

## Systematic Green Solvent Selection for the Hydroformylation of Long-Chain Alkenes

Steffen Linke, Kevin McBride, and Kai Sundmacher\*

Cite This: *ACS Sustainable Chem. Eng.* 2020, 8, 10795–10811

Read Online

ACCESS |



Metrics &amp; More



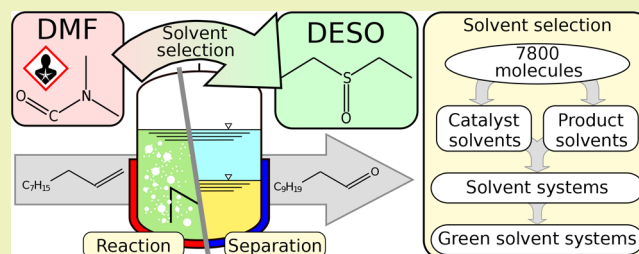
Article Recommendations



Supporting Information

**ABSTRACT:** When selecting solvents, not only thermodynamic criteria but also environmental, health, and safety (EHS) properties should be considered. To this end, we present a fully automated solvent selection approach for the homogeneous catalyzed hydroformylation of 1-decene in a temperature-dependent thermomorphic multicomponent system (TMS) using the conductor-like screening model for real solvents (COSMO-RS). The state-of-the-art TMS uses *N,N*-dimethylformamide (DMF) as the catalyst carrier, which is developmentally toxic, thereby demanding replacement. In order to evaluate the EHS properties holistically, 30 predictive EHS models for 15 different properties are considered in the solvent selection framework, as well as the reliability of the predictions. From a database containing approximately 7800 molecules, the most promising candidates are identified. The best solvent among them, diethyl sulfoxide (DESO), is predicted as outperforming DMF as a catalyst solvent in the hydroformylation. The evidence suggests that DESO may be a broadly applicable environmentally benign substitute solvent for DMF.

**KEYWORDS:** Green chemistry, solvent selection, COSMO-RS, homogeneous catalysis, liquid extraction, computer-aided molecular design



## INTRODUCTION

In 2015, the United Nations passed the “Sustainable Development Goals”,<sup>1</sup> with one of the proposed challenges being the establishment of a sustainable economy. In order to meet these goals, the chemical industry needs to facilitate the use of renewables as feedstock. Biorefineries could potentially meet the global chemical demand without creating a food shortage despite a growing world population.<sup>2</sup> Using biomaterials, it is beneficial to utilize the synthesis power provided by nature instead of gasifying the raw material and beginning chemical synthesis from scratch.<sup>3</sup> These new paths to chemical products require new synthesis methods dealing with liquids while avoiding high temperatures and pressures that tend to degrade the delicate renewable feedstocks. Homogeneous catalysis is a promising technology that can fulfill these constraints resulting in an inherent need for liquid solvents.<sup>4</sup>

Solvents take on a various number of tasks in chemical processes. They can dissolve reactants and catalysts to enable reactions, improve conversion, selectivity, and reaction rates, or facilitate separations such as in liquid extraction.<sup>5</sup> In light of these numerous applications, there is also a desire to select solvents systematically. Ideally, this selection is fast, cost-efficient, and has low experimental effort.

The selection of solvents follows different objectives. First, the general solvent functionality has to be considered: solubilities, separability, miscibility behavior, product stabiliz-

ing abilities, etc.<sup>6</sup> Also, practical criteria have to be considered such as reaction stability of the solvent, suitable viscosity, and appropriate boiling and melting points in order to be in the liquid state. Besides these functional and practical properties, environmental, health, and safety (EHS) criteria are increasingly more important in the selection process.<sup>7</sup> These criteria include, for example, fish toxicity, carcinogenicity, and the flash point. The incorporation of the EHS criteria in the initial stage of solvent selection is additionally motivated by regulative pressure. Within the European Union, the REACH legislation (Registration, Evaluation, and Authorization of Chemicals) defines the legal basis for handling chemicals depending on their EHS properties.<sup>8</sup>

Recently, the hydroformylation of long-chain olefins, which can be produced from renewable feedstocks,<sup>9</sup> was carefully investigated in order to design an efficient process using homogeneous catalysis.<sup>10–15</sup> The process is based on a temperature-controlled switchable solvent system called a thermomorphic multicomponent system (TMS) enabling the reaction and the separation of the product from the catalyst.

Received: April 6, 2020

Revised: May 18, 2020

Published: June 24, 2020



*N,N*-Dimethylformamide (DMF) was proposed initially as a promising solvent candidate by chemical intuition and was later identified as one of the best performing solvents from a thermodynamic point of view.<sup>16,17</sup> However, DMF is a developmental toxic substance, and is included in the REACH list of components of very high concern. Thus, there is a need to find alternative solvents in order to replace it.<sup>18</sup>

Consequently, this work proposes a new solvent selection approach to identify not only the best functional solvents based on their TMS properties but also on their EHS characteristics in a systematic and fully predictive manner. The *in silico* models applied to predict the EHS properties fulfill the requirements by the REACH legislation,<sup>19</sup> which underlines their acceptance in regulative processes. The approach is completely automated from the definition of the molecular space to the generation of a result report and also allows one to flexibly change the screening criteria as the time-consuming calculations results are stored within a database for reuse. The methodology is applied for the hydroformylation of long-chain olefins in order to find a “green” replacement solvent for DMF.

## ■ BACKGROUND

**Solvent Selection Approaches.** A general introduction to the field of molecular design for chemical processes and commonly used methods is given by Adjiman et al.<sup>20</sup> In their work, they present the potential of using the computational molecular design methods in integrated process design. The specific process objective includes various molecular property estimation models that can be tailored to the process needs and desired characteristics (such as EHS criteria). Such process objectives range from identifying molecules with desirable thermodynamic properties,<sup>21</sup> investigating EHS performance<sup>22,23</sup> to evaluating the economic potential of a process where the solvent structure is variable. This combination of solvent selection and process development is known as integrated solvent and process design.<sup>24–29</sup> For the property estimation (see the next section for details), pure QM calculations,<sup>30,31</sup> COSMO-RS/-SAC,<sup>32,33</sup> or group contribution (GC) methods<sup>29,34</sup> are commonly used. For the prediction of EHS properties, other techniques common to the machine learning community are more frequently applied, e.g., *k*-nearest neighbor methods.<sup>19</sup> Molecular design methods for solvent selection have been applied to some specific applications such as reactions<sup>35–37</sup> and separations in general, the latter including adsorption and absorption,<sup>38,39</sup> extraction,<sup>21,40</sup> and crystallization.<sup>41</sup>

Finally, three conceptual procedures can be distinguished for computer-aided solvent selection. First, solvents can be screened using a list or database for identifying alternatives.<sup>17,42</sup> Second, due to the discrete nature of GC methods, mixed-integer programming,<sup>43</sup> generalized disjunctive programming,<sup>44,45</sup> or genetic algorithms<sup>25,46</sup> can be used to compose new molecular structures whose properties can then be evaluated using the corresponding property estimation model. Third, the molecular targeting approach<sup>47</sup> where the parameters of the solvent in the property model are treated as degrees of freedom in a nonlinear optimization problem. The solution in the parameter space is then mapped to known molecules contained in a database or using a group contribution method.<sup>34,48</sup> A more detailed overview of molecular design methods is given by Austin et al. and Ng et al.<sup>49,50</sup>

Regarding CAMD including EHS aspects, Stefanis et al. were among the first to consider the environmental impact in a solvent design problem using a GC model to predict fish toxicity<sup>51</sup> and later ozone depletion potential.<sup>22</sup> As an alternative to the use of GC models, a database approach was proposed where selected EHS criteria are checked after the CAMD problem has been solved. Consequently, it was ensured that the defined EHS requirements are met.<sup>52</sup> Xu and Diwekar discussed the propagation of model uncertainties of three EHS properties to the final selection of solvents and pointed out the significant effect of these uncertainties.<sup>53</sup> In recent years, the simultaneous solvent and process design in combination with EHS properties has been investigated on the one hand;<sup>54,55</sup> on the other hand, one focus has been on the consideration of a larger set of EHS properties so that environmental, health, and safety categories have been explicitly addressed.<sup>54,56–59</sup> Most recently, EHS properties and life cycle assessment descriptors were considered in CAMD for the identification of more sustainable CO<sub>2</sub> capture solvents. A brief overview over the current CAMD tools and the inclusion of EHS properties has been recently published.<sup>60</sup>

When EHS properties are considered in CAMD problems, they are usually not translated in the practically important metrics: the hazard classification of “The Globally Harmonized System of Classification and Labelling of Chemicals” (GHS). These categories form the basis for the handling of chemicals and play an important role in daily life of laboratories and industry. Furthermore, apart from Xu and Diwekar,<sup>53</sup> the reliability of the EHS predictions is not considered in the decision-making process despite the fact that these properties are difficult to predict. Therefore, in this work, we propose a solvent design approach that includes a large set of EHS properties to cover the holistic picture of the EHS performance. These properties are evaluated regarding the GHS categories, and the uncertainty of the predictions is taken into account.

The previous work by McBride et al. is of particular importance for the current contribution. Here, a screening methodology was developed to identify optimal solvents for a TMS used to facilitate the hydroformylation of 1-dodecene. They showed that DMF is one of the most promising candidates for the recovery of the catalyst from a thermodynamic point of view.<sup>17</sup> Despite the extraction power of DMF, a computer-aided process design revealed that a multistage extraction process is still needed in order to ensure that the catalyst recovery is economically feasible.<sup>10</sup> However, using multistage extraction also allows one to explore the use of more environmentally benign solvents. The lower extraction power of a new solvent can be counterbalanced with an increased number of extraction stages. Recently, preliminary results using a new solvent selection methodology for such a process have been presented.<sup>10</sup> The work presented here is the further development of this approach.

**Thermodynamic Property Prediction.** Thermodynamic property predictions, including especially the behavior of mixtures, are prerequisites for computer-aided solvent selection.<sup>20</sup> Group contribution (GC) methods are a well-established concept for predicting fluid phase thermodynamics and are often used for molecular design problems.<sup>49</sup> One of the most widely used representatives is UNIFAC.<sup>61,62</sup> However, for the present work, UNIFAC cannot be applied since the partition of the catalyst–ligand complex is one main issue for

the solvent design. The complex ligand molecule BiPhePhos (Figure 2) cannot be predicted using UNIFAC.

Another popular approach for the prediction of thermodynamic properties is the conductor-like screening model for real solvents (COSMO-RS).<sup>63,64</sup> The continuum solvation model COSMO constructs the surface of a molecule as a cavity of small surface segments and calculates their electrostatic charge from the electron density by quantum mechanical calculation. The charged segments of all molecules in the mixture get paired according to the COSMO-RS theory, and thus, the thermodynamics of nonideal liquid mixtures can be estimated. By treating molecules as a whole in a quantum mechanical calculation, COSMO and COSMO-RS can in principle predict also complex molecules such as BiPhePhos. Therefore, COSMO/COSMO-RS is the chosen method for property estimation. An overview of COSMO and COMSO-RS and a comparison to UNIFAC can be found in Klamt et al.<sup>65–67</sup>

**EHS Property Prediction.** The prediction of EHS properties is commonly based on quantitative structure–activity relationships (QSAR). First, molecular descriptors that correlate with the EHS property of interest are identified using experimental data. These descriptors, e.g., molecular size, polarity, and octanol–water partition coefficient, need to be sourced from some predictive method in order to be used in a computational framework. This means that a mathematical model is built to compute the EHS property with the descriptors as input. The model can take various forms, and many machine learning techniques have been used to derive such models. The simplest example would be linear regression. Modern, more sophisticated approaches include *k*-nearest neighbors, support vector machines, and artificial neural networks, among others.<sup>19</sup>

It is remarkable that such QSAR models are used in the authorization of new chemicals according to the REACH legislation as long as they fulfill certain criteria regarding their development, validation, and transparency. This fact shows a broader acceptance of predictions made using these models, and it seems reasonable to implement them in solvent design problems.

To make predictions of the EHS properties, the VEGA toolbox is used in this work. This software contains 33 QSAR models encompassing 15 different EHS properties, which are given in Table 1. One of the most interesting and useful features of these models is the reliability score.

More specifically, each prediction is assessed according to its prediction quality, which is classified as “experimental”, “good”, “moderate”, or “low”. This valuation is derived based on different criteria specific to the modeling technique. Frequently used criteria are

- Are there certain structures that preferably lead to outliers (e.g., nitrogen groups)?
- Are structurally similar molecules included in the training set?
- Are the molecular descriptors of the predicted molecule within the range of the test set values?

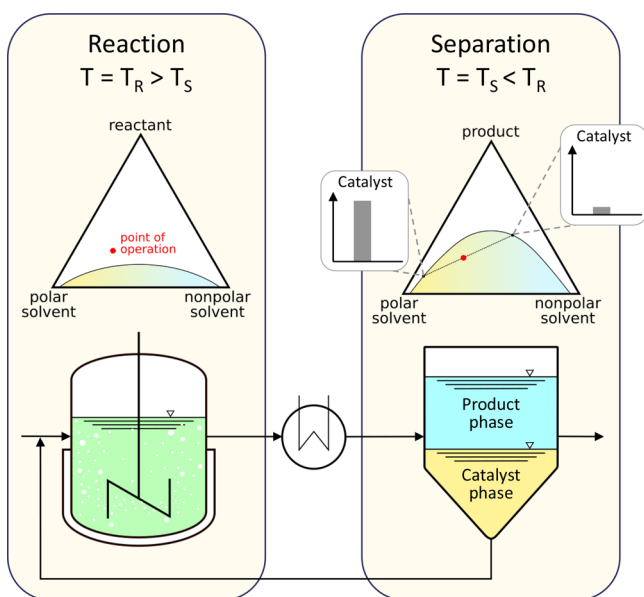
The reliability score is useful in accessing the trustworthiness of a prediction. In particular, when many different EHS properties are considered, there is a risk that many potential solvents will be rejected due to incorrect model predictions, as will be demonstrated in this work. Therefore, the consideration of a reliability measure has a very important benefit to avoid such false rejections.

**Table 1. List of the Models in the VEGA Toolbox to Predict the EHS Properties of Substances**

property	model origin	version	reference
mutagenicity	CONSENSUS	1.0.2	
	CAESAR	2.1.13	68
	SarPy/IRFMN	1.0.7	69
	ISS	1.0.2	70, 71
	KNN/Read-Across	1.0.0	
carcinogenicity	CAESAR	2.1.9	72
	ISS	1.0.2	73–75
	IRFMN/Antares	1.0.0	
	IRFMN/ISSCAN-CSX	1.0.0	
developmental toxicity	CAESAR	2.1.7	76
	PG	1.0.0	77
endocrine disruptor potential	IRFMN	1.0.1	78
	IRFMN/CERAPP	1.0.0	79
skin sensitization	CAESAR	2.1.6	80
hepatotoxicity	IRFMN	1.0.0	81
fish acute toxicity	SarPy/IRFMN	1.0.2	
	KNN/Read-Across	1.0.0	
	NIC	1.0.0	
daphnia toxicity	EPA (96h)	1.0.7	82
	EPA (48h)	1.0.7	83
	DEMETER	1.0.4	83
bee acute toxicity	KNN/IRFMN	1.0.0	84
bioaccumulation factor	CAESAR	2.1.14	85
	Meylan	1.0.3	86
	KNN/Read-Across	1.1.0	
biodegradability	Arnot/EpiSuite	1.0.0	87, 88
	IRFMN	1.0.9	
persistence (sediment)	IRFMN	1.0.0	89
persistence (soil)	IRFMN	1.0.0	89
persistence (water)	IRFMN	1.0.0	89
log <i>P</i> (octanol–water)	Meylan/Kowwin	1.1.4	90
	MLogP	1.0.0	91, 92
	ALogP	1.0.0	93–95

**Thermomorphic Multicomponent Systems (TMSs).** A TMS is composed of at least two solvents that form a single phase at a high temperature and decompose into two liquid phases at a lower temperature. At the higher temperature, a homogeneously catalyzed reaction is carried out in a single fluid phase. This helps to avoid mass diffusion limitations of the reactant into the catalyst phase. Subsequently, the post-reaction mixture is cooled down, leading to its decomposition into two liquid phases. Here, the catalyst and products are separated from each other due to differences in their respective partitioning behaviors. The TMS principle is depicted in Figure 1. The concept of TMS was introduced by Bergbreiter et al.<sup>96</sup> and has since been adapted for numerous homogeneously catalyzed reactions.<sup>97,98</sup> Obviously, the solvents for a TMS should be tailor-made for each reaction in order to fulfill the thermodynamic criteria needed to obtain the desired temperature switchable liquid–liquid equilibrium (LLE). This means that component solvents need to be carefully selected, which in turn requires a guideline for selection. To this end, experimentally based solvent selection approaches for the TMS were proposed by Behr et al.,<sup>99,100</sup> and a computational methodology was proposed by McBride et al.<sup>17</sup> The focus of this work is to identify green solvents for a TMS to facilitate the hydroformylation of long-chain olefins using computer-aided molecular design.

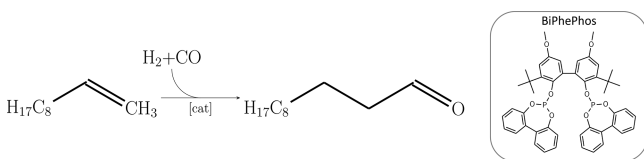




**Figure 1.** Illustration of the TMS principle: The reaction takes place at high temperature  $T_R$  in a single liquid phase (left box); the separation of the catalyst and the products is achieved by a liquid–liquid equilibrium at lower temperature  $T_S$  (right box).

**Hydroformylation of Long-Chain Olefins.** Hydroformylation is the reaction of olefins to aldehydes by the addition of hydrogen and carbon monoxide to the double bond. Short-chain olefins such as *n*-butene are produced efficiently via the Ruhrchemie/Rhône-Poulenc process.<sup>3</sup> In this process, water is used as the catalyst-harboring solvent, in which butene is dispersed as a gas. As the reaction proceeds, the aldehyde product begins to form a second nonpolar phase upon the reaction mixture, from which it is easily decanted. Despite the simplicity and efficiency of this process, it becomes less so with increasing chain length of the olefin reactant due to lower solubility in the aqueous phase. Thus, this concept cannot be applied for the production of long-chain olefins, especially in the case of 1-decene.

To overcome the limitations of poor reactant solubility, a solvent system utilizing the TMS concept is particularly applicable. The catalytic complex used in our previous work<sup>10,17,101</sup> and in this present article is composed of rhodium as the catalytic center and BiPhePhos (BPP) as the ligand. The chemical structure of BPP can be seen on the right-hand side of Figure 2. The ligand steers the selectivity to the desired terminal aldehyde by suppressing the hydroformylation to the branched aldehyde and hydrogenation to the alkane.<sup>102</sup> The left-hand side of Figure 2 shows the desired hydroformylation reaction of 1-decene to 1-undecanal, the example reaction considered in this work.<sup>103,104</sup> Typical reaction conditions are in a temperature range of 105–115 °C<sup>11</sup> and



**Figure 2.** Left: hydroformylation reaction of 1-decene to the terminal aldehyde 1-undecanal with synthesis gas. Right: chemical structure of the ligand BiPhePhos.

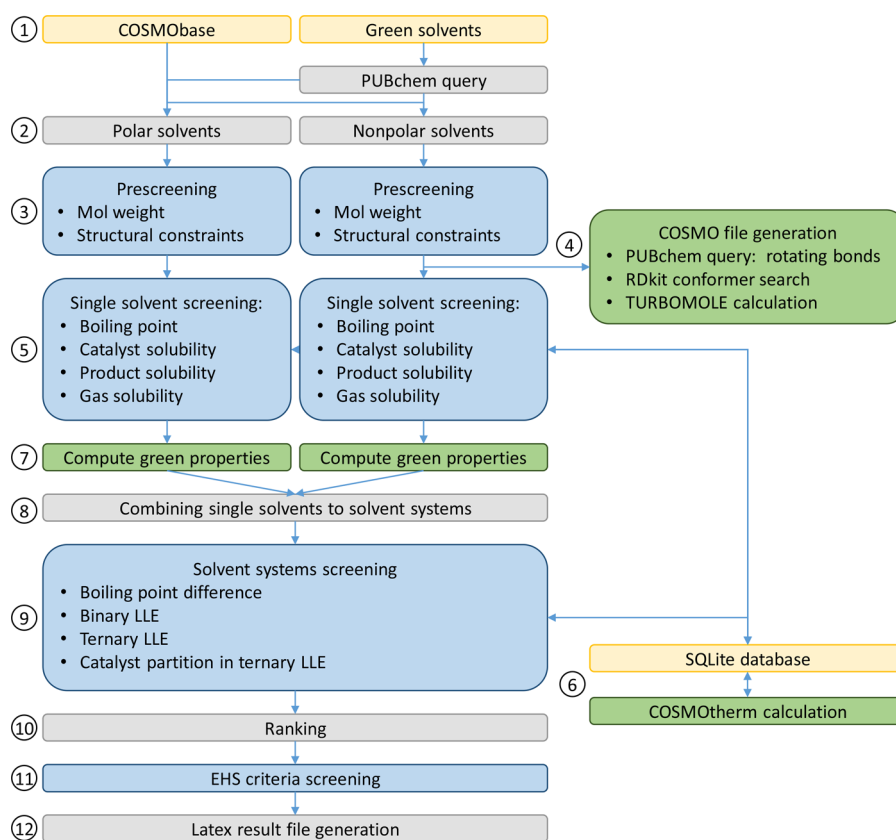
a pressure of 20 bar at an equimolar composition of hydrogen and carbon monoxide.<sup>14</sup>

**Soft- and Hardware.** Quantum chemical density functional theory calculations were performed with TURBOMOLE 7.1<sup>105</sup> using the “calculate” interface (version 2.1, 2009) for molecules that are found in the COSMObase 13.01 database.<sup>106</sup> RDKit (version 2018.09.2) was applied for determining unique SMILES codes and for generating conformers.<sup>107,108</sup> Thermodynamic predictions were made using COSMOtherm with the parameter set “BP\_TZVPD\_FINE\_18.ctd”,<sup>106</sup> and all results were stored in SQLite3 databases. EHS properties were estimated by VEGA (calculation core version: 1.2.4)<sup>109</sup> with the exception of the flash point, which was predicted by COSMOtherm. The interface between Python and VEGA was implemented via KNIME, version 3.5.1.<sup>110</sup> The entire screening methodology is implemented in Python 3.7 and runs on a Linux Ubuntu 16.04 computer (Intel i7-6700 at 3.4 GHz, 8 GB RAM).

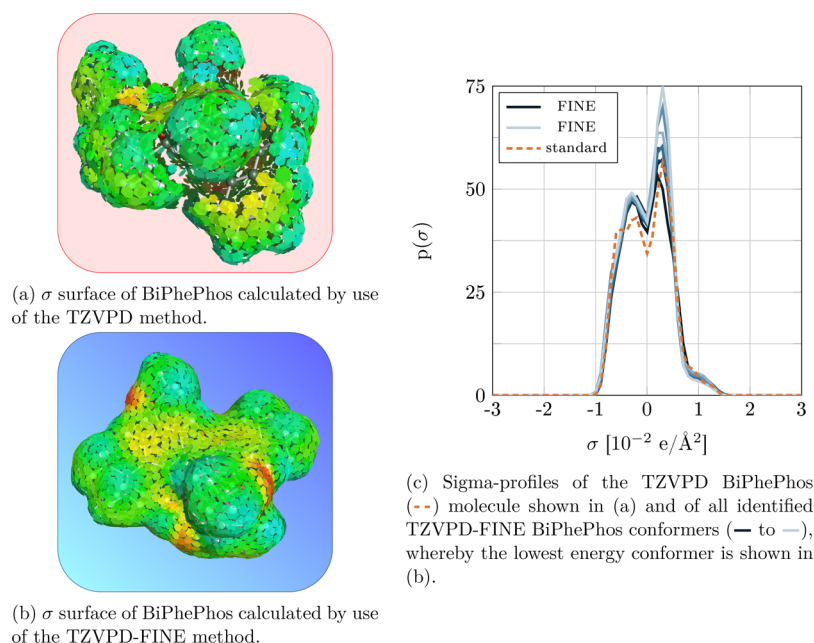
**Screening Methodology.** The solvent selection approach presented here is based on a database search using COSMO-RS for the prediction of thermodynamic properties. The overall goal is to identify solvent systems comprising two solvents, a polar and a nonpolar one, which enable the separation of the catalyst complex from the product of a homogeneous catalyzed reaction according to the TMS working principle. These solvents should also fulfill the selected EHS criteria. The general approach is presented in Figure 3. In the following, we explain the different steps of the screening procedure.

**Modeling of the Catalyst.** Before the solvent screening can start, the catalytic complex needs to be modeled with COSMO in order to predict its thermodynamic behavior using COSMO-RS. The rhodium–BiPhePhos catalyst–ligand complex is shown in Figure 2 on the right-hand side. It should be pointed out that COSMO-RS is the only approach that can predict the thermodynamic behavior of such a complex molecule without using experimental data. The assumption is made that the ligand alone dominates the phase partition behavior of the catalyst–ligand complex and can therefore represent the catalyst complex in the following calculations. This is justified by two reasons. The first is that the overwhelming majority of the catalyst–complex surface belongs to the ligand and not to the metal atom. Second, the ligand is usually always found in excess in the reactor to ensure stable selectivity. This means that each rhodium atom is necessarily attached to a ligand.

The validity of this assumption was already confirmed in previous studies.<sup>17,111</sup> In this same study, one conformer of BiPhePhos was modeled at the BP86-def2-TZVPD level but not using the FINE algorithm that would enable a smoother segmentation of the surface. In the present work, new QM calculations were performed using the FINE method. Additionally, conformers were also identified using RDKit. On the left-hand side of Figure 4, the surface segments of the COSMO models of the ligand using the TZVPD and TZVPD-FINE levels are presented. It is obvious that the FINE method delivers a more accurate shape of the molecule (Figure 4b). This is especially true for the valleys of the surface, which are completely covered with segments, whereas the standard algorithm shows gaps. One has the expectation that the FINE method will provide more accurate results. However, in this specific case, this is not what happens, and the results for the DMF-decane system are given in Table 2, where the partition



**Figure 3.** Overview of the structure of the solvent screening procedure: definition of the databases (step 1), derivation of the solvent design space for the polar and nonpolar solvent candidates (step 2), prescreening of the polar and nonpolar solvent candidates (step 3), procedure for the quantum chemical calculation (step 4), thermodynamic screening of the polar and nonpolar solvent candidates (step 5), prediction of thermodynamic properties using COSMOtherm (step 6), prediction of the environmental, health, and safety properties (step 7), composition of the polar and nonpolar solvents to solvent systems (step 8), thermodynamic screening of the solvent systems using COSMOtherm (step 9), ranking of the finally identified functional solvent systems according to thermodynamic criteria (step 10), environmental, health, and safety screening (step 11), and generation of a result report (step 12).



**Figure 4.** Left: charged segment surfaces of BiPhePhos (BPP) using the standard cavity construction algorithm (a) and the FINE algorithm (b). Right: histograms of the charged surface segments ( $\sigma$ -profile) for the molecule shown in panel (a) and for all eight identified conformers of the molecule in panel (b).

coefficients of the ligand between DMF and *n*-decane at infinite dilution are shown.

**Table 2.** Partition Coefficients  $P = x_{\text{BPP}}^{\text{nonpolar phase}}/x_{\text{BPP}}^{\text{polar phase}}$  of BiPhePhos (Modeled at the TZVPD and TZVPD-FINE Levels) within an LLE Comprising DMF (0.45 g %), *n*-Decane (0.45 g %), and Tridecanal (0.1 g %) at 25 °C

method	BiPhePhos loss [%]	partition coefficient [–]
TZVPD	3.35	0.0531
TZVPD-FINE and multiple conformers	70.70	3.6926
experimental	5 <sup>a</sup>	

<sup>a</sup>Experimental data taken from Brunsch and Behr.<sup>111</sup>

Surprisingly, the ligand modeled at the FINE level prefers the nonpolar phase over the DMF-rich phase. This prediction does not coincide with experimental data.<sup>17</sup> An explanation can be deduced by looking at the  $\sigma$ -profiles of BiPhePhos seen in Figure 4c. The  $\sigma$ -profile is the histogram of the screening charges around the molecule. Using the FINE method, the  $\sigma$ -profile of BiPhePhos contains more nonpolar segments (between  $-0.01$  and  $0.01$  e/Å<sup>2</sup>) (Figure 4c solid lines) than the same molecule calculated using the standard cavity construction algorithm (Figure 4c dashed line). This means that the valleys of BiPhePhos, which are only covered with segments using the FINE method, are rather neutral. Thus, BiPhePhos modeled with FINE will prefer more nonpolar environments in contrast to the molecule with the standard cavity construction. Exactly, this is reflected by the predicted partition coefficients in Table 2. One theory for this observation that the BiPhePhos with the standard cavity construction algorithm provides better results could be that the surface frequently found in the valleys of BiPhePhos does not come into contact with the surrounding molecules. These concave surface parts may not be attainable for surrounding molecules, and therefore, these areas may not contribute significantly to the partition behavior of BiPhePhos. As a result of this analysis, the calculations of BiPhePhos in the following solvent screening method were conducted using the standard cavity construction method.

**Thermodynamic Screening.** The screening begins by defining the molecular design space, which is determined by the available databases (see Figure 3, step 1). In this work, the COSMObase 13.01 database is used along with an additional database of green molecules previously identified by Moity et al.<sup>23</sup> Additionally, the popular DMF replacement solvent Cyrene<sup>112</sup> is included for comparison. For this set of additional molecules, only the SMILES codes are required as input, or if CAS numbers are specified, the SMILES codes are generated from an online query to PubChem.<sup>113</sup> To begin, the databases are homogenized using unique SMILES codes for all molecules generated with RDKit. Subsequently, this database is used as solvent design space for the polar and the nonpolar solvents resulting in two individual lists (Figure 3, step 2).

**Prescreening.** In an initial prescreening step, the molecules in both lists are checked for appropriate molecular weights and structural feasibility (step 3). The structural constraints are checked by looking at each molecule's name and SMILES code. Molecules are excluded for containing specific elements like fluorine (part the green screening due to environmental considerations), for having carbon double bonds (C=C) due to

reactivity, or for containing ions since the TMS concept is based on uncharged molecules as solvents and ions can only be predicted with lower accuracy.

**Quantum Mechanical COSMO Calculation.** After this prescreening, thermodynamic properties such as the boiling points of the solvents are investigated using COSMO-RS. For this, COSMO files need to be generated for the additional molecules not included in the COSMObase database (step 4). In order to obtain results with the same quality as the molecules in COSMObase, conformers are generated for the new molecules using RDKit and the conformer generation strategy proposed by Ebejer et al.<sup>108</sup> Conformer candidates are created by randomly rotating parts of the molecule, then a force field geometry optimization is performed, and repetitive molecular structures are sorted out. It is recommended to couple the number of potentially conformers under investigation with the number of rotating bonds contained within the molecule. Consequently, another PubChem query is performed to determine the number of rotating bonds. Just like in the COSMObase database, the nine conformers with the lowest energy are considered for the DFT calculation as a tradeoff between accuracy and computational time. For the geometry optimization at the QM level, the BP-86 functional with the def-TZVP basis set is used while applying the COSMO boundary condition and the standard COSMO cavity construction. Afterward, a single point calculation with the optimized geometry is performed using the more sophisticated def2-TZVPD basis set and the FINE cavity construction algorithm. The resulting COSMO files are stored in a database to avoid redundancies.

**Single Solvent Screening.** With the COSMO files available, the polar and nonpolar solvents are screened separately (step 5). At first, normal boiling points are predicted, and solvents with a boiling point that is either too high or too low are rejected. The maximum boiling point for the solvents is set at 20 °C below the boiling point of the product undecanal. This is done to reduce the chance of facing an expensive and difficult distillation of the solvents from the product. Second, the solubility of the catalyst is predicted in each solvent at infinite dilution. Hereby, a lower limit is applied for the polar solvents as a screening criterion, and an upper limit is applied for the nonpolar solvents. For the product solubility, it is the other way around: An upper limit is applied for the polar solvents, and a lower limit is applied for the nonpolar solvents. Finally, the gas solubilities of both hydrogen and carbon monoxide are taken into account. This is an important consideration because these gases are consumed in the hydroformylation and also are critical for enabling the active species of the catalyst.<sup>114</sup> The amount of each gas dissolved in the liquid phase for both gases should be as high as possible in both solvents. The result of each calculation is stored in an SQLite database (step 6), which can be conveniently recalled in subsequent screening runs using different rejection criteria.

**Solvent System Screening.** Now, the combination of both solvents into a solvent system comes into play (step 8). The number of polar and nonpolar solvents should be kept within reasonable limits; otherwise, a combinatorial explosion in the number of binary systems may occur. In the current screening approach, this is prevented by the careful screening steps taken for single solvents.

The first screening step for the binary solvent systems is the computation of the liquid–liquid equilibrium. The calculation of the LLE is much slower than the previous calculation of the



thermodynamic properties. The two rejection criteria used here are the general existence of an LLE at the separation temperature and a specific threshold for the cross solubility of the polar solvent into the nonpolar phase. This mutual solubility is taken into account because increasing amounts of polar solvent in the nonpolar phase lead to an increase in the amount of catalyst–ligand leaching into the nonpolar phase. This leaching behavior is exactly what we desire to inhibit. Another screening step investigated was the complete miscibility at reaction temperature. However, reliable predictions of the temperature dependency of the liquid–liquid equilibria are rare, and equilibrium thermodynamics generally tend to overestimate the critical solution point. In principle, such predictions are possible in COSMOtherm, but in order to avoid incorrect decisions that remove interesting solvents, this screening step is not implemented.

After the investigation of the binary LLEs, the ternary LLEs with the product as the third component are calculated. The screening criteria for the rejection of solvent systems are the number of LLE points found within a predefined set of mixture concentrations and the partition coefficient of the product. Finally, for all remaining TMSs, the distribution of the catalyst is calculated between all LLE points of the ternary system at infinite dilution. A threshold for the maximum partition coefficient of all LLE points is used as the screening criterion.

**Environmental, Health, and Safety Screening.** The EHS properties are calculated in step 7 of Figure 3 before the single solvents are combined into binary systems. The VEGA models used for each EHS property are provided in Table 1. VEGA calculations can be made in two ways: either by using the graphical user interface (GUI) or via an application programming interface (API).

Using the GUI enables one to access each model listed in Table 1. For the on-demand calculations made during the screening, the KNIME software provides an API to VEGA enabling external scripts to have access to VEGA. However, the endocrine disruptor (IRFMN/CERAPP) model, the hepatotoxicity model, and the bee toxicity model are not accessible with the latter. Therefore, the GUI was used to make EHS predictions for all molecules in the solvent design space in advance.

The EHS models deliver results in various forms either as continuous values (such as for LD<sub>50</sub>) or categorical outputs (like “mutagenic” or “toxic”). In order to generalize these different responses, a color code is introduced that ranks each result into five classes from best to worst: “green”, “green-yellow”, “yellow”, “yellow-red”, and “red”. The screening criteria for the different EHS properties are then specified relating to that color code. For each property, a specific boundary (such as only being green or yellow-green) can be applied. For the sake of completeness, it must be said that no screening criterion was applied for the octanol–water partition coefficient because this property is only a descriptor for bioaccumulation for which special models are available in VEGA (further information for this is given in the Supporting Information). After evaluating the VEGA models, the EHS characteristics of the remaining solvent candidates must still be checked using their safety data sheets. The VEGA models are very useful as a screening tool, but as with any kind of model, mistakes cannot be excluded.

The EHS screening is the last screening step in the screening process 3, step 11, so that all functional solvents are identified beforehand. This has the benefit that the green solvent systems

can be immediately compared to the functional solvent systems, and the performance loss when using a green solvent is revealed.

**Ensemble Models for Green Properties.** Some of the models in Table 1 describe the same property. Since different modeling approaches and different data are used in these cases, each model contains some useful information and should be considered in the screening procedure. To reduce the results to a single value for each property, models predicting the same property are combined to form a single ensemble model.

The VEGA toolbox already includes an ensemble model for mutagenicity. A weighted hard voting is performed by using the individual model results and weighing them with the predicted reliability. In this publication, similar ensemble models are introduced for all properties in Table 1 predicted by more than a single model. For each one of the end points mutagenicity, carcinogenicity, developmental toxicity, and endocrine disruptor potential, a weighted hard voting between the two categories “hazardous” and “harmless” is performed. For each category, a score  $S_+$  (eq 1) and a score  $S_-$  (eq 2) are computed by translating the reliabilities into numerical values and adding them together. These numerical values are in accordance with the consensus model for mutagenicity included in VEGA and are shown in Table 3.

$$S_+ = \sum R_{i,+} \quad (1)$$

$$S_- = \sum R_{j,-} \quad (2)$$

**Table 3. Numerical Values for the Different Reliabilities to Build the Weighted Hard Voting Ensemble Models for the Green Properties**

reliability	numeric weight R
experimental	1
good	0.9
moderate	0.6
low	0.2

If the condition  $S_+ \geq S_-$  holds, the ensemble response is considered hazardous. Otherwise, it is considered harmless. If experimental results are available, then only these values are taken into account to prevent potentially erroneous predictions from outweighing experimental data.

For fish toxicity, the models KNN/Read-Across, NIC, and EPA (96h) provide a numerical value for the concentration, but the SarPy/IRFMN model predicts categorically. Thus, the numerical responses are classified according to the criteria used to define these categories, and an ensemble model is built analogous to the others. The same procedure is also applied to biodegradability. In the case of daphnia toxicity and the bioaccumulation factor, all available models provide numerical predictions. Therefore, the response of the ensemble model is simply computed as the arithmetic mean using the reliabilities of the individual models as weighting factors.

Similar to the reliability score for the individual property models, the ensemble model for mutagenicity in VEGA introduces an index representing the quality of the ensemble prediction. The so-called consensus score  $C$  indicates if the individual models coincide and if their reliability is either high or low. The summed reliability of the ensemble model result  $S_+$

or  $S_-$  is divided by the total number of models  $n$  taken into account for the voting

$$C = \frac{S_+}{n} \text{ or } C = \frac{S_-}{n} \quad (3)$$

This means that additional corresponding predictions can decrease the ensemble model's reliability. For instance, we assume that an ensemble model is built out of the two individual models, and for a substance, the predictions are "positive" with good and moderate reliability. The consensus score would be 0.75. Expanding the ensemble model with the addition of another individual model delivering a positive prediction of moderate reliability would decrease the consensus score to 0.7. Although a new piece of information supports the ensemble results, the reliability decreases. Instead of using this definition of the consensus score, this publication uses a different method to determine the reliability of the consensus score. The difference of  $S_+$  and  $S_-$  is computed, and this result is assigned a reliability according to Table 3 such that corresponding results do not lead to a decrease in the reliability. For end points with multiple hazard levels, the positive score  $S_+$  includes all the hazard levels, while  $S_-$  includes only the harmless votes. This also leads to a more conservative prediction for the reliability of the ensemble models.

**Overall "Greenness" Score.** After introducing the ensemble models, there is still a set of 15 different properties describing the EHS properties. In order to condense these into one benchmark number, an overall greenness score (OGS) is proposed that shows at a glance how well the EHS criteria are met. All properties contributing to this score are classified into one of the five classes listed in Table 4. Afterward, the

**Table 4. Conversion of Color Codes into Numerical Values for Computing the Overall Greenness Score**

color code	numerical value
green	1
green-yellow	0.75
yellow	0.5
yellow-red	0.25
red	0

numerical representations of the classes are summed up and normalized by the number of models that contributed to the sum. For graphical representation, the OGS can be expressed again by the color code presented in Table 4. Naturally, a screening criterion can also be applied for the OGS according to the color code.

**Ranking of the Solvent Systems and Result Presentation.** The TMS systems identified as being functional (based only on their thermodynamic properties) are considered to be promising candidates for separating the product and recovering the catalyst. If the green screening criteria are applied, then the selected subset of TMS systems should conform to the chosen EHS criteria within the model uncertainty. These TMS systems are then ranked only according to their thermodynamic performance. These include the partition coefficients of the product and the catalyst complex between the polar and nonpolar phases for each tie line contained within the specific ternary phase diagram. For the catalyst separation, the maximum partition coefficient is used as a descriptor. For the product separation, the average of the partition coefficients

over the entire LLE domain is used as the descriptor. In order to rank the solvent systems, the two descriptors are normalized between all solvent systems and are geometrically averaged to generate a single number. This number is used as the final ranking metric. The optimal point for the best solvent would be a score of "one" for both individual descriptors. Thus, the shorter the distance to the optimal point, the better the anticipated performance of the solvent system.

All results generated during screening are automatically included in a final report that contains the identified solvent systems as well as the rejected ones (including the specific reasons why the solvent failed). The complete result reports are included as part of the Supporting Information (files 1–4). Some of these results are partially presented in the next section.

## RESULTS AND DISCUSSION

The methodology described in the previous section is applied to the hydroformylation of 1-decene. In total, four different screenings are performed with different selection criteria

- a full screening of the entire solvent design space,
- a reduced screening with *n*-decane fixed as the nonpolar solvent,
- the same screening as before but using the green solvent selection criteria, and
- the same screening as before but with relaxed green solvent selection criteria.

The complete result report for each screening with all graphs and tables can be found in the Supporting Information (files 1–4). The screening process starts with 7758 molecules available in COSMObase and 138 molecules from the green solvent database provided by Moity et al.<sup>23</sup> Also, as mentioned previously, the solvent Cyrene is included. Since several molecules in the green solvent database are also found in the COSMObase database, the actual solvent design space is composed of 7813 unique molecules in total.

**Functional TMS.** Starting with the screening of the entire solvent design space, the different screening steps and the rejection criteria are listed in Table 5. The number of suitable solvents and solvent systems after each step are visualized in Figure 5.

Hereby, subfigures (a) and (b) show the number of potential polar and nonpolar solvents, respectively. The combination of these solvents into binary solvent systems and the subsequent screening is shown in plot (c) of Figure 5. Even though only 31 polar and 35 nonpolar solvents pass the single solvent selection process, their combination leads to a total of 1085 possible binary solvent systems. This means that 1085 binary LLEs need to be computed, which is manageable within a few days. The subsequent calculation of the 382 ternary LLEs is even more time demanding, requiring about 1 week of computational time using the hardware mentioned in the **Soft- and Hardware** section. In the end, 358 feasible solvent systems are identified using only the thermodynamic selection criteria before applying any EHS criteria. This quite long list could be easily shortened by applying stricter selection criteria. However, before any modification of the screening, the result list is analyzed in the following paragraphs.

The identified solvent systems are mapped in Figure 6. The averaged partition coefficient of 1-undecanal over the entire LLE domain is used as the descriptor for the product separation performance on the abscissa axis. The same is



**Table 5.** List of the Screening Steps with the Abbreviations Used in Figure 5 and the Corresponding Criteria for Screening of the Entire Solvent Design Space

abbreviation	property	constraint for rejection
polar solvents		
MW	molecular weight	MW > 200 g/mol
structure	structural constraints	carbon double/triple bonds, ions
BP	boiling point	BP < 20 °C, BP > 203 °C
$S_{\min}^{\text{cat}}$	relative solubility for BiPhePhos	$S^{\text{cat}} < 11.28$
$S_{\max}^{\text{prod}}$	relative solubility for undecanal	$S^{\text{prod}} > -0.63$
$H_{\text{H}_2}$	Henry coefficient for hydrogen	$H_{\text{H}_2} > 1400$ bar
$H_{\text{CO}}$	Henry coefficient for carbon monoxide	$H_{\text{CO}} > 3000$ bar
nonpolar solvents		
MW	molecular weight	MW > 200 g/mol
structure	structural constraints	carbon double/triple bonds, ions
BP	boiling point	BP < 20 °C, BP > 203 °C
$S_{\max}^{\text{cat}}$	relative solubility for BiPhePhos	$S^{\text{cat}} > 9.81$
$S^{\text{prod}}$	relative solubility for undecanal	$S_{\min}^{\text{prod}} < -0.23$
$H_{\text{H}_2}$	Henry coefficient for hydrogen	$H_{\text{H}_2} > 1000$ bar
$H_{\text{CO}}$	Henry coefficient for carbon monoxide	$H_{\text{CO}} > 1000$ bar
solvent systems		
LLE	binary LLE	$x_{\text{pol}}^{\text{nonpolar phase}} > 0.15$
LLE	ternary LLE	LLE points < 2, tie line slope $\leq 0$
$p^{\text{prod}}$	undecanal partition	$p_{\text{averaged}}^{\text{prod}} \leq 0$
$p^{\text{cat}}$	BiPhePhos partition	$p_{\text{max}}^{\text{cat}} \leq 0$

done for the catalyst separation ability using the maximum partition coefficient of BiPhePhos as the descriptor on the ordinate axis. Both descriptors are normalized. The color of the circles indicates the OGS for the polar solvent in the TMS and reveals how well the EHS properties are satisfied. It is noticeable that points of the same color are found close to each other in certain areas. These points belong to solvent systems using the same polar solvent that has been paired with varying nonpolar solvents. Consequently, the influence of the polar solvent is more decisive for the TMS performance than the influence of the nonpolar solvent. This observation is confirmed by looking at the number of identified polar and nonpolar systems. In total, only 12 different polar solvents are present in the 358 solvent systems, having been paired with 34 different nonpolar solvents. This means that the nonpolar candidates are less diverse in their properties than the polar solvents. The systems defining the Pareto front in this diagram are given in Table 6.

First, DMF appears twice on the Pareto front. This is a strong confirmation that the methodology can successfully identify a functioning TMS. Diethyl sulfoxide, dimethylcyanamide, and *N,N*-dimethylacetamide are the three other polar solvents found along the Pareto front.

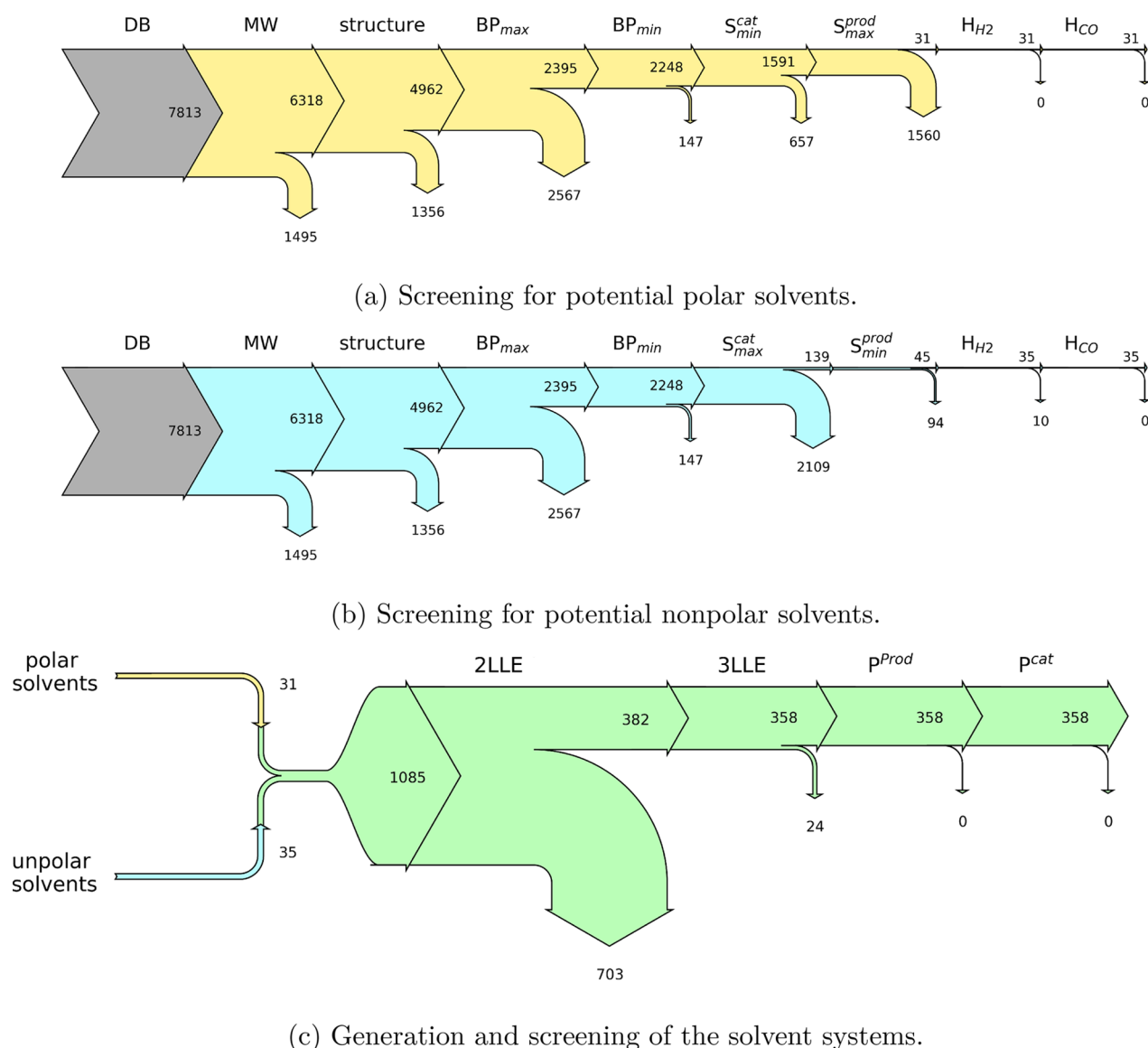
When looking at the nonpolar solvents, all 34 candidates are long-chain alkanes except for silaneoctyltrimethyl, tetraethylsilane, tetraethylgermane, and hexamethyldisilane. Altogether, simple nonpolar molecules are desired. Longer-chain molecules like dodecane (which has been also considered for a TMS used in hydroformylation) are not on the candidate list due to the boiling point constraint. Interestingly, a slightly branched structure (e.g., a methyl or ethyl group) seems to be beneficial compared to linear alkanes. This is a general trend

for the nonpolar solvents. As stated in the Introduction, the main concern of this contribution is the identification of replacements for the developmental toxic, polar solvent DMF. Therefore, it is reasonable to choose one long-chain alkane as the nonpolar solvent. By doing this, the screening criteria for the polar solvents can be relaxed since there is no more worry about a combinatorial explosion in the number of binary systems to consider. This can be seen as a decomposition step: The reduction of nonpolar solvent candidates to one gives the freedom to expand the search for a polar solvent. The improved potential of a methyl group on the alkane backbone is rather small and raises the costs for such solvents. Therefore, a linear alkane is the most reasonable choice. Since *n*-decane is already the state-of-the-art nonpolar solvent, it is chosen as the representative nonpolar solvent. However, for larger-scale experiments and investigations, the use of an alkane with only technical purity is without concern and may be slightly beneficial (at least in terms of operating costs) if the impurities contain mainly branched alkanes.

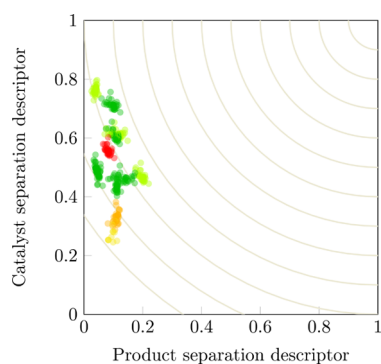
In the following, the screening results using decane as the nonpolar solvent and the relaxed thermodynamic screening criteria for the polar solvents are discussed. The relaxation affects the screening thresholds for the relative solubilities of the product ( $S_{\max}^{\text{prod}} = 0$ ), the catalyst complex ( $S_{\min}^{\text{cat}} = 10.5$ ), and the mutual cross solubility of the polar solvent in the nonpolar phase ( $x_{\text{pol, max}}^{\text{nonpolar phase}} = 0.5$ ). These adjustments lead to the investigation of 655 polar solvents for the solvent system generation instead of the previous 31. Applying the same criteria for these solvent systems leads to a short-list of 124 functional solvent systems. The Pareto plot of the identified systems is shown in Figure 7. Again, the product and the catalyst separation performances are antagonistic properties, and an optimal solvent fulfilling both properties to a high degree cannot be found. Compared to the screening of the entire solvent space in Figure 5a, a wider range for both properties is covered. The reason is that the performance of the solvent systems is more dependent on the polar solvent, and now, more polar solvents are considered for the composition of solvent systems.

In the following discussion, the EHS criteria are applied to the identified solvent systems in order to identify these TMSs best satisfying these constraints.

**Green TMS.** If all of the properties listed in Table 1 including the flash point for the polar solvents are strictly applied, there are no remaining TMSs. This means that in order to find any TMS, the strict EHS screening needs to be relaxed. There are two approaches that can be taken here. Either an acceptable hazard level for specific EHS criteria can be set, or the reliability of the VEGA predictions can be taken into account. The latter is preferable if we argue that predictions with a low reliability are not trustworthy enough to reject a solvent. Thus, predictions with low reliability are ignored for the following screening processes. This also affects the results of the ensemble models. In order to take the reliability of an ensemble model into account, the sum of the individual model reliabilities must be rated at least moderate, as explained in the Screening Methodology section. After applying the EHS criteria with these relaxed conditions, 14 solvent systems remain. Consequently, this means that 110 solvent systems, more than 88% of all systems considered, are rejected according to polar solvent's EHS properties. There is still a large reduction in the number of candidates despite the omission of the EHS predictions with low reliability. The



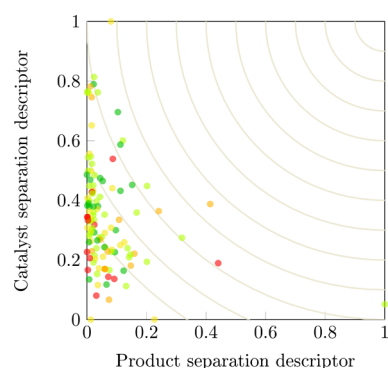
**Figure 5.** Screening procedures for the polar (a), the nonpolar solvents (b), and the solvent systems (c). The start points are the databases (DB) with the included molecules or the first-hand identified polar and nonpolar solvents. For each screening step, the numbers of remaining and rejected solvents (systems) are given at the arrowhead. The abbreviations are broken down in Table 5.



**Figure 6.** Functional TMS after screening the entire design space for the polar and nonpolar solvents. The color indicates how well the polar solvent satisfies the EHS criteria from green (best) to red (worst) corresponding to the overall greenness score introduced in the screening methodology section.

**Table 6.** Solvent Systems Defining the Pareto Front after Screening the Entire Solvent Design Space Using the Thermodynamic Criteria Listed in Table 5

polar solvent	nonpolar solvent
diethyl sulfoxide	2-methylundecane
diethyl sulfoxide	silanoctyltrimethyl
diethyl sulfoxide	3-methylundecane
dimethylformamide	3-methylundecane
dimethylcyanamide	tetraethylsilane
diethyl sulfoxide	tetraethylsilane
dimethylformamide	silanoctyltrimethyl
diethyl sulfoxide	tetraethylgermane
dimethylcyanamide	3-methylundecane
dimethylcyanamide	2-methylundecane
dimethylcyanamide	silanoctyltrimethyl
<i>N,N</i> -dimethylacetamide	tetraethylsilane
<i>N,N</i> -dimethylacetamide	2-methylundecane



**Figure 7.** Functional TMS after screening for the polar solvents using *n*-decane as the nonpolar solvent. The color indicates how well the polar solvent satisfies the EHS criteria from green (best) to red (worst) corresponding to the overall greenness score introduced in the Screening Methodology section.

candidates are listed with their chemical structure in the upper part of Table 7.

The chemical structures of the solvents and their safety data sheets (SDSs), if available, need to be finally checked by hand. Acetylperoxide and 3-(methylamino)propionitrile need to be rejected due to their structures since the peroxide or the secondary amine is not stable during the reaction, respectively. 1-Propanesulfonyl chloride reacts with water such that it

cannot be practically used as a solvent. Diethylcyanamide is labeled as being toxic in its material safety data sheet and is thus not suitable. The component 1,3-dioxol-2-one has several hazard potentials: toxic, harmful to organs, poisonous to aquatic organisms, and corrosive. This component is not revealed as hazardous by the VEGA models because the predictions for it are mostly of low reliability. However, this result is important in that it shows that the predictive models can still fail to identify substances with multiple hazards. At last, 1-mercapto-2-propanone is rejected since it is solid at room temperature. No indications for unacceptable hazardous properties are found for diethyl sulfoxide (DESO) and dimethylfurazan (DMFZ). For the other seven components, either no safety data sheet could be found, or the desired data was not available. In order to investigate a few more solvents than these two, the second relaxation strategy for the EHS screening is also applied. There is an overwhelming number of reasonable possibilities to relax some of the EHS constraints, which always end in some compromise. Here, the following approach is chosen: The skin sensitization model is not taken into account because the requirement that a solvent has no skin sensitization potential at all seems quite restrictive. Second, the flash point constraint is relaxed such that only solvents with a predicted flash point below 23 °C are rejected. This corresponds to the acceptance of flammable liquids of category 3 regarding the GHS classification, while more

**Table 7. Identified Green Polar Solvents for a TMS with *n*-Decane with Their Chemical Structures<sup>a</sup>**

Polar solvent	Structure	SDS	Polar Solvent	Structure	SDS
diethyl sulfoxide		✓	propanesulfonyl-chloride		✓
diethylcyanamide		✓	2-keto-4-butanethiol		
<i>n,n</i> -dimethylmethane-sulfinamide			3-(methylamino)-propionitrile		✓
cyanicacidethylester			acetylperoxide		
<i>o</i> -methyldimethyl-thiocarbamate			1,3-dioxol-2-one		✓
dimethylfurazan		✓	1-mercapto-2-propanone		✓
methylphosphonyl-dichloride			1,2,3-thiadiazole		
5-methyl-2(3h)-furanone		✓	methoxyaceto-nitrile		✓
tetrahydro-4h-pyran-4-one		✓	4,5-dihydro-3(2h)-thiophenone		✓

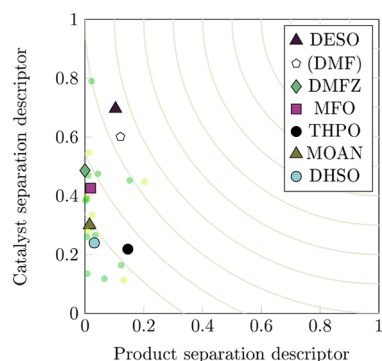
<sup>a</sup>Additionally, the availability of the safety data sheets (SDSs) is listed. The upper part shows that all candidates form the strict EHS screening, and the lower part shows that the final solvents form the relaxed EHS screening.



hazardous components of categories 1 and 2 are still excluded. The reason for choosing this relaxation is that a suitable midpolar solvent must contain elements of higher electro-negativity. Nitrogen and oxygen could both be suitable, but nitrogen-containing components often have negative effects on human health, while oxygen is likely to lower the flash point. Here, an attempt is made to find solvents without health risks at the expense of flammability, which is why the flash point condition has been relaxed. Applying these relaxed EHS constraints leads to the identification of 10 additional candidates (see the Supporting Information for further information). Four of them pass the evaluation of the chemical structure and the safety data sheet. These solvents, namely, 5-methyl-2(3h)-furanone (MFO), tetrahydro-4h-pyran-4-one (THPO), methoxyacetonitrile (MOAN), and 4,5-dihydro-3(2h)thiophenone (DHSO), can be found in the lower part of Table 7. It has to be mentioned that THPO was already found as a potential solvent in a former study<sup>10</sup> and was investigated in process optimization.<sup>29</sup>

The evaluation of the SDSs underscores the importance of performing this final checking step based on available experimental data since some hazardous components have not been predictively identified as such. On the other hand, green solvents could be mistakenly classified as hazardous and wrongly rejected. In order to reduce the chance of such wrong decisions, the demand for the VEGA reliability score could be increased resulting in a longer list for the SDS check.

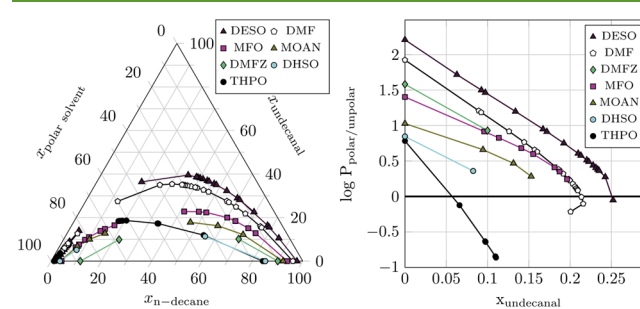
The following discussion focuses on the six identified green solvents and the comparison to a benchmark system using DMF. In Figure 8, the identified solvent systems are explicitly



**Figure 8.** Functional and green TMS after screening for the polar solvents using *n*-decane as the nonpolar solvent. The system using DMF is shown for benchmark purposes.

shown in the Pareto plot mapping product and catalyst separation abilities. The performance of DESO is very encouraging as it is predicted as performing very similarly to DMF. It shows a better predicted separation for the catalyst than the benchmark and only a slightly reduced product separation potential. Since DESO already showed up on the Pareto front of all previous screenings, it should also be competitive with other solvents that are selected only according to purely thermodynamic criteria. Even more encouraging in terms of reaction stability, the hydroformylation was already performed successfully in a TMS using the closely related molecule dimethyl sulfoxide (DMSO).<sup>115</sup> However, DMSO was shown to lead to high catalyst leaching and was outperformed by DMF in this regard. In the present screening, DMSO is sorted out due to its low

relative solubility for the catalyst, underscoring again the functionality of the COSMO-RS predictions for catalyst leaching. DESO, however, is less polar than DMSO due to it having two ethyl groups instead of methyl groups, which promotes higher catalyst solubility due to its decreased polarity. Another interesting property is the boiling temperature, which is important for the energy costs for the solvent regeneration. With 187 °C, the predicted boiling point (using COSMOtherm) of DESO is increased compared to the experimental value for DMF (153 °C). For a fair comparison in terms of economics, a process optimization would need to be performed. It may happen that the higher boiling point leads to higher costs than the savings caused by the better catalyst extraction power. Despite its higher boiling point than DMF, DESO is by far the most promising candidate. The Pareto plot shows that the other green candidates have a significantly weaker performance in terms of catalyst and product separation because they are further away from the optimal point (1, 1). In order to investigate the thermodynamic behavior in detail, Figure 9a depicts the ternary LLEs of

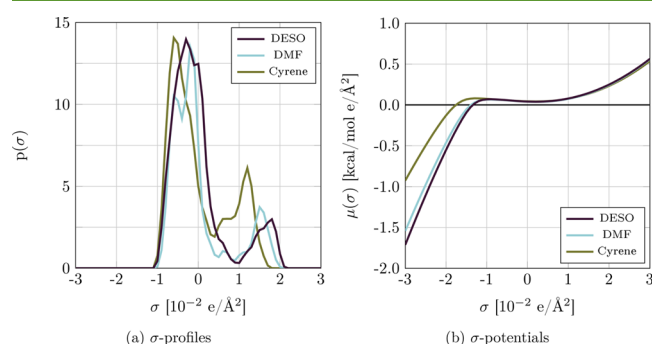


**Figure 9.** Ternary LLE predicted by COSMO-RS at 25 °C of *n*-decane, undecanal, and different polar solvents (a). Partition coefficients of the catalyst for each tie line in panel (a) over the product amount in the mixture for an equimolar phase composition (b).

the polar solvents with *n*-decane and the product undecanal at 25 °C representing the separation conditions in the hydroformylation process. DESO and DMF have the largest miscibility gaps; however, more decisive for a large partition of the product is the slope of the tie lines. Therefore, DESO and also THPO provide a similar separation efficiency for the product to DMF, something that the Pareto plot has already suggested. For the other candidates, the different partition of the product within the two liquid phases is much less pronounced. The amount of product, which is recycled with the catalyst phase back into the reactor, is consequently high and reduces the efficiency of a TMS-based process.

The catalyst partitioning behavior for each of the solvent systems is shown in Figure 9b. The figure more precisely shows the logarithmic partition coefficient of BiPhePhos for each computed tie line over the total amount of product in both phases. Hereby, an equimolar composition of both phases is assumed when computing the total amount of product so that the overall composition of the mixture is found at the middle of the tie line. DESO shows the best catalyst separation and even outperforms DMF over the entire figure. The other solvents are again less efficient but would be feasible to be used as a catalyst solvent.

Because of the highly promising performance of DESO, a comparison of its molecular characteristics seems to be appropriate. Again, the  $\sigma$ -profiles and  $\sigma$ -potentials are utilized. Figure 10 shows the profiles and potentials for DESO, DMF,



**Figure 10.**  $\sigma$ -profiles (a) and  $\sigma$ -potentials (b) for the solvents diethyl sulfoxide (DESO), dimethylformamide (DMF), and Cyrene.

and Cyrene. For DESO, there are three conformers available, but for better representation, only the profile of the lowest-energy conformer is shown. Cyrene is depicted in the figure since it is a popular green replacement solvent for DMF and was proven to be a good midpolar solvent for many different applications.<sup>112</sup> It was part of the solvent design space, but it was rejected due to its high boiling point of 227 °C. However, in general, Cyrene can be seen as a green benchmark for the DMF replacement. The  $\sigma$ -profiles of all three molecules are similar to each other; however, the overlapping of DESO and DMF is nearly perfect. For Cyrene, the peak of the positive screening charges is higher but is shifted toward more moderate charges, which leads one to expect that this would have different solvent properties. Similar results are also observed with the  $\sigma$ -potentials. Here, in the case of DESO and DMF, the energy reduction is stronger for surface elements with a negative screening charge than for Cyrene. Once again, DESO and DMF show a nearly perfect overlapping. In total, the figures reveal that DESO is more similar to DMF than Cyrene. This suggests that DESO may also be a convenient green DMF replacement in other applications as well. Of course, the EHS predictions made using VEGA are not sufficient to properly label DESO as green. This needs to be proven experimentally. However, the EHS models do not provide a reason for suspicion that it is hazardous. Additionally, DMSO is a widely used solvent without EHS complications and is even used as a medical carrier.<sup>116</sup> Therefore, it is reasonable to assume that DESO would also have good experimental EHS characteristics.

## CONCLUSIONS

A solvent screening methodology is proposed for identifying thermodynamically and environmentally benign solvent systems to facilitate the separation of a homogeneous catalyst via liquid–liquid extraction. One of the primary goals was to identify a green replacement solvent for the developmentally toxic dimethylformamide (DMF). The desired solvent systems operate according to the thermomorphic multicomponent system (TMS) concept such that the reaction is conducted in one liquid phase followed by the biphasic separation of the homogeneous catalyst and the product upon cooling. The COSMO-RS theory is used for predicting thermodynamic properties including the flash point, and EHS properties are

predicted using VEGA, which contains several QSAR models for 14 EHS end points. Ensemble models are introduced when multiple models are available for the same property in order to use as much information as possible. The reliability measure for the predictions made by VEGA is used to identify and ignore uncertain predictions to reduce the risk of falsely rejected solvents. Nevertheless, EHS predictions were verified using the safety data sheets of the identified solvents. By storing the calculated thermodynamic properties in SQLite databases, the screening procedure can be repeated with different criteria with negligible computation time. Ambiguous predictions were observed for the ligand molecule BiPhePhos using the FINE cavity construction method, while the standard cavity construction method provides reasonable results. It was concluded that for such a large molecule, the assumption that the complete surface of the molecule is in contact with solvent molecules might not hold. Precisely spoken, using a BiPhePhos molecule without surface segments in concave parts of the surface leads to decisive better results.

Diethyl sulfoxide (DESO) was identified as a highly promising solvent even outperforming the benchmark DMF in terms of catalyst extraction power at a comparable product separation efficiency. An analysis using COSMO-RS reveals that DESO shows very similar characteristics to DMF and may be able to replace DMF in other applications. In our future work, the identified solvent systems will be studied experimentally in order to verify the computational predictions of COSMO-RS and to investigate the hydroformylation reaction performance in each TMS.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acssuschemeng.0c02611>.

The supporting material includes the result files automatically generated after the screening procedure. The results are given for the three screenings using different constraints. The screening for the polar and nonpolar solvents (PDF)

The screening for the polar solvents with *n*-decane as the nonpolar solvent (PDF)

The same screening as the second but using the green prediction models to exclude hazardous molecules using strict screening constraints (PDF)

The same screening as the second but using relaxed constraints for the flash point and the skin sensitization (PDF)

A brief statement is given why the octanol–water partition is not considered as an EHS property (TXT)

## AUTHOR INFORMATION

### Corresponding Author

Kai Sundmacher – Chair for Process Systems Engineering, Otto-von-Guericke University, 39106 Magdeburg, Germany; Process Systems Engineering, Max Planck Institute for Dynamics of Complex Technical Systems, 39106 Magdeburg, Germany; [orcid.org/0000-0003-3251-0593](https://orcid.org/0000-0003-3251-0593); Email: [sundmacher@mpi-magdeburg.mpg.de](mailto:sundmacher@mpi-magdeburg.mpg.de)

## Authors

Steffen Linke – Chair for Process Systems Engineering, Otto-von-Guericke University, 39106 Magdeburg, Germany;  
orcid.org/0000-0003-3935-5936

Kevin McBride – Department of Chemical Engineering, Carnegie Mellon University, 15213 Pittsburgh, Pennsylvania, United States

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acssuschemeng.0c02611>

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - TRR 63 "Integrated Chemical Processes in Liquid Multiphase Systems" (Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) - TRR 63 "Integrierte chemische Prozesse in flüssigen Mehrphasensystemen" (Teilprojekt B9) - 56091768). S.L. is also affiliated with the International Max Planck Research School for Advanced Methods in Process and Systems Engineering - IMPRS ProEng at the Max Planck Institute for Dynamics of Complex Technical Systems Magdeburg.

## REFERENCES

- (1) United Nations *Transforming Our World: The 2030 Agenda For Sustainable Development*; 2015, <https://sustainabledevelopment.un.org/post2015/transformingourworld/publication>, last accessed: 21.06.2020.
- (2) Pfennig, A. Sustainable Bio- or CO<sub>2</sub> economy: Chances, Risks, and Systems Perspective. *ChemBioEng Rev* **2019**, *6*, 90–104.
- (3) Behr, A.; Neubert, P. *Applied homogeneous catalysis*; John Wiley & Sons: Weinheim, 2012.
- (4) Behr, A.; Vorholt, A. J. *Homogeneous catalysis with renewables; Catalysis by metal complexes*; Springer: Cham, 2017; Vol. volume 39.
- (5) Mohammad, A.; Inamuddin *Green Solvents I*; Springer: Netherlands, Dordrecht, 2012, DOI: 10.1007/978-94-007-1712-1.
- (6) Wypych, G. *Handbook of solvents*; ChemTec Publishing: Toronto and New York, 2001.
- (7) Byrne, F. P.; Jin, S.; Paggiola, G.; Petchey, T. H. M.; Clark, J. H.; Farmer, T. J.; Hunt, A. J.; Robert McElroy, C.; Sherwood, J. Tools and techniques for solvent selection: green solvent selection guides. *Sustainable Chem. Processes* **2016**, *4*, 7.
- (8) European Chemical Agency *Registration, Evaluation, Authorisation and Restriction of Chemicals: REACH*. 18.12.2006; <https://echa.europa.eu/regulations/reach/understanding-reach>, last accessed: 21.06.2020.
- (9) Marvey, B. B. Sunflower-based feedstocks in nonfood applications: Perspectives from olefin metathesis. *Int. J. Mol. Sci.* **2008**, *9*, 1393–1406.
- (10) McBride, K.; Linke, S.; Xu, S.; Sundmacher, K. Computer Aided Design of Green Thermomorphic Solvent Systems for Homogeneous Catalyst Recovery. *Comput.-Aided Chem. Eng.* **2018**, *44*, 1783–1788.
- (11) Jokiel, M.; Rätze, K. H. G.; Kaiser, N. M.; Künnemann, K. U.; Hollenbeck, J.-P.; Dreimann, J. M.; Vogt, D.; Sundmacher, K. Miniplant-Scale Evaluation of a Semibatch-Continuous Tandem Reactor System for the Hydroformylation of LongChain Olefins. *Ind. Eng. Chem. Res.* **2018**, *58*, 2471–2480.
- (12) Jokiel, M.; Kaiser, N. M.; Kováts, P.; Mansour, M.; Zähringer, K.; Nigam, K. D. P.; Sundmacher, K. Helically coiled segmented flow tubular reactor for the hydroformylation of long-chain olefins in a thermomorphic multiphase system. *Chem. Eng. J.* **2019**, *377*, 120060.
- (13) Rätze, K. H. G.; Jokiel, M.; Kaiser, N. M.; Sundmacher, K. Cyclic operation of a semibatch reactor for the hydroformylation of long-chain olefins and integration in a continuous production process. *Chem. Eng. J.* **2019**, *377*, 120453.
- (14) Dreimann, J. M.; Warmeling, H.; Weimann, J. N.; Künnemann, K.; Behr, A.; Vorholt, A. J. Increasing selectivity of the hydroformylation in a miniplant: Catalyst, solvent, and olefin recycle in two loops. *AIChE J.* **2016**, *62*, 4377–4383.
- (15) Schäfer, E.; Brunsch, Y.; Sadowski, G.; Behr, A. Hydroformylation of 1-Dodecene in the Thermomorphic Solvent System Dimethylformamide/Decane. Phase Behavior–Reaction Performance–Catalyst Recycling. *Ind. Eng. Chem. Res.* **2012**, *51*, 10296–10306.
- (16) Behr, A. *Angewandte homogene Katalyse*; Wiley-VCH: Weinheim, 2008.
- (17) McBride, K.; Gaide, T.; Vorholt, A.; Behr, A.; Sundmacher, K. Thermomorphic solvent selection for homogeneous catalyst recovery based on COSMO-RS. *Chem. Eng. Process.: Process Intensif.* **2016**, *99*, 97–106.
- (18) European Chemical Agency *Candidate List of substances of very high concern for Authorisation*; 2008, <https://echa.europa.eu/en/candidate-list-table>, last accessed: 21.06.2020.
- (19) Benfenati, E. *Theory, guidance and applications on QSAR and REACH*; 2013.
- (20) Adjiman, C. S.; Galindo, A.; Pistikopoulos, E. N.; Georgiadis, M. C.; Dua, V. *Molecular systems engineering*; John Wiley & Sons: Weinheim, 2010.
- (21) Gmehling, J.; Schedemann, A. Selection of Solvents or Solvent Mixtures for Liquid–Liquid Extraction Using Predictive Thermodynamic Models or Access to the Dortmund Data Bank. *Ind. Eng. Chem. Res.* **2014**, *53*, 17794–17805.
- (22) Pistikopoulos, E. N.; Stefanis, S. K. Optimal solvent design for environmental impact minimization. *Comput. Chem. Eng.* **1998**, *22*, 717–733.
- (23) Moity, L.; Durand, M.; Benazzouz, A.; Pierlot, C.; Molinier, V.; Aubry, J.-M. Panorama of sustainable solvents using the COSMO-RS approach. *Green Chem.* **2012**, *14*, 1132–1145.
- (24) Papadopoulos, A. I.; Linke, P. Multiobjective molecular design for integrated process-solvent systems synthesis. *AIChE J.* **2006**, *52*, 1057–1070.
- (25) Zhou, T.; McBride, K.; Zhang, X.; Qi, Z.; Sundmacher, K. Integrated solvent and process design exemplified for a Diels-Alder reaction. *AIChE J.* **2015**, *61*, 147–158.
- (26) Jens, C. M.; Nowakowski, K.; Scheffczyk, J.; Leonhard, K.; Bardow, A. CO from CO<sub>2</sub> and fluctuating renewable energy via formic-acid derivatives. *Green Chem.* **2016**, *18*, 5621–5629.
- (27) Zhou, T.; Zhou, Y.; Sundmacher, K. A hybrid stochastic–deterministic optimization approach for integrated solvent and process design. *Chem. Eng. Sci.* **2017**, *159*, 207–216.
- (28) Fleitmann, L.; Scheffczyk, J.; Schäfer, P.; Jens, C.; Leonhard, K.; Bardow, A. Integrated Design of Solvents in Hybrid Reaction-Separation Processes Using COSMO-RS. *Chem. Eng. Trans.* **2018**, *2018*, 599–564.
- (29) Keßler, T.; Kunde, C.; Linke, S.; McBride, K.; Sundmacher, K.; Kienle, A. Systematic Selection of Green Solvents and Process Optimization for the Hydroformylation of Long-Chain Olefins. *Processes* **2019**, *7*, 882.
- (30) Sheldon, T. J.; Folić, M.; Adjiman, C. S. Solvent Design Using a Quantum Mechanical Continuum Solvation Model. *Ind. Eng. Chem. Res.* **2006**, *45*, 1128–1140.
- (31) Struebing, H.; Ganase, Z.; Karamertzanis, P. G.; Sioukrou, E.; Haycock, P.; Piccione, P. M.; Armstrong, A.; Galindo, A.; Adjiman, C. S. Computer-aided molecular design of solvents for accelerated reaction kinetics. *Nat. Chem.* **2013**, *5*, 952–957.
- (32) Austin, N. D.; Samudra, A. P.; Sahinidis, N. V.; Trahan, D. W. Mixture design using derivative-free optimization in the space of individual component properties. *AIChE J.* **2016**, *62*, 1514–1530.
- (33) Scheffczyk, J.; Fleitmann, L.; Schwarz, A.; Lampe, M.; Bardow, A.; Leonhard, K. COSMO-CAMD: A framework for optimization-based computer-aided molecular design using COSMO-RS. *Chem. Eng. Sci.* **2017**, *159*, 84–92.



- (34) Lampe, M.; Stavrou, M.; Schilling, J.; Sauer, E.; Gross, J.; Bardow, A. Computer-aided molecular design in the continuous-molecular targeting framework using groupcontribution PC-SAFT. *Comput. Chem. Eng.* **2015**, *81*, 278–287.
- (35) Zhou, T.; Qi, Z.; Sundmacher, K. Model-based method for the screening of solvents for chemical reactions. *Chem. Eng. Sci.* **2014**, *115*, 177–185.
- (36) Struebing, H.; Obermeier, S.; Sioumkrou, E.; Adjiman, C. S.; Galindo, A. A QM-CAMD approach to solvent design for optimal reaction rates. *Chem. Eng. Sci.* **2017**, *159*, 69–83.
- (37) Amar, Y.; Schweidtmann, A. M.; Deutsch, P.; Cao, L.; Lapkin, A. Machine learning and molecular descriptors enable rational solvent selection in asymmetric catalysis. *Chem. Sci.* **2019**, *10*, 6697–6706.
- (38) Salazar, J.; Diwekar, U.; Joback, K.; Berger, A. H.; Bhowan, A. S. Solvent Selection for Post-Combustion CO<sub>2</sub> Capture. *Energy Procedia* **2013**, *37*, 257–264.
- (39) First, E. L.; Hasan, M. M. F.; Floudas, C. A. Discovery of novel zeolites for natural gas purification through combined material screening and process optimization. *AIChE J.* **2014**, *60*, 1767–1785.
- (40) Blumenthal, L. C.; Jens, C. M.; Ulbrich, J.; Schwering, F.; Langrehr, V.; Turek, T.; Kunz, U.; Leonhard, K.; Palkovits, R. Systematic Identification of Solvents Optimal for the Extraction of 5-Hydroxymethylfurfural from Aqueous Reactive Solutions. *ACS Sustainable Chem. Eng.* **2016**, *4*, 228–235.
- (41) Wang, J.; Lakerveld, R. Integrated solvent and process design for continuous crystallization and solvent recycling using PC-SAFT. *AIChE J.* **2018**, *64*, 1205–1216.
- (42) Burghoff, B.; Goetheer, E. L. V.; de Haan, A. B. COSMO-RS-Based Extractant Screening for Phenol Extraction As Model System. *Ind. Eng. Chem. Res.* **2008**, *47*, 4263–4269.
- (43) Burger, J.; Papaioannou, V.; Gopinath, S.; Jackson, G.; Galindo, A.; Adjiman, C. S. A hierarchical method to integrated solvent and process design of physical CO<sub>2</sub> absorption using the SAFT- $\gamma$  M ie approach. *AIChE J.* **2015**, *61*, 3249–3269.
- (44) Jonuzaj, S.; Akula, P. T.; Kleniati, P.-M.; Adjiman, C. S. The formulation of optimal mixtures with generalized disjunctive programming: A solvent design case study. *AIChE J.* **2016**, *62*, 1616–1633.
- (45) Jonuzaj, S.; Gupta, A.; Adjiman, C. S. The design of optimal mixtures from atom groups using Generalized Disjunctive Programming. *Comput. Chem. Eng.* **2018**, *116*, 401–421.
- (46) Venkatasubramanian, V.; Chan, K.; Caruthers, J. M. Computer-aided molecular design using genetic algorithms. *Comput. Chem. Eng.* **1994**, *18*, 833–844.
- (47) Bardow, A.; Steur, K.; Gross, J. Continuous-Molecular Targeting for Integrated Solvent and Process Design. *Ind. Eng. Chem. Res.* **2010**, *49*, 2834–2840.
- (48) Stavrou, M.; Lampe, M.; Bardow, A.; Gross, J. Continuous Molecular Targeting—Computer-Aided Molecular Design (CoMT—CAMD) for Simultaneous Process and Solvent Design for CO<sub>2</sub> Capture. *Ind. Eng. Chem. Res.* **2014**, *53*, 18029–18041.
- (49) Austin, N. D.; Sahinidis, N. V.; Trahan, D. W. Computer-aided molecular design: An introduction and review of tools, applications, and solution techniques. *Chem. Eng. Res. Des.* **2016**, *116*, 2–26.
- (50) Ng, L. Y.; Chong, F. K.; Chemmangattuvalappil, N. G. Challenges and opportunities in computer-aided molecular design. *Comput. Chem. Eng.* **2015**, *81*, 115–129.
- (51) Stefanis, S. K.; Buxton, A.; Livingston, A. G.; Pistikopoulos, E. N. A methodology for environmental impact minimization: Solvent design and reaction path synthesis issues. *Comput. Chem. Eng.* **1996**, *20*, S1419–S1424.
- (52) Hostrup, M.; Harper, P. M.; Gani, R. Design of environmentally benign processes: integration of solvent design and separation process synthesis. *Comput. Chem. Eng.* **1999**, *23*, 1395–1414.
- (53) Xu, W.; Diwekar, U. M. Multi-objective integrated solvent selection and solvent recycling under uncertainty using a new genetic algorithm. *Int. J. Environ. Pollut.* **2007**, *29*, 70.
- (54) Cignitti, S.; Mansouri, S. S.; Woodley, J. M.; Abildskov, J. Systematic Optimization-Based Integrated Chemical Product—Process Design Framework. *Ind. Eng. Chem. Res.* **2018**, *57*, 677–688.
- (55) Ooi, J.; Ng, D. K. S.; Chemmangattuvalappil, N. G. A Systematic Molecular Design Framework with the Consideration of Competing Solvent Recovery Processes. *Ind. Eng. Chem. Res.* **2019**, *58*, 13210–13226.
- (56) Vanderveen, J. R.; Patiny, L.; Chalifoux, C. B.; Jessop, M. J.; Jessop, P. G. A virtual screening approach to identifying the greenest compound for a task: application to switchable-hydrophilicity solvents. *Green Chem.* **2015**, *17*, S182–S188.
- (57) Chang, S. S. L.; Kong, Y. L.; Lim, W. X.; Ooi, J.; Ng, D. K. S.; Chemmangattuvalappil, N. G. Design of alternate solvent for recovery of residual palm oil: simultaneous optimization of process performance with environmental, health and safety aspects. *Clean Technol. Environ. Policy* **2018**, *20*, 949–968.
- (58) Neoh, J. Q.; Chin, H. H.; Mah, A. X. Y.; Aboagwa, O. A.; Thangalazhy-Gopakumar, S.; Chemmangattuvalappil, N. G. Design of bio-oil additives using mathematical optimisation tools considering blend functionality and sustainability aspects. *Sustainable Prod. Consumption* **2019**, *19*, S3–63.
- (59) Ten, J. Y.; Hassim, M. H.; Chemmangattuvalappil, N. G. Integration of safety and health aspects in a simultaneous process and molecular design framework. *Chem. Eng. Res. Des.* **2020**, *153*, 849–864.
- (60) Chemmangattuvalappil, N. G. Development of solvent design methodologies using computer-aided molecular design tools. *Curr. Opin. Chem. Eng.* **2020**, *27*, 51–59.
- (61) Fredenslund, A.; Jones, R. L.; Prausnitz, J. M. Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE J.* **1975**, *21*, 1086–1099.
- (62) Gmehling, J.; Li, J.; Schiller, M. A modified UNIFAC model. 2. Present parameter matrix and results for different thermodynamic properties. *Ind. Eng. Chem. Res.* **1993**, *32*, 178–193.
- (63) 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.
- (64) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. Refinement and Parametrization of COSMO-RS. *J. Phys. Chem. A* **1998**, *102*, S074–S085.
- (65) Klamt, A. The COSMO and COSMO-RS solvation models. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2017**, *8*, No. e1338.
- (66) Klamt, A.; Eckert, F.; Arlt, W. COSMO-RS: an alternative to simulation for calculating thermodynamic properties of liquid mixtures. *Annu. Rev. Chem. Biomol. Eng.* **2010**, *1*, 101–122.
- (67) Eckert, F.; Klamt, A. Fast solvent screening via quantum chemistry: COSMO-RS approach. *AIChE J.* **2002**, *48*, 369–385.
- (68) Ferrari, T.; Gini, G. An open source multistep model to predict mutagenicity from statistical analysis and relevant structural alerts. *Chem. Cent. J.* **2010**, *4*, S2.
- (69) Ferrari, T.; Cattaneo, D.; Gini, G.; Golbamaki Bakhtyari, N.; Manganaro, A.; Benfenati, E. Automatic knowledge extraction from chemical structures: the case of mutagenicity prediction. *SAR QSAR Environ. Res.* **2013**, *24*, 365–383.
- (70) Benigni, R.; Bossa, C. Mechanisms of chemical carcinogenicity and mutagenicity: a review with implications for predictive toxicology. *Chem. Rev.* **2011**, *111*, 2507–2536.
- (71) Benigni, R.; Bossa, C.; Tcheremenskaia, O. In vitro cell transformation assays for an integrated, alternative assessment of carcinogenicity: a data-based analysis. *Mutagenesis* **2013**, *28*, 107–116.
- (72) Fjodorova, N.; Vračko, M.; Novič, M.; Roncaglioni, A.; Benfenati, E. New public QSAR model for carcinogenicity. *Chem. Cent. J.* **2010**, *4*, S3.
- (73) Benigni, R.; Bossa, C.; Netzeva, T.; Rodomonte, A.; Tsakovska, I. Mechanistic QSAR of aromatic amines: new models for discriminating between homocyclic mutagens and nonmutagens, and validation of models for carcinogens. *Environ. Mol. Mutagen.* **2007**, *48*, 754–771.

- (74) Benigni, R.; Bossa, C.; Jeliaskova, N.; Netzeva, T.; Worth, A. *The Benigni / Bossa rulebase for mutagenicity and carcinogenicity - a module of Toxtree*; 2008.
- (75) Benigni, R.; Bossa, C.; Tcheremenskaia, O. Nongenotoxic carcinogenicity of chemicals: mechanisms of action and early recognition through a new set of structural alerts. *Chem. Rev.* **2013**, *113*, 2940–2957.
- (76) Cassano, A.; Manganaro, A.; Martin, T.; Young, D.; Piclin, N.; Pintore, M.; Bigoni, D.; Benfenati, E. CAESAR models for developmental toxicity. *Chem. Cent. J.* **2010**, *4*, S4.
- (77) Wu, S.; Fisher, J.; Naciff, J.; Lauferweiler, M.; Lester, C.; Daston, G.; Blackburn, K. Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. *Chem. Res. Toxicol.* **2013**, *26*, 1840–1861.
- (78) Roncaglioni, A.; Piclin, N.; Pintore, M.; Benfenati, E. Binary classification models for endocrine disrupter effects mediated through the estrogen receptor. *SAR QSAR Environ. Res.* **2008**, *19*, 697–733.
- (79) Mansouri, K.; et al. CERAPP: Collaborative Estrogen Receptor Activity Prediction Project. *Environ. Health Perspect.* **2016**, *124*, 1023–1033.
- (80) Chaudhry, Q.; Piclin, N.; Cotterill, J.; Pintore, M.; Price, N. R.; Chrétien, J. R.; Roncaglioni, A. Global QSAR models of skin sensitizers for regulatory purposes. *Chem. Cent. J.* **2010**, *4*, S5.
- (81) Pizzo, F.; Lombardo, A.; Manganaro, A.; Benfenati, E. A New Structure-Activity Relationship (SAR) Model for Predicting Drug-Induced Liver Injury, Based on Statistical and Expert-Based Structural Alerts. *Front. Pharmacol.* **2016**, *7*, 442.
- (82) Martin, T. M.; Young, D. M. Prediction of the acute toxicity (96-h LC50) of organic compounds to the fathead minnow (*Pimephales promelas*) using a group contribution method. *Chem. Res. Toxicol.* **2001**, *14*, 1378–1385.
- (83) Benfenati, E. *Quantitative structure-activity relationships (QSAR) for pesticide regulatory purposes*; Elsevier: Amsterdam and London, 2007.
- (84) Como, F.; Carnesecchi, E.; Volani, S.; Dorne, J. L.; Richardson, J.; Bassan, A.; Pavan, M.; Benfenati, E. Predicting acute contact toxicity of pesticides in honeybees (*Apis mellifera*) through a k-nearest neighbor model. *Chemosphere* **2017**, *166*, 438–444.
- (85) Zhao, C.; Boriani, E.; Chana, A.; Roncaglioni, A.; Benfenati, E. A new hybrid system of QSAR models for predicting bioconcentration factors (BCF). *Chemosphere* **2008**, *73*, 1701–1707.
- (86) Meylan, W. M.; Howard, P. H.; Boethling, R. S.; Aronson, D.; Printup, H.; Gouchie, S. Improved method for estimating bioconcentration/bioaccumulation factor from octanol/water partition coefficient. *Environ. Toxicol. Chem.* **1999**, *18*, 664–672.
- (87) Arnot, J. A.; Mackay, D.; Parkerton, T. F.; Bonnell, M. A Database of Fish Biotransformation Rates for Organic Chemicals. *Environ. Toxicol. Chem.* **2008**, *27*, 2263–2270.
- (88) Arnot, J. A.; Mackay, D.; Bonnell, M. Estimating metabolic biotransformation rates in fish from laboratory data. *Environ. Toxicol. Chem.* **2008**, *27*, 341–351.
- (89) Manganaro, A.; Pizzo, F.; Lombardo, A.; Pogliaghi, A.; Benfenati, E. Predicting persistence in the sediment compartment with a new automatic software based on the k-Nearest Neighbor (k-NN) algorithm. *Chemosphere* **2016**, *144*, 1624–1630.
- (90) Meylan, W. M.; Howard, P. H. Atom/fragment contribution method for estimating octanol-water partition coefficients. *J. Pharm. Sci.* **1995**, *84*, 83–92.
- (91) Moriguchi, I.; Hirono, S.; LIU, Q.; Nakagome, I.; Matsushita, Y. Simple Method of Calculating Octanol/Water Partition Coefficient. *Chem. Pharm. Bull.* **1992**, *40*, 127–130.
- (92) Moriguchi, I.; Hirono, S.; Nakagome, I.; Hirano, H. Comparison of reliability of log P values for drugs calculated by several methods. *Chem. Pharm. Bull.* **1994**, *42*, 976–978.
- (93) Ghose, A. K.; Crippen, G. M. Atomic Physicochemical Parameters for ThreeDimensional Structure-Directed Quantitative Structure-Activity Relationships I Partition Coefficients as a Measure of Hydrophobicity. *J. Comput. Chem.* **1986**, *7*, 565–577.
- (94) Viswanadhan, V. N.; Reddy, M. R.; Bacquet, R. J.; Erion, M. E. Assessment of methods used for predicting lipophilicity: Application to nucleosides and nucleoside bases. *J. Comput. Chem.* **1993**, *14*, 1019–1026.
- (95) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. Prediction of Hydrophobic (Lipophilic) Properties of Small Organic Molecules Using Fragmental Methods: An Analysis of ALOGP and CLOGP Methods. *J. Phys. Chem. A* **1998**, *102*, 3762–3772.
- (96) Bergbreiter, D. E.; Liu, Y.-S.; Osburn, P. L. Thermomorphic Rhodium(I) and Palladium(0) Catalysts. *J. Am. Chem. Soc.* **1998**, *120*, 4250–4251.
- (97) Dreimann, J. M.; Faßbach, T. A.; Fuchs, S.; Fürst, M. R. L.; Gaide, T.; Kuhlmann, R.; Ostrowski, K. A.; Stadler, A.; Seidensticker, T.; Vogelsang, D.; Warmeling, H. W. F.; Vorholt, A. J. Vom Laborkuriosum zum kontinuierlichen Prozess: Die Entwicklung thermomorpher Lösungsmittelsysteme. *Chem. Ing. Tech.* **2017**, *89*, 252–262.
- (98) Bianga, J.; Künnemann, K. U.; Gaide, T.; Vorholt, A. J.; Seidensticker, T.; Dreimann, J. M.; Vogt, D. Thermomorphic Multiphase Systems - Switchable Solvent Mixtures for the Recovery of Homogeneous Catalysts in Batch and Flow Processes. *Chem. – Eur. J.* **2019**, 11586–11608.
- (99) Behr, A.; Henze, G.; Obst, D.; Turkowski, B. Selection process of new solvents in temperature-dependent multi-component solvent systems and its application in isomerising hydroformylation. *Green Chem.* **2005**, *7*, 645–649.
- (100) Behr, A.; Kleyensteiber, A.; Becker, M. A novel approach to selecting thermomorphic multicomponent solvent systems (TMS) for hydroaminomethylation reactions. *Chem. Eng. Process.: Process Intensif.* **2013**, *69*, 15–23.
- (101) McBride, K.; Kaiser, N. M.; Sundmacher, K. Integrated reaction-extraction process for the hydroformylation of long-chain alkenes with a homogeneous catalyst. *Comput. Chem. Eng.* **2017**, *105*, 212–223.
- (102) Markert, J.; Brunsch, Y.; Munkelt, T.; Kiedorf, G.; Behr, A.; Hamel, C.; Seidel-Morgenstern, A. Analysis of the reaction network for the Rh-catalyzed hydroformylation of 1-dodecene in a thermomorphic multicomponent solvent system. *Appl. Catal., A* **2013**, *462–463*, 287–295.
- (103) Kiedorf, G.; Hoang, D. M.; Müller, A.; Jörke, A.; Markert, J.; Arellano-Garcia, H.; Seidel-Morgenstern, A.; Hamel, C. Kinetics of 1-dodecene hydroformylation in a thermomorphic solvent system using a rhodium-biphenylphosphos catalyst. *Chem. Eng. Sci.* **2014**, *115*, 31–48.
- (104) Hentschel, B.; Kiedorf, G.; Gerlach, M.; Hamel, C.; Seidel-Morgenstern, A.; Freund, H.; Sundmacher, K. Model-Based Identification and Experimental Validation of the Optimal Reaction Route for the Hydroformylation of 1-Dodecene. *Ind. Eng. Chem. Res.* **2015**, *54*, 1755–1765.
- (105) Turbomole Program Package For Electronic Structure Calculations; 1989-2007, <http://www.turbomole.com>.
- (106) COSMOlogic; <http://www.cosmologic.de>.
- (107) RDKit: Open-Source Cheminformatics Software; <https://www.rdkit.org/>, last accessed: 23.06.2020.
- (108) Ebejer, J.-P.; Morris, G. M.; Deane, C. M. Freely available conformer generation methods: How good are they? *J. Chem. Inf. Model.* **2012**, *52*, 1146–1158.
- (109) Benfenati, E.; Manganaro, A.; Gini, G. VEGA-QSAR: AI inside a platform for predictive toxicology; CEUR Workshop Proceedings: 2013,
- (110) Preisach, C.; Burkhardt, H.; Schmidt-Thieme, L.; Decker, R. *Data Analysis, Machine Learning and Applications: Proceedings of the 31st Annual Conference of the Gesellschaft für Klassifikation EV, Albert-Ludwigs Universität Freiburg, March 7–9, 2007*, 1st ed.; Studies in classification, data analysis, and knowledge organization; Springer Science & Business Media: s.l., 2008.
- (111) Brunsch, Y.; Behr, A. Temperature-controlled catalyst recycling in homogeneous transition-metal catalysis: minimization of catalyst leaching. *Angew. Chem., Int. Ed.* **2013**, *52*, 1586–1589.

(112) Camp, J. E. Bio-available Solvent Cyrene: Synthesis, Derivatization, and Applications. *ChemSusChem* **2018**, *11*, 3048–3055.

(113) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* **2019**, *47*, D1102–D1109.

(114) Jörke, A.; Seidel-Morgenstern, A.; Hamel, C. Rhodium-BiPhePhos catalyzed hydroformylation studied by operando FTIR spectroscopy: Catalyst activation and rate determining step. *J. Mol. Catal. A: Chem.* **2017**, *426*, 10–14.

(115) Gaide, T.; Jörke, A.; Schlipköter, K. E.; Hamel, C.; Seidel-Morgenstern, A.; Behr, A.; Vorholt, A. J. Isomerization/hydroformylation tandem reaction of a decene isomeric mixture with subsequent catalyst recycling in thermomorphic solvent systems. *Appl. Catal., A* **2017**, *532*, 50–56.

(116) Capriotti, K.; Capriotti, J. A. Dimethyl sulfoxide: history, chemistry, and clinical utility in dermatology. *J. Clin. Aesthetic Dermatol.* **2012**, *5*, 24–26.