Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Chemical Engineering

Computer-aided molecular and processes design based on quantum chemistry: current status and future prospects

Christoph Gertig[1], Kai Leonhard[1] and André Bardow[1,2]

Computer-Aided Molecular Design (CAMD) enables the automated exploration of chemical space and thus offers great possibilities for efficient design of chemical products. The key to reliable CAMD is a sound prediction of the properties of desired products, where quantum chemistry-based (quantum chemical, QC) prediction methods offer unique opportunities. In this article, we discuss CAMD methods based on QC and highlight two important fields of application: the design of solvents and of molecular catalysts. Screening of separation solvents based on physical property targets can be regarded as established by now. However, the integration of molecular design and process design remains an important challenge. For the design of reactive systems, transition state theory provides a sound basis. However, efficient CAMD methods and tools based on quantum chemistry are still in their infancy. Recent results and the unexplored opportunities of quantum chemistry make the development of QC-based CAMD methods a promising field of research.

Addresses
[1] Institute of Technical Thermodynamics, RWTH Aachen University, Aachen, Germany
[2] Institute of Energy and Climate Research - Energy Systems Engineering (IEK-10), Forschungszentrum Jülich, Jülich, Germany

Corresponding author:
Bardow, André (andre.bardow@ltt.rwth-aachen.de)

Current Opinion in Chemical Engineering 2019, **27**:89–97

This review comes from a themed issue on **Frontiers of chemical engineering**

Edited by **Rafiqul Gani** and **Lei Zhang**

For a complete overview see the Issue and the Editorial

Available online 14th January 2020

https://doi.org/10.1016/j.coche.2019.11.007

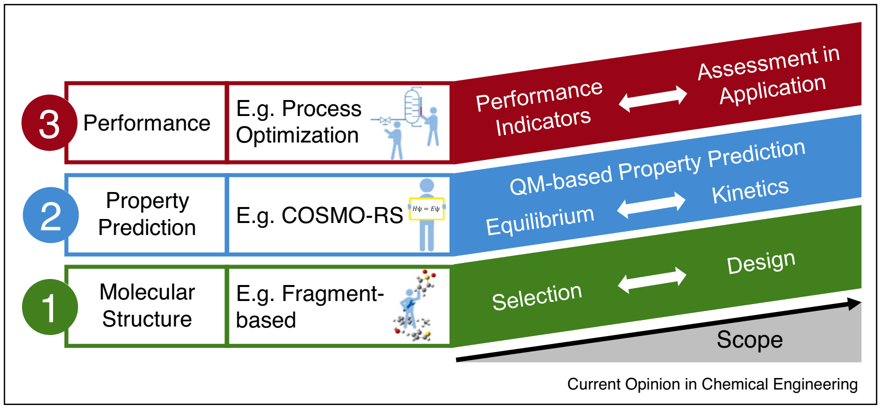
Introduction

The world is constantly demanding more chemical products [1]. At the same time, innovative chemical products and processes should be developed fast and efficiently due to global competition [2]. However, searching new molecules that best meet product requirements can be very tedious due to the vast chemical design space [3] that is estimated to contain up to 10^{200} organic molecules [4]. To explore large molecular search spaces, computer-aided molecular design (CAMD) [5•] methods have been developed. CAMD methods have been successfully applied in chemical product design [6•] of, for example, solvents, ionic liquids, catalysts, heat exchange fluids, and polymers (see Papadopoulos *et al.* for a comprehensive review [7••]).

Computer-aided molecular design requires three building blocks (Figure 1):

1 A method to explore chemical space is required to obtain the structure of molecules considered as chemical products. For example, candidate molecules can be generated and optimized based on a set of molecule fragments as building blocks. Strictly, the term ‘molecular design’ refers to the exploration of a molecular design space where not only known, but also promising new molecules are identified [5•]. While such design methods have a larger scope, we also consider the selection of candidates from databases of known molecules (screening) in this review.
2 A reliable method for property prediction is key to evaluating the performance of candidate molecules. Property prediction in CAMD is often based on group contribution (GC) [7••] methods like UNIFAC. While group contribution methods are straightforward to implement and computationally efficient, they also face significant limitations [8•,9]. First, GC methods assume that the properties of a molecule can be described based on its functional groups, neglecting the 3-dimensional structure. Second, parameters for every functional group considered by a GC method have to be fitted to extensive experimental data. This requirement practically limits the accessible molecular design space. Third, several parameter sets or GC methods are often required to predict different properties. Quantum chemistry-based (quantum chemical, QC) methods overcome the discussed limitations. Nevertheless, the use of QC in CAMD has been limited for a long time because of the computational requirements. However, both the available computational power and computationally efficient QC methods have improved considerably during the past decade. Thus, CAMD based on QC is becoming increasingly important and is the focus of this article. Some QC-based property prediction methods are routinely used in CAMD already today, for example, the Conductor-like Screening Model for Realistic Solvation (COSMO-RS) [10].

Figure 1



Current Opinion in Chemical Engineering

Overview of building blocks required to set up CAMD.

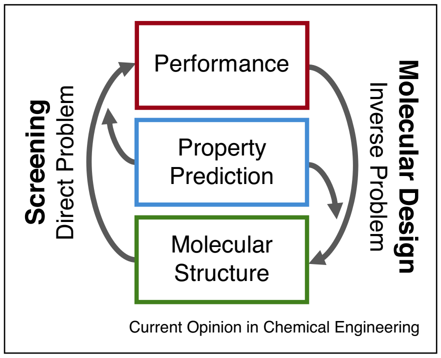
3. Considering the predicted properties of candidate molecules, their expected product performance must be evaluated. Many CAMD methods evaluate candidate molecules using simple performance indicators derived from physical properties. However, such performance indicators may not capture tradeoffs required for optimal performance [11]. Therefore, molecules should be assessed based on their predicted performance in the intended application. For example, the performance of designed solvent molecules should ideally be evaluated by optimizing the process that uses the solvents.

Recent developments for the three building blocks of CAMD using QC are discussed before we highlight CAMD applications in two important areas: the design of solvents and molecular catalysts. With this opinion paper, we aim to clarify the current research frontier. In particular, we show that screening of databases containing known molecules to match thermodynamic equilibrium property targets can be regarded as established method supported by suitable software. In addition, prediction of solvent effects on reaction kinetics has reached a high level of maturity with transition state theory as method of choice. Further development is required for the actual design of new molecules and for integrating CAMD with process design as discussed in the following.

Computer-aided molecular design

Computer-aided molecular design aims at finding the best molecule for a certain application [5•]. This aim corresponds to the solution of an inverse problem (Figure 2): performance measures are specified as design objectives and property prediction methods are used to design the molecular structure that shows the best performance. Thus, CAMD is sometimes referred to as 'inverse property prediction' or 'inverse performance prediction'.

Figure 2



Current Opinion in Chemical Engineering

Scheme of molecular design as inverse problem and screening as direct problem.

The inverse problem can be formulated as mathematical optimization problem [5•,12] as proposed by Gani [13]:

$$
\begin{array}{lll}
\min_{x,y} & f(x,y) & \text{performance} \\
s.t. & h(x,y) \leq 0 & \text{(in)equality constraints} \Rightarrow \text{properties and application} \\
 & g(y) \leq 0 & \text{(in)equality constraints} \Rightarrow \text{molecular structure}
\end{array}
\tag{1}
$$

Here, $f(x,y)$ is a design objective used to measure performance depending on continuous variables x (e.g. process conditions) and the molecular structure y of the designed molecule. The equality constraints in $h(x,y)$ represent equations for property prediction and for the intended application (e.g. process models). Inequality constraints in $h(x,y)$ restrict, for example, molecular and thermodynamic properties. Equality and inequality constraints $g(y)$ ensure chemical feasibility of the designed molecular structures and are often also used to limit the explored design space, for example, to structures with certain numbers of atoms or structures containing specific combinations of functional groups. While the constraints $g(y)$ are usually linear, the constraints $h(x, y)$ generally include nonlinear equations.

CAMD based on group contribution methods has the advantage that analytical derivatives can usually be calculated for the objective function $f(x, y)$ and the constraints $h(x, y)$ and $g(y)$ with respect to the molecular structure y (represented by the set of functional groups comprising the molecule). This is not possible for QC-based CAMD due to the inherent discrete nature of atoms in a molecule. Thus, optimization algorithms used to solve the design Problem (1) have to circumvent the need for these analytical derivatives. Besides optimization, enumerative algorithms are often used in CAMD to test all candidates from a database (screening) or a molecular design space (generate-and-test). Enumerative algorithms provide good solutions but have to test all candidates in the database/design space. An excellent review of solution approaches for CAMD is given by Papadopoulos *et al.* [7••].

Quantum chemical methods for property prediction in CAMD

There are different groups of QC methods used to predict properties [14,15]:

1. Methods like Hartree-Fock (HF) are based on the numerical solution of the Schrödinger equation (wave function methods) [14]. Modern post-HF methods offer the highest accuracy at the highest computational cost [16]. For small molecules with 1 or 2 non-hydrogen atoms, Zheng *et al.* [16] found mean unsigned errors in calculated energies of about 2 kJ/mol with the coupled cluster method CCSD(T) and large basis sets, often called the 'gold standard' of quantum chemistry [17].
2. Quantum chemical methods based on density functional theory (DFT) compute properties from the distribution of electron density [18]. Compared to post-HF methods, DFT methods typically offer higher computational efficiency, but lower accuracy [19]. Zheng *et al.* [16] found mean unsigned errors of about 8–24 kJ/mol for typical DFT methods like B3LYP but at 10^3 to 10^4 lower computational cost than for CCSD(T).
3. Semi-empirical methods approximate wave function methods and density functional tight binding methods (DFTB) approximate DFT [20]. Empirical approximations significantly reduce accuracy, but greatly enhance computational efficiency. Unsigned errors take values of up to 120 kJ/mol for semi-empirical methods while computational cost was found lower by factors around 10^8 compared to CCSD(T) [16].

Importantly, the computational effort required for the different methods scales differently with the size of the studied molecules such that the relative computational effort depends on molecular size. While methods with low computational requirements are not accurate over a broad range of systems and properties, they may still perform well in specific cases with a much higher accuracy than given above. However, appropriate validation data are usually required to identify a suitable low-cost method for a given system and such data are often unavailable. Only high-level methods like CCSD(T) are suited to compute properties with high accuracy for a broad range of systems.

QC methods predict molecular properties like geometry and energy and vibrational analyses determine vibrational states. However, CAMD requires not only molecular properties but also thermodynamic and kinetic quantities. These quantities arise in systems with many molecules and are determined by thermochemical calculations based on statistical thermodynamics and the results from QC [21,22]. QC calculations performed in vacuum consider no intermolecular interactions, which results in an ideal gas treatment of thermodynamic properties. Many applications in chemical product design need properties of liquids. Continuum solvation models (CSM) allow for the calculation of the required solvation properties considering the liquid environment of a molecule as a continuum. Important examples of CSM-based prediction methods are SMD [23], COSMO-RS [10] and COSMO-SAC [24].

Often, desired properties are difficult to calculate based on physical relationships but still correlate well with molecular structure and descriptors from QC. Such properties can be predicted based on QC by quantitative structure-property relationships (QSPR) [25•]. A general procedure has been established to develop and use QSPR models [25•]: To develop a QSPR, coefficients are fitted to training data and subsequently validated using validation data. Afterwards, the QSPR is used to predict properties based on molecular descriptors. This procedure also applies to QSPRs with descriptors from QC.

QSPRs with descriptors from QC have been used to predict a variety of properties of chemical products including melting points, boiling points, densities, vapor pressures, solubilities, surface tensions [25•], reaction kinetics [26], optical properties [27], lubricity [28] and the effectiveness of pharmaceuticals [29].

The QC-based prediction methods described above have been included into CAMD to design chemical products. In the next sections, we discuss recent CAMD methods for the design of two important chemical products: solvents and molecular catalysts.

QC-based CAMD of solvents for separations

CAMD of separation solvents based on QC is commonly based on performance indicators derived from thermodynamic properties (Figure 3). Farahipour *et al.* [30] design ionic liquids (IL) for CO_2 capture using the absorption-desorption index (ADI) to measure performance. The ADI is derived from Henry coefficients predicted with COSMO-RS to indicate absorption capacity and the ease of desorption at elevated temperatures. The authors select ten promising ILs by screening a database with 212 cations and 63 anions. In contrast, Ahmad *et al.* [31] use a generate-and-test approach to design solvents for chemical absorption of CO_2. The enthalpy of chemisorption reactions is predicted using the DFT method B3LYP and employed as indicator for the effort required for solvent recovery. Scheffczyk *et al.* [32] use a genetic optimization algorithm to design solvents for extraction processes based on 3D molecule fragments. Performance indicators like distribution coefficients are derived from activity coefficients predicted with COSMO-RS and used as performance measure.

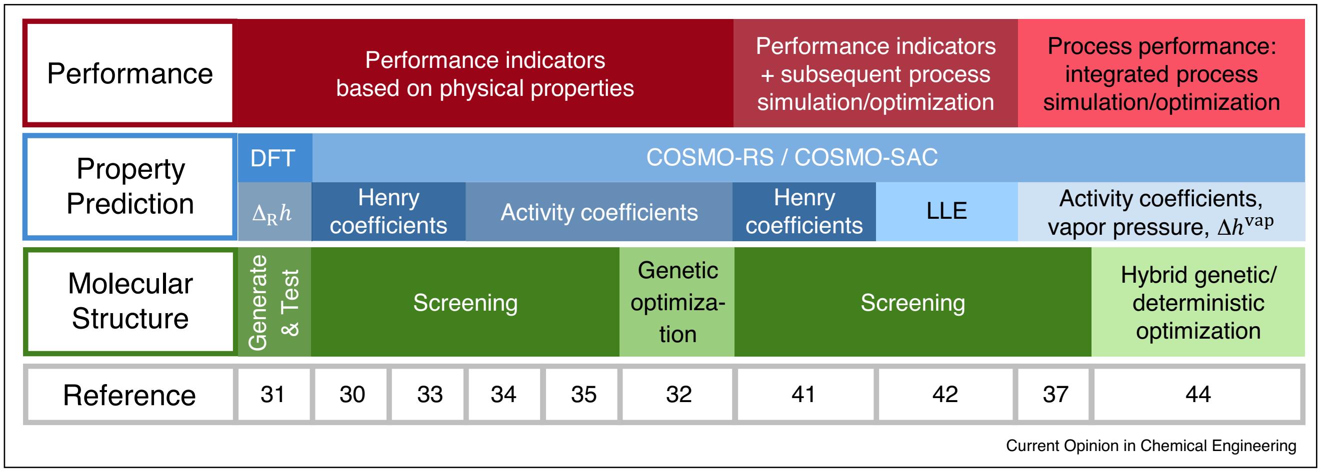
To increase the reliability of the results, some authors include experimental testing of promising candidates as final step of their design. Yu *et al.* [33] select ILs for natural gas dehydration in a COSMO-RS-based database screening and perform a dehydration experiment to confirm the performance of the top candidate. Fang *et al.* [34] select ILs as solvents for extractive distillation in a screening using COSMO-SAC and perform experiments to confirm the enhancement of relative volatility. Song *et al.* [35] perform a multi-stage screening to find ILs as extraction solvents, where the first screening step is based on COSMO-RS and the final screening step evaluates all remaining candidates by experiments. Although quantitative agreement between prediction and experiment cannot always be expected, the experimental investigations to validate design results show that QC-based CAMD is able to catch the right trends and identify promising molecules.

QC-based screening has been widely applied to find separation solvents, with ILs for absorption processes as common application [36]. COSMO-RS and COSMO-SAC are the prominent property prediction methods (Figure 3). The search for solvents based on predicted thermodynamic properties is now very time-efficient since databases with results of QC calculations are readily available. Screenings evaluate thousands of candidates in a few hours [37]. The QC-based screening of separation solvents based on thermodynamic properties has reached a high level of maturity due to the availability of established methods (COSMO-RS/SAC) with software tools and databases.

Going beyond screening of known molecular structures to design new molecules increases the computational effort, as QC calculations are required for newly designed molecules. The whole CAMD procedure may then take up to several days. Some authors avoid this computational effort by partly replacing QC with GC methods [38–40]. GC methods considerably speed up the design and integrate well into efficient integer optimization methods, but remove some advantages of QC (e.g. independence of fitted functional groups). Genetic algorithms using molecular fragments have provided the only route to directly generate novel molecules as input for QC calculations [32]. More efficient optimization algorithms for QC-based molecular design are therefore an important research target.

While most QC-based CAMD methods evaluate solvents based on performance indicators only (Figure 3), important tradeoffs may be neglected such that it is beneficial to evaluate solvents based on their processes performance. Some authors assess process performance after the screening/design [41,42]. However, the optimal solvent may strongly depend on the process conditions

Figure 3



Reference	31	30	33	34	35	32	41	42	37	44
Performance	Performance indicators based on physical properties						Performance indicators + subsequent process simulation/optimization		Process performance: integrated process simulation/optimization	
Property Prediction	DFT	COSMO-RS / COSMO-SAC								
	$\Delta_R h$	Henry coefficients		Activity coefficients			Henry coefficients	LLE	Activity coefficients, vapor pressure, Δh^{vap}	
Molecular Structure	Generate & Test	Screening				Genetic optimization	Screening			Hybrid genetic/ deterministic optimization

Current Opinion in Chemical Engineering

Overview of discussed CAMD methods for separation solvent design.

and vice versa [11]. Therefore, it is increasingly recognized that the process-based assessment should be integrated into the CAMD procedure. Scheffczyk *et al.* [37] screen solvents for extraction-distillation processes based on COSMO-RS and evaluate every candidate solvent using pinch-based models of distillation and extraction columns [43]. Energy demand and solvent loss are minimized. In subsequent work, Scheffczyk *et al.* [44•] integrate solvent and process design based on COSMO-RS. A hybrid optimization scheme is applied using genetic optimization for the solvent and deterministic optimization for process conditions. Again, pinch-based models are employed for efficient process-based assessment of every candidate solvent during design. Liquid-liquid equilibrium (LLE) experiments are performed confirming the potential of designed solvents. Assessment on the application level has been shown to improve design results. However, efficient methods and software tools for integrated solvent and process design based on QC are still lacking.

QC-based CAMD of reaction solvents

The design of optimal solvents for reaction-based processes requires property prediction to quantify the influence of solvents on reaction equilibrium and kinetics. QC methods are an excellent choice especially when it comes to computing reaction kinetics. To predict solvent-dependent reaction rate constants, two QC-based approaches have been used in CAMD: QSPR models where QC is used to obtain descriptors and/or training data and prediction with conventional transition state theory (TST) [14]. With TST, reaction rate constants are calculated from the activation barrier in Gibbs free energy between the reactants and a transition state. According to conventional TST, this transition state is a first-order saddle point on the minimum-energy path from the reactants to the desired products and can be identified by a geometry optimization.

Several methods for the design of reaction solvents employ predicted reaction rate constants as performance indicator. In pioneering work, Struebing *et al.* [45•,46] design solvents based on a set of selected functional groups. A QSPR model is used with descriptors from GC methods to obtain reaction rate constants. The coefficients of the QSPR are fitted to rate constants predicted by TST. The required activation barriers are predicted with DFT methods and SMD. The QSPR model is refined during the CAMD procedure using rate constants of top solvents from previous optimization steps predicted with TST. Thereby, the computational effort is reduced compared to using TST for predicting the rate constants in all candidate solvents. The prediction is, however, still partly based on GC methods. Austin *et al.* [47] also design solvents based on a set of groups. They use TST and activation barriers from DFT and COSMO-RS to calculate rate constants. An additional parameter is fitted to experimental data to enable quantitative prediction. QC calculations are partly replaced by GC methods, which expedites the design but restricts the design space to the available groups. Gertig *et al.* [48] propose a method for reaction solvent design based on pure prediction of rate constants. Solvent-independent contributions to the activation barriers are computed with post-HF methods and solvent-dependent contributions with COSMO-RS. The solvent is optimized with a genetic algorithm based on 3D molecule fragments, facilitating the QC-based prediction.

As for separation solvents, performance evaluation based on performance indicators does not always reflect process performance and may result in suboptimal solvents. Thus, a process-based assessment is desirable. Zhou *et al.* [49•] evaluate candidates based on the difference in final concentrations of desired and undesired products determined by reactor simulations. A QSPR with COSMO-RS-based descriptors is fitted to experimental data and used to predict rate constants. The required input for COSMO-RS is calculated with a GC method instead of QC. Inaccuracies in the prediction are taken into account using robust optimization in the solvent design. The same authors use their QSPR in another integrated design of solvent and process with profit calculated using process and cost models as performance measure [26]. Gertig *et al.* [50] extend their solvent design method based on pure prediction with QC [48] to an integrated design of solvents and processes. The integrated design problem is solved employing a hybrid optimization scheme: the solvent molecular structure is optimized with a genetic optimization algorithm based on 3D molecule fragments. The optimal process conditions and process performance are determined by deterministic process optimizations.

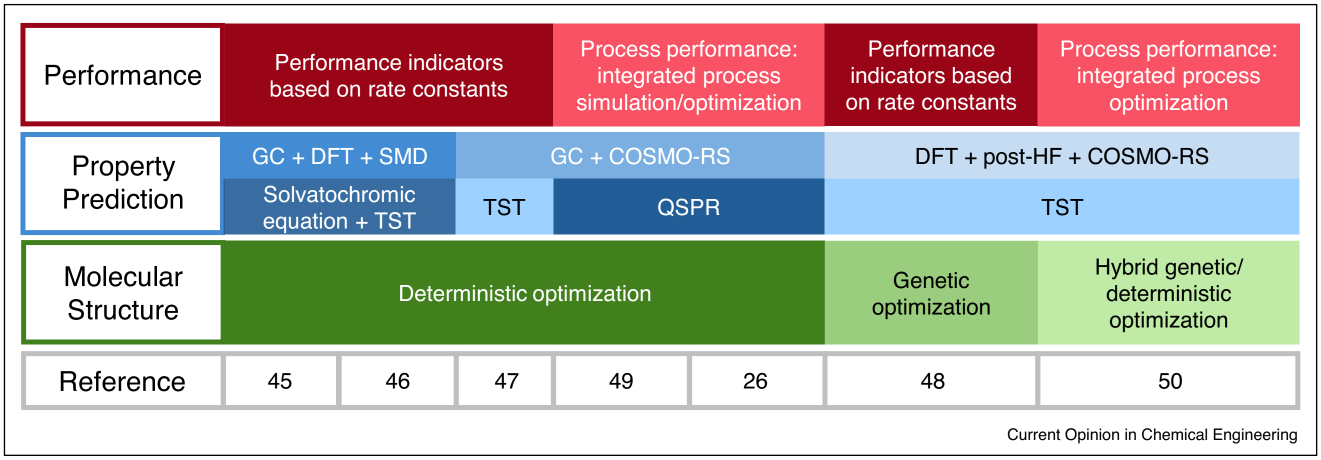
In comparison to CAMD for separation solvents, the design of reaction solvents is much less developed. However, TST seems to be established as method of choice to predict reaction kinetics (Figure 4). The increased computational cost now also seems manageable such that the first fully predictive CAMD of reaction solvents based on high-level QC could recently be demonstrated [48]. Further research is needed for efficient CAMD methods, and for the integration with process design.

Towards CAMD of molecular catalysts

CAMD of catalysts is still considered as one of the ‘holy grails in chemistry’ [51]. However, approaches towards automated in silico catalyst design are emerging. Here, we focus on the design of molecular (homogeneous) catalysts based on QC. For approaches to design heterogeneous catalysts, the reader is referred to the recent review of Freeze *et al.* [52].

Property prediction is now needed to assess catalytic effects. These effects have been studied at different

Figure 4



Performance	Performance indicators based on rate constants			Process performance: integrated process simulation/optimization		Performance indicators based on rate constants	Process performance: integrated process optimization
Property Prediction	GC + DFT + SMD		GC + COSMO-RS			DFT + post-HF + COSMO-RS	
	Solvatochromic equation + TST		TST	QSPR		TST	
Molecular Structure	Deterministic optimization					Genetic optimization	Hybrid genetic/ deterministic optimization
Reference	45	46	47	49	26	48	50

Current Opinion in Chemical Engineering

Overview of discussed CAMD methods for reaction solvent design.

levels of abstraction by optimizing a catalytic environment of reacting molecules. In early work, Tantillo *et al.* [53] propose a method called 'Theozymes' to study catalysis by enzymes. In Theozymes, the transition state is determined for the considered reaction. Subsequently, functional groups representing the catalyst are placed around the transition state and the positioning of these groups is optimized to minimize the activation barrier. Besides the original work of Tantillo *et al.*, there are more recent studies based on Theozymes [54]. Dittner and Hartke [55] study catalytic effects by optimizing an abstract catalytic environment around transition states. This approach is excellent to study catalytic effects but does not design actual catalyst molecules.

The methods discussed so far provide valuable insights into catalytic effects but different methods are required to design the molecular structure of catalyst molecules. Chang *et al.* [56] design Ni complexes to catalyze the CO/CO_2 conversion. Selected groups of the catalyst molecules are optimized to lower the activation energy of the rate-limiting reaction step. This activation energy is computed using semi-empirical tight binding linear combination of atomic potentials (TB-LCAP). In a refinement step after the design, the most promising catalyst candidates are further evaluated with DFT.

As for solvents, promising candidate catalysts may be missed if the optimization of catalyst and process is completely separated. Thus, we recently proposed a method for the integrated design of molecular catalysts and processes. Optimal catalyst molecules are designed based on a library of 3D molecule fragments. For process-based assessment, the process performance of every candidate catalyst is determined by individual process optimizations. The reaction kinetics achieved with the designed catalysts are calculated with transition state theory. The activation barriers as well as other required thermodynamic quantities are predicted with DFT, post-HF methods and COSMO-RS. The integration of catalyst and process design enables the direct design of catalyst structures catching all relevant tradeoffs to find optimal solutions.

Further developments are desirable towards reliable and broadly applicable design of molecular catalysts based on process performance. A major challenge for catalyst design is the search for the transition states required to use TST. Usually, QC methods need a good initial guess of the transition state geometry. Thus, current automated catalyst design methods are either limited to relatively small molecular design spaces or study an abstract catalytic environment without designing real catalyst structures. Methods that automate the generation of reaction networks and transition state geometries have already been developed [57–61] and work in specific cases. However, more generally applicable methods are needed for catalyst design.

Conclusion and outlook

In this opinion paper, we discuss CAMD methods based on QC. QC methods are not limited to fitted functional groups and can predict a broad range of properties independent of the availability of experimental data. Thus, QC methods are suited to explore vast molecular design spaces. Increasing computational power as well as the development of accurate and efficient QC methods has led to an increasing number of QC-based CAMD methods in recent years and we expect this trend to continue in the near future.

Two important applications of CAMD in chemical product design are highlighted: the design of solvents as well as the design of molecular catalysts. The design of separation solvents has been studied extensively leading to a large number of publications during the past decade, especially in the field of ILs for gas absorption. Large-scale screenings of thousands of known solvents can now be regarded as established due to available methods and software based on COSMO-RS/SAC. These screenings, however, usually employ thermodynamic properties as performance indicator. Integrating molecular design with process-based assessment of solvents has been demonstrated but efficient and standardized methods and tools are still missing.

Compared to the design of separation solvents, reaction solvent design has been studied less extensively. Transition state theory seems to provide the method of choice to predict reaction kinetics. Only few methods integrate the design of reaction solvents with process design. However, we consider the integration of solvent and process design as important future direction for both separation and reaction solvents. For GC-based CAMD, such integrated design tools are becoming available [62,63].

Compared to solvent design, CAMD of molecular catalysts is a quite unexplored field where we expect promising developments in the near future. There is still a major research need for reliable methods for automatic identification of transition states.

In our view, chemical engineers are only beginning to explore the power of quantum chemistry to design molecules. The spectrum of possible applications reaches much further than classical reaction and separation processes discussed here. Recently, QC-based prediction of environmental properties [64,65] has been demonstrated. Integration of these methods into CAMD would allow the design of environmentally friendly molecules. This paper will hopefully invite practitioners to employ available methods and researchers to push the frontier for CAMD using quantum chemistry.

Conflict of interest statement

Nothing declared.

Acknowledgements

C.G. gratefully acknowledges funding from the German Federal Ministry of Education and Research within the Project Carbon2Polymers (03EK30442C). K.L. and A.B. gratefully acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – Exzellenzcluster 2186 "The Fuel Science Center", ID: 390919832.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Levy P, Fernandez-Pales A, Vass T, Schroeder A, Baylin-Stern A, Kim T-Y, Feuvre PL, Bahar H, Ritchie J: *International Energy Agency*. 2019. [Online]. Available: https://www.iea.org/tcep/industry/chemical/#demand [Accessed 16 July 2019].

2. Grützner T, Schnider C, Zollinger D, Seyfang BC, Künzle N: **Reducing time to market by innovative development and production strategies**. *Chem Eng Technol* 2016, **39**:1835-1844.

3. Reymond J-L: **The chemical space project**. *Acc. Chem. Res.* 2015, **48**:722-730.

4. Fink T, Bruggesser H, Reymond JL: **Virtual exploration of the small-molecule chemical universe below 160 daltons**. *Angew Chem Int* 2005, **44**:1504-1508.

5. • Austin ND, Sahinidis NV, Trahan DW: **Computer-aided molecular design: An introduction and review of tools, applications, and solution techniques**. *Chem Eng Res Design* 2016, **116**:2-26.
Very instructive review of CAMD fundamentals.

6. • Ng KM, Gani R: **Chemical product design: Advances in and proposed directions for research and teaching**. *Comput Chem Eng* 2019, **126**:147-156.
Discussion of the state of the art and future directions for chemical product design including both research and teaching.

7. •• Papadopoulos AI, Tsivintzelis I, Linke P, Seferlis P: **Computer-Aided Molecular Design: Fundamentals, Methods, and Applications**. In *Elsevier Reference Module in Chemistry, Molecular Sciences and Chemical Engineering.* Edited by Reedijk J. Waltham, MA: Elsevier; 2018 http://dx.doi.org/10.1016/B978-0-12-409547-2.14342-2. 26-Jul-18.
Very comprehensive review of CAMD including property prediction, algorithms and applications.

8. • Gani R: **Group contribution-based property estimation methods: advances and perspectives**. *Curr Opin Chem Eng* 2019, **23**:184-196.
Discussion of state of the art group contribution methods, current limitations and future prospects.

9. Gmehling J: **Present status and potential of group contribution methods for process development**. *J Chem Thermodyn* 2009, **41**:731-747.

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

11. Adjiman CS, Galindo A, Jackson G: **Molecules matter: the expanding envelope of process design**. *Comput Aided Chem Eng* 2014, **34**:55-64.

12. Zhang L, Babi DK, Gani R: **New vistas in chemical product and process design**. *Annu Rev Chem Biomol Eng* 2016, **7**:557-582.

13. Gani R: **Chemical product design: challenges and opportunities**. *Comput Chem Eng* 2004, **28**:2441-2457.

14. Vereecken L, Glowacki DR, Pilling MJ: **Theoretical chemical kinetics in tropospheric chemistry: methodologies and applications**. *Chem Rev* 2015, **115**:4063-4114.

15. Mata RA, Suhm MA: **Benchmarking quantum chemical methods: are we heading in the right direction?** *Angew Chem Int Ed* 2017, **56**:11011-11018.

16. Zheng J, Zhao Y, Truhlar DG: **The DBH24/08 database and its use to assess electronic structure model chemistries for chemical reaction barrier heights**. *J Chem Theory Comput* 2009, **5**:808-821.

17. Řeza'č J, Hobza P: **Describing noncovalent interactions beyond the common approximations: how accurate is the "Gold Standard," CCSD (T) at the complete basis set limit?** *J Chem Theory Comput* 2013, **9**:2151-2155.

18. Bretonnet JL: **Basics of the density functional theory**. *AIMS Mater Sci* 2017, **4**:1372-1405.

19. Goerigk L, Hansen A, Bauer C, Ehrlich S, Najibi A, Grimme S: **A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions**. *Phys Chem Chem Phys* 2017, **19**:32184-32215.

20. Bannwarth C, Ehlert S, Grimme S: **GFN2-xTB—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-**

dependent dispersion contributions. *J Chem Theory Comput* 2019, **15**:1652-1671.

21. Paulechka E, Kazakov A: **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* 2017, **121**:4379-4387.

22. Umer M, Leonhard K: **Ab initio calculations of thermochemical properties of methanol clusters**. *J Phys Chem A* 2013, **117**:1569-1582.

23. Marenich AV, Cramer CJ, Truhlar DG: **Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions**. *J Phys Chem B* 2009, **113**:6378-6396.

24. Hsieh C-H, Sandler SI, Lin S-T: **Improvements of COSMO-SAC for vapor–liquid and liquid–liquid equilibrium predictions**. *Fluid Phase Equilibria* 2010, **297**:90-97.

25. • Nieto-Draghi C, Fayet G, Creton B, Rozanska X, Rotureau P, de Hemptinne JC, Ungerer P, Roussau B, Adamo C: **A general guidebook for the theoretical prediction of physicochemical properties of chemicals for regulatory purposes**. *Chem Rev* 2015, **115**:13093-13164.

Instructive introduction to QSPR and comprehensive review of property prediction with QSPRs.

26. Zhou T, McBride K, Zhang X, Qi Z, Sundmacher K: **Integrated solvent and process design exemplified for a Diels–Alder reaction**. *AIChE J* 2015, **61**:147-158.

27. Kapusta K, Sizochenko N, Karabulut S, Okovytyy S, Voronkov E, Leszczynski J: **QSPR modeling of optical rotation of amino acids using specific quantum chemical descriptors**. *J Mol Model* 2018, **24**:59.

28. Weinebeck A, Kaminski S, Murrenhoff H, Leonhard K: **A new QSPR-based prediction model for biofuel lubricity**. *Tribol Int* 2017, **115**:274-284.

29. De Benedetti PG, Fanelli F: **Multiscale quantum chemical approaches to QSAR modeling and drug design**. *Drug Discov Today* 2014, **19**:1921-1927.

30. Farahipour R, Mehrkesh A, Karunanithi AT: **A systematic screening methodology towards exploration of ionic liquids for CO_2 capture processes**. *Chem Eng Sci* 2016, **145**:126-132.

31. Ahmad MZ, Hashim H, Mustaffa AA, Maarof H, Yunus NA: **Design of energy efficient reactive solvents for post combustion CO_2 capture using computer aided approach**. *J Clean Prod* 2018, **176**:704-715.

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

33. Yu G, Dai C, Wu L, Lei Z: **Natural gas dehydration with ionic liquids**. *Energy Fuels* 2017, **31**:1429-1439.

34. Fang J, Zhao R, Su W, Li C, Liu J, Li B: **A molecular design method based on the COSMO-SAC model for solvent selection in ionic liquid extractive distillation**. *AIChE J* 2016, **62**:2853-2869.

35. Song Z, Hu X, Zhou Y, Zhou T, Qi Z, Sundmacher K: **Rational design of double salt ionic liquids as extraction solvents: separation of thiophene/n-octane as example**. *AIChE J* 2019, **65** e16625.

36. Shang D, Liu X, Bai L, Zeng S, Xu Q, Gao H, Zhang X: **Ionic liquids in gas separation processing**. *Curr Opin Green Sustain Chem* 2017, **5**:74-81.

37. Scheffczyk J, Redepenning C, Jens CM, Winter B, Leonhard K, Marquardt W, Bardow A: **Massive, automated solvent screening for minimum energy demand in hybrid extraction–distillation using COSMO-RS**. *Chem Eng Res Des* 2016, **115**:433-442.

38. Austin ND, Sahinidis NV, Trahan DW: **A COSMO-based approach to computer-aided mixture design**. *Chem Eng Sci* 2017, **159**:93-105.

39. Zhang J, Peng D, Song Z, Zhou T, Cheng H, Chen L, Qi Z: **COSMO-descriptor based computer-aided ionic liquid design for separation processes. Part I: modified group contribution methodology for predicting surface charge density profile of ionic liquids**. *Chem Eng Sci* 2017, **162**:355-363.

40. Peng D, Zhang J, Cheng H, Chen L, Qi Z: **Computer-aided ionic liquid design for separation processes based on group contribution method and COSMO-SAC model**. *Chem Eng Sci* 2017, **159**:58-68.

41. Bechtel S, Song Z, Zhou T, Vidakovic-Koch T, Sundmacher K: **Integrated process and ionic liquid design by combining flowsheet simulation with quantum-chemical solvent screening**. *Comput Aided Chem Eng* 2018, **44**:2167-2172.

42. Song Z, Zhou T, Qi Z, Sundmacher K: **Systematic method for screening ionic liquids as extraction solvents exemplified by an extractive desulfurization process**. *ACS Sustain Chem Eng* 2017, **5**:3382-3389.

43. Redepenning C, Recker S, Marquardt W: **Pinch-based shortcut method for the conceptual design of isothermal extraction columns**. *AIChE J* 2017, **63**:1236-1245.

44. • Scheffczyk J, Schfer P, Fleitmann L, Thien J, Redepenning C, Leonhard K, Marquardt W, Bardow A: **COSMO-CAMPD: a framework for integrated design of molecules and processes based on COSMO-RS**. *Mol Syst Design Eng* 2018, **3**:645-657.

Integrated design of solvents and processes including experimental validation.

45. • Struebing H, Ganase Z, Karamertzanis PG, Siougkrou E, Haycock P, Piccione PM, Armstrong A, Galindo A, Adjiman CS: **Computer-aided molecular design of solvents for accelerated reaction kinetics**. *Nat Chem* 2013, **11**:952.

Pioneering work on QC-based reaction solvent design.

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

47. Austin ND, Sahinidis NV, Konstantinov IA, Trahan DW: **COSMO-based computer-aided molecular/mixture design: a focus on reaction solvents**. *AIChE J* 2018, **64**:104-122.

48. Gertig C, Kröger LC, Fleitmann L, Scheffczyk J, Bardow A, Leonhard K: **Rx-COSMO-CAMD: computer-aided molecular design of reaction solvents based on predictive kinetics from quantum chemistry**. *Ind Eng Chem Res* 2019, **58**:22835-22846 http://dx.doi.org/10.1021/acs.iecr.9b03232.

49. • Zhou T, Lyu Z, Qi Z, Sundmacher K: **Robust design of optimal solvents for chemical reactionsA combined experimental and computational strategy**. *Chem Eng Sci* 2015, **137**:613-625.

Robust optimization of solvents taking the uncertainties in property prediction into account.

50. Gertig C, Leonhard K, Bardow A: **Integrated design of solvents and processes based on reaction kinetics from quantum chemical prediction methods**. *Comput Aided Chem Eng* 2019, **46**:415-420.

51. Poree C, Schoenebeck F: **A holy grail in chemistry: computational catalyst design: feasible or fiction?** *Acc Chem Res* 2017, **50**:605-608.

52. Freeze JG, Kelly HR, Batista VS: **Search for catalysts by inverse design: artificial intelligence, mountain climbers, and alchemists**. *Chem Rev* 2019, **119**:6595-6612.

53. Tantillo DJ, Jiangang C, Houk KN: **Theozymes and compuzymes: theoretical models for biological catalysis**. *Curr Opin Chem Biol* 1998, **2**:743-750.

54. Hare SR, Pemberton RP, Tantillo DJ: **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* 2017, **139**:7485-7493.

55. Dittner M, Hartke B: **Globally optimal catalytic fields–inverse design of abstract embeddings for maximum reaction rate acceleration**. *J Chem Theory Comput* 2018, **14**:3547-3564.

56. Chang AM, Rudshteyn B, Warnke I, Batista VS: **Inverse design of a catalyst for aqueous CO/CO_2 conversion informed by the Nill–iminothiolate complex**. *Inorg Chem* 2018, **57**:15474-15480.

57. Döntgen M, Przybylski-Freund MD, Kröger LC, Kopp WA, Ismail AE, Leonhard K: **Automated discovery of reaction pathways, rate constants, and transition states using reactive molecular dynamics simulations**. *J Chem Theory Comput* 2015, **11**:2517-2524.

58. Döntgen M, Schmalz F, Kopp WA, Kröger LC, Leonhard K: **Automated chemical kinetic modeling via hybrid reactive molecular dynamics and quantum chemistry simulations**. *J Chem Inf Model* 2018, **58**:1343-1355.

59. Simm GN, Vaucher AC, Reiher M: **Exploration of reaction pathways and chemical transformation networks**. *J Phys Chem A* 2018, **123**:385-399.

60. Dewyer AL, Argüelles AJ, Zimmermann PM: **Methods for exploring reaction space in molecular systems**. *WIREs Comput Mol Sci* 2018, **8** e1354.

61. Grambow CA, Jamal A, Li Y-P, Green WH, Zádor J, Suleimanov YV: **Unimolecular reaction pathways of a γ-ketohydroperoxide from combined application of automated reaction discovery methods**. *J Am Chem Soc* 2018, **140**:1035-1048.

62. Schilling J, Gross J, Bardow A: **Integrated design of ORC process and working fluid using process flowsheeting software and PC-SAFT**. *Energy Procedia* 2017, **129**:129-136.

63. Liu Q, Zhang L, Liu L, Du J, Tula AK, Eden M, Gani R: **OptCAMD: an optimization-based framework and tool for molecular and mixture product design**. *Comput Chem Eng* 2019, **124**:285-301.

64. Kleinekorte J, Kröger L, Leonhard K, Bardow A: **A neural network-based framework to predict process-specific environmental impacts**. *Comput Aided Chem Eng* 2019, **46**:1447-1452.

65. Calvo-Serrano R, González-Miquel M, Guillén-Gosálbez G: **Integrating COSMO-based σ-profiles with molecular and thermordynamic attributes to predict the life cycle environmental impact of chemicals**. *ACS Sustain Chem Eng* 2019, **7**:3575-3583.