term was described in section 2.1.3 and the residual term is calculated by

$$G^{\mathrm{E}}_{\mathrm{res}} = -RT \sum_{i=1}^{N_{\mathrm{c}}} q_i x_i \ln\left(\sum_{j=1}^{N_{\mathrm{c}}} \psi_j \tau_{j,i}\right) \tag{2.42}$$

with

$$\tau_{j,i} = \exp\left(-\frac{u_{j,i} - u_{i,i}}{RT}\right) \tag{2.43}$$

The temperature dependence of the interaction parameters $\tau_{j,i}$ is described with the Boltzmann factor, where $u_{j,i}$ is the interaction energy between molecule i and j.

In application of the Wilson, NRTL and UNIQUAC model to multi-component mixtures, only the parameters of all binary subsystems need to be known. These parameters for all three models

can be determined by fitting to experimental data of binary systems. Thus, these models can be applied predictively to multi-component mixtures when experimental data of all binary subsystems are available. However, this also limits the predictive applicability of the models, since they require a certain amount of experimental data related to the system of interest.

Since only one interaction energy between molecules of components i and j is distinguished, no orientation-dependent molecular potentials can be considered. Like the Guggenheim lattice model, the Wilson, NRTL and UNIQUAC models also cannot take the full three-dimensional molecular structure into account. In addition, and in contrast to the Guggenheim model, the other 3 models determine the local compositions $x_{j,i}$ independently, although they should fulfill both Eq. 2.24 and Eq. 2.25. This inconsistency has been criticized by different authors, for example, McDermott and Ashton [30] and means that a reliable prediction of multi-component mixtures cannot be guaranteed [31]. Because of this inconsistency, however, these semi-empirical models are very flexible, which is the reason for their broad application in chemical industries.

2.2.3 Group-contribution models for the excess Gibbs energy

The model parameters of all semi-empirical models described in section 2.2.2 must be determined by a fit to experimental data. This, of course, is very time-consuming and also does not allow for the description of systems for which no experimental data are available. The idea of group-contribution models is to reduce the innumerable number of different molecules to a manageable number of different functional groups. The thermodynamic behavior of molecules is then traced back to the properties of these functional groups.

With this concept almost all practical relevant substances can be considered an agglomerate of relatively few, adequately chosen functional groups. Based on this basic concept, different models have been developed. All popular models such as UNIFAC, ASOG, DISQUAC and GTASQUAC have in common that a combinatorial and a residual contribution are distinguished according to Eq. 2.1. The description of the residual contribution is normally based on the quasi-chemical approach that is now formulated for interactions between functional groups. For this, all group-contribution methods assume that the functional groups are independent of each other, as illustrated in Fig. 2.3. This assumption means that information about the three-dimensional molecular structure is lost. Group-contribution methods are therefore unable to account for the full three-dimensional structure of molecules.

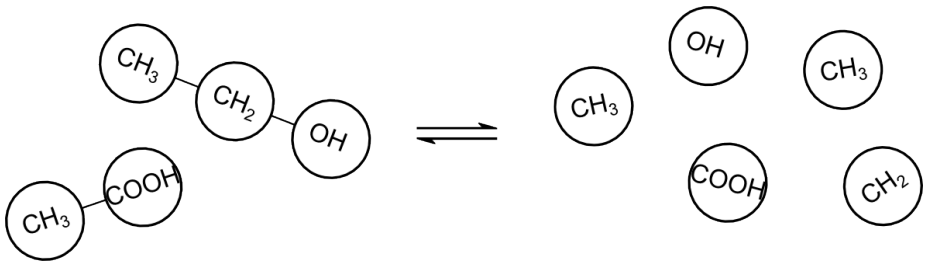


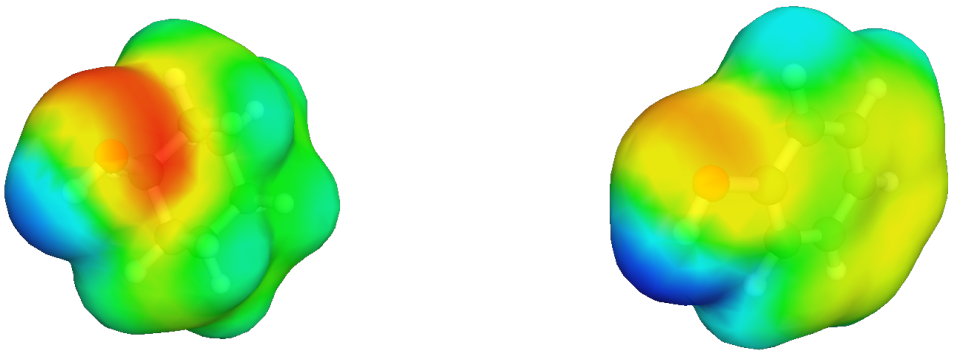
Figure 2.3: Assumption of independent groups in group-contribution models.

In this section, the popular UNIFAC model and two of its modifications are presented. This model shows the same inconsistencies concerning the local composition as the semi-empirical models presented in section 2.2.2. Furthermore, the GTASQUAC and DISQUAC models are briefly discussed. These are both physical consistent models. First, however, the definition of functional groups is discussed.

Definition of functional groups

The first step in developing a group-contribution model is the definition of functional groups. According to investigations by Wu and Sandler [32, 33], the definition of functional groups plays a major role in the performance of group-contribution models. They claim that the geometry of a functional group should be independent of the molecule in which the group occurs, each atom in a functional group should have approximately the same charge in all molecules in which the group occurs and the group should be approximately electroneutral. The charge of a functional group is determined by the sum of the partial atomic charges that are created due to the asymmetric distribution of electrons in chemical bonds. In addition, Wu and Sandler claim that each functional group should be the smallest entity such that a molecule can be divided into a collection of electroneutral groups. None of the above mentioned G^{E}-models define groups in a way which fulfills all these requirements [25].

Fig. 2.4 shows the surface-charge density of two molecules that both contain an OH-group. It can be clearly seen that the surface-charge density on the oxygen atom is different for both molecules. The OH-groups in both molecules are thus to be treated differently, since the charge of the O-atom is different in both groups. In models such as UNIFAC a distinction is therefore made between (a) an OH-group connected to an aliphatic and (b) an aromatic structure [34].

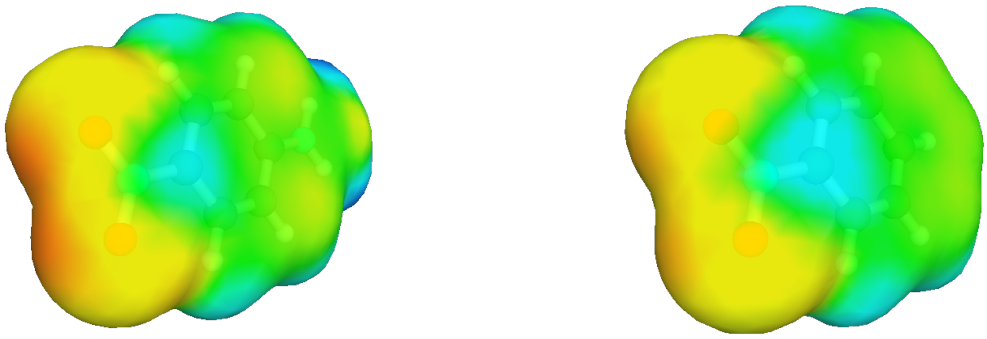


(a) cyclohexanol (b) phenol

Figure 2.4: Proximity effect on OH-group.

In the example in Fig. 2.4, the different charge on the O-atom is caused by the proximity of other groups. The phenyl group causes the charge of the oxygen atom to become delocalized, leading to a smaller charge of the O-atom in comparison to the O-atom in cyclohexanol. Fig. 2.5 shows another example of this so-called proximity effect [35]. Here, the amino-group in 4-aminonitrobenzene causes a slightly stronger polarization of the nitro-group in the same molecule in comparison to the nitro-group in nitrobenzene. In UNIFAC, a nitro-, an amino- and several aromatic CH-groups are distinguished for these molecules. However, to consider the effect of polarization of the nitro-group properly, a new group would have to be introduced containing both the nitro- and the amino-group as well as at least part of the phenyl-ring. The introduction of such large groups to account for all proximity effects in all molecules would, however, result in a dramatic increase in the number of groups and thus in the number of group-

interaction parameters. The advantage of the low number of parameters in the group-contribution method would therefore be lost. What is more, the interaction potential of such large groups might become so complex that it could no longer be adequately described [35].



(a) 4-aminonitrobenzene (b) nitrobenzene

Figure 2.5: Example of the proximity effect on a nitro-group.

Wu and Sandler's [32] requirement for the definition of groups, that the group should be approximately electroneutral, means that positively and negatively charged surface areas need to be summarized in one group. Nonetheless, this is a crude approximation of the true physical features of a group. The thermodynamic behavior of a system significantly depends on the interactions between the differently charged surface segments. This definition of groups therefore means that certain effects cannot be described adequately. An example of such effects are hydrogen bonds, since, e.g., the OH group is rather large. The surface area of a hydrogen atom in the OH group is significantly smaller and shows a high surface-charge density. It is this small part of the surface of the OH group that needs to be in contact with the part of the surface of a hydrogen-bond acceptor that also shows a high surface-charge density but of the opposite sign. Thus, for the formation of a hydrogen bond, a high order of molecular orientation is required. However, because of the relatively large OH group, this high order of orientation upon a hydrogen bond cannot be modeled by group-contribution methods. Surface-segment models that divide the molecular surface into segments according to the surface-charge density, as described in section 2.2.4, overcome this deficiency of group-contribution models.

The UNIFAC model

The UNIFAC model, first proposed by Fredenslund et al. [34], is the group-contribution equivalent of UNIQUAC that was described in section 2.2.2. The combinatorial term of UNIFAC is identical to the combinatorial term of UNIQUAC that was discussed in section 2.1.3. Formulated in terms of activity coefficients, it reads

$$\ln \gamma_{\mathrm{comb},i} = \ln \frac{\phi_i}{x_i} + \frac{z}{2} q_i \ln \frac{\psi_i}{\phi_i} + l_i - \frac{\phi_i}{x_i} \sum_{j=1}^{N_c} x_j l_j \tag{2.44}$$

with

$$l_i = \frac{z}{2}(r_i - q_i) - (r_i - 1) \tag{2.45}$$

Parameters r_i and q_i are calculated as the sum of the group volume and surface-area parameters R_k and Q_k

$$r_i = \sum_{k=1}^{N_{\mathrm{gr},i}} \nu_k^{(i)} R_k \tag{2.46}$$

$$q_i = \sum_{k=1}^{N_{\mathrm{gr},i}} \nu_k^{(i)} Q_k \tag{2.47}$$

where $N_{\mathrm{gr},i}$ is the number of different groups in component i. $\nu_k^{(i)}$ is the number of groups of type k in component i. R_k and Q_k are obtained from the group volume and surface area V_k and A_k given by Bondi [12]

$$R_k = \frac{V_k}{V_{\mathrm{ref}}} \tag{2.48}$$

$$Q_k = \frac{A_k}{A_{\mathrm{ref}}} \tag{2.49}$$

V_{ref} and A_{ref} are determined from Eqs. 2.16 and 2.17 with Eq. 2.18. z is assumed to equal 10.

For the residual contribution, instead of considering interactions between molecules as in UNIQUAC, interactions between functional groups are considered. The residual contribution in terms of activity coefficients is calculated by

$$\ln \gamma_{\mathrm{res},i} = \sum_{k=1}^{N_{\mathrm{gr}}} \nu_k^{(i)} \left(\ln \Gamma_k - \ln \Gamma_k^{(i)} \right) \tag{2.50}$$

where Γ_k is the group-activity coefficient, and $\Gamma_k^{(i)}$ is the activity coefficient of group k in a reference solution containing only molecules of type i.

The group-activity coefficient Γ_k is calculated by

$$\ln \Gamma_k = Q_k \left(1 - \ln \left(\sum_{m=1}^{N_{\mathrm{gr}}} \Psi_m \tau_{m,k} \right) - \sum_{m=1}^{N_{\mathrm{gr}}} \left(\frac{\Psi_m \tau_{k,m}}{\sum_{n=1}^{N_{\mathrm{gr}}} \Psi_n \tau_{n,m}} \right) \right) \tag{2.51}$$

where Ψ_m is the surface-area fraction of group m

$$\Psi_m = \frac{Q_m X_m}{\sum_{n=1}^{N_{\mathrm{gr}}} Q_n X_n} \tag{2.52}$$

with X_m as the mole fraction of group m

$$X_m = \frac{\sum_{i=1}^{N_{\mathrm{c}}} \nu_m^{(i)} x_i}{\sum_{i=1}^{N_{\mathrm{c}}} \sum_{n=1}^{N_{\mathrm{gr},i}} \nu_n^{(i)} x_i} \tag{2.53}$$

N_{gr} is the number of different groups in the mixture. The group-interaction parameter $\tau_{m,n}$ is calculated by

$$\tau_{m,n} = \exp\left(-\frac{a_{m,n}}{T} \right) \tag{2.54}$$

The group-interaction energy parameters $a_{m,n}$ are not symmetrical. Thus, per binary mixture of groups, the model has two parameters. On the webpage of the UNIFAC consortium, an overview of the group-interaction parameters determined so far is given [36]. Since many group-interaction parameters have not yet been determined, the applicability of UNIFAC as a universal predictive G^{E}-model is limited. The UNIFAC model also shows the same physical inconsistencies regarding the local composition as UNIQUAC, as discussed in section 2.2.2. Because of this inconsistency, the predictive quality of UNIFAC is limited.

Two modifications of UNIFAC that try to improve the predictive quality of UNIFAC have become popular. The modified UNIFAC (Dortmund) model by Weidlich et al. [17] and the modified UNIFAC (Lyngby) model by Larsen et al. [18] were proposed approximately simultaneously. Both the combinatorial as well as the residual contribution were modified empirically in both modifications. The modifications to the combinatorial contribution were discussed in section 2.1.3. The more relevant modifications are those made to the residual contribution. The group-interaction parameter $a_{m,n}$ is considered to be temperature dependent. In the modification by the Dortmund group, a polynomial approach is applied

$$a_{m,n} = A_{m,n} + B_{m,n}T + C_{m,n}T^2 \tag{2.55}$$

whereas in the modification by the Lyngby group the parameter is described by

$$a_{m,n} = A'_{m,n} + B'_{m,n}(T - T_0) + C'_{m,n}\left(T \ln \frac{T}{T_0} + T - T_0\right) \tag{2.56}$$

with T_0 as reference temperature. Larsen et al. [18] chose $T_0 = 298.15$ K. Because of this temperature dependence, both modifications show six parameters per binary mixture of groups. This large number of parameters increases the effort required for parameterization, but also makes both modifications extremely flexible for the description of thermophysical properties. Many of the parameters of the modified UNIFAC (Dortmund) model in particular have been determined by fitting to experimental data, resulting in this modification frequently being used in industrial applications. The webpage of the UNIFAC consortium gives an overview of the determined group-interaction parameters of the modified UNIFAC (Dortmund) model [36].

The GTASQUAC and DISQUAC model

UNIFAC and its modifications show the same physical inconsistency regarding the local composition as UNIQUAC. This inconsistency, however, allows the local composition to be solved analytically. This in turn results in shorter computation times and is therefore advantageous. Lacmann et al. [37] propose using a Taylor-series approximation for the quasi-chemical approach. For the resulting GTASQUAC model (Group contribution TAylor Series approximation for QUAsi-Chemical equilibria), the local compositions can also be determined analytically, and the model does not have the same inconsistency as the UNIFAC model. Because of the Taylor-series approximation, however, the model only gives an approximation of the quasi-chemical equations.

GTASQUAC distinguishes a combinatorial and a residual contribution to G^{E} using Eq. 2.1. The combinatorial term is identical to that of the modified UNIFAC (Lyngby) model, Eq. 2.22. For the residual term, first the excess enthalpy is defined. The local composition is defined by the quasi-chemical approach and is approximated with a Taylor series. Then the Gibbs-Helmholtz

equation

$$H^{\mathrm{E}} = \left(\frac{\partial \frac{G^{\mathrm{E}}}{T}}{\partial \frac{1}{T}} \right)_{p,x_i} \tag{2.57}$$

is integrated to obtain a rather long expression for the excess Gibbs energy [37]. The advantage of GTASQUAC is its physical consistency and the fact that it needs fewer parameters per group pair than the modified UNIFAC models.

A special contribution to the development of physical-consistent group-contribution methods was achieved by the work of Kehiaian and his co-workers in the scope of the TOM project (Thermodynamics of Organic Mixtures [38]) [35, 39–41]. They presented the DISQUAC model [39] that is based on Barker's theory [20], according to which interactions between molecules take place between contact points on the surface of molecules. Barker's theory can be considered as an early version of a group-contribution method that is physically consistent. Whereas Barker assumes specific contact points on the molecular surface, Kehiaian assigns these contact points to a contact area that basically corresponds to a functional group.

Apart from the combinatorial contribution, two more contributions to G^{E} are distinguished in the DISQUAC model:

$$G^{\mathrm{E}} = G^{\mathrm{E}}_{\mathrm{comb}} + G^{\mathrm{E}}_{\mathrm{disp}} + G^{\mathrm{E}}_{\mathrm{int}} \tag{2.58}$$

$G^{\mathrm{E}}_{\mathrm{disp}}$ describes dispersive interactions and $G^{\mathrm{E}}_{\mathrm{int}}$ describes stronger interactions such as polar interactions. $G^{\mathrm{E}}_{\mathrm{int}}$ is based on the quasi-chemical approach.

Although Gonzales et al. [42–44] showed that DISQUAC can describe polar as well as associating mixtures, the primary use of the model is not the comprehensive description of different phase equilibria. Instead, DISQUAC allows an intensive analysis of thermophysical properties of mixtures of components of a homologous series, since functional groups are mainly defined according to their functionality. These functional groups are different from the groups in UNIFAC. Kehiaian and Marongui [45] showed that DISQUAC, in contrast to UNIFAC, allows a detailed investigation of interaction parameters in dependence of molecular structure and can thus take proximity effects better into account. This is why DISQUAC is especially suited to describing effects of the molecular structure, since the proximity of other functional groups is accounted for in the interaction-energy parameters. Although the proximity of other groups is taken into account for the interaction-energy parameters, to describe the excess properties DISQUAC also assumes that the functional groups are independent. For this reason, DISQUAC also cannot take all the information on the three-dimensional molecular structure into account and can therefore not describe effects like multiple contact points. What is more, the model is very complex and only a small number of interaction parameters have been determined so far, which is why the model does not allow for a broad range of application.

2.2.4 Surface-segment models for the excess Gibbs energy

Section 2.2.3 showed that the definition of universal functional groups is problematic. Proximity effects require that more and sometimes bigger functional groups are defined. This, however, makes the group-contribution method less efficient and leads to difficulties in the description of the group-interaction potential. Furthermore, the requirement for functional groups to be approximately electroneutral requires positively and negatively charged surface areas to be summarized in one functional group [32]. This causes difficulties for the description of hydrogen bonds, among other things.

Surface-segment models overcome these difficulties by distinguishing only surface segments with different surface-charge densities. This way a functional group can be resolved in more detail. Molecular contacts are modeled as contacts between surface segments. This approach is very similar to Barker's idea of distinguishing contact points on the molecular surface [20]. Surface-segment models relate a surface area to these contact points, where the surface area determines the contact probability of the segment.

The most popular surface-segment model with a wide range of application in chemical industries is the COSMO-RS model by Klamt [46]. This model shows only universal parameters and requires information from an a-priori quantum-chemical calculation with the COSMO model [47]. Because of this, the model is predictive and applicable to almost any kind of substance. The thermodynamic part of COSMO-RS is identical to GEQUAC [46]. The formulation of GEQUAC, however, is more similar to the formulation of the quasi-chemical equations of section 2.2.1 and will therefore be discussed first. It will be shown that when the set of non-linear equations of GEQUAC is reformulated according to a solution algorithm by Larsen and Rasmussen [48], the typical COSMO-RS formulation results. Since the quasi-chemical theory is applied to the surface segments, the surface segments are considered to be independent. In the case of the surface-segment models this approximation is referred to as the free-segment approximation. Because of this free-segment approximation, surface-segment models also cannot take the full three-dimensional structure of molecules into account.

The GEQUAC model

GEQUAC (Group-surface Explicit QUAsi-Chemical theory) was developed to enable a physically founded description of both non-associating and associating liquid mixtures. In doing this, the weak preference of certain contacts according to the quasi-chemical theory was strengthened, in such a way that not only polar interactions but also the behavior of associating mixtures could be described [49–51].

For GEQUAC also a combinatorial and a residual contribution to the excess Gibbs energy according to Eq. 2.1 are distinguished. The combinatorial term is identical to the combinatorial term of UNIQUAC that was discussed in section 2.1.3. For the residual term, N_s different surface-segment types are distinguished. The residual Gibbs energy is calculated by

$$G_\mathrm{res} = \frac{1}{4} z q_\mathrm{tot} \sum_{m=1}^{N_\mathrm{s}} \sum_{n=1}^{N_\mathrm{s}} \psi_m \psi_{n,m} \left(\omega^\mathrm{h}_{n,m} - T\,\omega^\mathrm{s}_{n,m} + 2RT \ln \frac{\psi_{n,m}}{\psi_n} \right) \tag{2.59}$$

and needs to be evaluated both for the mixture and the pure components. The residual excess Gibbs energy is then given by

$$G^\mathrm{E}_\mathrm{res} = G_\mathrm{res} - \sum_{i=1}^{N_\mathrm{c}} x_i G^0_{\mathrm{res},i} \tag{2.60}$$

Equation 2.60 does not contain the contribution by the ideal mixture to the Gibbs energy, since this contribution is already included in the combinatorial term $G^\mathrm{E}_\mathrm{comb}$. q_tot is calculated by

$$q_\mathrm{tot} = \sum_{i=1}^{N_\mathrm{c}} x_i q_i \tag{2.61}$$

and ψ_n is the surface-area fraction of surface-segment type n in the mixture that is defined analogous to the surface-area fraction of groups in the UNIFAC model, Eq. 2.52. $\psi_{n,m}$ is the normalized fraction of contacts originating from surface segments of type m with surface segments of

type n. $z\omega^{\mathrm{h}}_{n,m}$ is the interchange enthalpy and $z\omega^{\mathrm{s}}_{n,m}$ the interchange entropy of an m-n contact. The ω parameters are defined as

$$\omega^{\mathrm{h}}_{n,m} = h_{n,m} + h_{m,n} - h_{m,m} - h_{n,n} \tag{2.62}$$

and

$$\omega^{\mathrm{s}}_{n,m} = s_{n,m} + s_{m,n} - s_{m,m} - s_{n,n} \tag{2.63}$$

with $h_{m,n}$ and $s_{m,n}$ as the interaction enthalpy and entropy of the m-n contact. Symmetry applies, thus

$$h_{n,m} = h_{m,n} \tag{2.64}$$

and

$$s_{n,m} = s_{m,n} \tag{2.65}$$

Equations 2.24 and 2.25 from the quasi-chemical approach are adapted to a formulation with surface-area fractions to

$$\sum_{n=1}^{N_{\mathrm{s}}} \psi_{n,m} = 1 \quad \text{for all} \quad m = 1, \ldots N_{\mathrm{s}} \tag{2.66}$$

and

$$\psi_m \psi_{n,m} = \psi_n \psi_{m,n} \tag{2.67}$$

respectively. The contact fractions $\psi_{n,m}$ are determined by minimizing the Gibbs energy. This leads to the condition that

$$\left(\frac{\partial G_{\mathrm{res}}}{\partial \psi_{m,n}} \right)_{T,p,x_i,\psi_{o,p} \neq \psi_{m,n}} = 0 \tag{2.68}$$

Equations 2.66 and 2.67 need to be considered for the derivative in Eq. 2.68. This results in an equation similar to Eq. 2.32:

$$\frac{\psi_{n,m}\psi_{m,n}}{\psi_{m,m}\psi_{n,n}} = \exp\left(-\frac{\omega^{\mathrm{h}}_{n,m} - T\omega^{\mathrm{s}}_{n,m}}{RT} \right) \tag{2.69}$$

A detailed derivation of the GEQUAC equations is given by Egner [49]. Equations 2.66, 2.67 and 2.69 form a non-linear set of equations, for which Larsen and Rasmussen proposed a modified Newton-Raphson solution algorithm [48]. Details on how to solve the GEQUAC-equation system with this algorithm are given in appendix 7.1.

In the case of GEQUAC, the parameter b_m that was introduced by Larsen and Rasmussen can be considered the activity coefficient of surface-segment type m. The chemical potential of surface-segment type m is then given by

$$\mu^{\mathrm{s}}_m = RT \ln b_m \tag{2.70}$$

The reformulation of the quasi-chemical equations according to the Newton-Raphson method which was modified by Larsen and Rasmussen results in

$$\mu^{\mathrm{s}}_m = -RT \ln\left(\sum_{n=1}^{N_{\mathrm{s}}} \psi_n \exp \frac{\mu^{\mathrm{s}}_n - \frac{1}{2}\omega^{\mathrm{h}}_{n,m} + \frac{1}{2}T\omega^{\mathrm{s}}_{n,m}}{RT} \right) \tag{2.71}$$

which is an implicit equation for μ^{s}_m. This segment-based formulation of the quasi-chemical equations is applied in COSMO-RS, which will be described in the following section.

The COSMO and COSMO-RS model

In 1995, Klamt presented the COnductor-like Screening MOdel for Real Solvents (COSMO-RS) [52] as an improvement on his COSMO model [47]. COSMO belongs to the class of so-called continuum solvation models that neglect the atomic structure of the solvent and treat it as a dielectric continuum with the permittivity ε. Such models allow investigation of the influence of the solvent on the charge distribution of a molecule due to electrostatics, but are not able to describe all effects of real solvents.

In COSMO, a solute molecule is considered to be situated inside a cavity in an ideal conductor. For the cavity construction, each atom type is assigned a radius and the union of the corresponding atom-centered spheres is considered as the interior of the cavity. Since this means the cavity can have defects at the intersection between two atoms, a smoothing algorithm is applied. Details about cavity construction for COSMO are given by Klamt [46].

The surface of the cavity is considered to be the solvent accessible surface. The electric field arising from the nuclei as well as from the electrons of the solute molecule is screened by the polarization of the continuum. The effect of this polarization can be represented by the surface-charge density distribution it produces on the inner surface of the cavity. To do this, the area of the solvent-accessible surface is divided into segments and for each segment a screening charge is determined. The choice of an ideal conductor for the continuum leads to a remarkably simple expression for the screening charges and the screening energy, because the resulting electrostatic potential Φ must equal zero for every point $\vec{r}$ on the surface of the cavity:

$$\Phi(\vec{r}) = 0 \tag{2.72}$$

The charge distribution on the surface of the cavity influences the charge distribution of the molecule. Quantum mechanics is applied to determine the structure of the molecule as well as its charge distribution inside the cavity. These are both consistent with the charge distribution on the surface of the cavity. For this quantum-mechanical calculation, there are different levels that can be applied. A very popular method is to apply density functional theory (DFT) [53, 54].

COSMO thus provides the molecular structure, the molecular volume (cavity volume), and a segmented molecular surface with its charge-density distribution and surface area. Klamt derived a model to describe the Gibbs energy based on this information [52]. A crucial assumption of COSMO-RS is that the surface segments that are used for the calculation can be considered to be independent and to have a constant charge density. To determine the charge density of the surface segments for the COSMO-RS calculation from the COSMO surface-charge densities $\tilde{\sigma}$, an averaging according to

$$\sigma_\nu = \frac{\sum_\mu \frac{\tilde{\sigma}_\mu r_\mu^2 r_{\mathrm{av}}^2}{r_\mu^2 + r_{\mathrm{av}}^2} \exp\left(-\frac{d_{\mu,\nu}^2}{r_\mu^2 + r_{\mathrm{av}}^2}\right)}{\sum_\mu \frac{r_\mu^2 r_{\mathrm{av}}^2}{r_\mu^2 + r_{\mathrm{av}}^2} \exp\left(-\frac{d_{\mu\nu}^2}{r_\mu^2 + r_{\mathrm{av}}^2}\right)} \tag{2.73}$$

is performed [55]. This averaging also helps to get rid off artifacts from the COSMO calculation due to the cavity surface not always being completely closed [56]. r_{av} is an averaging radius that is fitted to experimental data and that characterizes the size of an effective surface segment. r_μ is the radius of a circle with the same surface area as the surface segment μ and $d_{\mu,\nu}$ is the distance between surface segments μ and ν.

Because of the so-called free segment approximation, surface segments with an equal surface-charge density can now be merged. A distribution function $p(\sigma)$ is introduced that describes the amount of surface in a system with a surface-charge density between σ and $\sigma + d\sigma$. This

distribution function is called the σ-profile. The σ-profile of a mixture can be determined from the σ-profiles of the pure components $p_i(\sigma)$

$$p(\sigma) = \sum_i^{N_c} x_i p_i(\sigma) \tag{2.74}$$

For COSMO-RS, a combinatorial and a residual contribution to G^E are distinguished using Eq. 2.1. For the chemical potential of a component i, this results in

$$\mu_i = \mu_{\text{comb},i} + \mu_{\text{res},i} \tag{2.75}$$

As for GEQUAC, the combinatorial contribution to the chemical potential is determined using a Staverman-Guggenheim equation

$$\mu_{\text{comb},i} = -RT \left(\lambda \ln \bar{A}_{\text{cosmo}} + 1 - \frac{V_{\text{cosmo},i}}{\bar{V}_{\text{cosmo}}} + \ln \frac{V_{\text{cosmo},i}}{\bar{V}_{\text{cosmo}}} + \frac{z_{\text{cosmo}}}{2} \frac{A_{\text{cosmo},i}}{A_{\text{cosmo,ref}}} \left(1 - \frac{V_{\text{cosmo},i} \bar{A}_{\text{cosmo}}}{\bar{V}_{\text{cosmo}} A_{\text{cosmo},i}} + \ln \frac{V_{\text{cosmo},i} \bar{A}_{\text{cosmo}}}{\bar{V}_{\text{cosmo}} A_{\text{cosmo},i}} \right) \right) \tag{2.76}$$

where $A_{\text{cosmo,ref}}$ is a reference surface area. Klamt defined it as the partial surface area of an ethylene unit [57]. Instead of using the molecular surface area and volume as in UNIQUAC and the group-contribution methods, the volume and surface area of the cavity of the COSMO calculation are used for the molecular surface area and volume $A_{\text{cosmo},i}$ and $V_{\text{cosmo},i}$ of component i. λ and z_{cosmo} are considered parameters that were determined from a fit to experimental data [57]. The average molecular surface area and volume in the mixture are defined as

$$\bar{A}_{\text{cosmo}} = \sum_{i=1}^{N_c} x_i A_{\text{cosmo},i} \tag{2.77}$$

and

$$\bar{V}_{\text{cosmo}} = \sum_{i=1}^{N_c} x_i V_{\text{cosmo},i} \tag{2.78}$$

For the residual term, Klamt derived an equation without knowing about the quasi-chemical approach [46]. His original notation therefore deviates from typical quasi-chemical notation. For the residual part of the chemical potential of a surface-segment type with the surface-charge density σ, he writes

$$\mu^s(\sigma) = -RT \ln \left(\int \mathrm{d}\sigma' p(\sigma') \exp \left(\frac{\mu^s(\sigma') - a_{\text{eff}} e(\sigma, \sigma')}{RT} \right) \right) \tag{2.79}$$

and with the chemical potential of all segment-types, the residual chemical potential of component i is calculated by

$$\mu_{\text{res},i} = \int p_i(\sigma) \mu^s(\sigma) \, \mathrm{d}\sigma \tag{2.80}$$

The formulation of Eq. 2.79 is similar to the formulation of GEQUAC in Eq. 2.71. The differences between both equations are the approximation of the integral by a sum and the interaction term. COSMO-RS with $e(\sigma, \sigma')$ only considers an energy contribution, whereas GEQUAC additionally considers an entropic contribution. Since the entropy parameter $\omega^s_{n,m}$ of GEQUAC,

however, cannot be determined a-priori, it is often set to zero [25]. Because of the ω term, Eq. 2.71 includes the interaction energies between surface segments of the same type that are not included in Eq. 2.79. This difference is caused by the reference state of pure surface segments that was introduced for GEQUAC [50]. However, this reference state is irrelevant for the excess properties and the solution of the quasi-chemical equations. The thermodynamics of COSMO-RS are thus identical to that of GEQUAC, as was also shown by Klamt [46].

a_{eff} in Eq. 2.79 is defined as the average molecular contact area and is a model parameter that was fitted to experimental data. One would expect that a_{eff} is related to the radius r_{av} of an effective surface segment of Eq. 2.73. However, for reasons that have not yet been identified, much better results for COSMO-RS are obtained when both parameters are considered to be independent [46].

The interaction energy of a molecular contact $e(\sigma,\sigma')$ depends on the charge density σ and σ' of the two surface segments involved and consists of two terms

$$e(\sigma,\sigma') = e_{\text{misfit}}(\sigma,\sigma') + e_{\text{hb}}(\sigma,\sigma') \tag{2.81}$$

The reference state of COSMO-RS is that of the ideally screened system (COSMO). A system is ideally screened even if around each molecule only a very thin layer of ideal conductor is present [57]. To describe the difference between such an ideally screened system and a real system, the conductor has to be removed from the system completely. For a surface segment contact, the amount of energy required to remove the piece of ideal conductor between the two surface segments is therefore calculated. This energy is called the misfit energy [57] and is calculated by

$$e_{\text{misfit}}(\sigma,\sigma') = \frac{\alpha'}{2}\left(\sigma+\sigma'\right)^2 \tag{2.82}$$

Since the system's energy is minimal in the ideally screened state, the misfit term is always positive. α' is a model parameter and is fitted to experimental data.

If the surface-charge densities of the two surface segments are sufficiently large and have an opposite sign, a hydrogen bond can occur. The interaction energy of such a contact exceeds the normal polar interaction energy. In COSMO-RS, this excess interaction energy is accounted for by an additional hydrogen bonding term

$$e_{\text{hb}}(\sigma,\sigma') = c_{\text{hb}} \min\left(0, \min\left(0, \sigma_{\text{don}} + \sigma_{\text{hb}}\right) \max\left(0, \sigma_{\text{acc}} - \sigma_{\text{hb}}\right)\right) \tag{2.83}$$

with

$$\sigma_{\text{don}} = \min(\sigma,\sigma') \tag{2.84}$$

and

$$\sigma_{\text{acc}} = \max(\sigma,\sigma') \tag{2.85}$$

c_{hb} and σ_{hb} are both universal parameters that are fitted to experimental data. σ_{hb} is a threshold for hydrogen bonding. Eq. 2.83 thus does not contribute to the interaction energy, unless the more negative of the two screening charge densities is less than the threshold - σ_{hb}, and unless the more positive exceeds σ_{hb}.

Due to the popularity of the COSMO-RS model in chemical engineering, a number of reimplementations have been developed [58–61]. Although these reimplementations differ slightly with respect to parameterization and the details of the implementation, most of them are referred to as COSMO-RS in the literature. All reimplementations developed so far are less complete and less detailed than COSMOtherm [62]. All available comparisons indicate that COSMOtherm is more accurate than these reimplementations [63].

3 Improvement of the combinatorial contribution

The combinatorial contribution to the excess Gibbs energy describes mixing effects that are only caused by differences in size and shape and not by interaction energies. This contribution therefore only accounts for entropic effects and not for enthalpic effects. It needs to be considered for all mixtures with components that significantly differ in size and shape.

In UNIQUAC [11], the combinatorial contribution is described with the physically founded Staverman-Guggenheim equation, Eq. 2.13. The UNIQUAC combinatorial term, however, cannot always adequately describe combinatorial effects in real mixtures. Because of this, various empirical modifications for the combinatorial term have been proposed. These were discussed in section 2.1.3. Since such empirical modifications are not satisfactory, here a physically well-founded modification of the UNIQUAC combinatorial term is derived.

3.1 Modification of the UNIQUAC combinatorial term

Pfennig [64] investigated the Guggenheim combinatorial entropy term by comparing results from this model with results from lattice-based computer simulations. He showed that the model performs perfectly well for inflexible chain molecules. This was not surprising, since the Guggenheim model was derived exactly for such a lattice system and molecules that do not fold back on themselves. This agreement between lattice simulation and the Guggenheim model shows that the model equation is in principle well suited to describe the combinatorial contribution of chain molecules in a mixture. Thus, one would not necessarily expect that changing the model structure in an empirical fashion would improve performance.

The question thus arises why good results are obtained when the Staverman-Guggenheim model is compared to results from computer simulations and why the UNIQUAC combinatorial term compares badly with experimental results? A closer look at Eqs. 2.4 and 2.12 reveals that the reference volume and area cancel out in these equations, while the reference area does not in Eq. 2.14. Therefore, the second term in Eq. 2.13 depends on the choice of the standard segment. z is usually set to 10, which is not problematic, since it is zq_i that needs to be determined meaningfully. As a consequence of these considerations, $A_{\mathrm{vdW,ref}}$ should actually be regarded as a model parameter that cannot be set arbitrarily, as was done by Abrams and Prausnitz [11]. Therefore, here $A_{\mathrm{vdW,ref}}$ is considered a model parameter and it is replaced in the equation by a general reference surface area A^* of a new standard segment. This modification allows an improvement in the application of the Staverman-Guggenheim model to real mixtures, while maintaining the physically founded form of the model structure.

To keep the change to the model equation minimal, Eq. 2.13 can be reformulated to

$$G^{\mathrm{E}}_{\mathrm{comb}} = RT\sum_i^{N_c} x_i \ln\frac{\phi_i}{x_i} + \frac{1}{2}RT\sum_i^{N_c} z\frac{A_{\mathrm{vdW,ref}}}{A^*}q_i x_i \ln\frac{\psi_i}{\phi_i} \tag{3.1}$$

where q_i, ϕ_i and ψ_i are determined as in UNIQUAC.

The new reference surface now needs to be determined by fitting Eq. 3.1 to experimental data of molecules that differ in size. It is then assumed that the value of the new reference surface does not depend any further on the system that is considered and that A^* is thus a general constant in the equation. Since the combinatorial term does not take any energetic effects of mixtures and pure components into account, data of mixtures that behave nearly athermally should be used for the fit, because for such systems the influence of A^* is maximal.

3.2 Experimental data

No real binary mixture behaves completely athermally. Mixtures of *n*-alkanes, however, behave nearly athermally. Fig. 3.1 shows the heat of mixing of a binary mixture consisting of *n*-hexane and *n*-dodecane at various temperatures. Note that the heat of mixing is small, as it should be for an almost athermal mixture. The behavior of systems of *n*-alkanes that differ in size is thus dominated by effects that are described with the combinatorial term and so these mixtures were used to determine the reference surface of the new standard segment. For the fit, experimental H^{E} and γ^{∞} data of binary *n*-alkane mixtures were used. All data were taken from the DECHEMA Chemistry Data Series [15, 65].

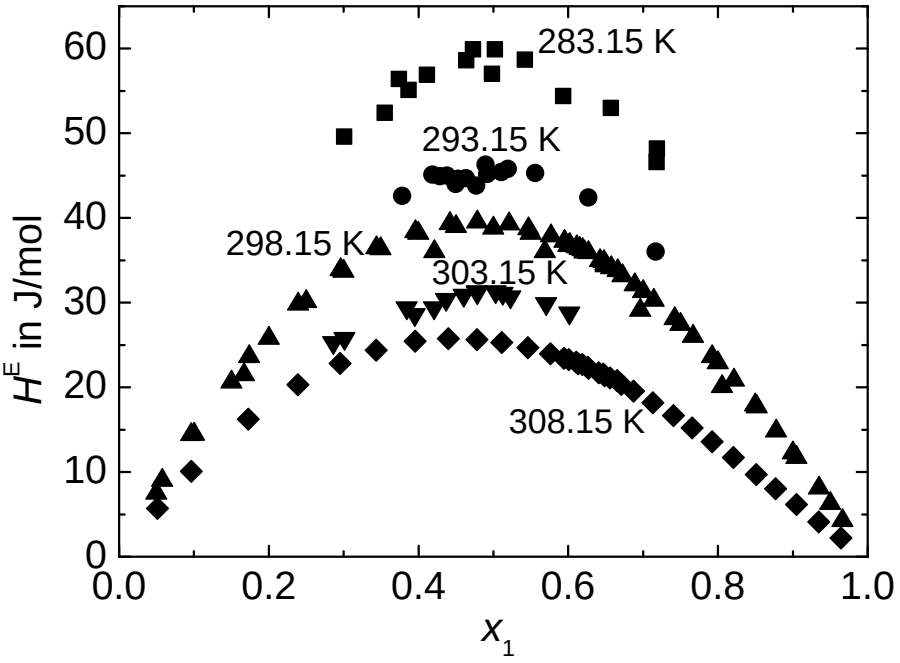


Figure 3.1: Experimental heat of mixing data for the binary *n*-hexane (1) + *n*-dodecane (2) mixture. All experimental data were taken from the DECHEMA data series [65].

As chain length increases for *n*-alkanes, the molecular flexibility can lead to a back bending of the chain to form intramolecular contacts. This changes the number of external contacts zq_i of a molecule, as discussed in Pfennig [64]. This effect of the flexibility of molecules is not

accounted for in the Staverman-Guggenheim combinatorial term as it is used here, so to obtain accurate results, data of long *n*-alkanes were omitted from the fit. Analyzing the tendency for backbending simulated with a geometric molecular model showed that the effect on zq_i may still be acceptable for *n*-hexadecane [64] and since many data for systems with *n*-hexadecane are available, *n*-hexadecane was still included in the data set for fitting. Thus, *n*-butane was the shortest and *n*-hexadecane the longest *n*-alkane considered.

All data used for the fit were carefully selected. The procedure for selecting the γ^∞ data involved performing a preliminary fit of Eq. 3.1 to all the data for one solute and comparing the model results with the experimental values. First, the model was checked for systematic deviations. For example, the model should not underestimate the experimental values for short solvents and overestimate them for long solvents. Such behavior was not detected, so next the data points for individual solvents were considered. If a data point clearly differed from the others, it was excluded from the data set. If several data points from the same reference were found to deviate from the rest, all of the data points of that reference were excluded from the fit, since in this case systematical errors in the remaining data points could not be excluded.

Since a considerable amount of data is required to determine whether the model shows systematic deviations or a data point differs from the rest, this selection procedure could only be performed for the solutes *n*-pentane up to *n*-octane. For the other solutes, the model was fitted to the data for *n*-pentane to *n*-octane and then extrapolated. Since the model is extrapolated, the selection criteria for these solutes were relaxed slightly. In total, about 23% of the original data were excluded by the selection procedure.

The H^E data cannot be easily depicted graphically, since in addition to the temperature and the type of components, the composition of the mixture varies. Instead of evaluating all the data for mixtures with one common component simultaneously, all binary systems were evaluated individually. To do so, all data points of a binary system were plotted in a single diagram. If enough data points were available to unambiguously identify data points that differed from the rest, these data points were excluded from the data set.

Tab. 3.1 shows the number of data points of activity coefficients at infinite dilution that were used for the fit after the selection procedure. The temperature of the data points varies between 280.15 K and 453.15 K. Tab. 3.2 shows the number of selected experimental data points for the heat of mixing of binary *n*-alkane mixtures. The temperature of these data points varies between 283.15 K and 349.15 K. In appendix 7.2, two tables that contain all data sets that were used for the fit can be found.

Table 3.1: Number of selected experimental data points of activity coefficients at infinite dilution of binary *n*-alkane mixtures. Temperature range between 280.15 K and 453.15 K.

solvent	solute						
	C4	C5	C6	C7	C8	C9	C10
C5	0	0	0	0	1	0	0
C6	0	1	0	0	0	0	0
C7	0	2	6	0	1	0	0
C8	0	6	3	0	0	0	0
C9	0	1	1	0	0	0	0
C10	0	0	5	0	0	0	0
C12	0	8	0	0	0	0	0
C16	4	14	60	35	13	9	2

Table 3.2: Number of selected experimental data points of heat of mixing of binary *n*-alkane mixtures. Temperature range between 283.15 K and 349.15 K.

component 1	component 2				
	C7	C8	C10	C12	C16
C5	0	0	3	0	0
C6	22	13	88	142	172
C7	0	0	0	19	16
C8	0	0	0	30	3
C10	0	0	0	19	4

3.3 Residual term

Although the heat of mixing of *n*-alkane mixtures is low, it still has to be accounted for if the reference surface of the new standard segment is to be determined accurately. Certain entropic excess contributions due to the small but strongly direction-dependent molecular potentials also have to be considered, as discussed by Patterson et al. [66, 67]. Therefore, the new combinatoric term in Eq. 3.1 is supplemented with a residual term intended to cover these effects. The residual term is developed empirically and only to reproduce well the selected experimental data for the fit. The resulting model is then fitted to the selected experimental data to obtain the value for the reference surface of the new standard segment.

The residual term was developed on the basis of a simple regular solution model, in which CH_2 and CH_3 segments of *n*-alkanes are differentiated energetically. For a binary mixture, this model reads

$$H^{\mathrm{E}} = \frac{1}{2} z \tilde{\omega} q_1 q_2 \frac{x_1 x_2}{x_1 q_1 + x_2 q_2} (\theta_{13} - \theta_{23})^2 \tag{3.2}$$

where θ_{i3} is the surface fraction of the CH_3 groups in the molecule i. $z\tilde{\omega}$ characterizes the interchange energy. $\tilde{\omega}$ is defined similarly to Eq. 2.33 by

$$\tilde{\omega} = 2\varepsilon_{23} - \varepsilon_{22} - \varepsilon_{33} \tag{3.3}$$

where ε_{23} is the interaction energy between a CH_2 and a CH_3 group. ε_{22} and ε_{33} are the interaction energies between two CH_2 and two CH_3 groups respectively.

It was shown by de Matos Alves [68] that Eq. 3.2 shows systematic errors when describing the heat of mixing of the selected *n*-alkane mixtures. These systematic errors result from the fact that Eq. 3.2 produces asymmetric curves when plotted over x_1 whereas most of the excess enthalpy data are symmetric. For example, in Fig. 3.1 it can be seen that there is only slight deviation from a symmetric course for the heat of mixing of the system *n*-hexane + *n*-dodecane.

Analysis showed that the overestimated asymmetry is the result of the denominator in Eq. 3.2. Thus, to avoid overestimated asymmetry of the model, the denominator of Eq. 3.2 was omitted, which can be regarded as skipping the q_i in the denominator. Then, with the available experimental H^{E} data, a correlation for $\tilde{\omega}$ was developed as a function of T and q_i. To do so, several plots of the experimental data over different variables were studied and several correlations were derived from these plots. These correlations were all investigated further. The best result was obtained with

$$H^{\mathrm{E}}_{\mathrm{res}} = \frac{1}{2} z x_1 x_2 q_1 q_2 (\theta_{13} - \theta_{23})^2 \left(a + b (q_1 - q_2) q_1^2 + \frac{c}{T} \right) \tag{3.4}$$

where a, b, and c are parameters. Eq. 3.4 was integrated using the Gibbs-Helmholtz equation, Eq. 2.57, to obtain a term for the residual excess Gibbs energy. The integration constant was determined by using the fact that for

$$\hat{T} = -\frac{c}{a + b(q_1 - q_2)q_1^2} \tag{3.5}$$

$H^{\mathrm{E}}_{\mathrm{res}}(x_1, \hat{T})$ equals zero for all x_1. This means that at $T = \hat{T}$ the mixture behaves ideally, since it shows no heat of mixing. Because of the lack of energetic effects in the mixture at this temperature, no energy related entropy effects can occur. This leads to the condition that $S^{\mathrm{E}}_{\mathrm{res}}(x_1, T)$ and, therefore, $G^{\mathrm{E}}_{\mathrm{res}}(x_1, T)$ must equal zero at $T = \hat{T}$. For $G^{\mathrm{E}}_{\mathrm{res}}(x_1, T)$

$$G^{\mathrm{E}}_{\mathrm{res}} = \frac{1}{2} z x_1 q_1 x_2 q_2 \left(\theta_{13} - \theta_{23}\right)^2 \left(a + b\left(q_1 - q_2\right) q_1^2 + \frac{c}{T} - T\left(\frac{c}{2T^2} - \frac{\left(a + b\left(q_1 - q_2\right) q_1^2\right)^2}{2c}\right)\right) \tag{3.6}$$

resulted.

Fitting the parameters a, b, and c to the selected experimental H^{E} data results in an average absolute deviation of 3.51 J/mol per data point. Fig. 3.2 shows a comparison of experimental heat of mixing data with results from Eq. 3.4 for the binary *n*-hexane + *n*-dodecane system. The correlation describes the experimental data well. In particular, the temperature dependence is properly depicted. Only at the lowest temperature are the experimental data slightly overestimated, and at the highest temperature the correlation shows a slight deviation due to the asymmetry of the experimental data that cannot be represented by the model.

A comparison of the result with Eq. 3.2, which is also included in Fig. 3.2, shows that Eq. 3.4 is a considerable improvement, since Eq. 3.2 overestimates the asymmetry dramatically and cannot describe temperature dependence. However, it should be stressed again that the aim here was only to find a suitable expression that represents the excess over the combinatorial contribution as accurately as possible in order to gain deeper insight into the behavior of the available combinatorial models. Thus, Eq. 3.4 can only be regarded as an auxiliary function with little molecular-thermodynamic meaning, but with the ability to represent reasonably the selected data for the fit with a minimal number of parameters.

3.4 Results and discussion

Substituting Eqs. 3.1 and 3.6 into Eq. 2.1 results in:

$$G^{\mathrm{E}} = \frac{1}{2} z x_1 q_1 x_2 q_2 \left(\theta_{13} - \theta_{23}\right)^2 \left(a + b\left(q_1 - q_2\right) q_1^2 + \frac{c}{T} - T\left(\frac{c}{2T^2} - \frac{\left(a + b\left(q_1 - q_2\right) q_1^2\right)^2}{2c}\right)\right) + RT \sum_i^{N_c} x_i \ln \frac{\phi_i}{x_i} + \frac{1}{2} RT \sum_i^{N_c} z \frac{A_{\mathrm{vdW,ref}}}{A^*} q_i x_i \ln \frac{\psi_i}{\phi_i} \tag{3.7}$$

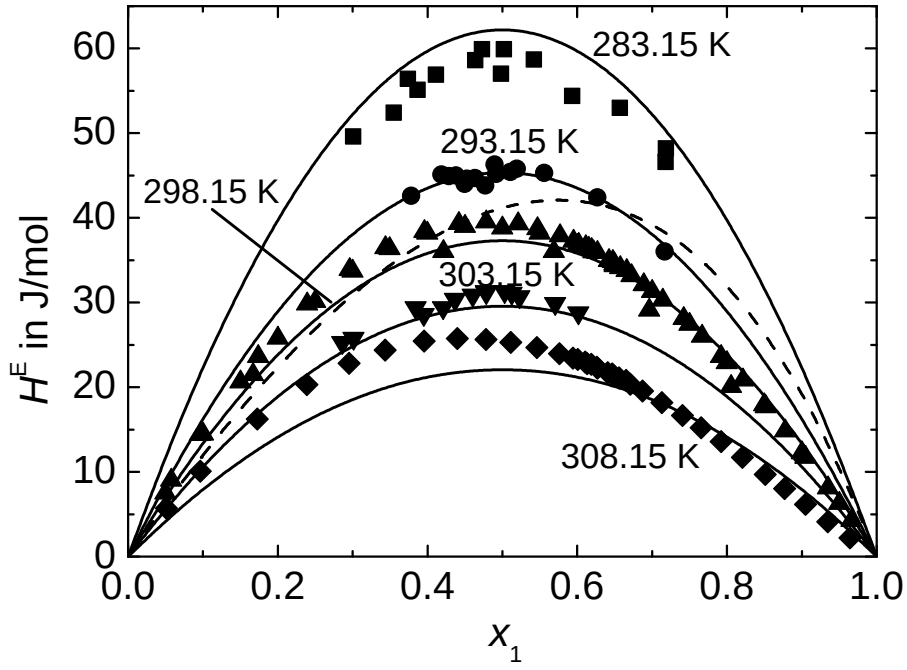


Figure 3.2: Comparison of experimental heat of mixing data with results from Eq. 3.4 (solid lines) and Eq. 3.2 (dashed line) for the binary *n*-hexane (1) + *n*-dodecane (2) mixture. All experimental data were taken from the DECHEMA data series [65].

The reference surface of the new standard segment A^* and the correlation parameters a, b, and c were fitted to the selected H^E and γ^∞ data simultaneously. The Levenberg-Marquardt algorithm (LMDIF) [69] and the objective function

$$\Delta = \frac{1}{\Delta_{H^E}} \sum_i \left(H^E_{\mathrm{exp},i} - H^E_{\mathrm{calc},i} \right)^2 + \frac{1}{\Delta_{\gamma^\infty}} \sum_j \left(\frac{\gamma^\infty_{\mathrm{exp},j} - \gamma^\infty_{\mathrm{calc},j}}{\gamma^\infty_{\mathrm{exp},j}} \right)^2 \tag{3.8}$$

were used.

The H^E data were fitted on the basis of absolute deviations, because when relative deviations are used, the fit tends to reproduce systems with small H^E values much better than systems with higher H^E values. The γ^∞ data were fitted on the basis of relative deviations. The contributions of H^E to the objective function were weighted with a factor

$$\Delta_{H^E} = \sum_i \left(H^E_{\mathrm{exp},i} - H^E_{\mathrm{calc},i} \right)^2 \tag{3.9}$$

which equals the sum of squared residuals resulting from fitting Eq. 3.7 only to the experimental H^E data. The contributions of γ^∞ to the objective function were weighted with

$$\Delta_{\gamma^\infty} = \sum_j \left(\frac{\gamma^\infty_{\mathrm{exp},j} - \gamma^\infty_{\mathrm{calc},j}}{\gamma^\infty_{\mathrm{exp},j}} \right)^2 \tag{3.10}$$

which equals the sum of squared residuals resulting from fitting Eq. 3.7 only to the experimental γ^∞ data. In this way, the different number of data for the different properties as well as their considerably different magnitudes of deviation are accounted for.

Table 3.3 shows the mean squared residuals for the different combinatorial terms. Tab. 3.3 also shows the mean squared residuals for Eq. 3.7 after the fit as well as the mean squared residual of only the combinatorial part of Eq. 3.7 after the fit. The new model describes the selected γ^∞ and H^E data very well. The fit is better than any other model for the combinatorial contribution. Evaluation of only the combinatorial part of Eq. 3.7 shows a slightly smaller mean squared residual for the γ^∞ data compared to the complete model. This small difference confirms that the non-combinatorial contribution to γ^∞ for *n*-alkane mixtures is small, which is why these mixtures were chosen for the fit. Because of the good description of H^E including its temperature dependence, the small residual part of the γ^∞ data is also described well. As a consequence, it can be assumed that the new reference segment for the combinatorial contribution is determined significantly from the data.

Table 3.3: Mean squared residuals for the different combinatorial terms and for the new model.

model	$\bar{\Delta}_{\gamma^\infty}$ %	$\bar{\Delta}_{H^\mathrm{E}}$ J/mol
UNIQUAC [11]	15.22	-
Huyskens [16]	5.12	-
Kikic [13]	3.88	-
modified UNIFAC (Dortmund) [17]	2.52	-
modified UNIFAC (Lyngby) [18]	4.87	-
Eq. 3.7	2.26	3.59
only the combinatorial part of Eq. 3.7	2.15	-

Equation 3.7 and its combinatorial part, Eq. 3.1, are added to Fig. 2.1 and plotted in Fig. 3.3. Comparison of the full model with only the combinatorial part of Eq. 3.7 shows that the contribution by the residual part of Eq. 3.7 is small. Since the residual part of Eq. 3.7 is temperature dependent, the model was only plotted for 300 K, whereas the experimental data vary between 293.15 K and 453.15 K. Eq. 3.7 was fitted to experimental data of *n*-alkanes with up to 16 carbon atoms. The full Eq. 3.7 has therefore only been extrapolated for up to 25 carbon atoms in the solvent, since it cannot be expected that the empirically formulated residual part of Eq. 3.7 can be extensively extrapolated.

The other models only consider the combinatorial contribution. To allow a fair comparison, the combinatorial part of Eq. 3.7 was plotted over the whole range with the reference surface, as determined from the fit of Eq. 3.7. It is obvious that the combinatorial part of Eq. 3.7 describes the data well, although it was not fitted directly to the data and only data from mixtures with *n*-alkanes with up to 16 carbon atoms were considered for the fit. It should be stressed that the characteristic general slope as a function of carbon number is better described by the new model compared to the other models. While, compared to the other models, the experimental data typically lie above the models for low carbon numbers and tend to lie below the models for higher carbon number – at least if the data for low carbon number are better described – the overall behavior is depicted better by the combinatorial part of Eq. 3.7 with the new reference segment.

The difference between the combinatorial part of Eq. 3.7 and the Flory-Huggins model is

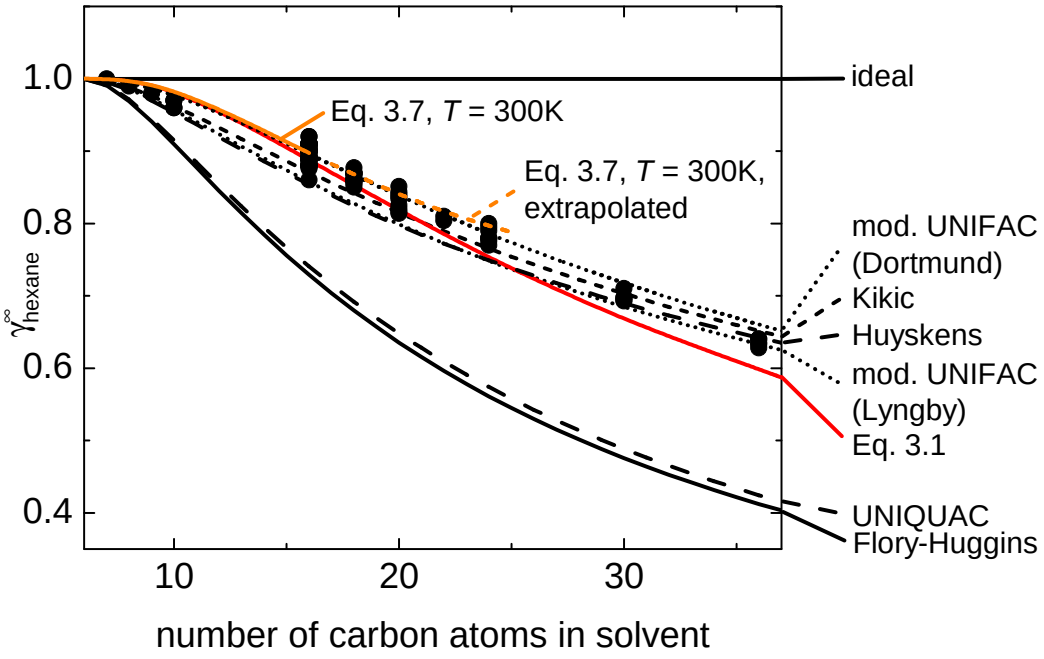


Figure 3.3: Comparison of the different combinatorial terms and Eq. 3.7 with experimental data of the activity coefficient of *n*-hexane at infinite dilution in other *n*-alkanes. All experimental data were taken from the DECHEMA data series [15].

due to the second term of Eq. 3.1. This term of the Staverman-Guggenheim equation is often considered a higher order correction term that is usually assumed to be small. However, it is shown here that this term is not negligible if applied properly. The result shows that a meaningfully determined reference segment for the UNIQUAC combinatorial term, while maintaining the physically founded form of the equation, leads to a result which is at least as good as that obtained using the purely empirical modifications of the original models.

Table 3.4 shows the parameters of Eq. 3.7 that resulted from the fit. The new reference surface area is 1.932×10^4 m^2/mol, corresponding to a radius of 0.05053 nm. These values are significantly smaller than those of the reference segment originally defined for UNIQUAC by Abrams and Prausnitz, which was characterized by a radius of 0.1818 nm and a surface area of 2.5×10^5 m^2/mol.

Table 3.4: Resulting values for the parameters of Eq. 3.7.

parameter	value
a	-303.1 J/mol
b	-0.3413 J/mol
c	9.355×10^4 JK/mol
A^*	1.932×10^4 m^2/mol

Abrams and Prausnitz defined the size of the standard segment according to the idea that the

reference segment should correspond to a real chemical unit. However, the model is based on a lattice picture of a fluid. This means that it is assumed that all molecules can be divided into standard segments of equal size. It is also assumed that molecules can only occupy discrete positions in space that are determined by the structure of the lattice and the size of the standard segment. An additional assumption is that all molecules can be arranged on the lattice in such a way that no lattice site remains empty. These assumptions obviously differ significantly from reality. Molecules can have sizes that do not differ by exactly one standard segment and space can be regarded as being continuous. Molecules can therefore in principle occupy an infinite number of positions in space. Thus, a discrepancy between the model and reality must be expected at some point. The result shows that the model can describe reality with a much smaller standard segment than that proposed by Abrams and Prausnitz. The new standard segment thus no longer corresponds to a real chemical unit.

It has to be pointed out that the smaller standard segment does not contradict the assumptions in the Staverman model. As was discussed in section 2.1.2, Staverman derived his model without an equation that relates q to r. Therefore, in Staverman's model there is no limitation to the number of internal contact sites a segment can have. Segments with no external contact segments may occur, which makes possible to build up molecules with a large number of small segments.

According to Bondi, the van der Waals radius of a hydrogen atom equals 0.120 nm [70]. The new standard segment is thus a little smaller than half the size of a hydrogen atom. Since the combinatorial term describes size and shape effects, a new physical interpretation of the standard segment is that its size can be regarded as the order of the geometrical dimension at which molecules "feel" their geometric details. With this new interpretation, the size of the new standard segment appears to be reasonable, since on the surface of a molecule at the intersection between two covalently bonded atoms, radii smaller than the van der Waals radii of atoms can occur.

4 The MOQUAC model

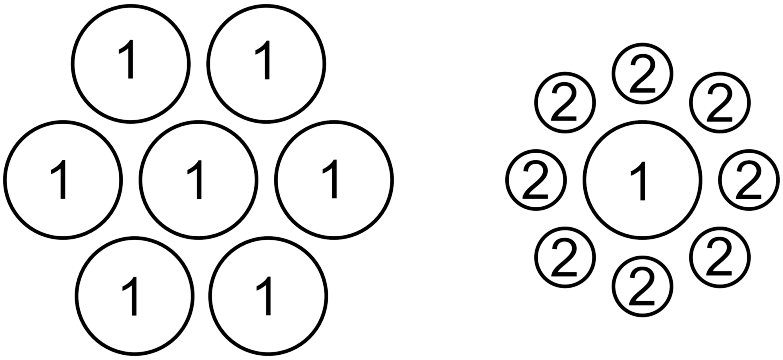
In this chapter, a new model for the excess Gibbs energy of liquid mixtures of uncharged components is derived. The model is based on molecular orientations and considers molecular contacts explicitly. Since for the derivation of the model the quasi-chemical theory is applied, the new model is called MOQUAC, which stands for "Molecular Orientation based QUAsi-Chemical theory". MOQUAC can account for the full three-dimensional molecular structure because, contrary to state-of-the-art G^{E}-models, with MOQUAC no assumptions are necessary that lead to the loss of all information on molecular geometry. Therefore, this model can describe effects like multiple contact points, which is important for the description of components with several strongly interacting functional groups, such as bio-based components.

4.1 Derivation of MOQUAC

Liquids are densely-packed systems of molecules and the behavior of such systems is generally very complex. A model is a simplified abstract view of this complex reality and serves to describe the behavior of liquids approximately. In order to produce the simplifications in a model, assumptions need to be made. One widely accepted assumption is that liquids can be treated as if their molecules were arranged on a lattice [19]. Here the structure of liquids is considered to be quasi-crystalline. Whereas in a crystal each molecule is surrounded by a definite invariable number of nearest neighbors, this number is not definite in a liquid. Nevertheless, it can be assumed that the number of nearest neighbors has a fairly well-defined average value. For MOQUAC it is assumed that component i has z_i nearest neighbors, where z_i generally depends on the local composition, which again depends on the size and shape of the components, their global concentration, the temperature and the interaction energies in the system. For example, a large molecule will generally have more nearest neighbors at infinite dilution in a solvent with small molecules than it will have in a pure state, as is illustrated in Fig. 4.1. However, here it is assumed that z_i is constant and independent of size, shape, composition, temperature and interaction energies. For most investigations in this work, the z_i were set to 10.

A molecule generally interacts with all other molecules in its vicinity. The interaction potential of uncharged molecules is usually of short range so the interaction energy with a direct neighbor is significant for the behavior of a system, but the interaction energy soon loses significance as the distance to the neighboring molecule increases. For MOQUAC it is thus assumed that only the interactions of a molecule with its z_i nearest neighbors need to be considered. The nearest neighbors form the first solvation shell of a molecule. If charged molecules, i.e. ions, had to be modeled, interactions beyond the first solvation shell would also have to be considered.

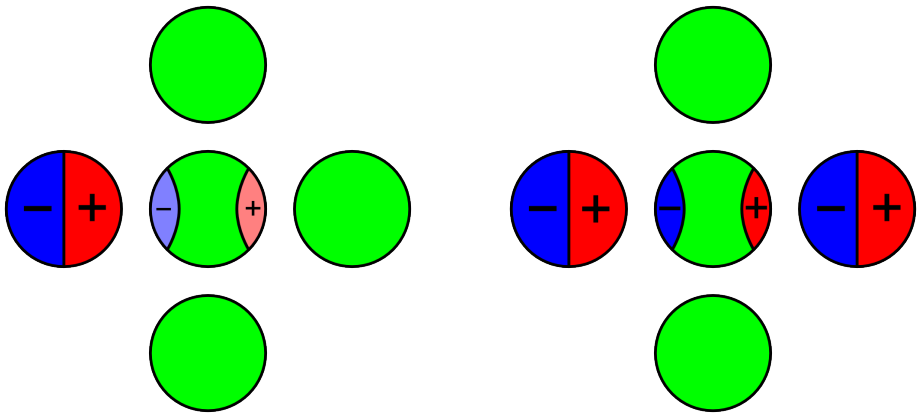
In addition, for MOQUAC it is assumed that an interaction between two directly neighboring molecules is unaffected by other interactions in the system. Nonetheless, this is generally not true since a molecule can, for example, become more polarized by the interaction with one of its neighboring molecules, which causes the interaction energy with its other neighboring molecules to change. This is illustrated in Fig. 4.2. The central molecule in Fig. 4.2a is a neutral molecule



(a) pure component 1 (b) infinite dilution of 1 in 2

Figure 4.1: Number of nearest neighbors of a central molecule in a pure state and at infinite dilution for molecules differing in size.

that is surrounded by three neutral molecules and a molecule with a permanent dipole. Areas of a molecule with no charge are green and areas with a positive or a negative charge are red or blue respectively. In the configuration that is shown in Fig. 4.2a, the molecule with the permanent dipole induces a small dipole on the central molecule. This is indicated by the small charged areas on the central molecule. In Fig. 4.2b the neutral molecule on the right is replaced by a molecule with a permanent dipole. Both molecules with a permanent dipole in this configuration induce a stronger dipole on the central molecule than is the case in Fig. 4.2a. Because of the stronger induced dipole, the interaction energy of the central molecule with its left neighbor is affected. Thus, the interactions do interfere with one another here, but this effect is neglected for MOQUAC.



(a) weak induced dipole (b) strong induced dipole

Figure 4.2: Interference of interactions.

The example in Fig. 4.2b also makes clear that the interaction energy between two directly neighboring molecules generally depends on the orientation of both molecules. If the right molecule with the permanent dipole in Fig. 4.2b is rotated by 180°, then the interaction en-

ergy of this molecule with the central molecule is affected. Molecular orientation is considered in MOQUAC. Thus, no averaging of the interaction energy over the orientations of a molecule is performed, in contrast to models like NRTL, Wilson or UNIQUAC. Only by considering the orientation can the three-dimensional molecular structure be accounted for and effects like multiple contact points be considered.

In reality, a molecule can assume an infinite number of different orientations. For MOQUAC this infinite number of orientations is discretized to a finite number of orientations N_i^{or} per component i. Here, an orientation k is defined by a direction of orientation and an angle of rotation around this direction of orientation. As an example for the definition of orientations, in Fig. 4.3 an idealized molecule with the shape of a cube is shown. The six faces of the cube are all subdivided into four equal-sized surface segments and all segments are numbered. The infinite number of different orientations of this cubic molecule can be discretized to, for example, 24 orientations, where for each orientation when looking at a face of the cube the quadrant indicated by the arrow is occupied by one of the 24 surface segments. In this case there are six directions of orientation, one for each face of the cube facing the front, and around each direction of orientation four rotations can occur. In Fig. 4.4 the first two orientations of the cube are shown.

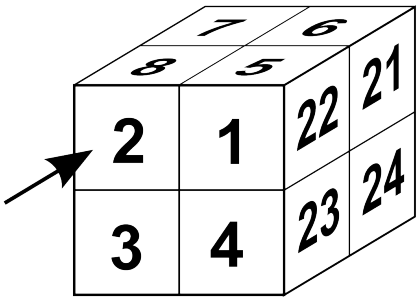
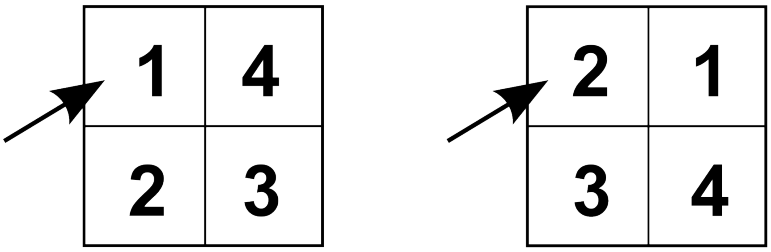


Figure 4.3: Cubic molecule with each face subdivided into 4 equal-sized surface segments.



(a) orientation 1 (b) orientation 2

Figure 4.4: Orientation 1 and 2 of the cubic molecule of Fig. 4.3.

In a system of molecules, each orientation i,k occurs with a certain probability. This probability generally depends on the interactions in the system, since certain orientations allow for energetic more favorable contacts than other orientations. $p_{i,k}^{\mathrm{R}}$ is the frequency of occurrence of orientation k of a molecule of type i in a system at infinite temperature, i.e. when interaction energies are irrelevant. In the case of the cubic molecule in the example above, the random frequency of occurrence of each orientation equals $1/24$. The orientations for the molecule in Fig.

4.3 can also be defined in a different way. For example, the first two orientations can be united to one single orientation. The number of different orientations is then reduced to 23 instead of 24 and the random frequency of occurrence of this larger orientation would be twice as high as the random frequency of occurrence of the other 22 orientations. In general, the random frequency of occurrence of an orientation i,k thus depends on the definition of the orientations that are distinguished for a molecule of type i. For MOQUAC it is assumed that the frequency of occurrence of orientation k of a molecule of type i does not change as a result of the interactions in a system, i.e. at a finite temperature. Additionally, the frequencies of occurrence of orientation are normalized, such that

$$\sum_{k=1}^{N_i^{\mathrm{or}}} p_{i,k}^{\mathrm{R}} = 1 \tag{4.1}$$

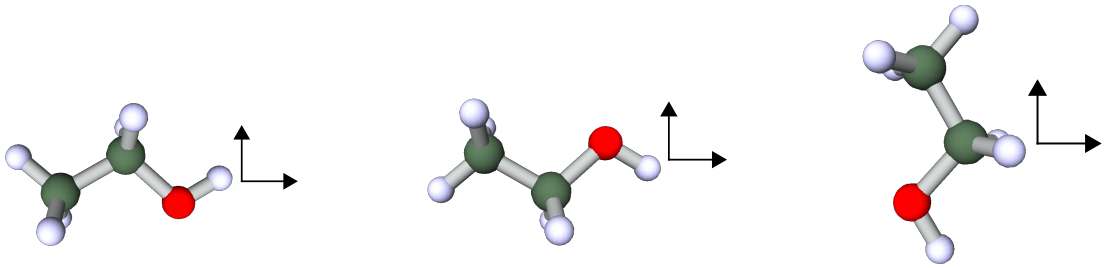
applies.

To characterize a contact between two neighboring molecules as a function of their orientation, contact types are defined in MOQUAC. To realize this for each contact, an individual reference system is introduced and the orientation of both molecules relative to this reference system is determined. The reference system itself is determined by the connection line between the centers of mass of both molecules. The direction of orientation of a molecule coincides with this connection line and points towards the neighboring molecule. A direction perpendicular to the connection line between the centers of mass is also required to evaluate the rotational orientation. The choice of this direction is arbitrary but should be constant for the contact. Although a molecule has only one global orientation in space, since for every contact an individual reference system is introduced, one orientation per contact thus z_i relative orientations are distinguished for each molecule of component i. The frequency of occurrence $p_{i,k}^{\mathrm{R}}$ of an orientation i,k is independent of the choice of the reference system and is therefore independent of the contact under consideration. By this specification it is not the absolute orientation of each molecule in space that is characterized but rather the relative orientation between two molecules in contact with each other.

A molecular contact is considered completely characterized by the type i and j of both molecules involved in the contact and their orientation k and l relative to the reference system of the contact. A contact of type j,l,i,k is a contact that originates from the orientation k of a central molecule of type i with the orientation l of a surrounding molecule of type j. Of course, this contact is identical to a contact of type i,k,j,l. The only difference is the molecule that is chosen as reference for the contact.

For example, in Fig. 4.5, three different orientations of an ethanol molecule are given. The arrows next to each molecule in Fig. 4.5 indicate the reference system. The horizontal arrows show the direction of orientation and the vertical arrows allow assignment of a rotational orientation to each molecule. Orientation 2 differs from orientation 1 by rotation and orientation 3 shows a different direction. In Fig. 4.6, two contacts between ethanol molecules are shown. Between both molecules of each pair of contacting molecules, a part of the connection line through the centers of mass is shown. In addition, for each contact, a direction perpendicular to the connection line is given. Together with the connection line this defines the reference system. The direction of orientation of a molecule should coincide with the connection line through the centers of mass of both contacting molecules and point towards the other molecule. The directions of orientation are also shown in the figure. With the definition of both reference systems, the relative orientation of the central molecule equals orientation 1 of Fig. 4.5 when considering the horizontal contact and equals orientation 3 of Fig. 4.5 when considering the vertical contact. The

orientation of the molecule on the right equals orientation 2 and the orientation of the molecule on the top equals orientation 3 of Fig. 4.5. Thus a 2-1 and a 3-3 contact are shown.



(a) orientation 1 (b) orientation 2 (c) orientation 3

Figure 4.5: Example of three orientations of an ethanol molecule.

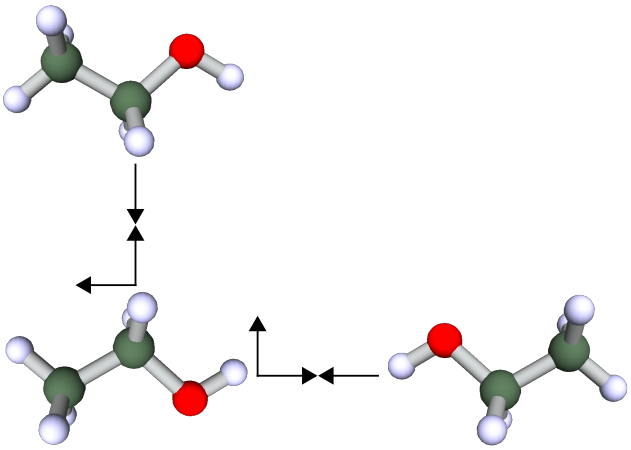


Figure 4.6: An example of a 2-1 and a 3-3 contact between ethanol molecules. The orientations are defined in Fig. 4.5.

The relative orientations k and l relative to the reference system of the contact and the type of molecules i and j defines the interaction energy of the two neighboring molecules $u_{j,l,i,k}$, which, generally speaking, also depends on other quantities such as the distance between both molecules and the temperature. For the derivation of the MOQUAC model in a first approximation it is assumed that the interaction energy for given relative orientations can be represented by one average value. Basically the same assumption is made for the original UNIFAC model [34], whereas in the UNIFAC modifications, the average interaction energy is considered temperature

dependent [17, 18]. In COSMO-RS, the misfit part of the interaction energy is considered temperature independent, whereas the hydrogen bonding part of the interaction energy is considered temperature dependent [46].

Because a contact of type j,l,i,k is equal to a contact of type i,k,j,l, the interaction energy is symmetrical for the MOQUAC model:

$$u_{j,l,i,k} = u_{i,k,j,l} \tag{4.2}$$

COSMO-RS also uses symmetric interaction energies, whereas for UNIFAC and its modifications, asymmetric interaction energies are assumed. For MOQUAC, the interaction energies between two molecules are parameters that for real components can be described by an additional model that describes the interaction energy, for example, as a general function of the charge densities of the interacting surfaces. Such a model is described in chapter 5.

To determine the energy of a liquid system of molecules, the frequency of occurrence of contacts of type j,l,i,k in the system needs to be known. For this, $\psi_{j,l,i,k}$ is defined as the fraction of molecular contacts originating from the orientation k of a molecule of type i with the orientation l of a molecule of type j. The $\psi_{j,l,i,k}$ are normalized such that for all $i \in [1,\ldots,N^{\mathrm{c}}]$ and $k \in [1,\ldots,N_i^{\mathrm{or}}]$

$$\sum_{j=1}^{N^{\mathrm{c}}} \sum_{l=1}^{N_j^{\mathrm{or}}} \psi_{j,l,i,k} = 1 \tag{4.3}$$

applies. If N is the number of molecules in the system, then Nx_iz_i is the number of contacts that originate from the molecules of component i. $Nx_iz_ip_{i,k}^{\mathrm{R}}$ is then the number of times the relative orientation k of component i can be identified in the system when considering all contacts. $Nx_iz_ip_{i,k}^{\mathrm{R}}\psi_{j,l,i,k}$ is the number of molecular contacts that originate from the orientation k of a molecule of type i with the orientation l of a molecule of type j. Of course, the number of contacts originating from the orientation k of a molecule of type i with the orientation l of a molecule of type j is equal to the number of contacts originating from the orientation l of a molecule of type j with the orientation k of a molecule of type i. Therefore, the symmetry conditions

$$z_ix_ip_{i,k}^{\mathrm{R}}\psi_{j,l,i,k} = z_jx_jp_{j,l}^{\mathrm{R}}\psi_{i,k,j,l} \tag{4.4}$$

can be defined for all $i,j \in [1,\ldots,N^{\mathrm{c}}]$, $k \in [1,\ldots,N_i^{\mathrm{or}}]$ and $l \in [1,\ldots,N_j^{\mathrm{or}}]$. This symmetry condition is equivalent to Eq. 2.25 of the lattice model by Guggenheim.

Since for the derivation of MOQUAC it is assumed that to describe a system of molecules only contacts between two directly neighboring molecules need to be considered, the energy of a liquid system U is calculated by summing the interaction energies of all contacts between directly neighboring molecules:

$$U = \sum_{i=1}^{N^{\mathrm{c}}} \sum_{k=1}^{N_i^{\mathrm{or}}} \sum_{j=1}^{N^{\mathrm{c}}} \sum_{l=1}^{N_j^{\mathrm{or}}} \frac{1}{2} z_ix_ip_{i,k}^{\mathrm{R}}\psi_{j,l,i,k}u_{j,l,i,k} \tag{4.5}$$

The factor $1/2$ in Eq. 4.5 follows from the fact that Eqs. 4.2 and 4.4 apply and that due to the four sums each contact is actually counted twice.

The goal of the derivation here is not only to describe the energy of a liquid system but describe its excess Gibbs energy G^{E}. G^{E} is defined by:

$$G^{\mathrm{E}} = G - RT\sum_{i=1}^{N_{\mathrm{c}}} x_i \ln x_i - \sum_{i=1}^{N_{\mathrm{c}}} x_iG_i^0 \tag{4.6}$$

The second term on the right hand side of Eq. 4.6 is the ideal mixture contribution to the Gibbs energy. G_i^0 is the Gibbs energy of pure component i. Thus, with an expression for the Gibbs energy G that can be applied to pure components and to mixtures, the excess Gibbs energy can be determined with Eq. 4.6. For the Gibbs energy, a contribution by the enthalpy H and the entropy S can be distinguished:

$$G = H - TS \tag{4.7}$$

A common assumption for the derivation of G^{E}-models is that liquid systems are considered to be independent of pressure. With this assumption, the enthalpy H of a liquid system equals its energy U, as was specified in Eq. 2.28. With Eq. 2.28

$$H = \sum_{i=1}^{N^{\mathrm{c}}} \sum_{k=1}^{N_i^{\mathrm{or}}} \sum_{j=1}^{N^{\mathrm{c}}} \sum_{l=1}^{N_j^{\mathrm{or}}} \frac{1}{2} z_i x_i p_{i,k}^{\mathrm{R}} \psi_{j,l,i,k} u_{j,l,i,k} \tag{4.8}$$

results from Eq. 4.5. Since the enthalpy and energy are equal, the Gibbs energy G of a liquid system is also equal to its free energy A:

$$G = A \tag{4.9}$$

The free energy can be determined from the canonical partition function Q with

$$nA = -k_{\mathrm{B}} T \ln Q \tag{4.10}$$

where n is the amount of substance of the liquid system in mole. The canonical partition function is arrived at by summing the Boltzmann factor of every microstate ξ of the liquid system:

$$Q = \sum_{\xi} \exp\left(-\frac{nU_\xi}{k_{\mathrm{B}} T}\right) \tag{4.11}$$

U_ξ is the energy of the system at the microstate ξ. Because of the lattice view of a liquid, a microstate is determined by the arrangement of all molecules on the lattice. Every configuration of molecules on the lattice results in a distribution of the different contact types. This distribution can be characterized by the vector $\vec{\psi}$ that contains the contact fraction of each contact type j,l,i,k, i.e. $\psi_{j,l,i,k}$.

According to Eq. 4.5, U only depends on $\vec{\psi}$. Two different microstates that have identical $\vec{\psi}$ thus have identical U. Microstates with identical $\vec{\psi}$ can be summarized and instead of summing over all microstates, Q can be determined by summing over all different $\vec{\psi}$:

$$Q = \sum_{\vec{\psi}} g(\vec{\psi}) \exp\left(-\frac{nU(\vec{\psi})}{k_{\mathrm{B}} T}\right) \tag{4.12}$$

$g(\vec{\psi})$ is the degeneracy of $U(\vec{\psi})$, i.e. the number of microstates with the same $\vec{\psi}$. Due to the fact that energy and enthalpy are equal, Eq. 4.12 can also be written as

$$Q = \sum_{\vec{\psi}} g(\vec{\psi}) \exp\left(-\frac{nH(\vec{\psi})}{k_{\mathrm{B}} T}\right) \tag{4.13}$$

where $H(\vec{\psi})$ is determined with Eq. 4.8.

The sum of Eq. 4.13 has a very large number of summands. In statistical thermodynamics it is common practice to approximate such a sum with its leading summand [71] if the number of possible states is large. This applies here, so $\vec{\psi}^*$ has to be determined where the summand of Eq. 4.13 is maximal. With this assumption

$$Q = g(\vec{\psi}^*) \exp\left(-\frac{nH(\vec{\psi}^*)}{k_B T}\right) \tag{4.14}$$

can be written for Eq. 4.13. Eqs. 4.9 and 4.10 are applied, with

$$G = H(\vec{\psi}^*) - \frac{k_B}{n} T \ln\left(g(\vec{\psi}^*)\right) \tag{4.15}$$

as the result. A comparison of Eq. 4.15 and Eq. 4.7 shows that S is calculated by:

$$S = \frac{k_B}{n} \ln\left(g(\vec{\psi}^*)\right) \tag{4.16}$$

Now $g(\vec{\psi})$ can be described by a similar approach to Guggenheim's [19], where the number of different ways to allocate all contacts on the lattice is considered. The result of this approach is

$$S = S_{comb} + R \sum_{i=1}^{N^c} \sum_{k=1}^{N_i^{or}} \sum_{j=1}^{N^c} \sum_{l=1}^{N_j^{or}} \frac{1}{2} x_i z_i p_{i,k}^R \psi_{j,l,i,k} \ln\left(\frac{x_j z_j p_{j,l}^R}{\psi_{j,l,i,k} \sum_{m=1}^{N^c} x_m z_m}\right) \tag{4.17}$$

For S_{comb}, different models can be used. In this work, Staverman's equation, Eq. 2.9, is used with the new reference segment of chapter 3 to determine the structural parameters r and q. The coordination number z of the combinatorial term in this work was always set to 10. A detailed derivation of Eq. 4.17 is given in appendix 7.3.

The residual part of the entropy S_{res} is defined as

$$S_{res} = S - S_{comb} \tag{4.18}$$

$\Leftrightarrow$

$$S_{res} = R \sum_{i=1}^{N^c} \sum_{k=1}^{N_i^{or}} \sum_{j=1}^{N^c} \sum_{l=1}^{N_j^{or}} \frac{1}{2} x_i z_i p_{i,k}^R \psi_{j,l,i,k} \ln\left(\frac{x_j z_j p_{j,l}^R}{\psi_{j,l,i,k} \sum_{m=1}^{N^c} x_m z_m}\right) \tag{4.19}$$

Appendix 7.5 shows that Eqs. 4.8 and 4.17 fulfill the Gibbs-Helmholtz equation

$$H = \left(\frac{\partial\left(\frac{H-TS}{T}\right)}{\partial\left(\frac{1}{T}\right)}\right)_{p,x_i} \tag{4.20}$$

and that thus the enthalpy and entropy terms of MOQUAC are consistent. Substituting Eqs. 4.8 and 4.17 into Eq. 4.7 results in

$$G = \sum_{i=1}^{N^c} \sum_{k=1}^{N_i^{or}} \sum_{j=1}^{N^c} \sum_{l=1}^{N_j^{or}} \frac{1}{2} z_i x_i p_{i,k}^R \psi_{j,l,i,k} u_{j,l,i,k} - T S_{comb} - TR \sum_{i=1}^{N^c} \sum_{k=1}^{N_i^{or}} \sum_{j=1}^{N^c} \sum_{l=1}^{N_j^{or}} \frac{1}{2} x_i z_i p_{i,k}^R \psi_{j,l,i,k} \ln\left(\frac{x_j z_j p_{j,l}^R}{\psi_{j,l,i,k} \sum_{m=1}^{N^c} x_m z_m}\right) \tag{4.21}$$

The condition for $\vec{\psi}^*$ that the summand of Eq. 4.13 is maximal is equivalent to the condition that Q of Eq. 4.14 is maximal and thus G of Eq. 4.15 or Eq. 4.21 respectively is minimal. A necessary condition for this minimum is that

$$\left(\frac{\partial G}{\partial \psi_{n,p,m,o}}\right)_{T,x_i} = 0 \tag{4.22}$$

applies. The ψ^* that fulfill Eq. 4.22 and simultaneously fulfill Eqs. 4.3 and 4.4 describe a representative system state which has an internal equilibrium of contacts for given macroscopic variables T and x_i. Under consideration of Eqs. 4.3 and 4.4, Eq. 4.22 results in:

$$\frac{\psi_{j,l,i,k}\psi_{i,k,j,l}}{\psi_{i,k,i,k}\psi_{j,l,j,l}} = \exp\left(-\frac{u_{j,l,i,k}+u_{i,k,j,l}-u_{j,l,j,l}-u_{i,k,i,k}}{RT}\right) \tag{4.23}$$

A detailed derivation of Eq. 4.23 is given in appendix 7.4. Since ψ is a measure for the concentration of a contact type, Eq. 4.23 is very similar to the law of mass action of chemical reactions. Because of this, the approach is referred to as quasi-chemical.

Eqs. 4.3, 4.4 and 4.23 form a set of non-linear equations that has to be solved to determine the representative contact fractions of the different contact types, $\psi^*_{j,l,i,k}$, in the liquid system. The structure of the set of non-linear equations is typical for the quasi-chemical theory and the same modified Newton-Rapson method by Larsen and Rasmussen [48] can be applied to solve the set of equations as was used for GEQUAC (see appendix 7.1). With $\vec{\psi}^*$ the Gibbs energy of the liquid system can be determined with Eq. 4.21. This method can be used to determine the Gibbs energy of mixtures as well as of pure components.

To arrive at the excess Gibbs energy, a combinatorial and a residual contribution to the Gibbs energy

$$G = G_{\text{comb}} + G_{\text{res}} \tag{4.24}$$

are distinguished. G_{comb} is defined as

$$G_{\text{comb}} = -TS_{\text{comb}} \tag{4.25}$$

such that for the residual contribution

$$\begin{aligned} G_{\text{res}} = &\sum_{i=1}^{N^{\text{c}}}\sum_{k=1}^{N_i^{\text{or}}}\sum_{j=1}^{N^{\text{c}}}\sum_{l=1}^{N_j^{\text{or}}}\frac{1}{2}z_i x_i p^{\text{R}}_{i,k}\psi_{j,l,i,k}u_{j,l,i,k} \\ &- TR\sum_{i=1}^{N^{\text{c}}}\sum_{k=1}^{N_i^{\text{or}}}\sum_{j=1}^{N^{\text{c}}}\sum_{l=1}^{N_j^{\text{or}}}\frac{1}{2}x_i z_i p^{\text{R}}_{i,k}\psi_{j,l,i,k}\ln\left(\frac{x_j z_j p^{\text{R}}_{j,l}}{\psi_{j,l,i,k}\sum_{m=1}^{N^{\text{c}}} x_m z_m}\right) \end{aligned} \tag{4.26}$$

results. The excess Gibbs energy is then determined by Eq. 2.1 with Eqs. 2.13 and 2.60.

The derivation of the new G^{E}-model MOQUAC is thus complete. Because of the consideration of different contact types, MOQUAC allows consideration of the effect of the three-dimensional molecular structure on the interaction energies. The interaction energies are, however, parameters for MOQUAC that have to be determined by an additional model. If these interaction energies can be described predictively then MOQUAC can also be applied predictively to describe the behavior of liquid systems. However, before an attempt is made to apply MOQUAC to mixtures of real components, MOQUAC will be verified by a comparison to simulation results of lattice systems.

4.2 Verification of MOQUAC by comparison to simulation results

A simple system fulfilling the assumptions made for deriving MOQUAC is a lattice system with molecules allowing for different interactions. If the model was derived correctly it should thus be able to describe simulation results of such a lattice system. In particular, MOQUAC should be able to describe systems whose behavior is determined by the three-dimensional molecular structure. To evaluate the performance of MOQUAC, results using MOQUAC were compared to (a) results from lattice simulations and (b) results from GEQUAC. GEQUAC serves as a benchmark since it is equal to the thermodynamic part of COSMO-RS, as was discussed in section 2.2.4.

The simulation results of the work by Pielen [25] were used for the evaluation of MOQUAC. Pielen considered a system consisting of cubes each with 6 faces that could differ in their interactions. These cubes were placed on a cubic lattice with 10 lattice positions in each spatial direction. In the simulations, periodic boundary conditions were applied and only contacts between direct neighbors were accounted for.

The faces of each cube are all subdivided into 4 segments of equal size. Figure 4.3 shows a three-dimensional image and Fig. 4.7 shows a flat projection of the molecular model Pielen applied for his simulations. In the lattice, the faces of two neighboring cubes always completely overlap, so no offset occurs. Thus, a surface segment always interacts with one other surface segment and cannot interact with more than one surface segment simultaneously.

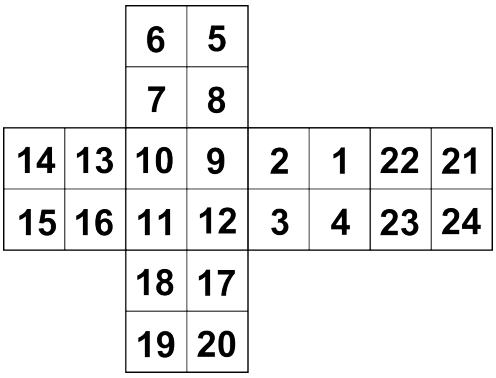


Figure 4.7: Flat projection of the general form of a molecule used in the Monte-Carlo simulation.

Pielen applied a Monte-Carlo algorithm for his simulations, where a Monte-Carlo step consisted of (a) interchanging two randomly chosen molecules (translation step) and (b) the 90° rotation of one randomly chosen molecule in one of six possible directions of rotation (rotation step). The direction of rotation was also randomly chosen. After both a translation step and a rotation step, the total energy U of the system was evaluated. If the new energy $U(2)$ was less than the energy before the step $U(1)$, the new state was accepted. If the new energy was higher, the Boltzmann factor

$$B = \exp\left(-\frac{U(2)-U(1)}{RT}\right) \tag{4.27}$$

was determined. The new state was accepted if the Boltzmann factor B was higher than a random number between 0 and 1. Otherwise the step was reversed.

At the beginning of a simulation, the molecules were distributed randomly on the lattice according to the given global composition. Then, 5×10^7 Monte-Carlo steps were conducted to equilibrate the system. After the equilibration phase, the system's energy was determined after each accepted step, in order to determine the average system energy. This evaluation phase also consisted of 5×10^7 Monte-Carlo steps. By conducting simulations at different temperatures, the temperature dependence of the energy of the system was determined. Since the lattice experiences no pressure dependence, the system's energy equals the system's enthalpy. Pielen determined the Gibbs energy from the enthalpy by integrating the Gibbs-Helmholtz equation [25].

Two different systems were simulated, one system with and one without coupled interactions. The system without coupled interactions allows a fair comparison with GEQUAC, because this model is not capable of describing coupled interactions, as discussed in chapter 2. By comparing MOQUAC's results with the simulation results for the system with coupled interactions, the improvement in the description of such systems by MOQUAC can be evaluated.

4.2.1 Simulation of a system without coupled interactions

The system without coupled interactions consists of two components. Component 1 has a positively charged face, a negatively charged face and four uncharged faces. A flat projection of this component is shown in Fig. 4.8a. The positively charged surface segments have the index 1, the negatively charged surface segments have the index 2 and the neutral surface segments the index 3 to 6. Component 2 is completely neutral. A flat projection of this component is shown in Fig. 4.8b. All surface segments of component 2 have the index 0.

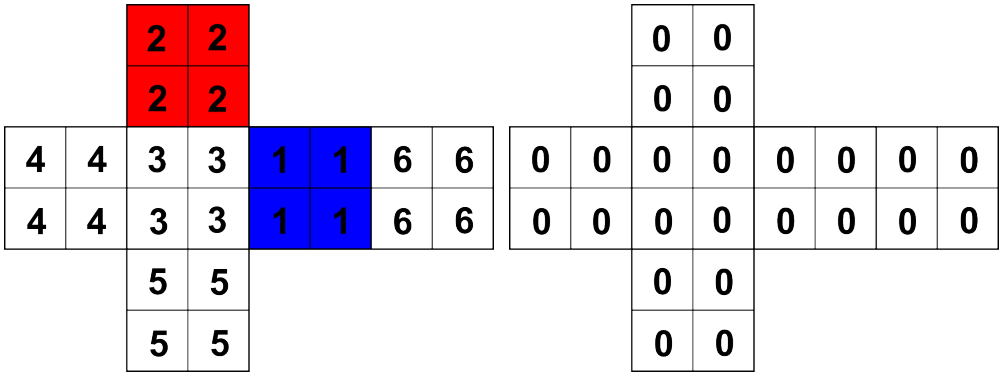


Figure 4.8: Flat projection of (a) component 1 and (b) component 2 in the system without coupled interactions. The blue sites are negatively charged, the red sites are positively charged and the colorless sites are neutral. Component 2 is also used in the system with coupled interactions.

Table 4.1 shows the interaction energies for the different surface-segment contacts used for the simulation. Equally charged surface segments experience a repulsive force and thus show positive interaction energy. Unequally charged surface segments experience an attractive force and thus show negative interaction energy. All interactions with a neutral surface segment show no interaction energy.

For MOQUAC it is assumed that 6 directions need to be considered for each component. The random probability of occurrence of each orientation $p^{\mathrm{R}}_{i,k}$ is $1/6$. $z_1 = z_2 = 6$, since each molecule

Table 4.1: Segment-segment interaction energy for both the system with and without coupled interactions.

$u_{12}/R = u_{21}/R$	-1200 K
$u_{11}/R = u_{22}/R$	1200 K
all other u_{ij}/R	0 K

has 6 nearest neighbors. For a molecular contact, four equally charged surface segments are always interacting simultaneously. The interaction energy divided by R is thus equal to -4800 K, 4800 K or 0 K. For GEQUAC the 4 segments of 1 face are united, leading to identical interaction energies. For GEQUAC z is also set equal to 6.

Figures 4.9 and 4.10 show a comparison of the excess properties between the model results and the results from the lattice simulation at 1600 K and 3200 K for the system without coupled interactions. MOQUAC and GEQUAC give identical results for this system and at both temperatures both models describe the data from the simulation very well.

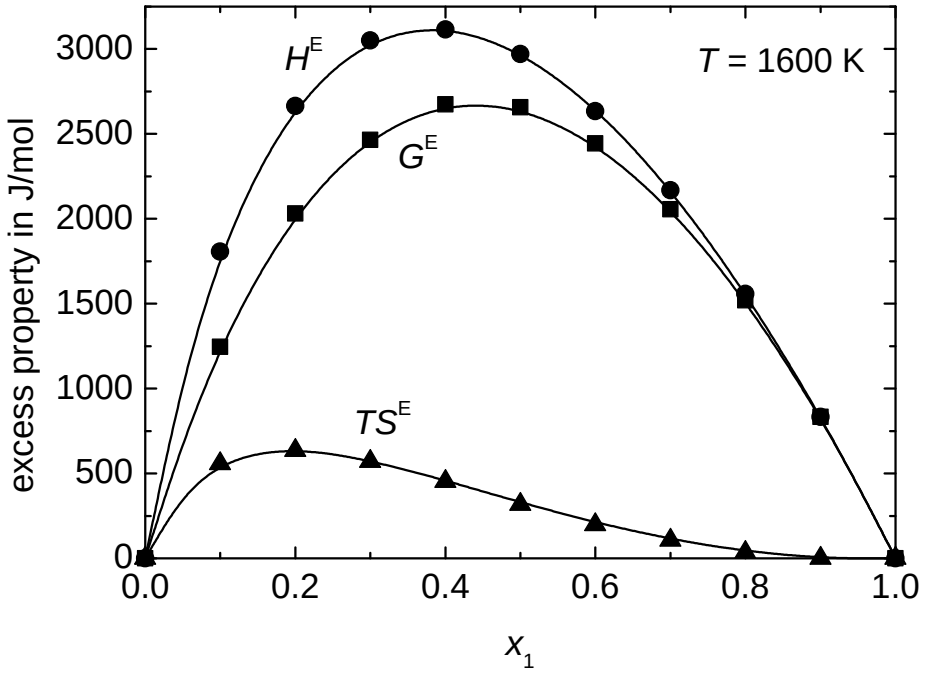


Figure 4.9: Comparison of results from simulation (symbols) [25] with results from MOQUAC and GEQUAC (lines) for the system without coupled interactions. The MOQUAC and GEQUAC result are identical. The GEQUAC results are identical to those obtained with the thermodynamic part of COSMO-RS.

The high temperatures for the simulation were chosen because u/RT needs to have an appropriate value, since this term determines the Boltzmann factor. To determine whether the choice of the interaction energies was appropriate, the resulting excess properties were evaluated. Since here the excess properties show realistic values, it was concluded that the choice of the interaction energies was adequate [25].

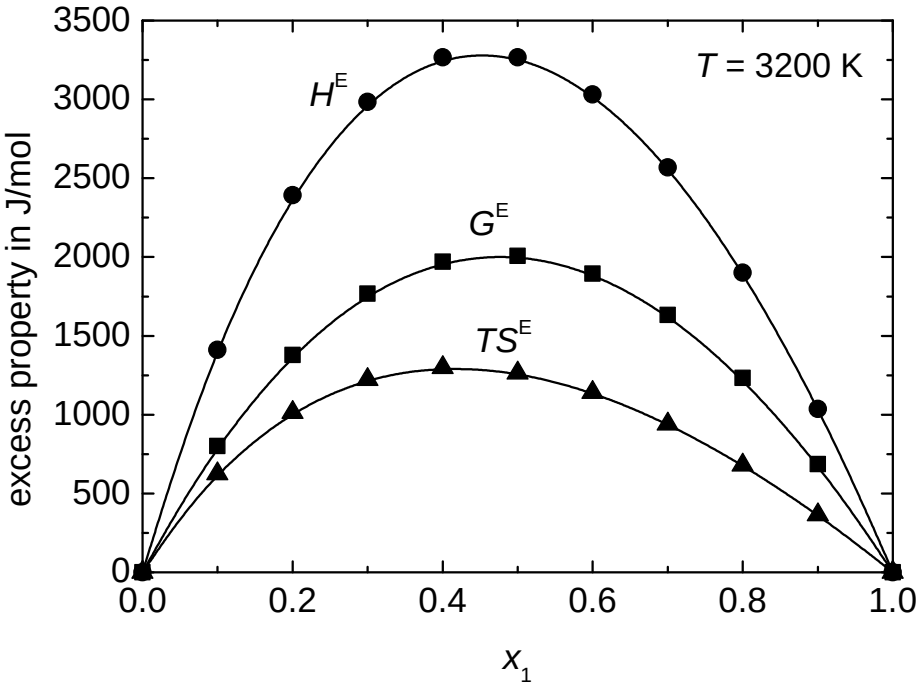


Figure 4.10: Comparison of results from simulation (symbols) [25] with results from MOQUAC and GEQUAC (lines) for the system without coupled interactions. The MOQUAC and GEQUAC result are identical. The GEQUAC results are identical to those obtained with the thermodynamic part of COSMO-RS.

4.2.2 Simulation of a system with coupled interactions

The system with coupled interactions also consists of two components. Component 1 has one positively charged and one negatively charged surface segment on the same face, as shown in Fig. 4.11. The positively charged site has the index 1 and is red, the negatively charged site has the index 2 and is blue. The remaining sites with the indices 3 to 24 are neutrally charged and are colorless. For two interacting molecules of component 1, a favorable configuration exists, where simultaneously two pairs of attracting surface-segment contacts exist. This configuration is shown in Fig. 4.12.

Component 2 is the same component as before again being completely neutral. Table 4.1 shows the interaction energies of the different surface-segment pairs.

For MOQUAC, 6 directions of orientation are distinguished. Per direction of orientation, a molecule can now have 4 different orientations, due to rotations. Thus, 24 orientations are considered for each component, $N_1^{\mathrm{or}} = N_2^{\mathrm{or}} = 24$. For example, orientation 1 of component 1 occurs when surface segment 1 is in the upper left position when looking at the face of the cube that contains surface segment 1. Orientation 2 of component 1 occurs when surface segment 2 is in the upper left position when looking at the face of the cube that contains surface segment 2, etc.. This was already illustrated in Fig. 4.3. The random probability of occurrence of each orientation, $p_{i,k}^{\mathrm{R}}$, is $1/24$. Again, $z_1 = z_2 = 6$, since each molecule has 6 nearest neighbors. Table 4.2 shows the interaction energies for the system with coupled interactions.

For GEQUAC, 24 surface segments are distinguished per component. The surface fraction

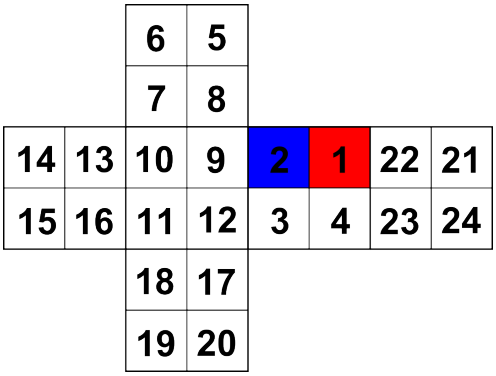


Figure 4.11: Flat projection of component 1 in the system with coupled interactions. The blue site (index 1) is negatively charged, the red site (index 2) is positively charged and all other sites are neutral.

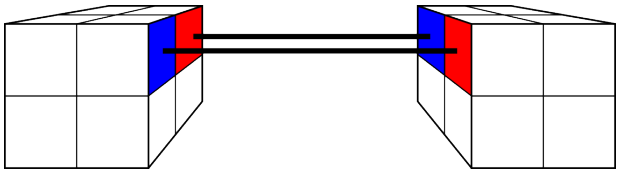


Figure 4.12: Most favorable configuration of component 1 in the system with coupled interactions.

of each surface segment is $1/24$. The coordination number is again set to 6 and the interaction energies for the segment-segment contacts are as given in Tab. 4.1.

Figures 4.13 and 4.14 show a comparison of the excess properties between the model results and the results from the lattice simulation at 1600 K and 1920 K. In addition to successfully describing the data from a simulation of a system without coupled interactions, MOQUAC is also able to describe the data from the simulation for the system with coupled interactions very well. GEQUAC and thus the thermodynamic part of COSMO-RS, on the other hand, give a very poor description of the simulation results. In contrast to MOQUAC, GEQUAC and the thermodynamic part of COSMO-RS are not capable of describing the coupled interaction of surface segments 1 and 2 of component 1. The coupled interaction in this case leads to a very attracting molecular contact. The high temperatures for the simulation were again chosen such that u/RT has an adequate value.

Table 4.2: MOQUAC contact-interaction energy for the system with coupled interactions.

$u_{1,1,1,1}/R$	-2400 K
$u_{1,1,1,4}/R = u_{1,4,1,1}/R$	1200 K
$u_{1,2,1,1}/R = u_{1,1,1,2}/R$	1200 K
$u_{1,2,1,3}/R = u_{1,3,1,2}/R$	1200 K
$u_{1,2,1,4}/R = u_{1,4,1,2}/R$	-2400 K
$u_{1,3,1,3}/R$	-2400 K
$u_{1,3,1,4}/R = u_{1,4,1,3}/R$	1200 K
all other $u_{j,l,i,k}/R$	0 K

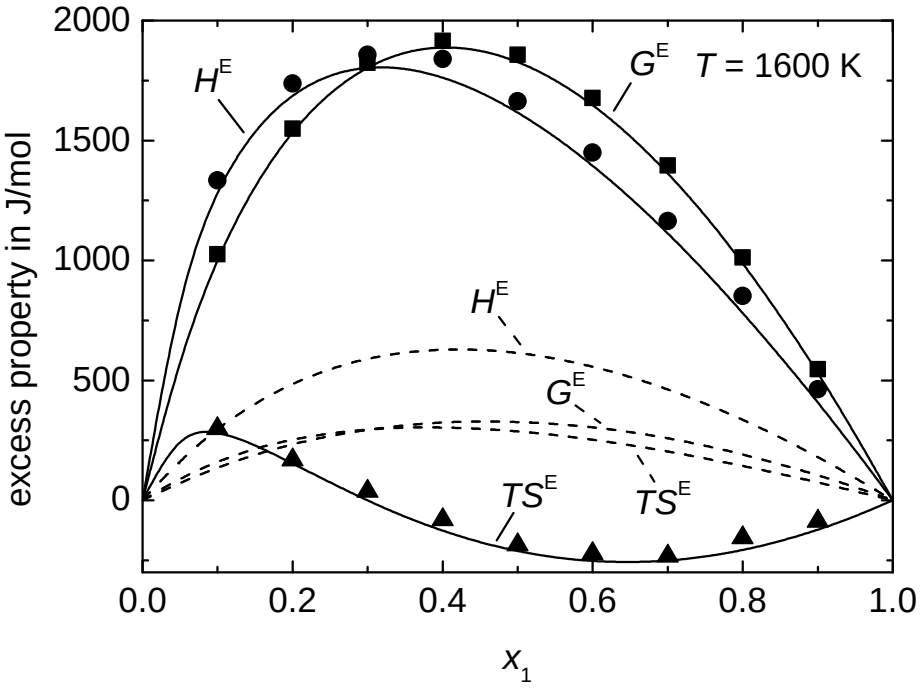


Figure 4.13: Comparison of results from simulation (symbols) [25] with results from MOQUAC (solid line) and GEQUAC (dashed line) for the system with coupled interactions. The GEQUAC results are identical to those obtained with the thermodynamic part of COSMO-RS.

By comparing MOQUAC with results from lattice simulations, it has thus been shown that MOQUAC can take the three-dimensional structure of molecules into account and that MOQUAC is therefore able to describe coupled interactions correctly. When there are no coupled interactions, MOQUAC yields identical results to GEQUAC and the thermodynamic part of COSMO-RS.

4.3 Accounting for conformers

For MOQUAC, isomers, especially stereoisomers, are easy to distinguish because of their different ways of interacting with other types of molecules. The three-dimensional structure of a molecule is accounted for by MOQUAC by the way a molecule interacts with other molecules. MOQUAC considers the structure of molecules to be rigid. While this is a good approximation for many simple chemical compounds, e.g. ethane, methanol, dimethylether, etc., most more complex molecules are flexible and thus have more than one relevant conformer, i.e. they have relevant metastable energy minima in addition to the total energy minimum. For molecules with a large number of rotatable bonds, the number of conformers can easily increase exponentially [46].

Pfennig [64] showed that the flexibility of a polymer chain can be modeled by distinguishing different polymer conformations. A similar approach is applied here to account for the flexibility of molecules. To do this, in MOQUAC a set of N_i^{conf} rigid conformers of a component i is considered. This set of conformers needs to be selected adequately to represent the most frequently

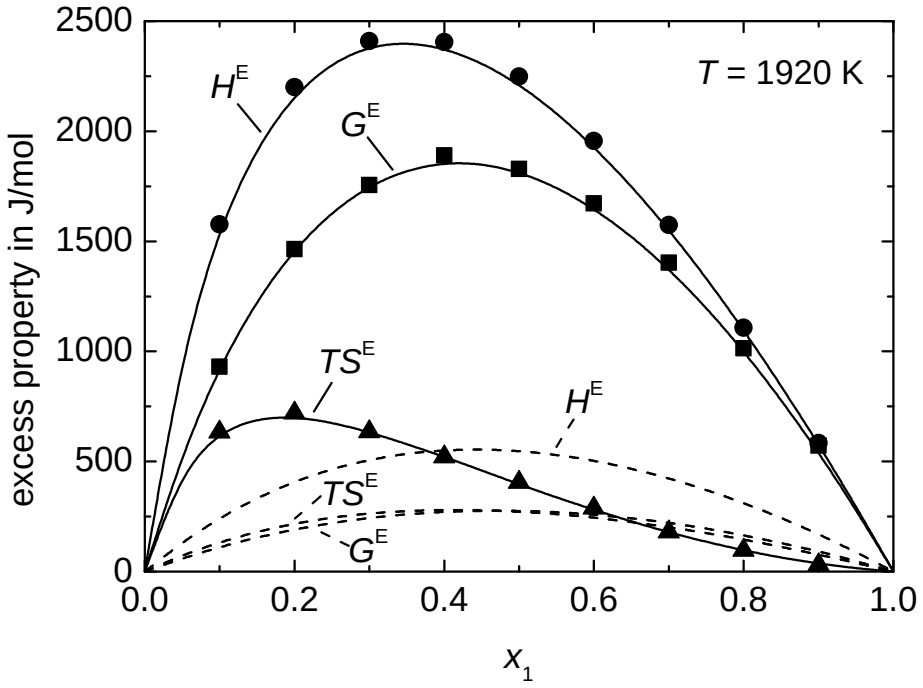


Figure 4.14: Comparison of results from simulation (symbols) [25] with results from MOQUAC (solid line) and GEQUAC (dashed line) for the system with coupled interactions. The GEQUAC results are identical to those obtained with the thermodynamic part of COSMO-RS.

occurring conformers of the component in the real system. Buggert et al. discuss different tools to identify conformers for the use with COSMO-RS [72]. Similar approaches can also be used to find the set of conformers for use with MOQUAC.

As in Pfennig [64], the probability of occurrence of a conformer is determined from the condition that all conformers of a component are in chemical equilibrium. Thus, for all r and $q \in [1,\ldots,N_i^{\mathrm{conf}}]$

$$\mu_{i,r} = \mu_{i,q} \tag{4.28}$$

applies, where $\mu_{i,r}$ is the chemical potential of conformer r of component i. This chemical potential is given by

$$\mu_{i,r} = \left(\frac{\partial nG}{\partial n_{i,r}}\right)_{T,n_{i,q\neq r},n_{j\neq i}} + \mu_{i,r}^{\mathrm{f}} \tag{4.29}$$

where $n_{i,r}$ is the amount of substance of conformer r of component i found in the mixture in equilibrium. G considers the contribution of the interactions and form and shape effects to the chemical potential and is given by Eq. 4.21. To find the chemical potential of a component, the contribution of the formation of the component $\mu_{i,r}^{\mathrm{f}}$ from some reference state also needs to be considered. The choice of this reference state is arbitrary, but the same reference state must be used for all types of molecules. In the contribution by formation no interactions between molecules should be considered, since these are already considered in the MOQUAC expression for G. For $\mu_{i,r}^{\mathrm{f}}$ an enthalpic and an entropic contribution are distinguished

$$\mu_{i,r}^{\mathrm{f}} = H_{i,r}^{\mathrm{f}} - TS_{i,r}^{\mathrm{f}} \tag{4.30}$$

where $H^{\mathrm{f}}_{i,r}$ is the enthalpy of formation and $S^{\mathrm{f}}_{i,r}$ the entropy of formation of conformer r of component i. The enthalpy and entropy of formation of each conformer can be determined from a quantum-chemical calculation, e.g. with the HyperChem program [73]. To a first approximation, $H^{\mathrm{f}}_{i,r}$ can also be set equal to the molecular energy without the dielectric energy, $E^{\mathrm{COSMO}}_{i,r}$, determined from a COSMO calculation

$$H^{\mathrm{f}}_{i,r} \approx E^{\mathrm{COSMO}}_{i,r} \tag{4.31}$$

and $S^{\mathrm{f}}_{i,r}$ can be set equal to

$$S^{\mathrm{f}}_{i,r} \approx R \ln m_{i,r} \tag{4.32}$$

where $m_{i,r}$ is the multiplicity of a molecule of conformer r of component i, i.e. the number of different ways to build the same structure [72]. These approximations are equivalent to those chosen for COSMO-RS [46].

In addition, for the chemical equilibrium between all conformers as described by Eq. 4.28, the sum of all mole fractions $x_{i,r}$ of the conformers of component i must equal the global mole fraction x_i of component i:

$$\sum_{r=1}^{N_i^{\mathrm{conf}}} x_{i,r} = x_i \tag{4.33}$$

4.4 Selection of molecular orientations

The quality of description of real systems using MOQUAC greatly depends on the appropriateness of the molecular orientations used. If, for example, a molecular orientation that can lead to a hydrogen bond is not accounted for, then the model will not describe the effect of this hydrogen bond on the behavior of the system. If, on the other hand, only very few orientations are accounted for and the orientation that can lead to a hydrogen bond is among them, the random frequency of occurrence $p^{\mathrm{R}}_{i,k}$ of this orientation will probably be assumed too high and as a result MOQUAC will determine an occurrence of the hydrogen bond too frequently.

All contact fractions $\psi_{j,l,i,k}$ are unknowns for the set of non-linear equations of MOQUAC. By applying the Newton-Rapson method which was modified by Larsen and Rasmussen [48] for this solution, the number of unknowns is reduced to one unknown per orientation. Thus, for a system of N^{c} different components, a total of

$$N^{\mathrm{or}} = \sum_{i=1}^{N^{\mathrm{c}}} N_i^{\mathrm{or}} \tag{4.34}$$

unknowns result. The set of equations has to be solved iteratively, as described in appendix 7.1, where in each step of the iteration, the inverse matrix of an N^{or}-by-N^{or} matrix needs to be determined. This is the most time-consuming step of the calculation and scales with a factor $N^{\mathrm{or}3}$. Thus, to keep the required computation time for solving the set of non-linear equations low, the number of orientations that are distinguished for a system should be kept low.

From these considerations it can be concluded that careful selection of the orientations that are considered for a system is of great importance. All relevant orientations that determine the behavior of a system need to be accounted for, but at the same time the number of different orientations that are considered should be kept as low as possible. It is also clear that the orientations that have to be considered per component depend on the system. Acetone, for example,

is able to form hydrogen bonds with chloroform, which strongly determines the behavior of this system. With *n*-heptane, on the other hand, acetone cannot form hydrogen bonds. For the system acetone + chloroform it is thus important to consider exactly those orientations of acetone that can lead to a hydrogen bond with chloroform. For the system acetone + *n*-heptane the choice of orientations for acetone will probably be less critical.

MOQUAC assumes independent molecular contacts. Because of this, two orientations of one molecule can be merged if they show identical interaction energies with all other orientations in the system. Proof of this, together with an example, is given in appendix 7.6. This feature is used for the definition of orientations of each component in a system.

The procedure is to first consider a fine grid of orientations per component and determine all interaction energies in the system based on that grid. Then, for each component it is evaluated if there are orientations that have similar energies with all other orientations in the system. Such orientations are then merged. In this way, the number of orientations per component is reduced, in order to reduce the computation time required for solving the MOQUAC equations. This procedure only needs to be done once per system. For the merging of two orientations, it is not necessary that they show absolutely identical interaction energies with all other orientations in the system. As a general rule, orientations which behave similarly can be merged.

An orientation of a component is given by a direction of orientation and a rotation. When N_i^{dir} directions and N_i^{rot} rotations are considered, then

$$N_i^{\mathrm{or}} = N_i^{\mathrm{dir}} N_i^{\mathrm{rot}} \tag{4.35}$$

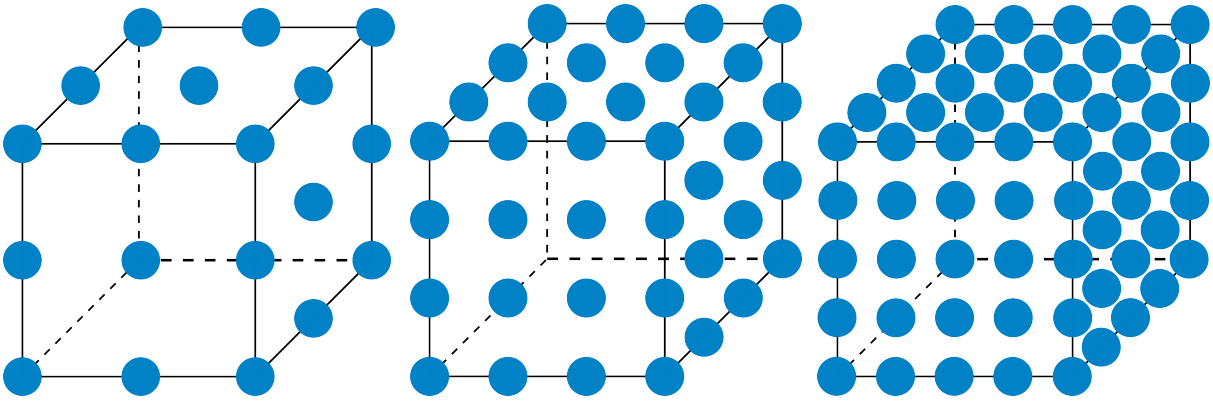
different orientations per component result. For example for the molecules in Fig. 4.7 that were used in the Monte-Carlo simulations in section 4.2, 24 orientations with 6 directions of orientation and 4 rotations per direction of orientation were considered.

Generally, in this work an attempt was made to define the directions of orientation for a molecule in such a way that a fairly homogeneous distribution of the directions of orientation results. With the definition that was applied, if the center of mass of the molecule were positioned in the center of a cube, lines originating from the center of mass and pointing in each direction of orientation would penetrate the surface of the cube at N_i^{dir} different sites. These sites are homogeneously distributed on the surface of the cube. Figure 4.15 shows these sites for 26, 56 and 98 directions of orientation. To achieve such a homogeneous distribution of the directions of orientation, only 4, 6, or $6s^2 - 12s + 8\,(\mathrm{for}\, s \geq 2)$ directions of orientation can be considered. In addition, N_i^{rot} equally distributed rotations are distinguished around each direction of orientation.

Starting from the fine distribution of orientations, it is determined which orientations can be merged for each component. For the merging or clustering of similar orientations to clusters, different clustering techniques can be used. Clustering is a method of unsupervised learning, and a common technique for statistical data analysis used in many fields [74]. For the clustering of similar orientations in this work an agglomerative algorithm was used, i.e. to begin with, each orientation is considered a cluster that consists of one element and progressively similar clusters are merged. The similarity $d_{A,B}$ of two clusters A and B was determined with

$$d_{A,B} = \min d_{(x),(y)} \qquad \text{for } \forall\, x \in A, y \in B \tag{4.36}$$

where x and y are elements of the cluster A and B respectively. The application of Eq. 4.36 results in single-linkage clustering [74]. This means that for every element in a cluster, at least one other element is less than or exactly a threshold value $d_{\max}$ away. There are also other criteria



(a) 26 directions (b) 56 directions (c) 98 directions

Figure 4.15: Cubes used for the definition of directions of orientation.

for similarity between two clusters that can be applied, but as a preliminary step in this paper only single-linkage clustering was considered.

To measure the distance between two elements $d_{(i,k),(i,m)}$, i.e. between two orientations i,k and i,m, different definitions can be applied. A general equation for such a distance measure is given by Bacher [74]. In the MOQUAC notation it reads

$$d_{(i,k),(i,m)} = \left(\sum_{j=1}^{N^{\mathrm{c}}} \sum_{l=1}^{N_j^{\mathrm{or}}} \left| u_{j,l,i,k} - u_{j,l,i,m} \right|^{v} \right)^{1/w} \tag{4.37}$$

where v and w are metric parameters. The parameter v weights the differences in individual variables. A larger value for v results in larger differences in few variables being more strongly weighted than small differences in many variables. The parameter w has the function of scaling the sum term of Eq. 4.37 back to the original scale. According to Bacher, as a general rule w is set to v [74].

Another definition for the distance between two orientations that was also used in this paper is

$$d_{(i,k),(i,m)} = \max \left| u_{j,l,i,k} - u_{j,l,i,m} \right| \qquad \text{for } \forall\, j \in [1,\ldots,N^{\mathrm{c}}], l \in [1,\ldots,N_j^{\mathrm{or}}] \tag{4.38}$$

In Eq. 4.38, the interaction energies of the orientations i,k and i,m with all other orientations in the system are compared and $d_{(i,k),(i,m)}$ is set to the largest deviation. Eq. 4.38 was mainly used in this work because it allows for a very fast algorithm. To show that the orientations i,k and i,m are not similar and can thus not be merged, it is sufficient to show that for at least one orientation in the system, the distance between the orientations i,k and i,m exceeds $d_{\max}$.

An example can help illustrate the algorithm of clustering as follows. Tab. 4.3 gives the interaction energies of a binary system with two orientations per component. Evaluating Eq. 4.38 results in $d_{(1,1),(1,2)} = 50$ and $d_{(2,1),(2,2)} = 125$. When the threshold value $d_{\max}$ is chosen to equal 50, then the two orientations of component 1 can be merged but the two orientations of component 2 cannot. Only when $d_{\max}$ is at least 125 can both the two orientations of component 1 and of component 2 be merged.

After determining which orientations can be merged into a cluster, the different clusters are considered the new orientations. The frequency of occurrence of a cluster is determined by summing the frequency of occurrence $p_{i,k}^{\mathrm{R}}$ of all orientations that were merged into the cluster.

Table 4.3: Interaction energies of a binary system with two orientations per component.

orientation	1,1	1,2	2,1	2,2
1,1	100	150	210	320
1,2	150	150	225	325
2,1	210	225	100	200
2,2	320	325	200	325

The interaction energies of the new orientations are determined by averaging over all relevant interaction energies. To calculate these averages each interaction energy should be weighted with the corresponding $p^{\mathrm{R}}_{i,k}$. However, since in this work the interaction energies and also the random probability of occurrence of each orientation are very similar, the arithmetic mean interaction energy was used throughout this work. Tab. 4.4 shows the averaged interaction energies for the example in Tab. 4.3 when $d_{\max}$ is set to 50 and the two orientations of component 1 are merged.

Table 4.4: Interaction energies of the binary system after merging the orientations of component 1 from Tab. 4.3.

orientation	1,1	2,1	2,2
1,1	137.5	217.5	322.5
2,1	217.5	100	200
2,2	322.5	200	325

The interaction energy is used for the evaluation of similarity between two orientations. Therefore, the model that is used to describe the interaction energy influences the clustering result. It is important that this interaction-energy model gives a smooth dependence of the interaction energy of the orientation. This is because it is only in this way that it can be guaranteed that for very similar orientations, similar interaction energies result. In the following chapter, a simple empirical interaction-energy model is presented. The clustering of orientations is investigated with this model.

5 Predictive description of orientation dependent interaction energies

For the application of MOQUAC, all interaction energies u need to be known. For a mixture that consists of N^{c} components, MOQUAC requires a total of

$$N^{u}_{\mathrm{tot}} = \sum_{i=1}^{N^{\mathrm{c}}} \sum_{j=1}^{N^{\mathrm{c}}} N^{\mathrm{or}}_{i} N^{\mathrm{or}}_{j} \tag{5.1}$$

different interaction energies. In Eq. 5.1, all different combinations of orientations of all components are considered. With the symmetry condition for the interaction energies, Eq. 4.2, the number of unknown interaction energies can be reduced to

$$N^{u} = \sum_{i=1}^{N^{\mathrm{c}}} \frac{1}{2} \left((N^{\mathrm{or}}_{i})^{2} + N^{\mathrm{or}}_{i} \right) + \sum_{i=1}^{N^{\mathrm{c}}} \sum_{j=i+1}^{N^{\mathrm{c}}} N^{\mathrm{or}}_{i} N^{\mathrm{or}}_{j} \tag{5.2}$$

Eq. 5.2 is thus the number of different interaction energies that need to be specified for applying MOQUAC. If all N^{or}_{i} equaled one, only one interaction energy per pair of components would be needed. In this case, MOQUAC would be simplified to Guggenheim's lattice-based model [19] described in section 2.2.1 and would have as many parameters as, for example, the semi-empirical Wilson, NRTL and UNIQUAC models. Since what makes MOQUAC advantageous is the fact that it enables consideration of the orientation dependence of the interaction energy, the number of parameters N^{u} for describing a real system will typically be significantly larger than the number of parameters of these semi-empirical models. If MOQUAC's model parameters had to be adjusted to experimental data, MOQUAC would be of no practical use. Therefore, the interaction energies must be determined with a separate model that contains only a small number of parameters. If this interaction-energy model can also be applied predictively, the combination of both the interaction-energy model and MOQUAC allows to truly predict G^{E}.

In the following, contributions to the interaction energy between two molecules are discussed. Some models that describe this interaction energy are also discussed and a simple empirical, predictive interaction-energy model is presented. This model is used to show the applicability of MOQUAC to real mixtures. The combination of MOQUAC with this simple interaction-energy model is compared to some existing predictive G^{E}-models.

5.1 Considerations about the molecular interaction energy

The interaction energy between two molecules is the result of various factors [4]. Molecules generally show a permanent charge distribution so if two molecules are in contact, there are Coulomb interactions between the charge distributions of both molecules. Molecules are also

generally polarizable, which means that their charge distribution is not fixed but can be influenced by the charge distribution of other molecules. This can lead to induced Coulomb interactions where induced charges interact with permanent charge distributions or with other induced charges.

Apart from Coulomb interactions, dispersion interactions also contribute to the interaction energy between two molecules. Dispersion forces are weak intermolecular forces that arise from the interactive forces between instantaneous charge distributions. Dispersion interactions are a result of the polarizability of molecules. Generally, the polarizability and thus the dispersion interactions increase with increasing molar mass. Since the dispersion interactions help to keep molecules together in a liquid state, e.g., the boiling point generally increases with increasing molar mass. This trend is exemplified by the halogens (from smallest to largest: F_2, Cl_2, Br_2, I_2). At room temperature fluorine and chlorine are gases, bromine is a liquid, and iodine is a solid. In contrast to dispersion forces, the Pauli exclusion principle always causes a repulsive force between two molecules. This force prevents a molecule from collapsing and also prevents two molecules from coming too close to one another.

Specific forces or chemical forces also contribute to the interaction energy between two molecules [4]. In contrast to the other, physical, contributions, chemical forces can be satisfied. One well known example of a chemical force is a hydrogen bond. For the formation of a hydrogen bond, a hydrogen atom that is covalently bonded to a highly electronegative atom is required as donor for the bond. Such a hydrogen atom always shows a high partial charge and can therefore cause strong Coulomb interactions. However, when it interacts with an adequate hydrogen-bond acceptor, a hydrogen bond is formed which generally shows a higher interaction energy than the Coulomb interaction alone. After the formation of a hydrogen bond, the hydrogen atom is saturated and will not form another hydrogen bond simultaneously.

There are several models that allow a predictive description of interaction energies. A common approach for molecular dynamics is, for example, to describe dispersion and repulsive forces with a Lennard-Jones potential and describe Coulomb interactions with partial charges on the molecule. Siepmann [75–83], for example, developed the united-atom approach, which is a method for a general description of interaction energies of different kinds of molecules by dividing the molecules into functional groups. Each functional group has a Lennard-Jones center and can additionally have a point charge. Because of the group approach, this approach can be applied predictively if all group parameters are determined. The parameters of each group have to be determined from a fit to experimental data.

For the predictive description of interaction energies for state-of-the-art G^{E}-models, two approaches can be distinguished. One approach that is applied in UNIFAC type models is, like the united-atom approach, based on group contributions and describes the interaction energy between two functional groups. But instead of introducing parameters for each group, for each combination of functional groups, parameters are introduced. For each combination of functional groups, the original UNIFAC model uses two parameters and in the case of the modified UNIFAC models by the Lyngby and Dortmund group, a total of up to six parameters can be used to describe the interaction energy. With their higher number of parameters, the modified UNIFAC models include a temperature dependence in the description of the group-interaction energy. To describe a broad range of different types of molecules, many different functional groups need to be defined and since each different combination of functional groups contributes to the total number of parameters, a rather large number of parameters are required for this approach to be universally applicable in predicting the behavior of systems. Since all parameters need to be determined from a fit to experimental data, a large data basis is required for this fit.

Another approach to the predictive description of interaction energies for G^E-models is based on surface-charge densities. In COSMO-RS, the interaction energy between two charged surface segments is determined. For the interaction energy between two surface segments, a misfit and a hydrogen-bond contribution are distinguished, as was discussed in section 2.2.4. The misfit contribution results from considering the energy required to remove a thin layer of ideal conductor from the system that screened the charge distribution of each molecule [57]. The hydrogen-bond contribution is an empirical term that accounts for the additional energy of a hydrogen bond to the Coulomb interaction. The advantage of the COSMO-RS approach is that its energy function is universal meaning the model can be applied to any type of surface segment and therefore molecule.

The relatively few parameters of the COSMO-RS energy function are determined from a fit to experimental data. Apart from these parameters, a quantum-chemical COSMO calculation of each component is also required in order to determine its surface-charge distribution. This COSMO calculation can be rather time consuming, but only needs to be performed once per component. The COSMO-RS approach is therefore a very efficient predictive approach for describing the interaction energy between two surface segments.

However, the UNIFAC approach and the COSMO-RS approach only describe the interaction energy between either functional groups or surface segments. Without additional assumptions, these approaches cannot be applied directly to determine the interaction energy between two molecules. To do this, a criterion needs to be introduced to determine which functional groups or which surface segments interact simultaneously upon a molecular contact.

The interaction energy between two molecules can be determined ab initio by a quantum-chemical calculation. Hellmann et al. [84], for example, describe the interaction energy between two methane molecules as a function of the relative orientation of both molecules. These calculations are, however, very time-consuming and this means that universal application of this method is limited to very small molecules for now.

In order to apply MOQUAC to real components, the decision was taken to develop an approach to describe the interaction energy between two molecules based upon surface-charge distributions from COSMO calculations. Contrary to group-contribution methods and similar to the COSMO-RS approach, this approach promises a predictive description of the interaction energies between two molecules for a broad range of different types of molecules and with only a small number of universal parameters.

5.2 Interaction-energy model

To show that MOQUAC can be applied predictively to mixtures of real components, a simple empirical model to describe the interaction energies between real components was derived. To enable the model to be used for many applications it uses, as COSMO-RS does, results from quantum chemical COSMO calculations [47]. COSMO calculations provide the surface-charge distribution and geometry of molecules, i.e. the position of each atom in the molecule. Since dispersion interactions are very similar in mixtures and pure components, they were disregarded for the description of G^E and related quantities. For the interaction-energy model, only Coulomb interactions are taken into account.

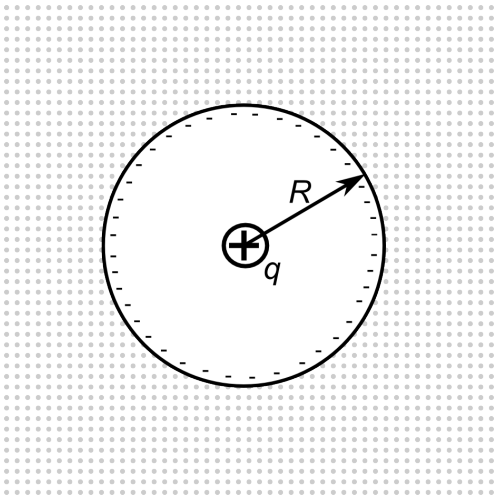
COSMO considers each molecule to be situated within a cavity inside an ideal conductor. The charge distribution of the molecule induces a charge distribution on the surface of the cavity, as described in section 2.2.4. To describe this charge distribution, the cavity surface is divided into

several surface segments. The charge distribution of the cavity surface is given by the charges of the surface segments. The fact that the conductor is ideal ensures that the charge distribution of the molecules is screened ideally. The result of a COSMO calculation gives no information on the charge distribution of the molecule itself. Only the charge distribution on the surface of the cavity is given.

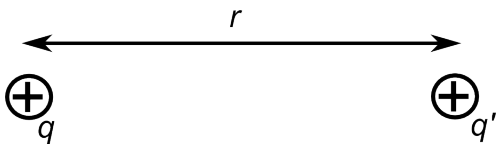
From the charge, the surface-charge density σ of a surface segment is determined through division by the surface area of the segment. However, the cavity surface is not always perfectly closed, so the surface areas and thus the charge densities of the surface segments can be erroneous. To eliminate such artifacts from the COSMO result, the surface-charge densities $\tilde{\sigma}$ as determined from a COSMO calculation are averaged with Eq. 2.73. Eq. 2.73 and the value for r_{av} of 0.050 nm were taken from Klamt [55]. In the following, an expression to describe the Coulomb interactions based on the averaged surface-charge densities σ is derived.

To describe the interaction energy between two molecules, it is assumed that the charge distribution of a molecule can be approximated by a set of point charges that are positioned at the center of mass of each atom. In Fig. 5.1a, one single point charge q is considered inside a spherically shaped cavity with the radius R in an ideal conductor. The total charge on the cavity surface due to the ideal screening equals $-q$ and this charge is distributed homogeneously if the point charge is positioned at the center of the cavity. The surface-charge density σ on the surface of the cavity thus equals

$$\sigma = \frac{-q}{4\pi R^2} \tag{5.3}$$



(a) A point charge q inside a cavity in an ideal conductor.



(b) System of two point charges q.

Figure 5.1: Model concept for the derivation of a model to describe the interaction energy based on surface-charge densities.

Now, a system of two point charges q and q' at a distance r is considered. Figure 5.1b shows this system. According to Coulomb's law, the interaction energy u of these two point charges

equals

$$u = \frac{qq'}{4\pi\varepsilon_0 r} \tag{5.4}$$

where ε_0 is the vacuum permittivity. To describe this interaction energy with the surface-charge densities, Eq. 5.4 can be rewritten as

$$u = \frac{-q}{4\pi R^2}\frac{-q'}{4\pi R'^2}\frac{16\pi^2 R^2 R'^2}{4\pi\varepsilon_0 r} \tag{5.5}$$

Eq. 5.3 is substituted into Eq. 5.5, which results in

$$u = \frac{4\pi R^2 R'^2}{\varepsilon_0 r}\sigma\sigma' \tag{5.6}$$

Introducing the variable c^{p}

$$c^{\mathrm{p}} = \frac{4\pi R^2 R'^2}{\varepsilon_0 r} \tag{5.7}$$

that only depends on some constants and on the distance between the two point charges as well as on the COSMO-radii R and R' allows Eq. 5.6 to be rewritten as

$$u = c^{\mathrm{p}}\sigma\sigma' \tag{5.8}$$

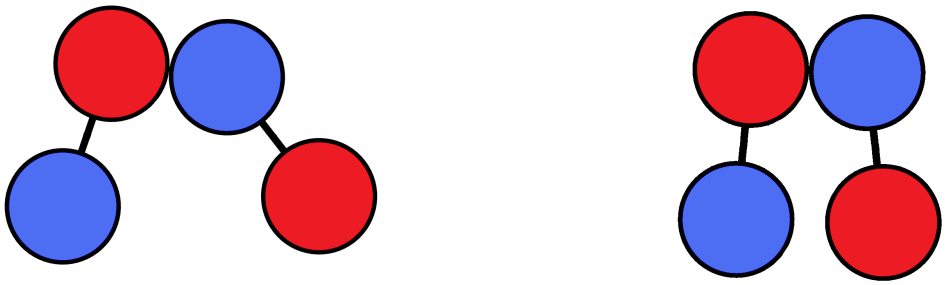
Fig. 5.2 shows two molecular contacts of two simple molecules. The charge distribution of each molecule in Fig. 5.2 can be represented by two point charges. The total Coulomb interaction energy of such a molecular contact is then determined by summing the Coulomb interaction energies of all pairs of point charges with the condition that each point charge of a pair belongs to a different molecule.

Eq. 5.4 gives the Coulomb interaction energy of a pair of point charges. It can be seen from this equation that the Coulomb interaction energy scales with $1/r$. Thus, the smaller the distance between a pair of point charges, the larger the Coulomb interaction energy. Generally, the closer two point charges are, the more they will contribute to the interaction energy of a molecular contact. For the interaction-energy model as an approximation for the determination of the interaction energy between two molecules, only those pairs of point charges are accounted for whose point charges are close enough to each other.

To introduce a criterion for two point charges to be close enough in order to be considered for the interaction energy, first the distance between two interacting molecules needs to be defined. For the interaction-energy model, it is arbitrarily assumed that when two molecules are in contact, they touch at their respective van der Waals surface. The van der Waals surface encloses the van der Waals volume of a molecule that cannot be occupied by other molecules. Here it is assumed that the molecules interact with their van der Waals surfaces defined by the molecular structure and the van der Waals radii as given by Bondi [12]. This is illustrated in Fig. 5.2, which shows that there is no free space assumed between two contacting molecules and that thus the free volume of the system is reduced to a minimum. Tab. 5.1 shows the van der Waals radius of the three atom types relevant for this work.

For two point charges to contribute to the interaction energy of a molecular contact, the distance between both point charges must be less than or equal to a threshold value $d_{i,j}^{\mathrm{ap}}$. The distance $d_{i,j}^{\mathrm{ap}}$ is defined as

$$d_{i,j}^{\mathrm{ap}} = f^{\mathrm{vdW}}(r_i^{\mathrm{vdW}} + r_j^{\mathrm{vdW}}) \tag{5.9}$$



(a) (b)

Figure 5.2: Molecular contact of two simple molecules.

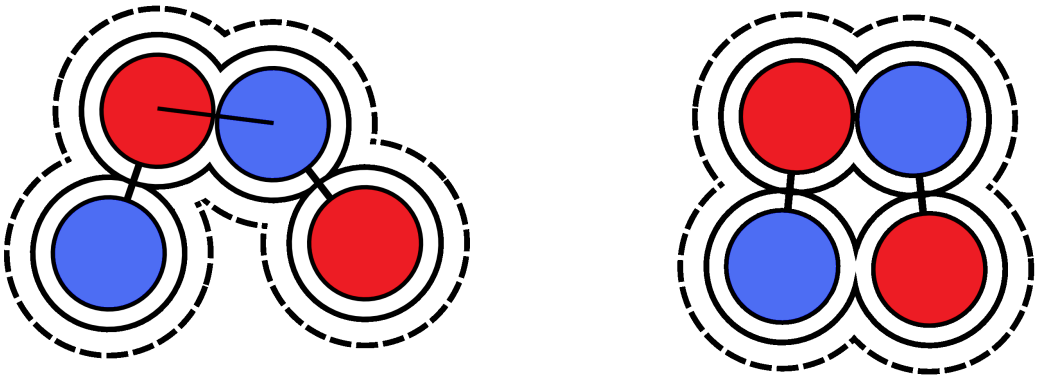
Table 5.1: Van der Waals radius according to Bondi [12] and COSMO radii.

atom type	van der Waals radius Å	COSMO radius Å
H	1.20	1.30
C	1.70	2.00
O	1.52	1.72

where r_i^{vdW} and r_j^{vdW} are the van der Waals radii of the atoms of type i and j respectively. f^{vdW} is the scaling factor of the van der Waals radius for the contact condition.

Fig. 5.3 shows the two molecular contacts together with the van der Waals radius that is scaled with a factor f^{vdW} (solid lines) and the COSMO radius (dashed lines). The COSMO radii are used for the construction of the cavity surface for which the result of a COSMO calculation gives the charge distribution. In Tab. 5.1, the COSMO radii of hydrogen, carbon and oxygen atoms are shown. According to the above definition for a contact between two point charges, only atoms for which the scaled van der Waals volumes overlap or touch have to be considered in order to calculate the interaction energy. Thus, in Fig. 5.3a there is one and in Fig. 5.3b there are two relevant pairs of atoms or point charges. To keep the number of parameters for the interaction-energy model as low as possible, only one scaling factor for the van der Waals radius for the contact condition is used. According to the picture of the interaction energy, this scaling factor could also be made a function of the strength of the point charge, because according to Eq. 5.4 the contribution of a pair of point charges does not only depend on the distance between the point charges but also on the charges of the point charges that are involved. The strength of a point charge could in a first approximation, e.g., be accounted for by distinguishing the type of atom it originates from.

Each pair of atoms that is relevant for the interaction energy of the molecular contact forms a contact point. There can be more than one contact point upon a molecular contact. $N^{\mathrm{cp}}_{j,l,i,k}$ is defined as the number of contact points of the molecular contact j,l,i,k. Since the condition for a contact point assures that the distance r between two point charges of all relevant contact points is similar and also the COSMO radii of the three atom types relevant for this work are not very different in size, it is assumed that c^{p} of Eq. 5.8 can be considered a constant model parameter.



(a) one relevant contact point

(b) two relevant contact points

Figure 5.3: Molecular contact of two simple molecules. The dashed circle indicates the COSMO radius and the additional solid circle the scaled van der Waals radius of an atom.

This is a rather rough simplification which yields a very simple interaction-energy model with a minimal number of adjustable parameters.

Eq. 5.8 describes the interaction energy between two point charges based on surface charge densities. The interaction energy of a molecular contact is then determined by summing the contribution of each contact point:

$$u_{j,l,i,k} = \sum_{m=1}^{N^{\mathrm{cp}}_{j,l,i,k}} c^{\mathrm{p}} \sigma_m \sigma'_m \tag{5.10}$$

For Eq. 5.10, the σ_m and σ'_m are representative surface-charge densities of the contact points. To determine the representative surface-charge densities, the connection line between the center of mass of the two contacting atoms is considered. This connection line is indicated for the contact points in Fig. 5.3a. The representative surface-charge densities are determined by determining the representative COSMO surface segments. This is done by separating both molecules along the connection line through the centers of mass of both contacting atoms until the molecules touch at their COSMO volume. This condition is illustrated in Fig. 5.4. The COSMO surface segments that are closest to the point on the COSMO surface where the connection line between the center of mass of the two contacting atoms penetrates the COSMO surface are the representative COSMO surface segments. These are marked dark yellow in Fig. 5.4.

To determine which COSMO surface segment is closest to the point were the connection line penetrates the COSMO surface, the dot product of the vector pointing from the center of mass of the atom to the center of mass of the other atom of the atom pair with the vector pointing from the center of mass of the atom to the center of its representative COSMO surface segment is used. This dot product is largest of all similar dot products with all other COSMO surface segments.

For the interaction-energy model presented, several rather rough assumptions were made. These assumptions cause that the interaction-energy model only has two universal parameters c^{p} and f^{vdW}. These parameters now have to be determined by a fit to carefully selected experimental data. Then, in principle, the combination of the interaction-energy model with MOQUAC can be applied predictively to any kind of liquid mixture.

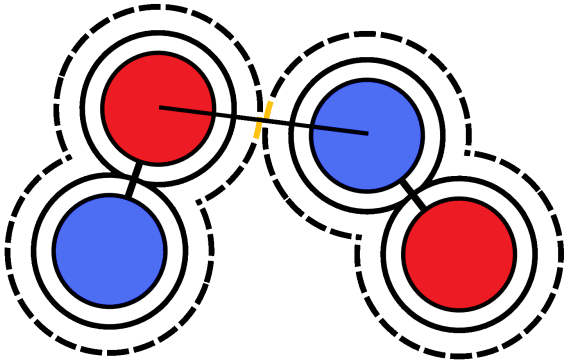


Figure 5.4: Representative surface segments of a contact point.

5.3 Selection of experimental data

In order to fit the parameters of the interaction-energy model and to show the predictive applicability of the combination of the interaction-energy model with MOQUAC, experimental data on the excess enthalpy and the isothermal vapor-liquid equilibrium of alkane + ketone and alkane + alcohol systems were selected. Alkane + ketone systems are highly non-ideal, but do not show association. Alkane + alcohol systems are also highly non-ideal, but are associating, i.e. aggregates are formed in the system. The simultaneous description of highly non-ideal associating and non-associating systems can be considered a challenge for models based on the quasi-chemical approximation [51]. Since the behavior of these systems is predominantly determined by polar interactions, disregarding the dispersion interactions in the interaction-energy model is justified.

All experimental data were taken from the Dechema Data Series [85, 86]. For the selection of the data sets used for the fit and to show the predictive applicability of MOQUAC used together with the interaction-energy model, binary systems were selected for which consistent data on the excess enthalpy and on the isothermal vapor-liquid equilibrium at a similar temperature are available. Binary systems of *n*-pentane up to *n*-decane combined with either ethanol up to 1-heptanol or 2-propanone up to 2-octanone or 3-pentanone up to 3-octanone were considered. By plotting all available data, data sets with equally distributed data points and with no obvious errors were identified. A data set obviously has errors if its data points show a strong scatter or include immediately identifiable outliers. For the vapor-liquid equilibrium data, it was necessary for both pure-component vapor pressures to be part of the data set and both consistency tests of the Dechema Data Series had to have been passed. A further condition for the data sets was that both the pressure p as well as the gas-phase composition y were measured. The pure-component vapor pressures of the evaluated data sets were plotted and compared to experimental data from the DIPPR database [27]. The pure-component vapor pressures of the selected data sets were not allowed to deviate significantly from the data of the DIPPR database. Following these steps enabled three binary alkane + ketone systems and three alkane + alcohol systems to be selected. The selected data sets for the fit of the parameters of the interaction-energy model and to show the predictive applicability of the combination of the interaction-energy model and MOQUAC are listed in Tab. 5.2.

Table 5.2: Selected data sets for the fit of the parameters of the interaction-energy model and to show the predictive applicability of the combination of the interaction-energy model with MOQUAC.

system	data type	T in K	original source
n-heptane + acetone	H^{E}	298.15	[87]
	VLE	313.15	[88]
n-heptane + 2-butanone	H^{E}	298.15	[89]
	VLE	318.15	[90]
n-heptane + 3-pentanone	H^{E}	298.15	[89]
	VLE	353.15	[91]
n-hexane + ethanol	H^{E}	298.15	[92]
	VLE	318.15	[92]
n-hexane + 1-butanol	H^{E}	288.15	[93]
	VLE	332.53	[94]
n-hexane + 1-pentanol	H^{E}	303.15	[95]
	VLE	298.15	[96]

5.4 Fitting the interaction-energy model parameters

The fitting of the parameters of the interaction-energy model was done with the help of the "gefit" program package that was developed at the Chair of Thermal Process Engineering at RWTH Aachen University. The program package was augmented with MOQUAC and the interaction-energy model. In appendix 7.7, a flowchart of the expanded gefit program is discussed. For the fit, the Levenberg-Marquardt algorithm (LMDIF) [69] is used. The objective function is:

$$\Delta = \frac{1}{\Delta H^{\mathrm{E}}} \sum_i \left(H^{\mathrm{E}}_{\mathrm{exp},i} - H^{\mathrm{E}}_{\mathrm{calc},i} \right)^2 + \frac{1}{\Delta \mathrm{VLE}} \sum_j \left[\left(y_{\mathrm{exp},j} - y_{\mathrm{calc},j} \right)^2 + \left(\frac{p_{\mathrm{exp},j} - p_{\mathrm{calc},j}}{p_{\mathrm{exp},j}} \right)^2 \right] \tag{5.11}$$

The H^{E} and y data were fitted based on absolute deviations because when relative deviations are used, the fit tends to reproduce smaller values better than higher values. For the vapor pressure, relative deviations were used because the range of the vapor pressure differs for different systems. The contributions of H^{E} to the objective function were weighted with a factor

$$\Delta H^{\mathrm{E}} = \sum_i \left(H^{\mathrm{E}}_{\mathrm{exp},i} - H^{\mathrm{E}}_{\mathrm{calc},i} \right)^2 \tag{5.12}$$

which results from evaluating the model only with the experimental H^{E} data. The contributions of y and p to the objective function were weighted with a factor

$$\Delta \mathrm{VLE} = \sum_j \left[\left(y_{\mathrm{exp},j} - y_{\mathrm{calc},j} \right)^2 + \left(\frac{p_{\mathrm{exp},j} - p_{\mathrm{calc},j}}{p_{\mathrm{exp},j}} \right)^2 \right] \tag{5.13}$$

resulting from evaluating the model only with the experimental vapor-liquid equilibrium data. In this way the considerably different magnitudes of contributions by the two different types of data were accounted for.

No conformers were considered for the fit and hence the conformer with the lowest energy state was selected for the calculation. The COSMO files required for the calculation at BP-TZVP level were taken, if available, from the COSMObase database, which was supplied with the COSMOtherm program package (Version C 2.1, COSMOlogic, Leverkusen, Germany). Otherwise they were determined with the TURBOMOLE program (Version 6.0, COSMOlogic, Leverkusen, Germany) using the software GaussView [97] to assemble the molecule. The DFT/COSMO calculation was performed with the BP86 density functional and a TZVP basis set combination.

98 directions of orientation were considered for each component. These 98 directions of orientation were distributed rather homogeneously to try to account for all relevant orientations of a molecule. For the definition of the directions of orientation, the cube from Fig. 4.15c was used. The center of mass of the molecule is positioned in the origin of co-ordinates and the initial orientation of each component was taken as given in the COSMO result. If one imagines a cube around the molecule which also has its center in the origin of co-ordinates, then the directions of orientation would penetrate the cube in the points shown in Fig. 4.15c. The random probabilities of occurrence of all directions of orientation are not equal since this distribution of directions is not exactly homogeneous. To determine the probability of occurrence of a direction of orientation, a Monte-Carlo method was applied where for 10,000,000 randomly determined directions, it was evaluated which of the 98 directions of orientation is most similar. Around each direction of orientation, 8 different equally distributed rotations are distinguished, which leads to a total of 784 different orientations per component. The frequency of occurrence of each orientation, $p^{\mathrm{R}}_{i,k}$, was determined by dividing the related probability of occurrence of the direction of orientation by 8.

The evaluation of one data point took approximately 10 minutes (Intel® Xeon® cpu x5570, 2.93 GHz) when 784 directions per component were considered. In order to reduce the computation time, similarly behaving orientations were clustered. For the clustering, the single-linkage method described in section 4.4 was applied. This means that Eq. 4.36 was applied to determine the similarity of two clusters. For the measure of distance, Eq. 4.38 was used because the application of this equation allows for the fastest algorithm, as was discussed in section 4.4.

After the clustering, the orientations in one cluster were merged and the average interaction energies of the cluster were determined. To do this, each interaction energy should have been weighted with the random probability of occurrence of its related orientation. However, since both the interaction energies and the random probability of occurrence of each orientation were similar, the arithmetic mean interaction energy was used instead.

The orientations in one cluster are only similarly behaving, not identically, which leads to a deviation between the MOQUAC result with and without clustering. As the maximum allowed distance between two clusters increases, the deviation between the result with and without clustering also increases. The maximum distance allowed between two orientations, $d_{\max}$, was set to 1 kJ/mol, a setting which a number of preinvestigations had suggested to be appropriate. Section 5.6 shows that for this value there is a significant decrease in computation time with a negligible deviation to the result without clustering.

To further reduce the required computation time for the fit, only 3 data points per data set were considered. These were the data points for a liquid composition in mole fraction x_1 of approximately 0.25, 0.50 and 0.75. The exact composition of each data point is given in Tab. 5.3. The pure-component vapor pressures were set equal to the experimental values given in the data set. The coordination number z_i of each component was set to 10, which is a common value for liquids. The coordination number z of the combinatorial contribution also equals 10, since with this value the new and optimized standard segment was determined. The structural

parameters r and q were taken from the table in the Dechema Data Series and corrected for the new reference segment of chapter 3. Their values are given in Tab. 5.4.

Table 5.3: Liquid composition x_1 of the data points that were actually considered for the parameter fitting.

system	H^E data	VLE data
n-heptane (1) + acetone (2)	0.2500; 0.5000; 0.7500	0.2571; 0.4920; 0.7386
n-heptane (1) + 2-butanone (2)	0.2500; 0.5000; 0.7500	0.2470; 0.5038; 0.7236
n-heptane (1) + 3-pentanone (2)	0.2500; 0.5000; 0.7500	0.2510; 0.4170; 0.7980

Table 5.4: Structural parameters r and q for the selected components for the fit with the new reference segment.

component	r	q
n-heptane	240.85	56.884
acetone	119.79	30.228
2-butanone	151.18	37.215
3-pentanone	182.57	44.203

5.4.1 Results of the fit and discussion

The parameters f^{vdW} and c^{p} of the interaction-energy model were adjusted to the experimental data of the alkane + ketone mixtures. The objective function, Eq. 5.11, is not continuous and thus has a discontinuous derivative with respect to f^{vdW}. The reason for this behavior is that each contact point contributes to the interaction energy of a molecular contact as specified in Eq. 5.10. A small increase in f^{vdW} can cause that for a molecular contact an additional contact point results. This inclusion of an additional contact point leads to a jump in the interaction energy of the molecular contact. To nevertheless use a derivative-based optimization procedure, f^{vdW} was varied in the range from 1.00 to 1.50 and kept fixed while c^{p} was fitted to the experimental data. It was found that for higher values of f^{vdW} the fit has convergence problems and good results are not obtained.

First, c^{p} was fitted only to the H^E data and only to the VLE data separately. ΔH^E and ΔVLE were then determined using Eqs. 5.12 and 5.13 at $f^{\mathrm{vdW}} = 1.00$ and the c^{p} that resulted from the respective fit. For these values, ΔH^E equals 1.6531×10^5 and ΔVLE equals 4.7580×10^{-2}. ΔH^E and ΔVLE were used to normalize the sums of squared residuals of the fits to the individual data types and were also used for the simultaneous fit to both the H^E and VLE data. Figure 5.5 shows the normalized minimal sum of squared residuals for all different fits as a function of f^{vdW}. The result in Fig. 5.5 shows that Δ of the fit to only the VLE data varies only minimal in the range from 1.00 to 1.30 with a minimum of around 1.2 and then starts to increase. The sum of squared residuals of the fit to only the H^E data and, because of that, also of the fit to both the H^E and VLE data simultaneously significantly increases from about $f^{\mathrm{vdW}} = 1.05$.

In Fig. 5.6, a detailed view of Fig. 5.5 is shown. In this detailed view it can be seen that Δ of the fit to only the H^E data has a local maximum at $f^{\mathrm{vdW}} \approx 1.0125$ and a local minimum at $f^{\mathrm{vdW}} \approx 1.03$. For higher values of f^{vdW} the sum of squared residuals steadily increases. The sum of squared residuals of the fit to both data types has a global minimum at $f^{\mathrm{vdW}} = 1.040$.

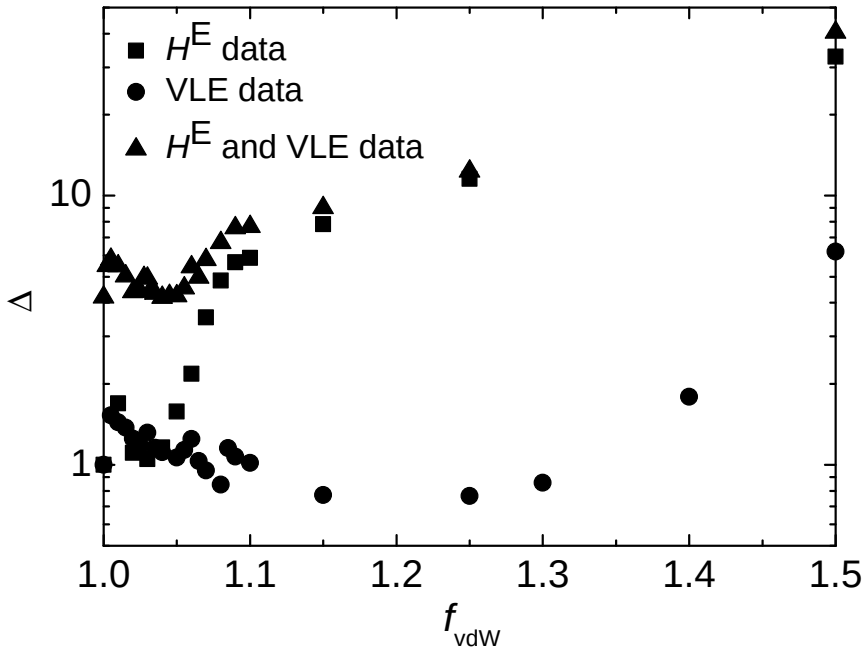


Figure 5.5: Normalized minimal sum of squared residuals as a function of f^{vdW} for the fit only to H^E data, only to VLE data and to both H^E and VLE data simultaneously.

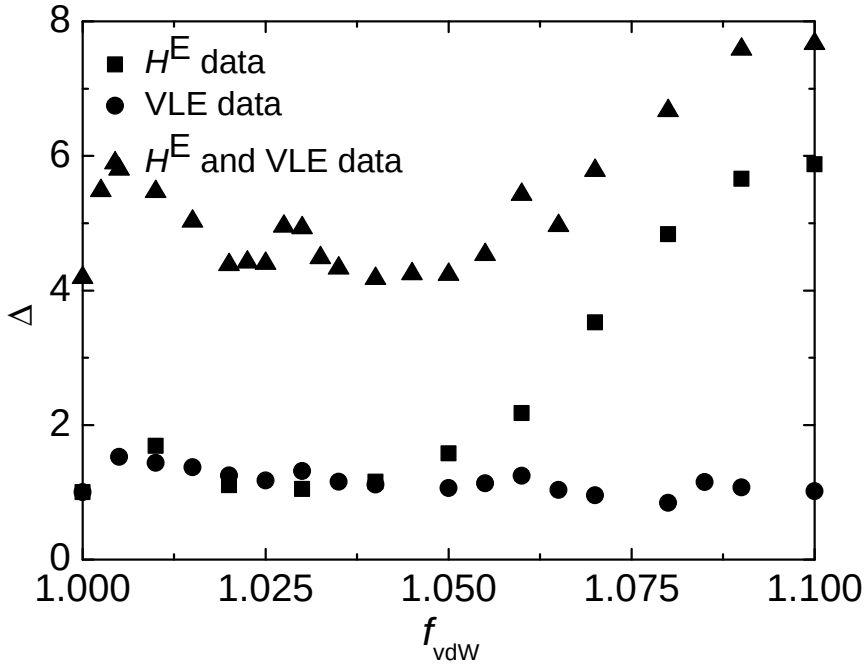


Figure 5.6: Detailed view of Fig. 5.5. Normalized minimal sum of squared residuals as a function of f^{vdW} for the fit only to H^E data, only to VLE data and to both H^E and VLE data simultaneously.

The course of the sum of squared residuals of the fit only to VLE data scatters somewhat but shows a relatively constant value in this region. The course of the sum of squared residuals of the fit to H^{E} and VLE data simultaneously shows a corresponding scatter.

To see how strong the influence of f^{vdW} is on the number of contact points of a molecular contact, the average number of contact points of all relevant molecular contacts of the three alkane + ketone systems as a function of f^{vdW} was also evaluated. Figs. 5.7 and 5.8 shows the result of this investigation. The average number of contact points appears to have a continuous course. What is particularly noteworthy is the fact that, contrary to the sum of squared residuals of the fit only to VLE data, the average number of contact points here does not show local extrema. Because each alkane + ketone system has 2,458,624 different molecular contacts, the discontinuities in the average number of contact points are probably evened out.

The result of this investigation additionally shows that at the optimal value of $f^{\mathrm{vdW}} = 1.040$, the average number of contact points equals approximately 1.5. On average, each molecular contact is thus characterized by more than one contact point. Contact points can thus not be considered to be independent of one another and therefore single-contact models are based on erroneous assumptions.

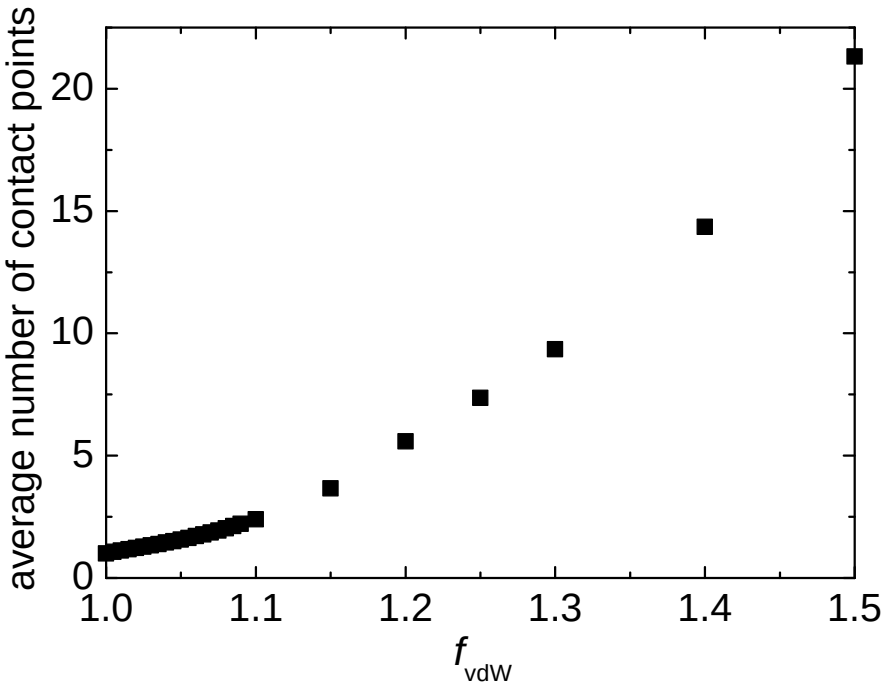


Figure 5.7: Average number of contact points of all relevant molecular contacts of the three alkane + ketone systems in dependence of f^{vdW}.

Figure 5.9 shows the parameter c^{p} as a function of f^{vdW} for the fit to the three objective functions. The value of the parameter c^{p} decreases with an increasing f^{vdW}, which is caused by the strong increase in the number of contact points with an increasing f^{vdW}, which is compensated for by a decrease in c^{p}. The results in the figure show that for smaller values of f^{vdW}, the optimal value of the parameter c^{p} is different for a fit only to H^{E} data as compared to a fit only to VLE data. The optimal value of the parameter c^{p} for a simultaneous fit to H^{E} and VLE data lies almost

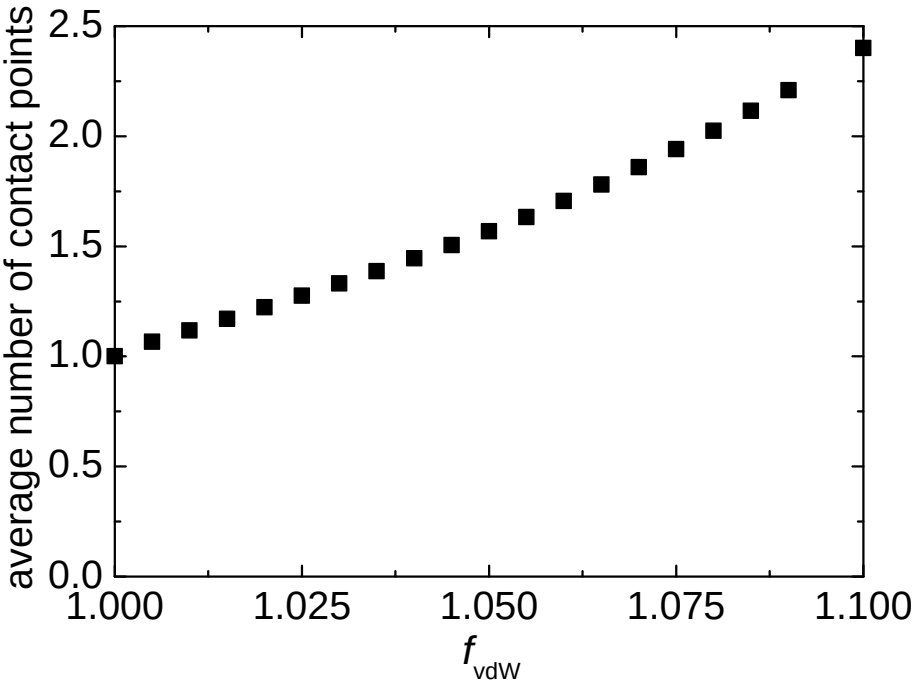


Figure 5.8: Detailed view of Fig. 5.7. Average number of contact points of all relevant molecular contacts of the three alkane + ketone systems in dependence of f^{vdW}.

exactly between the values of the fits to the individual data types. This suggests that the choice of the relative weights of the individual objective functions was appropriate.

With Figs. 5.5 and 5.9, the optimal set of parameters for the description of the experimental data can be determined as shown in Tab. 5.5. In the following, the last values simultaneously optimized for H^{E} and VLE are chosen.

Table 5.5: Resulting sets of parameters from the three different fits to the selected experimental data of the alkane + ketone systems.

data	f^{vdW}	c^{p}
	-	$\mathrm{J\AA}^4/\mathrm{mole}^2$
only H^{E}	1.000	7.6289×10^7
only VLE	1.250	2.4039×10^7
H^{E} and VLE	1.040	5.4086×10^7

The mean squared residuals for the different data types are given by the following definitions:

$$\bar{\varepsilon}_{H^{\mathrm{E}}} = \sqrt{\frac{\sum_{k=1}^{N^{H^{\mathrm{E}}}} \left(H_k^{\mathrm{E,exp}} - H_k^{\mathrm{E,calc}} \right)^2}{N^{H^{\mathrm{E}}}}} \tag{5.14}$$

$$\bar{\varepsilon}_{y} = \sqrt{\frac{\sum_{k=1}^{N^{\mathrm{VLE}}} \sum_{i=1}^{N^{\mathrm{c}}} \left(y_{i,k}^{\mathrm{exp}} - y_{i,k}^{\mathrm{calc}} \right)^2}{N^{\mathrm{VLE}} N^{\mathrm{c}}}} \tag{5.15}$$

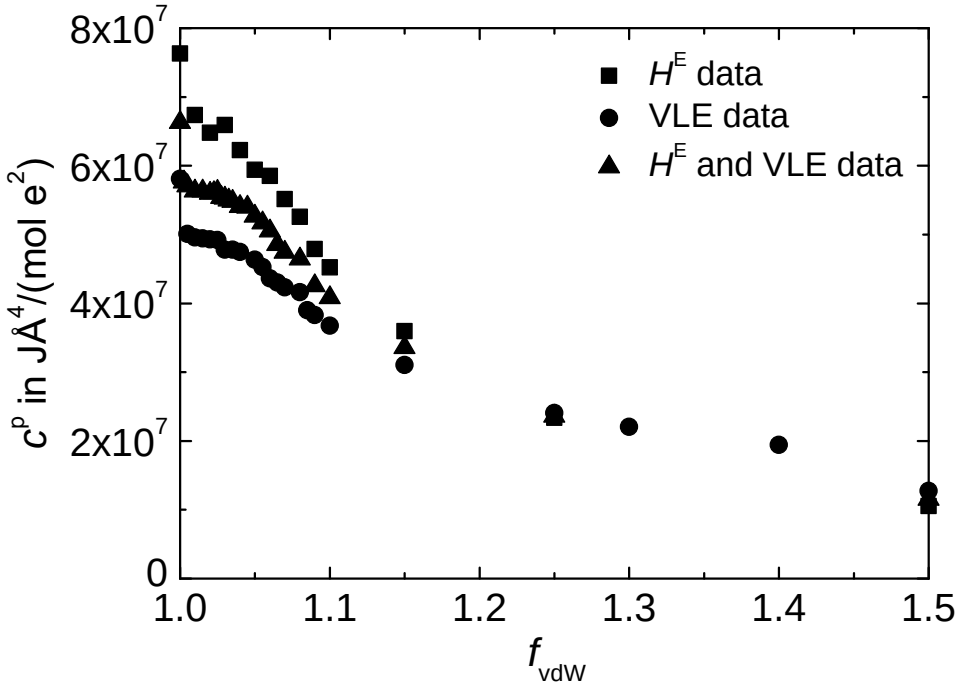


Figure 5.9: The parameter c^{p} as a function of f^{vdW} for the fit to only the H^{E} data, only the VLE data and both the H^{E} and VLE data simultaneously.

and

$$\bar{\varepsilon}_p = \sqrt{\frac{\sum_{k=1}^{N^{\mathrm{VLE}}} \left(\frac{p_k^{\mathrm{exp}} - p_k^{\mathrm{calc}}}{p_k^{\mathrm{exp}}} \right)^2}{N^{\mathrm{VLE}}}} \tag{5.16}$$

In Tab. 5.6, the mean squared residuals for the data points that were selected for the parameter fit are given. This table also shows the mean squared residuals of the COSMO-RS model and the modified UNIFAC (Dortmund) model with the same data. The result shows that the combination of the interaction-energy model with the MOQUAC model and the parameters that were determined from the fit simultaneous to H^{E} and VLE data describes the H^{E} data significantly better and the VLE data somewhat worse than the COSMO-RS model. The modified UNIFAC (Dortmund) model gives the best description of all data and is better than both other models.

Table 5.6: Mean squared residuals for the data points that were selected for the parameter fit.

model	$\bar{\varepsilon}_{H^{\mathrm{E}}}$ in J/mol	$\bar{\varepsilon}_p$ -	$\bar{\varepsilon}_y$ -
MOQUAC (fit both to H^{E} and VLE data)	200.26	0.09765	0.02237
MOQUAC (fit only to H^{E} data)	135.53	-	-
MOQUAC (fit only to VLE data)	-	0.05084	0.02698
COSMO-RS	360.60	0.06480	0.01935
modified UNIFAC (Dortmund)	46.09	0.02340	0.01133

In Fig. 5.10, a comparison of experimental H^{E} data with the model results for the system 2-

butanone + *n*-heptane is given. Modified UNIFAC (Dortmund) describes the experimental data very well compared to the other two models, while MOQUAC is slightly better than COSMO-RS. The slight asymmetry in the experimental data is described well by the modified UNIFAC (Dortmund) and COSMO-RS model, but somewhat overestimated by MOQUAC. For the other two systems, a similar result for the H^{E} data was obtained. These results are shown in appendix 7.8 in Figs. 7.3 to 7.5.

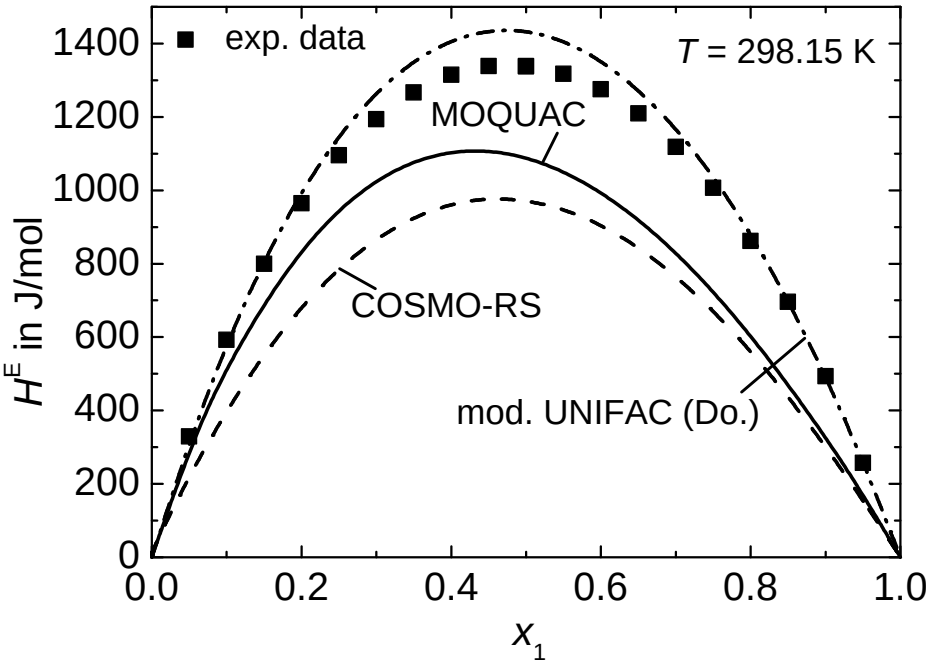


Figure 5.10: Comparison of experimental data with the model results for the system 2-butanone (1) + *n*-heptane (2) at $T = 298.15\,\mathrm{K}$.

Fig. 5.11 for the system acetone + *n*-heptane and Fig. 5.12 for the system *n*-heptane + 3-pentanone show a comparison of experimental data on the vapor pressure with the model results. The system acetone + *n*-heptane is the system for which VLE data are worst described with MOQUAC, since with MOQUAC demixing for this system at 313.15 K is predicted. The modified UNIFAC (Dortmund) model gives a very good description of both systems and is better than the other two models for both systems. COSMO-RS underestimates the experimental data of both systems and gives the worst description of the system *n*-heptane + 3-pentanone. In appendix 7.8, all remaining plots for the VLE data are shown in Figs. 7.3 to 7.11.

It has to be pointed out that the parameters of the interaction-energy model were fitted to the selected experimental data, and that the parameters of COSMO-RS and the modified UNIFAC (Dortmund) model were not. Despite this, MOQUAC does not describe the systems considerably better. However, for the application of MOQUAC only two parameters have been fitted. So although MOQUAC does not give a perfect description of the three selected alkane + ketone systems, the result is nonetheless considered satisfactory and is considered to demonstrate the applicability of MOQUAC to real mixtures.

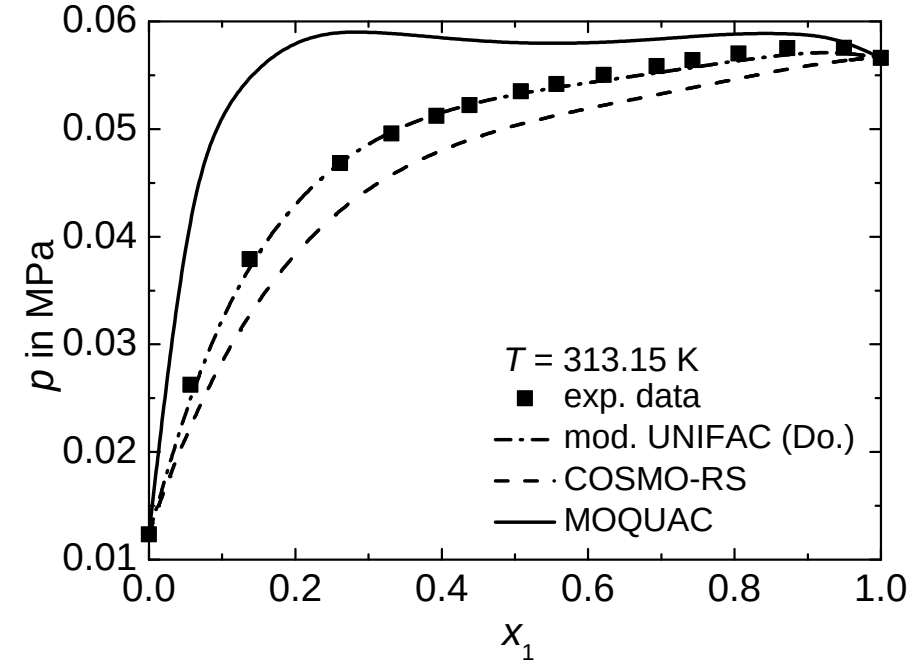


Figure 5.11: Comparison of experimental data with the model results for the system acetone (1) + *n*-heptane (2) at $T = 313.15\,\mathrm{K}$.

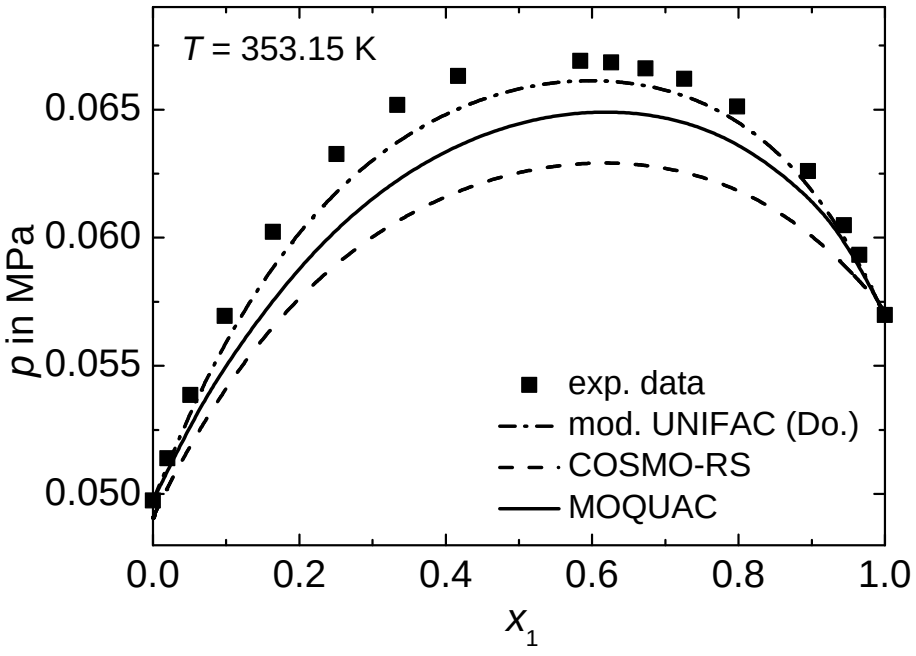


Figure 5.12: Comparison of experimental data with the model results for the system *n*-heptane (1) + 3-pentanone (2) at $T = 353.15\,\mathrm{K}$.

5.5 Predictive application of MOQUAC

To evaluate the predictive applicability of the combination of the interaction-energy model with MOQUAC, the combination of both models was applied to alkane + alcohol systems. Alcohols are strongly associating, which makes alkane + alcohol systems very challenging to model, especially for models based on the quasi-chemical approach [98]. Alcohols can form hydrogen bonds, which are very strong interactions that require a specific orientation of the contacting molecules. The COSMO-RS model accounts for hydrogen bonding by an additional term that takes the excess of the interaction energy over the polar interactions into account. The interaction-energy model that is used in this work does not have such an additional term for hydrogen bonding, however.

The combination of the interaction-energy model with MOQUAC was applied to predict H^{E} and VLE data of the three selected alkane + alcohol systems with the parameters that resulted from the fit. The calculation was performed with the same assumptions as for the parameter fitting. Since no conformers were considered, again the conformer with the lowest energy was selected for the calculation.

Tab. 5.7 shows the structural parameters of the relevant components. These parameters were determined with the new reference segment described in chapter 3.

Table 5.7: Structural parameters r and q for the selected components with the new reference segment.

component	r	q
n-hexane	209.46	49.897
ethanol	98.006	25.518
1-butanol	160.79	39.493
1-pentanol	192.18	46.481

As for the parameter fitting, only three data points per system were considered in the evaluation of the different model results. The compositions x_1 of the selected data points are given in Tab. 5.8.

Table 5.8: Liquid composition x_1 of the data points that were considered for the evaluation of the models.

system	H^{E} data	VLE data
n-hexane (1) + ethanol (2)	0.2577; 0.5077; 0.7672	0.2625; 0.4975; 0.7380
n-hexane (1) + 1-butanol (2)	0.2500; 0.5500; 0.7500	0.2190; 0.4886; 0.7849
n-hexane (1) + 1-pentanol (2)	0.2029; 0.5140; 0.8044	0.2000; 0.5000; 0.8000

Tab. 5.9 shows the mean squared residuals for the data points that were selected for the prediction. The result shows that MOQUAC predicts the H^{E} data slightly worse than COSMO-RS but also that it predicts the VLE data very well. MOQUAC even describes the vapor-pressure data slightly better than COSMO-RS. The best prediction is given by the modified UNIFAC (Dortmund) model, although in the case of the modified UNIFAC (Dortmund) model it is probably not a pure prediction. It is very likely that the experimental data of the three systems were used for the fitting of the model parameters.

Table 5.9: Mean squared residuals for the data points that were selected for the prediction.

model	$\bar{\varepsilon}_{H^E}$ in J/mol	$\bar{\varepsilon}_p$ -	$\bar{\varepsilon}_y$ -
MOQUAC	650.83	0.05448	0.01583
COSMO-RS	451.02	0.06654	0.00625
modified UNIFAC (Dortmund)	46.38	0.01400	0.00773

Fig. 5.13 shows a comparison of experimental H^E data with the model results for the system *n*-hexane + 1-pentanol. Both MOQUAC and COSMO-RS greatly overestimate the experimental data, whereas the modified UNIFAC (Dortmund) model gives a rather good description of the data. The experimental data shows a strong asymmetry that is overestimated by modified UNIFAC (Dortmund) and slightly underestimated by the MOQUAC and COSMO-RS models. For the other two systems, a similar result for the H^E data was obtained. These results are shown in appendix 7.9 in Figs. 7.12 and 7.13.

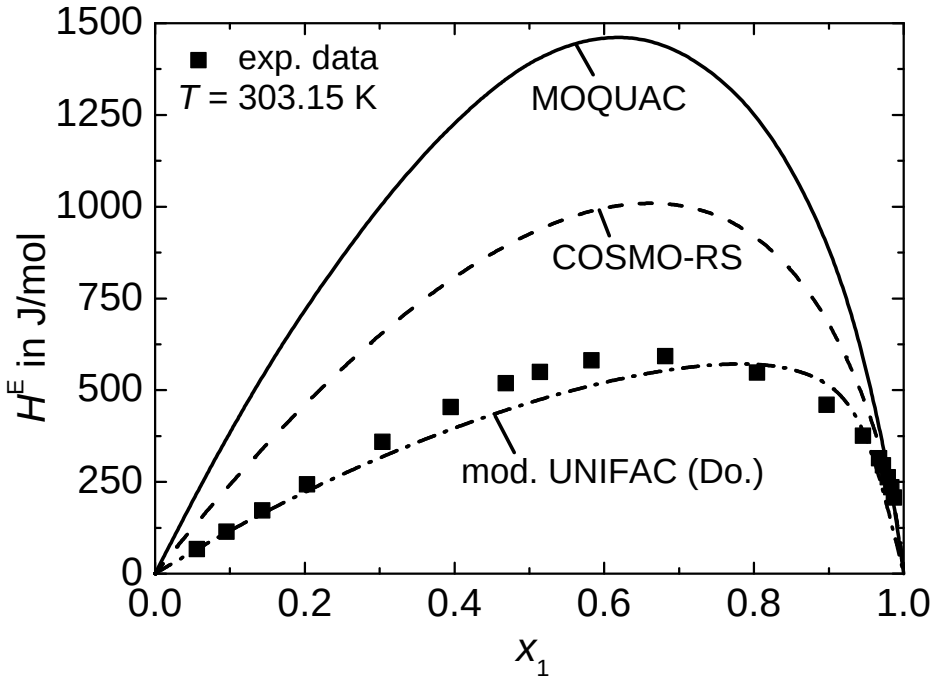


Figure 5.13: Comparison of experimental data with the model results for the system *n*-hexane (1) + 1-pentanol (2) at $T = 303.15\,\mathrm{K}$.

Figs. 5.14 and 5.15 show the comparison for the VLE data of the system *n*-hexane + 1-pentanol. MOQUAC's description of the VLE data of this system is remarkably good and only slightly overestimates the vapor pressure for most of the data points. Except for the data point at a high 1-pentanol concentration, the *y* data is very well predicted by the MOQUAC model. The modified UNIFAC (Dortmund) model gives the best description of the VLE data and the performance of the COSMO-RS model is similar to that of MOQUAC. In appendix 7.9 in Figs. 7.12 to 7.20 the plots of the comparisons of the experimental data with the different model results

for the other alkane + alcohol systems are given.

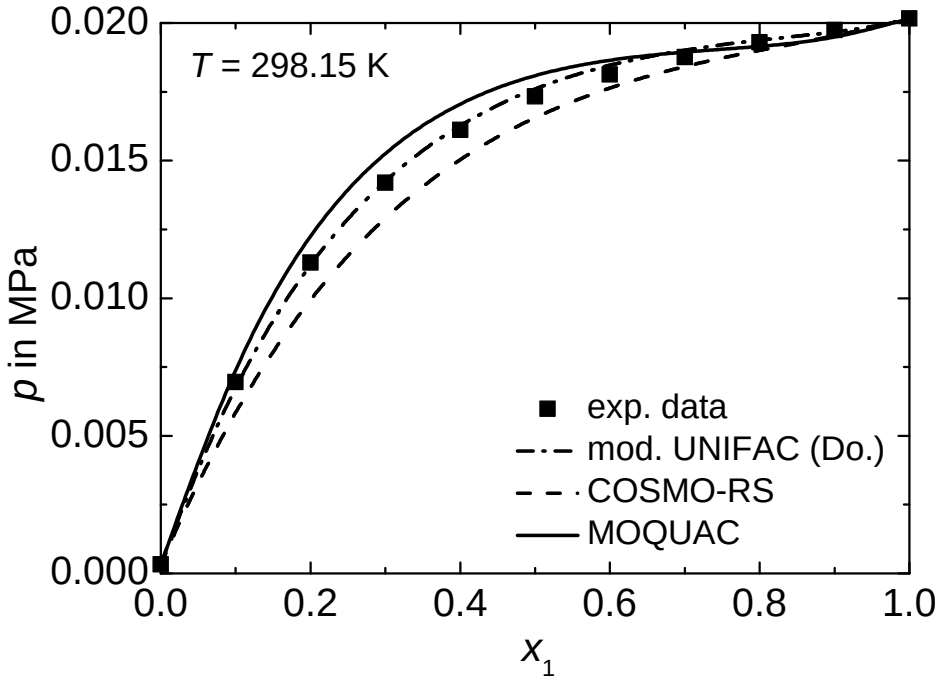


Figure 5.14: Comparison of experimental data with the model results for the system *n*-hexane (1) + 1-pentanol (2) at $T = 298.15\,\mathrm{K}$.

The combination of the interaction-energy model with MOQUAC enabled a surprisingly good prediction of experimental data of alkane + alcohol systems. The interaction-energy model is rather simple and contrary to COSMO-RS does not contain an additional term for hydrogen bonds. In addition, the model was only fitted to polar systems that show neither association nor hydrogen bonds. Despite this, MOQUAC is able to describe hydrogen bonding and association. This result shows that MOQUAC accounts for structural information appropriately.

Because of the simplicity of the interaction-energy model, the results of the predictive application of MOQUAC to real systems must be considered very promising. Both interaction-energy model and MOQUAC can still be improved in a variety of ways, however. For example for the interaction-energy model a more realistic distance dependence of the average interatomic-interaction energy as well as an additional term for dispersion interactions need to be considered. For the MOQUAC model, the choice of the orientations that are considered for each component can be further optimized. Furthermore, conformers of each component can be considered for the calculations. There is thus great potential for improving the predictive description of thermophysical data with MOQUAC. What is more, since MOQUAC can be used to determine the enthalpy and Gibbs energy of pure components, the application of the model is not limited to excess properties. The model can, in principle, also be used for the prediction of pure-component properties such as vapor pressures and enthalpies of vaporization.

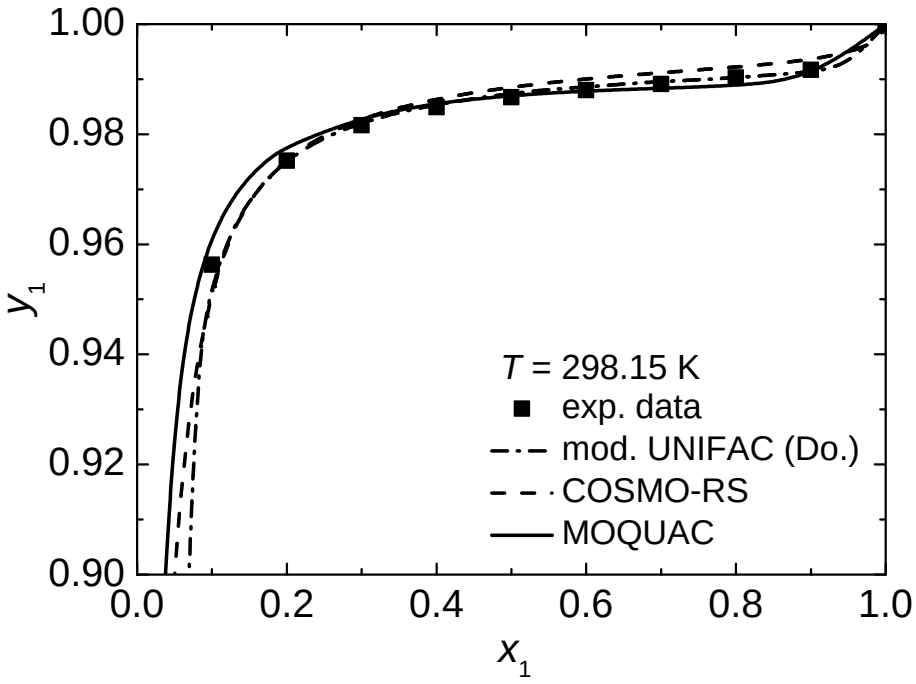


Figure 5.15: Comparison of experimental data with the model results for the system *n*-hexane (1) + 1-pentanol (2) at $T = 298.15\,\mathrm{K}$.

5.6 Clustering of orientations

To reduce the computation time for the parameter fitting, similarly behaving orientations were clustered. Similarity in the behavior of orientations is defined based on the values for the interaction energies. The parameters of the interaction-energy model therefore influence the number of clusters that are formed. Since the maximum distance allowed between two orientations $d_{\max}$ was kept fixed during a fit, the deviation of the model result with and without clustering varies, as the parameters of the interaction-energy model vary. To prevent this effect negatively influencing the result of the parameter fitting, the maximum allowed distance between two clusters was set to a rather small value of 1 kJ/mol. This value was determined from a number of preinvestigations which are discussed in this section.

For the fit, the distance between two orientations was determined with Eq. 4.38. Alternatively Eq. 4.37 with, for example,

$$v = w = 2 \tag{5.17}$$

could have been used to determine the similarity in behavior between two orientations. To evaluate the difference between the two definitions of distance, a system of two components is considered. The first component is acetone and the second component is an imaginary component with a charge density of each surface segment equal to zero. With this definition of component 2, only the interaction of acetone with itself determines which orientations of acetone can be merged. The error that is induced by summarizing similarly behaving orientations in clusters is

defined as

$$\varepsilon_H = \frac{H_1^{0,\text{with clustering}} - H_1^{0,\text{without clustering}}}{H_1^{0,\text{without clustering}}} \times 100\,\% \tag{5.18}$$

where H_1^0 is the enthalpy of pure component 1, i.e., acetone.

Section 4.4 showed that the computation time that is required to solve MOQUAC scales with $N^{\text{or}3}$. For component 2, only 1 orientation needs to be distinguished, since all interaction energies between two molecules that involves a component 2 equal zero. For acetone, 784 orientations are distinguished, as was also the case for the parameter fitting. The number of orientations for acetone after the clustering is therefore a measure for the computation time that is required to solve the MOQUAC equation system for this binary system.

In Fig. 5.16, the error ε_H is plotted versus the number of orientations N_1^{or} of acetone after the clustering for the two different definitions of distance between two orientations. Figure 5.16 was generated by varying $d_{\max}$. As can be seen, both definitions yield similar results. With the maximum allowed distance of 1 kJ/mol for Eq. 4.38, the number of clusters of acetone equals 652 with a negligible error ε_H smaller that 1.0×10^{-4} %.

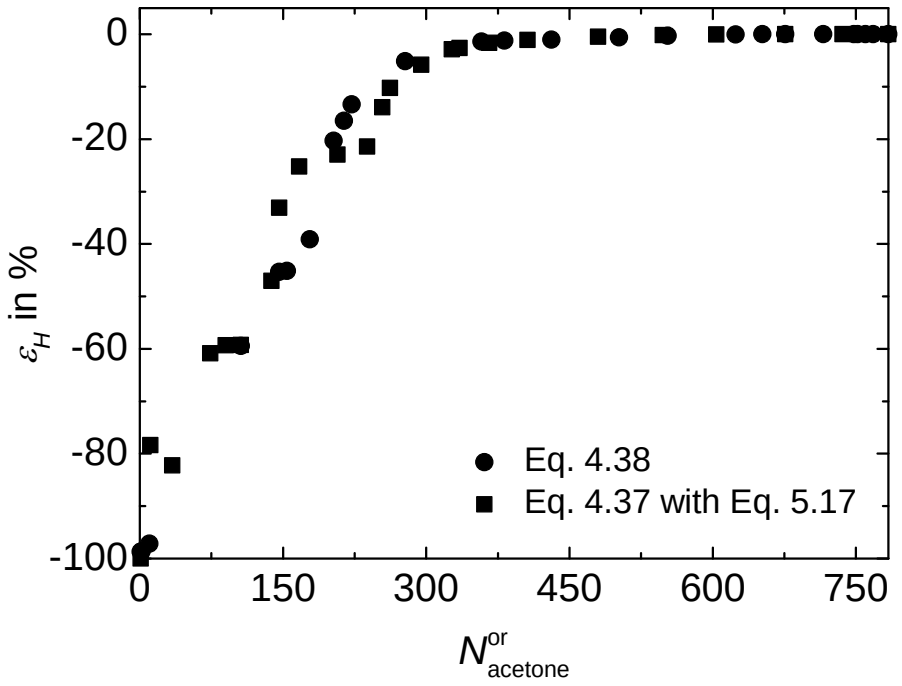


Figure 5.16: The difference between the model result with and without clustering as a function of the number of orientations after the clustering.

If the three data points for both the H^{E} and VLE data of all three alkane + ketone systems selected are evaluated with the optimal values for the parameters of the interaction-energy model, a total computation time of 181 minutes is required (Intel® Xeon® cpu x5570, 2.93 GHz). With clustering with $d_{\max} = 1$ kJ/mol, this computation time is reduced to only 61 minutes. The difference in the sum of squared residuals using Eq. 5.11 with the same values for ΔH^{E} and ΔVLE that were also applied for the fit is less than 2.0×10^{-2} %. Thus, with a negligible error

clustering enabled the computation time for the parameter fitting to be reduced by a factor of approximately three.

6 Summary

Modeling of the excess Gibbs energy enables description of the thermodynamic behavior of molecules in liquid mixtures. This is required for many applications in chemical engineering, such as the description of phase equilibria for, for example, the design and optimization of thermal separation processes. The three-dimensional molecular structure influences the behavior of molecules, since it determines if effects like steric hindrance or multiple contact points can occur. These effects are especially important for molecules with several strongly interacting functional groups. The three-dimensional molecular structure also needs to be accounted for if enantiomers have to be distinguished. Bio-based components often have several strongly interacting functional groups and show enantiomerism so consideration of the three-dimensional molecular structure is of special significance for such molecules.

In this work it has, however, been shown that state-of-the-art G^{E}-models are not able to take this three-dimensional molecular structure into account. It can thus not be expected that state-of-the-art G^{E}-models perform well for mixtures of components with several strongly interacting functional groups. This explains why these models need further development, especially due to the fact that a shift towards greater use of renewable resources in the chemical industry is likely.

G^{E}-models often distinguish between a combinatorial and a residual contribution where, with the combinatorial contribution, only effects of size and shape are described. All energetic effects, on the other hand, are accounted for by the residual contribution. This work has presented an improved model for the combinatorial contribution and a new model for the residual contribution.

The combinatorial term of UNIQUAC is based on the physically founded Staverman-Guggenheim model. When applied to real mixtures, the UNIQUAC combinatorial term performs badly, which is why several empirical modifications of the combinatorial contribution have previously been proposed. In this work, a physically founded improvement of the UNIQUAC combinatorial term is presented. This improvement performs similarly to or better than the empirical modifications.

The improvement of the combinatorial term consists of adjusting the size of the standard segment. While Abrams and Prausnitz chose the size of the spherical standard segment of UNIQUAC rather arbitrarily [11], the size of the new standard segment is determined by a fit to carefully selected experimental data of *n*-alkane mixtures. The new standard segment is, however, still considered to be spherical. Since *n*-alkane mixtures show a slight deviation from athermal behavior, a term to describe the residual contribution is also considered for the fit. In this way, the size of the standard segment is determined with high significance from the data. The residual contribution is described with an empirical model that describes the influence of temperature and size on the behavior of *n*-alkane mixtures. The parameters of this model and the size of the standard segment are determined by a fit to experimental H^{E}-data and data on activity coefficients at infinite dilution.

The new standard segment has a radius which is 3.6 times smaller than that proposed by Abrams and Prausnitz. The size of the new standard segment is regarded as a scale on which molecules “feel” their geometric details. The new radius is a little less than that of a hydrogen atom so this new interpretation seems plausible.

For the residual contribution, a new model called MOQUAC was derived. MOQUAC is based on the quasi-chemical approach and explicitly takes the orientation of molecules upon a molecular contact into account. By doing this, the full three-dimensional molecular structure can be accounted for. MOQUAC can be considered a systematic, physically founded further development of today's G^{E}-models. A comparison of MOQUAC with results from lattice simulations has shown that, contrary to state-of-the-art G^{E}-models, MOQUAC can accurately describe molecules that have complex interactions, for example, interactions with more than one contact point.

For MOQUAC to be able to be applied to real components, a model is required that describes the interaction energy between two components as a function of their orientation. A model that is capable of describing this interaction energy predictively is derived. This model uses information from a quantum-chemical calculation with COSMO. Since COSMO calculations can be performed for many different molecules, the interaction-energy model and thus MOQUAC can be applied to a broad range of molecules.

The interaction-energy model only has two parameters that are fitted to experimental data of *n*-alkane + ketone systems. After the fit, the combination of the interaction-energy model with MOQUAC describes these systems satisfactorily with a similar degree of accuracy to COSMO-RS. The results are also compared to results using the modified UNIFAC (Dortmund) model, which gives the best description of the selected *n*-alkane + ketone mixtures.

With the combination of the interaction-energy model and MOQUAC, a surprisingly good prediction of some experimental data of *n*-alkane + alcohol systems was achieved. This result is remarkable since, contrary to the systems that were used for the parameter fit, in alkane + alcohol systems, hydrogen bonding and association occur. MOQUAC in combination with the rather simple interaction-energy model again predicts the data of these systems with a similar degree of accuracy to COSMO-RS, although COSMO-RS has an additional specific term for hydrogen bonding something the interaction-energy model does not.

Because of the current simplicity of the interaction-energy model, the results of the predictive application of MOQUAC to real systems are very promising. Nonetheless, both the interaction-energy model and MOQUAC can still be improved in a variety of ways. For example, for the interaction-energy model, a more realistic dependence of distance of the average interatomic interaction energy as well as an additional term for dispersion interactions can be considered. For MOQUAC, the choice of the orientations that are taken into account for each component can be further optimized. In addition, conformers of each component can be considered for the calculations. There is thus a great potential for further improving the predictive description of thermo-physical data with MOQUAC.

7 Appendix

7.1 Solving the quasi-chemical equations for GEQUAC

In this section, a means of solving the quasi-chemical equations for GEQUAC is presented. Equations 2.66, 2.67 and 2.69 form the non-linear set of quasi-chemical equations of GEQUAC, for which Larsen and Rasmussen proposed a modified Newton-Raphson solution algorithm [48].

A non-random coefficient $\Gamma_{n,m}$ is defined by

$$\psi_{n,m} = \psi_n \Gamma_{n,m} \tag{7.1}$$

and Eq. 7.1 is substituted into Eq. 2.67

$$\Gamma_{m,n} = \Gamma_{n,m} \tag{7.2}$$

Introducing $\tau_{n,m}$ as

$$\tau_{n,m} = \exp\left(-\frac{\omega^{\mathrm{h}}_{n,m} - T\,\omega^{\mathrm{s}}_{n,m}}{2RT}\right) \tag{7.3}$$

and substituting both Eqs. 7.1 and 7.3 into Eq. 2.69 results in

$$\frac{\Gamma_{n,m}\Gamma_{m,n}}{\Gamma_{m,m}\Gamma_{n,n}} = \tau^2_{n,m} \tag{7.4}$$

Larsen and Rasmussen introduced a parameter b_m as

$$b_m = \sqrt{\Gamma_{m,m}} \tag{7.5}$$

With Eqs. 7.2 and 7.5, Eq. 7.4 can be written as

$$\Gamma_{n,m} = b_m b_n \tau_{n,m} \tag{7.6}$$

Substituting Eq. 7.6 into Eq. 7.1 and applying this to Eq. 2.66 results in

$$\sum_{n=1}^{N_{\mathrm{s}}} \psi_n b_m b_n \tau_{n,m} = 1 \tag{7.7}$$

$\Leftrightarrow$

$$\frac{1}{b_m} = \sum_{n=1}^{N_{\mathrm{s}}} \psi_n b_n \tau_{n,m} \tag{7.8}$$

In the solution algorithm, a vector $\vec{f}$ is introduced that needs to be minimized

$$f_m = \sum_{n=1}^{N_s} \psi_n b_n \tau_{n,m} - \frac{1}{b_m} \tag{7.9}$$

In the k-th iteration cycle

$$\vec{f}_k = \mathbf{B}\vec{b}_k - \vec{y}_k \tag{7.10}$$

$$\mathbf{J} = \mathbf{B} + \mathbf{D} \tag{7.11}$$

and

$$\vec{b}_{k+1} = \vec{b}_k - \mathbf{J}^{-1}\vec{f}_k \tag{7.12}$$

need to be calculated. It are:

- $\vec{f}$ the solution vector that needs to be minimized $(f_1, f_2, \ldots, f_{N_s})^{\mathrm{T}}$
- $\vec{b}$ the vector containing the parameters $(b_1, b_2, \ldots, b_{N_s})^{\mathrm{T}}$
- $\vec{y}$ a vector with the inverse b parameters $(1/b_1, 1/b_2, \ldots, 1/b_{N_s})^{\mathrm{T}}$
- $\mathbf{J}$ the Jacobian matrix
- $\mathbf{D}$ a diagonal matrix with the elements $\mathbf{D}(m,m) = 1/b_m^2$
- $\mathbf{B}$ a matrix with the elements $\mathbf{B}(m,n) = \psi_n \tau_{n,m}$

Since only the diagonal elements of the Jacobian matrix depend on the variables b_m, the Jacobian matrix can be determined very efficiently. Larsen and Rasmussen's solution algorithm also reduces the number of variables in the equation system consisting of Eqs. 2.66, 2.67 and 2.69 from N_s^2 to N_s.

The number of iterations that are required to solve the equation system depends on the initial guess of the unknown variables. Larsen and Rasmussen [48] describe two initial guesses. For GEQUAC, the supposedly simpler initial guess is applied [25], where the $b_{i,0}$ are set to

$$b_{i,0} = \frac{1}{\sum_{j=1}^{N_s} \psi_n \tau_{n,m}} \tag{7.13}$$

This is in fact the exact solution for infinite dilution of solutes in a pure solvent.

7.2 Experimental data sets for the fit of the new standard segment

Tables 7.1 and 7.2 contain the data sets used for the fit of the new standard segment.

DSET	ori. source	solute	solvent	T in K
10001	[99]	*n*-hexane	*n*-hexadecane	304.85, 315.35
10002	[99]	*n*-heptane	*n*-hexadecane	304.85, 315.35, 324.45
10009	[100]	*n*-hexane	*n*-hexadecane	293.15, 298.15, 303.15, 313.15, 323.15, 333.15
10010	[100]	*n*-heptane	*n*-hexadecane	293.15, 298.15, 303.15, 313.15, 323.15, 333.15
10015	[101]	*n*-hexane	*n*-hexadecane	293.15, 303.15, 313.15
10016	[101]	*n*-heptane	*n*-hexadecane	293.15, 303.15, 313.15
10019	[102]	*n*-butane	*n*-hexadecane	303.15, 313.15, 323.15
10020	[103]	*n*-pentane	*n*-dodecane	298.15, 298.15, 298.15, 280.15, 280.15, 280.15
10034	[104]	*n*-hexane	*n*-hexadecane	298.15
10035	[104]	*n*-heptane	*n*-hexadecane	298.15
10041	[105]	*n*-pentane	*n*-hexadecane	313.15, 343.15, 363.15
10042	[105]	*n*-hexane	*n*-hexadecane	313.15, 343.15, 363.15
10043	[105]	*n*-heptane	*n*-hexadecane	313.15, 343.15, 363.15
10044	[105]	*n*-octane	*n*-hexadecane	313.15, 343.15, 363.15
10045	[105]	*n*-nonane	*n*-hexadecane	313.15, 343.15, 363.15
10046	[105]	*n*-decane	*n*-hexadecane	343.15, 363.15
10059	[106]	*n*-pentane	*n*-hexane	293.15
10060	[106]	*n*-hexane	*n*-heptane	293.15
10061	[106]	*n*-pentane	*n*-octane	293.15
10065	[102]	*n*-butane	*n*-hexadecane	303.15
10066	[102]	*n*-pentane	*n*-hexadecane	303.15
10067	[102]	*n*-hexane	*n*-hexadecane	303.15
10068	[102]	*n*-heptane	*n*-hexadecane	303.15
10087	[102]	*n*-hexane	*n*-hexadecane	293.15, 313.15, 333.15
10094	[106]	*n*-pentane	*n*-heptane	293.15
10095	[106]	*n*-hexane	*n*-heptane	293.15
10096	[107]	*n*-pentane	*n*-octane	293.15
10097	[106]	*n*-pentane	*n*-octane	293.15
10102	[108]	*n*-pentane	*n*-dodecane	293.15, 303.15
10104	[109]	*n*-pentane	*n*-hexadecane	298.15
10105	[110]	*n*-pentane	*n*-hexadecane	323.15
10106	[111]	*n*-hexane	*n*-hexadecane	293.15
10107	[112]	*n*-hexane	*n*-hexadecane	293.15, 303.15, 333.15
10108	[113]	*n*-hexane	*n*-hexadecane	313.15, 323.15, 333.15
10110	[114]	*n*-hexane	*n*-hexadecane	293.15, 303.15, 313.15, 323.15, 333.15, 343.15
10112	[115]	*n*-octane	*n*-pentane	298.15
10115	[116]	*n*-pentane	*n*-heptane	304.15

10116	[117]	*n*-hexane	*n*-heptane	298.15
10118	[118]	*n*-hexane	*n*-heptane	298.15, 333.15, 373.15
10119	[117]	*n*-octane	*n*-heptane	298.15
10121	[119]	*n*-pentane	*n*-octane	303.15
10122	[120]	*n*-pentane	*n*-octane	333.15
10123	[121]	*n*-pentane	*n*-octane	313.15
10124	[119]	*n*-hexane	*n*-octane	303.15
10125	[120]	*n*-hexane	*n*-octane	333.15
10126	[122]	*n*-hexane	*n*-octane	303.15
10129	[120]	*n*-pentane	*n*-nonane	333.15
10130	[120]	*n*-hexane	*n*-nonane	333.15
10135	[123]	*n*-hexane	*n*-hexadecane	298.15
10136	[124]	*n*-hexane	*n*-hexadecane	308.15
10137	[110]	*n*-hexane	*n*-hexadecane	323.15
10138	[125]	*n*-hexane	*n*-hexadecane	333.15, 393.15, 453.15
10139	[126]	*n*-hexane	*n*-hexadecane	363.15, 393.15, 423.15
10140	[110]	*n*-heptane	*n*-hexadecane	323.15
10141	[109]	*n*-heptane	*n*-hexadecane	298.15
10142	[126]	*n*-heptane	*n*-hexadecane	363.15, 393.15, 423.15
10143	[125]	*n*-heptane	*n*-hexadecane	393.15, 453.15
10144	[110]	*n*-octane	*n*-hexadecane	323.15
10145	[126]	*n*-octane	*n*-hexadecane	363.15, 393.15, 423.15
10146	[125]	*n*-octane	*n*-hexadecane	453.15
10147	[126]	*n*-nonane	*n*-hexadecane	363.15, 393.15, 423.15
10149	[124]	*n*-pentane	*n*-hexadecane	308.15
10150	[127]	*n*-pentane	*n*-hexadecane	308.15
10152	[127]	*n*-hexane	*n*-hexadecane	308.15
10153	[124]	*n*-hexane	*n*-hexadecane	308.15, 323.15
10155	[109]	*n*-heptane	*n*-hexadecane	298.15
10156	[110]	*n*-heptane	*n*-hexadecane	323.15
10157	[128]	*n*-heptane	*n*-hexadecane	333.15, 363.15, 393.15, 423.15
10158	[126]	*n*-heptane	*n*-hexadecane	393.15, 453.15
10159	[110]	*n*-octane	*n*-hexadecane	323.15
10160	[126]	*n*-octane	*n*-hexadecane	363.15, 393.15, 423.15
10161	[125]	*n*-octane	*n*-hexadecane	453.15
10162	[126]	*n*-nonane	*n*-hexadecane	363.15, 393.15, 423.15
10191	[119]	*n*-hexane	*n*-decane	303.15, 333.15
10192	[122]	*n*-hexane	*n*-decane	303.15
10194	[120]	*n*-hexane	*n*-decane	333.15, 343.15
10201	[129]	*n*-pentane	*n*-hexadecane	293.15
10202	[119]	*n*-pentane	*n*-hexadecane	298.15, 303.15, 313.15, 333.15
10203	[130]	*n*-pentane	*n*-hexadecane	323.15
10204	[131]	*n*-hexane	*n*-hexadecane	293.15, 303.15, 333.15
10205	[129]	*n*-hexane	*n*-hexadecane	293.15
10207	[132]	*n*-hexane	*n*-hexadecane	298.15
10208	[133]	*n*-hexane	*n*-hexadecane	303.15
10209	[119]	*n*-hexane	*n*-hexadecane	303.15, 333.15

10210	[134]	*n*-hexane	*n*-hexadecane	303.15
10211	[116]	*n*-hexane	*n*-hexadecane	312.85
10212	[135]	*n*-hexane	*n*-hexadecane	315.35
10213	[130]	*n*-hexane	*n*-hexadecane	323.15
10214	[136]	*n*-hexane	*n*-hexadecane	333.15, 393.15
10215	[136]	*n*-hexane	*n*-hexadecane	333.45
10216	[135]	*n*-heptane	*n*-hexadecane	305.35
10217	[130]	*n*-heptane	*n*-hexadecane	323.15
10218	[119]	*n*-heptane	*n*-hexadecane	333.15

Table 7.1: Selected experimental γ^{∞} data. All data was taken from the DECHEMA Chemistry Data Series [15]. DSET is the data set number used for the fitting program.

DSET	ori. source	component 1	component 2	T in K (number of data points)
10023	[137]	*n*-hexane	*n*-decane	298.15 (26), 308.15 (31)
10093	[138]	*n*-hexane	*n*-dodecane	298.15 (37), 308.15 (37)
10281	[139]	*n*-pentane	*n*-decane	293.15 (3)
10284	[140]	*n*-hexane	*n*-heptane	298.15 (22)
10286	[141]	*n*-hexane	*n*-octane	298.15 (4)
10287	[142]	*n*-hexane	*n*-octane	298.15 (9)
10288	[143]	*n*-hexane	*n*-decane	298.15 (7)
10290	[139]	*n*-hexane	*n*-decane	293.15 (5)
10291	[144]	*n*-hexane	*n*-decane	298.15 (19)
10294	[145]	*n*-hexane	*n*-dodecane	283.15 (14)
10295	[145]	*n*-hexane	*n*-dodecane	293.15 (15)
10296	[145]	*n*-hexane	*n*-dodecane	303.15 (15)
10297	[141]	*n*-hexane	*n*-dodecane	298.15 (5)
10298	[146]	*n*-hexane	*n*-dodecane	298.15 (19)
10299	[147]	*n*-hexane	*n*-hexadecane	293.15 (4)
10300	[148]	*n*-hexane	*n*-hexadecane	313.15 (10)
10301	[148]	*n*-hexane	*n*-hexadecane	324.15 (5)
10302	[148]	*n*-hexane	*n*-hexadecane	333.15 (5)
10303	[148]	*n*-hexane	*n*-hexadecane	349.15 (5)
10304	[149]	*n*-hexane	*n*-hexadecane	298.15 (10)
10305	[150]	*n*-hexane	*n*-hexadecane	293.15 (13)
10306	[150]	*n*-hexane	*n*-hexadecane	303.15 (4)
10307	[150]	*n*-hexane	*n*-hexadecane	313.15 (15)
10308	[150]	*n*-hexane	*n*-hexadecane	323.15 (4)
10309	[151]	*n*-hexane	*n*-hexadecane	313.15 (43)
10310	[152]	*n*-hexane	*n*-hexadecane	298.15 (28)
10311	[152]	*n*-hexane	*n*-hexadecane	303.15 (26)
10316	[153]	*n*-heptane	*n*-dodecane	298.15 (19)
10317	[154]	*n*-heptane	*n*-hexadecane	298.15 (4)
10318	[154]	*n*-heptane	*n*-hexadecane	323.15 (8)
10319	[139]	*n*-heptane	*n*-hexadecane	293.15 (4)
10321	[155]	*n*-octane	*n*-dodecane	298.15 (11)
10322	[153]	*n*-octane	*n*-dodecane	298.15 (19)

10323	[139]	*n*-octane	*n*-hexadecane	293.15 (3)
10326	[153]	*n*-decane	*n*-dodecane	298.15 (19)
10327	[139]	*n*-decane	*n*-hexadecane	293.15 (4)

Table 7.2: Selected experimental H^{E} data. All data was taken from the DECHEMA Chemistry Data Series [86]. DSET is the data set number used for the fitting program.

7.3 Derivation of the MOQUAC entropy term

In this section, the derivation of Eq. 4.17 from Eq. 4.16 using a similar approach to Guggenheim [19] is shown. $\vec{\psi}^*$ characterizes the distribution of contacts for which the summand of Eq. 4.13 is maximal and $g(\vec{\psi}^*)$ is the number of microstates with the same $\vec{\psi}^*$. As with Guggenheim's quasi-chemical theory [19], for MOQUAC the number of distinguishable microstates of a liquid system that consists of N molecules is assumed to equal the number of distinguishable arrangements of molecules on the lattice. $g(\psi)$ is thus determined by the number of distinguishable ways N molecules can be arranged on a lattice such that the given distribution of ψ results.

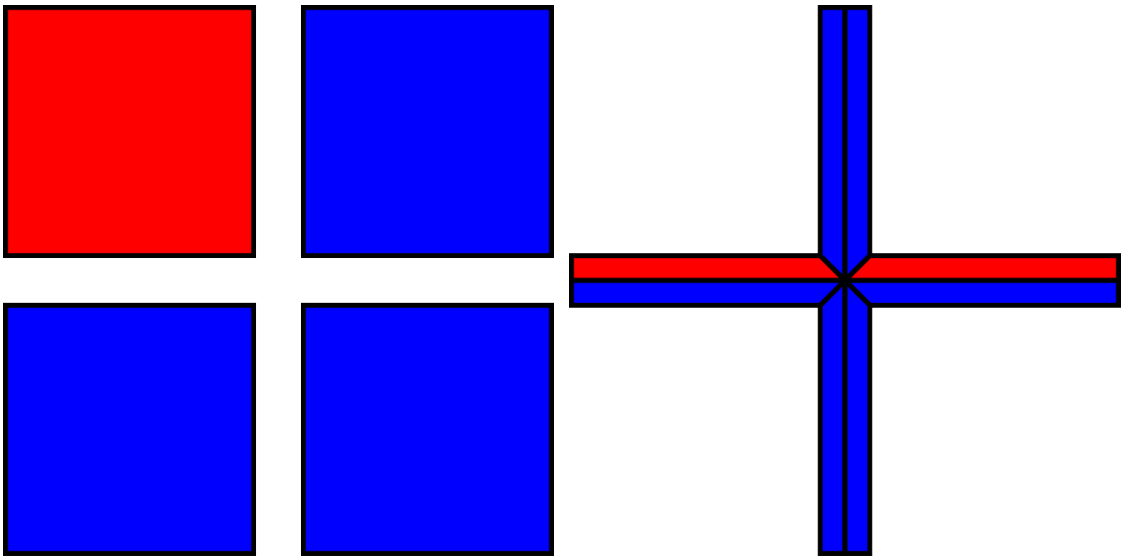
Following Guggenheim, the number of distinguishable arrangements of molecules on the lattice is described by the number of distinguishable ways the molecular contacts can be arranged on the lattice $\tilde{g}(\psi)$. To determine $\tilde{g}(\psi)$, it is additionally assumed that all molecular contacts can be placed independently on the lattice. Thus, the number of different ways to arrange all contacts on the lattice has to be determined, where two contacts j,l,i,k cannot be distinguished from one another. This problem can be represented by a mental exercise where, for example, three white balls and seven black balls are drawn from an urn. When the balls may not be returned to the urn once drawn, then $10!/(3!7!)$ different ways of drawing these ten balls exist. In a similar manner, $\tilde{g}(\psi)$ is calculated by

$$\tilde{g}(\psi) = \frac{\left(\frac{1}{2}\sum_{i=1}^{N^{\mathrm{c}}} N x_i z_i\right)!}{\prod_{i=1}^{N^{\mathrm{c}}}\prod_{k=1}^{N_i^{\mathrm{or}}}\prod_{j=1}^{N^{\mathrm{c}}}\prod_{l=1}^{N_j^{\mathrm{or}}}\left(\frac{1}{2}N x_i z_i p_{i,k}^{\mathrm{R}} \psi_{j,l,i,k}\right)!} \tag{7.14}$$

The factor $\frac{1}{2}$ appears in both the numerator and the denominator, because otherwise each contact is counted twice. The orientation of a contact is also distinguished, so for the placement on the lattice, the contact j,l,i,k is distinguished from the contact i,k,j,l.

$g(\psi)$ cannot be set equal to $\tilde{g}(\psi)$, because to do this, Eq. 7.14 has two shortcomings. First the assumption that all molecular contacts can be placed independently on the lattice is mistaken. The contacts are related to molecules and it is the molecules rather than the contacts that can be placed independently on the lattice. Certain configurations of molecular contacts will not be consistent with any of the configurations of molecules on the lattice. This is illustrated by Fig. 7.1. The four colored molecules in Fig. 7.1a can only be arranged in four different ways on the lattice. Fig. 7.1a shows one of these configurations. For all four configurations, there are two blue-blue and two blue-red contacts and according to Eq. 7.14 there are 4!/(1!1!2!) = 12 different ways of arranging the four contacts on the lattice. Thus, 8 of the arrangements of the contacts must give a configuration that is not consistent with any of the arrangements of the molecules on the lattice. One example of such an inconsistent configuration is shown in Fig. 7.1b. The cross represents the four contacts and the colors indicate the type of contact. The contact on the left as well as the contact on the right are blue-red contacts and the upper as well as the lower contact are blue-blue contacts. In the configuration of contacts, the upper left molecule should

be red according to the right contact and should be blue according to the upper contact. As this is a contradiction, the configuration of contacts shown in Fig. 7.1b is inconsistent with all configurations of molecules on the lattice.



(a) one red and three blue molecules distributed on four lattice sites

(b) a configuration of only the molecular contacts that is not consistent with any configuration of the four molecules on the lattice

Figure 7.1: Example of an inconsistent configuration of molecular contacts on a lattice.

To exclude configurations of molecular contacts on the lattice that are not consistent with any configuration of molecules on the lattice, higher approximations can be made. Guggenheim, for example, considered triangular triplets of molecular contacts or tetrahedral quadruplets instead of pairs [19]. For the derivation of MOQUAC, the assumption of independently placeable contacts is maintained. The first shortcoming is thus not corrected for.

The second shortcoming of Eq. 7.14 that does not allow $g(\psi)$ to be set directly equal to $\tilde{g}(\psi)$ is that if the $\tilde{g}(\psi)$ for all distributions of ψ are summed, then the number of all distinguishable ways of arranging the molecules on the lattice W should result:

$$W = \sum_{\text{all}\,\psi} g(\psi) \tag{7.15}$$

Summing over all $\tilde{g}(\psi)$ would overestimate W

$$W < \sum_{\text{all}\,\psi} \tilde{g}(\psi) \tag{7.16}$$

as is indicated in the example in Fig. 7.1. To correct for the second shortcoming, a scaling factor h for $\tilde{g}(\psi)$ is introduced and

$$g(\psi) = h\tilde{g}(\psi) \tag{7.17}$$

is written. Introducing Eq. 7.17 into Eq. 7.15 then results in:

$$W = \sum_{\text{all}\,\psi} h\tilde{g}(\psi) \tag{7.18}$$

To determine the scaling factor, h, W must be defined. For the modeling of G^{E}, it is typical to describe W with the equation for combinatorial entropy. For example, with his equation for the combinatorial entropy (Eq. 2.9), Staverman describes the number of different ways of arranging molecules on a lattice, considering their size and shape. Boltzmann's equation relates the molar entropy S to the number of microstates corresponding to a macrostate by

$$nS = k_{\mathrm{B}} \ln W \tag{7.19}$$

W can thus be formulated with Eq. 2.9 and Eq. 7.19:

$$W = \exp\left(\frac{nS_{\mathrm{comb}}}{k_{\mathrm{B}}}\right) \tag{7.20}$$

Eq. 7.18 is set equal to Eq. 7.20 to determine the scaling factor h:

$$\sum_{\mathrm{all}\,\psi} h\tilde{g}(\psi) = \exp\left(\frac{nS_{\mathrm{comb}}}{k_{\mathrm{B}}}\right) \tag{7.21}$$

$\Leftrightarrow$

$$h = \frac{\exp\left(\frac{nS_{\mathrm{comb}}}{k_{\mathrm{B}}}\right)}{\sum_{\mathrm{all}\,\psi} \tilde{g}(\psi)} \tag{7.22}$$

Introducing Eq. 7.14 into Eq. 7.22 results in

$$h = \frac{\exp\left(\frac{nS_{\mathrm{comb}}}{k_{\mathrm{B}}}\right)}{\sum_{\mathrm{all}\,\psi} \frac{\left(\frac{1}{2}\sum_{i=1}^{N^{\mathrm{c}}} N x_i z_i\right)!}{\prod_{i=1}^{N^{\mathrm{c}}} \prod_{k=1}^{N_i^{\mathrm{or}}} \prod_{j=1}^{N^{\mathrm{c}}} \prod_{l=1}^{N_j^{\mathrm{or}}} \left(\frac{1}{2} N x_i z_i p_{i,k}^{\mathrm{R}} \psi_{j,l,i,k}\right)!}} \tag{7.23}$$

The sum in the denominator of Eq. 7.23 can be approximated by its leading term, which according to Hála and Boublík [71] is the most likely distribution, i.e. random distribution, of all contacts. For a random distribution of contacts

$$\psi_{j,l,i,k} = \frac{x_j z_j p_{j,l}^{\mathrm{R}}}{\sum_{m=1}^{N^{\mathrm{c}}} x_m z_m} \tag{7.24}$$

so

$$h = \frac{\exp\left(\frac{nS_{\mathrm{comb}}}{k_{\mathrm{B}}}\right)}{\frac{\left(\frac{1}{2}\sum_{i=1}^{N^{\mathrm{c}}} N x_i z_i\right)!}{\prod_{i=1}^{N^{\mathrm{c}}} \prod_{k=1}^{N_i^{\mathrm{or}}} \prod_{j=1}^{N^{\mathrm{c}}} \prod_{l=1}^{N_j^{\mathrm{or}}} \left(\frac{1}{2} N x_i z_i p_{i,k}^{\mathrm{R}} \frac{x_j z_j p_{j,l}^{\mathrm{R}}}{\sum_{m=1}^{N^{\mathrm{c}}} x_m z_m}\right)!}} \tag{7.25}$$

$\Leftrightarrow$

$$h = \frac{\exp\left(\frac{nS_{\mathrm{comb}}}{k_{\mathrm{B}}}\right) \prod_{i=1}^{N^{\mathrm{c}}} \prod_{k=1}^{N_i^{\mathrm{or}}} \prod_{j=1}^{N^{\mathrm{c}}} \prod_{l=1}^{N_j^{\mathrm{or}}} \left(\frac{1}{2} N x_i z_i p_{i,k}^{\mathrm{R}} \frac{x_j z_j p_{j,l}^{\mathrm{R}}}{\sum_{m=1}^{N^{\mathrm{c}}} x_m z_m}\right)!}{\left(\frac{1}{2}\sum_{i=1}^{N^{\mathrm{c}}} N x_i z_i\right)!} \tag{7.26}$$

can be written for the scaling factor.

Substituting Eqs. 7.14, 7.17 and 7.26 into Eq. 4.16 results in

$$S = \frac{k_\mathrm{B}}{n} \ln \left(\frac{\exp\left(\frac{nS_\mathrm{comb}}{k_\mathrm{B}}\right) \left(\frac{1}{2}\sum_{i=1}^{N^\mathrm{c}} N x_i z_i\right)! \prod_{i=1}^{N^\mathrm{c}} \prod_{k=1}^{N_i^\mathrm{or}} \prod_{j=1}^{N^\mathrm{c}} \prod_{l=1}^{N_j^\mathrm{or}} \left(\frac{1}{2} N x_i z_i p_{i,k}^\mathrm{R} \frac{x_j z_j p_{j,l}^\mathrm{R}}{\sum_{m=1}^{N^\mathrm{c}} x_m z_m}\right)!}{\left(\frac{1}{2}\sum_{i=1}^{N^\mathrm{c}} N x_i z_i\right)! \prod_{i=1}^{N^\mathrm{c}} \prod_{k=1}^{N_i^\mathrm{or}} \prod_{j=1}^{N^\mathrm{c}} \prod_{l=1}^{N_j^\mathrm{or}} \left(\frac{1}{2} N x_i z_i p_{i,k}^\mathrm{R} \psi_{j,l,i,k}\right)!} \right) \tag{7.27}$$

$\Leftrightarrow$

$$S = \frac{k_\mathrm{B}}{n} \ln \left(\frac{\exp\left(\frac{nS_\mathrm{comb}}{k_\mathrm{B}}\right) \prod_{i=1}^{N^\mathrm{c}} \prod_{k=1}^{N_i^\mathrm{or}} \prod_{j=1}^{N^\mathrm{c}} \prod_{l=1}^{N_j^\mathrm{or}} \left(\frac{1}{2} N x_i z_i p_{i,k}^\mathrm{R} \frac{x_j z_j p_{j,l}^\mathrm{R}}{\sum_{m=1}^{N^\mathrm{c}} x_m z_m}\right)!}{\prod_{i=1}^{N^\mathrm{c}} \prod_{k=1}^{N_i^\mathrm{or}} \prod_{j=1}^{N^\mathrm{c}} \prod_{l=1}^{N_j^\mathrm{or}} \left(\frac{1}{2} N x_i z_i p_{i,k}^\mathrm{R} \psi_{j,l,i,k}\right)!} \right) \tag{7.28}$$

$\Leftrightarrow$

$$S = S_\mathrm{comb} + \frac{k_\mathrm{B}}{n} \ln \left(\prod_{i=1}^{N^\mathrm{c}} \prod_{k=1}^{N_i^\mathrm{or}} \prod_{j=1}^{N^\mathrm{c}} \prod_{l=1}^{N_j^\mathrm{or}} \frac{\left(\frac{1}{2} N x_i z_i p_{i,k}^\mathrm{R} \frac{x_j z_j p_{j,l}^\mathrm{R}}{\sum_{m=1}^{N^\mathrm{c}} x_m z_m}\right)!}{\left(\frac{1}{2} N x_i z_i p_{i,k}^\mathrm{R} \psi_{j,l,i,k}\right)!} \right) \tag{7.29}$$

When Stirling's approximation

$$\ln n! = n \ln n - n \tag{7.30}$$

for large factorials is applied, Eq. 7.29 can be written as:

$$\begin{aligned} S = S_\mathrm{comb} + \\ \frac{k_\mathrm{B}}{n} \sum_{i=1}^{N^\mathrm{c}} \sum_{k=1}^{N_i^\mathrm{or}} \sum_{j=1}^{N^\mathrm{c}} \sum_{l=1}^{N_j^\mathrm{or}} \left(\frac{1}{2} N x_i z_i p_{i,k}^\mathrm{R} \frac{x_j z_j p_{j,l}^\mathrm{R}}{\sum_{m=1}^{N^\mathrm{c}} x_m z_m} \ln \left(\frac{1}{2} N x_i z_i p_{i,k}^\mathrm{R} \frac{x_j z_j p_{j,l}^\mathrm{R}}{\sum_{m=1}^{N^\mathrm{c}} x_m z_m} \right) - \right. \\ \left. \frac{1}{2} N x_i z_i p_{i,k}^\mathrm{R} \frac{x_j z_j p_{j,l}^\mathrm{R}}{\sum_{m=1}^{N^\mathrm{c}} x_m z_m} - \frac{1}{2} N x_i z_i p_{i,k}^\mathrm{R} \psi_{j,l,i,k} \ln \left(\frac{1}{2} N x_i z_i p_{i,k}^\mathrm{R} \psi_{j,l,i,k} \right) + \frac{1}{2} N x_i z_i p_{i,k}^\mathrm{R} \psi_{j,l,i,k} \right) \end{aligned} \tag{7.31}$$

Applying Eqs. 4.1 and 4.3 and separating summands results in

$$\begin{aligned} S = S_\mathrm{comb} + \frac{k_\mathrm{B}}{n} \sum_{i=1}^{N^\mathrm{c}} \sum_{k=1}^{N_i^\mathrm{or}} \left(-\frac{1}{2} N x_i z_i p_{i,k}^\mathrm{R} + \frac{1}{2} N x_i z_i p_{i,k}^\mathrm{R} \right) + \\ \frac{k_\mathrm{B}}{n} \sum_{i=1}^{N^\mathrm{c}} \sum_{k=1}^{N_i^\mathrm{or}} \sum_{j=1}^{N^\mathrm{c}} \sum_{l=1}^{N_j^\mathrm{or}} \left(\frac{1}{2} N x_i z_i p_{i,k}^\mathrm{R} \frac{x_j z_j p_{j,l}^\mathrm{R}}{\sum_{m=1}^{N^\mathrm{c}} x_m z_m} \ln \left(\frac{1}{2} N x_i z_i p_{i,k}^\mathrm{R} \frac{x_j z_j p_{j,l}^\mathrm{R}}{\sum_{m=1}^{N^\mathrm{c}} x_m z_m} \right) - \right. \\ \left. \frac{1}{2} N x_i z_i p_{i,k}^\mathrm{R} \psi_{j,l,i,k} \ln \left(\frac{1}{2} N x_i z_i p_{i,k}^\mathrm{R} \psi_{j,l,i,k} \right) \right) \end{aligned} \tag{7.32}$$

$\Leftrightarrow$

$$S = S_{\mathrm{comb}} + \frac{k_{\mathrm{B}}}{n} \sum_{i=1}^{N^{\mathrm{c}}} \sum_{k=1}^{N_i^{\mathrm{or}}} \sum_{j=1}^{N^{\mathrm{c}}} \sum_{l=1}^{N_j^{\mathrm{or}}} \left(\frac{1}{2} N x_i z_i p_{i,k}^{\mathrm{R}} \frac{x_j z_j p_{j,l}^{\mathrm{R}}}{\sum_{m=1}^{N^{\mathrm{c}}} x_m z_m} \ln \left(\frac{1}{2} N x_i z_i p_{i,k}^{\mathrm{R}} \frac{x_j z_j p_{j,l}^{\mathrm{R}}}{\sum_{m=1}^{N^{\mathrm{c}}} x_m z_m} \right) - \frac{1}{2} N x_i z_i p_{i,k}^{\mathrm{R}} \psi_{j,l,i,k} \ln \left(\frac{1}{2} N x_i z_i p_{i,k}^{\mathrm{R}} \psi_{j,l,i,k} \right) \right) \quad (7.33)$$

$\Leftrightarrow$

$$\begin{aligned} S = S_{\mathrm{comb}} + & \\ \frac{k_{\mathrm{B}}}{n} \sum_{i=1}^{N^{\mathrm{c}}} \sum_{k=1}^{N_i^{\mathrm{or}}} \sum_{j=1}^{N^{\mathrm{c}}} \sum_{l=1}^{N_j^{\mathrm{or}}} & \left(\frac{1}{2} N x_i z_i p_{i,k}^{\mathrm{R}} \frac{x_j z_j p_{j,l}^{\mathrm{R}}}{\sum_{m=1}^{N^{\mathrm{c}}} x_m z_m} \ln \left(\frac{x_j z_j p_{j,l}^{\mathrm{R}}}{\sum_{m=1}^{N^{\mathrm{c}}} x_m z_m} \right) - \frac{1}{2} N x_i z_i p_{i,k}^{\mathrm{R}} \psi_{j,l,i,k} \ln \left(\psi_{j,l,i,k} \right) \right) + \\ \frac{k_{\mathrm{B}}}{n} \sum_{i=1}^{N^{\mathrm{c}}} \sum_{k=1}^{N_i^{\mathrm{or}}} \sum_{j=1}^{N^{\mathrm{c}}} \sum_{l=1}^{N_j^{\mathrm{or}}} & \left(\left(\frac{1}{2} N x_i z_i p_{i,k}^{\mathrm{R}} \frac{x_j z_j p_{j,l}^{\mathrm{R}}}{\sum_{m=1}^{N^{\mathrm{c}}} x_m z_m} - \frac{1}{2} N x_i z_i p_{i,k}^{\mathrm{R}} \psi_{j,l,i,k} \right) \ln \left(\frac{1}{2} N x_i z_i p_{i,k}^{\mathrm{R}} \right) \right) \end{aligned} \quad (7.34)$$

Again applying Eqs. 4.1 and 4.3 results in

$$\begin{aligned} S = S_{\mathrm{comb}} + & \frac{k_{\mathrm{B}}}{n} \sum_{j=1}^{N^{\mathrm{c}}} \sum_{l=1}^{N_j^{\mathrm{or}}} \left(\frac{1}{2} N x_j z_j p_{j,l}^{\mathrm{R}} \ln \left(\frac{x_j z_j p_{j,l}^{\mathrm{R}}}{\sum_{m=1}^{N^{\mathrm{c}}} x_m z_m} \right) \right) - \\ & \frac{k_{\mathrm{B}}}{n} \sum_{i=1}^{N^{\mathrm{c}}} \sum_{k=1}^{N_i^{\mathrm{or}}} \sum_{j=1}^{N^{\mathrm{c}}} \sum_{l=1}^{N_j^{\mathrm{or}}} \left(\frac{1}{2} N x_i z_i p_{i,k}^{\mathrm{R}} \psi_{j,l,i,k} \ln \left(\psi_{j,l,i,k} \right) \right) + \\ & \frac{k_{\mathrm{B}}}{n} \sum_{i=1}^{N^{\mathrm{c}}} \sum_{k=1}^{N_i^{\mathrm{or}}} \left(\left(\frac{1}{2} N x_i z_i p_{i,k}^{\mathrm{R}} - \frac{1}{2} N x_i z_i p_{i,k}^{\mathrm{R}} \right) \ln \left(\frac{1}{2} N x_i z_i p_{i,k}^{\mathrm{R}} \right) \right) \end{aligned} \quad (7.35)$$

$\Leftrightarrow$

$$\begin{aligned} S = S_{\mathrm{comb}} + & \frac{k_{\mathrm{B}}}{n} \sum_{j=1}^{N^{\mathrm{c}}} \sum_{l=1}^{N_j^{\mathrm{or}}} \left(\frac{1}{2} N x_j z_j p_{j,l}^{\mathrm{R}} \ln \left(\frac{x_j z_j p_{j,l}^{\mathrm{R}}}{\sum_{m=1}^{N^{\mathrm{c}}} x_m z_m} \right) \right) - \\ & \frac{k_{\mathrm{B}}}{n} \sum_{i=1}^{N^{\mathrm{c}}} \sum_{k=1}^{N_i^{\mathrm{or}}} \sum_{j=1}^{N^{\mathrm{c}}} \sum_{l=1}^{N_j^{\mathrm{or}}} \left(\frac{1}{2} N x_i z_i p_{i,k}^{\mathrm{R}} \psi_{j,l,i,k} \ln \left(\psi_{j,l,i,k} \right) \right) \end{aligned} \quad (7.36)$$

The second term on the right-hand side of Eq. 7.36 is expanded with Eq. 4.3

$$\begin{aligned} S = S_{\mathrm{comb}} + & \frac{k_{\mathrm{B}}}{n} \sum_{i=1}^{N^{\mathrm{c}}} \sum_{k=1}^{N_i^{\mathrm{or}}} \sum_{j=1}^{N^{\mathrm{c}}} \sum_{l=1}^{N_j^{\mathrm{or}}} \left(\frac{1}{2} N x_j z_j p_{j,l}^{\mathrm{R}} \psi_{i,k,j,l} \ln \left(\frac{x_j z_j p_{j,l}^{\mathrm{R}}}{\sum_{m=1}^{N^{\mathrm{c}}} x_m z_m} \right) \right) - \\ & \frac{k_{\mathrm{B}}}{n} \sum_{i=1}^{N^{\mathrm{c}}} \sum_{k=1}^{N_i^{\mathrm{or}}} \sum_{j=1}^{N^{\mathrm{c}}} \sum_{l=1}^{N_j^{\mathrm{or}}} \left(\frac{1}{2} N x_i z_i p_{i,k}^{\mathrm{R}} \psi_{j,l,i,k} \ln \left(\psi_{j,l,i,k} \right) \right) \end{aligned} \quad (7.37)$$

and the symmetry condition, Eq. 4.4, is applied so Eq. 7.37 results in

$$S = S_{\text{comb}} + \frac{k_{\text{B}}}{n} \sum_{i=1}^{N^{\text{c}}} \sum_{k=1}^{N_i^{\text{or}}} \sum_{j=1}^{N^{\text{c}}} \sum_{l=1}^{N_j^{\text{or}}} \left(\frac{1}{2} N x_i z_i p_{i,k}^{\text{R}} \psi_{j,l,i,k} \ln \left(\frac{x_j z_j p_{j,l}^{\text{R}}}{\sum_{m=1}^{N^{\text{c}}} x_m z_m} \right) \right) - \frac{k_{\text{B}}}{n} \sum_{i=1}^{N^{\text{c}}} \sum_{k=1}^{N_i^{\text{or}}} \sum_{j=1}^{N^{\text{c}}} \sum_{l=1}^{N_j^{\text{or}}} \left(\frac{1}{2} N x_i z_i p_{i,k}^{\text{R}} \psi_{j,l,i,k} \ln \left(\psi_{j,l,i,k} \right) \right) \tag{7.38}$$

$\Leftrightarrow$

$$S = S_{\text{comb}} + \frac{k_{\text{B}}}{n} \sum_{i=1}^{N^{\text{c}}} \sum_{k=1}^{N_i^{\text{or}}} \sum_{j=1}^{N^{\text{c}}} \sum_{l=1}^{N_j^{\text{or}}} \left(\frac{1}{2} N x_i z_i p_{i,k}^{\text{R}} \psi_{j,l,i,k} \ln \left(\frac{x_j z_j p_{j,l}^{\text{R}}}{\psi_{j,l,i,k} \sum_{m=1}^{N^{\text{c}}} x_m z_m} \right) \right) \tag{7.39}$$

Using the fact that n and N

$$n = \frac{N}{N_{\text{A}}} \tag{7.40}$$

as well as k_{B} and R are related by the Avogadro constant

$$k_{\text{B}} = \frac{R}{N_{\text{A}}} \tag{7.41}$$

Eq. 7.39 can be written as

$$S = S_{\text{comb}} + R \sum_{i=1}^{N^{\text{c}}} \sum_{k=1}^{N_i^{\text{or}}} \sum_{j=1}^{N^{\text{c}}} \sum_{l=1}^{N_j^{\text{or}}} \left(\frac{1}{2} x_i z_i p_{i,k}^{\text{R}} \psi_{j,l,i,k} \ln \left(\frac{x_j z_j p_{j,l}^{\text{R}}}{\psi_{j,l,i,k} \sum_{m=1}^{N^{\text{c}}} x_m z_m} \right) \right) \tag{7.42}$$

Eq. 7.42 is equal to Eq. 4.17.

7.4 Derivation of the MOQUAC model equation

In this section Eq. 4.23 is derived. To simplify the derivation, the reference state

$$G_{\text{ref}} = H_{\text{ref}} \tag{7.43}$$

with

$$H_{\text{ref}} = \sum_{i=1}^{N^{\text{c}}} \sum_{k=1}^{N_i^{\text{or}}} \frac{1}{2} z_i x_i p_{i,k}^{\text{R}} u_{i,k,i,k} \tag{7.44}$$

is introduced, which is subtracted from Eq. 4.21:

$$G^* = G - G_{\text{ref}} \tag{7.45}$$

$\Leftrightarrow$

$$G^* = \sum_{i=1}^{N^c}\sum_{k=1}^{N_i^{or}}\sum_{j=1}^{N^c}\sum_{l=1}^{N_j^{or}}\frac{1}{2}z_i x_i p_{i,k}^R \psi_{j,l,i,k}\left(u_{j,l,i,k} + TR\ln\frac{\psi_{j,l,i,k}\sum_{q=1}^{N^c} x_q z_q}{p_{j,l}^R x_j z_j}\right) - TS_{comb} - \sum_{i=1}^{N^c}\sum_{k=1}^{N_i^{or}}\frac{1}{2}z_i x_i p_{i,k}^R u_{i,k,i,k} \quad (7.46)$$

Note that introducing this reference state neither influences the result for the excess Gibbs energy nor for the local composition, since the reference state is independent of $\psi_{n,p,m,o}$.

Expanding the last term in Eq. 7.46 with Eq. 4.3 allows to rewrite Eq. 7.46 to

$$G^* = \sum_{i=1}^{N^c}\sum_{k=1}^{N_i^{or}}\sum_{j=1}^{N^c}\sum_{l=1}^{N_j^{or}}\frac{1}{2}z_i x_i p_{i,k}^R \psi_{j,l,i,k}\left(u_{j,l,i,k} - u_{i,k,i,k}\right) + \sum_{i=1}^{N^c}\sum_{k=1}^{N_i^{or}}\sum_{j=1}^{N^c}\sum_{l=1}^{N_j^{or}}\frac{1}{2}RT z_i x_i p_{i,k}^R \psi_{j,l,i,k}\ln\frac{\psi_{j,l,i,k}\sum_{q=1}^{N^c} x_q z_q}{p_{j,l}^R x_j z_j} - TS_{comb} \quad (7.47)$$

With Eqs. 4.4 and 4.2, Eq. 7.47 can be written as

$$G^* = \sum_{i=1}^{N^c}\sum_{k=1}^{N_i^{or}}\sum_{j\geq i}^{N^c}\sum_{\substack{l=1\\ l>k,\\ \text{if } j=i}}^{N_j^{or}}\frac{1}{2}z_i x_i p_{i,k}^R \psi_{j,l,i,k}\left(u_{j,l,i,k} + u_{i,k,j,l} - u_{j,l,j,l} - u_{i,k,i,k}\right) + \sum_{i=1}^{N^c}\sum_{k=1}^{N_i^{or}}\sum_{j=1}^{N^c}\sum_{l=1}^{N_j^{or}}\frac{1}{2}RT z_i x_i p_{i,k}^R \psi_{j,l,i,k}\ln\frac{\psi_{j,l,i,k}\sum_{q=1}^{N^c} x_q z_q}{p_{j,l}^R x_j z_j} - TS_{comb} \quad (7.48)$$

Substituting

$$\omega_{j,l,i,k} = u_{j,l,i,k} + u_{i,k,j,l} - u_{j,l,j,l} - u_{i,k,i,k} \quad (7.49)$$

into Eq. 7.48 results in:

$$G^* = \sum_{i=1}^{N^c}\sum_{k=1}^{N_i^{or}}\sum_{j\geq i}^{N^c}\sum_{\substack{l=1\\ l>k,\\ \text{if } j=i}}^{N_j^{or}}\frac{1}{2}z_i x_i p_{i,k}^R \psi_{j,l,i,k}\omega_{j,l,i,k} + \sum_{i=1}^{N^c}\sum_{k=1}^{N_i^{or}}\sum_{j=1}^{N^c}\sum_{l=1}^{N_j^{or}}\frac{1}{2}RT z_i x_i p_{i,k}^R \psi_{j,l,i,k}\ln\frac{\psi_{j,l,i,k}\sum_{q=1}^{N^c} x_q z_q}{p_{j,l}^R x_j z_j} - TS_{comb} \quad (7.50)$$

Splitting Eq. 7.50 up in an enthalpic and an entropic contribution according to Eq. 4.7 results in

$$H^* = \sum_{i=1}^{N^c} \sum_{k=1}^{N_i^{or}} \sum_{j\geq i}^{N^c} \sum_{\substack{l=1 \\ l>k, \\ \text{if } j=i}}^{N_j^{or}} \frac{1}{2} z_i x_i p_{i,k}^R \psi_{j,l,i,k} \omega_{j,l,i,k} \tag{7.51}$$

and

$$S^* = \sum_{i=1}^{N^c} \sum_{k=1}^{N_i^{or}} \sum_{j=1}^{N^c} \sum_{l=1}^{N_j^{or}} \frac{1}{2} RT z_i x_i p_{i,k}^R \psi_{j,l,i,k} \ln \frac{\psi_{j,l,i,k} \sum_{q=1}^{N^c} x_q z_q}{p_{j,l}^R x_j z_j} - T S_{comb} \tag{7.52}$$

Now Eq. 7.50 is partially derived with respect to $\psi_{n,p,m,o}$:

$$\left(\frac{\partial G^*}{\partial \psi_{n,p,m,o}} \right)_{T,x_i,\psi_{j,l,i,k} \neq \psi_{n,p,m,o}} = \underbrace{\sum_{i=1}^{N^c} \sum_{k=1}^{N_i^{or}} \sum_{j\geq i}^{N^c} \sum_{\substack{l=1 \\ l>k, \\ \text{if } j=i}}^{N_j^{or}} \frac{1}{2} z_i x_i p_{i,k}^R \frac{\partial \psi_{j,l,i,k}}{\partial \psi_{n,p,m,o}} \omega_{j,l,i,k}}_{a} + \sum_{i=1}^{N^c} \sum_{k=1}^{N_i^{or}} \sum_{j=1}^{N^c} \sum_{l=1}^{N_j^{or}} \frac{1}{2} RT z_i x_i p_{i,k}^R \frac{\partial \psi_{j,l,i,k}}{\partial \psi_{n,p,m,o}} \left(\ln \frac{\psi_{j,l,i,k} \sum_{q=1}^{N^c} x_q z_q}{p_{j,l}^R x_j z_j} + 1 \right) \tag{7.53}$$

The partial derivation of Eq. 4.4 results in

$$z_i x_i p_{i,k}^R \frac{\partial \psi_{j,l,i,k}}{\partial \psi_{n,p,m,o}} = z_j x_j p_{j,l}^R \frac{\partial \psi_{i,k,j,l}}{\partial \psi_{n,p,m,o}} \tag{7.54}$$

Substituting Eq. 7.54 into Eq. 7.53 results in

$$\left(\frac{\partial G^*}{\partial \psi_{n,p,m,o}} \right)_{T,x_i,\psi_{j,l,i,k} \neq \psi_{n,p,m,o}} = a + \underbrace{\sum_{i=1}^{N^c} \sum_{k=1}^{N_i^{or}} \sum_{j\geq i}^{N^c} \sum_{\substack{l=1 \\ l>k, \\ \text{if } j=i}}^{N_j^{or}} \frac{1}{2} RT z_i x_i p_{i,k}^R \frac{\partial \psi_{j,l,i,k}}{\partial \psi_{n,p,m,o}} \ln \frac{\psi_{j,l,i,k} \psi_{i,k,j,l} \left(\sum_{q=1}^{N^c} x_q z_q \right)^2}{p_{j,l}^R x_j z_j p_{i,k}^R x_i z_i}}_{b} + \sum_{i=1}^{N^c} \sum_{k=1}^{N_i^{or}} \frac{1}{2} RT z_i x_i p_{i,k}^R \frac{\partial \psi_{i,k,i,k}}{\partial \psi_{n,p,m,o}} \ln \frac{\psi_{i,k,i,k} \sum_{q=1}^{N^c} x_q z_q}{p_{i,k}^R x_i z_i} + \underbrace{\sum_{i=1}^{N^c} \sum_{k=1}^{N_i^{or}} \sum_{j=1}^{N^c} \sum_{l=1}^{N_j^{or}} \frac{1}{2} RT z_i x_i p_{i,k}^R \frac{\partial \psi_{j,l,i,k}}{\partial \psi_{n,p,m,o}}}_{c} \tag{7.55}$$

The partial derivation of Eq. 4.3 results in

$$\sum_{j=1}^{N^{\mathrm{c}}} \sum_{l=1}^{N_j^{\mathrm{or}}} \frac{\partial \psi_{j,l,i,k}}{\partial \psi_{n,p,m,o}} = 0 \tag{7.56}$$

With Eq. 7.56, c equals zero and Eq. 7.55 can be written as

$$\left(\frac{\partial G^*}{\partial \psi_{n,p,m,o}}\right)_{T,x_i,\psi_{j,l,i,k}\neq\psi_{n,p,m,o}} = a+b- \sum_{i=1}^{N^{\mathrm{c}}} \sum_{k=1}^{N_i^{\mathrm{or}}} \sum_{j=1}^{N^{\mathrm{c}}} \sum_{\substack{l=1 \\ l\neq k, \\ \text{if } j=i}}^{N_j^{\mathrm{or}}} \frac{1}{2} RT z_i x_i p_{i,k}^{\mathrm{R}} \frac{\partial \psi_{j,l,i,k}}{\partial \psi_{n,p,m,o}} \ln \frac{\psi_{i,k,i,k} \sum_{q=1}^{N^{\mathrm{c}}} x_q z_q}{p_{i,k}^{\mathrm{R}} x_i z_i} \tag{7.57}$$

Splitting up the terms of the sum results in

$$\left(\frac{\partial G^*}{\partial \psi_{n,p,m,o}}\right)_{T,x_i,\psi_{j,l,i,k}\neq\psi_{n,p,m,o}} = a+b- \underbrace{\sum_{i=1}^{N^{\mathrm{c}}} \sum_{k=1}^{N_i^{\mathrm{or}}} \sum_{j\geq i}^{N^{\mathrm{c}}} \sum_{\substack{l=1 \\ l>k, \\ \text{if } j=i}}^{N_j^{\mathrm{or}}} \frac{1}{2} RT z_i x_i p_{i,k}^{\mathrm{R}} \frac{\partial \psi_{j,l,i,k}}{\partial \psi_{n,p,m,o}} \ln \frac{\psi_{i,k,i,k} \sum_{q=1}^{N^{\mathrm{c}}} x_q z_q}{p_{i,k}^{\mathrm{R}} x_i z_i}}_{d} - \sum_{i=1}^{N^{\mathrm{c}}} \sum_{k=1}^{N_i^{\mathrm{or}}} \sum_{j\leq i}^{N^{\mathrm{c}}} \sum_{\substack{l=1 \\ l<k, \\ \text{if } j=i}}^{N_j^{\mathrm{or}}} \frac{1}{2} RT z_i x_i p_{i,k}^{\mathrm{R}} \frac{\partial \psi_{j,l,i,k}}{\partial \psi_{n,p,m,o}} \ln \frac{\psi_{i,k,i,k} \sum_{q=1}^{N^{\mathrm{c}}} x_q z_q}{p_{i,k}^{\mathrm{R}} x_i z_i} \tag{7.58}$$

$\Leftrightarrow$

$$\left(\frac{\partial G^*}{\partial \psi_{n,p,m,o}}\right)_{T,x_i,\psi_{j,l,i,k}\neq\psi_{n,p,m,o}} = a+b-d- \sum_{i=1}^{N^{\mathrm{c}}} \sum_{k=1}^{N_i^{\mathrm{or}}} \sum_{j\leq i}^{N^{\mathrm{c}}} \sum_{\substack{l=1 \\ l<k, \\ \text{if } j=i}}^{N_j^{\mathrm{or}}} \frac{1}{2} RT z_i x_i p_{i,k}^{\mathrm{R}} \frac{\partial \psi_{j,l,i,k}}{\partial \psi_{n,p,m,o}} \ln \frac{\psi_{i,k,i,k} \sum_{q=1}^{N^{\mathrm{c}}} x_q z_q}{p_{i,k}^{\mathrm{R}} x_i z_i} \tag{7.59}$$

With Eq. 7.54, Eq. 7.59 can be written as

$$\left(\frac{\partial G^*}{\partial \psi_{n,p,m,o}}\right)_{T,x_i,\psi_{j,l,i,k}\neq\psi_{n,p,m,o}} = a+b-d- \sum_{i=1}^{N^{\mathrm{c}}} \sum_{k=1}^{N_i^{\mathrm{or}}} \sum_{j\geq i}^{N^{\mathrm{c}}} \sum_{\substack{l=1 \\ l>k, \\ \text{if } j=i}}^{N_j^{\mathrm{or}}} \frac{1}{2} RT z_i x_i p_{i,k}^{\mathrm{R}} \frac{\partial \psi_{j,l,i,k}}{\partial \psi_{n,p,m,o}} \ln \frac{\psi_{j,l,j,l} \sum_{q=1}^{N^{\mathrm{c}}} x_q z_q}{p_{j,l}^{\mathrm{R}} x_j z_j} \tag{7.60}$$

Substituting again a, b and d and uniting terms results in

$$\left(\frac{\partial G^*}{\partial \psi_{n,p,m,o}}\right)_{T,x_i,\psi_{j,l,i,k}\neq\psi_{n,p,m,o}} =$$

$$\sum_{i=1}^{N^c}\sum_{k=1}^{N_i^{or}}\sum_{j\geq i}^{N^c}\sum_{\substack{l=1\\ l>k,\\ \text{if } j=i}}^{N_j^{or}} \frac{1}{2} z_i x_i p_{i,k}^R \frac{\partial \psi_{j,l,i,k}}{\partial \psi_{n,p,m,o}} \left(\omega_{j,l,i,k} + RT \ln \frac{\psi_{j,l,i,k}\psi_{i,k,j,l}\left(\sum_{q=1}^{N^c} x_q z_q\right)^2}{p_{j,l}^R x_j z_j p_{i,k}^R x_i z_i} - \right.$$

$$\left. RT \ln \frac{\psi_{i,k,i,k}\sum_{q=1}^{N^c} x_q z_q}{p_{i,k}^R x_i z_i} - RT \ln \frac{\psi_{j,l,j,l}\sum_{q=1}^{N^c} x_q z_q}{p_{j,l}^R x_j z_j} \right) \quad (7.61)$$

$\Leftrightarrow$

$$\left(\frac{\partial G^*}{\partial \psi_{n,p,m,o}}\right)_{T,x_i,\psi_{j,l,i,k}\neq\psi_{n,p,m,o}} =$$

$$\sum_{i=1}^{N^c}\sum_{k=1}^{N_i^{or}}\sum_{j\geq i}^{N^c}\sum_{\substack{l=1\\ l>k,\\ \text{if } j=i}}^{N_j^{or}} \frac{1}{2} z_i x_i p_{i,k}^R \frac{\partial \psi_{j,l,i,k}}{\partial \psi_{n,p,m,o}} \left(\omega_{j,l,i,k} + RT \ln \frac{\psi_{j,l,i,k}\psi_{i,k,j,l}}{\psi_{i,k,i,k}\psi_{j,l,j,l}} \right) \quad (7.62)$$

Substituting Eq. 7.62 into Eq. 4.22 results in

$$\sum_{i=1}^{N^c}\sum_{k=1}^{N_i^{or}}\sum_{j\geq i}^{N^c}\sum_{\substack{l=1\\ l>k,\\ \text{if } j=i}}^{N_j^{or}} \frac{1}{2} z_i x_i p_{i,k}^R \frac{\partial \psi_{j,l,i,k}}{\partial \psi_{n,p,m,o}} \left(\omega_{j,l,i,k} + RT \ln \frac{\psi_{j,l,i,k}\psi_{i,k,j,l}}{\psi_{i,k,i,k}\psi_{j,l,j,l}} \right) = 0 \quad (7.63)$$

Since the partial derivatives in Eq. 7.63 are independent of each other, each bracketed term needs to be equal to zero, thus

$$\frac{\psi_{j,l,i,k}\psi_{i,k,j,l}}{\psi_{i,k,i,k}\psi_{j,l,j,l}} = \exp\left(-\frac{\omega_{j,l,i,k}}{RT}\right) \quad (7.64)$$

must be fulfilled. Substituting $\omega_{j,l,i,k}$ in Eq. 7.64 leads to

$$\frac{\psi_{j,l,i,k}\psi_{i,k,j,l}}{\psi_{i,k,i,k}\psi_{j,l,j,l}} = \exp\left(-\frac{u_{j,l,i,k} + u_{i,k,j,l} - u_{i,k,i,k} - u_{j,l,j,l}}{RT}\right) \quad (7.65)$$

which is equal to Eq. 4.23.

7.5 Proof of the consistency of MOQUAC with the Gibbs-Helmholtz equation

In this section the consistency of the enthalpy term, Eq. 4.8, and the entropy term, Eq. 4.17, of MOQUAC with the Gibbs-Helmholtz equation is shown. Since the reference state of Eqs. 7.43 and 7.44 obviously fulfills the Gibbs-Helmholtz equation, it is sufficient to show that H^* and S^*

are consistent with the Gibbs-Helmholtz equation. For this Eqs. 7.51 and 7.52 are substituted into the Gibbs-Helmholtz equation as given by Eq. 4.20:

$$H^* = \sum_{i=1}^{N^c}\sum_{k=1}^{N_i^{or}}\sum_{j\geq i}^{N^c}\sum_{\substack{l=1\\ l>k,\\ \text{if } j=i}}^{N_j^{or}} \frac{1}{2} z_i x_i p_{i,k}^R \left(\frac{\partial \psi_{j,l,i,k}}{\partial \frac{1}{T}} \frac{\omega_{j,l,i,k}}{T} + \psi_{j,l,i,k}\omega_{j,l,i,k} \right) + \sum_{i=1}^{N^c}\sum_{k=1}^{N_i^{or}}\sum_{j=1}^{N^c}\sum_{l=1}^{N_j^{or}} \frac{1}{2} R z_i x_i p_{i,k}^R \frac{\partial \psi_{j,l,i,k}}{\partial \frac{1}{T}} \left(\ln \frac{\psi_{j,l,i,k}\sum_{q=1}^{N^c} x_q z_q}{p_{j,l}^R x_j z_j} + 1 \right) \quad (7.66)$$

$\Leftrightarrow$

$$\begin{aligned} H^* = & \underbrace{\left[\sum_{i=1}^{N^c}\sum_{k=1}^{N_i^{or}}\sum_{j\geq i}^{N^c}\sum_{\substack{l=1\\ l>k,\\ \text{if } j=i}}^{N_j^{or}} \frac{1}{2} z_i x_i p_{i,k}^R \frac{\partial \psi_{j,l,i,k}}{\partial \psi_{n,p,m,o}} \omega_{j,l,i,k} \right]}_{e} \frac{1}{T} \frac{\partial \psi_{n,p,m,o}}{\partial \frac{1}{T}} + \\ & \underbrace{\left[\sum_{i=1}^{N^c}\sum_{k=1}^{N_i^{or}}\sum_{j=1}^{N^c}\sum_{l=1}^{N_j^{or}} \frac{1}{2} RT z_i x_i p_{i,k}^R \frac{\partial \psi_{j,l,i,k}}{\partial \psi_{n,p,m,o}} \ln \frac{\psi_{j,l,i,k}\sum_{q=1}^{N^c} x_q z_q}{p_{j,l}^R x_j z_j} \right]}_{f} \frac{1}{T} \frac{\partial \psi_{n,p,m,o}}{\partial \frac{1}{T}} + \\ & \underbrace{\left[\sum_{i=1}^{N^c}\sum_{k=1}^{N_i^{or}}\sum_{j=1}^{N^c}\sum_{l=1}^{N_j^{or}} \frac{1}{2} RT z_i x_i p_{i,k}^R \frac{\partial \psi_{j,l,i,k}}{\partial \psi_{n,p,m,o}} \right]}_{g} \frac{1}{T} \frac{\partial \psi_{n,p,m,o}}{\partial \frac{1}{T}} + \\ & \sum_{i=1}^{N^c}\sum_{k=1}^{N_i^{or}}\sum_{j\geq i}^{N^c}\sum_{\substack{l=1\\ l>k,\\ \text{if } j=i}}^{N_j^{or}} \frac{1}{2} z_i x_i p_{i,k}^R \psi_{j,l,i,k}\omega_{j,l,i,k} \end{aligned} \quad (7.67)$$

It is

$$e + f + g = \left(\frac{\partial G^*}{\partial \psi_{n,p,m,o}} \right)_{T,x_i,\psi_{j,l,i,k}\neq\psi_{n,p,m,o}} \quad (7.68)$$

which according to Eq. 4.22 and as discussed in appendix 7.4 is equal to zero:

$$e + f + g = 0 \quad (7.69)$$

Eq. 7.67 can thus be simplified to

$$H^* = \sum_{i=1}^{N^c}\sum_{k=1}^{N_i^{or}}\sum_{j\geq i}^{N^c}\sum_{\substack{l=1\\ l>k,\\ \text{if } j=i}}^{N_j^{or}} \frac{1}{2} z_i x_i p_{i,k}^R \psi_{j,l,i,k}\omega_{j,l,i,k} \quad (7.70)$$

which is identical to H^* as given in Eq. 7.51. The consistency of MOQUAC with the Gibbs-Helmholtz equation is thus shown.

7.6 Proof that in MOQUAC equally behaving orientations can be merged

In this section it is shown that in MOQUAC two equally behaving orientations of one component can be merged. Two orientations i,k and i,m of one component are considered to be equally behaving if for all $j \in [1,\ldots,N^{\mathrm{c}}]$ and $l \in [1,\ldots,N_j^{\mathrm{or}}]$

$$u_{j,l,i,k} = u_{j,l,i,m} \tag{7.71}$$

and

$$u_{i,k,i,k} = u_{i,m,i,m} \tag{7.72}$$

applies.

To prove that two equally behaving orientations i,k and i,m can be merged, the ratio $\psi_{j,l,i,k}/\psi_{j,l,i,m}$ is determined. For this first with Eq. 4.4, Eq. 7.65 is written as

$$\frac{x_i z_i p_{i,k}^{\mathrm{R}} \psi_{j,l,i,k} \psi_{j,l,i,k}}{x_j z_j p_{j,l}^{\mathrm{R}} \psi_{i,k,i,k} \psi_{j,l,j,l}} = \exp\left(-\frac{u_{j,l,i,k} + u_{i,k,j,l} - u_{i,k,i,k} - u_{j,l,j,l}}{RT}\right) \tag{7.73}$$

$\Leftrightarrow$

$$\psi_{j,l,i,k} = \sqrt{\frac{x_j z_j p_{j,l}^{\mathrm{R}} \psi_{i,k,i,k} \psi_{j,l,j,l}}{x_i z_i p_{i,k}^{\mathrm{R}}}} \exp\left(-\frac{u_{j,l,i,k} + u_{i,k,j,l} - u_{i,k,i,k} - u_{j,l,j,l}}{2RT}\right) \tag{7.74}$$

which, when substituting Eq. 4.2, results in

$$\psi_{j,l,i,k} = \sqrt{\frac{x_j z_j p_{j,l}^{\mathrm{R}} \psi_{i,k,i,k} \psi_{j,l,j,l}}{x_i z_i p_{i,k}^{\mathrm{R}}}} \exp\left(-\frac{2u_{j,l,i,k} - u_{i,k,i,k} - u_{j,l,j,l}}{2RT}\right) \tag{7.75}$$

With Eq. 7.75, for the ratio

$$\frac{\psi_{j,l,i,k}}{\psi_{j,l,i,m}} = \sqrt{\frac{\psi_{i,k,i,k} p_{i,m}^{\mathrm{R}}}{\psi_{i,m,i,m} p_{i,k}^{\mathrm{R}}}} \exp\left(-\frac{2u_{j,l,i,k} - 2u_{j,l,i,m} - u_{i,k,i,k} + u_{i,m,i,m}}{2RT}\right) \tag{7.76}$$

results.

With Eqs. 7.71 and 7.72, Eq. 7.76 simplifies to

$$\frac{\psi_{j,l,i,k}}{\psi_{j,l,i,m}} = \sqrt{\frac{\psi_{i,k,i,k} p_{i,m}^{\mathrm{R}}}{\psi_{i,m,i,m} p_{i,k}^{\mathrm{R}}}} \tag{7.77}$$

Now equation 7.74 is substituted into Eq. 4.3 to yield

$$\sum_{j=1}^{N^{\mathrm{c}}}\sum_{l=1}^{N_j^{\mathrm{or}}}\sqrt{\frac{z_j x_j p_{j,l}^{\mathrm{R}}\psi_{i,k,i,k}\psi_{j,l,j,l}}{z_i x_i p_{i,k}^{\mathrm{R}}}}\exp\left(-\frac{2u_{j,l,i,k}-u_{j,l,j,l}-u_{i,k,i,k}}{2RT}\right)=1 \tag{7.78}$$

$\Leftrightarrow$

$$\sum_{j=1}^{N^{\mathrm{c}}}\sum_{l=1}^{N_j^{\mathrm{or}}}\sqrt{\frac{\psi_{i,k,i,k}}{p_{i,k}^{\mathrm{R}}}}\sqrt{\frac{z_j x_j p_{j,l}^{\mathrm{R}}\psi_{j,l,j,l}}{z_i x_i}}\exp\left(-\frac{2u_{j,l,i,k}-u_{j,l,j,l}-u_{i,k,i,k}}{2RT}\right)=1 \tag{7.79}$$

$\Leftrightarrow$

$$\sqrt{\frac{\psi_{i,k,i,k}}{p_{i,k}^{\mathrm{R}}}}=\frac{1}{\sum_{j=1}^{N^{\mathrm{c}}}\sum_{l=1}^{N_j^{\mathrm{or}}}\sqrt{\frac{z_j x_j p_{j,l}^{\mathrm{R}}\psi_{j,l,j,l}}{z_i x_i}}\exp\left(-\frac{2u_{j,l,i,k}-u_{j,l,j,l}-u_{i,k,i,k}}{2RT}\right)}} \tag{7.80}$$

With Eqs. 7.71 and 7.72, from Eq. 7.80

$$\sqrt{\frac{\psi_{i,k,i,k}}{p_{i,k}^{\mathrm{R}}}}=\sqrt{\frac{\psi_{i,m,i,m}}{p_{i,m}^{\mathrm{R}}}} \tag{7.81}$$

can be derived. With Eq. 7.81, Eq. 7.77 can be simplified to

$$\psi_{j,l,i,k}=\psi_{j,l,i,m} \tag{7.82}$$

All contact fractions of the orientation i,k and i,m are thus equal. With this result, Eq. 4.21 is evaluated. It can be seen that, because Eqs. 7.71 and 7.82 apply, the orientations i,k and i,m can be merged to a new orientation i,k'. The random frequency of occurrence of this new orientation $p_{i,k'}^{\mathrm{R}}$ is given by

$$p_{i,k'}^{\mathrm{R}}=p_{i,k}^{\mathrm{R}}+p_{i,m}^{\mathrm{R}} \tag{7.83}$$

7.6.1 Example of merging of orientations

A binary system of cubic molecules is considered. Each face of a molecule has 4 surface segments of equal size. Figure 4.7 shows a flat projection of the general form of these component types. Several surface segments of component 1 have a surface charge that is specified in Fig. 7.2. The surface segments of component 2 are all equal 0. It is assumed that the interaction energy between two contacting surface segments is given by the product of the charges of both surface segments.

For each component, 24 orientations are distinguished: 6 directions of orientation, given by the normal of each face, and 4 rotations around each direction of orientation. When two molecules are in contact, always the two faces of the molecules completely overlap. Upon a contact the four surface segments of the contacting face always interact with one other surface segment. Since a binary mixture is considered, in total there are 48 times 48 different contact types, where the interaction energies of all contact types that involve at least one component 2 are equal to zero. Before the merging of equally behaving orientations, the random frequency of occurrence of each orientation equals $1/24$.

Table 7.4 shows the interaction energy of all contact types that involve two components 1. For the remaining contact types it is indicated that the interaction energy is equal 0. It is obvious that

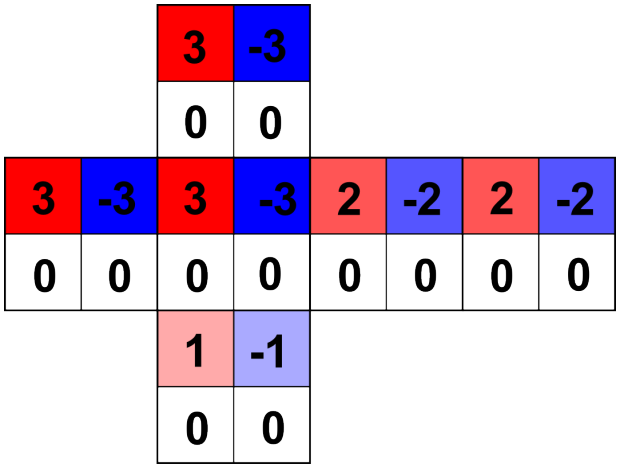


Figure 7.2: Flat projection of component 1 for the example of merging of orientations.

Tab. 7.4 is symmetric, because Eq. 4.2 applies. Having a closer look at Tab. 7.4 reveals that also all 4-by-4 sub-matrices show a characteristic structure. In each diagonal from top right to bottom left the interaction energies are identical.

According to Eqs. 7.71 and 7.72, two orientations are equally behaving, if the two lines that belong to the orientations are identical. Because of symmetry also the two columns that belong to the orientations are identical. If this is the case, the orientations can be merged. This is equivalent to merging the equal lines and columns in the interaction-energy matrix. Tab. 7.4 shows that orientation 1 is equal to orientations 5 and 9, 2 is equal to 6 and 10, 3 is equal to 7 and 11, 4 is equal to 8 and 12, 13 is equal to 17, 14 is equal to 18, 15 is equal to 19 and 20 is equal to 24. After merging the related lines and columns, the new interaction matrix of Tab. 7.3 results. The random frequencies of occurrence of the new orientations are given in Tab. 7.5.

Table 7.3: Interaction-energy matrix after the merging of similar orientations.

i		1	1	1	1	1	1	1	1	1	1	1	1	2
	k	1	2	3	4	5	6	7	8	9	10	11	12	1
1	1	-18	9	0	9	-12	6	0	6	-6	3	0	3	0
1	2	9	0	9	-18	6	0	6	-12	3	0	3	-6	0
1	3	0	9	-18	9	0	6	-12	6	0	3	-6	3	0
1	4	9	-18	9	0	6	-12	6	0	3	-6	3	0	0
1	13	-12	6	0	6	-8	4	0	4	-4	2	0	2	0
1	14	6	0	6	-12	4	0	4	-8	2	0	2	-4	0
1	15	0	6	-12	6	0	4	-8	4	0	2	-4	2	0
1	16	6	-12	6	0	4	-8	4	0	2	-4	2	0	0
1	21	-6	3	0	3	-4	2	0	2	-2	1	0	1	0
1	22	3	0	3	-6	2	0	2	-4	1	0	1	-2	0
1	23	0	3	-6	3	0	2	-4	2	0	1	-2	1	0
1	24	3	-6	3	0	2	-4	2	0	1	-2	1	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 7.4: Interaction-energy matrix before the merging of equally behaving orientations.

i		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2
	k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	1
1	1	-18	9	0	9	-18	9	0	9	-18	9	0	9	-12	6	0	6	-12	6	0	6	-6	3	0	3	0
1	2	9	0	9	-18	9	0	9	-18	9	0	9	-18	6	0	6	-12	6	0	6	-12	3	0	3	-6	0
1	3	0	9	-18	9	0	9	-18	9	0	9	-18	9	0	6	-12	6	0	6	-12	6	0	3	-6	3	0
1	4	9	-18	9	0	9	-18	9	0	9	-18	9	0	6	-12	6	0	6	-12	6	0	3	-6	3	0	0
1	5	-18	9	0	9	-18	9	0	9	-18	9	0	9	-12	6	0	6	-12	6	0	6	-6	3	0	3	0
1	6	9	0	9	-18	9	0	9	-18	9	0	9	-18	6	0	6	-12	6	0	6	-12	3	0	3	-6	0
1	7	0	9	-18	9	0	9	-18	9	0	9	-18	9	0	6	-12	6	0	6	-12	6	0	3	-6	3	0
1	8	9	-18	9	0	9	-18	9	0	9	-18	9	0	6	-12	6	0	6	-12	6	0	3	-6	3	0	0
1	9	-18	9	0	9	-18	9	0	9	-18	9	0	9	-12	6	0	6	-12	6	0	6	-6	3	0	3	0
1	10	9	0	9	-18	9	0	9	-18	9	0	9	-18	6	0	6	-12	6	0	6	-12	3	0	3	-6	0
1	11	0	9	-18	9	0	9	-18	9	0	9	-18	9	0	6	-12	6	0	6	-12	6	0	3	-6	3	0
1	12	9	-18	9	0	9	-18	9	0	9	-18	9	0	6	-12	6	0	6	-12	6	0	3	-6	3	0	0
1	13	-12	6	0	6	-12	6	0	6	-12	6	0	6	-8	4	0	4	-8	4	0	4	-4	2	0	2	0
1	14	6	0	6	-12	6	0	6	-12	6	0	6	-12	4	0	4	-8	4	0	4	-8	2	0	2	-4	0
1	15	0	6	-12	6	0	6	-12	6	0	6	-12	6	0	4	-8	4	0	4	-8	4	0	2	-4	2	0
1	16	6	-12	6	0	6	-12	6	0	6	-12	6	0	4	-8	4	0	4	-8	4	0	2	-4	2	0	0
1	17	-12	6	0	6	-12	6	0	6	-12	6	0	6	-8	4	0	4	-8	4	0	4	-4	2	0	2	0
1	18	6	0	6	-12	6	0	6	-12	6	0	6	-12	4	0	4	-8	4	0	4	-8	2	0	2	-4	0
1	19	0	6	-12	6	0	6	-12	6	0	6	-12	6	0	4	-8	4	0	4	-8	4	0	2	-4	2	0
1	20	6	-12	6	0	6	-12	6	0	6	-12	6	0	4	-8	4	0	4	-8	4	0	2	-4	2	0	0
1	21	-6	3	0	3	-6	3	0	3	-6	3	0	3	-4	2	0	2	-4	2	0	2	-2	1	0	1	0
1	22	3	0	3	-6	3	0	3	-6	3	0	3	-6	2	0	2	-4	2	0	2	-4	1	0	1	-2	0
1	23	0	3	-6	3	0	3	-6	3	0	3	-6	3	0	2	-4	2	0	2	-4	2	0	1	-2	1	0
1	24	3	-6	3	0	3	-6	3	0	3	-6	3	0	2	-4	2	0	2	-4	2	0	1	-2	1	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 7.5: The random frequencies of occurrence of the orientations of component 1 after the merging of similar orientations.

orientation k	$p^{\mathrm{R}}_{1,k}$
1, 2, 3 and 4	$3/24$
5, 6, 7 and 8	$2/24$
9, 10, 11 and 12	$1/24$

7.7 Flowchart of the expanded gefit program

The gefit program package is a collection of different fortran subroutines that were originally implemented by Pfennig [156]. The gefit program serves to simultaneously fit parameters of a G^{E}-model to different kinds of experimental data. Three input files are required for the program: a "run", a "mix" and a "pur" file. In the run file details about the parameter fitting such as the data sets that are used for the fit or the starting values of the model parameters are specified. The mix file contains experimental data of mixtures and the pur file contains pure-component informations. A run of the gefit program shows the following general course of action:

- start *gefit*
- read input
- start minimization
  - evaluate objective function
    - call *gex*
  - check convergence: iterate or stop minimization
- write output
- stop *gefit*

For the evaluation of the objective function, different thermo-physical properties of the mixture such as the heat of mixing or the vapor-liquid equilibrium need to be determined. Generally, all thermo-physical properties are determined with a G^{E}-model. Thus, it is necessary to call the subroutine *gex*, which contains an implementation of the selected G^{E}-model. MOQUAC was added to the *gex* subroutine. In the following, the course of action of the *gex* subroutine for MOQUAC is shown:

- define orientations
- upon first call $\rightarrow$ call *morp*
- determine u if required $\rightarrow$ call *mobce*
- cluster similar orientations
- call *moquac*

First, the subroutine *gex* defines the orientations that are distinguished for each component. Upon the first call of the subroutine, the subroutine *morp* is called. This subroutine calculates the random frequency of occurrence of each orientation for every component $p_{i,k}^{\mathrm{R}}$ with a Monte-Carlo method. This Monte-Carlo method was described in section 5.4.

The interaction energy for each combination of orientations depends on the parameters of the interaction-energy model. If these parameters were changed during the course of the minimization, the interaction energies need to be determined anew. The interaction energies are determined in the subroutine *mobce*. In the following, the course of action of this subroutine is described:

- read COSMO files
- average surface-charge densities
- determine u of all contact types

First, the COSMO files are read. These files contain information on molecular geometry and charge distribution. Then, the COSMO charge densities are averaged to eliminate artifacts of the COSMO calculation.

After all interaction energies have been determined, similarly behaving orientations are clustered in the *gex* subroutine. For each cluster, average interaction energies are determined. With the reduced number of orientations and with the average interaction energies, the *moquac* subroutine is called. In this subroutine, the MOQUAC equations are solved and all excess properties as well as the activity coefficient of every component are determined. For the determination of the activity coefficients and the excess enthalpy, analytical equations are used. These results are returned to gefit for the determination of thermo-physical properties for the parameter fit.

7.8 Comparison of experimental data with different model results for the system alkane + ketone

Figures 7.3 to 7.11 show all plots of the comparison of the combination of the interaction-energy model with MOQUAC to all the experimental data points of the selected data sets of the alkane + ketone systems. The interaction-energy model is evaluated with the parameters that were determined from the simultaneous fit to H^{E} and VLE data of alkane + ketone systems. Also the results of COSMO-RS and the modified UNIFAC (Dortmund) model are shown.

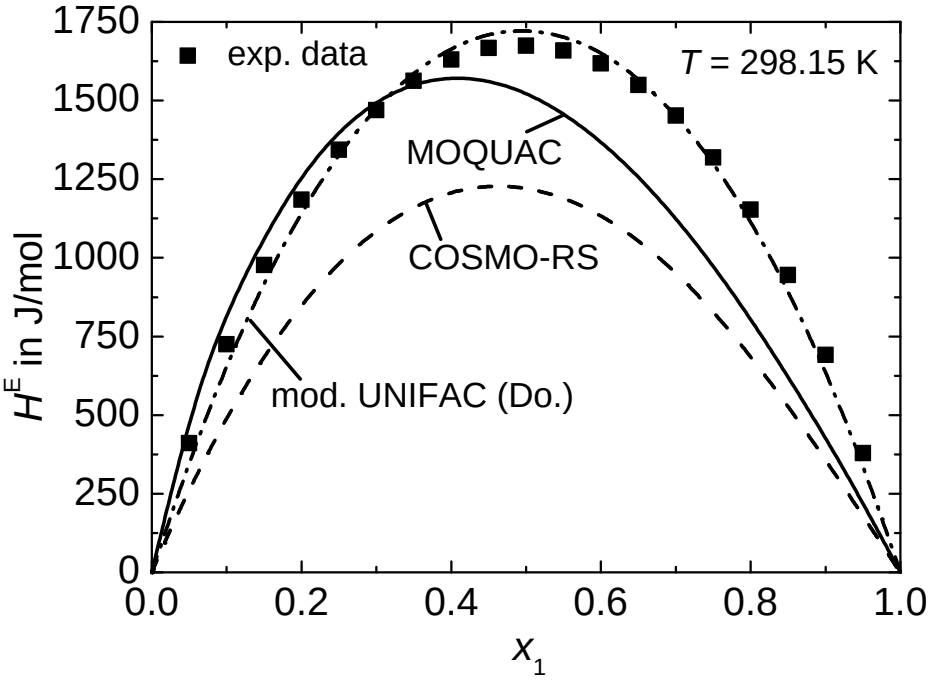


Figure 7.3: Comparison of experimental data with the model results for the system acetone (1) + *n*-heptane (2) at $T = 298.15\,\mathrm{K}$.

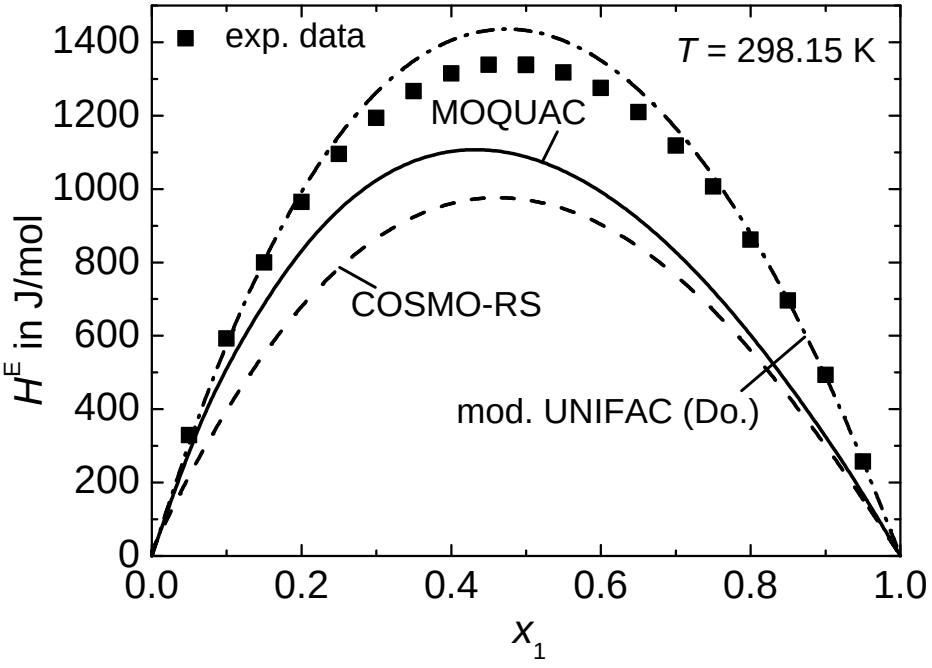


Figure 7.4: Comparison of experimental data with the model results for the system 2-butanone (1) + *n*-heptane (2) at $T = 298.15\,\mathrm{K}$.

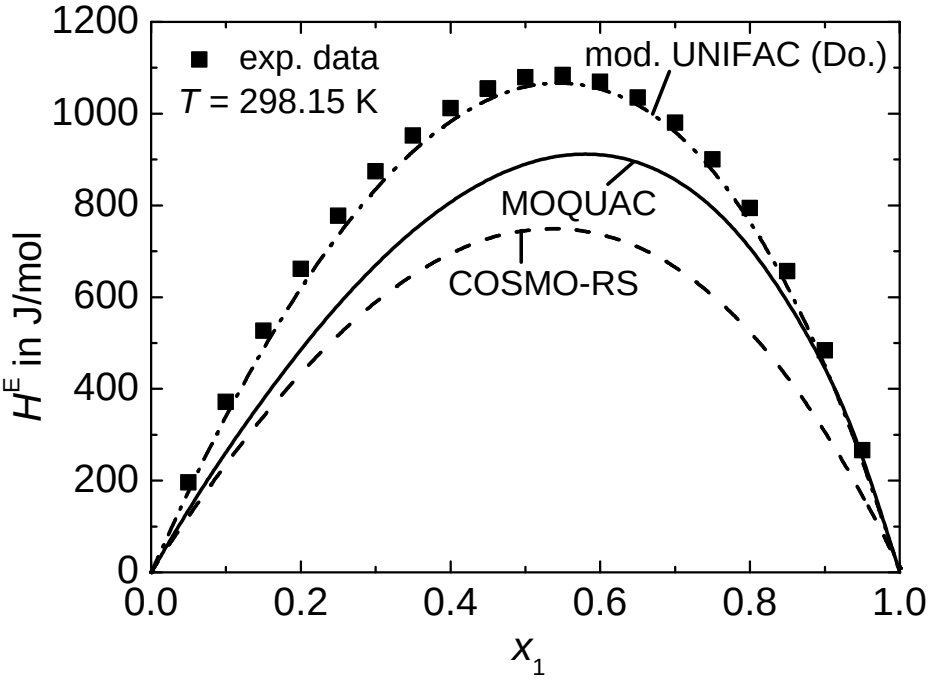


Figure 7.5: Comparison of experimental data with the model results for the system 3-pentanone (1) + *n*-heptane (2) at $T = 298.15\,\mathrm{K}$.

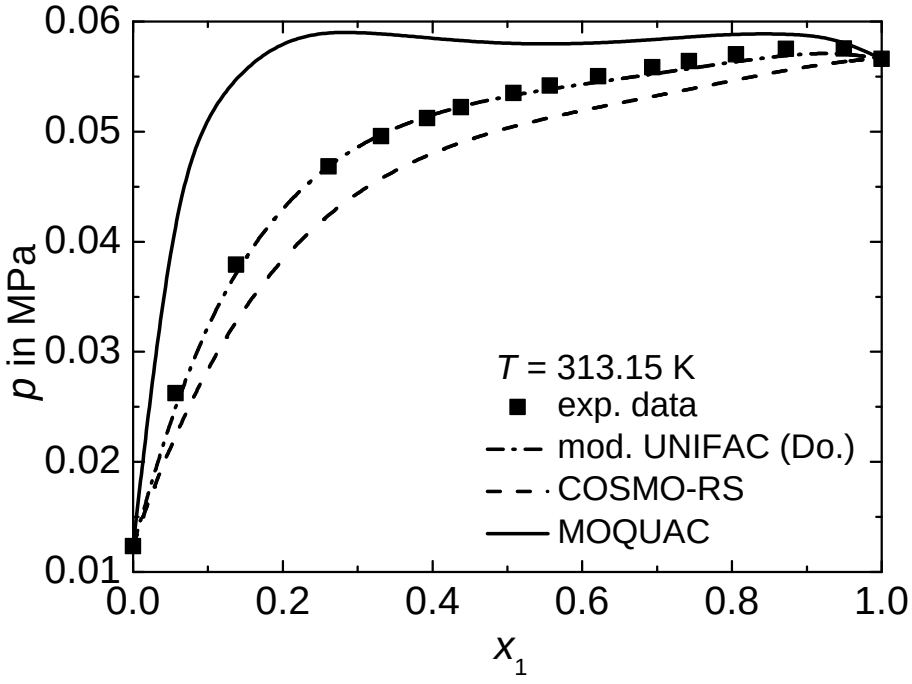


Figure 7.6: Comparison of experimental data with the model results for the system acetone (1) + *n*-heptane (2) at $T = 313.15\,\mathrm{K}$.

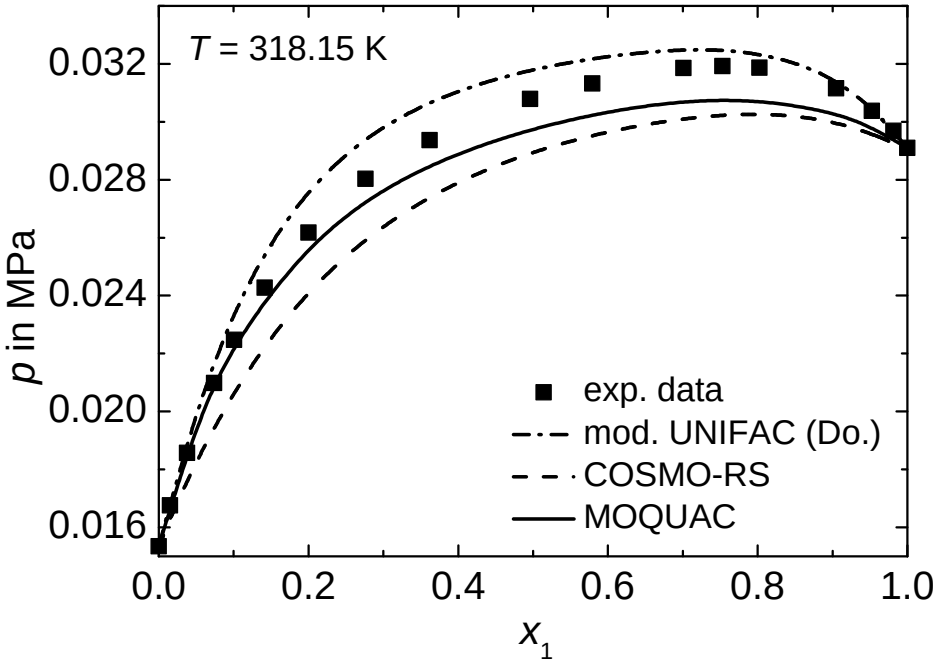


Figure 7.7: Comparison of experimental data with the model results for the system 2-butanone (1) + *n*-heptane (2) at $T = 318.15\,\mathrm{K}$.

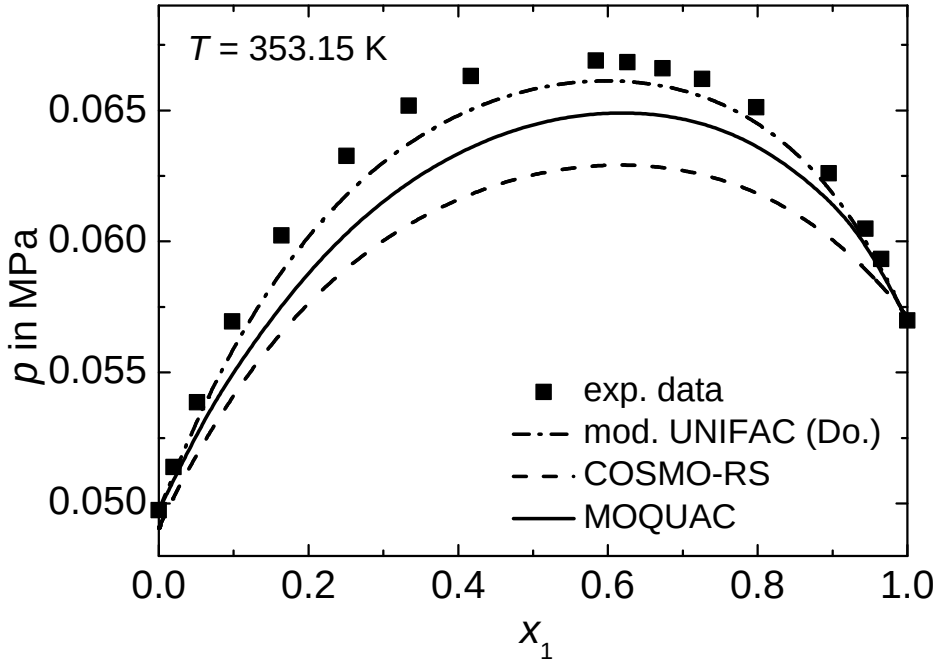


Figure 7.8: Comparison of experimental data with the model results for the system *n*-heptane (1) + 3-pentanone (2) at $T = 353.15\,\mathrm{K}$.

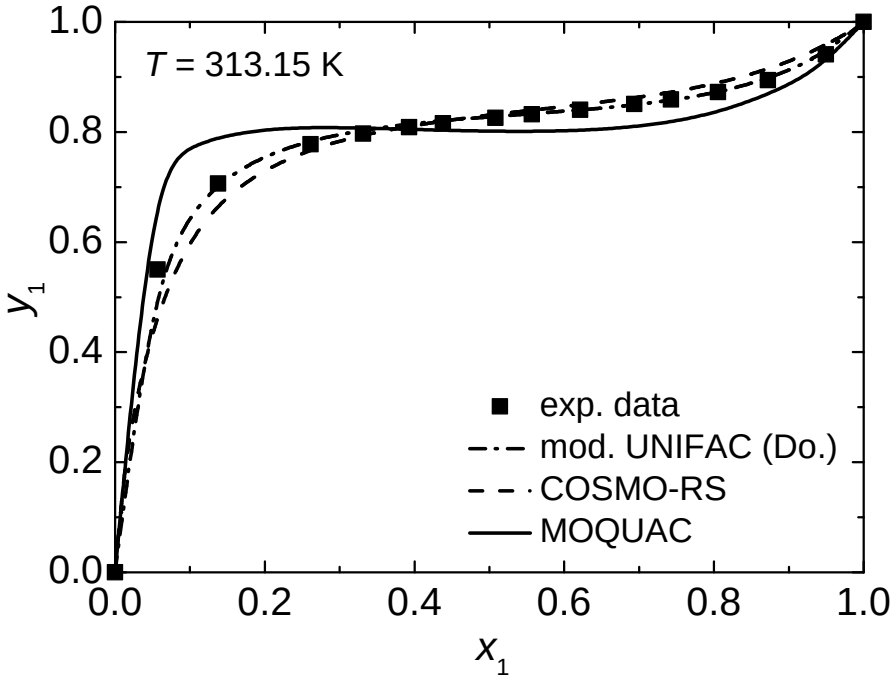


Figure 7.9: Comparison of experimental data with the model results for the system acetone (1) + *n*-heptane (2) at $T = 313.15\,\mathrm{K}$.

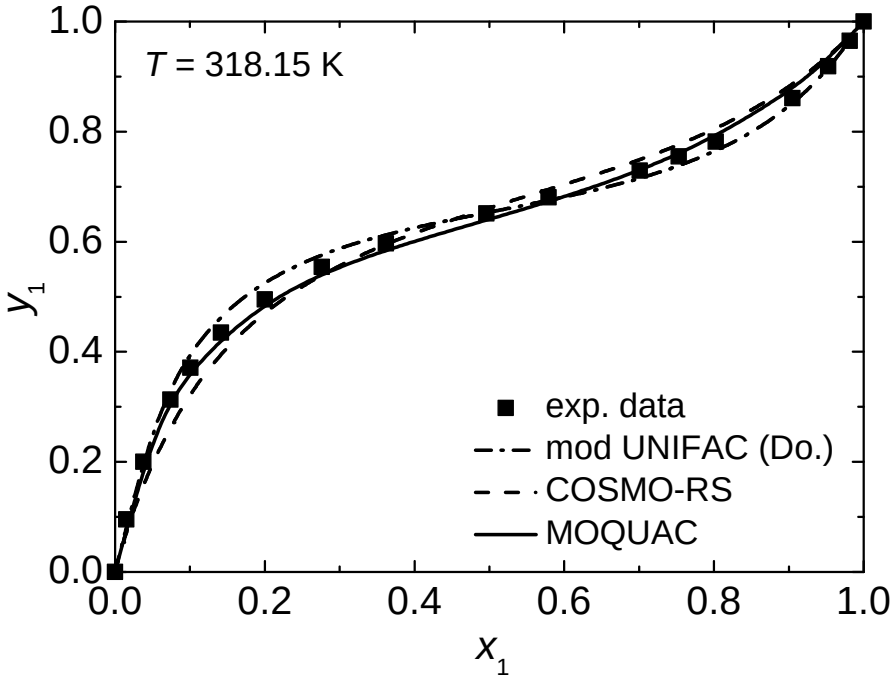


Figure 7.10: Comparison of experimental data with the model results for the system 2-butanone (1) + *n*-heptane (2) at $T = 318.15\,\mathrm{K}$.

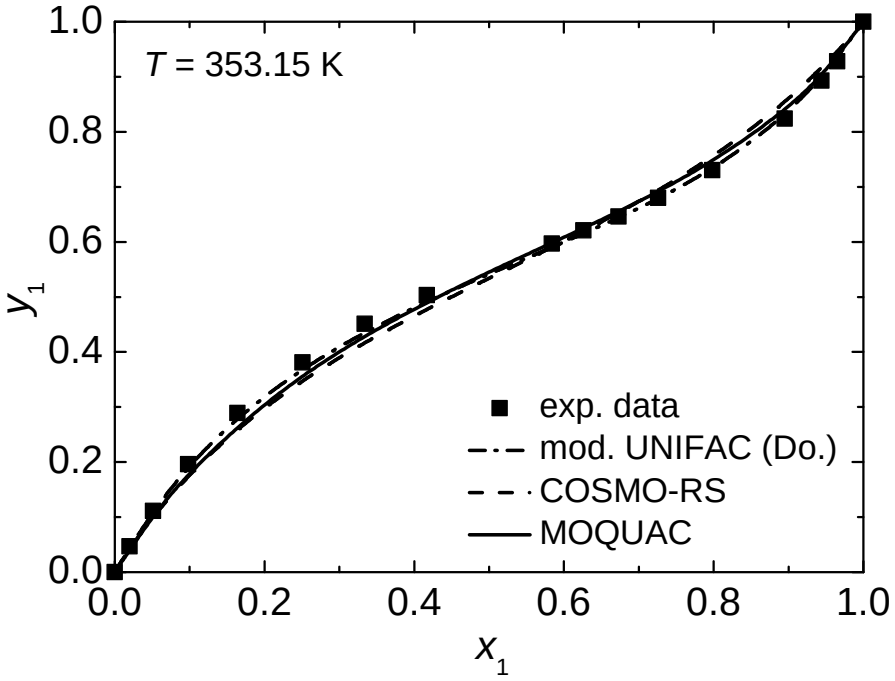


Figure 7.11: Comparison of experimental data with the model results for the system *n*-heptane (1) + 3-pentanone (2) at $T = 353.15\,\mathrm{K}$.

7.9 Comparison of experimental data with different model results for the system alkane + alcohol

Figures 7.12 to 7.20 show all plots of the comparison of the combination of the interaction-energy model with MOQUAC to all the experimental data points of the selected data sets of the alkane + alcohol systems. The interaction-energy model is evaluated with the parameters that were determined from the simultaneous fit to H^{E} and VLE data of alkane + ketone systems. Also the results of COSMO-RS and the modified UNIFAC (Dortmund) model are shown.

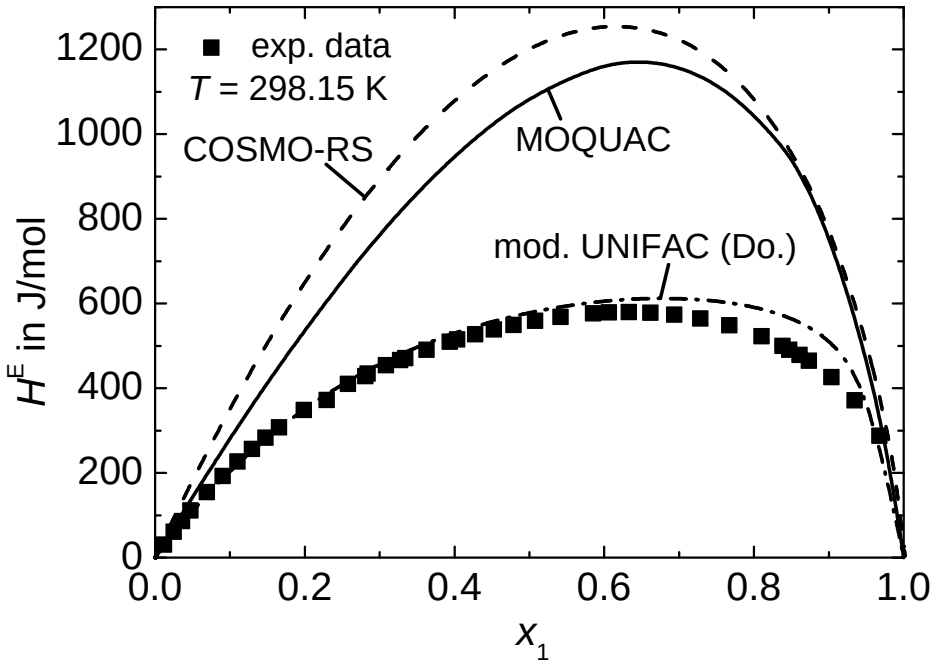


Figure 7.12: Comparison of experimental data with the model results for the system *n*-hexane (1) + ethanol (2) at $T = 298.15\,\mathrm{K}$.

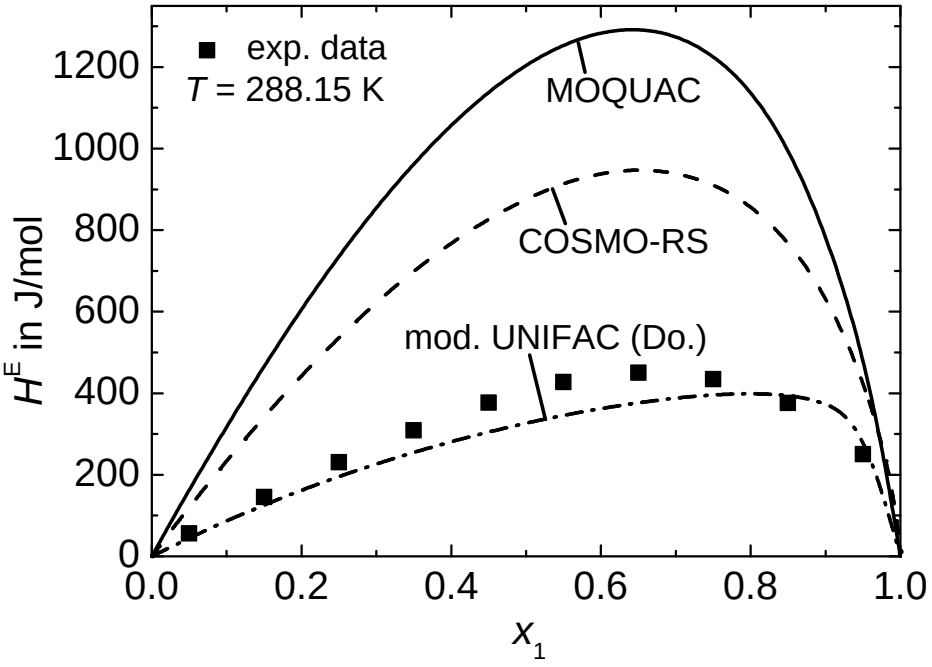


Figure 7.13: Comparison of experimental data with the model results for the system *n*-hexane (1) + 1-butanol (2) at $T = 288.15\,\mathrm{K}$.

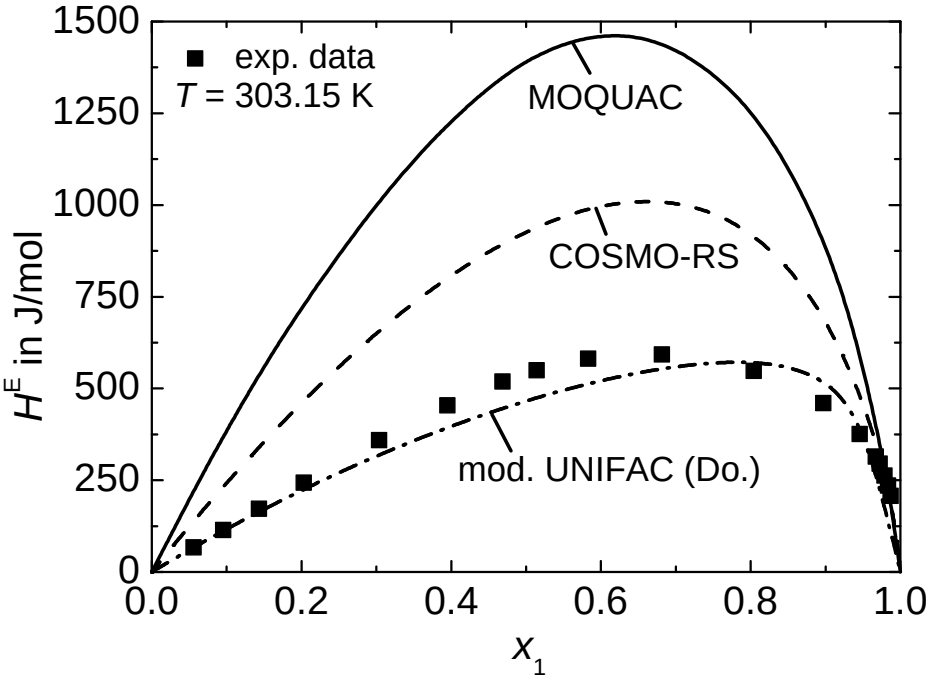


Figure 7.14: Comparison of experimental data with the model results for the system *n*-hexane (1) + 1-pentanol (2) at $T = 303.15\,\mathrm{K}$.

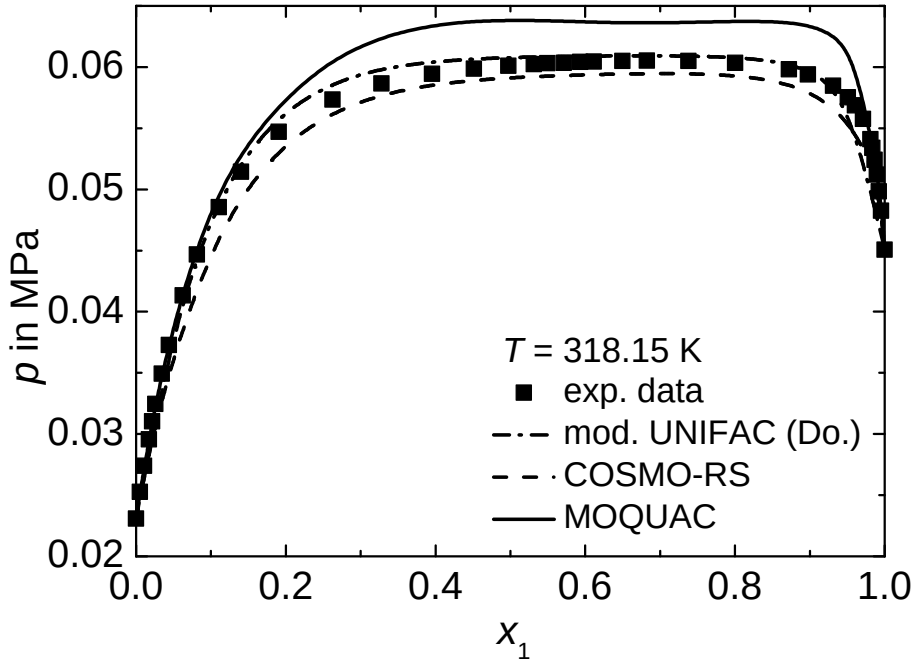


Figure 7.15: Comparison of experimental data with the model results for the system *n*-hexane (1) + ethanol (2) at $T = 318.15\,\mathrm{K}$.

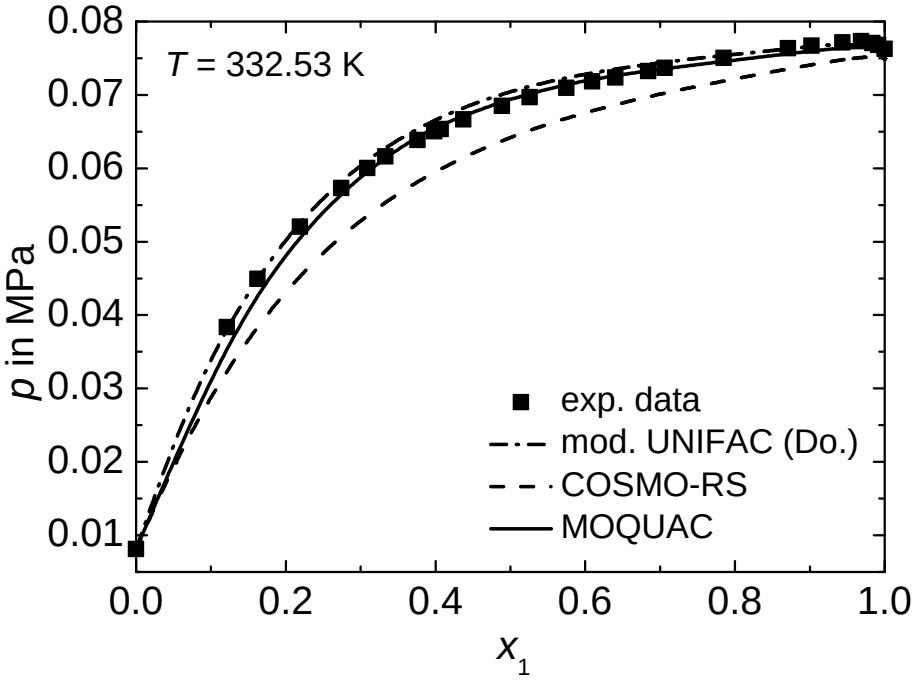


Figure 7.16: Comparison of experimental data with the model results for the system *n*-hexane (1) + 1-butanol (2) at $T = 332.53\,\mathrm{K}$.

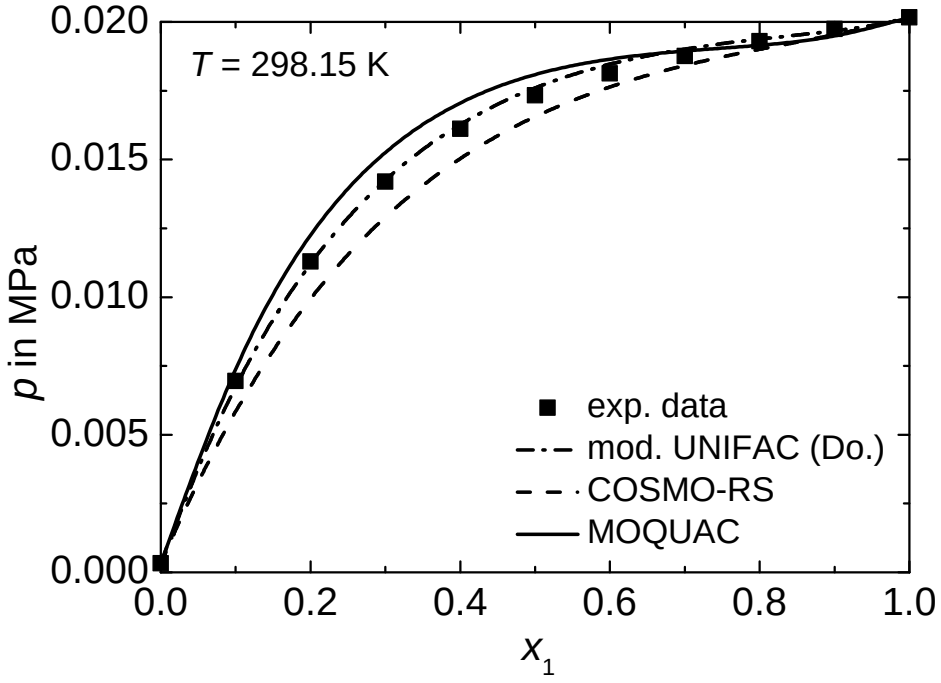


Figure 7.17: Comparison of experimental data with the model results for the system *n*-hexane (1) + 1-pentanol (2) at $T = 298.15\,\mathrm{K}$.

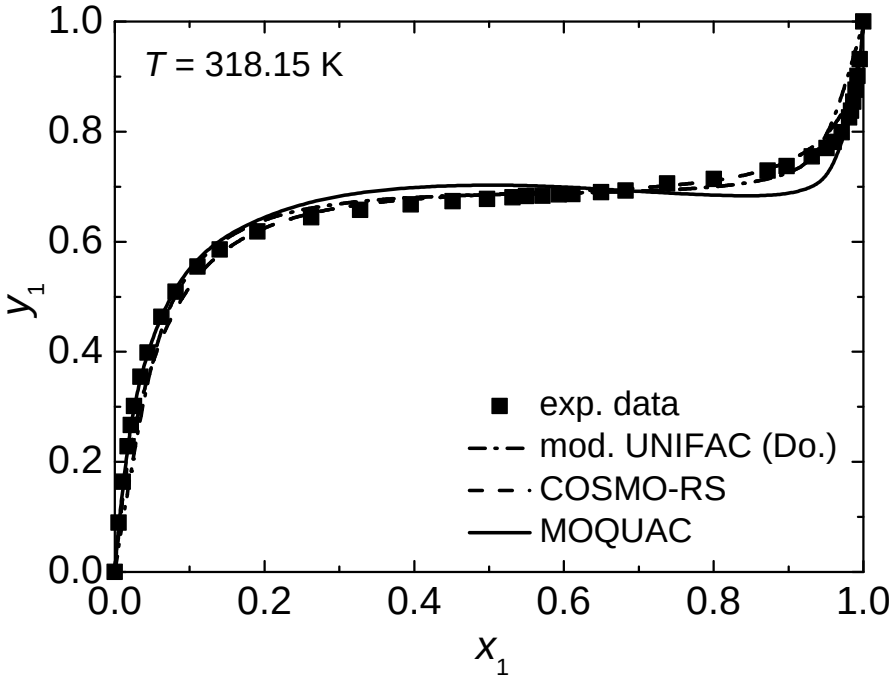


Figure 7.18: Comparison of experimental data with the model results for the system *n*-hexane (1) + ethanol (2) at $T = 318.15\,\mathrm{K}$.

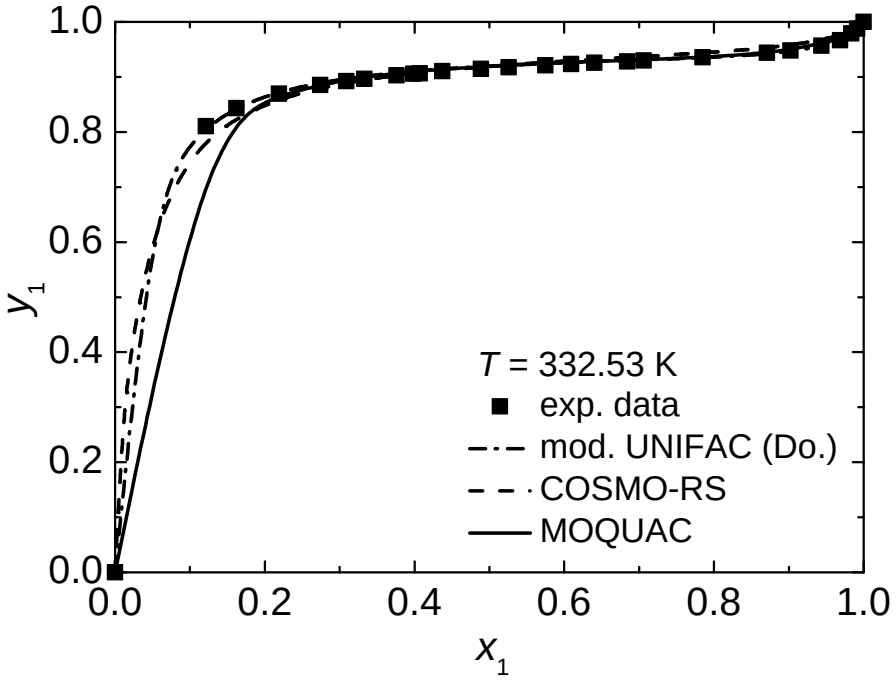


Figure 7.19: Comparison of experimental data with the model results for the system *n*-hexane (1) + 1-butanol (2) at $T = 332.53\,\mathrm{K}$.

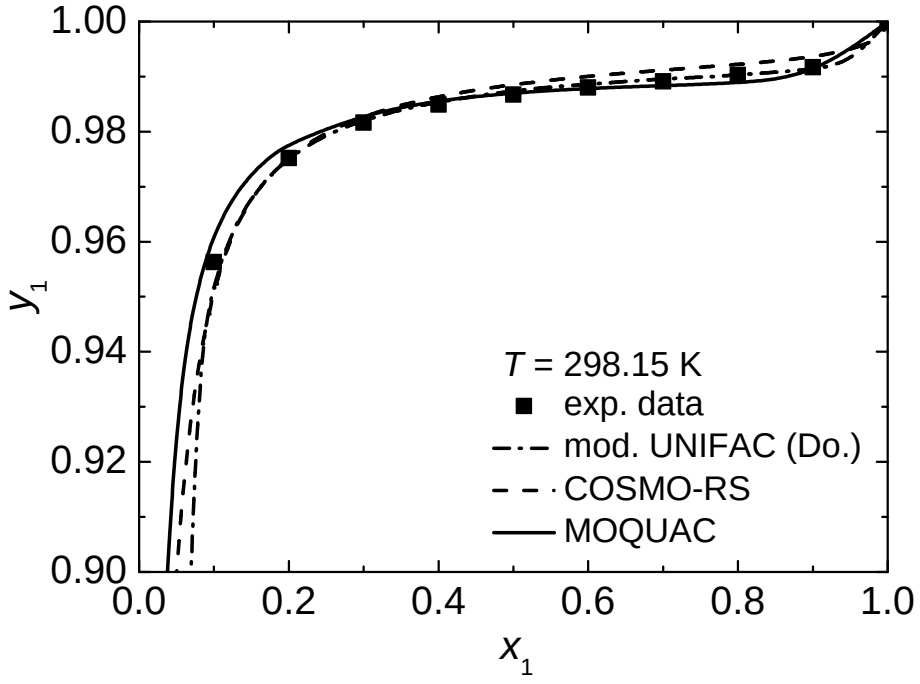


Figure 7.20: Comparison of experimental data with the model results for the system *n*-hexane (1) + 1-pentanol (2) at $T = 298.15\,\mathrm{K}$.

Nomenclature

Symbols

a	K	UNIFAC model parameter
a	m^2	COSMO-RS model parameter
a	$\frac{\mathrm{J}}{\mathrm{mol}}$	correlation parameter
A	$\frac{\mathrm{J}}{\mathrm{mol}}$	free energy
A	m^2	molecular surface area (1 index)
A	K	modified UNIFAC model parameter (2 indices)
A^*	m^2	surface area of the new standard segment
b	-	auxiliary variable
b	$\frac{\mathrm{J}}{\mathrm{mol}}$	correlation parameter
B	-	modified UNIFAC model parameter (2 indices), Boltzmann factor
$\mathbf{B}$	-	auxiliary matrix
c	$\frac{\mathrm{Jm}^2}{\mathrm{molA}^2\mathrm{s}^2}$	COSMO-RS parameter
c	$\frac{\mathrm{JK}}{\mathrm{mol}}$	correlation parameter
c	$\frac{\mathrm{Jm}^4}{\mathrm{molA}^2\mathrm{s}^2}$	interaction-energy model parameter
C	-	modified UNIFAC model parameter (2 indices)
d	m	distance
d	$\frac{\mathrm{J}}{\mathrm{mol}}$	similarity
$\mathbf{D}$	-	auxiliary matrix
e	$\frac{\mathrm{J}}{\mathrm{molm}^2}$	energy of a surface-segment contact
f	-	scaling factor, objective function
g	-	degeneracy
G	$\frac{\mathrm{J}}{\mathrm{mol}}$	Gibbs energy
G	-	NRTL model parameter (2 indices)
h	$\frac{\mathrm{J}}{\mathrm{mol}}$	enthalpy of a molecular contact or surface-segment contact
h	-	scaling factor of the degeneracy
H	$\frac{\mathrm{J}}{\mathrm{mol}}$	enthalpy
$\mathbf{J}$	-	Jacobian matrix
k_B	$\frac{\mathrm{J}}{\mathrm{molK}}$	Boltzmann constant
l	-	UNIFAC parameter
m	-	multiplicity
n	mol	amount of substance
N	-	number
N_A	-	Avogadro constant
N^u	-	number of interaction energies
p	Pa	pressure
p	-	frequency of occurrence, σ-profile
q	-	structural area parameter

q	As	charge
Q	-	structural surface area parameter of a functional group (1 index)
Q	-	canonical partition function
r	-	structural volume parameter
r	m	radius, distance
$\vec{r}$	m	position vector
R	$\frac{\mathrm{J}}{\mathrm{molK}}$	universal gas constant
R	-	structural volume parameter of a functional group (1 index)
R	m	radius
s	$\frac{\mathrm{J}}{\mathrm{molK}}$	entropy of a molecular contact or surface-segment contact
S	$\frac{\mathrm{J}}{\mathrm{molK}}$	entropy
t	K	temperature
T	K	temperature
u	$\frac{\mathrm{J}}{\mathrm{mol}}$	interaction energy of a molecular contact
U	$\frac{\mathrm{J}}{\mathrm{mol}}$	energy
v	-	metric parameter
V	$\frac{\mathrm{m}^3}{\mathrm{mol}}$	molecular volume
w	-	metric parameter
W	-	number of distinguishable ways of arranging molecules on a lattice
x	-	liquid mole fraction (1 index), local composition (2 indices)
X	-	group fraction
$\vec{y}$	-	auxiliary vector
z	-	coordination number, number of nearest neighbors

Greek symbols

α	parameter of the NRTL and COSMO-RS model
γ	activity coefficient
Γ	activity coefficient of a group (1 index), non-random coefficient (2 indices)
Δ	sum of squared residuals
ε	error, permittivity
ε_0	vacuum permittivity
θ	surface-area fraction
λ	parameter of COSMO-RS
Λ	parameter of the Wilson model
μ	chemical potential
ν	parameter of the UNIFAC model
π	mathematical constant pi
ρ	parameter of the combinatorial contribution
σ	surface-charge density
τ	model parameter of the NRTL and UNIFAC model, energy parameter
ϕ	volume fraction
Φ	electrostatic potential
ψ	surface-area fraction (1 index), local composition, contact fraction
Ψ	group surface-area fraction
ω	interaction parameter

Indices

i	mainly used to index components
j	mainly used to index components
k	mainly used to index orientations or groups
l	mainly used to index orientations
m	used to index components, groups, surface segments or contact points
n	used to index components, groups or surface segments
o	used to index surface segments or orientations
p	used to index surface segments or orientations
r	used to index components
μ	used to index surface segments
ν	used to index surface segments
ξ	used to index a microstate

Superscripts

0	pure component
∞	infinity
ap	atom pair
arr	arrangement
c	component
conf	conformer
cp	contact point
ct	contact type
dir	direction
E	excess
f	formation
h	enthalpy
L	liquid
or	orientation
p	polar
rot	rotation
R	random
s	entropy, surface segment
$\bar{}$	average value
$\vec{}$	vector

Subscripts

0	reference value
acc	acceptor
av	average
c	component
calc	calculated
comb	combinatorial
disp	dispersion
don	donor

eff	effective
exp	experimental
gr	group
hb	hydrogen bond
int	polar interactions
max	maximum
ref	reference
res	residual
s	surface segment
tot	total
vdW	van der Waals

Abbreviations

DFT	density functional theory
Do.	Dortmund
exp.	experimental
mod.	modified
TOM	thermodynamics of organic mixtures
VLE	vapor-liquid equilibrium

Bibliography

[1] J. Gmehling, UNIFAC - ein wichtiges Werkzeug für die chemische Industrie, Software-Entwicklung in der Industrie, Springer Verlag, 1987.

[2] E. Klemm, G. Emig, Technische Chemie - Einführung in die chemische Reaktionstechnik, Springer Berlin Heidelberg New York, 2005.

[3] J. M. Briggs, T. B. Nguyen, W. L. Jorgensen, Monte Carlo Simulations of Liquid Acetic Acid and Methyl Acetate with the OPLS Potential Functions, J. Phys. Chem. 95 (1991) 3315–3322.

[4] J. M. Prausnitz, R. N. Lichtenthaler, E. Gomes de Azevedo, Molecular Thermodynamics of Fluid-Phase Equilibria, Prentice-Hall Inc., 2 edn., 1986.

[5] P. J. Flory, Thermodynamics of High Polymer Solutions, J. Chem. Phys. 9 (1941) 660–661.

[6] M. L. Huggins, Solutions of Long Chain Compounds, J. Chem. Phys. 9 (1941) 440.

[7] A. Pfennig, Thermodynamik der Gemische, Springer Verlag Berlin Heidelberg New York, 2004.

[8] E. A. Guggenheim, Statistical Thermodynamics of Mixtures with Zero Energies of Mixing, Proc. Roy. Soc. London A 183 (1944) 203–212.

[9] I. C. Sanchez, R. H. Lacombe, An Elementary Molecular Theory of Classical Fluids. Pure Fluids, J. Phys. Chem. 80 (1976) 2352–2362.

[10] A. J. Staverman, The Entropy of High Polymer Solutions, Recl. Trav. Chim. Pays-Bas 69 (1950) 163–174.

[11] D. S. Abrams, J. M. Prausnitz, Statistical Thermodynamics of Liquid Mixtures: A New Expression for the Excess Gibbs Energy of Partly or Completely Miscible Systems, AIChE J. 21 (1) (1975) 116–128.

[12] A. Bondi, Physical Properties of Molecular Crystals, Liquids and Glasses, Wiley, New York, 1968.

[13] I. Kikic, P. Alessi, P. Rasmussen, A. Fredenslund, On the Combinatorial Part of the UNIFAC and UNIQUAC Models, Can. J. Chem. Eng. 58 (1980) 253–258.

[14] M. D. Donohue, J. M. Prausnitz, Combinatorial Entropy of Mixing Molecules that Differ in Size and Shape. A Simple Approximation for Binary and Multicomponent Mixtures, Can. J. Chem. 53 (1975) 1586–1592.

[15] J. Gmehling, D. Tiegs, A. Medina, M. Soares, J. Bastos, P. Alessi, I. Kikic, M. Schiller, J. Menke, Activity Coefficients at Infinite Dilution, DECHEMA Chemistry Data Series, 1986-2008.

[16] P. L. Huyskens, M. C. Haulait-Pirson, A New Expression for the Combinatorial Entropy of Mixing in Liquid Mixtures, J. Mol. Liq. 31 (1985) 135–151.

[17] U. Weidlich, J. Gmehling, A Modified UNIFAC Model 1 Prediction of VLE, H^{E} and γ_{∞}, Ind. Eng. Chem. Res. 26 (1987) 1372–1381.

[18] B. L. Larsen, P. Rasmussen, A. Fredenslund, A Modified UNIFAC Group-Contribution Model for Prediction of Phase Equilibria and Heats of Mixing, Ind. Eng. Chem. Res. 26 (1987) 2274–2286.

[19] E. A. Guggenheim, Mixtures, Clarendon Press, Oxford, 1952.

[20] J. Barker, Cooperative Orientation Effects in Solutions, J. Chem. Phys. 20 (1952) 1526 – 1532.

[21] J. Barker, Cooperative Orientation in Solutions. The Accuracy of the Quasi-Chemical Approximation, J. Chem. Phys. 21 (1953) 1391 –1394.

[22] G. M. Wilson, Vapor-Liquid-Equilibrium. 11. A New Expression for the Excess Free Energy of Mixing, J. Am. Chem. Soc. 86 (1964) 127–130.

[23] Y. Hu, E. G. Azevedo, J. M. Prausnitz, The Molecular Basis for Local Compositions in Liquid Mixture Models, Fluid Phase Equilib. 13 (1983) 351–360.

[24] K. Lucas, Molecular Models for Liquids, Cambridge University Press, 2007.

[25] G. Pielen, Detaillierte molekulare Simulationen und Parameterstudie für ein ternäres Gemisch zur Weiterentwicklung des GEQUAC-Modells, Ph.D. thesis, RWTH University - AVT - Thermal Process Engineering, 2005.

[26] H. Renon, J. M. Prausnitz, Local Compositions in Thermodynamic Excess Functions for Liquid Mixtures, AIChE J. 14 (1968) 135–144.

[27] A. I. of Chemical Engineers, Design Institute for Physical Property Data ®, URL `http://www.aiche.org/DIPPR/index.aspx`, 2009.

[28] R. L. Scott, Corresponding States Treatment of Nonelectrolyte Solutions, J. Chem. Phys. 25 (1956) 193–205.

[29] G. Maurer, J. M. Prausnitz, Derivation and Extension of UNIQUAC Equation, Fluid Phase Equilib. 2 (2) (1978) 91–99.

[30] C. McDermott, N. Ashton, Note on the Definition of Local Composition, Fluid Phase Equilib. 1 (1977) 33–35.

[31] S. I. Sandler, Models for Thermodynamic and Phase Equilibria Calculations, CRC Press, 1993.

[32] S. I. Sandler, H. S. Wu, Use of ab Initio Quantum Mechanics Calculations in Group Contribution Methods. 1. Theory and the Basis for Group Identifications, Ind. Eng. Chem. Res. 30 (1991) 881–889.

[33] S. I. Sandler, H. S. Wu, Use of ab Initio Quantum Mechanics Calculations in Group Contribution Methods. 2. Test of New Groups in UNIFAC, Ind. Eng. Chem. Res. 30 (1991) 889–897.

[34] A. Fredenslund, R. L. Jones, J. M. Prausnitz, Group-Contribution Estimation of Activity Coefficients in Nonideal Liquid Mixtures, AIChE J. 21 (1975) 1086–1099.

[35] H. Kehiaian, Group Contribution Methods for Liquid Mixtures: a Critical Review, Fluid Phase Equilib. 13 (1983) 243–252.

[36] The UNIFAC Consortium, URL `http://unifac.ddbst.de/`, 2011.

[37] B. Lacmann, B. Schäfer, D. Brennecke, J. Li, A. Pfennig, Calculations of Multicomponent Phase Equilibria with a New Group Contribution Method, Phys. Chem. Chem. Phys. 1 (1999) 115–119.

[38] H. Kehiaian, Thermodynamik flüssiger Mischungen von Kohlenwasserstoffen mit verwandten Substanzen, Berichte der Bunsen-Gesellschaft 81 (1977) 908–921.

[39] H. V. Kehiaian, J. P. E. Grolier, G. C. Benson, Thermodynamics of Organic Mixtures. A Generalized Quasichemical Theory in Terms of Group Surface Interactions, J. Chim. Phys. 75 (1978) 1031–1048.

[40] H. Kehiaian, S. I. Sandler, Thermodynamic Properties of Binary Mixtures Containing Aldehydes. 2 Liquid-Vapor equilibria in Normal or Branched Alkanal + Normal Alkane Mixtures: Analysis in Terms of a Quasichemical Group-Contribution Model, Fluid Phase Equilib. 17 (1984) 139–145.

[41] H. Kehiaian, Thermodynamics of Binary Liquid Organic Mixtures, Pure & Appl. Chem. 57 (1985) 15–30.

[42] J. A. Gonzalez, I. G. de la Fuente, J. C. Cobos, C. Casanova, Characterization of the Alcohol/Alkanol Contacts and Prediction of Excess Functions of Ternary Systems of Two *n*-Alkan-1-Ols and One *n*-Alkane Using DISQUAC, Fluid Phase Equilib 78 (1992) 61–80.

[43] J. A. Gonzalez, I. G. de la Fuente, J. C. Cobos, C. Casanova, Solid-Liquid Equilibria Using DISQUAC. Prediction for 1-Alkanol + *n*-Alkane Systems, Fluid Phase Equilib 94 (1994) 167–179.

[44] J. A. Gonzalez, I. G. de la Fuente, J. C. Cobos, Thermodynamics of Mixtures Containing Linear Monocarboxylic Acids ii. Binary Systems Showing Cross-Association Between Components: DISQUAC Characterization of Monocarboxylic Acid + 1-Alkanol + Linear Monocarboxylic Acid Mixtures, Fluid Phase Equilib. 135 (1997) 1–21.

[45] H. Kehiaian, B. Marongiu, A Comparative Study of Thermodynamic Properties and Molecular Interactions in Mono- and Polychloroalkane + *n*-Alkane or + Cyclohexane Mixtures, Fluid Phase Equilib. 40 (1988) 23–78.

[46] A. Klamt, COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design, Elsevier, 2005.

[47] A. Klamt, G. Schüürmann, 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 2 (1993) 799–805.

[48] B. I. Larsen, P. Rasmussen, A Comparison between the Quasichemical Model and Two-Fluid Local-Composition Models, Fluid Phase Equilib. 28 (1986) 1–11.

[49] K. Egner, J. Gaube, A. Pfennig, GEQUAC, an Excess Gibbs Energy Model for Simultaneous Description of Associating and Non-Associating Liquid Mixtures, Ber. Bunsen Ges. Phys. Chem. 101 (1997) 209–218.

[50] K. Egner, Ein neues Modell zur Berechnung der Exzessgrößen stark nichtidealer Mischungen und Messungen ternärer Exzessenthalpien, Ph.D. thesis, RWTH University - AVT - Thermal Process Engineering, 1998.

[51] K. Egner, J. Gaube, A. Pfennig, GEQUAC, an Excess Gibbs Energy Model Describing Associating and Nonassociating Liquid Mixtures by a New Model Concept for Functional Groups, Fluid Phase Equilib. 158-160 (1999) 381–389.

[52] A. Klamt, Conductor-Like Screening Model for Real Solvents: A new Approach to the Quantitative Calculation of Solvation Phenomena, J. Phys. Chem. 99 (1995) 2224–2235.

[53] W. Kohn, P. Hohenberg, Inhomogeneous Electron Gas, Phys. Rev. 136 (1964) B864–B871.

[54] W. Kohn, L. J. Sham, Self-Consistent Equations Including Exchange and Correlation Effects, Phys. Rev. 140 (1965) A1133–A1138.

[55] A. Klamt, J. Volker, T. Buerger, J. C. W. Lohrenz, Refinement and Parametrization of COSMO-RS, J. Phys. Chem. A 102 (1998) 5074–5085.

[56] A. Klamt, personal communication, 2011.

[57] A. Klamt, F. Eckert, COSMO-RS: A Novel and Efficient Method for the a Priori Prediction of Thermophysical Data of Liquids, Fluid Phase Equilib. 172 (2000) 43–72.

[58] S. T. Lin, I. Sandler, A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model, Ind. Eng. Chem. Res. 41 (2002) 899–913.

[59] H. Grensemann, J. Gmehling, Performance of a Conductor-Like Screening Model for Real Solvents Model in Comparison to Classical Group Contribution Methods, Ind. Eng. Chem. Res. 44 (2005) 1610–1624.

[60] T. Banerjee, M. K. Singh, A. Khanna, Prediction of Binary VLE for Imidazolium Based Ionic Liquid Systems Using COSMO-RS, Ind. Eng. Chem. Res. 45 (2006) 3207–3219.

[61] C. C. Pye, T. Ziegler, E. van Lenthe, J. N. Louwen, An Implementation of the Conductor-Like Screening Model of Solvation Within the Amsterdam Density Functional Package - Part 2 COSMO for Real Solvents, Can. J. Chem. 87 (2009) 790–797.

[62] F. Eckert, A. Klamt, COSMOtherm software, COSMOlogic GmbH, Leverkusen, Germany, 2011.

[63] A. Klamt, The COSMO and COSMO-RS Solvation Models, Wiley Interdisciplinary Reviews: Computational Molecular Science (2011) n/a–n/aISSN 1759-0884, URL `http://dx.doi.org/10.1002/wcms.56`.

[64] A. Pfennig, Thermodynamic Modeling of Polymer Solutions with a Modified Staverman Equation, Macromol. Chem. Theory Simul. 3 (1994) 389–407.

[65] C. Christensen, J. Gmehling, P. Rasmussen, U. Weidlich, T. Holderbaum, Heats of Mixing Data Collection, DECHEMA Chemistry Data Series, 1984-1991.

[66] D. Patterson, P. Tancrede, P. Bothorel, Interactions in Alkane Systems by Depolarized Rayleigh Scattering and Calorimetry. Part 1. Orientational Order and Condensation Effects in *n*-Hexadecane + Hexane and Nonane Isomers., J. Chem. Soc., Faraday Trans. 2 73 (1977) 15–28.

[67] D. Patterson, P. Tancrede, P. Bothorel, P. de St. Romain, Interactions in Alkane Systems by Depolarized Rayleigh Scattering and Calorimetry. Part 2. Orientational Order in Systems Containing Hexadecane Isomers of Different Structure., J. Chem. Soc., Faraday Trans. 2 73 (1977) 29–39.

[68] M. de Matos Alves, Parameterbestimmung eines physikalisch fundierten Modells zur Beschreibung der Entropie einer athermischen Mischung, Master's thesis, RWTH University - AVT - Thermal Process Engineering, 2010.

[69] J. More, B. Garbow, K. Hillstrom, Minpack, Argonne National Laboratory, www.netlib.org/minpack/, 2011.

[70] A. Bondi, Van der Waals Volumes and Radii, J. Phys. Chem. 68 (1964) 441–451.

[71] E. Hála, T. Boublik, Einführung in die statistische Thermodynamik, Tschechoslowakische Akademie der Wissenschaften, 1967.

[72] M. Buggert, C. Cadena, L. Mokrushina, I. Smirnova, E. J. Maginn, W. Arlt, COSMO-RS Calculations of Partition Coefficients: Different Tools for Conformational Search, Chem. Eng. Technol. 32 (2009) 977–986.

[73] HyperChem, URL `http://www.hyper.com/`, 2011.

[74] J. Bacher, Clusteranalyse: anwendungsorientierte Einführung, R. Oldenbourg Verlag München Wien, 1996.

[75] M. G. Martin, J. I. Siepmann, Transferable Potentials for Phase Equilibria. 1. United-Atom Description of *n*-Alkanes, J. Phys. Chem. B 102 (1998) 2569–2577.

[76] M. G. Martin, J. I. Siepmann, Novel Configurational-Bias Monte Carlo Method for Branched Molecules. Transferable Potentials for Phase Equilibria. 2. United-Atom Description of Branched Alkanes, J. Phys. Chem. B 103 (1999) 4508–4517.

[77] B. Chen, I. Siepmann, Transferable Potentials for Phase Equilibria. 3. Explicit-Hydrogen Description of Normal Alkanes, J. Phys. Chem. B 103 (1999) 5370–5379.

[78] C. D. Wick, M. G. Martin, J. I. Siepmann, Transferable Potentials for Phase Equilibria. 4. United-Atom Description of Linear and Branched Alkenes and Alkylbenzenes, J. Phys. Chem. B 104 (2000) 8008–8016.

[79] B. Chen, J. J. Potoff, S. J. I., Monte Carlo Calculations for Alcohols and Their Mixtures with Alkanes. Transferable Potentials for Phase Equilibria. 5. United-Atom Description of Primary, Secondary, and Tertiary Alcohols, J. Phys. Chem. B 105 (2001) 3093–3104.

[80] J. M. Stubbs, J. J. Potoff, J. I. Siepmann, Transferable Potentials for Phase Equilibria. 6. United-Atom Description for Ethers, Glycols, Ketones, and Aldehydes, J. Phys. Chem. B 108 (2004) 17596–17605.

[81] C. D. Wick, J. M. Stubbs, N. Rai, J. I. Siepmann, Transferable Potentials for Phase Equilibria. 7. Primary, Secondary, and Tertiary Amines, Nitroalkanes and Nitrobenzene, Nitriles, Amides, Pyridine, and Pyrimidine, J. Phys. Chem. B 109 (2005) 18974–18982.

[82] N. Lubna, G. Kamath, J. J. Potoff, N. Rai, J. I. Siepmann, Transferable Potentials for Phase Equilibria. 8. United-Atom Description for Thiols, Sulfides, Disulfides, and Thiophene, J. Phys. Chem. B 109 (2005) 24100–24107.

[83] N. Rai, J. I. Siepmann, Transferable Potentials for Phase Equilibria. 9. Explicit Hydrogen Description of Benzene and Five-Membered and Six-Membered Heterocyclic Aromatic Compounds, J. Phys. Chem. B 111 (2007) 10790–10799.

[84] R. Hellmann, E. Bich, E. Vogel, Ab Initio Intermolecular Potential Energy Surface and Second Pressure Virial Coefficients of Methane, J. Chem. Phys. 128 (2008) 214303.

[85] J. Gmehling, U. Onken, W. Arlt, P. Grenzheuser, U. Weidlich, B. Kolbe, J. Rarey, Vapor-Liquid Equilibrium Data Collection, DECHEMA Chemistry Data Series, 1991-2009.

[86] C. Christensen, J. Gmehling, P. Rasmussen, U. Weidlich, T. Holderbaum, Heats of Mixing Data Collection, DECHEMA Chemistry Data Series, 1984-1991.

[87] Y. Akamatsu, H. Ogawa, S. Murakami, Molar Excess Enthalpies, Molar Excess Volumes and Molar Isentropic Compressions of Mixtures of 2-Propanone With Heptane, Benzene and Trichloromethane at 298.15 k, Thermochim. Acta 113 (1987) 141–150.

[88] G. Kolasinska, M. Goral, J. Giza, Vapour-Liquid Equilibria and Excess Gibbs Free Energy in Binary Systems of Acetone with Aliphatic and Aromatic Hydrocarbons at 313.15K, Z. Phys. Chemie, Leipzig 263 (1982) 151–160.

[89] O. Kiyohara, Y. Handa, G. C. Benson, Thermodynamic Properties of Binary Mixtures Containing Ketones III. Excess Enthalpies of *n*-Alkanes + Some Aliphatic Ketones, J. Chem. Thermodyn. 11 (1979) 453–460.

[90] M. Takeo, K. Nishii, T. Nitta, T. Katayama, Isothermal Vapor-Liquid Equilibria for Two Binary Mixtures of Heptane with 2-Butanone and 4-4-Methyl-2-Pentanone Measured by a Dynamic Still with a Pressure Regulation, Fluid Phase Equilib. 3 (1979) 123–131.

[91] G. Geiseler, H. Koehler, Thermodynamisches Verhalten der Mischsysteme Methylaethylketoxim/*n*-Heptan, Diaethylketon/*n*-Heptan und Methylaethylketoxim/Diaethylketon, Ber. Bunsenges. Phys. Chem. 72 (1968) 697–706.

[92] S. J. O'Shea, R. H. Stokes, Activity Coefficients and Excess Partial Molar Enthalpies for Ethanol + Hexane from 283 to 318 K, J. Chem. Thermodynamics 18 (1986) 691–696.

[93] T. H. Nguyen, G. A. Ratcliff, Heats of Mixing of *n*-Alcohol - *n*-Alkane Systems at 15°C and 55°C, J. Chem. Eng. Data 20-3 (1975) 252–255.

[94] C. Berro, M. Rogalski, A. Peneloux, Excess Gibbs Energies and Excess Volumes of 1-Butanol - *n*-Hexane and 2-Methyl-1-Propanol - *n*-Hexane Binary Systems, J. Chem. Eng. Data 27 (1982) 352–355.

[95] C. G. Savini, D. R. Winterhalter, H. C. Van Ness, SEL. DATA MIXTURES, SER. A, INT. DATA SERIES 11.

[96] S. G. Sayegh, G. A. Ratcliff, Excess Gibbs Energies of Binary Systems of Isopentanol and *n*-Pentanol with Hexane Isomers at 25°C: Measurement and Prediction by Analytical Group Solution Model, J. Chem. Eng. Data 21-1 (1976) 71–74.

[97] R. Dennington, T. Keith, J. Millam, GaussView Version 5, semichem Inc. Shawnee Mission KS, 2009.

[98] B. Marongiu, S. Porcedda, A. Piras, D. Falconieri, M. R. Tine, The Disquac Model, Rev. Roum. Chim. 56 (2011) 453–464.

[99] P. Alessi, I. Kikic, A. Alessandrini, M. Fermegelia, Activity Coefficients at Infinite Dilution by Gas-Liquid Chromatography. 1. Hydrocarbons and *n*-Chloroparaffins in Organic Solvents, J. Chem. Eng. Data 27 (1982) 445–448.

[100] R. Castellis, E. Aranncibia, C. Castells, Thermodynamics of Hydrocarbon Solutions Using G.L.C. *n*-Hexane, *n*-Heptane, Benzene, and Toluene as Solutes Each at Infinite Dilution in *n*-Hexadecane, in *n*-Octadecane, and in *n*-Eicosane, J. Chem. Thermodyn. 22 (1990) 969–977.

[101] C. Chien, M. Kopecni, R. Laub, C. Smith, Solute Liquid-Gas Activity and Partition Coefficients with Mixtures of *n*-Hexadecane and *n*-Octadecane with *N*,*N*-Dibutyl-2-Ethylhexylamide Solvents, J. Phys. Chem. 85 (1981) 1864–1871.

[102] C. Hicks, C. Young, Activity Coefficients of C4-C8 *n*-Alkanes in C16-C32 *n*-Alkanes, Trans. Faraday Soc. 64 (1968) 2675–2682.

[103] T. Letcher, W. Moollan, The Determination of Activity Coefficients at Infinite Dilution Using G.L.C. with a Moderately Volatile Solvent (Dodecane) at the Temperatures 280.15 K and 298.15 K, J. Chem. Thermodyn. 27 (1995) 1025–1032.

[104] D. Richon, P. Antoine, H. Renon, Infinite Dilution Activity Coefficients of Linear and Branched Alkanes from C1 to C9 in *n*-Hexadecane by Inert Gas Stripping, Ind. Eng. Chem. Process Des. Dev. 19 (1980) 144–147.

[105] P. Snyder, J. Thomas, Solute Activity Coefficients at Infinite Dilution via Gas-Liquid Chromatography, J. Chem. Eng. Data 13 (1968) 527–529.

[106] J. Thomas, B. Newman, D. Wood, C. Eckert, Limiting Activity Coefficients of Nonpolar and Polar Solutes in both Volatile and Nonvolatile Solvents by Gas Chromatography, J. Chem. Eng. Data 27 (1982) 399–405.

[107] B. Dolezal, J. Horejs, R. Volf, R. Holub, unknown title, Sci Papers of the Prague Inst. of Chem. Techn. 17 (1982) 113.

[108] A. N. Genkin, Equation for Calculating *n*-Alkane Activity Coefficients at Infinite Dilution in Saturated Hydrocarbons of Normal Structure, Zh. Fiz. Khim. 51 (1967) 1798.

[109] I. Kikic, H. Renon, Extension of Chromatographic Method of Determination of Thermodynamic Properties, J. Sep. Sci. 11 (1976) 45–63.

[110] C. Wufeng, L. Lan, G. Hua, Z. Guo-Kan, Study on Activity Coefficients at Infinite Dilution in Mixed Stationary Liquids by Gas-Liquid Chromatography, Acta Chimica Sinica 42 (1984) 233.

[111] T. Letcher, G. Netherton, Prediction of Finite-Concentration Activity Coefficients from a Single G.L.C.-Determined Activity Coefficient at Infinite Dilution, J. Chem. Thermodyn. 7 (1975) 353–357.

[112] A. Cruickshank, M. Windsor, C. Young, The Use of Gas-Liquid Chromatography to Determine Activity Coefficients and Second Virial Coefficients of Mixtures. II. Experimental Studies on Hydrocarbon Solutes, Proc. Roy. Soc. London 295 (1966) 271.

[113] O. Wicarova, J. Novak, J. Janak, A Study on the Reliability of the Excess Enthalpies Calculated from Gas-Liquid Chromatographic Retention Data, J. Chromatogr. 65 (1972) 241.

[114] Y. Iwai, M. Yamashita, K. Kohaski, Y. Aray, Measurement and Prediction of Infinite Dilution Activity Coefficients for C6 Hydrocarbons in Heavy Paraffinic Hydrocarbons, Kagaku Kogaku Ronbunshu 14 (1988) 706.

[115] J. Park, A. Hussam, P. Couasnom, D. Frity, P. Carr, Experimental Reexamination of Selected Partition Coefficients from Rohrschneider's Data Set, Anal. Chem. 59 (1987) 1970–1976.

[116] S. Shen, S. Wang, J. Zhang, J. Shi, Determination and Application of Activity Coefficients at Infinite Dilution, Hua Kung Hsueh Pao (Peking) 39 (1988) 501.

[117] J. Rytting, L. Huston, T. Higuchi, Thermodynamic Group Contributions for Hydroxyl, Amino, and Methylene Groups, J. Pharm. Sci. 67 (1978) 615.

[118] C. Deal, E. Derr, Selectivity and Solvency in Aromatics Recovery, Ind. Eng. Chem. Process Des. Dev. 3 (1964) 394–399.

[119] A. Kwantes, G. Rijnders, The Determination of Activity Coefficients at Infinite Dilution by Gas-Liquid Chromatography, Gas Chromatog., Proc. Symposium, Amsterdam 1 (1958) 125–135.

[120] A. I. Belfer, Unsteady-State Chromatography, Neftekhimiya 12 (1972) 435.

[121] A. Belfer, D. Locke, Non-Steady-State Gas Chromatography for Activity Coefficient Measurements, Anal. Chem. 56 (1984) 2485–2489.

[122] M. Hoffstee, A. Kwantes, G. Rijnders, Determination of Activity Coefficients at Infinite Dilution by Gas-Liquid Chromatography, Proc. Intern. Symposium Distillation, Brighton, England 1 (1960) 105–109.

[123] J. Leroi, J. Masson, H. Renon, J. Fabries, H. Sannier, Accurate Measurement of Activity Coefficient at Infinite Dilution by Inert Gas Stripping and Gas Chromatography, Ind. Eng. Chem. Process Des. Dev. 16 (1977) 139–144.

[124] T. Letcher, F. Marsiciano, Thermodynamics of Hydrocarbon Mixtures Using Gas-Liquid Chromatography. The Determination of Activity Coefficients of Some Unsaturated C5 and C6 Unbranched Hydrocarbons in *n*-Octadecane, *n*-Octadec-1-ene, *n*-Hexadecane, and *n*-Hexadec-1-ene, J. Chem. Thermodyn. 6 (1974) 501–507.

[125] E. Turek, D. Arnold, R. Greenkorn, K. Chao, A Gas-Liquid Partition Chromatograph for the Accurate Determination of Infinite Dilution Activity Coefficients, Ind. Eng. Chem. Fundam. 18 (1979) 426–429.

[126] D. Arnold, a Group Contribution Model for Liquids and Their Mixtures with Experimental Determination of Infinite Dilution Activity Coefficients, Ph.D. thesis, Purdue University, USA, 1980.

[127] A. Cruickshank, B. Gainey, C. Young, Activity Coefficients of Hydrocarbons C4 to C8 in *n*-Octadecane at 35°C. Gas-Chromatographic Investigation, Trans. Faraday Soc. 64 (1968) 337.

[128] D. Arnold, R. Greenkorn, K. Chao, Infinite-Dilution Activity Coefficients for Alkanals, Alkanoates, Alkanes, and Alkanones in *n*-Octane, J. Chem. Eng. Data 27 (1982) 123–125.

[129] N. D. Gritchina, unknown title, Nepodviznye fazy v gazovoi chromatografii 1 (1970) 42–46.

[130] W. Chen, M. Chen, Y. Zheng, unknown title, Lan-Chou ta Hsueh Hsueh Pao 19 (1983) 84.

[131] R. Newman, J. Prausnitz, Polymer-Solvent Interactions from Gas-Liquid Partition Chromatography, J. Phys. Chem. 76 (1972) 1492–1496.

[132] D. Richon, F. Sorrentino, A. Voilley, Infinite Dilution Activity Coefficients by the Inert Gas Stripping Method: Extension to the Study of Viscous and Foaming Mixtures, Ind. Eng. Chem. Process Des. Dev. 24 (1985) 1160–1165.

[133] G. Alimova, N. Pavlyuk, A. Gaile, Study of the Selectivity of Solvents in the Dearomatization of Liquid Paraffins, Issled. v Obl. Khimii i Tekhnol. Produktov Pererab. Goryuch. Iskopaemykh, L. 1 (1980) 92–95.

[134] L. E. Reshetnikova, D. A. Vyakhirev, Gas Chromatographic Determination of Thermodynamic Characteristics of Infinitely Diluted Hexane, Cyclohexane, and Benzene Solutions. II, Trudy po Khimii i Khimicheskoi Tekhnologii 2 (1966) 247–254.

[135] Z. Gu, W. Zhao, Y. Zheng, J. Shi, unknown title, J. Chem. Eng. Chin. Univ. 5 (1991) 79.

[136] D. Arnold, M. Juang, K. Morgan, Infinite-Dilution Activity Coefficients of Organic Solutes in Fluorene and Dibenzofuran, J. Chem. Eng. Data 34 (1989) 161–165.

[137] K. Marsh, J. Ott, M. Costigan, Excess Enthalpies, Excess Volumes, and Excess Gibbs Free Energies for (*n*-Hexane + *n*-Decane) at 298.15 and 308.15 K, J. Chem. Thermodyn. 12 (1980) 343–348.

[138] J. Ott, K. Marsh, R. Stokes, Excess Enthalpies, Excess Volumes, and Excess Gibbs Free Energies for (*n*-Hexane + *n*-Dodecane) at 298.15 K and 308.15 K, J. Chem. Thermodyn. 13 (1981) 371–376.

[139] M. L. McGlashan, K. Morcom, Heats of Mixing of Some *n*-Alkanes, Trans. Faraday Soc. 57 (1961) 907–913.

[140] F. Kimura, G. C. Besnon, C. J. Halpin, Excess Enthalpies of Binary Mixtures of *n*-Heptane with Hexane Isomers, Fluid Phase Equilib. 11 (1983) 245–250.

[141] V. T. Lam, P. Picker, D. Patterson, P. Tankrede, Thermodynamic Effects of Orientational Order in Chain-Molecule Mixtures. 1. Heats of Mixing of Globular and Normal Alkanes, J. Chem. Soc. Faraday Trans. 70 (1974) 1465–1478.

[142] S. E. M. Hamam, M. K. Kumaran, G. C. Benson, Excess Enthalpies of Binary Mixtures of *n*-Octane with Each of the Hexane Isomers at 298.15 K, Fluid Phase Equilib. 18 (1984) 147–154.

[143] A. Heintz, R. Lichtenthaler, Excess Enthalpies of Liquid Alkane Mixtures at Pressures up to 500 bar I. Experimental Results, Ber. Bunsen Ges. Phys. Chem. 84 (1980) 727.

[144] S. E. M. Hamam, G. C. Benson, Excess Enthalpies of Binary Mixtures of *n*-Decane with Hexane Isomers, J. Eng. Chem. Data 31 (1986) 45–47.

[145] J. G. Fernandez-Garcia, C. G. Boissonas, Propriétés thermodynamiques des mélanges binaires. Chaleurs de mélange des *n*-alcanes ainsi que de leurs isomères, Helv. Chim. Acta 50 (1967) 1059–1068.

[146] S. E. M. Hamam, M. K. Kumaran, G. C. Benson, Excess Enthalpies and Excess Volumes of Each of the Mixtures: (*n*-Dodecane + an Isomer of Hexane) at 298.15 K, J. Chem. Thermodyn. 16 (1984) 537.

[147] J. G. Fernandez-Garcia, C. G. Boissonas, Calorimètre pour la mesure des chaleurs de mélange des liquides, Helv. Chim. Acta 49 (1966) 854–858.

[148] T. Holleman, Heats of Mixing of Liquid Binary Normal Alkane Mixtures, Physica 31 (1965) 49–63.

[149] J. A. Larkin, D. V. Fenby, T. S. Gilman, R. L. Scott, Heats of Mixing of Nonelectrolyte Solutions. III. Solutions of the Five Hexane Isomers with Hexadecane, J. Phys. Chem. 70 (1966) 1959–1963.

[150] M. L. McGlashan, K. W. Morcom, Thermodynamics of Mixtures of *n*-Hexane + *n*-Hexadecane. Part 1. Heats of mixing, Trans. Faraday Soc. 57 (1961) 581–587.

[151] K. Marsh, P. P. Organ, Excess Molar Enthalpies and Excess Molar Volumes for Three- and Four-Component *n*-Alkane Mixtures Simulating (*n*-Hexane + *n*-Hexadecane), J. Chem. Thermodyn. 17 (1985) 835–841.

[152] R. C. Miller, A. G. Williamson, Excess Molar Enthalpies for (*n*-Hexane + *n*-Hexadecane) and for Three- and Four-Component Alkane Mixtures Simulating this Binary Mixture, J. Chem. Thermodyn. 16 (1984) 793–799.

[153] M. K. Kumaran, G. C. Benson, Excess Enthalpies of *n*-Dodecane + *n*-Heptane, + *n*-Octane, and + *n*-Decane at 298.15 K, J. Chem. Thermodyn. 18 (1986) 993–996.

[154] G. W. Lundberg, Thermodynamics of Solutions XI. Heats of Mixing of Hydrocarbons, J. Chem. Eng. Data 9 (1964) 193–198.

[155] R. B. Grigg, J. R. Goates, J. Ott, Excess Volumes and Excess Enthalpies for (*n*-Dodecane + *n*-Octane) and Excess Volumes for (*n*-Dodecane + Cyclohexane) at 298.15 K, J. Chem. Thermodyn. 14 (1982) 101–102.

[156] A. Pfennig, 2007, personal communication, 2007.