Reaction Network Flux Analysis: Optimization-Based Evaluation of Reaction Pathways for Biorenewables Processing

A. Voll and W. Marquardt
Aachener Verfahrenstechnik - Process Systems Engineering, RWTH Aachen University, Turmstraße 46, 52064 Aachen, Germany

DOI 10.1002/aic.12704

Even though biomass is attracting increasing interest as a raw material in the chemical and the fuel industries, only few biobased production processes are yet established. At the same time a lot of new catalytic routes are proposed, but their potential in biorefinery applications is hard to predict. Reaction network flux analysis (RNFA) is introduced as a novel, rapid screening method which bridges the gap between chemo- or biocatalysis and process design by (1) systematically identifying and (2) subsequently analyzing and ranking the large number of alternative reaction pathways based on limited data. This optimization-based method helps to detect promising production routes as well as bottlenecks in possible pathways. The potential and the application of the RNFA methodology will be demonstrated by means of a case study for the production of the potential biofuel 3-methyl-tetrahydrofuran (3-MTHF) from the platform chemical itaconic acid (IA).

Keywords: reaction network flux analysis, chemical process design, biorenewables processing, mixed-integer programming, bioeconomy science

Introduction and Motivation

Due to the limited availability of fossil resources and due to the increasing effort to reduce carbon dioxide emissions, the switch from fossil to renewable feedstock for the production of chemicals and fuels has recently attracted significant attention. A sustainable realization of the switch from fossil to renewable raw materials will require considerable modifications in production processes and in the established value chains. Instead of reducing biomass to C_1 building blocks (e.g., by gasification) and synthesizing complex molecules out of it, processes and applications should be designed to refunctionalize the molecular structures present in native biomass[1,2] in order to exploit the synthesis power of nature to the extent possible. Such a strategy inevitably necessitates the exploration of new synthetic pathways and, consequently, leads to a new range of oxygenated platform chemicals and products. Recently, many reactions are analyzed for transforming biorenewable raw materials such as sugars or fatty acids into (platform) chemicals and fuel compounds (e.g.,[3,4]). While some bio- and chemocatalytic routes are novel, others have been known for a long time and are presently rediscovered or improved with respect to yield and selectivity (e.g.,[5–7]). However, most often these reactions are proposed by bio- or chemocatalysis research, which typically focuses on a single or few related reaction steps in laboratory experiments, while the perspective of the full reaction pathway from raw materials to products in a sequence of transformation steps is largely neglected. In particular, neither the pretreatment of the feedstock and the isolation of the byproducts are considered, nor the performance of the overall process is estimated.

Systematic process synthesis is an established discipline in chemical engineering which is documented in a number of classical textbooks such as those of Douglas,[8] Smith,[9] Blass[10] or Seider et al.[11] It has been successfully applied to many petrochemical process design problems including the synthesis of reactor systems, separation sequences or heat exchanger networks and the optimization of the overall process for economical performance or energy efficiency. More recently, process synthesis methodologies have also been applied to design processes transforming biorenewable feedstock into molecular products (see, e.g.,[1,13–15]). However, systematic process synthesis has rarely been integrated with fundamental research on novel reactions and catalysts, although conceptual design should be considered as early as possible in catalytic process development to achieve a "green production".[16] In this context, conceptual design should not start after the reaction pathway is fixed by chemical or biotechnological research, but it should rather support research on reaction steps and sequences by a comprehensive comparison of alternative pathways as early as possible to avoid unfavorable production routes right from the start (see, e.g.,[9]). Such a comparison is supposed to be difficult as most established methodological approaches require data, which are only available for actually operating processes,[17,18] but which are largely lacking during fundamental research. A

Additional Supporting Information can be found in the online version of this article

Correspondence concerning this article should be addressed to W. Marquardt at wolfgang.marquardt@avt.rwth-aachen.de.

number of methodologies have been reported in recent years for the generation of possible reaction pathways (or networks) based on known and conjectured alternatives for each individual reaction step (see, e.g.,[19–21]). None of these, however, considers the evaluation of the reaction pathways from the perspective of process performance, which is the key objective of systematic process design.

The evaluation of candidate reaction pathways in biorenewables processing can be related to flux analysis of a reaction network, a technique which is well established in other research fields, in particular in metabolic engineering. Flux-balance models describing the metabolism of a specific organism are routinely employed to support the design of experiments in order to identify the structure of a metabolic model, to distinguish between different operating modes of the organism or to locate the potential for metabolic reengineering of a pathway.[22,23] Among others, the method of extreme pathways is one of the most promising concepts.[24] Here, metabolic networks are modeled in terms of nodes (metabolites), and arcs (reactions) such that material flows can be traced from stationary flux balances for all the nodes in the network. As this modeling approach results in an (often largely) underdetermined system of linear equations, efficient linear programming (LP) strategies can be applied to find those flux distributions which maximize a biologically meaningful objective function.[25] Nonisolated optima—that is, optimal solutions with the same objective function value but different realizations of the decision variables, i.e., the flows—must be detected to fully understand the organism's behavior. Mixed-integer linear programming (MILP) algorithms have been shown to be well-suited for this purpose.[26] All possible flux distributions can be represented by linear combinations of these so-called extreme point solutions.

While the field of application and the overall goals are totally different, the general concepts of metabolic pathway analysis can be transferred to the analysis of reaction pathways from biorenewable raw materials to desired chemical or fuel products in order to get first insight into the nature of the process design problem. While the consideration of appropriate reaction pathways is part of standard procedures in process design (see, e.g.,[8] or[9]), a model-based procedure accounting for different performance criteria is still missing. Given the strong impact of the reaction pathway on the total cost of ownership of a process, it is especially important that all possible reaction pathways are detected and that their advantages and disadvantages can be assessed quantitatively in the very early process design stages. Material balances are obviously well-suited and should be used for a primary classification of the pathway alternatives. This way, not only the expected product yield, but also the number and quantity of all reactants and byproducts can be calculated to support the estimation of the raw material cost and the development of sufficient supply strategies. Then, preliminary decisions can be taken on the handling of high volume byproducts, whether they can be recycled, sold or used in an integrated biorefinery for further processing or for energy supply. In addition, the composition of the reaction product stream provides an indication about the separation effort required for the isolation and purification of byproducts. Furthermore, additional evaluation criteria can be connected to the mass balance. For example, enthalpies of formation can be utilized to calculate a possibly rough estimate of the energy needs of the future process, while toxicity and emission data help to predict its environmental impact. Since the mass balance enables the sizing and costing of the major equipment, a first economic evaluation can be performed, if cost data for raw material, products and catalysts are available. These concepts can also be combined to an integrated product and process design approach.[27] Overall, the rational process design can be significantly facilitated by systematically integrating this knowledge in the analysis of process alternatives.

In this article, we show how methods of metabolic pathway analysis can be transferred to the analysis of reaction pathways for future production networks based on biorenewable raw materials. It is explained, which analogies of metabolic and reactive pathways can be used and which adjustments for reactive systems must be made to derive the novel method of reaction network flux analysis (RNFA). After presenting the structure of the network model with its necessary constraints, adequate solution strategies are discussed, before criteria for the evaluation of reaction pathways are derived. Finally, the application of RNFA is illustrated by a case study for the production of 3-methyl-tetrahydrofuran (3-MTHF, $C_5H_{10}O$) from itaconic acid (IA, $C_5H_6O_4$), which has been identified as a platform chemical in future biobased value chains.[17] Altogether, the article presents RNFA as a rapid screening method for the systematic evaluation of reaction pathways as the first stage of process design.

Methodology

According to the principles of metabolic pathway analysis, RNFA is introduced as an optimization-based method to systematically identify and, subsequently, evaluate feasible and attractive reaction pathways in a given reaction network. A reaction pathway is defined as the sequence of all the required reaction steps connecting a (set of) starting and a (set of) target molecule(s). In order to detect these pathways, all possible reactions linking raw materials, intermediates and products are summarized in a network. Then, adequate optimization techniques are applied to enumerate all feasible reaction pathways according to a given objective such as, e.g., the maximization of product yield. Finally, the promising pathways are classified by various additional evaluation criteria pointing to reveal the characteristics of the alternatives and to elucidate the problems to be addressed in the next design phase. In the following paragraphs the different methodological aspects of RNFA are explained in detail.

Construction of reaction networks and data collection

First of all, a network of the considered reactive system must be setup as a basis for RNFA. As the analysis results are sensitive to structural changes in the network, all relevant substances and reactions should be included from the beginning. Therefore, it is important to properly define the scope of the analysis. A reaction network could, e.g., summarize all known reactions starting from a specific reactant or leading to a specific product. In addition, general objectives to be achieved by a reaction sequence can be formulated such as, e.g., an increase of the energy content along the reaction pathway. However, since modern multistage catalysis opens up a wide range of possibilities to realize a desired chemical synthesis step, it can hardly be guaranteed that all options are covered during data collection, especially if the reactive system is subject to ongoing catalysis research. The reaction network is not restricted to reactions which have been successfully demonstrated and reported in literature. Rather, speculative reaction steps can be conjectured on purpose by experienced synthesis chemists in order to analyze the consequences for a reaction pathway, if the proposed reaction steps were successfully shown in chemical research.

The generation of the reaction network should be started with an extensive literature research. Nowadays, web-based search engines and subject-specific databases facilitate the screening of the relevant journal and patent literature. Recently published research results must surely be integrated in the network, but review articles offer an important, ideally comprehensive source of information. It is also worth considering older publications because some interesting reactions might have been known for a long time, but lately lost attention due to a variety of reasons. Finally, the literature research is expanded by experts' knowledge such that the potential of innovative ideas and preliminary research results can be covered by RNFA early in the design lifecycle. Besides, the existing procedures for an automatic generation of reaction networks (see, e.g.,[19–21]) should be considered. In all these approaches, a set of generalized reaction types must be defined first. Each reaction type in the set is then instantiated by means of given reactants and their successive products to generate the network by some kind of rule-based combinatorial algorithm. The application of automatic procedures is particularly promising in case of complex raw materials. Especially, in biorenewables processing, the feedstock is of multicomponent nature, because complex mixtures often result from biomass pretreatment and subsequent depolymerization of the macromolecules. This feedstock may not have to be separated into its constituents, but could be converted directly into mixtures of target compounds, which might qualify as an end product. This scenario is very likely to occur in case of next generation biofuel production and very complex reaction networks inevitably result in such cases.

The structure of the reaction network including the stoichiometric equations is vital for RNFA. Beyond this essential input data, additional pieces of information about the network, the reactions and the substances can be included to improve the significance of the evaluation results. Different inputs are required depending on the expected level of detail of the analysis. For example, reaction yields and standard energies of formations are sufficient for preliminary mass and energy balancing, whereas the operating points of the units in the flow sheet must be known for a detailed analysis such as heat integration. Although the required thermodynamic property data are easily accessible in databases (e.g., DIPPR, Dechema or NIST) for common substances, experimental data are usually not available for industrially less relevant substances. In these cases, property prediction methods[28] have to be applied to derive an estimate of the property values of interest from molecular structure. According to the type of property and the desired accuracy, group contribution methods,[28–30] (targeted) quantitative structure property relations[31,32] or even molecular dynamic calculations[33,34] can be chosen. Nevertheless, if these estimates of physical property data are used, the uncertainty in the predictions should be considered in RNFA, for example, by means of a sensitivity analysis of the solution of the optimization problem[35] or of its reformulation as a robust (worst-case) optimization problem. Properties with significant impact on the evaluation results must be validated by experimental measurements for reliable conclusions.

Modeling of reaction networks

As soon as the network is setup it can be represented as a graph consisting of nodes and arcs. Here, nodes and arcs relate to the substances and the reactions of the network, respectively. Arcs are directed according to the direction of the reaction and connect more than two nodes, if a reaction has multiple reactants or products. Subsequently, the stationary material balances of the network can be formulated as

$$\boldsymbol{A} \cdot \boldsymbol{f} = 0. \tag{1}$$

Here $\boldsymbol{A} = (v_{i,j}) \in \mathbb{Q}^{\mathrm{sxr}}$ is the matrix of the stoichiometric coefficients $v_{i,j}$. The rows of $\boldsymbol{A}$ refer to the substances $I \in \{1,\ldots,s\}$ and its columns to the reactions $j \in \{1,\ldots,r\}$. Following the generally accepted convention, stoichiometric coefficients $v_{i,j}$ are positive, if a substance i is formed, and negative, if the substance i is consumed by a certain reaction j. The vector $\boldsymbol{f} \in \mathbb{R}^{\mathrm{r}}$ concatenates the molar fluxes of all reactions j, which are also referred to as normalized reaction rates or extents of reaction elsewhere.[36] Consequently, the (possibly normalized) molar flow rate of a substance i in a certain reaction j is obtained by multiplying the reactive flow f_j with the corresponding stoichiometric coefficient $v_{i,j}$. As the entries of the vector $\boldsymbol{f}$ are restricted to positive values, the matrix $\boldsymbol{A}$ must also include the raw material supply to the network and the removal of the end products in order to close the overall material balance in Eq. 1. To distinguish the different types of reactions, this equation can be structured as

$$[\boldsymbol{A}_1 \quad \boldsymbol{A}_2 \quad \boldsymbol{A}_3] \cdot \begin{bmatrix} \boldsymbol{f}_1 \\ \boldsymbol{f}_2 \\ \boldsymbol{f}_3 \end{bmatrix} = 0. \tag{2}$$

In the context of reaction networks, $\boldsymbol{A_2}$ summarizes the stoichiometric coefficients of all chemically founded reactions of the network, while $\boldsymbol{A_1}$ and $\boldsymbol{A_3}$ refer to pseudo-reactions to symbolize the supply of reactants and the removal of products, respectively. The reactions of $\boldsymbol{A_1}$ only have a single product and no reactants, so they can also be interpreted as sources of the network to provide the necessary amount of all required materials. In contrast, reactions of $\boldsymbol{A_3}$ can be seen as sinks because each of them has a stoichiometric coefficient of -1 for the product to be removed. Such removal reactions must be formulated for all substances which could possibly be byproducts in the network, i.e., for all substances which are formed by certain reactions but not (completely) consumed by successive reactions. This is the case for desired end products, but also for intermediates, which may accumulate in the network due to limited conversion. If the product range is not known exactly, removal reactions must be formulated for all substances in the network.

Therefore, an alternative, but completely equivalent formulation of the mass balance might be advantageous. In particular, Eq. 2 can be rewritten to result in

$$\boldsymbol{A}' \cdot \boldsymbol{f}' = b \quad \text{with} \quad \boldsymbol{A}' = [\boldsymbol{A}_1 \boldsymbol{A}_2], \ \boldsymbol{f}' = \begin{bmatrix} \boldsymbol{f}_1 \\ \boldsymbol{f}_2 \end{bmatrix}, \ \boldsymbol{b} = -\boldsymbol{A}_3 \cdot \boldsymbol{f}_3, \tag{3}$$

where all molar product flows are directly covered by the vector $\boldsymbol{b} \in \mathbb{R}^{\mathrm{s}}$. With this formulation, the addition of substances only leads to modifications in the rows of the

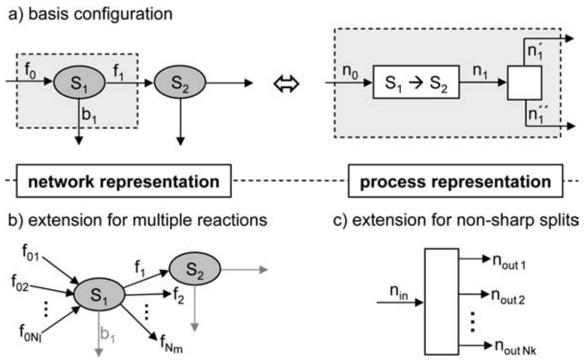

Figure 1. Illustration of the network and the corresponding process representation to derive yield constraints for (a) a basis configuration, (b) an extension to multiple supply and conversion reactions in the network representation and (c) an extension to a separation in the process splitting one feed into multiple outlet streams.

matrix $\boldsymbol{A}'$, but no additional removal reactions must be formulated.

Ideal reaction performance is described by the model (Eq. 2), but in reality reactions are limited in terms of conversion and selectivity or in terms of yield, which is defined as the product of the former.[36] Since these constraints affect the overall mass balance, they must be included in the network model. The following definitions are based on reaction yield since it subsumes the influence of both, selectivity and conversion. Roughly speaking, yield constraints can be formulated such that only a limited fraction of the incoming flows can be converted by each applicable reaction. However, the yield definition for the network model as a multireaction system is not trivial because the influencing factors, normally used to formulate yield constraints, are not directly available in the network representation, but are rather related to the processing conditions. Therefore, a relation between the reaction network representation and a suitable process representation will be derived to clarify the corresponding mathematical constraints for the reaction network model. To this end, important notation and concepts are introduced first by means of a simple example. Next, they are generalized to cover parallel and successive reactions with multiple reactants and products.

Figure 1a illustrates a simple cutout of a reaction network and the associated process flow sheet. The substance S_1 is converted by a reaction R_1 forming the substance S_2 as the only product. In both representations, the relevant steps are marked by the gray boxes, which have the same input and output streams. The variables f_j and b_i in the network representation have already been introduced as reactive flows, while the variables n_j in the process representation refer to the total molar flow. The variables of the two perspectives can be transferred into each other by replacing the molar flow rates n_j by the product of the reactive flow f_j, and the associated stoichiometric coefficient $\nu_{i,j}$. Consequently, the molar flow rate n_0 can be calculated by

$$n_0 = \nu_{1,0} f_0, \tag{4}$$

assuming that the substance S_1 is supplied by a reaction R_0, which is not explicitly shown in Figure 1. If reaction R_1 is not run under full yield, the reactor effluent n_1 is composed of the product S_2 and of unconverted reactant S_1 according to

$$n_1 = b_1 + \nu_{2,1} f_1. \tag{5}$$

In this case, a separation step is required in the process flow sheet to split the reactor effluent into streams of desired composition (i.e., pure components or mixtures of given specification). If an ideal separation is assumed, the molar flows n'_1 and n''_1 consist of the pure substances S_1 and S_2, respectively, and can be related to the network representations by

$$\begin{aligned} n_1' &= b_1, \\ n_1'' &= \nu_{2,1} f_1. \end{aligned} \tag{6}$$

An explicit separation step is not necessary in the network representation, since unconverted reactants are directly balanced in the product vector $\boldsymbol{b}$ such that reactions are limited to the conversion of and into pure substance streams.

Now, the yield of reaction R_1 can be defined according to the common convention as

$$Y_{R1} = \frac{n_1''}{n_0} = \frac{\nu_{2,1} f_1}{\nu_{1,0} f_0}, \tag{7}$$

by dividing the amount of desired product (n''_1) by the amount of the limiting reactant (n_0), which is considered to be a pure substance stream as well. As the yield coefficient Y_j is formulated on a molar basis, it can take values greater than 1, if several mole-equivalents of product are formed per mole of reactant. The yield coefficient can be normalized to values between 0 and 1 by division with the stoichiometric coefficients.[36]

The simple example must be generalized, because substance S_1 could be formed and consumed by more than one reaction as illustrated in Figure 1b. This generalization can be accommodated in terms of the yield definition by

$$Y_{R1} = \frac{\nu_{2.1} f_1}{\sum\limits_{l=1}^{N_1} \nu_{1,l} f_{0l} + \sum\limits_{m=2}^{N_m} \nu_{1,m} f_m}. \tag{8}$$

where all incoming reaction flows f_{0l} represent the reactant supply of the node S_1,while all exiting flows f_m refer to the consumption of this reactant. Since the exiting flows have negative stoichiometric coefficients by definition, the denominator determines the amount of substance that is available in reaction R_1. Note, that the counter m starts at 2 as the reactive flow of the considered reaction R_1 may not be included in the sum of exiting streams. The formulation in Eq. 8 captures both, successive and parallel reactions, since it can be applied to each step in a cascade of successive reactions, while parallel reactions are covered by the sum of exiting flows.

Finally, equations of type (Eq. 8) can be adopted to restrict the reactive flow of some reaction j by

$$f_j \leq Y_j \frac{\sum\limits_{l} \nu_{i,l} f_l + \sum\limits_{m,m \neq j} \nu_{i,m} f_m}{\nu_{i,j}} \quad \forall i \tag{9}$$

assuming that the yield coefficient Y_j is known. Eq. 9 is utilized as inequality to allow for the determination of a flexible flux distribution, which is not limited to the ideal case of maximal flow rates. These so-called yield constraints must be included in the optimization problem formulation for all exiting streams of each substance i excluding removal reactions, such that they are finally formulated for all reactions summarized in the matrix $\boldsymbol{A_2}$.

Yield coefficients are ideally available in literature. However, it must be kept in mind that these yields often stem from laboratory experiments and cannot directly be extrapolated to the process scale because additional limitations like heat- and mass-transfer effects have not been accounted for properly in the laboratory experiments. If no experimental or theoretical information can be found at all, a reasonable value can be chosen in the sense of a target yield. In this case, supplemental sensitivity analyses must be performed to determine the influence of such choice on the overall performance of the reaction network, in particular, on the yield of the target products. Alternatively, the yield coefficient(s) of one or more key reactions can be determined in the RNFA such that a desired product yield can be reached. Unfortunately, this leads to a nonlinear model with all the well-known computational drawbacks. In any case, the specified or computed yields of not yet established reactions define a target for catalysis research.

The yield constraints (Eq. 9) are derived under the assumption of a sharp split into pure components. However, the modeling framework can be expanded to cover non-sharp splits. To this end, the generalized concept of a separation step is depicted in Figure 1c, where the incoming molar flow n_{in} is split into N_k outlet streams $n_{\mathrm{out},k}$, each of which is potentially composed of a set of s components. This kind of separation can be modeled by

$$\begin{aligned} &n_{out\,k,i} = \mu_{k,i} n_{in\,i} \qquad k = 1,\ldots,N_k,\ i = 1,\ldots,s,\\ &\sum_{k=1}^{N_k} \mu_{k,i} = 1, \end{aligned} \tag{10}$$

introducing split factors $\mu_{k,i}$ for each outlet stream k and for each specific component i.

Of course, there are alternative ways of modeling yield constraints, which might be advantageous depending on the application. For example, the composition of the product mixture can be defined instead of fixing the yield coefficient. This turns out to be extremely beneficial, if a reaction cascade generates a mixture of defined components with a known composition or a known composition range, but with unknown reaction yields in the individual steps. Such alternative constraints can easily be implemented in the reaction network model to describe the performance of the reaction steps in a different way.

Formulation of the optimization problem and solution strategy

The variety of reaction pathways covered by the reaction network can be detected by solving the network model. In most cases the network contains more reactions than substances because a substance can normally be synthesized by more than one reaction. Hence, the corresponding system of mass balance equations is underdetermined and has an infinite number of solutions. In this context, a solution denotes the flux distribution associated with a certain reaction pathway containing all active reaction steps from raw material to target product. Instead of taking additional assumptions or fixing additional constraints, mathematical optimization techniques are applied to find the best possible reaction pathway according to some suitably chosen objective function. The optimization problem for the network model can be formulated as

$$\begin{aligned} &\min_{\boldsymbol{f}',\boldsymbol{b}} \left\{ \begin{matrix} \phi_1 \\ \vdots \\ \phi_n \end{matrix} \right\} \\ &s.t. \boldsymbol{A}' \cdot \boldsymbol{f}' = \boldsymbol{b}, \\ &\quad \boldsymbol{f}', \boldsymbol{b} \geq 0, \end{aligned} \tag{11}$$

ϕ_n represents the measure of the different objectives. The maximization of the overall product yield is certainly one appropriate objective function; alternative performance measures will be introduced in the next section. If more than one objective should be considered, special solution strategies must be applied. In general, the resulting multiobjective optimization problem can either be solved by optimizing the weighted sum of the single objectives or by minimizing one objective while constraining all others by a certain upper bound.[37] Multiple objectives are especially useful, if competing objectives should be investigated. For example, economic criteria have shown to be in conflict with thermodynamic or ecological criteria in process design (see, e.g.,[38–40]). In the context of RNFA the information for the calculation of detailed performance criteria is not available, such that more fundamental indicators must be defined. In principle, even the target molecule or the set of target molecules can be determined as part of the optimization procedure, but the optimization problem becomes easier, if they are chosen *a priori* by integrating RNFA in product design approaches.[27]

Most likely, several reaction pathways will fulfill the requirements and share the same value of the objective function. However, even though the pathways might have the same objective function value (i.e., the same product yield); they differ at least in their flux distributions. This means that a target molecule can be synthesized by multiple reaction routes. These so-called alternate optima are particularly plausible because ideal reaction yield coefficients are assumed, i.e., yield constraints as introduced by means of Eq. 9 are not present. Consequently, all optimal reaction pathways, i.e., all extreme point solutions of problem (Eq. 11), need to be known to enable a comprehensive evaluation. The infinitely many flux scenarios can then be represented by linear combinations of these extreme point solutions.[35] In order to obtain the complete set of alternate optima, different solution strategies have been developed (e.g.,[41,42]). Lee et al.[26] propose an algorithm, which is easy to implement. Here, binary variables are introduced to indicate active reaction steps. Then the model is integrated in a recursive mixed-integer problem (MIP), which includes constraints for changing the so-called basis variables in each iteration step. In this way new reaction pathways with different combinations of active reaction steps are identified until no other solution with the same objective function value is found.

The sketched solution strategies are widely used in metabolic pathway analysis (e.g.,[26,21]), but some specific characteristics must be kept in mind in a RNFA application. Yield constraints are not considered in the identification step, because they would cause a different objective function value for each pathway such that the pathways would not be

Table 1. Definition of Evaluation Criteria for RNFA

	evaluation criteria	Definition
mass balance criteria	raw materials	f_{1j}
	main reactant	$f_{1\text{reactant}}$
	hydrogen	f_{1H2}
	by-products	b_i
	water	b_{H2O}
	carbon dioxide	b_{CO2}
	product yield	$Y_{\text{target}} = \frac{b_{\text{target}}}{f_{1\,\text{reactant}}} \quad (12)$
	carbon efficiency	$\eta^{C} = \frac{b_{\text{target}} N_{C,\text{target}}}{\sum_{i=1}^{s} (\boldsymbol{A}_1 \cdot \boldsymbol{f}_1)_i N_{C,i}} \quad (13)$
	number of reaction steps	N_R
energy criteria	energy efficiency	
	combustion	$\eta^{E,com} = \frac{b_{\text{target}} H_{\text{com,target}}}{\sum_{i=1}^{s} (\boldsymbol{A}_1 \cdot \boldsymbol{f}_1)_i H_{\text{com},i}} \quad (14)$
	formation	$\eta^{E,form} = \frac{b_{\text{target}} H_{\text{form,target}}}{\sum_{i=1}^{s} (\boldsymbol{A}_1 \cdot \boldsymbol{f}_1)_i H_{\text{form},i}} \quad (15)$
cost criteria	investment costs	$\text{IC} = 3 \cdot 10^6 (\Delta E)^{0.84} \quad (16)$
	total annualized cost	$TAC = \frac{IC \cdot i}{1-(1+i)^{-n}} + \sum_{i=1}^{s} (\boldsymbol{A}_1 \boldsymbol{f}_1)_i C_i \quad (17)$
	raw material cost	$C_{\text{max,reactant}} = \frac{b_{\text{target}} C_{\text{target}} - \frac{IC \cdot i}{1-(1+i)^{-n}}}{f_{1\,\text{reactant}}} \quad (18)$
	total annualized revenues	$TAR = b_{\text{target}} C_{\text{target}} - TAC \quad (19)$

identified as alternate optima by the algorithm. Furthermore, the solution algorithms are limited to linear problems. They can only be utilized in RNFA, if linearity is retained in the pathway identification step. Hence, yield and other nonlinear constraints are added in a subsequent analysis step, where the promising reaction pathways can be categorized by a number of (nonlinear) evaluation criteria to be introduced in the following section. In summary, the solution strategy can be structured in the following steps:

1. Formulation of the optimization problem (Eq. 11) including the selection of an objective function;
2. Solution of the MIP problem resulting from a reformulation of Eq. 11 to identify all possible reaction pathways with the same value of the objective function;
3. Inclusion of yield constraints (Eq. 9) and calculation of alternate evaluation criteria for a fixed reaction pathway;
4. Evaluation and discussion of the reaction pathways based on selected evaluation criteria;
5. Transformation of the reaction flux data into process flows, using Eqs. 4–7 and Eq. 10, if desired.

Evaluation criteria for reaction networks

Suitable criteria for the evaluation of reaction pathways can easily be derived from the mass balance, which is the primary source of information in the early design stage. Its analysis helps to identify promising reaction pathways as well as reaction bottlenecks. An overview of possible evaluation criteria is presented in Table 1. Here, the criteria are divided into three groups, namely mass balance, energy and cost criteria. In the following the individual evaluation criteria will be explained and discussed in detail.

Reaction pathways usually differ in the use of main and auxiliary reactants as well as in their byproduct formation. In the first place, the flow rate of the reactant hydrogen f_{H2} and of the byproducts water b_{H2O}, and carbon dioxide b_{CO2} should be balanced in the context of biomass conversion. In contrast to fossil feedstock, biomass is a highly oxygenated compound which typically has to be reduced to less oxygenated products in order to fulfill the requirements of the intended application. Oxygen can be removed either in form of water or of carbon dioxide. While hydrogen has to be supplied to the process in the first case, a certain fraction of carbon is lost as byproduct instead of being converted to the actual product in the second case. Hydrogen can only be sustainably produced from solar water splitting as any process relying on reforming of fossil or renewable raw material will also contribute to a carbon loss due to release of carbon dioxide. This tradeoff must be considered comparing the different reaction pathway alternatives either indirectly by assessing the hydrogen demand and the carbon loss or be including the hydrogen generation pathway into the reaction network.

Furthermore, the mass balance of the reaction network also provides an estimate of the required additional reactants as well as of the expected byproducts and their molar amounts. The analysis, thus, diagnoses requirements on subsequent process design related to the supply of raw materials, the handling of byproducts and the separation tasks required. Generally, these evaluation criteria can be discussed considering two different scenarios: Either how much raw material is required to produce a certain amount of product or how much product can be produced from a defined amount of raw material.

Both aspects are covered by the overall product yield Y_{target}, which is certainly one of the most important performance indicators. According to Eq. 12 (see Table 1 for Eqs. 12–19), it relates the (molar) amount of the product component *target* to the amount of consumed limiting feedstock. If the yield coefficients of the individual reaction steps are chosen to be less than one— the most common scenario in practice—, the overall product yield also reflects the influence of the number of reaction steps in a reaction pathway. The number of reaction steps can also be counted as independent evaluation criterion because pathways with fewer reaction steps are more desirable at the same yield due to the lower operational and investment costs expected.

Even though the overall product yield is a widely-accepted evaluation criterion, it does not always reflect all aspects of a reaction pathway properly. Since the yield is only calculated based on the amounts of product and limiting main reactant, it is, e.g., not suitable for a comparison of synthesis routes, if they do not share common sets of main and auxiliary reactants and products. Therefore, more capable efficiency indicators must be introduced. One example is the carbon efficiency η^C, which is defined in Eq. 13 as the ratio of the molar amount of carbon in the desired product to the one in the raw materials, whereas the number of carbon atoms per molecule is given by N_C. The carbon efficiency can also be interpreted as atom economy as postulated in the principles of "green chemistry",[43] defined for carbon. Depending on the reaction pathway, this and similar evaluation criteria may be correlated with the overall product yield and also with each other such that it might be sufficient to involve only one of them in the analysis.

Besides the performance indicators derived from the mass balance, a number of criteria can be defined to evaluate the energy balance of a reaction pathway. The energy efficiency η^E can be determined based on Eqs. 14 or 15, where either the enthalpy of combustion H_{com}, or the enthalpy of formation H_{form} at standard conditions is employed. In both cases, the energy stored in the desired product is related to the energy supplied by all reactants. While the energy efficiency of combustion focuses on the future application (i.e., the combustion in an engine), the energy efficiency of formation rather stresses the quality of the production process. The use of enthalpies at standard conditions is inevitable, because the operating temperatures of the reactions are typically not known. This lack of information is also an argument against the use of more telling exergy efficiencies to replace energy efficiencies. Furthermore, even if operating temperatures and pressures of the reactions would be known, specific entropies required to evaluate exergy efficiencies are often not available and are hard to predict for the substances of interest.

As operating conditions are not known for most reaction steps, a simple energy balance for each reaction pathway can be based on the calculation of the reaction enthalpy ΔH_R given by

$$\Delta H_R = \sum_{i=1}^{s} b_i H_{\text{form},i} - \sum_{i=1}^{s} (\boldsymbol{A}_1 \boldsymbol{f}_1)_i H_{\text{form},i}, \qquad (20)$$

where $H_{\text{form},i}$ are the enthalpies of formation at standard conditions (25°C) for each substance i. Although this overall energy balance of a reaction route is dominated by the choice of the main raw material and the product,[27] candidate pathways differ in their energy balances due to different reaction yields as well as due to the diverse combinations of (auxiliary) reactants and byproducts. Again, this becomes especially important, if reaction routes are compared, which do not start from the same feedstock or which do not form the same main product. Simple energy balances, which can be stated with the knowledge available, can provide a foundation for more complex investigations. For instance, Huijbregts et al.[44] show that the cumulative energy demand (CED) is a good predictor for the environmental burden of a process,[45] because their process evaluations based on the CED lead to results similar to those from more detailed life-cycle assessment methodologies like the ecological footprint.[46,47] This is advantageous because the latter methodologies require inputs, which are not accessible in the early design phase.

Lange[48] also pursues the idea of deriving more complex criteria from a simple energy balance. In particular, he has introduced Eq. 16 to correlate the investment cost IC to the energy loss ΔE of a production system. The investment cost IC (in U.S-\$) are calculated for the year 1993, and the energy loss ΔE (in MW) is computed as the difference between the lower heating values of the feed streams to the plant and the one of the product stream. According to this definition, the energy loss ΔE can be determined from

$$\Delta E = \sum_{i=1}^{s} (\boldsymbol{A}_1 \boldsymbol{f}_1)_i \left| H_{com,i} \right| - b_{\text{target}} \left| H_{com,\text{target}} \right|, \qquad (21)$$

in the context of RNFA. This equation describes the influence of the chemical transformation, but does not focus on additional fuel inputs. If the investment cost IC is known from Eq. 16, the total annualized cost TAC can be estimated easily from Eq. 17 with the interest rate i (a fraction of 1), the integer plant lifetime n (in years), the raw material cost C_i (in \$/mol), and the material flow $\boldsymbol{f}_1$ (in mole/year).[49] Even though operating costs are not included at this point, the flux distributions provided by RNFA enable a preliminary economic evaluation. Of course, Eq. 17 also allows the estimation of product cost required for an economically benign production. For example, since transportation fuel prices must lie in a certain range to be competitive (e.g., about 1 \$/kg), it is also possible to predict the maximal raw material cost $C_{\text{max,reactant}}$ from Eq. 18 assuming that the feedstock cost is dominated by this factor. The process can only be economically favorable, if the real feedstock cost is expected to be lower than the calculated value, otherwise alternative supply or even production strategies must be discussed. Or, to put it another way, Eq. 18 can be reformulated to calculate the total annualized revenues (TAR), Eq. 19, as a profit measure, if reliable raw material costs are known. Altogether, the calculations of investment, total annualized cost as well as revenues and raw material cost show that important information can be gained from RNFA, if a basic energy balance is included.

In addition to the overall energy balance, reaction pathways differ in the sequence of reaction enthalpies along the pathway. Pathways can be compared by plotting the enthalpy per reaction step over the reaction sequence. Finley et al.[50] present these so-called thermodynamic landscapes for biodegradation pathways, but do not discuss them from a process perspective. In contrast to almost thermoneutral reaction steps, strongly exothermic or endothermic reaction steps require heat exchange, which implies the use of additional heat exchanger equipment and may lead to supplemental utility cost. Reactions with a strong heat tone are, therefore, less favorable from an economical perspective. Therefore, thermodynamic landscapes provide the basis for another RNFA performance indicator, although the determination of operating conditions and equipment dimensions would enhance the accuracy of the analysis.

The list of evaluation criteria can steadily be expanded, but indicators which require a lot of input data are not easily applicable in the early design phase.[18] Most often, the collection of data is extremely time-consuming and at the same time affected by uncertainties. For example, toxicities and catalyst cost have been considered in a former study[51] though their values are prone to significant uncertainty. Toxicity measures are defined according to risk and safety phrases, but material and safety data cannot be found for all substances. Catalyst costs, on the other hand, are hard to estimate since they change rapidly and scale nonlinearly with the globally required amount. Furthermore, for most of the reactions, a particular catalyst is not known, preventing a sound cost estimate by principle. Hence, in order to limit the effort for data collection, knowledge-intensive indicators should only be applied after a set of promising reaction pathways has been identified by means of easily accessible criteria, which can be derived from mass and energy balances.

More complex evaluation criteria are favorably applied for further discrimination of the promising reaction pathways. These criteria typically rely on a process perspective and hence form the interface between RNFA and conceptual process design. Examples for such criteria are feasibility and effort of the separations after each reaction step. The recycling of unconverted feedstock has to be accounted for to obtain reliable results. Of course, a first layout of the process structure must be provided, which in turn requires appropriate property data (e.g., vapor pressures) which have been hardly measured yet for biobased molecular products and are difficult to predict. Therefore, these evaluation criteria are often not applicable in pathway analysis.

In summary, all evaluation criteria should help to identify promising reaction pathways and to discuss tradeoffs between alternative production routes. Based on these results the number of promising solutions can be reduced while increasing the level of detail in the analysis. This way, future research reaction steps and catalysts can be guided efficiently by focusing on the still existing bottlenecks which may exist in otherwise promising pathways.

Illustrative Case Study - Biofuel Production

Sustainable processes for the conversion of whole plants into fuels are developed in the interdisciplinary research center "Tailor-Made Fuels from Biomass" (TMFB) at RWTH Aachen University.[52] The target biofuels are blends of well-defined oxygenated components with tailored properties for novel low-temperature combustion engines. Contrary to other approaches, selective bio- and chemocatalysis are applied to preserve the synthesis power of nature. In this context, itaconic acid (IA, $C_5H_6O_4$) is chosen as a key intermediate for the production of biofuels. IA is a C_5-dicarboxylic acid, which, for example, can be produced from glucose by aerobic fungal fermentation[53] and has been considered as one potential building block for both, specialty and commodity chemicals derived from biomass.[54]

Construction of the itaconic acid network and data collection

As an example, a reaction network toward fuel components starting from IA is built and evaluated in this case study. Even though the definition of relevant criteria for the selection of fuel components is an ongoing research topic,[27,55,56] it is well known that the energy content of biomass must be increased during biofuel production. Following this principle, reactions increasing the hydrogen to carbon ratio and decreasing the oxygen to carbon ratio are integrated in the network. Overall, the network should summarize the range of molecules, which can be produced from IA given this state of research in bio- and chemocatalysis. It has been built based on chemists' knowledge and on information retrieved from a detailed literature search using the research discovery tool SciFinder with access to the CAS databases and to MEDLINE.[57] Most of the reactions can also be found in a number of recently published review articles (e.g.,[5,6,58]). In order to limit the complexity of the network, different isomeric forms of the molecules are only considered, if they are crucial for the successive reactions. Currently, the reaction network is composed of 116 reactions and 80 substances. Although the reactions are reported in the literature, very little information on the reactants, products and the reactions exists. In particular, only a fourth of the starting, intermediate and target molecules are listed in the DIPPR database.[61] Furthermore, yield data are only available for less than 30% of the active reaction steps. In order to provide a standardized basis for the evaluation, the required properties are calculated by Joback's group contribution method[28] for all components while postponing the impact of inaccuracies in the property prediction on the results of RNFA at this point. For this case study, 3-methyl-tetrahydrofuran (3-MTHF, $C_5H_{10}O$) is chosen as a target molecule, because of its favorable physical and combustion properties and the promising results of preliminary engine tests.[59] Supporting information on the network may be found in the online version of this article.

Modeling, formulation of the optimization problem and identification of reaction pathways

To detect all possible reaction pathways the general optimization problem in Eq. 11 is refined to result in

$$\begin{aligned} &\max_{\boldsymbol{f}',\boldsymbol{b}} b_{\mathrm{MTHF}} \\ &\quad s.t. \boldsymbol{A}' \cdot \boldsymbol{f}' = \boldsymbol{b}, \\ &\quad f_{1IA} = \alpha, \\ &\quad \boldsymbol{f}', \boldsymbol{b} \geq 0. \end{aligned} \tag{22}$$

Its solution identifies (all) the reaction pathway(s) with the maximal molar yield of the target molecule 3-MTHF, which

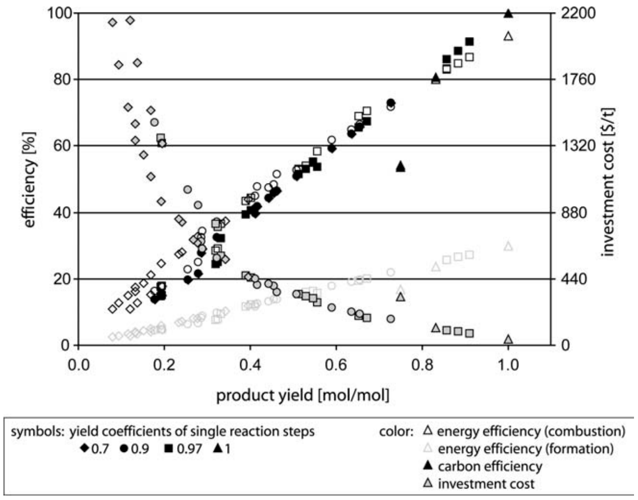

Figure 2. Correlation between the 3-MTHF yield and the performance indicators (energy and carbon, efficiency, annualized investment cost).

Hence, yield can be chosen as the representative criterion in the evaluation of the reaction pathways.

is one component of the product vector ***b***. The main feed stream to the network (i.e., IA), which is represented by the flow f_{1IA}, is fixed in order to avoid infinite reaction fluxes and to allow for a comparison of different solutions. The value α is chosen to be 1 mol/s in the study. As described before, the problem is first formulated without yield constraints to retain linearity of the problem and is then solved by MIP. Once all possible pathways are identified, yield constraints and evaluation criteria are added and the mass balances are calculated for all the predefined pathways. If no yield information is available, the yields of the corresponding reaction steps are fixed to reasonable values; the influence of these assumptions is analyzed subsequently.

In total 19 different reaction pathways can be found in the network solving the optimization problem[22] in the general algebraic modeling system (GAMS).[60] These pathways are not branched and consist of the minimal number of reaction steps for the production of the specified molecule.

Selection of evaluation criteria

Suitable evaluation criteria are selected next. In particular, the energy efficiency based on the combustion and formation enthalpy, the carbon efficiency as well as the annualized investment cost, which scales with the amount of MTHF produced, are calculated for all the 19 possible reaction pathways. Furthermore, four of these calculations are performed each assuming a different yield coefficient (0.7, 0.9, 0.97, 1) for every reaction step, where no yield information is available. The results are shown in Figure 2 plotting the different evaluation criteria over the product yield. The evaluation criteria are displayed on two ordinates, one for the efficiencies (left), and one for the cost information (right). While colors and fillings stand for the particular evaluation criteria, the symbols mark the calculations for the four yield coefficients. Thus, each data point in the diagram indicates the value of a specific evaluation criterion for one reaction pathway assuming a defined yield coefficient per reaction step. Six of 19 reaction pathways require one or more additional reactants like methanol or ethanol. However, these solutions do not particularly stand out as the evaluation criteria account for the use of these additives. Thus, they are not specially labeled. Figure 2 clearly shows that all evaluation criteria correlate with the overall yield of the product 3-MTHF almost independently of the reaction pathway. The remarkably strong correlation can be explained because all reaction pathways are based on the same main feedstock (IA) and form the same main product (3-MTHF). In this case study, the same results are obtained ranking the reaction pathways by any of the aforementioned-named criteria, such that the yield of 3-MTHF is applied as the representative indicator in the following analysis. A multicriteria evaluation is not required at this point.

At the same time it should be mentioned that such strong correlations do not apply for all evaluation criteria. As an example, the relative amount of hydrogen per unit of reactant (IA) and per unit of product (3-MTHF) is presented in Figure 3 for the setup described earlier. Even though a general trend can be diagnosed for both indicators (the former one increases, while the latter one decreases with 3-MTHF yield), the overall yield does not seem to be the dominant factor.

Evaluation of the identified reaction pathways

Even in the cumulative illustration of Figures 2 and 3, differences between the particular reaction pathways can be discovered. By considering all pathways with the same yield coefficient for the single reaction steps (marked by the same symbol), it can be shown that some pathways reach higher overall product yields and better efficiencies than others, even though the same performance is assumed for each reaction step. Hence, some pathways require a lower yield per reaction step to meet a certain desired product yield. However, the influence of the yield coefficient on the overall performance has to be examined, before promising pathways can be selected for further analysis.

The yield coefficients of the individual reaction steps are varied in the range from 0.5 to 1; for each value of the yield coefficient, the corresponding mass balances are calculated for all the reaction pathways. If the yield coefficient for a specific reaction is known, it is fixed to the reported value and kept unchanged. As a result, Figure 4 shows the two selected evaluation criteria, namely the 3-MTHF yield and the hydrogen consumption, plotted over the yield coefficient

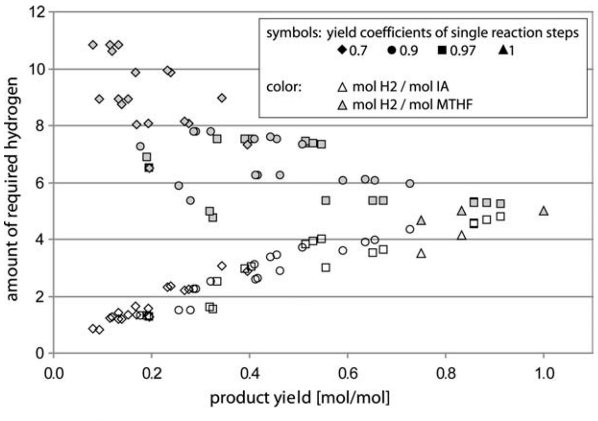

Figure 3. Compared to other evaluation criteria, there is a weaker correlation between the product yield of 3-MTHF and the hydrogen demand.

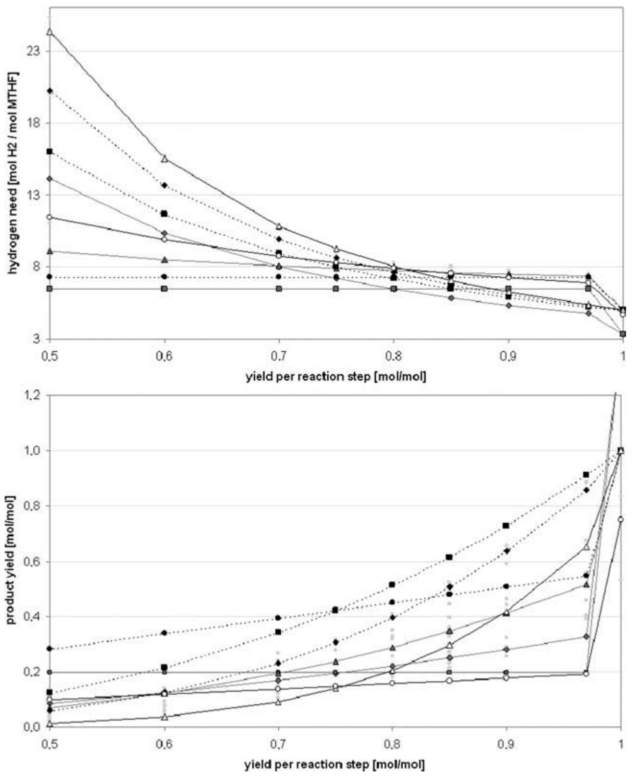

Figure 4. The influence of the yield coefficients for the individual reaction steps on the evaluation criteria differs for the different reaction pathways.

The graphs show this influence for selected pathways.

assumed for the single reaction steps. To demonstrate the diverse types of solutions found, only some representative reaction pathways are sketched in the figure. The same labeling for a pathway is used to consider the yield and the hydrogen need in the upper and lower half of Figure 4. The locations of all other solutions are either overlapping with the pathways shown or are marked by gray dots. All reaction pathways follow the same general trend. As expected, the overall yield of 3-MTHF increases with the yield coefficient of the individual reaction steps and even product yields greater than one can be obtained in those cases, where more than one mole of 3-MTHF can be formed per mole IA. In contrast, the amount of hydrogen required for the production of one mole of product decreases with an increasing yield coefficient because more hydrogen is then consumed in byproduct formation. However, the scenarios differ in both, their absolute yield and hydrogen demand for a particular yield coefficient, and their dependencies on the yield coefficients. Generally, two types can be distinguished. An example of each is given in Figure 4 marked by the bold lines. On the one hand the yield of 3-MTHF and the hydrogen need of some reaction pathways are strongly affected by minor variations in the yield coefficients deviating from one. The gradient decreases with the yield coefficient such that (almost) a plateau is reached (e.g., pathway labeled by gray squares). On the other hand, the performance of some reaction pathways deteriorates more slowly, but constantly over the entire yield range (e.g., pathway labeled by black squares). Obviously, reaction pathways with a high yield of 3-MTHF and a low hydrogen demand are favorable. However, it must be assessed which target yield is viable, both from a catalytic and a process engineering point of view or which target yield is required to realize an economically feasible production. In order to guarantee good performance for a wide range of yield, a certain reaction pathway might be chosen despite of its drawbacks under ideal conversion conditions. According to this policy and accepting the overall yield as the most important evaluation criterion, the five pathways with the highest yields of 3-MTHF are chosen for further analysis assuming a yield coefficient of 0.97 per reaction step. This scenario is selected to allow some operational losses. The yield curves of the chosen reaction pathways approximately follow the dashed lines in Figure 4 such that they still outperform the other solutions for a lower yield coefficient of 0.9, and they are even ranked in the upper third for a very low-yield coefficient of 0.7. A group of exactly five reaction pathways is taken into consideration for

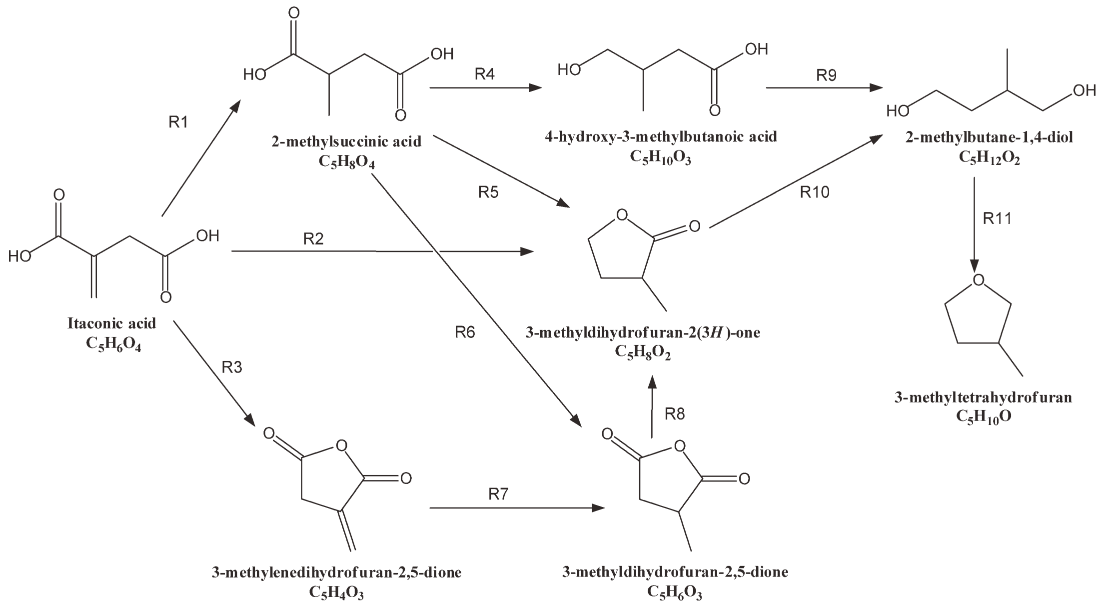

Figure 5. Reaction network summarizing the active reaction steps of the five most promising reaction pathways.

Table 2. Comparison of the Different Evaluation Criteria for Selected High-Yield and Low-Yield Reaction Pathways Assuming a Yield Coefficient of 0.97 per Reaction Step

rank of pathway	active reactions	mass balance				energy			cost	
		hydrogen need [mol/ mol IA]	water formation [mol/mol IA]	MTHF yield [mol/mol IA]	carbon efficiency [%]	energy of combustion [%]	energy of formation [%]	energy loss [MW]	investment cost [$/t]	max. IA cost [$/t]
1	2,10,11	4.79	2.85	0.9127	91.27	86.53	27.21	16.97	77.32	922.68
2	1,4,9,11	4.68	2.74	0.8853	88.53	84.68	26.39	19.72	87.69	912.31
3	1,5,10,11	4.68	2.77	0.8853	88.53	84.68	26.39	19.72	87.69	912.31
4	3,7,8,10,11	4.54	2.74	0.8587	85.87	83.05	25.60	22.24	97.03	902.97
5	1,6,10,11	4.57	2.71	0.8587	85.87	82.86	25.60	22.54	98.14	901.86
18	–	1.27	1.49	0.1957	14.91	17.73	4.66	505.66	1338.18	−338.18
19	–	1.32	1.44	0.1921	14.64	17.35	4.57	519.31	1368.45	−368.45

further analysis because they clearly stand out from the other solutions.

Figure 5 shows all substances and reactions which appear in any of the five selected pathways. The active reaction steps for each pathway are listed in Table 2. This table also summarizes the relevant data obtained by RNFA assuming a yield coefficient of 0.97 per reaction step. For comparison, the information is also provided for the two least promising pathways of the ranking. In contrast to the low-yield solutions, the top-ranked pathways carry reasonable 3-MTHF yields between 0.9127 and 0.8587. For the 3-MTHF production, none of the reaction pathways requires additional reactants besides IA and hydrogen, which is consumed in similar amounts by each pathway. The more 3-MTHF is produced, the higher is the hydrogen demand and at the same time the formation of water. For high-yield solutions a larger amount of feedstock reacts to the final product such that more hydrogen is needed for the successive deoxygenation along the pathway; in turn, more water is formed as a byproduct. Each reaction route expresses a different set of byproducts consisting of the substances along the pathway, which only appear in small amounts. By detecting promising pathways, the number of interesting reactions is already reduced to a manageable number (Figure 5), and can be further reduced by first focusing on reactions, which are included in several pathways. Particularly, reaction 11 is active in all pathways such that a sound performance of this reaction is of utmost importance.

Furthermore, all the efficiency criteria only vary in the range of 5% for the top-five solutions. Despite the low values in terms of energy of formation, the good results indicate a reasonable combination of reactants and products. The estimation of specific investment costs per unit of product supports this positive impression. The calculations of the energy loss and annualized specific investment cost are based on Eqs. 16 and 21, presuming an annual production of 100,000 tons of 3-MTHF, an interest rate of 8%, and a plant lifetime of 10 years. Updating the cost by means of the chemical engineering plant cost index, results in investment cost of 77–98 $/t of product. Note that this cost estimate does not account for the cost of the raw material IA, which is currently sold at about 4,000–40,000 $/t, depending on amount and purity. Therefore, the difference in the predicted investment cost should not be overestimated. Furthermore, the current cost of IA is well above the maximal raw material cost predicted by Eq. 18, if an average fuel price of 1 $/kg is requested. This fuel price is chosen in accordance with the work of McAloon et al.[62] where bioethanol prices well below 1 $/kg are postulated. Consequently, the raw material cost must be significantly reduced to render 3-MTHF an economically viable biofuel, for example, by an integrated production of IA from biomass or by switching to another less costly platform chemical.

In contrast to the closely-related top-five solutions, major differences can be found, if they are compared to the reaction pathways at the end of the ranking. Those pathways show very low yields of 3-MTHF, which comes along with a poor rating in all other criteria, such that these pathways are not acceptable for the realization of an economically efficient process. In particular, a consideration of the investment cost shows that even minor losses in the individual reaction steps may have significant effects, especially taking the specific investment cost per unit product into account. Here, cost under 30 $/t can be obtained by the best reaction pathways under full conversion. This stresses the importance of a comparison of the alternative reaction pathways in the early design phase to make sure that only the most promising pathways from biorenewables to next generation fuels are investigated.

Comparison of the thermodynamic landscapes for different reaction pathways

While the energy efficiency as well as the investment cost evaluated so far just consider the cumulated reaction enthalpy, Figure 6 shows the thermodynamic landscapes, i.e., the reaction enthalpy for every step in the sequence. The courses of the least (gray, dashed lines) as well as of the most promising

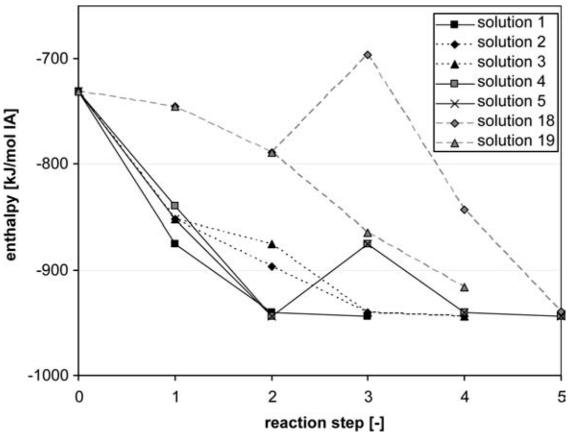

Figure 6. Sequences of reaction enthalpies illustrate the differences of selected reaction pathways with high-yield (solutions 1–5), and low-yield (solutions 18 and 19).

pathway alternatives are plotted for the conversion of one mole of IA. Even though the definition of an optimal sequence remains difficult, major differences can be observed between the low-yield and the high-yield alternatives, which in turn just differ slightly from each other. Besides, it is interesting to note that two of the promising pathways rely on an endothermic reaction step within the sequence. In these cases, process heat has to be either provided by a hot utility or by internal heat exchange, while in case where only exothermic reaction steps are present, the reaction heat can be exported and possibly reused. However, at least the reaction conditions must be considered for a more comprehensive assessment of the economical implications of the energy balance. Unfortunately, this information is lacking for many reaction steps.

The example of the production of 3-MTHF from IA demonstrates how RNFA can be used to enumerate alternative reaction pathways in a reaction network and, subsequently, rank them to identify promising and deficient solutions. In addition, a couple of issues can be identified which require more attention during the subsequent process design.

Discussion

The information gained by RNFA will be discussed from the perspective of four topics, namely (1) the supply of reactants, (2) the handling of byproducts, (3) the reaction performance, and (4) the refinement of the energy balance.

Most importantly, the source of the main reactant as well as the reaction additives to be supplied to the process has to be assessed. In some cases, the reactants can be produced from biomass within an integrated biorefinery, otherwise they must be provided by an external source. In the 3-MTHF case study, the promising pathways do not require any additional reactants besides IA and hydrogen. Nevertheless, a sustainable supply of the fairly large amount of hydrogen must be guaranteed either by steam reforming, pyrolysis of biomass residuals, electrolysis (with sustainable electricity sources like solar or wind energy) or a combination of photo- and thermophilic fermentation of a biorenewable substrate. Moreover, the analysis shows that the current market price of IA does not allow an economically feasible conversion of this substrate into 3-MTHF for use as a biofuel. Two options to resolve this problem need to be studied. First, a decreasing trend of the IA price may be realized if larger amounts are produced relying on existing process concepts. Second, an alternative production process for IA using native biomass may be found to result in an economically reliable overall process. In both cases, the production of the (auxiliary) reactants—hydrogen and IA—must be integrated in the reaction network in order to properly assess the additional effort in terms of raw material demand, investment costs, etc.

Water occurs as major byproduct in all reaction pathways resulting in strongly nonideal aqueous-organic mixture which call for efficient separation strategies. As all other byproducts (e.g., methylsuccinic acid, 4-hydroxyl-3-methylbutanoic acid, methylbutanediol) are formed in minor amounts, their costly separation from 3-MTHF might be avoided, if the organic byproducts do not negatively affect downstream reactions and if the finally resulting organic mixture could be directly used as a fuel blend. Otherwise, at least the feasibility of the assumed separation steps must be checked to assure the realization of the overall process. The estimation of the minimum energy demand would help to detail the assessment of the energy balance.

As expected, even minor losses in the reaction yield of the individual reaction steps add up and significantly reduce the efficiency of the entire process. Consequently, the yield coefficients of the identified key reactions must be verified and their performance must eventually be optimized by reactor as well as catalyst design to fulfill the assumed target yield of 0.97.

Last, the energy balance should be refined. So far, the discussion is based on the sequence of reaction enthalpies along the reaction pathways at standard conditions. In many cases, however, the enthalpies of vaporization and reaction range in the same order of magnitude such that the choice of the reaction phase (i.e., liquid or vapor phase) would not only lead to extensive modifications of the thermodynamic landscape, but would also strongly impact the process structure. Hence, it is most important to fix the reaction conditions in the next design step to facilitate a more precise estimation of the energy balance of the individual reaction steps, which is the basis for assessing the potential of internal heat integration and which determines the overall energy demand of the process.

Concluding Remarks

In this article, RNFA is introduced as an easy and efficient tool for the systematic identification and classification of reaction pathways. The reusability of the reaction network enables a fast calculation of scenarios with varying, e.g., the reaction yields, the feedstock composition or the performance criterion in the optimization problem. Promising pathways are identified from the set of solution provided by MIP in a subsequent analysis, which can be based on a variety of economic, energetic or ecological indicators deduced from the mass balance. These primary indicators can be refined by additional criteria related to conceptual process design including separation indices or energy demand. The combination of these two steps offers the basis for a rapid evaluation of alternatives in an early design stage by RNFA.

The case study demonstrates how the knowledge gained from a comprehensive mass balance helps to identify the most promising reaction pathways and the most serious bottlenecks in order to guide further research activities in a rational way. Therefore, this work provides valuable support for an evaluation of biobased processes during the early design phases, but it should be complemented with respect to several aspects in the future.

Since the necessary property data is often not available for all the substances of interest, adequate property prediction methods have to be applied. In this work, the property data are considered to be free of error to keep the focus on the illustration of the methods. The further development of RNFA methodology should look at an elaborate coverage of uncertainties in the property data.

Furthermore, the handling of mixtures has been excluded by postulating an ideal separation after each reaction step. This very conservative strategy can be refined, if constraints on the separation behavior are introduced similar to those on reaction performance. Then, separation tasks can be specified for subsequent process design or the feasibility of a specific separation technique can directly be tested as part of RNFA by applying appropriate indicators (see e.g.,[63]).

While a multicriteria evaluation has not been needed in the case study of this article, it should generally be considered for a balanced analysis of the ecological, economical and technical performance of the process alternatives. Such multicriteria evaluation can easily be incorporated in RNFA. Apart from that, this article only examines the conversion of a single platform chemical into a single product, but RNFA can also be utilized to evaluate more complex biorefinery concepts with a biobased feedstock and a portfolio of products.[64,65] RNFA is particularly useful for the comparison of reaction pathways, which differ in the reactants and/or products. It also has been successfully utilized in an integrated product and process design approach.[27] An automatic generation of the reaction network should be considered in future work, because the outcome of RNFA strongly depends on the completeness of the network.

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.

Literature Cited

1. Sanders J, Scott E, Weusthuis R, Mooibroek H. Bio-refinery as the bio-inspired process to bulk chemicals. *Macromol Biosci*. 2007;7:105–117.
2. Marquardt W, Harwardt A, Hechinger M, Kraemer K, Viell J, Voll A. The biorenewables opportunity - toward next generation process and product systems. *AIChE J*. 2010;56:2228–2235.
3. Bozell J. Connecting biomass and petroleum processing with a chemical bridge. *Science*. 2010;329:522–523.
4. Lange J-P, Price R, Ayoub P, Louis J, Petrus L, Clarke L, Gosselink H. Valeric Biofuels: A platform of cellulosic transportation fuels. *Angew Chem*. 2010;122:4581–4587.
5. Kamm B, Gruber P, Kamm M. *Biorefineries - Industrial Processes and Products*. Vol. 1 and 2. Weinheim: Wiley-VCH; 2005.
6. Corma A, Iborra S, Velty A. Chemical routes for the transformation of biomass into chemicals. *Chem Rev*. 2007;107:2411–2502.
7. Torres A, Daoutidis P, Tsapatsis M. Continuous production of 5-hydroxymethylfurfural from fructose: a design case study. *Energy Environ Sci*. 2010;3:1560–1572.
8. Douglas JM. *Conceptual Design of Chemical Processes*. New York: McGraw Hill; 1988.
9. Smith R. *Chemical Process Design*. New York: McGraw Hill; 1995.
10. Blass E. *Entwicklung Verfahrenstechnischer Prozesse*. Berlin: Springer; 1997.
11. Seider W, Seader J, Lewin D. *Process Design Principles*. New York: Wiley & Sons; 1998.
12. Ahmetovi E, Martin M, Grossmann I. Optimization of energy and water consumption in corn-based ethanol plants. *Ind Eng Chem Res*. 2010;49:7972–7982.
13. Geilen F, Engendahl B, Harwardt A, Marquardt W, Klankermayer J, Leitner W. Building a renewable supply chain: Selective and flexible transformation of biomass-derived platform chemicals. *Angew Chem*. 2010;49:5510–5514.
14. Gerber L, Gassner M, Maréchal F. *Integration of LCA in a thermo-economic model for multi-objective process optimization of SNG production from woody biomass*. In: Jezowski J, Thullie J, eds. *Proceedings of Escape 19*. Elsevier; 2009:1405–1410.
15. Kraemer K, Harwardt A, Bronneberg R, Marquardt W. Separation of butanol from acetone-butanol-ethanol fermentation by a hybrid extraction distillation process. *Comput Chem Eng*. 2011;35:949–963.
16. Poliakoff M, Licence P. Green chemistry. *Nature*. 2007;450:810–812.
17. Hermann B, Patel M. Today's and tomorrow's bio-based bulk chemicals from white biotechnology - a techno-economic analysis. *Appl Biochem Biotechnol*. 2007;136:361–388.
18. Sugiyama H, Fischer U, Hungerbühler K. Decision framework for chemical process design including different stages of environmental, health and safety assessment. *AIChE J*. 2008;54:1037–1053.
19. Broadbelt L, Pfaendtner J. Lexicography of kinetic modeling of complex reaction networks. *AIChE J*. 2005;51:2112–2121.
20. Li C, Henry C, Jankowski M, Ionita J, Hatzimanikatis V, Broadbelt L. Computational discovery of biochemical routes to specialty chemicals. *Chem Eng Sci*. 2004;59:5051–5060.
21. Mavrovouniotis M, Bonvin D. Towards design of reaction pathways. *AIChE Symposium Series*. 1995;304:41–51.
22. Schilling C, Schuster S, Palsson B, Heinrich R. Basic concepts and scientific applications in the post-genomic era. *Biotechnol Prog*. 1999;5:296–303.
23. Schuster S, Fell D, Dandekar T. Metabolic pathway analysis: A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat Biotechnol*. 2000;18:326–332.
24. Klamt K, Stelling J. Two approaches for metabolic pathway analysis? Trends in biotechnology. *Trends Biotechnol*. 2003;21:64–69.
25. Schilling C, Palsson B. The underlying pathway structure of biochemical reaction networks. *Proc Nat Acad Sci*. 1998;95:4193–4198.
26. Lee S, Phalakornkule C, Domach M, Grossmann I. Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Comput Chem Eng*. 2000;24:711–716.
27. Hechinger M, Voll A, Marquardt W. Towards an integrated design of biofuels and their production pathways. *Comp Chem Eng*. 2010;34:1909–1918.
28. Joback K, Reid R. Estimation of pure-component properties from group-contributions. *Chem Eng Commun*. 1987;57:233–243.
29. Lydersen A. Estimation of critical properties of organic compounds. *Eng Exp Stn Rep*. 1955:3.
30. Marrero J, Gani R. Group-contribution based estimation of pure component properties. *Fluid Phase Equilibria*. 2001;183–184:183–208.
31. Karelson M, Lobanov V, Katritzky A. Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem Rev*. 1996;96:1027–1043.
32. Brauner N, Stateva R, Cholakov G, Shacham M. Structurally "targeted" quantitative structure-property relationship method for property prediction. *Ind Eng Chem Res*. 2006;45:8430–8437.
33. Gubbins K, Moore J. Molecular modeling of matter: impact and prospects in engineering. *Ind Eng Chem Res*. 2010;49:3026–3046.
34. Maginn E, Elliott J. Historical perspective and current outlook for molecular dynamics as a chemical engineering tool. *Ind Eng Chem Res*. 2010;49:3059–3078.
35. Nocedal J, Wright S. *Numerical Optimization*. New York: Springer; 2006.
36. Missen R, Smith W. Yield, selectivity and all that. *Chem Eng Edu*. 2000:320–325.
37. Miettinen K. *Nonlinear Multiobjective Optimization*. Boston, MS: Kluwer Academic Publishers; 1999.
38. Gassner M, Maréchal F. Methodology for the optimal thermo-economic, multi-objective design of thermodchemical fuel production from biomass. *Comp Chem Eng*. 2009;33:769–781.
39. Martinez P, Eliceche A. *Biobjective optimization using Environmental and Economic Functions in Utility Plants*. In: Pierucci S, Buzzi Ferraris G, eds. *Proceedings of Escape 20*. Elsevier; 2010:853–858.
40. Ponce-Ortega J, Mosqueda-Jiménez F, Serna-Gonzáles M, Jiménez-Gutiérrez A, El-Halwagi M. A Property-based approach to the synthesis of material conservation networks with economic and environmental objectives. *AIChE J*. 2010.doi 10.1002/aic.12444.
41. Matheiss T, Rubin D. A survey and comparison of methods finding all vertices of convex polyhedral sets. *Math Operations Res*. 1980;5:167–185.
42. Swart G. Finding the convex hull facet by facet. *J Algorithms*. 1985;6:17–48.
43. Anastas P, Warner J. *Green Chemistry: Theory and Practice*. New York: Oxford University Press; 1998.
44. Huijbregts M, Hellweg S, Frischknecht R, Hendriks H, Hungerbühler K, Hendriks A. Cumulative energy demand as predictor for the environmental burden of commodity production. *Environ Sci Technol*. 2010;44:2189–2196.
45. Verein Deutscher Ingenieure. *Cumulative Energy Demand, Terms, Definitions, Methods of Calculations*. Düsseldorf: VDI Richtlinien 4600; 1997.
46. Consoli F, Allen D, Boustead I, Fava J, Franklin W, Jensen A, De Oude N, Parrish R, Perriman R, Postlethwaite D, Quay B, Séguin J, Vigon B. *Guidelines for Life-Cycle Assessment: A 'Code of Practice'*. Washington DC: Society of Environmental Toxicology and Chemistry; 1993.
47. Wackernagel M, Yount J. The ecological footprint: an indicator of progress toward regional sustainability. *Environ Monit Assess*. 1998;51:511–529.
48. Lange J-P. Fuels and chemicals manufacturing - guidelines for understanding and minimizing the production costs. *Cattech*. 2001;5:82–95.
49. Biegler L, Grossmann I, Westerberg A. *Systematic Methods of Chemical Process Design*. Upper Saddle River: Prentice Hall PTR; 1997.

50. Finley S, Broadbelt L, Hatzimanikatis V. Computational framework for predictive biodegradation. *Biotechnol Bioeng*. 2009;104:1086–1097.
51. Besler A, Harwardt A, Marquardt W. *Reaction networks - A rapid screening method*. In: Jezowski J, Thullie J, eds. *Proceedings of Escape 19*. Elsevier; 2009:243–248.
52. Wimmer T. Tailor-Made Fuels from Biomass - Excellence RWTH Aachen University establishes a Fuel Design Center. *Nature Jobs*. 2007:502–502.
53. Willke T, Vorlop K-D. Biotechnological production of itaconic acid. *Appl Microbiol Biotechnol*. 2001;56:289–295.
54. Werpy T, Petersen G. Top Value Added Chemicals from Biomass.-Vol 1. Results of Screening for Potential Candidates from Sugars and Synthesis Gas. Report of the U.S. Dept of Energy; 2004.
55. Müther M, Janssen A, Pischinger S, Klankermayer J, Leitner W. Maßgeschneiderte Biokraftstoffe für zukünftige Brennverfahren - Anforderungen und Lösungsansätze. In: 7. *VDI-Tagung Innovative Fahrzeugantriebe*. 2010.
56. Hechinger M, Marquardt W. Targeted QSPR for the prediction of the laminar burning velocity of biofuels. *Comp Chem Eng*. 2010;34:1507–1514.
57. SciFinder. American Chemical Society. 2007; www.cas.org.
58. Huber G, Iborra S, Corma A. Synthesis of transportation fuels from biomass: chemistry, catalysis, and engineering. *Chem Rev*. 2006;106:4044–4098.
59. Dept of Energy, Office of Energy Efficiency and Renewable Energy, 1999, 10 CFR Part 490 Alternative Fuel Transportation Program; P-series Fuels, Federal Register, 64, 26822.
60. GAMS - General Algebraic Modeling Systems. 2008; www.gams.-com.
61. DIPPR®801 database. Design Institute for Physical Properties. 2009. http://dippr.byu.edu.
62. McAloon A, Taylor F, Yee W, Ibsen K, Wooley R. *Determining the Cost of Producing Ethanol from Corn Starch and Lignocellulosic Feedstocks*. Report of the U.S Dept. of Agriculture and Energy; 2000.
63. Jaksland C, Gani R, Lien K. Separation process design and synthesis based on thermodynamic insights. *Chem Eng Sci*. 1995;50: 511–530.
64. Voll A, Harwardt A, Marquardt W. *Production of Biofuels - The Hydrogen Consumption as Performance Indicator in the Early Design Stage*. In: Stolten D, Grube T, eds. *Proceedings of the WHEC 18*. EnergieAgentur.NRW. 2010:233–239.
65. Cherubini F, Stromman A. Production of biofuels and biochemicals from lignocellulosic biomass: estimation of theoretical yields and efficiencies using matrix algebra. *Energy Fuels*. 2010;24: 2657–2667.

Manuscript received Feb. 23, 2011, and revision received May 16, 2011.