Model-Based Formulation of Biofuel Blends by Simultaneous Product and Pathway Design

Manuel Dahmen and Wolfgang Marquardt*†

Aachener Verfahrenstechnik—Process Systems Engineering, RWTH Aachen University, Turmstrasse 46, 52064 Aachen, Germany

Supporting Information

ABSTRACT: Oxygenated species obtained from the selective chemo-catalytic refunctionalization of lignocellulosic materials and rationally formulated mixtures thereof can be tailored to the needs of advanced internal combustion engine concepts like low temperature compression-ignition or highly boosted spark-ignition combustion. In the present contribution, we present a framework for model-based formulation of biofuel blends with tailored properties by considering the fuel’s molecular composition as the fundamental design degree of freedom. To this end, optimization-based reaction pathway screening is combined with mathematical fuel property prediction by means of structural group contribution and quantitative structure-property relationship modeling in order to formulate and solve an integrated product and pathway design problem. The model-based approach is envisioned (i) to guide fundamental experimental investigations of the combustion behavior of blended biofuels toward the most favorable mixtures and (ii) to identify promising conversion pathways for further elaboration by means of reaction engineering and conceptual process design. The latter is ultimately needed to bridge the gap from the mass- and energy-based molecular level analysis presented here to a process level analysis addressing the economics of the involved conversion and separation steps. A case study is dedicated to the identification of promising blends and respective production routes for the spark-ignition (SI) engine. The goal here is to produce a 100% bioderived fuel with tailored properties from hexoses and pentoses of lignocellulosic biomass by utilizing renewable hydrogen from carbon-free energy sources such as wind or solar to boost the energy of fuel produced for a given quantity of biomass supplied.

INTRODUCTION

Biofuels are considered to be one important technological option for mitigating the negative effects of global warming caused by excessive carbon dioxide emissions from the combustion of fossil fuels in ground and air transport. To increase the availability of feedstock and to avoid direct competition with food production, research has been directed to the conversion of lignocellulosic feedstock and to avoid direct competition with food production, and (iii) by the interactions between conversion pathways that connect these molecules refunctionalization of bioderived sugars in principle, (ii) by the numerous conversion pathways that connect these molecules with the biomass feedstock, and (iii) by the interactions between physicochemical fuel properties and the performance of an internal combustion engine. Hechinger, Voll, and Marquardt12 suggested to tackle this problem by combining computer-aided molecular design (CAMD)13 for de novo generation of biogenic fuel candidates exhibiting tailored properties10 with reaction network flux analysis (RNFA),14 i.e., an optimization-based methodology to identify the most promising routes for fuel production based on a network of competing reaction pathways. Until now, CAMD and RNFA have been applied consecutively or iteratively and have been mostly confined to the identification of pure component fuels.9,10,12,14−19 In the present contribution, the two approaches are brought together to solve an integrated product and pathway design problem by incorporating CAMD aspects, e.g., predictive fuel property models, into the optimization-based framework provided by RNFA.

The focus of this contribution lies on the rational formulation of well-defined blends of few bioderived fuel components. As such we will not consider thermochemical biofuel production pathways which typically yield complex multicomponent mixtures of (oxygenated) hydrocarbons. Instead, we will focus on biofuels made through hydrolysis of cellulose and hemicellulose following by selective synthesis of few oxygenated species via chemical and/or fermentation routes.2,4,7,20−22 The blends are formulated to maximize a process-related performance indicator, such as mass of fuel produced for a fixed input of biomass, while certain fuel quality requirements, including constraints on the derived cetane number, fuel density, or volatility, are met.

A comprehensive methodology for computer-aided mixture design and formulation of blended products was been proposed as the fundamental design degree of freedom. To this end, optimization-based reaction pathway screening is combined with mathematical fuel property prediction by means of structural group contribution and quantitative structure-property relationship modeling in order to formulate and solve an integrated product and pathway design problem. The model-based approach is envisioned (i) to guide fundamental experimental investigations of the combustion behavior of blended biofuels toward the most favorable mixtures and (ii) to identify promising conversion pathways for further elaboration by means of reaction engineering and conceptual process design. The latter is ultimately needed to bridge the gap from the mass- and energy-based molecular level analysis presented here to a process level analysis addressing the economics of the involved conversion and separation steps. A case study is dedicated to the identification of promising blends and respective production routes for the spark-ignition (SI) engine. The goal here is to produce a 100% bioderived fuel with tailored properties from hexoses and pentoses of lignocellulosic biomass by utilizing renewable hydrogen from carbon-free energy sources such as wind or solar to boost the energy of fuel produced for a given quantity of biomass supplied.

Received: January 11, 2017
Revised: March 1, 2017
Published: March 1, 2017

© 2017 American Chemical Society

DOI: 10.1021/acs.energyfuels.7b00118
Energy Fuels 2017, 31, 4096−4121
The blend design framework proposed in the present contribution is depicted in Figure 1 and comprises four stages. In stage 1, a set of so-called palette compounds has to be selected. The term *palette compound* refers to any compound that could be part of a rationally formulated blend, in principle. The set of palette compounds therefore defines the search space for the blend design. Palette compounds can either be assembled manually or generated automatically by the computer-aided fuel design approach proposed by Dahmen and Marquardt. In the present contribution, fuel property prediction techniques based on structural group contribution (QSPR) modeling developed within TMFB are extended to predict mixture properties such that even constraints on the distillation profile of a fuel can be considered explicitly in the blend design problem. The remainder of the paper is structured as follows: The blend design framework is outlined in the next section. Afterward, the concept of conversion pathway maps is introduced which is based on RNFA. Models for the prediction of mixture properties will be discussed in the subsequent section. This then allows to derive the problem formulation and to explain the solution strategy. A case study demonstrates the feasibility of the blend design approach. It is dedicated to the identification of a 100% renewable fuel mixture for the spark-ignition (SI) engine and respective production routes. Specifically, we investigate the scenario where the energy of fuel produced for a given amount of biomass shall be maximized. In order to meet this objective, renewable hydrogen from carbon-free energy sources such as wind or solar shall be provided. Finally, some conclusions and discussions are given.

## INTEGRATED PRODUCT AND PATHWAY DESIGN OF BIOFUELS

In the present contribution, the set of so-called palette compounds has to be selected. The term *palette compound* refers to any compound that could be part of a rationally formulated blend, in principle. The set of palette compounds therefore defines the search space for the blend design. Palette compounds can either be assembled manually or generated automatically by the computer-aided fuel design approach proposed by Dahmen and Marquardt. Criteria for deciding whether a molecule should be included in the set of palette compounds include pure component properties and yield data, as well as additional process-related data if available.

Based on the set of palette compounds, a conversion pathway map (CPM) is created in stage 2. The CPM holds all conversion pathways to be considered in fuel production. Its basic structure could also be generated algorithmically since the generator of molecular structures proposed by Dahmen and Marquardt keeps track of substrate/product relationships. This includes the case where a single product can be derived from multiple bioderived platform chemicals as well as the case where multiple platform molecules have to be refunctionalized and joined into a single target molecule. Known performance measures of catalytic systems, i.e., selectivity and conversion data, have to be assigned manually to each pathway included in the CPM.

### Table 1. Fuel Properties Considered in the Design of Biofuel Blends with Tailored Properties

<table>
<thead>
<tr>
<th>Fuel Property</th>
<th>Impact on Engine Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Derived cetane number (DCN) [−]</td>
<td>Spark-ignition engine: low DCN (&lt;20) correlates with high RON/MON → high knock-resistance → high efficiency; compression-ignition engine: DCN above ~40 required for autoignition</td>
</tr>
<tr>
<td>Oxygen content [wt %]</td>
<td>Engine-out soot emissions</td>
</tr>
<tr>
<td>Lower heating value [MJ/kg]</td>
<td>Mileage per unit mass of fuel</td>
</tr>
<tr>
<td>Liquid density [kg/m³]</td>
<td>Mixture formation → emissions</td>
</tr>
<tr>
<td>Latent heat of vaporization [kJ/kg, at T]</td>
<td>Spark-ignition engine: cold-start/run, charge cooling</td>
</tr>
<tr>
<td>Surface tension [mN/m]</td>
<td>Mixture formation → emissions</td>
</tr>
<tr>
<td>Kinematic viscosity [mm²/s]</td>
<td>Mixture formation → emissions</td>
</tr>
<tr>
<td>Reid vapor pressure [kPa]</td>
<td>Spark-ignition engine: cold-startability, vapor lock</td>
</tr>
<tr>
<td>Distillation profile [°C]</td>
<td>Oil dilution, driveability; spark-ignition engine: cold-start/run, vapor-lock</td>
</tr>
</tbody>
</table>

*c.f. the work of Dahmen and Marquardt* for a detailed discussion on the interactions between physico-chemical fuel properties and engine performance.

by Gani and co-workers. In their framework, a mixture design problem is formulated as a mixed-integer nonlinear program which is solved by means of a decomposition-based approach, i.e., the solution to the original problem is sought by considering a series of solvable subproblems and therefore by performing a sequential reduction of the number of blend candidates. This approach has been applied to a variety of case studies including the formulation of gasoline- and diesel-like fuels where the design objective has been to find optimal blend components for a so-called main ingredient, i.e., a fossil gasolene/diesel pseudocompound that must constitute the main part of all mixtures under consideration. In the aforementioned works, fixed cost factors ($/L) assigned to the individual blend components have been the only process-related criteria explicitly included in the optimization problem. Recently, Ng and co-workers have developed a systematic methodology for mixture design in the context of integrated biorefineries that includes the identification of optimal conversion pathways via superstructure optimization once an optimal mixture has been determined based on CAMD techniques. Instead of such a two-step approach, Daoutidis and co-workers have proposed a strategy that simultaneously identifies gasoline blends with desired properties and the corresponding chemistries by combining an automatic rule-based generator of reaction networks with an RNFA-like approach. The design problem is formulated as a mixed-integer nonlinear program which is solved by means of a decomposition-based approach. In the subsequent section, this then allows to derive the problem formulation and to explain the solution strategy. A case study demonstrates the feasibility of the blend design approach. It is dedicated to the identification of a 100% renewable fuel mixture for the spark-ignition (SI) engine and respective production routes. Specifically, we investigate the scenario where the energy of fuel produced for a given amount of biomass shall be maximized. In order to meet this objective, renewable hydrogen from carbon-free energy sources such as wind or solar shall be provided. Finally, some conclusions and discussions are given.

### INTEGRATED PRODUCT AND PATHWAY DESIGN OF BIOFUELS

The blend design framework proposed in the present contribution is depicted in Figure 1 and comprises four stages. In stage 1, a set of so-called palette compounds has to be selected. The term *palette compound* refers to any compound that could be part of a rationally formulated blend, in principle. The set of palette compounds therefore defines the search space for the blend design. Palette compounds can either be assembled manually or generated automatically by the computer-aided fuel design approach proposed by Dahmen and Marquardt. Criteria for deciding whether a molecule should be included in the set of palette compounds include pure component properties and yield data, as well as additional process-related data if available.

Based on the set of palette compounds, a conversion pathway map (CPM) is created in stage 2. The CPM holds all conversion pathways to be considered in fuel production. Its basic structure could also be generated algorithmically since the generator of molecular structures proposed by Dahmen and Marquardt keeps track of substrate/product relationships. This includes the case where a single product can be derived from multiple bioderived platform chemicals as well as the case where multiple platform molecules have to be refunctionalized and joined into a single target molecule. Known performance measures of catalytic systems, i.e., selectivity and conversion data, have to be assigned manually to each pathway included in the CPM.
Models based on QSPR and group additivity are utilized to predict pure component property data for palette compounds where experimental data are not available.

Stage 3 deals with setting up and solving the actual mixture design problem which is a nonlinear program (NLP). Different process-related objective functions, e.g., maximization of production yield or maximization of the energy of fuel produced for a fixed feed of biomass, can be considered here. Mathematical constraints include the pathway model, i.e., mass balances of all species together with selectivity data and conversion limits, as well as the property models to predict the blend’s fuel properties. Due to nonlinearity in the property models, a sequential solution strategy will be pursued.

Stage 4 is dedicated to a detailed analysis of the resulting mixtures. This includes the inspection of blend compositions, fuel properties and material flow diagrams. While a local, necessary phase stability criterion is included in the NLP formulation, a more rigorous stability analysis is performed a posteriori in stage 4 to check both necessary and sufficient conditions for miscibility. Multiobjective optimization and Pareto analysis can examine the trade-off between competing objectives, e.g., between maximization of yield and maximization of the fuel’s knock resistance. After analyzing the proposed product and pathway designs, the modeler may choose to refine the scenario definition, for instance, by making changes to the set of palette compounds and/or fuel specifications or by including additional conversion pathways. In the following sections, key elements of the blend design framework depicted in Figure 1 will be explained in more detail, starting with the concept of conversion pathway maps.

**PATHWAY MODEL**

**Extension of Reaction Network Flux Analysis.** Figure 2 shows the graphical illustration of an example conversion pathway map (CPM). The nodes of the CPM include biomass, hexoses, pentoses, platform molecules (shaded in gray), and palette compounds (encircled by dashed lines). Byproducts and intermediates are not considered here since the focus lies strictly on fuel production. Water, CO₂, and hydrogen are omitted in the graphical representation for the sake of simplicity.

---

**Figure 2.** Graphical illustration of an exemplary conversion pathway map (CPM). A platform molecule (shaded in gray) can also be a palette compound (encircled by dashed lines), like in the case of ethanol. Note that other molecules, i.e., intermediates, byproducts, water, CO₂, and hydrogen, are omitted in the graphical representation for the sake of simplicity.

---

**Figure 1.** Blend design framework: integrated product and pathway design.
stationary mole balances of a CPM comprising \( n_N \) nodes and \( n_P \) pathways can be formulated as

\[
A \mathbf{f} = \mathbf{b}
\]

(1)

Here, \( A \in \mathbb{R}^{n_N \times n_P} \) is the matrix of stoichiometric coefficients \( a_{i,k} \), where the rows correspond to the \( i \in \{1, \ldots, n_N\} \) nodes, i.e., reactants, and the columns correspond to the \( k \in \{1, \ldots, n_P\} \) pathways, i.e., reactions. The vector \( \mathbf{f} \) holds the molar fluxes over all pathways \( h \), whereas the vector \( \mathbf{b} \) describes the molar product flows.\(^{17}\)

In the original RNFA,\(^{14}\) yield constraints are formulated to account for the fact that ideal reaction performance as described by eq 1 is rarely achieved. These yield constraints act as conversion limits, i.e., selectivities \( S_h \) are assumed to be one and yields \( Y_h \) are effectively identical to conversions \( C_h \), because of \( Y_h = C_h S_h \). However, in practice, yields are more often constrained by low selectivity rather than by low conversion. In such cases, the yield constraints can lead to unrealistic results, especially in the context of mixture design where a single reactant can often be converted into a multitude of fuel components via independent pathways. To overcome this issue, both conversion \( C_h \) and selectivity \( S_h \) of each pathway \( h \) can be included in the model formulation. By introducing an artificial mismatch into the stoichiometric balance of a pathway \( h \), selectivities \( S_h \) smaller than one can be accounted for. For instance, the pseudo reaction for pathway \( P6 \) (cf. Figure 2)

\[
\text{C}_4\text{H}_8\text{O}_4 \; (\text{succinic acid}) + 2\text{H}_2
\rightarrow \; \text{P6}(\text{C}_4\text{H}_6\text{O}_2 \; (\gamma\text{-butyrolactone}) + 2\text{H}_2\text{O})
\]

(2)

recognizes the fact that unspecified byproducts and/or wastes are created if \( S_{p6} < 1 \). Thus, in the model, the selectivity information is encoded in the matrix \( A \).

Regarding the formulation of pseudo reactions it is therefore necessary to distinguish between coproducts that constitute fuel components and unspecified byproducts and/or waste. The fuel component coproduct is considered a palette compound. For instance, pathway \( P11 \) in Figure 2 has one main product \((\text{cylopentanone}) \) and two coproducts \((\text{cylopentanol and 2-methyltetrahydrofuran}) \). All three species constitute palette compounds and are balanced accordingly. Byproducts, most notably chemical products, and waste, however, are neglected in the pseudoreactions for two reasons. First, the focus here lies strictly on fuel production and, as such, the analysis of an integrated biorefinery that produces both fuels and chemicals is considered beyond the scope of this contribution, but offers an obvious extension of the suggested methodology. Second, whereas selectivity and conversion data for the main products of a conversion pathway can be extracted from the available literature in the great majority of cases, comprehensive (quantitative) information about the spectrum of byproducts is mostly lacking.

If the case of limited conversion shall be modeled, the following general form of a conversion constraint for a limiting reactant \( i \) in pathway \( h \) is used:

\[
0 \geq \sum_{k \neq h} a_{i,k} f_h - C_h
\]

(3)

The extension of RNFA from yield to selectivity and conversion limitation is best illustrated using a simple example, i.e., the two-step conversion of succinic acid \((\text{species A}) \) into \(\gamma\text{-butyrolactone (species B) and tetrahydrofuran (species C)} \) as depicted in Figure 3. A mole flow \( f_{p3} \) of 1 kmol/h succinic acid is supplied to this sample network. Note that both \(\gamma\text{-butyrolactone and tetrahydrofuran are considered palette compounds. The pathway yields } Y_{p6} \text{ and } Y_{p7} \text{ shall be 0.8 and 0.6, respectively. The stoichiometric matrix } A \text{ for this example reads}

\[
A = \begin{bmatrix}
1 & -1 & 0 \\
0 & 1 & -1 \\
0 & 0 & 1
\end{bmatrix}
\]

(4)

and the RNFA yield constraints\(^{17}\) for pathways \(P6\) and \(P7\) are formulated as

\[
Y_{p6} \geq \frac{u_{b,p6} f_{p6}}{\sum_{k \neq p6} u_{b,k} f_k} = \frac{1 f_{p6}}{f_{p3} + 0 f_{p7}} = \frac{f_{p6}}{f_{p3}} \Rightarrow Y_{p6} f_{p3} \leq Y_{p6} f_{p3}
\]

(5)

\[
Y_{p7} \geq \frac{u_{b,p7} f_{p7}}{\sum_{k \neq p7} u_{b,k} f_k} = \frac{1 f_{p7}}{0 f_{p3} + 1 f_{p6}} = \frac{f_{p7}}{f_{p6}} \Rightarrow Y_{p7} f_{p6} \leq Y_{p7} f_{p6}
\]

(6)

Considering the simple objective of maximizing the mole flow of a blend composed of B and C, the linear program can then be written as

\[
\text{max } (b_B + b_C)
\]

s.t.

\[
A \mathbf{f} = \mathbf{b}
\]

(7)

\[
f_{p3} = 1
\]

(8)

\[
f_{p6} - Y_{p6} f_{p3} \leq 0
\]

(9)

\[
f_{p7} - Y_{p7} f_{p6} \leq 0
\]

(10)

The solution to this problem, i.e., vectors \( \mathbf{f} \) and \( \mathbf{b} \) yielding the maximum product flow of 0.8 kmol/h \((\text{calculated as 0.5333 kmol/h of }\gamma\text{-butyrolactone and 0.2667 kmol/h of tetrahydrofuran}) \), is shown in Figure 3. Thus, the result is a binary blend composed of \(\gamma\text{-butyrolactone and tetrahydrofuran with the molar ratio of 2:1. Note that there is also unconverted succinic acid leaving the network (0.2 kmol/h).}

Separately accounting for selectivity \(S_h\) and conversion \(C_h\) allows to investigate the scenario where yield is actually limited due to selectivity and not due to conversion, i.e., \(S_{p6} = 0.8, S_{p7} = 0.6, C_{p6} = 1.0, \) and \(C_{p7} = 1.0 \). Then the stoichiometric matrix \( A \) needs to be modified to

\[
A = \begin{bmatrix}
1 & -1 & 0 \\
0 & 0.8 & -1 \\
0 & 0 & 0.6
\end{bmatrix}
\]

(12)
and the optimization problem is rewritten as

$$\max_{f_2 \geq 0, b_2 \geq 0} (b_8 + b_3)$$

s.t. $A \mathbf{f} = \mathbf{b}$

$$f_{P3} = 1$$

The solution to this scenario is shown in Figure 4. Instead of a binary mixture, pure $\gamma$-butyrolactone is produced here. Moreover, there is no succinic acid left.

Finally, assume that there is also a conversion limitation for pathway $P6$, i.e., $C_{P6} = 0.9$. Adapting the general form of a conversion constraint (cf. eq 3) to this example yields

$$0 \geq -\eta_{i,P6} f_{P6} - C_{P6} = f_{P6} - C_{P6} f_{P3} \geq 0 \Rightarrow f_{P6} \leq C_{P6} f_{P3}$$

Note that eq 16 gives a constraint that is identical to the one stated in eq 5, thus demonstrating that the original RNFA yield constraints act as conversion constraints. The optimization problem is reformulated as

$$\max_{f_2 \geq 0, b_2 \geq 0} (b_8 + b_3)$$

s.t. $A \mathbf{f} = \mathbf{b}$

$$f_{P6} - C_{P6} f_{P3} \leq 0$$

$$f_{P3} = 1$$

Figure 5 shows the solution to this problem. As expected, unconverted succinic acid (0.1 kmol/h) leaves the network. In combination with the selectivity constraints for pathways $P6$ and $P7$, the conversion limitation leads to a product flow of only 0.72 kmol/h (pure $\gamma$-butyrolactone).

**Biomass Supply.** It is assumed that a constant flow of biomass is supplied to the CPM via a pathway $S1$ to be included in the stoichiometric matrix $A$. Although being omitted in the graphical illustration of the CPM (cf. Figure 2), a hydrogen supply pathway $S2$ is added to matrix $A$, thus allowing to balance the hydrogen demand associated with the production of a specific fuel blend. Analogously, a CO$_2$ mole balance is included to assess the amount of CO$_2$ generated from the biomass.

The supplied biomass is fractionated and depolymerized into hexoses and pentoses in the first conversion step (cf. Figure 2, pathway P1). A variety of processing concepts, e.g., Organo- solv, Organocat, or mechanocatalytic strategies, have been proposed for this purpose. Although hemicellulose is a polymer of different sugars, its most abundant building block is xylose, i.e., a polymer of xylose. Therefore, hemicellulose is often pragmatically treated as a polymer of the monomer unit C$_5$H$_8$O$_4$, as can be seen from Table 2.

<table>
<thead>
<tr>
<th>Lignocellulosic Biomass and Corresponding Monomer Units</th>
<th>Fraction of Biomass</th>
<th>Corresponding Monomer Unit</th>
<th>Monomer Molar Weight [g/mol]</th>
</tr>
</thead>
<tbody>
<tr>
<td>cellulose (polymer of glucose)</td>
<td>0.477</td>
<td>C$<em>6$H$</em>{12}$O$_6$</td>
<td>162.14</td>
</tr>
<tr>
<td>hemicellulose (polymer of xylose)</td>
<td>0.351</td>
<td>C$_5$H$_8$O$_4$</td>
<td>132.12</td>
</tr>
<tr>
<td>lignin</td>
<td>0.20</td>
<td>C$<em>6$H$</em>{12}$O$_6$</td>
<td>180.20</td>
</tr>
</tbody>
</table>

For a fixed, hypothetical biomass composition of 50 wt % cellulose (polymer of glucose), 30 wt % hemicellulose (polymer of xylose), and 20 wt % lignin, 1 kmol of biomass monomers (∼154.71 kg of biomass) can theoretically yield 0.477 kmol of hexoses (C$_6$H$_{12}$O$_6$) and 0.351 kmol of pentoses (C$_5$H$_8$O$_4$) after hydrolysis. Therefore, one can formulate the pseudoreaction

$$\text{biomass} \rightarrow S_{PL,C6} \cdot 0.477 \text{hexoses(C}_6\text{H}_{12}\text{O}_6)$$

$$+ S_{PL,C5} \cdot 0.351 \text{pentoses(C}_5\text{H}_8\text{O}_4)$$

which, in the case of $S_{PL,C6} = S_{PL,C5} = 1$, would represent a perfect recovery of hexoses and pentoses. Note that the composition of beechwood is similar to the 50/30/20 composition used here. The selectivity constants $S_{PL,C5}$ and $S_{PL,C6}$ can be used to adjust the yields of hexoses and pentoses depending on the performance of a practical fractionation and depolymerization process. Pragmatically, we have set $S_{PL,C5} = S_{PL,C6} = 0.9$ based on a recent overview of different depolymerization approaches published by Luterbacher, Alonso, and Dumesic. Obviously, instead of a single conversion pathway P1, multiple fractionation and depolymerization pathways could be included in one CPM to account for the fact that different processing concepts might yield different shares of C$_5$ and C$_6$ sugars.

**Calculation of Process Performance Measures.** In order to compute process-related performance indicators, e.g., hydrogen demand, mass of fuel produced or lower heating value (LHV) energy efficiency, the entries of the solution vector $\mathbf{b}$ referring to the amounts of palette compounds produced need to be investigated.

The total molar flow of fuel produced is defined as

$$n_{\text{fuel}} = \sum_{i \in \text{PC}} b_i$$

*Corresponding Monomer Units* follows.

Figure 4. Modified RNFA formulation with selectivity constraints: two-step conversion of succinic acid (species A) into $\gamma$-butyrolactone (species B) and tetrahydrofuran (species C).

Figure 5. Modified RNFA formulation with selectivity and conversion constraints: Two-step conversion of succinic acid (species A) into $\gamma$-butyrolactone (species B) and tetrahydrofuran (species C).
where PC denotes the set of palette compounds included in the CPM. Similarly, the total mass flow of fuel produced can be derived from

\[ m_{fuel} = \sum_{i \in PC} b_i \cdot M_i \]  

(23)

where \( M_i \) denotes the molar weight of palette compound \( i \).

Then, mole \( z_i \) and mass \( \xi_i \) fractions of all blend components can be calculated from

\[ z_i = \frac{b_i}{\hat{n}_{fuel}} \quad \forall \ i \in PC \]  

(24)

\[ \xi_i = \frac{b_i \cdot M_i}{\hat{m}_{fuel}} \quad \forall \ i \in PC \]  

(25)

The total energy flow of fuel produced is defined as

\[ \hat{E}_{fuel} = \sum_{i \in PC} b_i \cdot M_i \cdot \text{LHV}_i \]  

(26)

where \( \text{LHV}_i \) is the lower heating value of blend component \( i \).

The stoichiometric hydrogen demand for fuel production is computed based on \( f_{S2} \), i.e., the molar flow of hydrogen supplied to the network. To this end, the constraint \( b_{H_2} = 0 \) is added making sure that there is no residual hydrogen leaving the network. It is convenient to normalize the mass-based hydrogen demand by division with the mass of fuel produced, i.e.,

\[ m_{H_2} = \frac{f_{S2} \cdot M_{H_2}}{\hat{m}_{fuel}} \]  

(27)

where \( M_{H_2} \) denotes the molar weight of hydrogen \((2.02 \text{ kg/kmol})\).

Similarly, the mole-based hydrogen consumption can be stated as

\[ n_{H_2} = \frac{f_{S2}}{\hat{n}_{fuel}} \]  

(28)

The amount of \( \text{CO}_2 \) generated from the carbohydrate fraction of the biomass is related to the energy of fuel produced, i.e.,

\[ m_{\text{CO}_2} = \frac{b_{\text{CO}_2} \cdot M_{\text{CO}_2}}{\hat{E}_{fuel}} \cdot 1000 \]  

(29)

where \( b_{\text{CO}_2} \) is the molar product flow of \( \text{CO}_2 \) leaving the network and \( M_{\text{CO}_2} \) denotes the molecular weight of carbon dioxide \((44.01 \text{ kg/kmol})\).

Given a fixed, hypothetical supply of biomass of 1 kmol/h \((\sim 154.71 \text{ kg/h})\), the LHV efficiency \( (\eta_{\text{LHV}}) \) can be computed from

\[ \eta_{\text{LHV}} = \frac{\hat{E}_{fuel}}{154.71 \cdot \text{LHV}_{\text{biomass}} + f_{S2} \cdot M_{H_2} \cdot \text{LHV}_{H_2}} \]  

(30)

where \( \text{LHV}_{H_2} \) denotes the lower heating value of hydrogen \((120 \text{ MJ/kg})\). Based on the dry biomass composition of 50 wt % cellulose \((\text{monomer } C_6H_{10}O_5)\), 30 wt % hemicellulose \((\text{monomer } C_6H_{12}O_5)\), and 20 wt % lignin \((\text{monomer } C_{10}H_{12}O_3)\), the lower heating value \( \text{LHV}_{\text{biomass}} \) in eq 30 is estimated as 18.6 MJ/kg with the help of a regression model proposed by Demirbas et al. Typical lower heating values for wood reported in the literature range from 18.6 to 18.9 MJ/kg depending on type and source.

### FUEL PROPERTY MODELS

Predictive models are required to evaluate pure component and mixture properties relevant to fuel design (cf. Table 1). From Table 3 it can be seen that with the exception of derived cetane number \( \text{(DCN)} \) and activity coefficient, QSPR models are used to determine the primary properties in the present contribution. Primary properties are those properties which are predicted solely based on the molecular structure of a compound. In contrast, secondary properties are derived on the basis of primary properties \((\text{and/or secondary properties})\) and additional relationships. The rationale behind choosing the respective property models is briefly summarized in the last column of Table 3. A particular feature of the QSPR and DCN models is their tailored range of applicability. This has been achieved by training on large, diverse sets of \( C_xH_yO_z \) structures and by including high-quality property data only. Both the DCN and the QSPR models have been validated carefully and some of them have been successfully used in fuel design before.

Statistical measures for the models can be found in the Supporting Information. It shall be noted that the outlined approach has been designed to be ”fully predictive”. We refer to a property model as being fully predictive, if the required component-specific input data can be estimated solely on the basis of the two-dimensional molecular graph with the given set of methods. This way, the blend design approach can also cope with novel compounds which have not been characterized experimentally yet.

Activity coefficients are required to describe the nonideal behavior of mixtures with regard to Reid vapor pressure and distillation profile, two important measures for fuel volatility. For the present contribution, the modified UNIFAC (Dortmund) model has been implemented in Matlab. UNIFAC has broad range of applicability; parameters for functional groups instead of molecules/mixtures.

### Table 3. Group Contribution (GC) and Quantitative Structure–Property Relationship (QSPR) Models Used to Predict Primary Properties

<table>
<thead>
<tr>
<th>method</th>
<th>properties</th>
<th>rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>GC model(^{4,5})</td>
<td>derived cetane number ( \text{DCN} ) ([-])</td>
<td>broad range of applicability ((\text{tailored for } C_xH_yO_z \text{ structures})); high-quality IQT (^{6}) only training data; cross-validation and external validation</td>
</tr>
<tr>
<td>QSPR models(^{4,5})</td>
<td>critical data ((\text{temperature } T_{\text{cr}} \text{ [K]}, \text{ pressure } p_{\text{cr}} \text{ [bar]}, \text{ volume } V_{\text{cr}} \text{ [cm}^3]/[\text{mol}]), \text{ normal boiling point } T_{\text{f}} \text{ [K]})); ( \text{density} \rho \text{ [kg/m}^3] ) at 298 K, ( \text{lower heating value} \text{ LHV} \text{ [MJ/kg]} ) at 298 K, ( \text{surface tension} \sigma \text{ [mN/m]} ) at 298 K, ( \text{dynamic viscosity} \mu \text{ [Pa s]} ) at 298 K</td>
<td>based on partial least-squares regression ((\text{PLSR}); \text{ cross-validation and external validation})); broad applicability ((\text{tailored for } C_xH_yO_z \text{ structures})); high-quality training data ((\text{DIPPR database} )^{5,6})</td>
</tr>
<tr>
<td>UNIFAC GC model (modified UNIFAC ((\text{Dortmund}))(^{55-58})</td>
<td>activity coefficient ( \gamma ) ([-])</td>
<td>broad range of applicability; parameters for functional groups instead of molecules/mixtures</td>
</tr>
</tbody>
</table>

\(^{4}\)Average relative errors for training and independent test sets are provided in the Supporting Information \((\text{SI})\). \(^{5}\)The Ignition Quality Tester \((\text{IQT})\) is a standardized constant-volume combustion chamber used to measure the derived cetane number \((\text{DCN}; \text{ASTM D6980})\).
been chosen because it is a widely used group contribution method that is fully predictive and applicable to the variety of molecular structures considered here. Moreover, it is of limited computational complexity to be used in an optimization-based approach. Different variants of the UNIFAC approach have been used extensively in the context of CAMD and mixture design before.\(^{24,25,61}\) Only recently, first CAMD and mixture design approaches based on SAFT\(^{60–67}\) and COSMO-RS\(^{68,69}\) have been presented. In contrast to the original UNIFAC model developed by Fredenslund et al.,\(^{70}\) the modified UNIFAC (Dortmund) model proposed by Gmehling and co-workers\(^{53–55}\) features temperature-dependent group interaction parameters and an empirically modified combinatorial part that is thought to offer a better description of asymmetric systems.\(^{55,71}\) In quantitative comparisons with the original UNIFAC formulation, modified UNIFAC (Dortmund) has demonstrated significantly improved results in different areas, including the prediction of vapor–liquid equilibria (VLE) and liquid–liquid equilibria (LLE) mole fractions.\(^{55,71}\) It is therefore used throughout the present contribution.

Linear blending rules are the simplest mixing rules and they are often used to describe fuel properties due to the lack of better alternatives.\(^{55,72}\) The error caused by the assumption of linear blending depends on the property and mixture under consideration. For instance, the error in case of the specific volume is typically less than 3.5%, whereas for the liquid viscosity of a mixture the linear rule can only be expected to yield the correct order of magnitude.\(^{7}\) The following linear and nonlinear mixing rules for the liquid density \(\rho_L\), the surface tension \(\sigma\) and the kinematic viscosity \(\nu\) have been compiled based on the compendium of property modeling published by Gmehling et al.\(^{71}\) The chosen models are considered to be among the best in class under the premise that they are fully predictive and applicable to the diversity of molecular structures considered here.

In order to approximate the liquid density \(\rho_L\) [kg/m\(^3\)] of a mixture, a linear mixing rule is applied to the specific volume \(\nu_L\) [m\(^3\)/kg] where the contribution of the excess volume is neglected:\(^{71}\)

\[
\nu_L = \sum_{i \in PC} \xi_i \rho_{L,i} \Rightarrow \rho_L = \left( \sum_{i \in PC} \xi_i \rho_{L,i} \right)^{-1}
\]

(31)

According to Gmehling et al.,\(^{71}\) this simplified density calculation typically leads to small errors and eq 31 should be preferred over a linear mixing rule for the density \(\rho_L = \sum_{i \in PC} \xi_i \rho_{L,i}\).

The surface tension \(\sigma\) of a blend is computed with the help of a Parachor-based mixing rule,\(^{71}\) i.e.,

\[
\sigma = \left( \rho_{\text{m}} \sum_{i \in PC} \xi_i \rho_{L,i} \frac{(\rho_{m,i})^{-1} \sigma_{ij}^{-1/4} + (\rho_{m,i})^{-1} \sigma_{ij}^{-1/4}}{2} \right)^4 \text{[mN/m]}
\]

(32)

where \(\rho_{\text{m}}\) [mol/cm\(^3\)] is the molar density of the blend.

The blend’s kinematic viscosity \(\nu\) is calculated by assuming linear mixing for the dynamic viscosity \(\mu\), i.e.,

\[
\nu = \frac{1}{\rho_L} \sum_{i \in PC} \xi_i \mu_i
\]

(33)

Some authors\(^{75,74}\) have proposed to employ the Clausius–Clapeyron equation (CCE) to derive an estimate for the latent heat of vaporization \(H_{\text{vap}}\) of fuel mixtures. Assuming the ideal gas law for the vapor phase and further assuming the gas volume to be much larger than the liquid volume, i.e., \(\nu_g \gg \nu_L\), the CCE is written as\(^{71}\)

\[
H_{\text{vap}} = \frac{RT^2}{\rho^S} \frac{dp^S}{dT}
\]

(34)

where the vapor pressure of the mixture \(p^S\), i.e., the bubble point pressure, at a given temperature \(T\) can be calculated from\(^{75}\)

\[
p^S(T) = \sum_{i \in PC} z_i \gamma(T, x) p_i^S(T)
\]

(35)

In eq 35, \(p_i^S\) and \(\gamma_i\) denote the vapor pressure and activity coefficient of component \(i\), respectively. Equation 35 can be differentiated to compute \(dp_i^S/dT\) in eq 34. However, strictly speaking, the Clausius–Clapeyron equation is valid for the phase transition of a pure compound at a constant temperature only.

Chupka et al.\(^{76}\) have recently shown that the application of CCE to ethanol–hydrocarbon mixtures will significantly underestimate the true heat of vaporization. This can be explained by the fact that the vapor pressure is predominantly influenced by the most volatile component, whereas the entire mixture is often wide-boiling.\(^{76}\) As an alternative approach, Chupka et al.\(^{76}\) have suggested to employ the linear mixing rule

\[
H_{\text{vap}} = \sum_{i \in PC} \xi_i H_{\text{vap},i}
\]

(36)

which is based on the enthalpies of vaporization \(H_{\text{vap},i}\) of the individual constituents of the blend. Such estimates were found to compare favorably with experimentally determined heats of vaporization, although the enthalpy of mixing is not accounted for by this approach.\(^{76,77}\) Based on these findings, eq 36 will be used to obtain an estimate for the heat of vaporization of a mixture in the present contribution, whereas the enthalpy of vaporization of a pure compound, i.e., \(H_{\text{vap},i}\) is computed from the Clausius–Clapeyron equation (cf. eq 34). To this end, the vapor pressure \(p_i^S\) of a pure compound at temperature \(T\) is calculated by means of the extended Antoine equation,\(^{71,75}\) i.e.,

\[
\ln(p_i^S(T)) = C_{\text{ant},1,i} + \frac{C_{\text{ant},2,i}}{T + C_{\text{ant},3,i}} + C_{\text{ant},4,i}T + C_{\text{ant},5,i} \ln(T) + C_{\text{ant},6,i}T^{C_{\text{ant},7,i}}
\]

(37)

if parameter values \(C_{\text{ant},1,i}\), \(C_{\text{ant},2,i}\), \(C_{\text{ant},3,i}\), \(C_{\text{ant},4,i}\), \(C_{\text{ant},5,i}\) and \(C_{\text{ant},7,i}\) are available. Alternatively, the vapor pressure is considered a secondary property and the Hoffman–Florin (HF) equation\(^{71,78}\) is applied as part of a fully predictive approach (cf. Supporting Information):

\[
p_i^S(T) = f_{\text{HF}}(T_{\text{crit},i}, P_{\text{crit},i}, T_{\text{boil},i}, T)[\text{kPa}]
\]

(38)

The material specific inputs to eq 38, i.e., the critical temperature \(T_{\text{crit},i}\), the critical pressure \(P_{\text{crit},i}\) and the normal boiling point \(T_{\text{boil},i}\) are primary properties that are predicted by means of QSPR (cf. Table 3).

The lower heating value LHV and the derived cetane number DCN of the blend are calculated based on the assumption of linear mixing as well, i.e.,

\[
\text{LHV} = \sum_{i \in PC} \xi_i \text{LHV}_i
\]

(39)
DCN = \sum_{i \in PC} z_i \cdot \text{DCN}_i \quad (40)

Note that in eq 40, the dimensionless DCN is weighted by mole fractions due to the fact that the mixture octane number (ON) of a gasoline surrogate model composed of \( n \)-heptane/iso-octane/toluene \(^9\) as well as the DCN of mixtures of furanic species and \( n \)-heptane \(^{40}\) have been observed to follow linear-by-mole blending rules. Nonlinear synergistic/antagonistic blending effects, however, have been reported for ON in case of ethanol and toluene. \(^{51,82}\) Although cetane number of diesel fuel is generally believed to closely follow a linear blending rule, \(^{83}\) in principle, nonlinear effects are plausible from a kinetics point of view, especially if fuels with different chemistries are investigated.

**Distillation Curve Model.** Fuel volatility is important for in-cylinder mixture formation. In both spark-ignition (SI) and compression-ignition engines, too heavy components can cause engine oil dilution. \(^{39,59,84}\) Especially in SI engines, high fuel volatility is critical in obtaining a well-mixed, ignitable fuel/air mixture. Engine cold-start and low ambient temperatures place particular demands on the fuel in this respect. \(^{11,37,39,41}\) Although the compression-ignition engine is generally capable of burning less volatile fuel, \(^{85}\) high volatilities seen in some biofuels and certain diesel blends have been linked to better mixture homogenization in low-temperature combustion concepts. \(^{1,33,86-90}\)

It shall be noted that the flash point is a critical volatility-related parameter of diesel fuel for safety reasons, as only at temperatures above the flash point (typically at least 55 °C) an ignitable mixture of gaseous fuel and air is formed. \(^{34,91,92}\) In case of gasoline, the vapor above the liquid fuel is typically too rich to be flammable as a result of gasoline’s high volatility. \(^{34}\)

The standard way of measuring fuel volatility is to perform a batch distillation according to ASTM D86. \(^{85}\) This yields a so-called distillation curve, i.e., a (graphical) relationship between the temperature of the fuel vapor and the volume fraction of the vapor and toluene. \(^{81,82}\) Although cetane number of diesel fuel is generally believed to closely follow a linear blending rule, \(^{83}\) in principle, nonlinear effects are plausible from a kinetics point of view, especially if fuels with different chemistries are investigated.

![Figure 6. Conceptual model of an idealized batch distillation process.](image)

**Figure 6.** Conceptual model of an idealized batch distillation process.

\[
\frac{dn_{D,i}}{dt} = N y_{L,i} \quad \forall \; i = 1, ..., n_C - 1 \quad (41)
\]

\[
\frac{dV_D}{dt} = \frac{\dot{N}}{\rho_m(T_D, y_{L,i})} \quad (45)
\]

\[
y_{L,i} = \frac{x_{L,i} \gamma(T_L, x_L) p_i^S(T_L)}{p} \quad \forall \; i = 1, ..., n_C \quad (46)
\]

\[
x_{L,i} = \frac{n_{L,i}}{n_L} \quad \forall \; i = 1, ..., n_C \quad (47)
\]

\[
x_{D,i} = \frac{n_{D,i}}{n_D} \quad \forall \; i = 1, ..., n_C \quad (48)
\]

\[
n_D = \sum_i n_{D,i} \quad (49)
\]

\[
n_L = \sum_i n_{L,i} \quad (50)
\]

\[
p = \sum_i x_{L,i} \gamma(T_L, x_L) p_i^S(T_L) \quad (51)
\]

\[
p = \sum_i x_{D,i} \gamma(T_D, x_D) p_i^S(T_D) \quad (52)
\]

Equations 41 and 42 represent the total mole balances for the liquid (L) in the distillation flask and for the distillate (D) in the distillate receiver, whereas eqs 43 and 44 are the component mole balances for the liquid and the distillate, respectively. \(^{75}\) At each point in time, the liquid is assumed to be at its bubble point. \(^{96}\) Mathematically, the vapor–liquid equilibrium can be written as eq 46, where \(y_{L,i}\) denote the mole fractions of the vapor which is assumed to be in equilibrium with the liquid. \(^{75}\) In eq 46, the notation \(x_L = [x_{L,1}, x_{L,2}, ..., x_{L,n_C}]^T\) is used to indicate that the mole fractions of all species are needed to calculate the activity coefficient \(y\) of species \(i\).

It is further assumed that the vapor stream \(N\) is immediately transferred to and totally condensed in the distillate receiver. \(^{96}\) The change in the distillate volume \(V_D\) is computed based on the liquid molar density of the condensate at the bubble point.
temperature $T_D$ of the distillate in the distillate receiver. This allows to derive a differential equation for the volume of the distillate (cf. eq 45). The molar fractions of the liquid and the distillate are defined by eqs 47 and 48, whereas eqs 49 and 50 represent the closure conditions for the liquid and the distillate, respectively. The bubble point temperatures $T_L$ and $T_D$ have to be calculated by iteratively solving eqs 51 and 52, i.e.,

$$ f(T) = \sum_i \gamma_i(T) \cdot p_i^S(T) - p = 0.75 $$

Beginning at $t = 0$ and setting $N = 1$, the entire liquid is distilled at $t = 1$. The distillation is assumed to take place at sea level atmospheric pressure ($p = 1.01325$ bar).

Again, the vapor pressure $p_i^S$ of a pure compound $i$ at temperature $T$ (cf. eqs 46, 51, and 52) is calculated by means of the extended Antoine equation (cf. eq 37). Alternatively, the HF equation (cf. eq 38) is applied as part of a fully predictive approach, if Antoine parameter values are not available.

The liquid molar density $\rho_m$ of the condensed vapor in eq 45 is described as

$$ \rho_m(T, X_{li}) = \left( \sum_i \frac{Y_i}{\rho_{mi}(T)} \right)^{-1} $$

where the liquid molar density $\rho_{mi}$ of species $i$ is computed from the DIPPR 105 equation

$$ \rho_{mi}(T) = \frac{C_{dip,1,i}}{C_{dip,2,i}^{1+(1-T/C_{crit,i})^{\delta_{1},a,i}}} $$

if parameter values $C_{dip,1,i}$, $C_{dip,2,i}$, $C_{dip,3,i}$ and $C_{dip,4,i}$ are available. Alternatively, the COSTALD method (cf. Supporting Information) is applied as part of a fully predictive approach.

$$ \rho_m(T) = f_{COSTALD}(T_{crit,i}, V_{crit,i}, P_{crit,i}, T_{boil,i}, T) \text{[kmol/m}^3] $$

The four material-specific inputs to eq 55, i.e., critical temperature $T_{crit,i}$, critical volume $V_{crit,i}$, critical pressure $P_{crit,i}$ and the normal boiling point $T_{boil,i}$ are primary properties that are predicted by means of QSPR (cf. Table 3).

Once the solution to eq 45 is available, the evolution of the volume fraction distilled $V_f$ is calculated based on the final volume of distillate $V_f(t = 1)$, i.e.,

$$ V_f(t) = \frac{V_f(t = 1)}{V_f(t = 1)} $$

Having built a similar model, Huber et al. observed a horizontal offset between experimental and computed data when plotting the temperature $T_D$ over the volume fraction $V_f$. This offset is thought to correspond to the fluid transit delay, i.e., the time it takes for the vapor formed above the liquid in the kettle to travel upward, to condense in the condenser and to be eventually collected in the distillate receiver. Huber et al. have proposed to correct for this delay by shifting the computed distillation profile leftwards by a constant volume increment $V_{shift}$ to be found by manual adjustment.

Figure 7 compares computed and measured ADC data for a diesel fuel surrogate proposed by Mueller et al. and compared with experimental data generated by the Advanced Distillation Curve (ADC) method. Experimental data are also taken from Mueller et al. Activity coefficients have been calculated by modified UNIFAC (Dortmund) Activity coefficients have been calculated by modified UNIFAC (Dortmund). Note: A volume shift $V_{shift}$ of $-0.05$ has been applied to compensate for the fluid transit delay.

Reid Vapor Pressure Model. The Reid vapor pressure (RVP) as defined in ASTM D323 is another important measure for the volatility of gasoline fuels. RVP is highly relevant for the ability of performing engine start when the engine is cold relative to its design operating temperature. It is also important to prevent vapor-lock, i.e., the formation of fuel vapor bubbles in the fuel delivery system which leads to abnormal fuel flow to the combustion chamber. Conceptually, the Reid vapor pressure is the pressure $P_{RVP}$ in a closed system in which the biphasic fuel is stored at a constant temperature of 37.8 °C with a constant vapor-to-liquid volume ratio $V_{FL}/V_L$ of four. Hatzioannidis et al. have shown that UNIFAC is generally capable of predicting the RVP of simulated gasoline blended with a variety of oxygenated additives and fuel extenders. More recently, Backhaus and Rothamer have employed UNIFAC for RVP prediction as part of a design approach for drop-in gasoline fuels satisfying certain volatility constraints. The Reid vapor pressure test can be modeled by the flash equations:

$$ n = n_L + n_V $$

$$ n_{z_i} = n_L x_i + n_V y_i $$

$$ y = K x_i $$

In eq 57, $n$ is the total moles, whereas $n_L$ and $n_V$ are the moles of the liquid and the vapor phase, respectively. The overall mole fractions $z_i$ are determined from eq 58 based on the mole fractions for the liquid phase $x_i$ and for the vapor phase $y_i$. The vapor–liquid equilibrium is described by the equilibrium ratio $K_i$ in eq 59. Inserting eqs 57 and 59 into eq 58 and performing rearrangement gives

$$ x_i = z_i/[1 + (K_i - 1) n_{z_i}/n] $$

The volume of the gas phase is determined from the ideal gas law, i.e.,

$$ V_g = n_{RT}/P_{RVP} $$

and the volume of the liquid phase is given by $V_L = n_L/\rho_{mol}$ where the molar density of the liquid $\rho_m$ is again calculated from eq 53. The vapor-to-liquid volume ratio of four then leads to the following ratio between $n_V$ and $n$:

$$ 4 = \frac{V_g}{V_L} = \frac{n_V RT \rho_m}{(n - n_V) P_{RVP}} \Leftrightarrow n_V = \frac{4 P_{RVP}}{RT \rho_m + 4 P_{RVP}} $$
Assuming that $x^*$ is the global minimizer of eq 67 such that the closure condition (cf. eq 68) holds, the necessary and sufficient condition for phase stability is $f_{tpd}(x^*) \geq 0$. Due to the strong nonlinear behavior of eq 67 and the existence of many local minima, several approaches to global optimization have been applied to the tangent plane stability problem in combination with equations of state, NRTL and UNIQUAC. In the present contribution, the solver globalsearch from Matlab’s Global Optimization toolbox is used to solve the problem defined by eqs 67 and 68. Globalsearch is an implementation of the OQNLP algorithm, i.e., a multistart heuristic algorithm combining scatter search with local NLP solvers. Ugray et al. have reported that the OQNLP algorithm is capable of finding global solutions to all 142 continuous-variable NLP test problems compiled by Floudas et al. However, it should be noted that due to its heuristic nature the algorithm cannot guarantee that the global minimum of the tangent plane distance function $f_{tpd}$ is identified.

## PROBLEM FORMULATION AND SOLUTION STRATEGY

After pathway and property models have been defined, the blend design problem can now be formulated as a nonlinear program (NLP). Ultimately, the purpose of fuel production is the provision of energy for transport applications. Therefore, the objective is chosen as the maximization of the energy of fuel produced (in terms of lower heating value) for a given feed of biomass. Then, the problem can be stated as follows:

**Objective function**

$$\min_{\varepsilon \geq 0 \geq b \geq 0} \sum_{j \in PC} b_i M_i LHV_j$$

s.t.  

**Pathway model**

- Mass balance $A \cdot f = b$
- Biomass supply $f_{bi} = 1$
- No residual $H_2$, $b_{H_2} = 0$

**Definitions**

- Mole fractions $z_i = b_i / \left( \sum_{i \in PC} b_i \right)$ for all $i \in PC$
- Molar mass $M = \sum_{j \in PC} z_j M_j$
- Mass fractions $\xi_i = z_j M_j / M$ for all $i \in PC$
- Mass density $\rho_L = 1 / \left( \sum_{j \in PC} \xi_j / \rho_{Lj} \right)$
- Molar density $\rho_m = \rho_L / M$
- Derived cetane number $DCN^{ab} \leq \sum_{j \in PC} z_j DCN_j \leq DCN^{ab}$

4105

DOI: 10.1021/acs.energyfuels.7b00118
Energy Fuels 2017, 31, 4096–4121
lower heating value \( LHV^b \leq \sum_{j \in \text{PC}} \xi_j LHV_j \leq LHV^{ab} \) (80)

enthalpy of vaporization \( H_{\text{vap}}^b \leq \sum_{j \in \text{PC}} \xi_j H_{\text{vap},j} \leq H_{\text{vap}}^{ab} \) (81)

oxygen content \( \text{wtO}_2^b \leq \sum_{j \in \text{PC}} \xi_j \text{wtO}_2 \leq \text{wtO}_2^{ab} \) (82)

liquid density \( \rho_L^b \leq \rho_L \leq \rho_L^{ab} \) (83)

kinematic viscosity \( \nu^b \leq \frac{1}{\rho_L^b} \sum_{j \in \text{PC}} z_j \mu_j \leq \nu^{ab} \) (84)

surface tension \( \sigma^b \) (85)

approx Reid vapor pressure \( p_{\text{RVP}}^b \)
\[
\leq \sum_{j \in \text{PC}} z_j \gamma_j(37.8 \degree C, \text{z}) \cdot \rho_j^b(37.8 \degree C) \\
\leq p_{\text{RVP}}^{ab} (86)
\]

distillation (10 mol % evap) \( T10(m)^b \)
\[
\leq \text{Dist-Model}_{T10(m)}(z) \\
\leq T10(m)^{ab} (87)
\]

distillation (50 mol % evap) \( T50(m)^b \)
\[
\leq \text{Dist-Model}_{T50(m)}(z) \\
\leq T50(m)^{ab} (88)
\]

distillation (90 mol% evap) \( T90(m)^b \)
\[
\leq \text{Dist-Model}_{T90(m)}(z) \\
\leq T90(m)^{ab} (89)
\]

stability constraint

local, necessary condition \( \lambda_i(\mathbf{H}(\Delta g(298 \text{K}, z))) \geq 0 \)
\[
i = 1, ..., n_C - 1 (90)
\]

The optimization variables are the flows \( \mathbf{f} \) and \( \mathbf{b} \). Thus, the task of the optimizer is to find a feasible combination of \( \mathbf{f} \) and \( \mathbf{b} \) that maximizes the energy of fuel produced (cf. eq 69). The property constraints are formulated on the basis of the \( n_C \) palette compounds. This means that, in principle, the solution to the optimization problem could be a blend of \( n_C \) components. In practical cases, however, the solution is typically composed of few species only, as the mole fractions of the other palette compounds become zero. Likewise, most entries of \( \mathbf{f} \) and \( \mathbf{b} \) are zero.

Constraints based on the pathway model (cf. eqs 71–73) are followed by a set of definitions (cf. eqs 74–78) which are given here for the sake of clarity, although, in principle, these could also be integrated into the subsequently listed property and stability constraints (cf. eqs 79–90). Note that the pure component properties, e.g., the derived cetane number, can be calculated a priori, whereas those properties depending on the composition, namely, activity coefficient, Reid pressure, distillation profile, and Gibbs energy of mixing, have to be recomputed in each step taken by the NLP solver. The efforts associated with the evaluation of those properties scale with the number of palette compounds. This means that also the total computational time for solving the blend design problem (cf. eqs 69–90) primarily depends on the number of palette compounds.

Two measures have been taken to speed up the calculations. First, inside the optimization problem, the Reid pressure (RVP) is approximated by the bubble point pressure \( p_{\text{bp}} \) (cf. eq 86). The correct RVP \( p_{\text{RVP}} \) is always computed from the system of nonlinear eqs 62 and 63, once the solution to the optimization problem has been obtained. From our experience with the case study below, the approximated RVP \( (p_{\text{bp}}) \) is typically between zero and 1 kPa higher than the correct RVP \( (p_{\text{RVP}}) \). Second, within the NLP the distillation curve is considered on a molar basis (instead of volume). This way, the idealized batch distillation model (cf. eqs 41–52) can be stripped down to:

\[
\frac{dN}{dt} = -N = -1 (91)
\]

\[
\frac{dx_{Li}}{dt} = \frac{1}{n_i} \left( x_{Li} - x_{Li} \gamma_i(T_{Li}, \mathbf{x}_i) \rho_i^S(T_{Li}) \right) \\
\quad \forall i = 1, ..., n_C - 1 (92)
\]

\[
p = \sum_i x_{Li} \gamma_i(T_{Li}, \mathbf{x}_i) \rho_i^S(T_{Li}) (93)
\]

where the initial values are set to \( x_{Li} = z_i \) (\( z_i \) is the blend composition defined in eq 74) and \( n_i = 1 \). The characteristic T10 temperature (cf. eq 87), i.e., the temperature where 10% of the fuel have been evaporated, is then simply \( T_{10}(t = 0.1) \). The solution of this smaller DAE system (cf. eqs 91–94) is approximated with the explicit Euler method based on a fixed grid. Compared with the error-controlled, variable step length solvers \textit{ode45} and \textit{ode15s} built into Matlab, the computational overhead as well as the number of activity coefficient calculations are significantly reduced, thereby greatly improving performance. The approximation error correlates with the chosen step length, i.e., the size of the mole increment (we have used 1 mol %), and is assessed by once running the error-controlled solver a posteriori. Although the distillation constraints are defined on molar basis in order to reduce the computational effort, the blend’s volumetric distillation curve is computed with eqs 41–52 a posteriori, because the volume basis is the more frequently used variant. Therefore, we will distinguish between \( T10(m)/T50(m)/T90(m) \) and \( T10(v)/T50(v)/T90(v) \) in the following, where \( m \) refers to mole and \( v \) refers to volume.

The local, necessary criterion for phase stability is integrated into the NLP formulation in an attempt to circumvent parts of miscibility gaps (cf. eq 90). The Hessian \( \mathbf{H} \) is approximated with finite differences and the eigenvalues are computed with Matlab’s \textit{eig} routine which is based on the QR algorithm. In order to avoid both the locally unstable and the metastable regions, a bilevel optimization problem would have to be formulated and solved. This is not done here for the sake of
simplicity and performance. Instead, each solution to eqs 69–90 is checked a posteriori by minimization of the tangent plane distance function, i.e., by solving a second but smaller NLP (cf. eqs 67 and 68). This way blends which turn out to be immiscible are discarded from the set of solutions.

Note that the optimization problem defined by eqs 69–90 represents a general formulation where each property is constrained by a lower and an upper bound. However, depending on the modeler’s intentions, the formulation can easily be modified for the design task at hand. Such modification might include removal of certain bounds. For instance, in the case study below, no upper bound is considered for the oxygen content and the lower heating value is not constrained at all. Instead of maximization of a process-related criteria like mass or energy of fuel produced, specific blend properties can be optimized as well. For instance, smaller values for the DCN correlate with higher efficiency of a spark-ignition engine. Thus, the objective function

\[ \min_{f \geq 0, b \geq 0} \sum_{j \in \text{PC}} z_j \text{DCN}_j \]  

(95)

may be chosen. Note that the inequality constraint

\[ \sum_{j \in \text{PC}} b_j M_j \geq n_{\text{fuel, min}} \]  

(96)

can be added simultaneously to ensure that a certain minimum mass flow of fuel \( n_{\text{fuel, min}} \) is still produced. Alternatively, if no more than \( n_{\text{H}_2, \text{max}} \) \( \text{mol}_{\text{H}_2}/\text{mol}_{\text{fuel}} \) of external hydrogen shall be used to upgrade the carbohydrates, the inequality constraint

\[ \frac{f_{S_2}}{\left( \sum_{j \in \text{PC}} b_j \right)} \leq n_{\text{H}_2, \text{max}} \]  

(97)

may be used.

Although the complete problem described by eqs 69–90 could be approached directly, it has been observed that if initial values for the full problem are derived systematically by first solving two simpler subproblems, the overall time required to find the solution can often be reduced. This sequential solution strategy is depicted in Figure 8. The first subproblem is derived from ignoring both the distillation profile constraints (cf. eqs 87–89) and the phase stability constraint (cf. eq 90) (cf. Figure 8, step 1). This nonlinear problem is small enough to be efficiently handled by GlobalSearch and the global solver is employed to decrease the risk of getting stuck at a local optimum that is not the global optimum. Once solution vectors \( f \) and \( b \) have been obtained, these are taken as initial guesses for the next subproblem that includes the phase stability constraint, but still lacks the distillation model (cf. Figure 8, step 2). Since the computation of the eigenvalues of the Hessian of \( \Delta g \) makes this subproblem significantly more computationally demanding, a local NLP solver is used here. If the initial values do not violate the phase stability constraint, completion of step 2 is extremely fast as the solution to subproblem 2 then is identical to the initial guesses. Finally, the results from step 2 are used to initialize the full problem (cf. Figure 8, step 3). Solving the full problem is by far the most computationally demanding step because of the distillation model, however, only if the supplied initial values violate the distillation profile constraints. With the proposed sequential strategy (cf. Figure 8), the total computational effort required to solve the blend design task therefore strongly depends on whether the path to the solution is somehow a fortuitous one.

If the CPM contains numerous palette compounds, blends of high complexity may arise. Here high complexity refers to a large number of components to be included in the final fuel. Trace amounts of a component \( i \) can easily be removed by adding a constraint \( b_i = 0 \) to the general problem formulation (cf. eqs 69–90). However, typically there is a multitude of blends with similar objective values and these blends can differ substantially with respect to composition and utilization of production pathways. Being able to select from a range of promising blends rather than having to rely on a single best solution may prove particularly valuable in this context. To obtain blends with a specified number \( (n_C) \) of components, subproblems can
be extracted from the CPM and solved in a systematic fashion. The enumeration-based strategy proposed in Figure 9 can be readily applied to generate binary, ternary and quaternary mixtures, if the total number of palette compounds to choose from \( n_{C_{\text{tot}}} \) is in the range of 20−30. Although in principal the strategy can also be applied to multicomponent mixtures with \( n_{C_{\text{tot}}} > 4 \), it quickly approaches its limits because of the combinatorial explosion encountered in step A (cf. Figure 9). The reason is that there are generally \( \frac{n_{C_{\text{tot}}}!}{n_{C}!(n_{C_{\text{tot}}}-n_{C})!} \) ways of choosing \( n_{C} \) palette compounds from a CPM that comprises a total number of \( n_{C_{\text{tot}}} \) palette compounds. Different subproblem definition strategies have to be developed to cope with such situations.

As can be seen from Figure 9, the workflow proposed here starts with the enumeration of all possible combinations of palette compounds (step A). Subsequently, the CPM is analyzed to discard those combinations that do not utilize both hexoses and pentoses (step B.1), because conversion of a single sugar fraction strongly limits the amount of fuel obtained. In step B.2, the feasibility of each combination is judged on the basis of pure component properties. For instance, if a binary blend shall have a mixture DCN smaller than 15, at least one of the two components needs to exhibit a DCN < 15. In step B.3, a first optimization problem is formulated based on the mixture property models. The objective of this problem is to find a feasible point, i.e., a specific composition that satisfies all imposed property constraints. Here it is still ignored whether this composition can actually be achieved in a process. In order to reduce the computational burden for the final design step, only those species and pathways are considered which eventually lead toward the palette compounds under investigation. An example for such CPM subpath extraction (step B.4) is depicted in Figure 10. Finally, the actual blend design and the rigorous miscibility check (steps C and D) are carried out by means of the sequential approach described before (cf. Figure 8).

**CASE STUDY: FUELS FOR THE SPARK-IGNITION ENGINE**

**Scenario Description.** A case study is dedicated to the identification of biofuels with tailored properties and their respective optimal production pathways. The target combustion system is a boosted direct injection spark-ignition engine, like those studied within TMFB previously.\(^{11,37-39}\) The objective is to maximize the energy of fuel produced (in terms of lower heating value) given a fixed feed of biomass (cf. eq 69). This can be achieved particularly well by considering production routes that (a) have high yields and (b) are capable of integrating major amounts of additional hydrogen. Since the amount of biomass that can be grown for the purpose of fuel production is limited, biomass hydrogenation by means of water electrolysis utilizing carbon-free energy sources such as wind or solar is seen as an opportunity to leverage the amount and energy of liquid hydrocarbons that can be produced from the biomass feedstock.\(^{119-122}\) Therefore, in this case study, we have decided not...
to constrain the supply of hydrogen, but instead to consider hydrogenation as an enabler for directing as many carbon as possible from the carbohydrates toward the fuel, thereby maximizing the energy of fuel produced.

As can be seen from Figure 11, twelve biobased intermediates (shaded in gray) which can be produced from hexoses and pentoses in large volumes, e.g., ethanol, acetic acid and furfural, are considered. These platforms have been chosen because they...
Energy & Fuels

![Image](https://example.com/image.png)

Article

by research octane number (RON)\textsuperscript{127} and motor octane number (MON).\textsuperscript{128} However, the applicability range of currently available RON/MON models is largely limited to nonoxygenated hydrocarbon species.\textsuperscript{129-137} Therefore, we make use of the observation that DCN and RON/MON are approximately inversely correlated.\textsuperscript{10,45,138} For instance, iso-octane as a reference compound for RON determination has the high RON of 100, whereas its experimental DCN is as low as 17.\textsuperscript{139} Similar to our previous publication,\textsuperscript{10} we have chosen DCN ≤ 20 to identify fuels with a knock-resistance that is sufficiently high to justify further investigation as a candidate fuel for the spark-ignition engine. This further investigation should include experimental RON/MON determination by means of ASTM D2699/ D2700.\textsuperscript{127,128}

Reid vapor pressure (RVP), T10(m), and T50(m) are constrained to mitigate problems with vapor lock, cold-starting and during the warm-up driving phase.\textsuperscript{34} Moreover, T90(m) is constrained to avoid lube oil dilution and excessive deposit formation.\textsuperscript{34} The upper and lower bounds for RVP, T10(m), T50(m), and T90(m) were chosen such that all distillation and vapor pressure classes of US gasoline fuel as defined in ASTM D4814-16\textsuperscript{59} are covered. The minimum Reid vapor pressure, which is important for engine cold-starting,\textsuperscript{34} has been taken from the European gasoline fuel standard EN 228,\textsuperscript{140} since the US standard does not explicitly include such a constraint. The enthalpy of vaporization is an important parameter for in-cylinder mixture formation in engine cold-start/run.\textsuperscript{11,37,39,41} It is related to one kg of air at stoichiometric conditions, because the atomic composition of a fuel dictates the mass of fuel to be injected.\textsuperscript{10,123} With the aim of enhancing spray breakup and atomization, limits for surface tension and viscosity are defined.\textsuperscript{39,141,142} Finally, a high fuel oxygen content (≥10%) is envisaged because fuel oxygen can effectively reduce the emissions of particulate matter by improving the local oxygen to carbon ratio in the air/fuel-mixture.\textsuperscript{143-145}

For this case study, pure-component property data for LHV, $\rho$, $\sigma$, and $\mu$ have been retrieved from the DIPPR 801 database.\textsuperscript{52} Only data that has been classified as “accepted” by DIPPR staff and that is thought to have an error of ≤10% have been queried. This covers approximately 74% of the required data for LHV, $\rho$, $\sigma$, and $\mu$. The remaining fraction has been predicted with the QSPR models listed in Table 3. The predictions can be found in the Supporting Information. In the case of vapor pressure and molar density, parameters for the extended Antoine equation (eq 37) and the DIPPR 105 equation (eq 54) have been retrieved from the Aspen Physical Property System pure component databank,\textsuperscript{146} whereas the other 13 DCNs had to be predicted with the fully predictive approach, i.e., critical data QSPR plus Hoffmann–Florin equation and/or COSTALD method, to the remaining instances (cf. Supporting Information). Pure-component enthalpy of vaporization $H_{\text{vap}}$ has been calculated from the vapor pressure curve via the Clausius–Clapeyron equation (cf. eq 34). UNIFAC parameters relevant to the species in the case study have been taken from different literature sources.\textsuperscript{43,99,100} Experimental DCNs for 11 compounds were found in literature,\textsuperscript{43,80,147} whereas the other 13 DCNs had to be predicted with the help of the group contribution model\textsuperscript{15} (cf. Supporting Information). Pragmatically, a value of 7 has been assigned to those compounds which do not ignite in the IQT as a consequence of the measurement limit, i.e., DCN ≈ 7.2.\textsuperscript{81,147}

Results. Maximization of the energy of fuel produced yields a five-component blend (referred to as blend A in the following) which is rich in 1-butanol (44 mol %), cyclopentane (31 mol %), and ethyl acetate (15 mol %). This design features a high degree of hydrogenation ($4.6 \text{ molH}_2/\text{molfuel}$) and yields 0.75 MJ$_{\text{fuel}}$/MJ$_{\text{biomass}}$. The corresponding material flow diagram is depicted in Figure 12. The C$_3$ sugars are converted into cyclopentane via pathways P4, P38, and P31, whereas the majority of the C$_5$ sugars is directed to succinic acid fermentation. The remaining C$_6$ sugars are fed to

<table>
<thead>
<tr>
<th>Table 4. Specifications for a Tailored Biofuel for a Boosted Direct-Injection Spark-Ignition Engine$^a$</th>
<th>fuel property</th>
<th>lower bound</th>
<th>upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>derived cetane number DCN [−]</td>
<td>20</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>oxygen content wtO$_2$ [wt %]</td>
<td>10</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>enthalpy of vaporization $H_{\text{vap}}$ [kJ/kg$_{\text{air}}$]</td>
<td>60</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>surface tension $\sigma$ [mN/m]</td>
<td>30</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>kinematic viscosity $\nu$ [mm$^2$/s]</td>
<td>0.5</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>approximate Reid vapor pressure $p_R$ [kPa]</td>
<td>45</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>distillation profile$^a$</td>
<td>100</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>temperature at 10% mole fraction evp T10(m) [°C]</td>
<td>45</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>temperature at 50% mole fraction evp T50(m) [°C]</td>
<td>70</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>temperature at 90% mole fraction evp T90(m) [°C]</td>
<td>125</td>
<td>125</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ See the work of Dahmen and Marquardt\textsuperscript{10} for a detailed discussion on the interactions between physico-chemical fuel properties and SI engine performance. $^a$Idealized open batch distillation (cf. Figure 6). By research results, 2-Methylfuran and 2,5-dimethylfuran can be made via chemocatalytic conversion of C$_5$/C$_6$ sugars and have been increasingly envisaged because the furan group interaction parameters are missing in the publicly available UNIFAC literature.\textsuperscript{100,126} At this time, the furan-derivative $\gamma$-valerolactone has been included in the case study because of its high knock-resistance,\textsuperscript{5,41} although its use as a fuel component is viewed with some criticism due to $\gamma$-valerolactone’s potential for abuse as a drug.\textsuperscript{41} With a normal boiling point of 207–208 °C,\textsuperscript{5} $\gamma$-valerolactone is also the least volatile compound in the case study.

The desired fuel properties are summarized in Table 4. The antiknock quality of practical fuels is measured most commonly...
ethanol fermentation. Ethanol, in this case, does not constitute a fuel component, but acts as a substrate for the production of acetic acid. The intermediates acetic acid and ethanol are finally converted into the fuel component ethyl acetate.

As can be seen from Figure 12, the carbon dioxide released in ethanol fermentation is fully reused in the succinic acid fermentation step. While CO₂ has zero enthalpy of combustion, the reuse of carbon allows for higher degrees of hydrogenation in subsequent steps. Thus, coprocessing of ethanol and succinic acid can be seen as an elegant way for chemical hydrogen storage in the case of SI fuels where a certain oxygen content is mandatory. It shall be noted that an analysis of the general trade-off between hydrogen demand [molH₂/molfuel] and energy of fuel produced [MJfuel/MJbiomass] can be found in the Supporting Information. The discussion there includes a blend that has been designed with the premise of zero hydrogen demand.

In case of blend A, succinic acid is converted into a mixture of ethanol and n-butane via pathways P23−P25 and P36 (cf. Figure 12). Cyclopentane and n-butane are responsible for the blend’s excellent front-end volatility, i.e., low T₁₀ and high Reid vapor pressure. Table 5 shows the estimated blend properties and Figure 13 depicts the distillation curve. Based on the pure component DCN values, the alcohols and the ethyl acetate are considered the most knock-resistant species in the mixture. Recently, Perez and Boehman have proposed a quantitative correlation between RON/MON and the ignition delay in the IQT, i.e., the standardized constant-volume combustion chamber that is used to measure DCN. If DCN is converted into raw IQT ignition delay (cf. the definitions given in ASTM D6890), the equations reported by Perez and Boehman can be applied to calculate a quantitative estimate for RON/MON. We will call these correlated RON (cRON) and correlated MON (cMON) in the following to distinguish them from true RON/MON. Since the accuracy of the IQT—RON/MON correlation has not been systematically assessed for a wider range of fuels, cRON and cMON should be generally taken with caution. For instance, the cRON for iso-octane is only 96.7, whereas its true RON is 100.

The cRON calculated for blend A is 92 (cf. Table 5). While RON91 constitutes the minimum fuel quality required for most modern passenger cars equipped with knock sensors, future fuels will probably require higher RONs because of the strive for higher engine efficiencies. In this context, the model-based design approach can be used to analyze the trade-off between DCN as a simplified measure for knock-resistance and the energy of fuel produced. To this end, a sequence of design problems is solved with decreasing upper bounds for the DCN. The resulting trade-off is visualized in Figure 14.

It can be seen that the energy of fuel produced in MJfuel per MJbiomass decreases as DCNs get smaller. However, for DCNs above 12, the correlation is generally weak. While there is no clear trend for mass and LHV energy efficiencies, hydrogen consumption follows the trend seen for energy of fuel produced. The reduced opportunity for hydrogenation in case of low DCN is accompanied by significant CO₂ emissions, because the emitted carbon can no longer be directed toward the fuel without violating the property constraints.

Figure 15 shows the material flow diagram for blend B, i.e., the blend with a DCN of 9 (cf. Figure 14) which translates into a cRON of 114 and a cMON of 104. The presumed high knock-resistance of this seven-components blend primarily results from the high fractions of ethanol, cyclopentanone, and 2-butanone.

Table 5. Predicted Properties of the Designs Discussed in This Case Study

<table>
<thead>
<tr>
<th>Blend</th>
<th>DCN</th>
<th>RON</th>
<th>MON</th>
<th>LHV</th>
<th>LHV energy efficiency [%]</th>
<th>Energy of fuel produced [MJfuel/molfuel]</th>
<th>Mole composition [mol %]</th>
<th>Blend composition [mol %]</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>114</td>
<td>104</td>
<td>92</td>
<td>35.5</td>
<td>77.9</td>
<td>0.67</td>
<td>1-butanol (44%), cyclopentane (31%), ethyl acetate (15%), 2,5-dimethyl-tetrahydrofuran (15%)</td>
<td>1-butanol (44%), cyclopentane (31%), ethyl acetate (15%), 2,5-dimethyl-tetrahydrofuran (15%)</td>
</tr>
<tr>
<td>B</td>
<td>126</td>
<td>104</td>
<td>92</td>
<td>81.8</td>
<td>81.8</td>
<td>0.47</td>
<td>1-butanol (44%), cyclopentanone (26%), 2,5-dimethyl-tetrahydrofuran (15%)</td>
<td>1-butanol (44%), cyclopentanone (26%), 2,5-dimethyl-tetrahydrofuran (15%)</td>
</tr>
<tr>
<td>C</td>
<td>126</td>
<td>104</td>
<td>92</td>
<td>78.1</td>
<td>78.1</td>
<td>0.54</td>
<td>1-butanol (44%), cyclopentane (26%), 2,5-dimethyl-tetrahydrofuran (15%)</td>
<td>1-butanol (44%), cyclopentane (26%), 2,5-dimethyl-tetrahydrofuran (15%)</td>
</tr>
<tr>
<td>D</td>
<td>126</td>
<td>104</td>
<td>92</td>
<td>78.1</td>
<td>78.1</td>
<td>0.38</td>
<td>1-butanol (44%), cyclopentane (26%), 2,5-dimethyl-tetrahydrofuran (15%)</td>
<td>1-butanol (44%), cyclopentane (26%), 2,5-dimethyl-tetrahydrofuran (15%)</td>
</tr>
</tbody>
</table>

*Computed with Matlab ode15s, error-controlled, variable step-length solver ode15s.*
The latter compound has an experimental RON/MON of 117/107. The production of 2-butanone, however, necessitates the inclusion of a separate 2,3-butanediol fermentation step in the design (cf. Figure 15). In contrast to the previous design, carbon dioxide released in ethanol fermentation is mostly emitted here (compare Figures 12 and 15). The low hydrogen consumption in the production of blend B (1.7 molH₂/mol fuel) is accompanied by a lower energy of fuel produced (0.61 MJ fuel/MJ biomass). However, the LHV energy efficiency is only slightly lower compared to the previous design. The estimated blend properties can be found in Table 5.

From the trade-off analysis in Figure 14 it appears that the blend with a DCN of 12 would represent a nice compromise between knock-resistance on the one hand and energy of fuel produced and CO₂ emissions on the other hand. However, it should be kept in mind that the inverse relationship between DCN and RON/MON holds only approximately. Moreover, linear mixing is assumed here and pure component DCNs of species that do not ignite in the IQT (e.g., ethanol, cyclopentanone) are assumed to be 7. In principle, the uncertainties of the outlined design approach could be reduced significantly if experimental RON and MON values were available for the pure compounds. It is also possible to derive nonlinear RON/MON mixing rules, once blending effects can be quantified based on comprehensive experimental RON/MON mixture data for the compounds of interest.

The complexities of blends A and B are quite high, meaning that both blend components and conversion pathways are numerous. The number of conversion pathways in a design can be taken as a first rough estimate of process complexity. Moreover, it is often assumed that the number of processing steps correlates with the capital cost of a production plant. With the aim of arriving at less complex designs, all optimal binary and ternary blends, which can be formulated based on the conversion pathway map depicted in Figure 11, have been generated. To this end, the strategy proposed in Figure 9 has been employed. The DCN

Figure 13. Predicted distillation curves for the blends discussed in this case study. The curves have been computed with Matlab’s ode15s error-controlled, variable step-length solver. The bars show the constraints for T10(m), T50(m), and T90(m) (cf. Table 4). Note that Vshift has been set to zero in an attempt to simulate an idealized batch distillation process.
constraint has now been tightened to DCN $\leq 15$, which corresponds to a cRON/cMON of $\geq 102/93$, in order to place particular focus on fuels especially well-suited for highly boosted, direct-injection SI engines. Figure 16 shows the number of active conversion steps and the energy of fuel produced for all optimal binary and ternary blends. The ideal design would be located in the lower right part of the graph. It can be noted that there is a ternary blend (called blend C in the following) which is slightly inferior to blend A in terms of energy of fuel produced. However, this blend is less complex. Indeed, the material flow diagram for blend C can be seen as a simplified version of design A (compare Figures 12 and 17). Instead of acetic acid fermentation and ethyl acetate production, ethanol is directly used as a fuel component here. Furthermore, the conversion of 1-butanol to n-butane has been eliminated which somewhat deteriorates the front-end volatility, as can be seen from Figure 13. Still, all fuel blending of biomass to fuels remains a complex design problem.
properties are well within the specified bounds (compare Tables 4 and 5). With a cRON of 103, blend C is expected to offer greater knock-resistance than blend A. At the same time, blend C retains a key property of blend A, i.e., the almost complete avoidance of carbon dioxide emissions that is enabled by concurrent ethanol and succinic acid fermentation (cf. Figure 17).

If the process complexity shall be reduced even further, Figure 16 shows that blend D represents a nice trade-off between the energy of fuel produced and the number of conversion steps. As can be seen from the material flow diagram in Figure 18, the majority of C6 sugars is still directed toward ethanol fermentation in this design. However, a significant fraction is also converted chemo-catalytically into 2,5-dimethyltetrahydrofuran, a compound with a high heating value. Knock-resistance is provided primarily by ethanol, and cyclopentane is required to satisfy the front-end volatility requirements. Two additional ternary blends are located in the vicinity of blend D in Figure 16 and their main components are ethanol and cyclopentane as well. This suggests that the third component of blend D, i.e., 2,5-dimethyltetrahydrofuran, can well be exchanged for tetrahydrofuran or ethyl acetate, if the proportions of ethanol and cyclopentane are adjusted accordingly (cf. Figure 16).

In summary, the case study has demonstrated the capabilities of the model-based approach for integrated product and pathway design of novel biofuel blends. The succinic acid pathway constitutes an elegant option for reusing CO2 that is produced elsewhere in the plant, e.g., in concurrent ethanol fermentation. This way more carbon reaches the fuel and, together with hydrogen produced from renewable electricity, this allows to increase the fuel energy available for combustion given a fixed input of biomass. The case study has also revealed the promising role of cyclopentanone production from furfural in the context of...
SI fuel synthesis. Cyclopentanone is considered an ideal fuel for highly boosted direct injection SI engines due to its extremely low autoignition reactivity and hence extreme knock-resistance. Further hydrogenation of cyclopentanone yields cyclopentane, a volatile compound which is expected to mitigate cold-start and cold-run problems typically observed in oxygenated fuels. The analysis of selected ternary blends has shown that, as a general rule, it is beneficial to generate and analyze a range of designs with similar properties, as these designs can differ considerably with respect to fuel components and production pathways employed. A subsequent, manual inspection from the perspective of early conceptual process design can then be used to discriminate among the alternatives. At the latest when the number of designs shrinks to a few, experimental validation of predicted blend properties should be performed. Finally, Figure 16 has also revealed the importance of performing combined product and pathway design. For though all binary and ternary blends plotted in this graph are considered fuels with tailored properties, only few of these are attractive from a process point of view, which means that the respective designs feature both high energy of fuel produced and limited complexity.

![Figure 18. Material flow diagram for the optimal ternary blend of ethanol (58 mol %), cyclopentane (27 mol %), and 2,5-dimethyltetrahydrofuran (15 mol %) (blend D, cf. Figure 16).](image)

<table>
<thead>
<tr>
<th>energy of fuel produced</th>
<th>mass of fuel produced</th>
<th>hydrogen demand</th>
<th>LHV</th>
<th>energy efficiency</th>
<th>CO₂ from biomass</th>
<th>derived cetane number (DCN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MJₖfuel/MJₖbiomass</td>
<td>kgₖfuel/kgₖbiomass</td>
<td>mol H₂/mol fuel</td>
<td>MJₖ</td>
<td>MJₖbiomass/MJₖ</td>
<td>g/MJₖfuel</td>
<td></td>
</tr>
<tr>
<td>0.65</td>
<td>0.36</td>
<td>2.2</td>
<td>0.56</td>
<td>12.16</td>
<td>15.0</td>
<td></td>
</tr>
</tbody>
</table>

DISCUSSIONS

In this paper we have presented an optimization-based approach for rational formulation of biofuel blends with tailored properties for combustion in future engine concepts. The blends are designed such that a process-related performance measure, i.e., the energy of fuel produced, is optimized. This has been achieved by stating and solving a nonlinear program that couples reaction flux analysis and mixture property prediction based on QSPR and group contribution modeling. Because feedstock cost is thought to dominate total biofuel production cost, selectivity data taken from the literature were assigned to the conversion pathways considered in our study. It was demonstrated that a multitude of product/pathway designs can be generated automatically, if subproblems are defined and solved in a systematic fashion. Such a strategy then yields a set of alternatives for further investigation by means of process network flux analysis (PNFA) and conceptual process design (CPD). PFNA and CPD are ultimately needed to bridge the gap from the mass- and energy-based molecular level analysis considered here to a process level analysis that deals with the economics of fuel production by taking into account aspects like separation of byproducts and solvents and energy recovery from waste streams.

A weakness of the computational approach is the fact that reliable liquid–liquid equilibria prediction is generally considered to be very difficult. Luckily, experimental validation of miscibility is quite straightforward. Removal of the phase stability criteria is conceivable, if experimental knowledge can rule out the presence of a miscibility gap. As a general rule, measured pure component property data should be used wherever possible to improve the accuracy of mixture property prediction. However, experimental data is often incomplete in the early screening phase. This underscores the value of the fully predictive approach taken here.
The blend design methodology is based on data and models which both are subject to uncertainty. For instance, pathway yield data often stem from lab-scale experiments and thus may change as a consequence of scale-up and transition from batch to continuous production. In case of RNFA, a comprehensive sensitivity study based on Monte Carlo analysis has revealed that differentiation of fuel candidates can become difficult or even impossible, if simultaneously occurring deviations of ±15% in all model parameters are assumed. For a practical blend design problem, logical next steps could be: The blend’s physicochemical properties should be measured in order to confirm the fuel’s quality. If measured and predicted fuel properties are too far apart, accuracy of pure component property data should be improved (by acquiring measurement data instead of using predictions). Relaxation or tightening of property constraints may be tried, if the solution lies on a constraint and the divergence of measured and predicted data is likely due to model inaccuracy, e.g., underestimation of DCN due to nonlinear mixing behavior. It is also conceivable to improve property models based on the characteristics of the specific mixture under investigation. One-at-a-time sensitivity analysis with regard to the selectivities of all pathways involved in the blend’s production can be performed to evaluate how the optimum of the objective function and the blend’s composition will be affected. As depicted in Figure 19, an iterative strategy, where the model-based design approach is combined with experimental and fundamental investigations, should be pursued aiming at gradual uncertainty reduction, model improvement and pathway debottlenecking. This is even more important since the fuel definition itself, i.e., the properties to be modeled and their target ranges, is continuously evolving as the fuel/engine interaction becomes more understood.

A potential problem that has not been addressed in this contribution is chemical reactivity among the blend components. For instance, if it is known that two components A and B tend to react with each other, precautions must be taken to avoid that both of them appear in the same blend. This could be achieved by stating a mixed-integer nonlinear program which, however, is significantly more challenging to solve. If the issue of reactivity is limited to few components, the CPM might also be decomposed into subproblems manually. In any case, reactivity, including the issue of long-term fuel stability, is probably difficult to judge in advance, thus requiring specialized testing as a part of experimental fuel quality validation. Practical fuel oxidation stability issues are unlikely to be limited to furanic species, as we have reported previously on the issue of high peroxide concentrations in oxygenated fuel samples being prepared for IQT testing. Thus, the formulation of certain oxygenated blends will likely require the use of stabilizers like butylated hydroxytoluene.

Throughout this paper, lignin has been treated as a source for process power and heat. Therefore, only carbon from the C₅ and C₆ sugars has been directed toward the fuel. However, lignin could also be added as a pseudocomponent to the CPM, as both lignin combustion generating CO₂ and lignin gasification yielding syngas bear the potential to improve the carbon efficiency and/or energy of fuel produced.

Although, for computational reasons, the number of palette compounds in our approach must be restricted to a reasonable number, network generator tools like RING might be used to generate alternative pathways and associated pathway performance data for integration into the formulation of the blend design problem. It shall also be mentioned that surrogate models for fossil fuels, i.e., gasoline and diesel fuel, can easily be integrated into the framework outlined above. For instance, the eight components diesel surrogate proposed by Mueller et al. has been rationally formulated to match the DCN, the liquid density, the distillation curve, and the distribution of key functional groups of a real diesel fuel. The property models used in the present contribution can be applied readily to mixtures of oxygenates and these eight hydrocarbon model compounds, thereby giving the opportunity to design biofuels for blending applications with fossil fuels.
The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.energyfuels.7b00118.

Case study: references for conversion pathways; statistics for QSPR and DCN group contribution models; application of HF equation; application of COSTALD method; trade-off analysis between hydrogen demand and energy of fuel produced (PDF)

Case study: stoichiometric matrix and primary property data (XLSX)

**ASSOCIATED CONTENT**

**Author Information**

**Corresponding Author**

*E-mail: w.marquardt@fz-juelich.de.*

**ORCID**

Wolfgang Marquardt: 0000-0002-4455-8378

**Present Address**


**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

This work was performed as part of the Cluster of Excellence "Tailor-Made Fuels from Biomass", which is funded by the Excellence Initiative by the German federal and state governments to promote science and research at German universities.

**NOMENCLATURE**

**Abbreviations**

ADC = Advanced Distillation Curve  
cMON = correlated motor octane number  
cRON = correlated research octane number  
CAMD = computer-aided molecular design  
CCE = Clausius–Clapeyron equation  
CPD = conceptual process design  
CPM = conversion pathway map  
DCN = derived cetane number  
IQT = Ignition Quality Tester  
LHV = lower heating value  
LLE = liquid–liquid equilibrium  
MON = motor octane number  
NLP = nonlinear program  
PNFA = process network flux analysis  
QSPR = quantitative structure–property relationship  
RNFA = reaction network flux analysis  
RON = research octane number  
RVP = Reid vapor pressure  
SI = spark-ignition  
TMFB = Tailor-Made Fuels from Biomass  
VLE = vapor–liquid equilibrium

**Greek Letters**

$\gamma$ = activity coefficient  
$\eta_{\text{LHV}}$ = LHV efficiency  
$\lambda$ = eigenvalue  
$\mu$ = dynamic viscosity  
$\nu$ = kinematic viscosity  
$\nu$ = stoichiometric coefficient  
$\xi$ = mass fraction blend  
$\rho_{l}$ = liquid density  
$\rho_{m}$ = molar density  
$\sigma$ = surface tension

**Latin letters**

$A$ = matrix of stoichiometric coefficients  
$b$ = product flow  
$B$ = vector of product flows  
$C$ = pathway conversion  
$C_{\text{ant}}$ = $C_{\text{ant}}$ = parameters for the extended Antoine equation  
$C_{\text{dip}}$ = $C_{\text{dip}}$ = parameters for the DIPPR 105 equation  
$\text{DCN} = \text{derived cetane number}$  
$E_{\text{fuel}}$ = energy flow of fuel produced  
$f$ = molar flow over pathway  
$f^*$ = vector of molar flows over pathways  
$\bar{f}_{\text{COSTALD}} = \text{COSTALD equation}$  
$\bar{f}_{\text{HF}} = \text{Hoffmann–Florin equation}$  
$\bar{f}_{\text{dip}}$ = tangent plane distance function  
$\Delta g$ = Gibb’s energy change of mixing  
$H$ = Hessian matrix  
$H_{\text{vap}}$ = enthalpy of vaporization  
$K$ = equilibrium ratio  
$LHV = \text{lower heating value}$  
$m_{\text{CO}}$ = mass of carbon dioxide per MJ of fuel  
$m_{\text{fuel}}$ = total mass flow of fuel produced  
$m_{\text{H}_{2}}$ = moles of hydrogen per mole of fuel  
$m_{\text{H}_{2,\text{max}}}$ = maximum hydrogen demand of fuel production (moles of hydrogen per mole of fuel)  
$n$ = number of nodes in a CPM  
$n_{\text{fuel}}$ = total molar flow of fuel produced  
$n_{\text{H}_{2}}$ = number of pathway conversion  
$n_{\text{H}_{2,\text{max}}}$ = maximum hydrogen demand of fuel production (moles of hydrogen per mole of fuel)  
$n_{\text{N}}$ = number of nodes in a CPM  
$n_{\text{pathways}}$ = number of pathways in a CPM  
$N_{\text{fl}}$ = mole flow of vapor  
$p$ = sea level atmospheric pressure  
$P_{\text{b}}$ = bubble point pressure (approximated Reid vapor pressure)  
$P_{\text{cr}}$ = critical pressure  
$P_{\text{v}}$ = vapor pressure  
$P_{\text{f}}$ = Reid vapor pressure  
$P_{\text{crit}}$ = critical pressure  
$P_{\text{v}}$ = enthalpy of vaporization  
$P_{\text{crit}}$ = critical pressure  
$P_{\text{v}}$ = normal boiling point  
$P_{\text{crit}}$ = critical temperature  
$T_{10}/T_{50}$ = temperature on the distillation curve where 10/50/90 mol % of the fuel have been evaporated  
$T_{20}/T_{50}$ = temperature on the distillation curve where 10/50/90 mol % of the fuel have been evaporated  
$T_{\text{v}}$ = specific volume  
$V$ = volume  
$V_{\text{cd}}$ = critical volume  
$V_{\text{i}}$ = volume fraction distilled  
$V_{\text{inc}}$ = normalized volume increment  
$w_{O_{2}}$ = oxygen content (weight fraction)  
$x = \text{mole fraction liquid phase}$  
$x = \text{vector of mole fractions liquid phase}$  
$y = \text{mole fraction vapor phase}$  
$\gamma_{y}$ = vector of mole fractions vapor phase  
$Y = \text{pathway yield}$

$\xi_{\text{c}}$ = mass fraction blend  
$\rho_{\text{L}}$ = liquid density  
$\rho_{\text{m}}$ = molar density  
$\sigma$ = surface tension


Selective and flexible transformation of biomass-derived platform chemicals by a multifunctional catalytic system. Angew. Chem., Int. Ed. 2010, 122, 5642−5646.


Model-based Identi...


(52) AICHe, DIPPR 801 Database version 6.0; Design Institute for Physical Properties (DIPPR), American Institute of Chemical Engineers (AICHe), USA, 2012; http://www.aiche.org/dippr.


DOI: 10.1021/acs.energyfuels.7b00118
Energy Fuels 2017, 31, 4096–4121


(92) DIN EN 590 Automotive fuels—Diesel—Requirements and test methods; German version, DIN Deutsches Institut fuer Normung e.V., 2014.


(97) Backhaus, J.; Rothamer, D. Design methodology for volatility characteristics of bio-derived drop-in gasoline fuels. Spring Technical Meeting of the Central States Section of the Combustion Institute, April 22–24, 2012.


(122) Connolly, D.; Mathiesen, B. V.; Ridjan, I. A comparison between renewable transport fuels that can supplement or replace biofuels in a 100% renewable energy system. Energy 2014, 73, 110–125.


(140) DIN EN 228 Automotive fuels - Unleaded petrol - Requirements and test methods; German version, DIN Deutsches Institut fuer Normung e.V., 2014.


