Computers and Chemical Engineering 153 (2021) 107438

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

CAT-COSMO-CAMPD: Integrated in silico design of catalysts and processes based on quantum chemistry

Christoph Gertig[a], Lorenz Fleitmann[a,b], Carl Hemprich[a], Janik Hense[a], André Bardow[a,b,c,*], Kai Leonhard[a,*]

[a] Institute of Technical Thermodynamics, RWTH Aachen University, Schinkelstrasse 8, Aachen 52062, Germany
[b] ETH Zürich, Department of Mechanical and Process Engineering, Energy & Process Systems Engineering, Tannenstrasse 3, Zürich 8092, Switzerland
[c] Institute of Energy and Climate Research (IEK-10), Wilhelm-Johnen-Strasse, Jülich 52428, Germany

ARTICLE INFO

Article history:
Received 15 January 2021
Revised 20 June 2021
Accepted 1 July 2021
Available online 8 July 2021

Keywords:
Catalysis
Kinetics
CAMD
Process design
Quantum chemistry
COSMO-RS

ABSTRACT

Catalysts are of paramount importance as most chemical processes would be uneconomical without suitable catalysts. Consequently, the identification of appropriate catalysts is a key step in chemical process design. However, the number of potential catalysts is usually vast. To suggest promising candidates for experimental testing, in silico catalyst design methods are highly desirable. Still, such computational methods are in their infancy. Moreover, simple performance indicators are commonly employed as design objective instead of evaluating the actual process performance enabled by considered catalysts. Here, we present the CAT-COSMO-CAMPD method for integrated in silico design of homogeneous molecular catalysts and processes. CAT-COSMO-CAMPD integrates design of molecular catalysts with process optimization, enabling a process-based evaluation of every designed candidate catalyst. Reaction kinetics of catalytic reactions are predicted by advanced quantum chemical methods. We demonstrate for a catalytic carbamate-cleavage process that CAT-COSMO-CAMPD successfully identifies catalyst molecules maximizing the predicted process performance.

1. Introduction

Catalysts play a key role in today's chemical industry. About 75% of chemical processes are based on catalytic reactions and over 90% of new processes developed in recent years use catalysts (Hagen, 2015). Catalysts may enhance reaction kinetics by orders of magnitude or provide the selectivity required for economic production of chemicals. Consequently, the selection of a suitable catalyst is a key step in chemical process design. Today, catalysts are usually selected based on experimental methods such as high-throughput experimental screenings or combinatorial chemistry. However, these methods are usually not target-oriented and may lead to a huge experimental effort. Moreover, the chemical design space of possible catalyst molecules is vast (Fink et al., 2005; Reymond, 2015). Therefore, it is not feasible to test all potential catalysts experimentally and the full potential of the chemical design space is thus likely not exploited. Consequently, it is highly desirable to develop in silico-methods to explore the chemical design space and to suggest the most promising candidates for experimental testing.

To identify the most promising molecules in large design spaces efficiently, Computer-Aided Molecular Design (CAMD) (Austin et al., 2016) methods have been developed. CAMD methods explore the chemical design space in silico, typically based on optimization algorithms that are employed to find the best candidates. As described in our recent review article (Gertig et al., 2020c), CAMD methods comprise of 3 building blocks:

1. An algorithm is required to explore the chemical design space and to suggest and optimize molecular structures (Papadopoulos et al., 2018). Typically, the design spaces are defined by a set of functional groups or molecule fragments. Structures of candidate molecules are generated from these groups or fragments. The optimization of these structures can be based on deterministic or stochastic optimization (Papadopoulos et al., 2018).
2. CAMD requires a sound prediction of such unknown properties, e.g., molecular and thermodynamic quantities as well as

* Corresponding authors.
E-mail addresses: abardow@ethz.ch (A. Bardow), Kai.Leonhard@ltt.rwth-aachen.de (K. Leonhard).

https://doi.org/10.1016/j.compchemeng.2021.107438

chemical properties such as reaction kinetics. The reason for the need to predict properties in CAMD is that CAMD methods are supposed to not only examine already known molecular structures, but also to suggest new candidate molecules with unknown properties. Thermodynamic properties are commonly predicted in CAMD based on group-contribution (GC) methods (Papadopoulos et al., 2018; Gmehling, 2009). GC methods offer the advantage of a straightforward implementation and computational efficiency (Gani, 2019). However, the chemical design space accessible with these methods is limited to the functional groups a GC method was trained for. Moreover, the prediction of different properties often requires several GC methods. In contrast, the computationally more demanding quantum chemical (QC) (Atkins and Friedman, 2011) methods are not limited to certain functional groups. In conjunction with thermochemistry (Paulechka and Kazakov, 2017; Umer and Leonhard, 2013), QC methods provide consistent predictions of molecular and thermodynamic properties as well as kinetics of chemical reactions (Vereecken et al., 2015). Due to the availability of increased computational capacities, even the use of advanced QC methods in CAMD has become feasible in recent years (Gertig et al., 2020c).

3. The performance of the designed molecules needs to be evaluated during CAMD based on a chosen objective. Commonly, CAMD methods use simple performance indicators as objective function defined based on predicted molecular, thermodynamic or chemical properties (Papadopoulos et al., 2018). However, such indicators may not capture all aspects and trade-offs relevant for the intended use of the designed structures. Thus, CAMD using simple performance indicators likely results in the design of sub-optimal molecules (Adjiman et al., 2014). Preferentially, CAMD should evaluate all candidate molecules directly based on their intended application (Gertig et al., 2020c). In the context of process design, the designed molecules are applied in processes e.g., as working fluids, solvents or catalysts (Papadopoulos et al., 2018). Thus, each candidate molecule should be evaluated using process optimizations to determine the process performance that can be reached using the candidate. The integration of process optimizations into the design procedure corresponds to the extension of CAMD to integrated Computer-Aided Molecular and Process Design (CAMPD) (Papadopoulos et al., 2018).

CAMPD methods have already been used extensively for integrated in silico design of molecules and processes, e.g., working fluids and Organic Rankine Cycles (Schilling et al., 2017; 2020; Linke et al., 2015) or extraction solvents and processes (Austin et al., 2017; Papadopoulos and Linke, 2009; Scheffczyk et al., 2018). A very comprehensive review of CAM(P)D applications was recently given by Papadopoulos et al. (2018). For reactive chemical processes, CAMD methods have been used to design reaction solvents that accelerate reaction kinetics (Gertig et al., 2019a; Struebing et al., 2013; 2017; Liu et al., 2019a; 2019b). Moreover, CAMPD methods have been developed for integrated design of reaction solvents and processes (Zhou et al., 2015; Gertig et al., 2020b; Zhang et al., 2020).

The main difference of CAMPD methods for non-reactive and for reactive processes lies in the building block property prediction: CAMPD methods for non-reactive processes are commonly based on GC methods for property prediction (Papadopoulos et al., 2018). In contrast, the prediction of reaction kinetics usually requires quantum chemistry. Quantum chemical methods in conjunction with thermochemistry and transition state theory (TST) (Vereecken et al., 2015; Eyring, 1935) have proven to be suited to predict reaction kinetics in CAM(P)D of reaction solvents and processes (Gertig et al., 2020c).

CAMPD methods for the integrated design of molecules and non-reactive processes have gained a high level of maturity and several CAM(P)D methods have also been developed for the design of reaction solvents. The use of computational methods has as well gained importance in the search for new catalysts during the past years as shown by several recent review articles. The review by Ahn et al. (2019) introduces organic and metalorganic catalysis as well as strategies for the use of computational methods in the search for new catalysts. Foscato and Jensen (2020) present an elaborated review of computational methods for the development of homogeneous catalysts including large-scale in silico screenings. Freeze et al. (2019) provide an extensive review of catalyst development strategies that especially includes heterogeneous catalysts. These reviews show that computational methods are nowadays used extensively to shed light on catalytic mechanisms, to explain experimental findings or to evaluate new catalysts before going to experiments. Moreover, various systematic approaches for the use of in silico methods in catalyst development have been proposed. Nevertheless, designing molecular catalysts in silico is still regarded as one of the “holy grails in chemistry” (Poree and Schoenebeck, 2017) and only few approaches to automated CAMD of catalyst molecules have been shown so far.

For in silico studies of catalysis, some authors have proposed to employ an abstract catalytic environment. In an early approach called “Theozymes”, Tantillo et al. (1998) determine the transition state (TS) of chemical reactions using quantum chemical methods. Subsequently, functional groups are chosen to represent the catalyst and placed around the transition state structure. The spatial positions of these functional groups are optimized in order to stabilize the TS. The stabilization of the TS reduces the activation barrier of the chemical reaction and accelerates reaction kinetics. Thus, the best possible catalytic activity is determined for the chosen functional groups. Recent studies (Hare et al., 2017) are still based on the original work of Tantillo et al. In a recent approach, Dittner and Hartke (2018, 2020) employ an abstract environment that is optimized in order to maximize the catalytic effect of electrostatic interactions. Approaches based on abstract representation of catalysts are well suited to study catalytic effects and provide a theoretical optimum of catalytic activity. However, no real catalyst structures are designed directly.

Few approaches have been proposed in literature to design catalyst molecular structures directly. Lin et al. (2005) design transition metal catalysts based on selected functional groups. To optimize the molecular structure of the catalysts, a suitable tabu search algorithm (Chavali et al., 2004) is employed. The properties of designed catalysts are predicted with Quantitative Structure-Property Relationships (QSPR) that are fitted to experimental data. Targets for properties such as density and toxicity are used as design objectives. Consequently, the method directly suggests catalyst structures, but does not optimize these structures based on the achieved catalytic performance. Chu et al. (2012) employ a quantitative structure-activity relationship (QSAR) model in catalyst design. This QSAR model relates the catalytic activity of Ruthenium catalyst complexes for olefin metathesis to descriptors such as bond distances, angles and partial charges (Occhipinti et al., 2006). New catalyst complexes are constructed from a fixed metal core, a list of ligand scaffolds and a variety of molecule fragments that can be attached to the ligand scaffolds. An evolutionary algorithm optimizes the structure of the catalyst. Krausbeck et al. (2017) propose a design method called “Molecular Scaffold Design” based on the idea that unstable, distorted structures occur during chemical reactions and that catalysts need to stabilize such structures to enhance reaction kinetics. The design starts with an unstable fragment with frozen geometry that includes distorted reactants, the core of the catalyst and additional binding sites. A list of atoms that may be added at binding sites is specified. Sub-

sequently, a set of candidates is generated by enumerating the different combinations of atoms that can be added at the binding sites and saturating the resulting structures with hydrogen atoms. These additional hydrogen atoms are replaced by binding sites in the subsequent iteration. The structures are scored with a measure of forces on the nuclei of the unstable fragment computed with quantum mechanical density functional theory (DFT). In an iterative procedure, new "onion shells" (Krausbeck et al., 2017) of atoms are constructed around promising candidates from previous steps until a structure is obtained where the forces on the nuclei in the unstable fragment vanish. More recently, Chang et al. (2018) designed Ni catalyst complexes for a catalytic CO/CO_2 conversion. In their design approach, selected groups of the ligands of the Ni complexes are optimized with the objective to minimize the activation energy of the rate-limiting reaction step. During the design procedure, the activation energies are predicted using the tight binding linear combination of atomic potentials (TB-LCAP) (Xiao et al., 2008) method. Promising candidates from the design are subsequently investigated in more detail using DFT.

The design approaches discussed above can be regarded as pioneering work towards in silico design of molecular catalysts. However, two important building blocks of a reliable, direct in silico design of molecular catalyst structures have not been completed, yet. First, a reliable prediction method is needed for the catalytic performance of candidate catalysts, i.e., the acceleration of the reaction kinetics by the designed catalysts. To be reliable, the prediction should employ high-level QC methods already during the design procedure. Second, the ultimate objective of chemical process design is maximum process performance rather than the acceleration of the chemical reactions. Thus, a process-based evaluation of each candidate catalyst is desired. For this purpose, process optimizations have to be integrated into the in silico design of molecular catalysts. In this work, we propose a CAMPD method called CAT-COSMO-CAMPD that integrates the discussed building blocks into the in silico design of molecular catalysts. The prediction of catalytic effects is based on TST and advanced QC methods such as DLPNO-CCSD(T) (Riplinger and Neese, 2013) and COSMO-RS (Klamt et al., 2010). This prediction is broadly applicable and not limited to certain functional groups. Thereby, large and diverse chemical design spaces can be explored. Optimal catalyst structures and process conditions are determined by a hybrid optimization scheme: The genetic optimization algorithm LEA3D (Douguet et al., 2005) generates and optimizes catalyst structures based on a library of 3D molecule fragments. Deterministic process optimizations maximize the performance of processes for each molecular catalyst considered during the design procedure. Thus, the desired process-based evaluation is ensured for all candidate catalysts. Currently, CAT-COSMO-CAMPD is applicable to the integrated design of homogeneous molecular catalysts and chemical processes involving gaseous and liquid phases. Potential extensions to other classes of catalysts are discussed in Section 4.

The proposed CAT-COSMO-CAMPD method is explained in detail in Section 2 of this article. Next, the application of CAT-COSMO-CAMPD to the case study of a catalytic carbamate-cleavage process is presented (Section 3). Subsequently, current limitations and future prospects of CAT-COSMO-CAMPD are discussed and conclusions are drawn (Section 4).

2. CAT-COSMO-CAMPD for integrated catalyst and process design

The integrated catalyst and process design problem is formulated as optimization problem specifying the generic CAMD problem discussed by Gani (2004):

$$
\begin{aligned}
\max_{x,y} \quad & f(x,\Theta,k) && \text{process-based objective} \\
\text{s.t.} \quad & k = h_1(x,\Theta) && \text{kinetic model} \\
& \Theta = h_2(x,y) && \text{thermodynamic property model} \\
& 0 = h_3(x,\Theta,k) && \text{process model} \\
& g_1(y) = 0 && \text{chemical feasibility} \\
& g_2(y) \le 0 && \text{chemical feasibility} \\
& c_1(\Theta) \le 0 && \text{constraints on thermodynamic properties} \\
& c_2(y) \le 0 && \text{constraints on molecular properties} \\
& c_3(x,\Theta,k) \le 0 && \text{constraints on the process} \\
& x \in X && \text{variable process conditions} \\
& y \in Y && \text{molecular structure of catalyst}
\end{aligned}
\tag{1}
$$

In Problem (1), $f(x,\Theta,k)$ represents the process-based objective (e.g., conversion or yield) that may depend on the variable process conditions x, on thermodynamic equilibrium properties Θ and on the reaction kinetics determined by reaction rate constants k. The objective is maximized by optimizing the variable process conditions x and the molecular structure of the catalyst molecule y. Rate constants k themselves also depend on x and Θ and are calculated using a kinetic model $h_1(x,\Theta)$. The thermodynamic equilibrium properties Θ depend on the variable process conditions x as well as on the molecular structure of the catalyst molecule y and are calculated using a thermodynamic property model $h_2(x,y)$. The equations of the process model are represented by $h_3(x,\Theta,k)$. Equality constraints $g_1(y)$ and inequality constraints $g_2(y)$ ensure chemical feasibility of the catalyst molecules, e.g., correct valency of all atoms in the molecule. Additionally, constraints on thermodynamic properties $c_1(\Theta)$ (e.g., minimal boiling point of the catalyst molecule), on molecular properties $c_2(y)$ (e.g., number of atoms in the molecule or restrictions on the combination of functional groups) and on the process $c_3(x,\Theta,k)$ may be used. The variable process conditions x contained in a range of allowed process conditions X represent the process-related degrees of freedom. Besides quantities like pressures and temperatures, these process-related degrees of freedom may e.g., also include vessel sizes or compositions of mixtures fed to the process. The molecular structure of the catalyst molecule y is contained in a chemical design space Y.

It should be noted that the rate constants $k = h_1(x,\Theta)$ in Problem (1) do not directly depend on the molecular structure of the catalyst molecule y. However, this missing direct dependence does not mean that catalysts do not impact the rate constants but rather reflects the way rate constants are calculated. Catalyst molecules influence rate constants by reducing so-called activation barriers that reactions need to overcome to take place. We regard these activation barriers as quantities associated with thermodynamic pseudo-equilibria and therefore include them in the thermodynamic equilibrium properties Θ. The calculation of reaction rate constants is explained in more detail in Section 2.1. Subsequently, the prediction of thermodynamic equilibrium properties (Section 2.2) and process modeling (Section 2.3) are described, before the solution approach to the optimization Problem (1) is presented (Section 2.4).

2.1. Quantum chemistry-based prediction of reaction kinetics

The kinetics of catalytic reactions are described by reaction rate constants k that indirectly depend on the structure of the used catalyst molecule y as explained above. The methods we use to predict rate constants were described in detail in earlier work

(Gertig et al., 2019, 2021; Kröger et al., 2017). An overview is given in the following without detailed derivations of all equations.

The rate constants k of elementary reactions are calculated based on conventional transition state theory (TST) (Vereecken et al., 2015) and the so-called Eyring Equation (Eyring, 1935):

$$k = \frac{k_B T}{h} (V_m)^{(n-1)} \exp\left(-\frac{\Delta G^{\ddagger}}{RT}\right). \tag{2}$$

In Eq. (2), k_B is the Boltzmann constant, T is the temperature at which the reaction takes place, h is Planck's constant and R is the gas constant. The molar volume of the reaction phase is represented by V_m and n denotes the reaction order defined as the sum of the stoichiometric coefficients of all reactants. The activation barrier $G^{\ddagger}$ is the difference in molar Gibbs free energy between the state of the reactants and a so-called transition state (TS). According to conventional TST, this transition state is a first-order saddle point in energy that is passed along the reaction path and can be determined using quantum chemical methods (Foresman and Frisch, 2015). The TS is an unstable state in pseudo-equilibrium with the reactant state.[1]

In case the rate constant k is calculated for a reaction taking place in a liquid phase, the activation barrier $G^{\ddagger}$ is split in different contributions:

$$\begin{aligned} k = \frac{k_B T}{h}\left(V_m^{\text{i.G.}}\right) & \quad {}^{(n-1)} \exp\left(-\frac{\Delta G_{\ddagger}^{\text{i.G.}}}{RT}\right) \\ & * \frac{\prod_i \gamma_i^{\text{uN}}}{\gamma_{\ddagger}^{\text{uN}}} \exp\left(-\frac{\Delta \tilde{G}_{\ddagger}^{\text{solv}} - \sum_i \Delta \tilde{G}_i^{\text{solv}}}{RT}\right) \\ = k^{\text{i.G.}} \frac{\prod_i \gamma_i^{\text{uN}}}{\gamma_{\ddagger}^{\text{uN}}} & \quad \exp\left(-\frac{\Delta \tilde{G}_{\ddagger}^{\text{solv}} - \sum_i \Delta \tilde{G}_i^{\text{solv}}}{RT}\right). \end{aligned} \tag{3}$$

This splitting into different contributions offers the advantage that the most appropriate methods can be chosen for the calculation of each contribution (Deglmann et al., 2009; Peters et al., 2008; Hellweg and Eckert, 2017; Coote, 2009). In Eq. (3), $V_m^{\text{i.G.}}$ is the molar volume of the reaction phase in the used ideal gas reference state and may be calculated using the ideal gas law. The activation barrier in the ideal gas reference state is denoted by $\Delta G_{\ddagger}^{\text{i.G.}}$. The first 3 terms shown in the upper line of Eq. (3) determine a reaction rate constant $k^{\text{i.G.}}$ in the ideal gas reference state:

$$k^{\text{i.G.}} = \frac{k_B T}{h}\left(V_m^{\text{i.G.}}\right)^{(n-1)} \exp\left(-\frac{\Delta G_{\ddagger}^{\text{i.G.}}}{RT}\right). \tag{4}$$

Solvation effects represent the non-ideal effects that the environment of the reacting species has on the rate constant k. These solvation effects are accounted for by two terms in Eq. (3): First, the ratio of the product of the unsymmetrically normalized activity coefficients γ_i^{uN} of all reactants i to the unsymmetrically normalized activity coefficient $\gamma_{\ddagger}^{\text{uN}}$ of the transition state. Second, an exponential term containing the different Gibbs free energies of solvation $\tilde{G}^{\text{solv}}$ of the reactants i and the transition state ‡, respectively. The exponential term accounts for the difference between the ideal gas reference state and a liquid reference state. The term with the unsymmetrically normalized activity coefficients γ^{uN} in turn accounts for the difference between the liquid reference state and the actual reaction mixture composition. This latter term is sometimes neglected, which corresponds to approximating the rate constant k by the rate constant at infinite dilution. The Gibbs free energies of solvation $\tilde{G}^{\text{solv}}$ are computed based on molar reference states. A suitable choice of the ideal gas reference state is the reaction temperature T and a reference pressure of $p^0 = 1$ bar ≈ 1 atm. Details about the use of reference states as well as a derivation of Eq. (3) can be found in our earlier work (Gertig et al., 2019a; 2020b).

To determine all quantities necessary to evaluate Eq. (3), the following computation scheme is applied:

1. Optimized geometries of reactants, catalysts and transition states are determined using the quantum-mechanical density functional theory (DFT) method B3LYP (Stephens et al., 1994) with empirical dispersion correction (Grimme et al., 2010) (B3LYP-D3) and TZVP basis set. Vibrational analysis is carried out subsequently. The B3LYP method is known for good accuracy in geometry optimization and frequency analysis in spite of the rather moderate computational resources required (Zheng et al., 2009; Gottschalk et al., 2018). The rigid rotor harmonic oscillator (RRHO) (Atkins and Friedman, 2011) model is used in the frequency analysis. The software Gaussian 09 (Frisch et al., 2013) is employed for both geometry optimization and frequency analysis.
2. To obtain accurate electronic energies, single point (SP) calculations are performed with the post-Hartree-Fock method DLPNO-CCSD(T) (Riplinger and Neese, 2013; Riplinger et al., 2013) with aug-cc-pVTZ basis set and TightPNO settings. The software ORCA (Neese, 2018) is used for these SP calculations.
3. Activation barriers $\Delta G^{\ddagger,\text{i.G.}}$ in the ideal gas reference state are determined by thermochemical calculations with GoodVibes (Funes-Ardoiz and Paton, 2018) based on the RRHO approximation. As the RRHO approximation can cause significant errors in calculated entropies in case of low frequencies, Grimme's quasi-harmonic treatment (Grimme, 2012) is employed to reduce these errors.
4. Next, the reaction rate constants $k^{\text{i.G.}}$ in the ideal gas reference state are computed using Eq. (4).
5. Optionally, tunneling corrections to $k^{\text{i.G.}}$ may be computed in case significant impact of tunneling on the reaction rate is expected. Tunneling corrections based on Eckart (1930) computed with the TAMkin package (Ghysels et al., 2010) have been found to be a reasonable choice.
6. The advanced solvation model COSMO-RS (Klamt et al., 2010; Klamt, 1995; Klamt et al., 1998) is used to calculate Gibbs free energies of solvation $\tilde{G}^{\text{solv}}$ of reactants, catalysts and transition states. The software turbomole (Ahlrichs et al., 1989; TURBOMOLE, 2015) is employed for COSMO (Klamt and Schüürmann, 1993) calculations with BP86 (Becke, 1988; Perdew, 1986a; 1986b) and def2-TZVP basis set. For this purpose, geometries are optimized with BP86/def2-TZVP in vacuum and the actual COSMO calculations are performed as SP calculations. The COSMO-RS calculations (COSMOtherm; Eckert and Klamt, 2002) to obtain the $\tilde{G}^{\text{solv}}$ values are performed subsequently.
7. To calculate the required unsymmetrically normalized activity coefficients γ^{uN}, we either employ COSMO-RS directly or a suitable surrogate model fitted to data from COSMO-RS.
8. Finally, reaction rate constants k in liquid phase are calculated using Eq. (3).

As discussed in our previous work (Gertig et al., 2021), it may be important to search for conformers of reacting species and transition states to obtain accurate reaction rate constants k. We perform this conformer search using rotor scans (Foresman and Frisch, 2015). For the integrated catalyst and process design with CAT-COSMO-CAMPD, we assume that it is sufficient to find the conformer with the lowest energy for all species by means of such scans. This assumption is considered a compromise between prediction accuracy and the effort and complexity of the computations performed during the integrated design. The rotor scans to search the most stable conformers are performed for reactants, products and solvents in advance of the actual design. As catalysts and tran-

[1] The term "pseudo-equilibrium" implies that the reactants are not in equilibrium with the products at the same time, although the back-reaction might proceed via the same transition state.

sition states change during the design, selected automated rotor scans are performed as explained in Section 2.4.

The expected uncertainty in the rate constants k computed with the methods described above was discussed in detail in previous work (Gertig et al., 2019, 2021; Kröger et al., 2017) and is thus only mentioned briefly here. Generally, we expect that the predicted rate constants k should agree with experimentally determined rate constants k^{exp} within one order of magnitude at a temperature of 25°C. The uncertainty is expected to decrease with increasing temperature. Furthermore, some errors cancel if the prediction is used for comparison of different candidate catalysts, which is advantageous in CAM(P)D where absolute values are less important than rankings and trends.

2.2. Prediction of thermodynamic equilibrium properties

All thermodynamic equilibrium properties Θ required to solve Problem (1) are computed based on the COSMO-RS (Klamt et al., 2010) model. These quantities typically include Gibbs free energies of solvation $\widetilde{G}^{\text{solv}}$, pure component vapor pressures p_i^S as well as activity coefficients γ and unsymmetrically normalized activity coefficients γ^{uN}.

We compute the Gibbs free energies of solvation $\widetilde{G}^{\text{solv}}$ at all temperatures of interest directly using COSMO-RS. This is practical because the $\widetilde{G}^{\text{solv}}$ are calculated for defined reference states and thus do not change with changing reaction mixture composition during process simulations and optimizations.

Pure component vapor pressures p_i^S are calculated using the Antoine equation (Pfennig, 2004). The required Antoine parameters for each component are computed with COSMO-RS.

In contrast to $\widetilde{G}^{\text{solv}}$ and p_i^S, the activity coefficients γ and unsymmetrically normalized activity coefficients γ^{uN} depend on the reaction mixture composition. Therefore, these quantities have to be re-evaluated frequently in case mixture compositions change. These re-evaluations would consume too much time if COSMO-RS was used directly during process simulations and optimizations. This time-consumption is not only caused by the solution of the COSMO-RS equations themselves, but also by required software interfacing. Thus, the NRTL (Renon and Prausnitz, 1968) activity coefficient model is used as surrogate model. NRTL parameters are automatically fitted to activity coefficient data generated for every system under consideration using COSMO-RS.

2.3. Process modeling

Process models are formulated based on balance equations and equations accounting for phase equilibria. Moreover, power laws (Levenspiel, 1999) are employed in conjunction with the predicted reaction rate constants k to describe the rates of elementary reactions. The predicted property data allows to formulate a wide range of process models. In the case study presented below, we consider semi-batch operation. In this case, the resulting set of equations is a differential-algebraic system of equations (DAE). The DAE system used in the case study is discussed in the supporting information. The process models are implemented in MATLAB (2019) and solved with the *ode15s* solver.

2.4. Solution approach of CAT-COSMO-CAMPD

In the following, the solution approach of the proposed CAT-COSMO-CAMPD method to the integrated catalyst and process design Problem (1) is explained. The solution approach of CAT-COSMO-CAMPD follows our quantum chemistry-based design methods for solvents (Scheffczyk et al., 2018; Gertig et al., 2019a; 2020b; Scheffczyk et al., 2017; Fleitmann et al., 2018). A preliminary and brief presentation of CAT-COSMO-CAMPD was already given at the conference corresponding to this special issue (Gertig et al., 2020a). CAT-COSMO-CAMPD employs a hybrid optimization scheme. The genetic optimization algorithm LEA3D (Douguet et al., 2005) is used to identify the structure y^* of the optimal catalyst molecule, whereas gradient-based process optimization is used to determine optimal values of the variable process conditions x^* for each considered catalyst.

The LEA3D algorithm generates 3D molecular structures based on a pre-defined library of 3D molecule fragments. Fragments are combined randomly in the initial step of the optimization to obtain a first generation of candidate molecules. Based on the performance of these initial candidates, LEA3D uses genetic operations to alter the structures and to suggest the next generation of candidates. New generations are iteratively suggested until a specified maximum number of generations is reached. By this procedure, the space of candidate molecules is systematically explored to determine the optimal structure y^*. Other convergence criteria suitable for genetic algorithms could be employed (Safe et al., 2004). Still, the used genetic algorithm is stochastic such that convergence to the global optimum cannot be guaranteed.

Working with 3D molecular structures in the optimization of catalyst molecules offers the major advantage that quantum chemistry-based property prediction can be employed in a straightforward way: QC methods generally require 3D starting geometries as input. QC-based property prediction offers several advantages (see Section 1: Quantum chemical methods and thermochemistry consistently predict a broad range of molecular, thermodynamic and chemical properties including reaction rate constants. In contrast to e.g., group-contribution methods, QC methods are not limited to previously fitted functional groups. Thus, a wide variety of 3D molecule fragments may be chosen when setting up the chemical design space.

Currently, CAT-COSMO-CAMPD requires the user to specify a so-called scaffold fragment among the other 3D fragments. This scaffold fragment contains the reactants as well as the catalytically active group contained in all catalysts designed in one CAT-COSMO-CAMPD run. Thus, during one design run, the mechanism of catalysis does not change. The scaffold fragment is employed to construct starting geometries for the search of transition states of the catalytic reactions with the designed catalysts. Further information about scaffold fragments is given in the subsequent description of the design procedure and an example is discussed in Section 3.1.

The complete CAT-COSMO-CAMPD procedure to solve the integrated design Problem (1) is shown in the flowchart in Fig. 1 and explained in the following:

1. First, several specifications have to be made:
   - The reaction network and the process under consideration have to be specified.
   - The catalytically active group is chosen. The designed catalysts are based on this group.
2. Quantum chemical and thermochemical calculations are performed for reactants and products in the specified reaction network as well as for solvents used in the process. These QC and thermochemical calculations are discussed in Section 2.1 (Steps 1–3 and 6). Moreover, a so-called reference system is defined (Fig. 1) that corresponds to a reaction system including a typical molecule with the catalytically active group as catalyst. QC calculations are employed to determine the transition state geometry for the reference system. From this TS geometry, the scaffold fragment is constructed (for an example, see Fig. 4). As already described above, the scaffold fragment contains the reactants and the catalytically active group of the catalyst. All atoms of the reactants and the catalytically active group are positioned in space as in the determined TS geometry of the

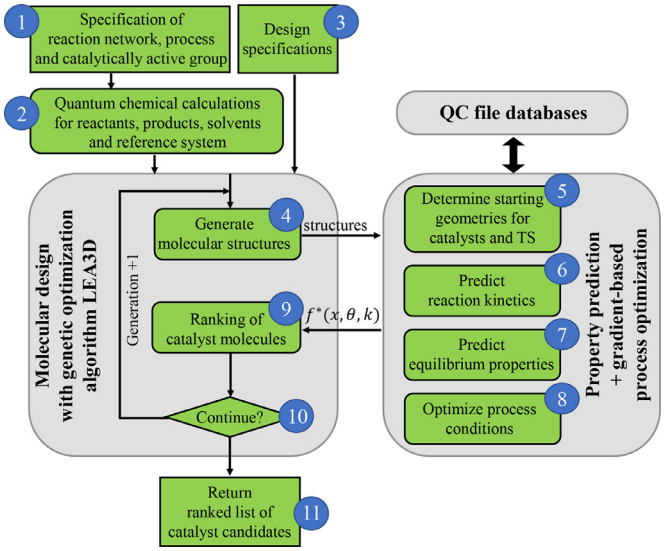

Fig. 1. Flowchart of the CAT-COSMO-CAMPD method for solution of the integrated catalyst and process design Problem (1). Used abbreviations: QC (Quantum Chemical), TS (Transition State).

reference system. The scaffold fragment is used in Step 3 when the chemical design space Y is defined.

3. Design specifications have to be made for the integrated catalyst and process design:
   - A fragment library is provided that contains various 3D molecule fragments including the catalyst scaffold fragment. From these 3D fragments, catalyst structures are constructed. The choice of fragments in the library determines the chemical design space Y for CAMD of catalysts. It is important to note that the resulting design space does not correspond to a set of possible catalysts, but to a set of possible transition states of the catalytic reaction under consideration. The reason is that the scaffold fragment contains not only the catalytically active group of the catalyst, but also the reactants. Consequently, the transition states are designed directly instead of the catalyst molecules. The direct design of the transition states is advantageous because starting geometries for the optimization of the transition state structures have to be provided. Obtaining suitable starting geometries of transition states can be considered the critical aspect of the automated prediction of rate constants k of catalytic reactions.
   - The user may choose rotor scans and additional pre-optimizations performed in Step 5.
   - Settings are required for the genetic algorithm LEA3D used for the optimization of catalyst molecular structures. These settings include the maximum number of generations, the number of candidate catalyst molecules per generation as well as the probabilities for genetic operations such as mutation and cross-over of candidates.
   - The process model $h_3(x, \Theta, k)$ is specified.
   - The process-based objective $f(x, \Theta, k)$ of the design is chosen.
   - The values of constant process parameters are assigned and process degrees of freedom x are selected.
   - The allowed range of operating conditions X is defined.
   - Constraints are set including constraints $c_1(\Theta)$ on thermodynamic properties, $c_2(y)$ on molecular properties and $c_3(x, \Theta, k)$ on the process.

   LEA3D inherently respects the constraints $g_1(y)$ and $g_2(y)$ to ensure chemical feasibility of designed molecules. Moreover, LEA3D ensures that each candidate catalyst molecule contains one scaffold fragment and thus one catalytically active group.
4. LEA3D suggests a generation of 3D molecular structures of transition states with candidate catalysts. The initial generation is created by random combination of the scaffold fragment with further molecule fragments from the fragment library. Candidates of subsequent generations are obtained from genetic operations such as mutation and cross-over that are applied to promising candidates of the preceding generation. The generated structures are passed to the automated property prediction and gradient-based process optimization.
5. Suitable starting geometries for the optimization of catalysts and transition states are determined. It is important to note that by the term "geometry" of a molecule or TS, we here understand the set of spatial positions of all atoms comprising the molecule or TS. In contrast, "molecular structures" include the connectivity of atoms as shown in structural formulas of molecules. In principle, 3D geometries of transition states are already provided by LEA3D in Step 4. However, it happens occasionally that these structures are not accurate enough for the subsequently used QC methods to work properly. Thus, a first pre-optimization improves the TS geometries employing the force field method MMFF94 (Halgren, 1996a; 1996b; 1996c) available in the chemical toolbox Open Babel (O'Boyle et al., 2011; Open Babel). As this first pre-optimization is not suited to handle transition states, the scaffold fragment is substituted by a dummy atom. Afterwards, the dummy is re-substituted by the scaffold fragment using translation and rotation operations in 3D cartesian coordinate space. Next, further pre-optimization is performed using Gaussian 09. Selected rotor scans are performed using the QC method AM1 (Dewar et al., 1985) to ensure that the minimum energy conformer of each TS is identified. The selection of rotor scans to perform is made by the user (Step 3). Optionally, a geometry optimization with the DFT method B3LYP that minimizes energy may follow the rotor scans. During rotor scans and energy minimization with Gaussian 09, it has to be ensured that no atomic distances are changed that correspond to bonds that break or form during the considered reaction. For this purpose, the "modredundant" option (Foresman and Frisch, 2015) is employed. The TS geometries obtained from the described pre-optimizations are used as starting geometries for subsequent QC calculations. Starting geometries of the catalyst molecules are extracted as a subset of the TS starting geometries.
6. To predict reaction kinetics of the catalytic reactions, reaction rate constants k are computed as explained in Section 2.1. Calculations for reactants, products and solvents performed in Step 2 are not repeated here.
7. Required thermodynamic equilibrium properties Θ are predicted as described in Section 2.2.
8. Gradient-based process optimizations are performed for processes with each candidate catalyst. For these processes, the process optimizations determine the optimal process performance $f^*(x^*, \Theta, k)$ as well as the corresponding optimal values of variable process conditions x^*. The interior point algorithm available in the built-in function *fmincon* in MATLAB is employed for process optimizations. The evaluations of the process model $h_3(x, \Theta, k)$ required during process optimization are performed with the solver *ode15s* as already mentioned in Section 2.3.

As the quantum chemical calculations performed in Steps 6 and 7 may be computationally demanding, the output files of these calculations are stored in dedicated QC file databases. If structures are suggested in Step 4 that were already considered previously, the demanding QC calculations can be skipped. It is important to

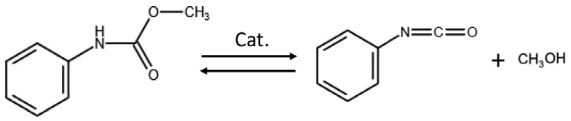

Fig. 2. Cleavage reaction of methyl phenyl carbamate (MPC) to phenyl isocyanate and methanol.

note here that the prediction of reaction kinetics (Step 6) already requires some thermodynamic equilibrium properties Θ. Moreover, certain equilibrium properties need to be re-evaluated when conditions such as reaction mixture composition change during dynamic processes or when variable process conditions x are changed during process optimizations (Step 8). Thus, Steps 6 to 8 are not performed strictly in the sequence displayed in Section 2.4.

9. The candidate catalysts of the current generation are ranked based on the determined values of the optimal process performance $f^*(x^*, \Theta, k)$.
10. A new generation of candidate structures is generated by LEA3D based on the previous generation using genetic operations (Step 4). The probabilities that certain candidates of the previous generation are selected as parents for candidates of the new generation are related to the process performance $f^*(x^*, \Theta, k)$. Steps 4 to 10 are repeated until the maximum number of generations specified in Step 3 is reached.
11. As output of the integrated catalyst and process design with CAT-COSMO-CAMPD, a ranked list of catalyst structures y including the corresponding values of $f^*(x^*, \Theta, k)$ and x^* is assembled.

Obtaining a ranked list instead of a single optimal catalyst as output of the design offers an important advantage: The user may choose candidates for experimental testing among several near-optimal candidates from the design. Thus, further criteria such as ease of synthesis, commercial availability or toxicity that were not considered in the design may be taken into account in the final choice.

3. Case study: integrated catalyst and carbamate-cleavage process design

To demonstrate the application of CAT-COSMO-CAMPD, an integrated catalyst and process design is performed for a catalytic carbamate-cleavage process (Wang et al., 2017) of methyl phenyl carbamate (MPC) to phenyl isocyanate and methanol (Fig. 2). Carbamate-cleavage reactions represent challenging steps in possible production routes to industrially important isocyanates (Six and Richter, 2003). One such production route that aims at CO_2-based isocyanate production (Kaiser et al., 2018) has been investigated in the research project (Carbon2Chem). The design of carbamate-cleavage processes is challenging because the cleavage reactions are strongly endothermic. Moreover, reaction equilibria strongly favor carbamate formation and are thus very unfavorable for carbamate-cleavage processes (Leitner et al., 2018). Typically, reaction temperatures of $T^R > 200°C$ are required to drive the reaction. To avoid fast back-reactions, continuous removal of the alcohol produced as by-product is required during carbamate-cleavage. In case of volatile alcohols such as methanol, this continuous removal can be ensured using stripping with inert nitrogen gas (Cao et al., 2015). A suitable process flowsheet already introduced in previous studies (Gertig et al., 2020b; 2020a; 2019b) is shown in Fig. 3. A semi-batch reactor is employed to carry out the catalytic cleavage reaction that takes place in the liquid phase using di-phenyl ether as solvent. Nitrogen is used for stripping to carry the formed volatile methanol out of the reactor. A flash is used

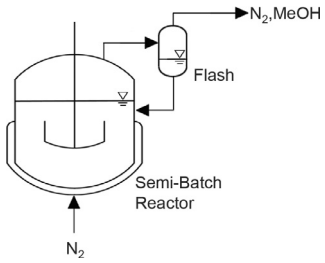

Fig. 3. Process flowsheet of the considered semi-batch catalytic carbamate-cleavage process.

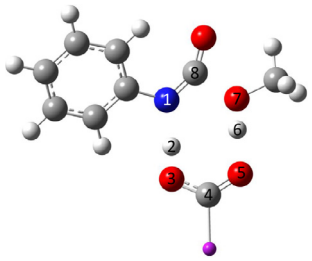

Fig. 4. Scaffold fragment used for the integrated catalyst and carbamate-cleavage process design with CAT-COSMO-CAMPD. The scaffold fragment contains the reactants and the catalytically active group. The violet dot marks the site where the scaffold fragment is connected to other 3D fragments in the design. This figure was generated using GaussView (Frisch et al., 2013).

to condense and recycle unintentionally removed isocyanate, carbamate, solvent and catalyst. Ideally, only nitrogen and methanol leave the semi-batch process with the top product stream of the flash. A process model $h_3(x, \Theta, k)$ for this process was developed as explained in Section 2.3 and is discussed in more detail in the supporting information. Required rate constants k of the catalytic cleavage reaction as well as thermodynamic equilibrium properties are computed as described in Sections 2.1 and 2.2. It was shown in previous work (Gertig et al., 2021) that the employed methods are suited to carbamate-cleavage reactions. Due to the challenging nature of carbamate-cleavage, it is expected to be difficult to obtain satisfying isocyanate yields in the considered semi-batch cleavage process, even when using a catalyst. Thus, the isocyanate yield is a good choice for the objective of the integrated catalyst and carbamate-cleavage process design. The specifications of the according design with CAT-COSMO-CAMPD are given in the following section.

3.1. Specifications for the integrated catalyst and carbamate-cleavage process design

The reaction and process under consideration are specified as described above. As catalytically active group, the carboxyl group is chosen that is known to have catalytic properties (Satchell and Satchell, 1975). The constructed scaffold fragment is shown in Fig. 4. As can be seen, the scaffold fragment contains the reactant MPC in a partially cleaved state as well as the carboxyl group of the catalyst that is designed. The carboxyl group acts as both proton acceptor and proton donor in a concerted reaction: The proton originally bound to the nitrogen of the carbamate (atom 2 in Fig. 4) is accepted and simultaneously, the proton initially contained in the carboxyl group (atom 6 in Fig. 4) is donated to the methoxy group of the carbamate to form the by-product methanol. The violet dot marks the anchor point where the scaffold fragment is connected to other 3D molecule fragments in the catalyst design.

No significant influence of tunneling is expected for the catalytic carbamate-cleavage. Thus, no tunneling corrections are computed. One rotor scan is performed for each candidate around

the bond between the carboxyl group and the rest of the catalyst (atom 4 and violet dot in Fig. 4; see step 5 in Section 2.4).

The yield used as objective $f(x, \Theta, k)$ of the integrated design is defined as the final moles of isocyanate present in the reactor divided by the initial moles of carbamate provided. This objective is maximized solving the integrated design Problem (1) in order to identify the optimal catalyst structure y^* and corresponding optimal values of variable process conditions x^*. The temperature T^F in the flash and the volume flow $\dot{V}_{N_2}$ of nitrogen fed to the reactor are chosen as variable process conditions for process optimizations. The reaction temperature T^R in the reactor is fixed to 473.15 K. This reaction temperature lies in the typical range of reaction conditions for carbamate-cleavage (Gertig et al., 2021; Wang et al., 2017). The reaction temperature is lower compared to our previous study of auto-catalytic carbamate-cleavage (Gertig et al., 2020b) to take into account that the designed catalysts accelerate the cleavage reaction. Higher temperatures are expected to accelerate the cleavage reaction, but may also lead to undesired side reactions. The pressure of the process is set to $p^{set} = 4\,\text{bar}$. The semi-batch process is allowed to run for 12 h. A reactor volume of $V^R = 1\,\text{m}^3$ as well as initial fractions of 15 mass-% MPC and 5 mass-% catalyst in the reaction mixture are chosen.

The constraints $g_1(y)$ and $g_2(y)$ (see Problem (1)) to ensure chemical feasibility of designed catalyst molecules are inherently respected by the LEA3D algorithm as already mentioned in Section 2.4. Moreover, constraints $c_2(y)$ ensure that each designed catalyst contains exactly one scaffold fragment and limit the number of non-hydrogen atoms in the designed catalysts to a maximum of 13. The operating range X is set to allow flash temperatures of $280\,\text{K} < T^F < 380\,\text{K}$ as well as nitrogen volume flows of $5 \times 10^{-5}\text{m}^3\text{s}^{-1} < \dot{V}_{N_2} < 1.5 \times 10^{-1}\text{m}^3\text{s}^{-1}$. The temperature range used for T^F likely allows for the use of cooling water. The optimal nitrogen volume flow $\dot{V}^*_{N_2}$ is sought between values close to zero and an upper bound that is expected to lie well above favorable values. To define the design space Y for catalyst design, various 3D alkyl, aryl, ether, ester, keto, nitrile, halide, sulfene, sulfide, and imine fragments are provided. The full list of fragments is given in the supporting information. After the initial generation of candidate catalyst molecules designed by random combination of fragments, the integrated catalyst and carbamate-cleavage process design is run for 6 further generations of candidates using a number of 12 candidates per generation.

3.2. Results of the integrated catalyst and carbamate-cleavage process design

The integrated catalyst and carbamate-cleavage process design with CAT-COSMO-CAMPD results in the optimized catalyst molecular structure y^* with SMILES (Weininger, 1988) code ClCOC(C(=O)O)C1CCCCC1. The corresponding 2D structural formula is shown in Fig. 5. The achieved objective function value $f^*(x*, \Theta, k)$ amounts to a yield of 21%. The corresponding optimal values x^* of the variable process conditions are a flash temperature of $T^F = 281.1\,\text{K}$ and a nitrogen volume flow of $\dot{V}_{N_2} = 5.4 \times 10^{-2}\text{m}^3\text{s}^{-1}$. The top catalyst molecule enables a considerably higher process performance compared to common carboxylic acids such as acetic acid (9% yield, also displayed in Fig. 5). In total, 33 candidate catalysts from the design shown in Fig. 5 meet all constraints. The full list of SMILES codes of these catalyst molecules can be found in the supporting information.

The obtained results show that CAT-COSMO-CAMPD successfully identifies molecular structures of catalysts that considerably improve the predicted process performance compared to common reference molecules. However, it is not clear at this point whether the whole integrated catalyst and process design was required to achieve this result or whether it would be sufficient to perform a less complex catalyst design optimizing the rate constant k of the catalytic cleavage reaction. To shed light on this question, a catalyst design that is not integrated with process optimization is presented in the following.

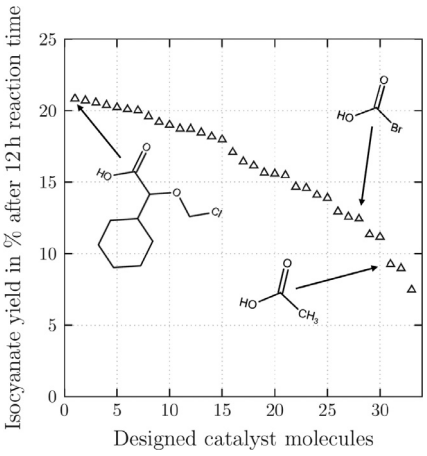

Fig. 5. Optimized yields achieved with 33 catalyst molecules from the design that meet all constraints.

3.3. Specifications for the catalyst design without integrated process optimization

Performing a less complex catalyst design without integrated process optimization corresponds to reducing the Computer-Aided Molecular and Process Design (CAMPD) to a Computer-Aided Molecular Design (CAMD). Thus, Problem (1) (see Section 2) simplifies: The objective function $f(\Theta, k)$ does not depend on any variable process conditions x and no process model h_3 is required any more. Consequently, also the possibility to set process constraints c_3 as well as an operation range X are removed. Except any specifications that are not relevant for the resulting CAMD problem, the catalyst design without integrated process optimization is specified as the integrated design (Section 3.1).

3.4. Results of the catalyst design without integrated process optimization

The catalyst design to maximize the rate constant k of the catalytic carbamate-cleavage reaction results in brominated formic acid with SMILES code OC(=O)Br as optimal catalyst molecule. The 2D structural formula is shown in Fig. 6. The corresponding predicted rate constant at 200°C amounts to $k = 4.43 \times 10^{-6}\ \text{m}^6\text{mol}^{-2}\text{s}^{-1}$. In total, 26 catalysts from the design meet all constraints (Fig. 6). While we suspect brominated formic acid to be unstable at reaction conditions, the results of the design with k as objective demonstrate the problems associated with simple performance measures: Using a catalyst that enables a high rate constant k may still lead to a poor process performance. Brominated formic acid as catalyst leads to the highest predicted rate constant k, but the process using this catalyst achieves only a relatively low yield of 13% after 12 h (see Fig. 5). The underlying reason for the poor performance of simple design objectives is that besides a high reaction rate constant, further aspects are important to reach maximum process performance. For example, the volatility of the catalyst influences catalyst loss due to stripping with nitrogen. Highly polar catalysts increase the overall polarity of the reaction mixture, which has adverse effects on the reaction rate depending on the amount of catalyst used. Moreover, the catalyst molecule influences activities of other substances in the mixture

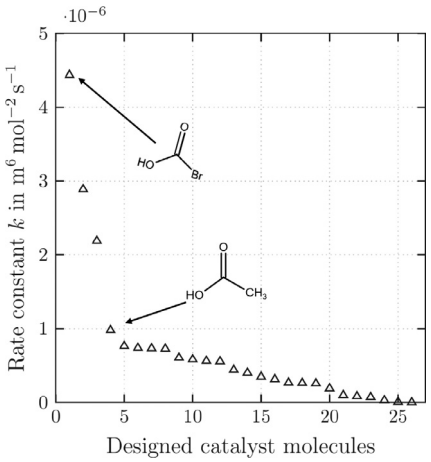

Fig. 6. Optimized predicted rate constants k of the catalytic carbamate-cleavage reaction achieved with 26 catalyst molecules from the design that meet all constraints.

and thus the vapor-liquid equilibria (VLE) in the process. Consequently, for example, the catalyst impacts the activity coefficient of methanol in the flash and can lead to unfavorable methanol recycling. In summary, several criteria need to be accounted for and trade-offs between these criteria have to be made in order to reach optimum process performance. Therefore, an integrated catalyst and process design is required that combines molecular design with process optimization. Still, evaluation criteria used to identify promising catalyst molecules might sometimes even go less far as to calculate reaction rate constants. For many reactions, simpler, heuristic criteria could be thought of. In the subsequent section, we discuss why such criteria may not only fail to identify catalysts enabling optimal process performance, but even fail to find catalysts that enable optimal rate constants.

3.5. Discussion of heuristic criteria for catalyst design

In many cases, heuristic criteria for catalyst design can be defined that allow assessing candidates with less effort than computing reaction kinetics or using process-based evaluation such as our design method CAT-COSMO-CAMPD. For the catalytic carbamate-cleavage reaction considered in the present case study, we discuss three possible examples of heuristic evaluation criteria:

- The barrier in electronic energy ΔE^{el} computed in vacuum often represents the largest contribution to the overall activation barrier $\Delta G^{\ddagger}$ in Gibbs free energy that the reaction has to overcome. Thus, a heuristic design objective for catalyst design could be to minimize ΔE^{el} of the reaction.
- The catalyst and the carbamate form a ring for concerted proton transport in the transition state of the catalytic carbamate-cleavage reaction (see Fig. 4, atoms 1–8). Thus, it could be supposed that the proton donor and/or acceptor properties of the carboxyl group determine catalytic activity. These donor/acceptor properties are related e.g., to the partial charges of the oxygen atoms of the carboxyl group of the catalyst. Therefore, the partial charges should be either strong or weak, depending on whether proton acceptance or donation is critical, or a certain value represents the optimal trade-off. Thus, another objective for catalyst design could be to maximize or minimize partial charges of the carboxyl oxygen atoms or to match a certain target value.
- There are assumptions mentioned in literature (Satchell and Satchell, 1975) that a nucleophilic character of the oxygen atom of the reactant's methoxy group stabilizes the transition state of the reaction. Thus, it is supposed that an important function of the catalyst is to increase this nucleophilic character. A high nucleophilicity should be associated with a high negative partial charge of the oxygen atom and should be observable in the transition state of the catalytic reaction.

The 3 criteria discussed above were evaluated for the catalysts designed in the design run presented in Section 3.4. All required quantities were extracted from the output of the DLPNO-CCSD(T) calculations (see Step 2 of the computation scheme explained in Section 2.1). Mulliken charges (Mulliken, 1955) are used to approximate the partial charges of atoms. Fig. 7 plots the criteria versus the logarithm of the reaction rate constant k achieved with the respective catalysts. For reasons discussed in the preceding sections, it cannot be expected that the heuristic criteria reflect the achievable process performance. Still, one could argue that using one of the heuristic criteria should at least result in the design of catalysts that optimize the reaction rate constant k of the catalytic carbamate-cleavage. However, if this was the case, the chosen criterion should correlate well with k.

As can be seen in the upper part of Fig. 7, there is a general trend of increasing rate constants k with decreasing barriers in electronic energy ΔE^{el} as expected. However, the correlation of k with ΔE^{el} is clearly not good enough for ΔE^{el} to serve as design objective. In particular, the catalyst leading to the lowest ΔE^{el} enables only a moderate rate constant k. The reason for the insufficient correlation is that important effects on k are not reflected by ΔE^{el}. Although representing a major contribution to the activation barrier $\Delta G^{\ddagger}$ that in turn determines the rate constant, evaluation of catalysts based on ΔE^{el} completely neglects entropic and solvation effects.

The partial charges of the oxygen atoms in the carboxyl group of the catalyst are shown in the middle part of Fig. 7. The term "donor O-atom" hereby refers to the oxygen atom initially bonded to a hydrogen atom. The correlation of these partial charges with k is not satisfactory. Indeed, there is a trend that the absolute values

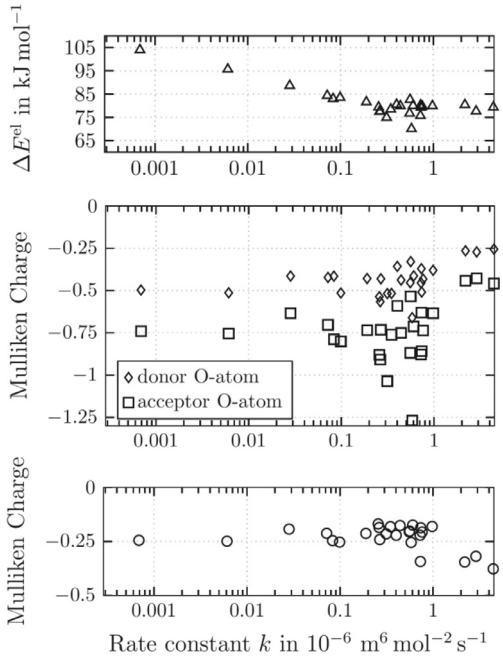

Fig. 7. Heuristic criteria vs. log of the rate constant k. Upper part: barrier in electronic energy ΔE^{el} calculated using DLPNO-CCSD(T) for the carbamate-cleavage with designed catalysts. Middle part: Mulliken charges of the 2 oxygen atoms in the carboxyl groups of designed catalysts. Lower part: Mulliken charge of the oxygen atom in the methoxy group of the carbamate (in TS).

of the charges of the oxygen atoms decrease with increasing rate constant k, which could be related to proton donor and acceptor properties of the catalysts. In fact, most of the highly ranked catalysts contain electron-withdrawing groups like halogens or nitrile groups, which explains the small absolute charge of the oxygen atoms in the carboxyl group. However, it can be seen in Fig. 7 that the trend is only weak and there is strong scattering.

The third suggested heuristic criterion, the nucleophilic character of the oxygen atom of the methoxy group, is evaluated based on the Mulliken charge of the O-atom in the TS. Interestingly, some of the catalysts that enable very high rate constants strongly increase the absolute charge of the O-atom, relating to an increased nucleophilicity. However, no general trend is observed over the whole range of k values. Thus, the third heuristic criterion is also no reliable design objective.

The above discussion shows why choosing heuristic criteria as design objectives is not reliable: Important effects are likely neglected by such criteria. The correlation of heuristic criteria with the ultimate goal of a catalyst design may be weak and there may be strong scattering.

In summary, two important conclusions are drawn from the presented results: First, a full computation of reaction kinetics is required in catalyst design. Heuristic criteria do not guarantee finding the catalyst structures that enable the largest catalytic effects. Second, various other effects besides accelerating reaction kinetics are important to achieve optimal process performance. Consequently, a process-based evaluation of candidate catalysts as used by CAT-COSMO-CAMPD is strongly recommended.

4. Conclusions and future perspectives

Here, we present CAT-COSMO-CAMPD as a method for integrated in silico design of catalysts and processes. The integrated design is formulated as optimization problem and a hybrid optimization scheme is employed to solve the design problem: The genetic optimization algorithm LEA3D is used to explore the chemical design space and to identify the most promising candidate catalyst molecules. Process conditions are optimized in gradient-based process optimizations. The LEA3D algorithm works with 3D structures and designs molecules based on a library of 3D molecule fragments. Thus, 3D structures of all considered molecules are available throughout the design and the direct use of quantum chemical methods for property prediction is facilitated. Consequently, CAT-COSMO-CAMPD avoids simplified property prediction methods and the need for extensive experimental data. Reaction rate constants are predicted based on transition state theory and thermodynamic properties of mixtures are computed using the advanced solvation model COSMO-RS.

The application of CAT-COSMO-CAMPD for integrated catalyst and process design is demonstrated for a catalytic carbamate-cleavage process. The results show that promising catalyst molecules are identified and processes are optimized successfully. Moreover, we show that the integration of molecular design with process optimization is required to achieve optimal process performance. In contrast, selection of catalysts based on predicted reaction rate constants or heuristic criteria likely fails to find the optimal catalyst structures.

So far, we have applied CAT-COSMO-CAMPD only to organic molecular catalysts. More sophisticated homogeneous catalysts such as catalyst complexes may be designed with the same approach, although the QC-based prediction of the catalytic activity of the complexes will be more demanding. First, larger system sizes will either require more computer power or using less sophisticated QC methods. Second, converging geometry optimizations during automated design may become more challenging for larger catalyst structures. Third, different possible spin multiplicities will have to be considered for some catalyst complexes.

The design of heterogeneous catalysts is not within the current scope of our method. For such catalysts, dedicated approaches are required to obtain candidate catalysts and to predict catalytic activity. For a review of such approaches, the reader is referred to Freeze et al. (2019). However, the integration of catalyst design with process optimization using a hybrid optimization scheme as introduced with CAT-COSMO-CAMPD may also be applied in the design of heterogeneous catalysts.

Possible extensions to CAT-COSMO-CAMPD also include more sophisticated process simulation and optimization. For the demonstration of CAT-COSMO-CAMPD in the case study presented in the previous section, a process model was implemented that contains the 2-phase reactor and a flash. Considering additional separation steps or even the complete downstream processing is straightforward in the sense that the additional steps may be implemented in the same process model (Scheffczyk et al., 2018). Of course, the process optimization is expected to become increasingly challenging with an increasing number of degrees of freedom.

In principle, it is also possible to call an external software for rigorous process simulation. In this case, the user could benefit from available libraries with detailed unit operation models and dedicated solution algorithms. In turn, the use of external programs requires suitable interfacing and thus needs to consider how data is transferred to and from the process simulation software. In addition, it must be ensured that no convergence problems occur that would require manual overrides and thus hinder the automation of the design. However, promising solutions for this approach have been presented (Scheffczyk et al., 2018; López et al., 2018; Navarro-Amorós et al., 2014). As an alternative to integrating external software into the design, sophisticated process simulations and optimizations could be used to examine promising candidates after the actual integrated design.

Current limitations of CAT-COSMO-CAMPD in the design of homogeneous catalysts are the accuracy and computational requirements of QC methods used as well as the need to know the mechanism of catalysis and the active group in advance. However, in silico design methods such as CAT-COSMO-CAMPD will benefit from the growth of computational capacity and the ongoing development of efficient but accurate QC methods. Thus, we expect that the design of increasingly complex systems will become possible. There are also promising developments in the field of automated generation of reaction networks (Döntgen et al., 2015; 2018; Simm et al., 2018; Dewyer et al., 2018) that may serve to overcome the need of CAT-COSMO-CAMPD to know the mechanism of catalysis in advance. Moreover, such an automated determination of reaction mechanisms could ease the consideration of multiple catalytically active groups and undesired side reactions. At present, none of the available methods seems to be accurate and efficient enough for a broad range of systems. However, improving such methods is subject of ongoing research.

Surrogate models (e.g., neural networks) may complement QC methods in property prediction during design. For the design of reaction solvents, a hybrid approach using both QC-based prediction of reaction rates and a surrogate model was already introduced by Struebing et al. (2013, 2017). In their work, only part of the candidate molecules are investigated using the full QC-based treatment. Most candidates are evaluated with a surrogate model that is constantly refined during the design procedure using the results from QC. However, one should be aware that the advantages of surrogates regarding computation time and computational capacity usually come with the disadvantage of a lower accuracy compared to the original model. Still, depending on the computational requirements and the number of candidates that need to be evaluated, also integrated catalyst and process design might

benefit from such a hybrid approach. Exploiting the recent advances in machine learning for catalyst design (dos Passos Gomes et al., 2021) could therefore be promising.

Despite the use of high-level QC methods, there is no guarantee that the prediction of catalytic activity does not overestimate the performance of candidates. Therefore, we recommend integrating in silico design with selected experiments in the future. A first step would be testing the top candidates obtained with CAT-COSMO-CAMPD experimentally as already shown for our approach to integrated design of solvents and separation processes (Scheffczyk et al., 2018). The experimental results could be used to confirm the catalytic activity and, if required, to modify the used combination of QC methods for a second design run with CAT-COSMO-CAMPD.

In summary, this article shows the promise of methods for integrated in silico design of catalysts and processes. The proposed CAT-COSMO-CAMPD method was demonstrated successfully in the presented case study and is generally applicable to various other systems. As computational design of molecular catalysts is still a quite unexplored field of research, we expect major developments in the coming years.

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

CRediT authorship contribution statement

Christoph Gertig: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **Lorenz Fleitmann:** Methodology, Software, Writing – review & editing. **Carl Hemprich:** Methodology, Software, Formal analysis, Visualization, Writing – review & editing. **Janik Hense:** Methodology, Formal analysis, Writing – review & editing. **André Bardow:** Conceptualization, Funding acquisition, Writing – review & editing, Supervision. **Kai Leonhard:** Conceptualization, Funding acquisition, Writing – review & editing, Supervision.

Acknowledgment

C.G., A.B. and K.L. gratefully acknowledge funding from the German Federal Ministry of Education and Research (BMBF) within the project Carbon2Polymers (03EK30442C). L.F. thanks the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for funding under Germany's Excellence Strategy - Cluster of Excellence 2186 "The Fuel Science Center" - ID: 390919832. Furthermore, the authors are grateful to J. Langanke, M. Leven and E. Erdkamp for valuable discussions. Simulations were performed with computing resources granted by RWTH Aachen University under projects rwth0284 and rwth0478.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.compchemeng.2021.107438.

References

Adjiman, C.S., Galindo, A., Jackson, G., 2014. Molecules matter: the expanding envelope of process design. Comput. Aided Chem. Eng. 34, 55–64. Elsevier
Ahlrichs, R., Bär, M., Häser, M., Horn, H., Kölmel, C., 1989. Electronic structure calculations on workstation computers: the program system turbomole. Chem. Phys. Lett. 162 (3), 165–169.
Ahn, S., Hong, M., Sundararajan, M., Ess, D.H., Baik, M.-H., 2019. Design and optimization of catalysts based on mechanistic insights derived from quantum chemical reaction modeling. Chem. Rev. 119 (11), 6509–6560.
Frisch et al., M. J., 2013. Gaussian 09, Revision D.01.
Atkins, P., Friedman, R., 2011. Molecular Quantum Mechanics, fifth ed. Oxford University Press, Oxford.
Austin, N.D., Sahinidis, N.V., Trahan, D.W., 2016. Computer-aided molecular design: an introduction and review of tools, applications, and solution techniques. Chem. Eng. Res. Des. 116, 2–26.
Austin, N.D., Sahinidis, N.V., Trahan, D.W., 2017. A COSMO-based approach to computer-aided mixture design. Chem. Eng. Sci. 159, 93–105.
Becke, A.D., 1988. Density-functional exchange-energy approximation with correct asymptotic behavior. Phys. Rev. A 38 (6), 3098.
Cao, Y., Li, H., Qin, N., Zhu, G., 2015. Kinetics of the decomposition of dimethylhexane-1, 6-dicarbamate to 1, 6-hexamethylene diisocyanate. Chinese J. Chem. Eng. 23 (5), 775–779.
Carbon2Chem, URL https://www.thyssenkrupp.com/carbon2chem/de/carbon2chem
Chang, A.M., Rudshteyn, B., Warnke, I., Batista, V.S., 2018. Inverse design of a catalyst for aqueous CO/CO2 conversion informed by the NiII–Iminothiolate complex. Inorg. Chem. 57 (24), 15474–15480.
Chavali, S., Lin, B., Miller, D.C., Camarda, K.V., 2004. Environmentally-benign transition metal catalyst design using optimization techniques. Comput. Chem. Eng. 28 (5), 605–611.
Chu, Y., Heyndrickx, W., Occhipinti, G., Jensen, V.R., Alsberg, B.K., 2012. An evolutionary algorithm for de novo optimization of functional transition metal compounds. J. Am. Chem. Soc. 134 (21), 8885–8895.
Coote, M.L., 2009. Quantum-chemical modeling of free-Radical polymerization. Macromol. Theory Simul. 18 (7–8), 388–400.
COSMOtherm, C3.0, release 1701, COSMOlogic GmbH & Co KG, Leverkusen. URL http://www.cosmologic.de
Deglmann, P., Müller, I., Becker, F., Schäfer, A., Hungenberg, K.-D., Weiß, H., 2009. Prediction of propagation rate coefficients in free radical solution polymerization based on accurate quantum chemical methods: vinylic and related monomers, including acrylates and acrylic acid. Macromol. React. Eng. 3 (9), 496–515.
Dewar, M.J., Zoebisch, E.G., Healy, E.F., Stewart, J.J., 1985. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. J. Am. Chem. Soc. 107 (13), 3902–3909.
Dewyer, A.L., Argüelles, A.J., Zimmerman, P.M., 2018. Methods for exploring reaction space in molecular systems. Wiley Interdiscip. Rev. Comput. Mol. Sci. 8 (2), e1354.
Dittner, M., Hartke, B., 2018. Globally optimal catalytic fields–inverse design of abstract embeddings for maximum reaction rate acceleration. J. Chem. Theory Comput. 14 (7), 3547–3564.
Dittner, M., Hartke, B., 2020. Globally optimal catalytic fields for a Diels–Alder reaction. J. Chem. Phys. 152 (11), 114106.
Döntgen, M., Przybylski-Freund, M.-D., Kröger, L.C., Kopp, W.A., Ismail, A.E., Leonhard, K., 2015. Automated discovery of reaction pathways, rate constants, and transition states using reactive molecular dynamics simulations. J. Chem. Theory Comput. 11 (6), 2517–2524.
Döntgen, M., Schmalz, F., Kopp, W.A., Kröger, L.C., Leonhard, K., 2018. Automated chemical kinetic modeling via hybrid reactive molecular dynamics and quantum chemistry simulations. J. Chem. Inf. Model. 58 (7), 1343–1355.
Douguet, D., Munier-Lehmann, H., Labesse, G., Pochet, S., 2005. LEA3D: a computer-aided ligand design for structure-based drug design. J. Med. Chem. 48 (7), 2457–2468.
Eckart, C., 1930. The penetration of a potential barrier by electrons. Phys. Rev. 35 (11), 1303–1309.
Eckert, F., Klamt, A., 2002. Fast solvent screening via quantum chemistry: COSMO-RS approach. AIChE J. 48 (2), 369–385.
Eyring, H., 1935. The activated complex in chemical reactions. J. Chem. Phys. 3 (2), 107–115.
Fink, T., Bruggesser, H., Reymond, J.-L., 2005. Virtual exploration of the small–molecule chemical universe below 160 daltons. Angew. Chem. Int. Ed. 44 (10), 1504–1508.
Fleitmann, L., Scheffczyk, J., Schäfer, P., Jens, C., Leonhard, K., Bardow, A., 2018. Integrated design of solvents in hybrid reaction-separation processes using COSMO-RS. Chem. Eng. 69.
Foresman, J.B., Frisch, Æ., 2015. Exploring Chemistry with Electronic Structure Methods, third ed. Gaussian, Inc., Wallingford, CT.
Foscato, M., Jensen, V.R., 2020. Automated in silico design of homogeneous catalysts. ACS Catal. 10 (3), 2354–2377.
Freeze, J.G., Kelly, H.R., Batista, V.S., 2019. Search for catalysts by inverse design: artificial intelligence, mountain climbers, and alchemists. Chem. Rev. 119 (11), 6595–6612.
Funes-Ardoiz, I., Paton, R. S., 2018. GoodVibes: GoodVibes 2.0.3. DOI:10.5281/zenodo.595246. 10.5281/zenodo.595246
Gani, R., 2004. Chemical product design: challenges and opportunities. Comput. Chem. Eng. 28 (12), 2441–2457.
Gani, R., 2019. Group contribution-based property estimation methods: advances and perspectives. Curr. Opin. Chem. Eng. 23, 184–196.
Gertig, C., Erdkamp, E., Ernst, A., Hemprich, C., Kröger, L.C., Langanke, J., Bardow, A., Leonhard, K., 2021. Reaction mechanisms and rate constants of auto-catalytic urethane formation and cleavage reactions. ChemistryOpen. accepted for publication

Gertig, C., Fleitmann, L., Hemprich, C., Hense, J., Bardow, A., Leonhard, K., 2020. Integrated in silico design of catalysts and processes based on quantum chemistry. Comput. Aided Chem. Eng. 48, 889–894. Elsevier

Gertig, C., Fleitmann, L., Schilling, J., Leonhard, K., Bardow, A., 2020. Rx-COSMO-CAMPD: enhancing reactions by integrated computer-Aided design of solvents and processes based on quantum chemistry. Chem. Ing. Tech. 92 (10), 1489–1500.

Gertig, C., Kröger, L., Fleitmann, L., Scheffczyk, J., Bardow, A., Leonhard, K., 2019. Rx-COSMO-CAMD: computer-aided molecular design of reaction solvents based on predictive kinetics from quantum chemistry. Ind. Eng. Chem. Res. 58 (51), 22835–22846.

Gertig, C., Leonhard, K., Bardow, A., 2019. Integrated Design of Solvents and Processes Based on Reaction Kinetics from Quantum Chemical Prediction Methods. Comput. Aided Chem. Eng. 46, 415–420. Elsevier

Gertig, C., Leonhard, K., Bardow, A., 2020. Computer-aided molecular and processes design based on quantum chemistry: current status and future prospects. Curr. Opin. Chem. Eng. 27, 89–97.

Ghysels, A., Verstraelen, T., Hemelsoet, K., Waroquier, M., van Speybroeck, V., 2010. TAMkin: A versatile package for vibrational analysis and chemical kinetics. J. Chem. Inf. Model. 50 (9), 1736–1750. doi:10.1021/ci100099g.

Gmehling, J., 2009. Present status and potential of group contribution methods for process development. J. Chem. Thermodyn. 41 (6), 731–747.

Gottschalk, H.C., Poblotzki, A., Suhm, M.A., Al-Mogren, M.M., Antony, J., Auer, A.A., Baptista, L., Benoit, D.M., Bistoni, G., Bohle, F., Dahmani, R., Firaha, D., Grimme, S., Hansen, A., Harding, M.E., Hochlaf, M., Holzer, C., Jansen, G., Klopper, W., Kopp, W.A., Kröger, L.C., Leonhard, K., Mouhib, H., Neese, F., Pereira, M.N., Ulusoy, I.S., Wuttke, A., Mata, R.A., 2018. The furan microsolvation blind challenge for quantum chemical methods: first steps. J. Chem. Phys. 148 (1), 014301.

Grimme, S., 2012. Supramolecular binding thermodynamics by dispersion-corrected density functional theory. Chem. Eur. J. 18 (32), 9955–9964.

Grimme, S., Antony, J., Ehrlich, S., Krieg, H., 2010. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J. Chem. Phys. 132 (15), 154104.

Hagen, J., 2015. Industrial Catalysis, 3rd WILEY-VCH.

Halgren, T.A., 1996. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. J. Comput. Chem. 17 (5–6), 490–519.

Halgren, T.A., 1996. Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. J. Comput. Chem. 17 (5–6), 553–586.

Halgren, T.A., 1996. Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. J. Comput. Chem. 17 (5–6), 616–641.

Hare, S.R., Pemberton, R.P., Tantillo, D.J., 2017. Navigating past a fork in the road: carbocation-π interactions can manipulate dynamic behavior of reactions facing post-transition-state bifurcations. J. Am. Chem. Soc. 139 (22), 7485–7493.

Hellweg, A., Eckert, F., 2017. Brick by brick computation of the gibbs free energy of reaction in solution using quantum chemistry and COSMO-RS. AIChE J. 63 (9), 3944–3954.

Kaiser, T., Rathgeb, A., Gertig, C., Bardow, A., Leonhard, K., Jupke, A., 2018. Carbon2Polymer–conceptual design of a CO2-based process for the production of isocyanates. Chem. Ing. Tech. 90 (10), 1497–1503.

Klamt, A., 1995. Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena. J. Phys. Chem. 99 (7), 2224–2235.

Klamt, A., Eckert, F., Arlt, W., 2010. COSMO-RS: an alternative to simulation for calculating thermodynamic properties of liquid mixtures. Annu. Rev. Chem. Biomol. Eng. 1, 101–122.

Klamt, A., Jonas, V., Bürger, T., Lohrenz, J.C., 1998. Refinement and parametrization of COSMO-RS. J. Phys. Chem. A 102 (26), 5074–5085.

Klamt, A., Schüürmann, G., 1993. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. J. Chem. Soc., Perkin Trans. 2 (5), 799–805.

Krausbeck, F., Sobez, J.-G., Reiher, M., 2017. Stabilization of activated fragments by shell-wise construction of an embedding environment. J. Comput. Chem. 38 (14), 1023–1038.

Kröger, L.C., Kopp, W.A., Leonhard, K., 2017. Prediction of chain propagation rate constants of polymerization reactions in aqueous NIPAM/BIS and VCL/BIS systems. J. Phys. Chem. B 121 (13), 2887–2895.

Leitner, W., Franciò, G., Scott, M., Westhues, C., Langanke, J., Lansing, M., Hussong, C., Erdkamp, E., 2018. Carbon2Polymer–chemical utilization of CO2 in the production of isocyanates. Chem. Ing. Tech. 90 (10), 1504–1512.

Levenspiel, O., 1999. Chemical Reaction Engineering, third ed. John Wiley & Sons, New York.

Lin, B., Chavali, S., Camarda, K., Miller, D.C., 2005. Computer-aided molecular design using Tabu search. Comput. Chem. Eng. 29 (2), 337–347.

Linke, P., Papadopoulos, A.I., Seferlis, P., 2015. Systematic methods for working fluid selection and the design, integration and control of organic Rankine cycles-a review. Energies 8 (6), 4755–4801.

Liu, Q., Zhang, L., Liu, L., Du, J., Meng, Q., Gani, R., 2019. Computer-aided reaction solvent design based on transition state theory and COSMO-SAC. Chem. Eng. Sci. 202, 300–317.

Liu, Q., Zhang, L., Tang, K., Feng, Y., Zhang, J., Zhuang, Y., Liu, L., Du, J., 2019. Computer-aided reaction solvent design considering inertness using group contribution-based reaction thermodynamic model. Chem. Eng. Res. Des. 152, 123–133.

López, C.A.M., Telen, D., Nimmegeers, P., Cabianca, L., Logist, F., Van Impe, J., 2018. A process simulator interface for multiobjective optimization of chemical processes. Comput. Chem. Eng. 109, 119–137.

MATLAB R2019b, 2019. R2019b, The MathWorks, Natick, MA, USA.

Mulliken, R.S., 1955. Electronic population analysis on LCAO–MO molecular wave functions. I. J. Chem. Phys. 23 (10), 1833–1840.

Navarro-Amorós, M.A., Ruiz-Femenia, R., Caballero, J.A., 2014. Integration of modular process simulators under the generalized disjunctive programming framework for the structural flowsheet optimization. Comput. Chem. Eng. 67, 13–25.

Neese, F., 2018. Software update: the ORCA program system, version 4.0. Wiley Interdiscip. Rev. Comput. Mol. Sci. 8 (1), e1327.

O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T., Hutchison, G.R., 2011. Open Babel: an open chemical toolbox. J. Cheminf. 3 (1), 33.

Occhipinti, G., Bjørsvik, H.-R., Jensen, V.R., 2006. Quantitative structure- activity relationships of ruthenium catalysts for olefin metathesis. J. Am. Chem. Soc. 128 (21), 6952–6964.

Open Babel, URL http://openbabel.org/wiki/Main_Page

Papadopoulos, A.I., Linke, P., 2009. Integrated solvent and process selection for separation and reactive separation systems. Chem. Eng. Process 48 (5), 1047–1060.

Papadopoulos, A.I., Tsivintzelis, I., Linke, P., Seferlis, P., 2018. Computer aided molecular design: fundamentals, methods and applications. Ref. Module Chem. Mol. Sci. Chem. Eng. doi:10.1016/B978-0-12-409547-2.14342-2.

dos Passos Gomes, G., Pollice, R., Aspuru-Guzik, A., 2021. Navigating through the maze of homogeneous catalyst design with machine learning. Trends Chem..

Paulechka, E., Kazakov, A., 2017. Efficient DLPNO–CCSD (T)-based estimation of formation enthalpies for C-, H-, O-, and N-containing closed-shell compounds validated against critically evaluated experimental data. J. Phys. Chem. A 121 (22), 4379–4387.

Perdew, J.P., 1986. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. Phys. Rev. B 33 (12), 8822.

Perdew, J.P., 1986. Erratum: density-functional approximation for the correlation energy of the inhomogeneous electron gas. Phys. Rev. B 34 (10), 7406.

Peters, M., Greiner, L., Leonhard, K., 2008. Illustrating computational solvent screening: prediction of standard gibbs energies of reaction in solution. AIChE J. 54 (10), 2729–2734.

Pfennig, A., 2004. Thermodynamik der Gemische. Springer Verlag, Berlin.

Poree, C., Schoenebeck, F., 2017. A holy grail in chemistry: computational catalyst design: feasible or fiction? Acc. Chem. Res. 50 (3), 605–608.

Renon, H., Prausnitz, J.M., 1968. Local compositions in thermodynamic excess functions for liquid mixtures. AIChE J. 14 (1), 135–144.

Reymond, J.-L., 2015. The chemical space project. Acc. Chem. Res. 48 (3), 722–730.

Riplinger, C., Neese, F., 2013. An efficient and near linear scaling pair natural orbital based local coupled cluster method. J. Chem. Phys. 138 (3), 034106.

Riplinger, C., Sandhoefer, B., Hansen, A., Neese, F., 2013. Natural triple excitations in local coupled cluster calculations with pair natural orbitals. J. Chem. Phys. 139 (13), 134101.

Safe, M., Carballido, J., Ponzoni, I., Brignole, N., 2004. On stopping criteria for genetic algorithms. In: Brazilian Symposium on Artificial Intelligence. Springer, pp. 405–413.

Satchell, D.P.N., Satchell, R.S., 1975. Acylation by ketens and isocyanates. a mechanistic comparison. Chem. Soc. Rev. 4 (2), 231–250.

Scheffczyk, J., Fleitmann, L., Schwarz, A., Lampe, M., Bardow, A., Leonhard, K., 2017. COSMO-CAMD: A framework for optimization-based computer-aided molecular design using COSMO-RS. Chem. Eng. Sci. 159, 84–92.

Scheffczyk, J., Schäfer, P., Fleitmann, L., Thien, J., Redepenning, C., Leonhard, K., Marquardt, W., Bardow, A., 2018. COSMO-CAMPD: a framework for integrated design of molecules and processes based on COSMO-RS. Mol. Syst. Des. Eng. 3 (4), 645–657.

Schilling, J., Horend, C., Bardow, A., 2020. Integrating superstructure-based design of molecules, processes, and flowsheets. AIChE J. 66 (5), e16903.

Schilling, J., Tillmanns, D., Lampe, M., Hopp, M., Gross, J., Bardow, A., 2017. From molecules to dollars: integrating molecular design into thermo-economic process design using consistent thermodynamic modeling. Mol. Syst. Des. Eng. 2 (3), 301–320.

Simm, G.N., Vaucher, A.C., Reiher, M., 2018. Exploration of reaction pathways and chemical transformation networks. J. Phys. Chem. A 123 (2), 385–399.

Six, C., Richter, F., 2003. Isocyanates, Organic. Ullmann's Encyclopedia of Industrial Chemistry. Wiley Online Library.

Stephens, P., Devlin, F., Chabalowski, C., Frisch, M.J., 1994. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. J. Phys. Chem. 98 (45), 11623–11627.

Struebing, H., Ganase, Z., Karamertzanis, P.G., Siougkrou, E., Haycock, P., Piccione, P.M., Armstrong, A., Galindo, A., Adjiman, C.S., 2013. Computer-aided molecular design of solvents for accelerated reaction kinetics. Nat. Chem. 5 (11), 952–957.

Struebing, H., Obermeier, S., Siougkrou, E., Adjiman, C.S., Galindo, A., 2017. A QM–CAMD approach to solvent design for optimal reaction rates. Chem. Eng. Sci. 159, 69–83.

Tantillo, D.J., Jiangang, C., Houk, K.N., 1998. Theozymes and compuzymes: theoretical models for biological catalysis. Curr. Opin. Chem. Biol. 2 (6), 743–750.

Umer, M., Leonhard, K., 2013. Ab initio calculations of thermochemical properties of methanol clusters. J. Phys. Chem. A 117 (7), 1569–1582.

Vereecken, L., Glowacki, D.R., Pilling, M.J., 2015. Theoretical chemical kinetics in tropospheric chemistry: methodologies and applications. Chem. Rev. 115 (10), 4063–4114.
Wang, P., Liu, S., Deng, Y., 2017. Important green chemistry and catalysis: non-phosgene syntheses of isocyanates–thermal cracking way. Chinese J. Chem. 35 (6), 821–835.
Weininger, D., 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci. 28 (1), 31–36.
Xiao, D., Yang, W., Beratan, D.N., 2008. Inverse molecular design in a tight-binding framework. J. Chem. Phys. 129 (4), 044106.
Zhang, L., Pang, J., Zhuang, Y., Liu, L., Du, J., Yuan, Z., 2020. Integrated solvent-process design methodology based on COSMO-SAC and quantum mechanics for TMQ (2, 2, 4-trimethyl-1, 2-H-dihydroquinoline) production. Chem. Eng. Sci. 226, 115894.
Zheng, J., Zhao, Y., Truhlar, D.G., 2009. The DBH24/08 database and its use to assess electronic structure model chemistries for chemical reaction barrier heights. J. Chem. Theory Comput. 5 (4), 808–821.
Zhou, T., McBride, K., Zhang, X., Qi, Z., Sundmacher, K., 2015. Integrated solvent and process design exemplified for a Diels–Alder reaction. AIChE J. 61 (1), 147–158.
TURBOMOLE V7.0.1 2015, A Development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH. TURBOMOLE GmbH, 1989–2007, since 2007, URl: http://www.turbomole.com