Universiteit
Leiden
The Netherlands

Pair configurations to molecular activity coefficients: PAC-MAC

Sweere, A.J.M.

Citation

Sweere, A. J. M. (2017, June 14). *Pair configurations to molecular activity coefficients: PAC-MAC*. Retrieved from https://hdl.handle.net/1887/49551

Version:	Not Applicable (or Unknown)
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/49551

Note: To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

The handle http://hdl.handle.net/1887/49551 holds various files of this Leiden University dissertation.

Author: Sweere, A.J.M.
Title: Pair configurations to molecular activity coefficients: PAC-MAC
Issue Date: 2017-06-14

Chapter I

Introduction

Miscibility of molecules plays an important role in various chemical areas ranging from food stabilization[1, 2] to enhanced oil recovery[3, 4] and from drug discovery[5, 6] to plastic development[7-9]. Numerous chemical experiments are available to determine miscibility properties, for example vapor-liquid equilibria (VLE) measurement[10, 11], the "slow-stirring" method[12] and inverse gas chromatography[13]. In case experiments are too time consuming, expensive, dangerous or simply impossible, we need alternative methods to predict the desired properties. Methods related to the prediction of thermodynamic miscibility come under the heading of molecular modeling: an umbrella term for all computational approaches to interpret, predict or replace chemical experiments.[14, 15]

If no assumptions are incorporated into a computational method, then experimental data will be predicted exactly. However, assumptions almost always have to be made in molecular modeling; even an exact solution of the Schrödinger equation for a single helium atom has never been found[16]. Moreover, the calculation of miscibility properties using sophisticated quantum mechanics simulations is only possible for mixtures of small molecules due to long calculation times.[17-20]

Computational methods can be classified into three different approaches: *empirical*, *semi-empirical* and *ab initio*. An *empirical* model is fully based on experimental findings, without any physical foundation. Chemical similarity searching is an example of an *empirical* approach since the calculated property of a molecular system fully depends on experimental properties of comparable systems obtained using experimentally derived similarity and/or distance coefficients.[21] In a *semi-empirical* model, experimental data is used to optimize physics-based functions. For example, the famous B3LYP functional contains 3 parameters fitted to experimental atomization energies, ionization potentials, proton affinities and atomic energies.[22] Within an *ab initio* (from the beginning) model, all introduced simplifications are fully derived from physical principles. Two examples of *ab initio* quantum chemistry methods are Hartree-Fock[23, 24] and Møller-Plesset perturbation theory[25].

The use of experimental data, to optimize a computational method, improves the accuracy and serves as justification of introduced simplifications. Therefore, useful *ab initio* models[26, 27] for the calculation of miscibility properties are rare; *empirical* and *semi-empirical* methods are more common. Various popular methods for the prediction of miscibility properties are described in the remainder of **Chapter I**.

MOLECULAR SIMULATIONS

Molecular Dynamics (MD) and Monte Carlo (MC) simulations use respectively Newton's equations of motion[28] and Boltzmann statistics[29] to simulate a molecular system on the atomistic scale. Interactions between the molecules are calculated using classical force fields to avoid the computational expensive calculation of the electron density.[30] Force fields consist of various pairwise potential energy functions with a corresponding set of parameters and provide an estimate of the potential energy surface derived from quantum mechanics. The parameters in force fields are

mainly derived from a combination of quantum mechanics and experimental data.[31-35] However, even if classical force fields are used, the calculation speed remains an important issue in molecular simulations: the required time to calculate a VLE diagram is often more than one day per chosen mole fraction on a single core.[36, 37] If faster calculations are desired, then assumptions with regard to reducing the number of degrees of freedom are inevitable. Possible solutions involve coarse-graining of the all-atom molecular structure by using softcore beads (DPD[38]), avoiding the explicit calculation of many-body interactions by using correlation functions (3D-RISM-HNC[39]), neglecting long-range interactions (molecular cluster calculations[40]) or calculating solely two-body interactions (pair sampling method[41]). A brief description of the latter two is given below.

Molecular cluster calculations

Akkermans has developed a simple and efficient method to predict Flory-Huggins χ-interaction parameters based on molecular cluster calculations.[40] The Flory-Huggins χ-interaction parameter is a dimensionless parameter related to the excess interaction energy of the two components in a binary mixture.[42, 43] Molecular clusters are generated using a Monte Carlo procedure by placing molecules randomly around a central reference molecule with touching Van der Waals surfaces.[44] Overlapping Van der Waals surfaces of neighboring molecules is avoided in order to obtain a realistic representation of molecular clusters as first coordination shell realizations. Four examples of generated molecular clusters are shown in Figure 1. An expression for the "cluster free energy" is used to calculate miscibility properties, including χ-interaction parameters and energies of mixing, from a sampled set of 10^6 molecular cluster energies calculated using the COMPASS force field[33]. The method is developed to estimate interaction parameters in mesoscale simulations such as DPD[38] and DDFT[45]. Flory-Huggins χ-interaction parameters are shown to be overestimated if the calculation is based solely on mixing energies because the neglected excess entropy of mixing is, in general, positive. With the inclusion of the excess entropy of mixing reasonable agreement with experiment is obtained despite the absence of long-range interactions. Moreover, the random sampling procedure in combination with the avoidance of overlapping neighboring molecules leads to unrealistic holes between the molecules in the coordination shell resulting in an obtained average coordination number being 50% smaller than in MD.[46]

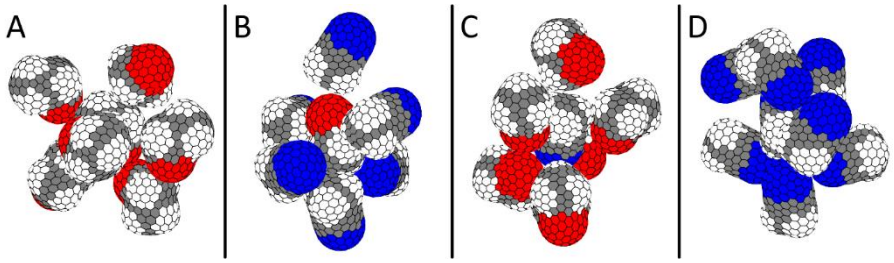

Figure 1. Examples of generated molecular clusters: a reference molecule ethanol surrounded by 6 ethanol molecules (A), a reference molecule ethanol surrounded by 6 acetonitrile molecules (B), a reference molecule acetonitrile surrounded by 6 ethanol molecules (C) and a reference molecule acetonitrile surrounded by 7 acetonitrile molecules (D).

Pair sampling method

Fan et al. have developed a theoretical model for the prediction of miscibility behavior of binary mixtures using a combination of Flory-Huggins theory and Monte Carlo sampling of molecular pair configurations.[41] The pair configurations are obtained, in accordance with molecular cluster calculations, by placing molecules randomly around a central molecule with touching Van der Waals surfaces.[44] Four examples of sampled pair configurations are shown in Figure 2. The coordination number is calculated by packing neighboring molecules around a central molecule without overlapping Van der Waals surfaces. Again, in accordance with molecular cluster calculations, the avoidance of overlapping neighboring molecules leads to unrealistic holes between the molecules resulting in small coordination numbers: around 7 on average. The obtained average coordination number and the Boltzmann weighted average intermolecular energy of the sampled pair configurations, calculated using classical force fields, for all combinations of two molecules are subsequently required for the calculation of χ-interaction parameters using the original Flory-Huggins lattice theory.[42, 43] Fan's pair sampling method is very similar to Akkermans' molecular cluster calculations; the two main differences involve the sampling of pair configurations instead of cluster configurations and the calculation of χ-interaction parameters using Flory-Huggins theory instead of a formulation of the free energy.[40] A simple Boltzmann weighted average over pair energies will overestimate the probability of favorable pair configurations because steric effects, included in cluster configurations, are neglected. The pair sampling method is incorporated in the Blends module of BIOVIA Materials Studio.[47]

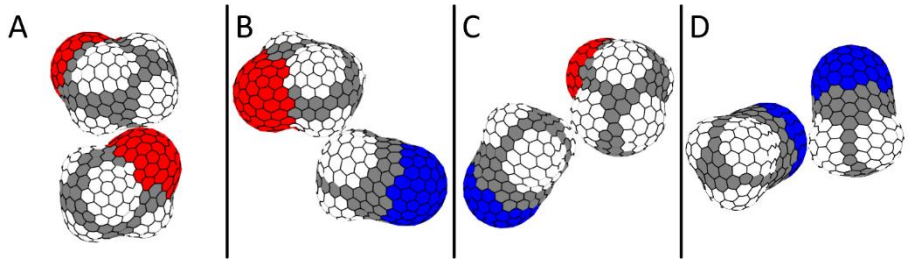

Figure 2. Examples of sampled pair configurations: ethanol – ethanol (A), ethanol – acetonitrile (B), acetonitrile – ethanol (C) and acetonitrile – acetonitrile (D).

QUASI-CHEMICAL APPROXIMATION

In 1944, Edward A. Guggenheim derived a combinatorial expression to approximate the pair distribution of adjacent molecules in a mixture of molecules on a lattice.[48] The resulting quasi-chemical approximation assumes the pairs of adjacent molecules to be Boltzmann distributed with the added constraint stating the total probability to obtain a molecular component in the distribution of pairs to be equal to the molecular fraction in the mixture. The quasi-chemical approximation is an improvement of the regular solution model.[49] The regular solution model assumes molecules in a mixture to be randomly distributed whereas the quasi-chemical approximation takes a non-random distribution of neighboring molecules into account in non-ideal solutions: energetically favorable pairs of adjacent molecules have a higher probability of occurrence than energetically unfavorable pairs of molecules. The expression is similar to the equation for the equilibrium concentrations of molecules in reversible chemical reactions. Two well-known *semi-empirical* models based on the quasi-chemical approximation of Guggenheim are UNIFAC[50] and COSMO-RS[51]. Both models are briefly explained below.

UNIFAC

The UNIQUAC Functional-group Activity Coefficients (UNIFAC) model is a group contribution method for the prediction of activity coefficients in nonelectrolyte multicomponent liquid mixtures.[50] The Universal Quasi-Chemical (UNIQUAC) equation is an activity coefficient model using two unary parameters per component in the mixture and two binary parameters per pair of components in the mixture.[52] UNIFAC combines UNIQUAC with a group contribution approach[53] by replacing the molecular unary and binary parameters with functional group unary and binary parameters. Group contribution methods are based on the "additive principle" stating a molecular property to be the summation of individual contributions of the functional groups.[54] The parameters within the method are optimized using experimental data and have little physical meaning.[55] The use of functional group parameters gives UNIFAC, in contrast to UNIQUAC, the ability to predict phase equilibria if experimental data is absent. Examples of other activity

coefficient models using a group contribution approach are GC-NRTL[56] and F-SAC[57, 58].

COSMO-RS

The Conductor-like Screening Model for Real Solvents (COSMO-RS) uses conductor screening charges on molecular surface panels for the prediction of activity coefficients in multicomponent liquid mixtures.[51] The screening charges are obtained by quantum mechanical calculations of single molecules in a perfect conductor. Tessellation of the molecular surface is based on the GEPOL algorithm.[59] The calculation of activity coefficients from the distribution of screening charges, a so-called σ-profile, of all molecules in the mixture is, in accordance with UNIFAC, based on the quasi-chemical approximation of Guggenheim.[48] In COSMO-RS, molecular mixtures are considered to be a collection of independent interacting surface panels.[60] A hypothetical relation between the 'surface-interaction energy density' of a pair of adjacent panels and the screening charges of both panels is formulated by Klamt. COSMO-RS has a much stronger physical basis than UNIFAC, resulting in far less empirical parameters: COSMO-RS contains only three universal parameters obtained by fitting experimental VLE data, indicating the importance of the molecular charge distribution in miscibility properties. The philosophy of COSMO-RS is copied by Lin and Sandler in the development of a very similar COSMO-SAC method.[61] The concept of COSMO-RS and COSMO-SAC is graphically illustrated in Figure 3.

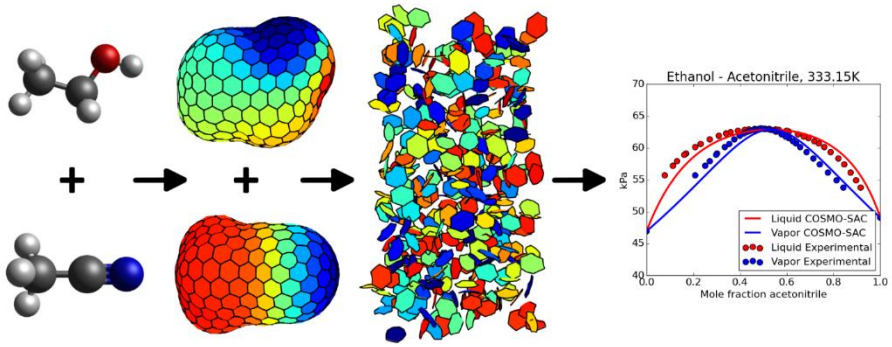

Figure 3. Graphical illustration of COSMO-RS and COSMO-SAC. Quantum mechanics is used to calculate screening charges on molecular surface panels. Subsequently, thermodynamic miscibility properties are obtained by assuming molecular mixtures to be a collection of independent interacting charged surface panels. COSMO-SAC data is extracted from the COSMO-SAC 2013 software developed by Sandler research group[62] and experimental data is obtained from Compostizo et al.[63]

QSAR

Quantitative Structure-Activity Relationship (QSAR) models and Quantitative Structure-Property Relationship (QSPR) models relate quantifiable molecular structural characteristics to respectively activities or other properties of molecular systems.[64] Four examples of structural characteristics are the number of carbon atoms[65], topological indices[66], functional groups[50, 53] and σ-profiles[51, 61]. A physical interpretation of the desired activities is sometimes completely absent in QSAR models, only structural characteristics as molecular descriptors with a possible influence are important.[67] Therefore, QSAR models are useful methods if molecular systems are too complex for a physical description. A reliable pre-defined training set and relevant molecular descriptors are two essential requirements for a useful QSAR.[68] The training set is used for calibration of the model and requires accurate measured activities of well-defined molecular systems. A proper selection of descriptors is important to avoid the presence of redundant molecular characteristics highlighting peculiarities of individual molecules and, as a consequence, losing track of major general tendencies for accurate predictions outside the training set.[69] The selection of descriptors in QSAR models can be performed using a statistical analysis or using a physical analysis.[70] In the statistical analysis, descriptors with a high correlation with the desired property are selected from a large set of possible descriptors, resulting in *empirical* QSAR models.[71] In the physical analysis, descriptors are selected based on justifiable physical principles, resulting in *semi-empirical* QSAR models.[66, 72, 73] Increasing the number of molecular descriptors and corresponding parameters always results in a better fit of the training set. However, increasing the number of descriptors and/or parameters not necessarily results in improved predictability outside the training set; the use of descriptors with low correlation to the desired property, collinearity between multiple descriptors and over-fitting of data are three examples of pitfalls resulting in reduced predictive capacity.[74] The weaknesses of QSAR models are expressed by a famous saying of John von Neumann: "With four parameters I can fit an elephant, and with five I can make him wiggle his trunk."[75, 76] As a rule of thumb, the number of data points should be at least 5 times bigger than the number of molecular descriptors.[77] Moreover, interpretability of the model is desired to avoid the use of QSAR as a "black box".[74, 78, 79] UNIFAC is basically an example of a QSAR model since functional groups are structural characteristics[68, 69] of molecules and the corresponding parameters are optimized using experimental data.

CHEMICAL SIMILARITY SEARCHING

Within chemical similarity searching, molecular properties are determined using experimental properties of comparable molecules.[21] The similar property principle of Johnson and Maggiora is the basis of chemical similarity searching and states: "Similar compounds have similar properties."[80] Two requirements are necessary for similarity searching: a large database with molecules and corresponding properties

and a protocol or expression to measure the similarity. The similarity measurement is performed by comparing structural descriptors of the molecule with corresponding descriptors of the molecules in the database. An example of a popular structural descriptor is a molecular fingerprint. A molecular fingerprint is a binary bit string representing the presences of a list of structural characteristics. Tanimoto coefficients are often used to quantify the similarity between two molecules.[81] A Tanimoto coefficients is comparable to the correlation coefficient between the fingerprints corresponding to the two molecules. A physical interpretation of the desired properties is completely absent in *empirical* chemical similarity searching. Therefore, chemical similarity searching is, in accordance with QSAR models, a useful method if molecular systems are too complex for a physical description, for example in the calculation of receptor binding affinities.[82]

PAC-MAC

Although all methods described above are well established in the field of miscibility prediction, they all have their drawbacks. Molecular simulations are computational expensive, the foregoing methods based on the quasi-chemical approximation lack an explicit incorporation of the molecular structure and *empirical* methods do not have the physical basis to be predictive outside the used experimental database. A useful *semi-empirical* model for the prediction of miscibility properties should comply with two essential requirements. Firstly, the model should be based on the underlying physics as much as possible for a good predictability outside the training set. Secondly, the model should be optimized using enough reliable experimental data for a good accuracy inside the training set. From our perspective, a combination of the strengths of multiple methods offers opportunities in the development of more sophisticated models. Moreover, the continuous increase of processor speed and RAM memory enables possibilities for new, high computational demanding, methods. Both reasons were the cause for our efforts in the development of the Pair Configurations to Molecular Activity Coefficients (PAC-MAC) model.[46] PAC-MAC is a force field based quasi-chemical method for the prediction of miscibility properties of multicomponent mixtures. Molecular structures are explicitly included in PAC-MAC by sampling molecular pair configurations based on Fan's pair sampling method.[41] Subsequently, molecular activity coefficients of multicomponent mixtures are obtained by minimization of an expression for the free energy, based on Guggenheim's quasi-chemical approximation[48], similar to UNIFAC[50] and COSMO-RS[51]. Realistic packing of the molecules is obtained with the addition of a constraint stating the molecular surface to be fully surrounded by neighboring molecules. PAC-MAC is very similar to Akkermans' molecular cluster calculations; the two main differences involve the realistic packing of molecules using a constraint instead of cluster configurations and the formulation of the free energy based on Guggenheim's quasi-chemical approximation instead of Boltzmann.[40] Furthermore, with the introduction of a few

tunable parameters optimized using experimental data, similar to QSAR[64], higher accuracy is obtained.

The PAC-MAC method is described in full detail in **Chapter II**. A proof of concept is given by comparing various calculated VLE diagrams with experimental data and results obtained using COSMO-SAC and Monte Carlo simulations. PAC-MAC is shown to be orders of magnitude faster than molecular simulations.

A comprehensive accuracy test is presented in **Chapter III** by comparing calculated free energies of mixing with an experimental database containing 1092 binary mixtures. The introduction of two adjustments to the model, regarding partial occupation of the molecular surface and a modification of the Staverman-Guggenheim free energy correction term, results in a significant reduction of the root mean squared error (RMSE). Nevertheless, the competitive UNIFAC model still outperforms PAC-MAC because the binary parameters in UNIFAC are optimized using experimental miscibility data whereas the force field parameters in PAC-MAC are not.

The treatment of polymers within the PAC-MAC model is extensively discussed in **Chapter IV**. High correlation is obtained in a comparison of calculated Flory Huggins χ-interaction parameters with experimental data for 779 polymer solutions. Predictions of solvent activities are of similar accuracy as with the COSMO-RS model.

Because PAC-MAC depends on the used force field, accurate force field parameters are of high relevance. The accuracy of the OPLS-AA force field[32] is tested in **Chapter V** by calculating free energies of mixing using thermodynamic integration (TI) of MD simulations. The RMSE of PAC-MAC is shown to be only slightly larger than obtained using TI of MD simulations, despite the addition of several assumptions. Moreover, by optimizing 50 force field parameters the RMSE of PAC-MAC is reduced below 0.1 k_BT.

REFERENCES

(1) Dickinson, E. Food Emulsions and Foams: Stabilization by Particles. *Curr. Opin. Colloid Interface Sci.* **2010**, *15*, 40-49.

(2) Dickinson, E. Use of Nanoparticles and Microparticles in the Formation and Stabilization of Food Emulsions. *Trends Food Sci. Tech.* **2012**, *24*, 4-12.

(3) Thomas, S. Enhanced Oil Recovery - An Overview. *Oil Gas Sci. Technol.-Rev. IFP* **2008**, *63*, 9-19.

(4) Kokal, S.; Al-Kaabi, A. In *Enhanced Oil Recovery: Challenges & Opportunities*, World Petroleum Council, 2010; 2010; pp 64-69.

(5) Laitinen, R.; Löbmann, K.; Strachan, C. J.; Grohganz, H.; Rades, T. Emerging Trends in the Stabilization of Amorphous Drugs. *Int. J. Pharm.* **2013**, *453*, 65-79.

(6) Lu, Y.; Park, K. Polymeric Micelles and Alternative Nanonized Delivery Vehicles for Poorly Soluble Drugs. *Int. J. Pharm.* **2013**, *453*, 198-214.

(7) Sionkowska, A. Current Research on the Blends of Natural and Synthetic Polymers as New Biomaterials: Review. *Prog. Polym. Sci.* **2011**, *36*, 1254-1276.

(8) Li, J.; Shanks, R. A.; Long, Y. Miscibility and Crystallisation of Polypropylene–Linear Low Density Polyethylene Blends. *Polymer* **2001**, *42*, 1941-1951.

(9) Zhao, L.; Choi, P. A Review of the Miscibility of Polyethylene Blends. *Mater. Manuf. Processes* **2006**, *21*, 135-142.

(10) Kamihama, N.; Matsuda, H.; Kurihara, K.; Tochigi, K.; Oba, S. Isobaric Vapor–Liquid Equilibria for Ethanol + Water + Ethylene Glycol and Its Constituent Three Binary Systems. *J. Chem. Eng. Data* **2012**, *57*, 339-344.

(11) Wohlfarth, C. Polymer Solutions. In *Landolt-Börnstein - Group VIII Advanced Materials and Technologies*, Lechner, M. D.; Arndt, K. F., Eds. Springer: Berlin, Germany, 2009; Vol. 6D1, p 3890.

(12) De Bruijn, J.; Busser, F.; Seinen, W.; Hermens, J. Determination of Octanol/Water Partition Coefficients for Hydrophobic Organic Chemicals with the "Slow-Stirring" Method. *Environ. Toxicol. Chem.* **1989**, *8*, 499-512.

(13) Voelkel, A.; Fall, J. Inverse Gas Chromatography. Relationship between Mass Activity Coefficient and Flory-Huggins Interaction Parameter in the Examination of Petroleum Pitches. *Chromatographia* **1997**, *44*, 197-204.

(14) Yates, J. T.; Johnson, J. K. *Molecular Physical Chemistry for Engineers*; University Science Books: Sausalito, California, 2007.

(15) Comba, P. *Modeling of Molecular Properties*; Wiley-VCH: Weinheim, Germany, 2011.

(16) Nakashima, H.; Nakatsuji, H. Solving the Schrödinger Equation for Helium Atom and its Isoelectronic Ions with the Free Iterative Complement Interaction (ICI) Method. *J. Chem. Phys.* **2007**, *127*, 224104.

(17) Kirchner, B.; Hutter, J. The Structure of a DMSO–Water Mixture from Car–Parrinello Simulations. *Chem. Phys. Lett.* **2002**, *364*, 497-502.

(18) Saharay, M.; Balasubramanian, S. Electron Donor−Acceptor Interactions in Ethanol−CO2 Mixtures: An ab Initio Molecular Dynamics Study of Supercritical Carbon Dioxide. *J. Phys. Chem. B* **2006**, *110*, 3782-3790.
(19) Glezakou, V.-A.; Rousseau, R.; Dang, L. X.; McGrail, B. P. Structure, Dynamics and Vibrational Spectrum of Supercritical CO2/H2O Mixtures from ab Initio Molecular Dynamics as a Function of Water Cluster Formation. *Phys. Chem. Chem. Phys.* **2010**, *12*, 8759-8771.
(20) Wang, C.; He, X.-T.; Zhang, P. Thermophysical Properties of Hydrogen-Helium Mixtures: Re-examination of the Mixing Rules via Quantum Molecular Dynamics Simulations. *Phys. Rev. E* **2013**, *88*, 033106.
(21) Willett, P.; Barnard, J. M.; Downs, G. M. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983-996.
(22) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648-5652.
(23) Hartree, D. R. The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods. *Math. Proc. Cambridge Philos. Soc.* **1928**, *24*, 89-110.
(24) Fock, V. Näherungsmethode zur Lösung des Quantenmechanischen Mehrkörperproblems. *Z. Physik* **1930**, *61*, 126-148.
(25) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618-622.
(26) Liu, Y.-P.; Kim, K.; Berne, B. J.; Friesner, R. A.; Rick, S. W. Constructing ab Initio Force Fields for Molecular Dynamics Simulations. *J. Chem. Phys.* **1998**, *108*, 4739-4755.
(27) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. Development of an Accurate and Robust Polarizable Molecular Mechanics Force Field from ab Initio Quantum Chemistry. *J. Phys. Chem. A* **2004**, *108*, 621-627.
(28) Newton, I. *Philosophiæ Naturalis Principia Mathematica*; Jussu Societatis Regiæ ac Typis Josephi Streater. Prostat apud plures Bibliopolas.: Londini, 1687.
(29) Boltzmann, L. *Studien über das Gleichgewicht der Lebendigen Kraft zwischen Bewegten Materiellen Punkten*; Wien, 1868.
(30) Monticelli, L.; Tieleman, D. P. Force Fields for Classical Molecular Dynamics. In *Biomolecular Simulations: Methods and Protocols*, Monticelli, L.; Salonen, E., Eds. Humana Press: Totowa, New Jersey, 2013; Vol. 924, pp 197-213.
(31) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. DREIDING: a Generic Force Field for Molecular Simulations. *J. Phys. Chem.* **1990**, *94*, 8897-8909.
(32) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225-11236.
(33) Sun, H. COMPASS: An ab Initio Force-Field Optimized for Condensed-Phase Applications - Overview with Details on Alkane and Benzene Compounds. *J. Phys. Chem. B* **1998**, *102*, 7338-7364.

(34) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765-784.
(35) Martin, M. G.; Siepmann, J. I. Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes. *J. Phys. Chem. B* **1998**, *102*, 2569-2577.
(36) Ferrando, N.; Lachet, V.; Boutin, A. Monte Carlo Simulations of Mixtures Involving Ketones and Aldehydes by a Direct Bubble Pressure Calculation. *J. Phys. Chem. B* **2010**, *114*, 8680-8688.
(37) Smith, W. R.; Figueroa-Gerstenmaier, S.; Skvorova, M. Molecular Simulation for Thermodynamic Properties and Process Modeling of Refrigerants. *J. Chem. Eng. Data* **2014**, *59*, 3258-3271.
(38) Groot, R. D.; Warren, P. B. Dissipative Particle Dynamics: Bridging the Gap Between Atomistic and Mesoscopic Simulation. *J. Chem. Phys.* **1997**, *107*, 4423-4435.
(39) Beglov, D.; Roux, B. An Integral Equation To Describe the Solvation of Polar Molecules in Liquid Water. *J. Phys. Chem. B* **1997**, *101*, 7821-7826.
(40) Akkermans, R. L. C. Mesoscale Model Parameters from Molecular Cluster Calculations. *J. Chem. Phys.* **2008**, *128*, 244904.
(41) Fan, C. F.; Olafson, B. D.; Blanco, M.; Hsu, S. L. Application of Molecular Simulation to Derive Phase Diagrams of Binary Mixtures. *Macromolecules* **1992**, *25*, 3667-3676.
(42) Flory, P. J. Thermodynamics of High Polymer Solutions. *J. Chem. Phys.* **1942**, *10*, 51-61.
(43) Huggins, M. L. Some Properties of Solutions of Long-Chain Compounds. *J. Phys. Chem.* **1942**, *46*, 151-158.
(44) Blanco, M. Molecular Silverware. I. General Solutions to Excluded Volume Constrained Problems. *J. Comput. Chem.* **1991**, *12*, 237-247.
(45) Fraaije, J. G. E. M.; van Vlimmeren, B. A. C.; Maurits, N. M.; Postma, M.; Evers, O. A.; Hoffmann, C.; Altevogt, P.; Goldbeck-Wood, G. The Dynamic Mean-Field Density Functional Method and its Application to the Mesoscopic Dynamics of Quenched Block Copolymer Melts. *J. Chem. Phys.* **1997**, *106*, 4260-4269.
(46) Sweere, A. J. M.; Fraaije, J. G. E. M. Force-Field Based Quasi-Chemical Method for Rapid Evaluation of Binary Phase Diagrams. *J. Phys. Chem. B* **2015**, *119*, 14200-14209.
(47) Akkermans, R. L. C.; Spenley, N. A.; Robertson, S. H. Monte Carlo Methods in Materials Studio. *Mol. Simul.* **2013**, *39*, 1153-1164.
(48) Guggenheim, E. A. Statistical Thermodynamics of Mixtures with Non-Zero Energies of Mixing. *Proc. R. Soc. London, Ser. A* **1944**, *183*, 213-227.
(49) Hildebrand, J. H. Solubility. XII. Regular Solutions. *Journal of the American Chemical Society* **1929**, *51*, 66-80.

(50) Fredenslund, A.; Jones, R. L.; Prausnitz, J. M. Group-Contribution Estimation of Activity Coefficients in Nonideal Liquid Mixtures. *AIChE J.* **1975**, *21*, 1086-1099.
(51) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995**, *99*, 2224-2235.
(52) Abrams, D. S.; Prausnitz, J. M. Statistical Thermodynamics of Liquid Mixtures: A New Expression for the Excess Gibbs Energy of Partly or Completely Miscible Systems. *AIChE J.* **1975**, *21*, 116-128.
(53) Van Krevelen, D. W.; Te Nijenhuis, K. *Properties of Polymers: Their Correlation with Chemical Structure; their Numerical Estimation and Prediction from Additive Group Contributions*; Elsevier: Amsterdam, the Netherlands, 2009; p 1030.
(54) Kolská, Z.; Randova, A.; Zábranský, M. Group Contribution Methods for Estimation of Selected Physico-Chemical Properties of Organic Compounds. In *Thermodynamics - Fundamentals and Its Application in Science*, Morales-Rodriguez, R., Ed. InTech Open Access Publisher: 2012; pp 135-162.
(55) da Silva, E. F. Modelling of Activity Coefficients using Computational Chemistry. Norwegian University of Science and Technology, Trondheim, Norway, 2002.
(56) Bouneb, N.; Meniai, A. H.; Louaer, W. Introduction of the Group Contribution Concept into the NRTL Model. *Energy Procedia* **2012**, *18*, 1182-1188.
(57) Soares, R. d. P.; Gerber, R. P. Functional-Segment Activity Coefficient Model. 1. Model Formulation. *Ind. Eng. Chem. Res.* **2013**, *52*, 11159-11171.
(58) Soares, R. d. P.; Gerber, R. P.; Possani, L. F. K.; Staudt, P. B. Functional-Segment Activity Coefficient Model. 2. Associating Mixtures. *Ind. Eng. Chem. Res.* **2013**, *52*, 11172-11181.
(59) Silla, E.; Tuñón, I.; Pascual-Ahuir, J. L. GEPOL: An Improved Description of Molecular Surfaces II. Computing the Molecular Area and Volume. *J. Comput. Chem.* **1991**, *12*, 1077-1088.
(60) Klamt, A. *COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*, 1st ed.; Elsevier: Amsterdam, the Netherlands, 2005.
(61) Lin, S.-T.; Sandler, S. I. A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model. *Ind. Eng. Chem. Res.* **2002**, *41*, 899-913.
(62) Xiong, R.; Sandler, S. I.; Burnett, R. I. An Improvement to COSMO-SAC for Predicting Thermodynamic Properties. *Ind. Eng. Chem. Res.* **2014**, *53*, 8265-8278.
(63) Compostizo, A.; Crespo-Colín, A.; Rubio, R. G. Isothermal Vapor−Liquid Equilibrium of Binary Mixtures Containing 1-Chlorobutane, Ethanol, or Acetonitrile. *J. Chem. Eng. Data* **2009**, *54*, 613-618.

(64) Hansch, C.; Maloney, P. P.; Fujita, T.; Muir, R. M. Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients. *Nature* **1962**, *194*, 178-180.
(65) Egloff, G.; Sherman, J.; Dull, R. B. Boiling Point Relationships among Aliphatic Hydrocarbons. *J. Phys. Chem.* **1940**, *44*, 730-745.
(66) Randic, M. Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609-6615.
(67) Nantasenamat, C.; Isarankura-Na-Ayudhya, C.; Naenna, T.; Prachayasittikul, V. A Practical Overview of Quantitative Structure-Activity Relationship. *EXCLI J.* **2009**, *8*, 74-88.
(68) Le, T.; Epa, V. C.; Burden, F. R.; Winkler, D. A. Quantitative Structure–Property Relationship Modeling of Diverse Materials Properties. *Chem. Rev.* **2012**, *112*, 2889-2919.
(69) Shahlaei, M. Descriptor Selection Methods in Quantitative Structure–Activity Relationship Studies: A Review Study. *Chem. Rev.* **2013**, *113*, 8093-8103.
(70) Perumareddi, J. R.; Sastri, V. S.; Roberge, P. R. Selection of Corrosion Inhibitors Based on Structural Features. In *Materials Performance Maintenance*, Pergamon: Oxford, 1991; pp 195-206.
(71) Golbraikh, A.; Tropsha, A. Beware of q2! *J. Mol. Graph. Model.* **2002**, *20*, 269-276.
(72) Kopp, H. Ueber die Vorausbestimmung einiger Physikalischen Eigenschaften bei Mehreren Reihen Organischer Verbindungen. *Justus Liebigs Ann. Chem.* **1842**, *41*, 169-189.
(73) Tanji, K. K.; Sullivan, J. J. *QSAR Analysis of the Chemical Hydrolysis of Organophosphorus Pesticides in Natural Waters*; W-843; University of California: Davis, 1995; p 38.
(74) Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; Kuz'min, V. E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Terfloth, L.; Gasteiger, J.; Richard, A.; Tropsha, A. QSAR Modeling: Where Have You Been? Where Are You Going To? *J. Med. Chem.* **2014**, *57*, 4977-5010.
(75) Dyson, F. A Meeting with Enrico Fermi. *Nature* **2004**, *427*, 297.
(76) Mayer, J.; Khairy, K.; Howard, J. Drawing an Elephant with Four Complex Parameters. *Am. J. Phys.* **2010**, *78*, 648-649.
(77) Zhou, Y.-x.; Xu, L.; Wu, Y.-p.; Liu, B.-l. A QSAR Study of the Antiallergic Activities of Substituted Benzamides and their Structures. *Chemometr. Intell. Lab. Syst.* **1999**, *45*, 95-100.
(78) Bajorath, J. *Chemoinformatics: Concepts, Methods, and Tools for Drug Discovery*; Humana Press: Totowa, New Jersey, 2004; Vol. 275.
(79) QE, G. The Report from the Expert Group on (Quantitative) Structure-Activity Relationships [(Q)SARs] on the Principles for the Validation of (Q)SARs. *Organ. Econ. Co-operation Dev. Paris* **2004**, *49*, 1-206.

(80) Johnson, M. A.; Maggiora, G. M. *Concepts and Applications of Molecular Similarity*; John Wiley & Sons: New York, 1990; p 393.
(81) Tanimoto, T. T. *An Elementary Mathematical Theory of Classification and Prediction*; International Business Machines Corporation: New York, 1958; p 10.
(82) Luo, M.; Wang, X. S.; Tropsha, A. Comparative Analysis of QSAR-based vs. Chemical Similarity Based Predictors of GPCRs Binding Affinity. *Mol. Inf.* **2016**, *35*, 36-41.