Chem Soc Rev

REVIEW ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Chem. Soc. Rev.*, 2018, **47**, 8307

Towards *operando* computational modeling in heterogeneous catalysis

Lukáš Grajciar, [a] Christopher J. Heard, [a] Anton A. Bondarenko, [b] Mikhail V. Polynski, [b] Jittima Meeprasert, [c] Evgeny A. Pidko *[bc] and Petr Nachtigall *[a]

An increased synergy between experimental and theoretical investigations in heterogeneous catalysis has become apparent during the last decade. Experimental work has extended from ultra-high vacuum and low temperature towards *operando* conditions. These developments have motivated the computational community to move from standard descriptive computational models, based on inspection of the potential energy surface at 0 K and low reactant concentrations (0 K/UHV model), to more realistic conditions. The transition from 0 K/UHV to *operando* models has been backed by significant developments in computer hardware and software over the past few decades. New methodological developments, designed to overcome part of the gap between 0 K/UHV and *operando* conditions, include (i) global optimization techniques, (ii) *ab initio* constrained thermodynamics, (iii) biased molecular dynamics, (iv) microkinetic models of reaction networks and (v) machine learning approaches. The importance of the transition is highlighted by discussing how the molecular level picture of catalytic sites and the associated reaction mechanisms changes when the chemical environment, pressure and temperature effects are correctly accounted for in molecular simulations. It is the purpose of this review to discuss each method on an equal footing, and to draw connections between methods, particularly where they may be applied in combination.

Received 14th May 2018

DOI: 10.1039/c8cs00398j

rsc.li/chem-soc-rev

[a] *Department of Physical and Macromolecular Chemistry, Faculty of Science, Charles University in Prague, 128 43 Prague 2, Czech Republic. E-mail: lukas.grajciar@natur.cuni.cz, petr.nachtigall@natur.cuni.cz, heardc@natur.cuni.cz*

[b] *TheoMAT group, ITMO University, Lomonosova 9, St. Petersburg, 191002, Russia*

[c] *Inorganic Systems Engineering group, Department of Chemical Engineering, Faculty of Applied Sciences, Delft University of Technology, Van der Maasweg 9, 2629 HZ Delft, The Netherlands. E-mail: e.a.pidko@tudelft.nl*

Lukáš Grajciar

Lukáš Grajciar received his MSc and PhD degrees in chemistry from the Charles University in Prague in 2009 and 2013, respectively, developing and applying dispersion-corrected DFT methods for adsorption in zeolites and metal–organic frameworks. At his postdoctoral position at Jena University in Germany, he became involved in development of high-performance algorithms for ab initio treatment of large molecules and periodic system within the TURBOMOLE program, including implementation of a new tool for global structure optimization of clusters in confinement. Currently, he is a researcher at the Charles University in Prague, investigating reactivity of zeolites using biased ab initio molecular dynamics.

Christopher J. Heard

Christopher Heard is a postdoctoral researcher at the Charles University in Prague. As part of the CUCAM project, he currently investigates the stability and reactivity of zeolitic and layered oxide materials with ab initio thermodynamic techniques. Previously, he completed his BA and MSci (2010) at the University of Cambridge, followed by PhD studies under Prof. Roy Johnston at the University of Birmingham (2014), developing and employing global optimization tools for free and oxide-supported metal clusters. This was followed by a postdoctoral position in computational surface science at Chalmers University in Sweden (until 2016), involving heterogeneous catalysis and microkinetic modelling.

1. Introduction

Most of the chemicals produced nowadays are obtained using processes based on catalysis. The on-going search for optimal process conditions and the most suitable catalyst is driven by various concerns, including (i) environmental impact, (ii) resource utilization, (iii) safety and (iv) overall process economy. While this has traditionally been the domain of experimental investigations, the input from computational investigations has been steadily increasing over the last 40 years. An increased synergy between theory and experiment has become apparent during the last decade, in particular, in the field of heterogeneous catalysis.

By definition a heterogeneous catalyst shifts the reference reaction onto a different free energy surface where the energy of critical transition states with respect to relevant intermediates becomes lower. Mechanisms of chemical reactions were traditionally explored within the concept of the potential energy surface (PES), considering simplified models of a catalytic system working under idealized conditions of, basically, infinite dilution. Such a heterogeneous catalysis model represents ultra-high vacuum conditions, for which calculations provide information at 0 K; we will refer to this model as the 0 K/UHV model. Strictly speaking, such a description corresponds to rather unrealistic reaction conditions and its validity decreases with increasing temperature and pressure. A great number of mechanisms have been proposed based on calculations with such a simplistic model and results were often at least in qualitative agreement with available experimental data. Computational results obtained with 0 K/UHV model correspond reasonably well with experimental data obtained for well-defined surfaces under UHV conditions. However, the overlap of such calculated data and catalytic experiments carried out under realistic conditions is

Anton A. Bondarenko

Anton A. Bondarenko was born in Vyborg, Russia, in 1995. He graduated from Saint-Petersburg State University with a bachelor's degree in biology in 2017 and enrolled to master's programme at ITMO University the same year, where he joined the group of theoretical chemistry (TheoMAT) headed by Prof. Pidko. His research interests are machine learning, chemoinformatics, and automation of chemical calculations.

Mikhail V. Polynski

Mikhail V. Polynski (Moscow, Russia, 1990) graduated from the Higher Chemical College of the Russian Academy of Sciences in 2013. In 2013–2017 he followed a PhD program at Lomonosov Moscow State University and carried out research in computational chemistry and catalysis at Zelinsky Institute of Organic Chemistry, Moscow, under the guidance of Prof. Valentine Ananikov. Since Fall 2017 he has been assisting Prof. Pidko in building and leading the theoretical chemistry group (TheoMAT) at IITMO University, St. Petersburg, Russia. His main research interests are automation of computational chemistry research, ab initio MD, and theory of catalysis.

Jittima Meeprasert

Jittima was born in Bangkok, Thailand. She received her BS degree in General Science and MS degree in Chemistry from Kasetsart University. She then worked as a research assistant in Nanoscale Simulation Laboratory at the National Nanotechnology Center. She is currently pursuing a PhD degree under the guidance of Professor Evgeny Pidko in the Department of Chemical Engineering at Delft University of Technology, The Netherlands. Her research interest focuses on computational heterogeneous catalysis.

Evgeny A. Pidko

Evgeny A. Pidko (Moscow, Russia, 1982) received his PhD from Eindhoven University of Technology in 2008, wherein in 2011–2017 he was an Assistant Professor of Catalysis for Sustainability. Since 2016 he has been a part-time professor of theoretical chemistry at ITMO University, St. Petersburg. Since Fall 2017 he has been an Associate Professor and head of the Inorganic Systems Engineering group at Delft University of Technology. In his research he combines theory and experiment to study mechanisms of homogeneous and heterogeneous catalysts and guide the development of new and improved catalyst systems relevant to sustainable chemistry and energy technologies.

rather small, and a good agreement between 0 K/UHV theory and catalytic experiments was often just fortuitous.

The success of the simple PES concept applied within the 0 K/UHV approximation can be expected only when the following assumptions hold: (i) the structure of the active site under realistic conditions is known (or correctly guessed), (ii) both the structure of the active site and the reaction mechanism do not depend on the surface coverage of individual reaction intermediates, (iii) the reaction mechanism found under nearly UHV conditions is not different from that at the realistic composition of the surrounding gas or liquid phase and (iv) temperature effects, including the transition from PES to free energy surface (FES), can be safely neglected. Unfortunately, all such assumptions are rarely satisfied at once. If the temperature is relatively low it follows that reactants, products and/or reaction intermediates are adsorbed on the surface; and in contrast, one can expect that the reaction proceeds on a clean catalyst surface only at elevated temperature.

A deeper atomistic insight into the reaction mechanisms, the catalyst structure/activity relationship and catalyst stability/transformation during the reaction greatly increases our chances to find the optimal catalyst for a particular process. The most detailed experimental evidence about the catalyst at the molecular level can be obtained by a combination of characterization techniques under UHV conditions. More and more information becomes available from experimental investigations gathered under the conditions of a model catalytic reaction – *in situ* conditions – and also under conditions where the applied catalytic process takes place – *operando* conditions. For details of experimental *in situ* and *operando* conditions see, *e.g.*, ref. 1–3. A great development of *in situ* and in particular *operando* experimental techniques for studying catalytic reactions in the last 20 years has brought an increasing amount of information about the state of the catalysts under realistic conditions.[4,5]

Among the most important findings emerging from such studies is the evidence of the dynamic nature of the catalyst surface, whose structure constantly changes under the catalytic reaction conditions. For example, in oxidation catalysis by supported metal nanoparticles, *in situ* and *operando* techniques revealed the formation of ultra-thin oxide layers covering the metal nanoparticles in an oxidizing atmosphere, which provide the active sites for the target catalytic reactions. Obviously, such an active site model could not be proposed based on the UHV surface science experiments or computations carried out in the 0 K/UHV regime. A problem of how the structure of the catalyst depends on the realistic chemical environment and temperature that are relevant for a particular process is thus the key for a proper understanding of catalysis at the molecular level and for a design of improved catalysts.[6–8]

Similar to the shift of experimental investigations in catalysis from UHV to *operando* conditions, theoretical investigations in the field of catalysis are moving more and more from 0 K/UHV models to computational *operando* investigations. In analogy with the experimental *operando* conditions, a computational *operando* model is defined by the following conditions: the structure of the active catalyst surface and the reaction coordinates must reflect realistic conditions during the reaction and a complex reaction network must be established (see Fig. 1 and corresponding text for more details). However, a transition from the 0 K/UHV to *operando* model dramatically influences the complexity of the problem and increases computational demands. A number of methods have been developed in the past few decades that ease the 0 K/UHV → *operando* transition and it is the goal of this review to discuss the current state of the computational investigations of catalysis, with the goal to enable the long-sought after paradigm of catalysis by design.

A huge gap between the 0 K/UHV models on one side and *operando* models on the other side cannot be overcome by a single computational method that would explicitly account for the whole complexity of the underlying phenomena. A multiscale modeling approach can be followed to construct a composite methodology that includes all the crucial physical phenomena. In our opinion, the following five methods appear to be the most important for bridging this gap: (i) global optimization techniques, (ii) *ab initio* constrained thermodynamics, (iii) biased MD simulations, (iv) microkinetic models of reaction networks. The fifth class of methods is a conceptually different approach that does not necessarily imply the explicit account of the complex physics of a catalyst system and yet holds great promise as a tool to enable catalysis by design. This class is the broad family of machine learning methods. The latest development of each of these five techniques is addressed individually in the following five sections of this review.

A transition from the 0 K/UHV to *operando* model is schematically depicted in Fig. 1. The 0 K/UHV model corresponds to the situation at the lower left corner, corresponding to vanishing partial pressures of reaction components (expressed in terms of chemical potentials) and low temperature. The *operando* model corresponds to the upper right corner. Going from bottom to top of the figure the reaction environment (in terms of chemical potentials and temperature) becomes more realistic. Any model improvement results in the increased complexity of the problem

Petr Nachtigall

Prof. Petr Nachtigall completed his PhD in 1995 at the University of Pittsburgh. He then moved to Prague where he held a research position in Academy of Sciences of Czech Republic. In 2010 he moved to the Faculty of Science of the Charles University where he is currently Head of the Department of Physical and Macromolecular Chemistry. His research is focused on the theoretical investigation of surface properties of solids, related mainly to gas adsorption and catalytic processes involving microporous and nano-structured materials.

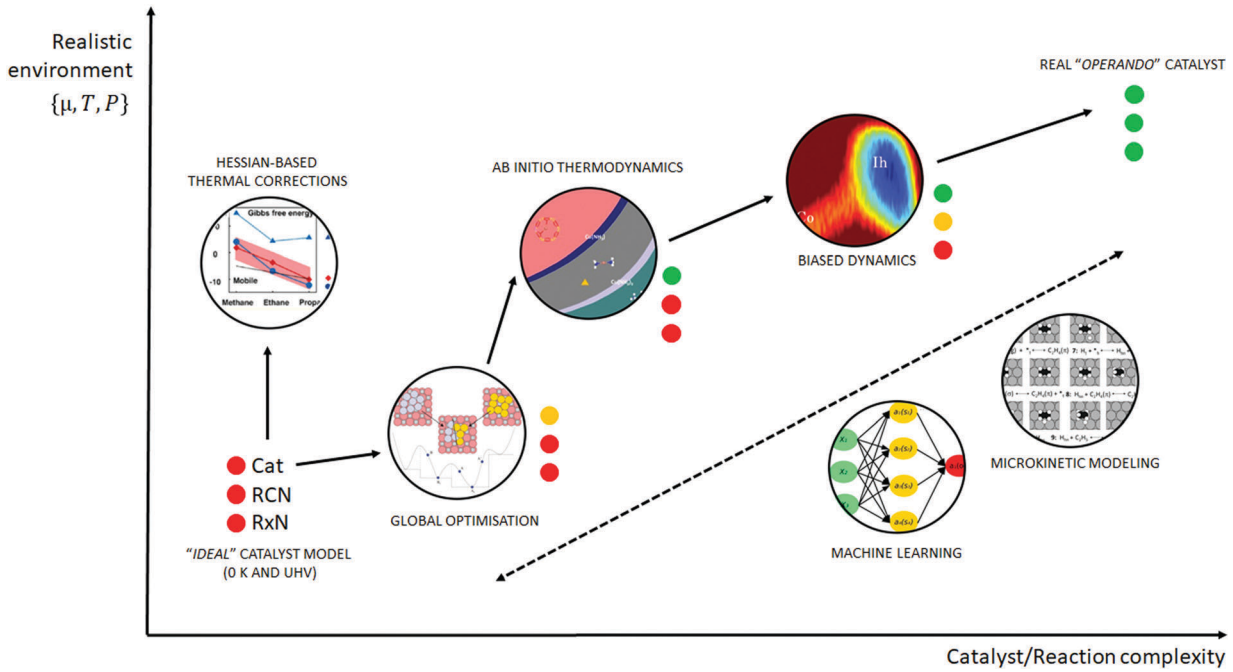

Fig. 1 Schematic of the various computational methods applied to heterogeneous catalysis, which lie between an idealised UHV model and a realistic, *operando* model. The traffic light key depicts the quality of each method with respect to catalyst model complexity (Cat), reaction coordinate accuracy (RCN) and reaction network complexity (RxN). Adapted with permission from Piccini *et al.*, *Journal of Physical Chemistry C*, 2015, **119**, 6128–6137, Copyright 2015, American Chemical Society, Vilhelmsen *et al.*, *Journal of Chemical Physics*, 2014, **141**, 044711, Copyright 2014, American Institute of Physics, Chen *et al.*, *Journal of Catalysis*, 2018, **358**, 179–186, Copyright 2018, Elsevier, Pavan *et al.*, *Journal of Chemical Physics*, 2015, **143**, 184304, Copyright 2015, American Institute of Physics, Heard *et al.*, *ACS Catalysis*, 2016, **6**, 3277–3286, Copyright 2016, American Chemical Society.

(from left to right), mostly in the number of configurations that are considered. Basics of the 0 K/UHV model include the following approximations: (i) idealized catalyst surface (denoted as Cat in Fig. 1), (ii) idealized reaction coordinates with minimum number of reactants on the PES at 0 K (reaction coordinate environment – RCE) and (iii) elementary reaction steps are considered (reaction network – RxN). All these approximations must be lifted to move forward to an *operando* model.

Methods presented in Fig. 1 from left to right start with Hessian-based thermal corrections, followed by a global optimization approach, *ab initio* constrained thermodynamics and biased MD; microkinetic modeling and machine learning techniques are taken off this order since they can be used at any level of the 0 K/UHV → *operando* transition. The order presented in Fig. 1 is motivated by the fact that if all extensions are applied for a particular system, they would be applied in the order presented in the figure, with the exception of Hessian-based thermal corrections. Hessian-based thermal corrections allow a proper transition from potential- to free-energy surfaces while the complexity of the system remains unchanged; they can be used either to improve the 0 K/UHV model or in combination with global optimization or *ab initio* constrained dynamics (improving the quality of partition functions). It is important to note that it is common to apply just one or two extensions (or even three in some cases) and by no means does it have to be the first methods from left to right. For example, it is rather common to combine Hessian-based thermal corrections directly with microkinetics. It depends on the particular problem under investigation as to which of the extensions is crucial. Global optimization techniques mostly help in finding relevant configurations when these are difficult or impossible to obtain from relevant experimental data. *Ab initio* thermodynamics is critically important for the investigation of catalyst surfaces that are changed in the reaction environment. Biased molecular dynamics (MD) techniques become essential for the localization of transition states in complex environments when these are strongly affected by the surrounding molecules. Microkinetic modeling of the reaction network is essential for situations in which a large number of reaction intermediates exist. Last but not least, machine learning techniques are emerging as a useful tool in rationalization of the system descriptors and finding important correlations in large data sets.

Each of the methods presented in Fig. 1 is designed to overcome part of the gap between 0 K/UHV and *operando* conditions. Each method is discussed in the following sections and each of the methods has been reviewed separately in recent years in a comprehensive way. It is the purpose of this review to discuss them on an equal footing with respect to the gap between 0 K/UHV and *operando*. It should be stressed that the simultaneous application of all these extensions is computationally prohibitive in a general sense. But it should be noted that it is often not necessary to apply all these model extensions for a particular catalytic system; instead it is important to identify which of the extensions is critical for the problem investigated.

2. Global optimization

2.1. Basic principles

Global optimization (GO) is a class of heuristic methods used to search the breadth of a cost function $E(\boldsymbol{X})$ defined by the multi-dimensional vector $\boldsymbol{X}$. The goal is to locate the local minimum $\tilde{\boldsymbol{X}}$, which globally minimizes that function, such that $E(\tilde{\boldsymbol{X}}) = \min[E(\boldsymbol{X})]$. In the case of computational chemistry, $\boldsymbol{X}$ is usually a 3N dimensional vector of the atomic coordinates of an N atom system, which returns the potential energy of the configuration. Thus, GO is aimed at finding the structure which has the lowest potential energy. The dimensionality of the landscape to be sought is often reduced considerably from 3N, either by constraining the position of certain atoms during local minimisation, by combining positions into collective variables (as in metadynamics, described in detail in Section 4), or by defining symmetry classes to group like atoms together. As it is impossible in practice to ensure that a putative global minimum is the true global minimum for all but trivial systems, convergence criteria on GO searches are applied, based either on stagnation of the diversity of structures, or some practical consideration, such as the number of minima found. The moveclass, by which the search of the energy landscape is undertaken, is another heuristic choice, described for a number of GO methods in Section 2.2. Combining a moveclass, a local minimisation algorithm, an acceptance criterion for new structures, and a convergence criterion for the search, a GO method can locate a library of low-lying minima relevant for the physico-chemical property of interest, or deliver the global minimum, to be used as the best estimate of the preferred structure of a system.

2.2. Global optimization methods

In condensed matter physics, we are primarily concerned with finding the stable phases of materials, and as such GO has been widely applied. GO provides a library of low energy configurations, which are useful for describing systems with strong structure–function relationships. In heterogeneous catalysis, the picture is more complicated, as catalytic reactions are often controlled kinetically, rather than thermodynamically, and involve transient species and dynamic restructuring of the active phase. Generally, low temperature, low pressure environments are most accurately reproduced by GO techniques. Hence, from the first reports of robust methods in the 1990s, GO has been developed and applied successfully in heterogeneous catalysis research in three main areas: (i) the optimization of vacuum phase nanoparticle and cluster structures, (ii) determination of stable, catalytically important surfaces and (iii) the adsorption, growth and migration of active catalytic particles upon substrates, which usually aim to connect to gas phase spectroscopy or surface science experiments. The variety of GO techniques will be covered in the following section, supported by appropriate examples that are relevant for catalysis. For systems which change strongly in structure or composition during a reaction, or interact strongly with the environment, GO is less valuable, which is why it is seldom applied, for example, to *operando* descriptions of systems with complex solvation chemistry.

A good GO method must balance the local and global aspects of searching the energy landscape. Local optimization methods serve to locate the configuration which corresponds to the local minimum of the potential energy well to which the current configuration belongs. The computational methods to achieve this are numerous and robust.[9,10] However, the global search is required to explore the breadth of the energy landscape efficiently, so as to capture all relevant structural classes. This is performed in a heuristic, system-specific manner. In a recent article by Jørgensen and colleagues, the balance between efficient global and local search is recast into the concept of "exploration *versus* exploitation".[11] They find that the optimal balance between exploration (finding new regions of configuration space) and exploitation (exhausting the local region to find all nearby low-lying minima) can enhance the GO efficiency for molecular structures. By contrast, it is found to be less powerful for surface GO, because the possible configurations are strongly templated by the layers below. Several good reviews exist for detailed examination of the technical aspects of global landscape search and optimization methods.[12,13] We will give a brief introduction to the more popular techniques, before describing the catalytic applications in more detail.

Basin-hopping (BH)[14] is a Monte Carlo based global optimization technique, and has a long history of application to materials science. This method belongs to a class of energy landscape-simplifying hypersurface deformation techniques, which remove barriers to energetically downhill steps, and vastly improve the ergodicity of the exploration. Extensive modifications to the original BH method have been developed since the late 1990s. One of the most notable examples is minima hopping (MH), from Goedecker,[15] which applies short bursts of molecular dynamics (MD) simulations between local optimization steps, with a variable temperature parameter to allow for escape from deep basins. This technique has been widely used, for example, in GO for large gold clusters (up to Au_{318}),[16] the discovery of a new, photocatalytically promising titania nanosheet isomer[17] and in determining the role of solvating water in electrochemical water oxidation catalysis over $IrO_2(110)$ (see Fig. 2).[18] Sicher found that the MD moves in MH are more efficient than saddle point search methods for escaping minima, requiring fewer force calculations to achieve the same success rate.[19,20] Another development is the parallel excitable walkers (PEW) method of Rossi,[21] which combines a modified tabu search to avoid stagnation in previously visited basins, with the benefits of multiple simultaneous searches. The walkers move in parallel on the same energy landscape and avoid sampling the same region of configuration space by dynamically repelling each other. Walkers are determined to be neighbours based on an order parameter for structural similarity. If two walkers are too close together, the Metropolis ratio is shifted to allow for more unfavourable uphill steps to be accepted. The advantage of the method over traditional tabu sampling is that isolated walkers retain the sampling efficiency of basin hopping, without wasting cpu time rejecting steps due to fixed energy penalties. As an example, Ferrando and coworkers have applied this method for the geometry optimization of binary transition metal nanoalloy clusters.[21,22]

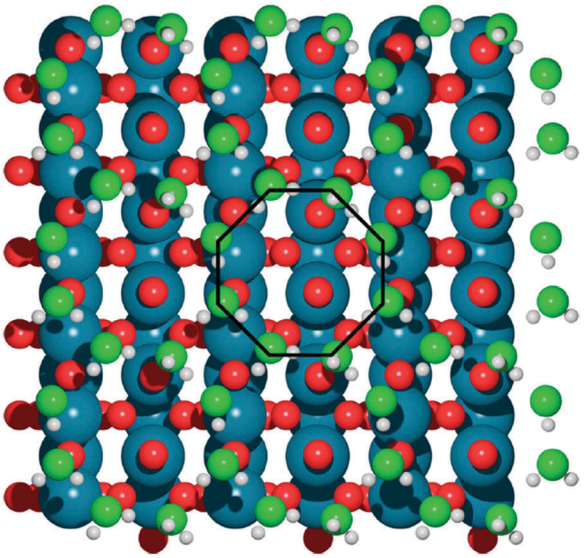

Fig. 2 The optimal, octahedral structure of a water bilayer upon IrO_2, determined with DFT minima-hopping. Water oxygen atoms are light green, surface oxygen atoms are red, iridium is blue and hydrogen is white. Reprinted with permission from ref. 18. Copyright 2017 American Chemical Society.

Nature inspired GO methods have also been extensively applied to catalytically relevant systems. These methods include, but are not limited to ant colony, artificial bee colony and particle swarm methods.[13] The most popular class of nature inspired methods by far has been the genetic or evolutionary algorithm (EA).[12,23,24] EAs employ a series of steps which mimic changes that occur to a population of individuals in nature, such as mating and mutation.[25] These steps interchange and modify the structures of a population of trial solutions on a generational basis, until convergence to the putative globally optimal structure. Convergence is defined heuristically, either by a preset choice of generations[26] or some stagnation criterion in either energies or structures.[23] Very recently, machine learning techniques have begun to be implemented into the selection step of EAs. Jørgensen *et al.* introduced a clustering method to the EA maintained within the atomic simulation environment (ASE).[27] This technique groups together elements of the population into classes based on a structural similarity metric. The selection of the next generation of structures is thus biased to focus on promising regions of configuration space. Two algorithms have been tested: one in which unexplored structural classes are biased towards, and one in which over-explored regions are biased against. Clustering was successfully applied both to gas phase molecules and the catalytically important anatase TiO_2(001) surface.

In both Monte Carlo and nature-inspired GO strategies, the choice of moveclass for exploring the energy landscape is crucial to the efficiency of the search, and much of the improvement of these methods comes from designing and combining effective moveclasses. For example, SHAKE moves[28] for both EA and BH methods involve moving all atoms by some fixed amount. A variant of the SHAKE move was introduced by Kim *et al.*,[29] in which each atom is moved during the BH step, but within a displacement range that is a function of the distance of the atom from the centre of the system. This modification mimics the increased diffusivity of atoms near or at the surface of a particle, and was found to enhance the efficiency of the GO search by a factor of 3.8 over standard BH for an AuPd cluster. Continuous symmetrisation techniques have been developed by several groups,[30,31] which improve search efficiency by biasing moves towards the completion of high symmetry structures. Schönborn *et al.* have reported the effectiveness of an "average offspring" EA mating method.[32] In this strategy, two parents are selected. For each atom of the first parent, the position of the nearest atom in the second parent is located, and the child is assigned an atom at the average of those positions. More recently, Vegge used radial cuts to improve the search for binary particles which show a tendency towards core–shell structures.[33] The optimal choice of moveclass is thus arrived at heuristically, but allows for flexibility in the GO method, to treat many types of system efficiently.

Moving beyond simply the location of the global minimum, pathway sampling methods combine discovery of the low-lying regions of the energy landscape with identification of paths that connect the minima together. In this way, one can move from static structure prediction to estimation of chemical properties. As with GO methods, the automated nature of path sampling allows for the avoidance of the biases of prior human intuition. Much work has gone into improving path sampling methods in two main ways. First are the algorithms which connect minima, such as eigenvector-following to follow soft-mode pathways, and their hybrid implementations.[13,34] Nudged elastic band (NEB) methods also belong to this group,[35] as do string methods.[36–38] Second, there are the global searches which utilise these minimum-connecting steps. Discrete path sampling is one example,[13,39] in which single or double ended path searches aim to find and connect adjacent minima and build up a picture of the energy landscape in an automatic manner. These search methods have even been applied beyond catalysis, for example, in recent work on the migration of lithium cations in the $Li_{0.5}MnO_2$ battery materials.[40] Another related example is the minima hopping guided pathway approach of Schäfer *et al.*[41] The stochastic surface walking method of Zhang *et al.* is a promising variant which was designed specifically for chemical reactions.[42] Connections between minima which are defined as reactants and products according to criteria such as bond connectivity are discovered so as to target promising reaction paths. The search mode may be biased according to particular reaction coordinates to speed up the search. This method has been recently applied to the water gas shift reaction on Cu(111), isolating a new mechanism for formic acid formation.

2.3. Free-standing particles

The simplest model for a catalytic nanoparticle is that of a free-standing cluster in a vacuum. This approximation is reasonable either as a first order interpretation of the particle under inert atmospheres, or under the assumption of ultrasoft landing on inert supports. Of course, in most applications, the role of surface, solvation and ligands is important. Nevertheless, a

great deal can be learned from knowledge of the geometric and electronic structure of vacuum-phase clusters, and this has traditionally been the starting point for nano-catalysis GO.

Early studies investigated the structures of model particles on the order of 100 atoms, utilising a range of empirically parameterized potentials (EP), such as Gupta[43,44] and others.[16,45] Focussing mainly on the structural motifs favoured by various mono- and bimetallic clusters, these studies revealed a complex landscape of preferred structures in the non-scalable size regime from sub-nm to a few nm, including disordered morphologies, polyicosahedra, prolate disk-like structures[44,46] and strained, even chiral particles.[47] Increasing computational resources and the development of robust density functional software packages have allowed for electronic structures to be seriously investigated in GO methods. Two phase optimization techniques, in which a pre-screening global optimization is undertaken at the forcefield level, followed by a reoptimization of the promising structures at the DFT level have become common.[48] However, the risk of the two phase approach is clear: those structures which are preferred with DFT, but not at the forcefield level, are screened out and lost. One way to minimise this effect is to parameterise a forcefield against DFT data. Such parameterisation has been applied extensively by Johnston and coworkers, for bimetallic clusters such as Au–Pd[49] and Cu–Ag.[50] For ultrasmall vacuum-phase model catalytic particles, the small size both requires and allows for GO with more accurate, electronic structure methods. The most well studied class of systems is that of gold, and doped gold clusters,[51–54] for which the high degree of relativistic s–d hybridisation is key. DFT-GO has even been used in conjunction with TD-DFT and ion mobility simulations to fingerprint isomers in a cluster beam.[55] In such an area, where particles exist transiently, or are difficult to isolate, DFT-GO can provide support. The additional complexity of the energy landscape for multicomponent cage systems necessitates an unbiased exploration of configuration space, such as the DFT tabu search for cationic Cu–Sn core–shell clusters[56] and the DFT-EA approach used for Bi–Sn cages.[57] For transition metals, the complex spin arrangements are difficult to predict, owing to the subtle balance between the magnetic moment and structure. As an example, consider ultrasmall Ru–Sn particles. Sn-Doping into noble metal particles is known to enhance catalytic activity, while reducing the manufacturing cost, by replacing some of the expensive platinum-group metal.[58] Paz-Borbòn and coworkers investigated the properties of Ru_{2n}, Sn_{2n} and $(RuSn)_n$ clusters ($n \leq 6$) towards the catalytic hydrogenation of ethylene with a DFT-BH approach, combined with NEB to determine reaction barriers.[59] They observed that the inclusion of tin drives a profound change in the structure, from cubic towards compact, Sn-capped structures, and a reduction in the total magnetic moment. These changes coincide with a decrease in the rate-determining step barrier, and thus an enhancement of reactivity, in agreement with experimental findings. An interesting development, which combines DFT-GO with *ab initio* constrained thermodynamics, was made by Scheffler *et al.* to isolate the global minima of clusters at finite temperatures in the presence of oxygen.[60,61] This method was applied to Mg_xO_y clusters over a range of sizes ($M < 16$), spin states and stoichiometries, finding a surprising preference for non-stoichiometric particles at small sizes.

2.3.1. Ligand-passivated particles. Another possible method to stabilize (sub)nanometer clusters and control their size distribution is to passivate them with (organic) ligands.[10] Passivation of the cluster can lead to two types of cluster–ligand complexes depending on the strength of the cluster–ligand interaction and cluster/ligand concentration:[62] (i) a simple association complex of the ligand with the cluster's global minimum (GM) (or few low-energy isomers), or (ii) a cluster–ligand complex with weak topological similarities to the cluster's gas-phase GM. The former type is amenable to a two-phase GO procedure. In this procedure, low-energy cluster isomers from the gas-phase are obtained first, followed by optimization of the position and orientation of the ligands, which decorate the cluster core. For the latter cluster–ligand type, often full GO of cluster–ligand species is necessary. Both types of cluster–ligand and GO approaches have been considered in GO studies of passivated clusters. These studies are dominated by two classes of systems, the thiolate-protected gold[63–65] and silver[25,66–68] clusters and the hydrogen-passivated silicon clusters.[24,50,69–73]

Thiolate-protected gold clusters have been used as a model system for metal nanoparticles because of their extraordinary stability[74] and availability of synthetic strategies able to prepare monodisperse clusters in high yields.[75,76] The interest peaked with crystal structure determination of $Au_{102}(SR)_{44}$[77] and $Au_{25}(SR)_{18}$[78] clusters, which were formed from a high-symmetry Au core capped by "staple" motifs -RS-(Au-RS)$_n$-Au-RS- (n = 0, 1). This "divide and protect" structural concept,[79] *i.e.* division of the cluster into the metal core and protecting ligands, has been utilized in the first two-phase GO studies[63,64] on $Au_{20}(SR)_{16}$ and $Au_{24}(SR)_{20}$. In these studies Pei *et al.* employed EP-driven BH for an Au core supplemented by manual construction of ligand protections of various lengths that still fulfill the constraints of the molecular formula, which was then followed by local DFT optimizations of the assembled $Au_n(SR)_m$ cluster. The "divide and protect" concept influenced also the first direct GA-based DFT-GO by Xiang *et al.*,[66] which they applied to the Ag^{7-} cluster, ligated with $(SCH_3)_4$ or $(DMSA)_4$. Their procedure involves performing mating and mutation steps on the metal cluster core, with only one ligand atom (sulphur) bound at the core surface, followed by the re-introduction of the remaining ligand chain for local geometry optimization, as depicted in Fig. 3. Before carrying out the local DFT optimization, a fast EP-based Monte Carlo run is used to reorient ligand chains to minimize their steric repulsion. Recently, a full DFT EA-GO of cluster–ligand species has been employed[65] to search for structures of $(AuL)_n$ (L = Cl, SH, SCH_3, PH_2, $P(CH_3)_2$, n = 1–13) clusters. The high ligand concentration (Au-to-L ratio 1 : 1) in these structures prevents formation of an Au core both invalidating the "divide and protect" concept for these stoichiometries and justifying the use of standard cluster GO implementation without any passivation-specific improvements/biases. Lastly, in a number of studies, Bonačić-Koutecký *et al.*[25,67,68] investigated thiolate-protected silver clusters using simulated annealing at the semi-empirical AM1 level to obtain candidate structures for the subsequent local DFT re-optimization. In their works, the

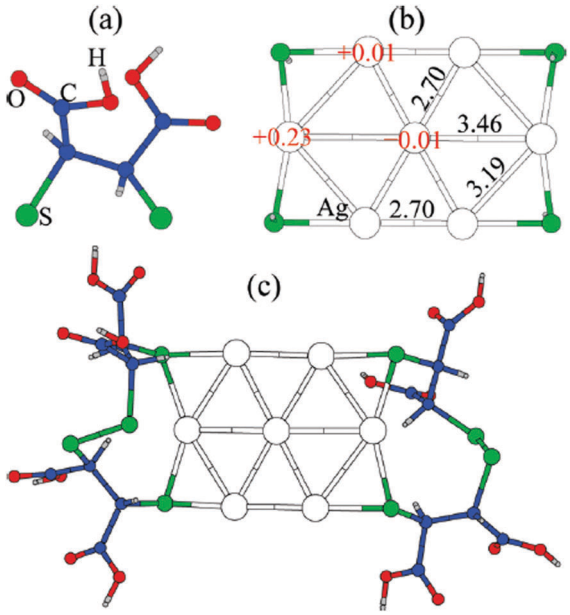

Fig. 3 Structure of the GM for $Ag_7(DMSA)_4^-$ determined with the GA of Xiang *et al.* (a) the structure of the DMSA ligand. (b) The GM determined with minimal sulphur capping atoms, showing Bader charges and Ag–Ag bond lengths, (c) the complete structure of the complex. Reprinted with permission from ref. 66. Copyright 2010, American Chemical Society.

authors illustrated how increasing ligand concentration lengthens the staple motifs -RS-(Ag-RS)$_n$-Ag-RS- until the whole Ag core is consumed and all that is left are the differently interconnected (Ag-RS) units.[25]

The second class of cluster–ligand systems investigated using GO methods are silicon nanoparticles, passivated with hydrogen. The justification for the interest in these systems comes from their potential use in optoelectronics, solar cells or photocatalysis, stemming from their strong size-dependent photoluminescence, photostability and the possibility of preparing highly monodisperse systems.[33,80] In addition, the biocompatibility of silicon and its flexible surface chemistry that facilitate water dispersibility and easy conjugation of DNA or protein probes have made these systems intriguing for bioimaging or for use as biosensors.[81] The pioneering work on GO of hydrogen-passivated silicon clusters has been done by Ge and Head who in a sequence of studies[24,69–71] gradually perfected their EA-based GO implementation. In their first study,[69] a standard two-phase GO was used, employing the AM1 method to locate several low-lying structures of $Si_{10}H_n$ (n = 4, 8, 2, 16, 20) and $Si_{14}H_{20}$, followed by re-optimization at the MP2 and DFT levels. For large, well-passivated $Si_{14}H_{20}$ clusters, the emergence of the bulk Si diamond-lattice structure as GM was observed. The only passivation-specific modification to standard cluster EA-GO involves generation of the initial population; first, the randomly generated Si core was created inside the cubic box and then the H atoms were randomly arranged near the box borderline, either outside or inside. In the follow-up study,[70] the authors tried to mitigate the incorrect AM1 energy ranking by proposing an iterative GO strategy involving two separate EAs invoked consecutively. One is the standard cluster EA (CEA) used for structure optimization at the AM1 level, while a second EA is used to reparametrize the AM1 method using a growing set of reference *ab initio* data (either DFT or MP2) obtained from local *ab initio* re-optimization of low-energy isomers from previous CEA runs. The two separate EAs are performed iteratively until the AM1 parameters give an energy ordering that is consistent with the accumulated *ab initio* database. Although such adaptive, on-the-fly re-parametrizations tailored for a specific problem at hand hold, in our opinion, great potential for the future (with availability of fast computers and robust fitting approaches[82]) (see Section 6 on machine learning approaches), in the early 2000s this approach was deemed prohibitively expensive[70] and was not pursued further. Rather, a fixed set of improved AM1 parameters, termed the GAM1 method, was obtained[24] from the Si_7H_{14} training set, and considered transferable to other Si_nH_m stoichiometries. The sequence of studies by Ge and Head concluded[71] with the introduction of new system-specific mutation operators such as SiH_3 removal, SiH_2 removal or H shift. New mutation operators combined with the previously re-parametrized[24] AM1 method improved convergence of GA, in particular for problematic cases with $Si_{14}H_{20}$ and Si_6H_6 stoichiometries. Recently, small H-passivated silicon clusters have also become the subject of a direct DFT EA-based GO investigation by Baturin *et al.*[50] using a cluster EA implementation in the USPEX code.[83] This study of $Si_{10}H_{2m}$ (m = 0–12) nanoclusters highlighted how hydrogen concentration, temperature and density of low-energy isomers affect the structural and compositional flexibility of the nanocluster ensemble, possibly making experimentally realizable cluster compositions highly non-uniform, both structurally and compositionally. Indeed, with increasing size of the cluster in question, the number of low-energy isomers explodes, making the comprehensive search for configurational space prohibitively expensive. Rather, it becomes necessary to employ strategies capable of obtaining representative structures under given experimental conditions (*e.g.*, concentration of the reactive compounds) with topological properties (*e.g.*, concentration of defects) consistent with experiment. Some work in this direction was done by Biswas *et al.*[72,73] who used EP-driven metadynamics simulations supplemented with simulated annealing runs at the non-self-consistent DFT level using the Harris functional[84] to obtain models of hydrogenated amorphous silicon (a-Si:H). Rather than obtaining the converged free-energy surface of such a large and complex system, the purpose of the metadynamics simulations in this study was to generate configurations with specific topological properties (*e.g.*, Si dangling-bond defects), which are consistent with experimental data from IR and NMR spectroscopies. In particular, the EP-driven metadynamics run, using the average coordination number of silicon atoms as a collective variable, produced an ensemble of a-Si structures with a defined number of undercoordinated Si atoms, which were passivated by hydrogen atoms, using a simple geometric construction to achieve maximal tetrahedral character of defective Si sites, and re-optimized at the non-self-consistent DFT level.

The final class of passivated clusters investigated at the GO level are the hydroxylated silica (silicon dioxide) clusters, a system of ubiquitous fundamental importance (*e.g.*, in mineral nucleation, growth and dissolution processes or in synthesis of nanoporous silicate materials such as zeolites), investigated by Bromley *et al.*[85,86] In their first study,[85] a standard two-level BH-GO scheme was employed, in which several tentative low-energy isomers of $(SiO)_n(H_2O)_m$ (n = 4, 8, 16, 24 and m up to $m/n \geq 0.5$) obtained at the EP level were re-optimized using DFT. In the follow-up study[86] on $(SiO)_n(H_2O)_m$ (n = 6, 8, 10, 12), the authors refined their BH-GO approach by employing a two-step local optimization in each BH step, termed a cascade basin hopping approach, where first a simple and computationally efficient EP is used to pre-optimize the new distorted structural candidate, followed by a more sophisticated EP to carry out full relaxation accounting for polarization and H-bonding. These GO investigations managed to capture two very distinct structural regimes in the $(SiO)_n(H_2O)_m$ system – while small clusters are progressively hydroxylated with increasing water content, the larger clusters tend to form dense amorphous clusters with hydrogen-bonded surface water molecules. This highlights the importance of un-biased (global) structure optimization approaches to correctly predict the structures of passivated clusters as a function of the cluster size or passivation degree.

2.4. Structures of catalyst surfaces

Extended exposed surfaces, which are crucial to heterogeneous catalysis, are usually more geometrically restricted than free particles, owing to the periodicity of the crystal and the presence of strong covalent bonds to the bulk below. As a result, GO for surfaces has received less attention computationally than isolated particles, with simplified periodic models deemed to be sufficient for most investigations. Such models are, however, problematic in exceptional cases. These cases include thin films, where the surface layer may be structurally and electronically distinct from the bulk, due, for example, to incomplete growth or undercoordination. In this area, GO methods are useful in combination with surface science experiments. Another case is for real catalysts, in which complex physicochemical conditions are present at the surface. Oxygen pressures, access to potential adsorbates and temperature can all drive dramatic changes to the surface layer. Metal oxide surfaces, which are often involved in oxidation, photo- and electrocatalysis, are particularly prone to such effects, and have received increased attention recently.

Development of GO methods for periodic condensed matter includes the periodic cut,[87] which is useful for combining parent structures for EA mating steps in systems with periodic boundary conditions and different supercells. Another improvement is the mating slab procedure of Chuang *et al.*,[41] in which a minority section at the top of the slab is chosen to mate between elements in the population, keeping the bulk-like layers below, fixed. The cutting plane was found to be optimal when unconstrained to pass through the cell centre, and kept away from the cell boundaries. The approach was applied for an illustrative example of silicon.

In order to support surface science experiments in the elucidation of complex surface phases, DFT-GO has developed as a useful characterisation tool. In an early example of DFT-GO, Sierka *et al.* determined the stable geometries of oxidised Mo(112) in the p(1 × 2) and p(1 × 3) structures[88] with an evolutionary algorithm, called the hybrid *ab initio* genetic algorithm (HAGA).[89] An oxygen-induced missing-row reconstruction was observed in both cases, which coexist over a wide range of oxygen partial pressures. Later work applied the same method towards the more complex O(2 × 3)–Mo(112) system,[90] again finding better agreement with experimental data than previous models. The HAGA method is similar to other EAs except that it can be used in a constant chemical potential mode, rather than the standard constant composition mode. This allows for the fitness determination of elements of the generation to be ranked by the approximate free energy of formation of the product state, rather than the total internal energy. This is beneficial when studying the chemical reactions that form oxide surfaces. For MoO_x surfaces, the relevant reaction is the formation of the metal oxide from the Mo(112) surface and molecular oxygen. Evolutionary algorithms continue to be used to elucidate structures of reactive oxide surfaces which have eluded experimental characterisation. The structure of the 4 × 1 reconstruction of SnO_2(110), which is active in oxidation catalysis of CO and of CO/NO, and as an activity-enhancing support for metal particles,[91] has been unknown since the 1980s. It was very recently determined, using a combination of DFT-GO and experimental surface X-ray diffraction by Merte *et al.*[92] The surface is found to be terminated by an ordered array of Sn_3O_3 clusters upon the bulk termination of SnO_2(110).

For complex catalytic surfaces under real conditions, defects, such as steps, vacancies and adgrowths are common. In these cases, GO may still provide insight. This is the case for the prototypical TiO_2(110) surface, which is an important and well-studied system for photocatalytic oxidation reactions.[93] Martinez *et al.* employed a DFT-EA to explore the local structure of the common $\langle 1\bar{1}1 \rangle$ and $\langle 001 \rangle$ step edges of TiO_2.[94,95] The authors found new step edge structures which are more stable than the bulk termination, and the presence of O vacancies, which are active in ethanol dissociation. Bechstein *et al.* applied a DFT-EA to explain the presence of reduced strand-like Ti_xO_y adgrowths at $TiO_2\langle 1\bar{1}1 \rangle$ step edges, which are observed in STM images.[96] By separating the strand, which is around 6 nm in length, into distinct regions (the connection region, the strand region and the end-of-strand region), they could unravel the structure of a large system, considering three separate global optimization investigations in parallel. In this way, the unbiased optimization of the local structure allowed for geometries to be discovered which are unexpected from prior chemical intuition, or which would be implausible to otherwise study, due to the vast configuration space available.

Surface GO methods are not only used in support of existing experiments. Theoretical investigations on surface structure have occasionally predicted stable phases of materials before experiment. An MH-GO study on free-standing TiO_2 nanosheets

has recently predicted a new honeycomb isomer that is lower in energy than those previously discovered.[17] Using an artificial neural network potential that was trained on DFT structures, the novel isomer was determined to have a good band alignment with the redox potential for water splitting, and thus is promising from the point of view of both synthesis and catalysis. 2D confining potentials were employed, which is a general procedure for 2D material GO. Randomness was maintained in the search path out of basins by choosing a soft mode, while relaxing the constraint to adopt only the lowest frequency eigenmode. In fact, this choice is the reason MH is found to be more efficient than saddle-point methods.[32] The latter methods either take the lowest mode, which is not guaranteed to correspond to a low energy path, or calculate paths along all eigenmodes, which is computationally demanding. For ultrathin films of AlN, BeO, GaN, SiC, ZnO, and ZnS, DFT calculations have suggested a graphitic structure for the thinnest films of each species, which convert to the polar $(0001)/(000\bar{1})$ above a certain number of layers, and are stabilised by charge transfer.[97] In a similar spirit, the simulated mechanical annealing method of Bromley and colleagues has been used to predict novel phases of catalytically important reduced cerium oxide surfaces,[98] ZnS nanosheets[99] and nanotubes.[100] The approach involves gradually compressing and expanding the system, locally relaxing the geometry at each point, and capturing any new local minima. The process is repeated for each new minimum until the structural space is exhausted.

2.5. Surface-deposited particles

In heterogeneous catalytic experiments and industrial applications, active catalysts are often made up of clusters and nanoparticles supported by stable, insulating surfaces. The role of the surface is much more complex than simply providing additional physisorption to stabilise the particle against sintering.[71,101] Major factors which must be present in GO investigations to treat surface-supported particles include: (i) the effects of lattice strain and epitaxy with the surface, and any particular effects of strong adsorption, (ii) charge transfer between surface and particle, defects and the possibility of sintering and particle migration, (iii) the effect of solvent, adsorbates and relevant reactions on the catalyst structure, and (iv) encapsulation, for catalytic particles contained within a confining environment. Examples of each issue are considered in the following section.

2.5.1. Strain, epitaxy and adsorption strength. Miyazaki and Inoue probed the effects of tuning the interaction strength between a cluster and a support with an early surface genetic algorithm.[102] Binary strings encoded the structural information, within a lattice model for discretizing space. The cluster atom size and relative strength of intra-particle and particle–surface interactions were modelled with a Lennard Jones potential. Particles which adopt icosahedral morphologies in the vacuum phase were found to wet the surface, forming either monolayer islands or condensed layered structures, depending on the interaction potential. For large cluster atoms, the potential well for cluster–surface bonding was narrow, and thus smeared out, inducing full surface wetting. This is an early example of lattice effects being directly responsible for cluster structure in GO investigations. In another early surface EA-GO study, Zhuang *et al.* determined the global minima of adatom clusters Al_n, Ni_n, Ag_n, Pd_n and Pt_n ($n \leq 40$) upon (111) surfaces,[103] with modified embedded atom potentials. Clusters generally favoured structures which maximised the number of nearest neighbours, but discrepancies were found for systems where the adatom–surface interaction was particularly strong, allowing for nearest neighbour bond-breaking to be compensated for by strong adsorption of edge sites to the surface. Recently, Eckhoff and colleagues extended the analysis of adsorption to generic pristine surfaces, focussing on the mechanical properties of the surface, and their effect in driving the geometry of the particle.[104] They report that the surface microstructure, defined by the lateral strain of the substrate, can have profound effects on the preferred cluster morphology on pristine supports. Stacking faults, twinning and reorientation of the cluster can all be observed in global minimum energy structures, along with reordering of the relative stability of structural motifs. It should be concluded that lattice mismatch and strain is sufficient to access the full range of possible adsorbate structures, even in the absence of surface roughness or defects. Lattice mismatch between the adsorbate and surface is important in the growth, structure and stability of particles, and has been studied in detail with two phase EP/DFT-GO methods for model surfaces and metal particles.[105,106] For cubic lattices, such as MgO(100), cubic phases develop in the adsorbed particle, which modify the facets exposed to potential gas phase reactants. The balance between surface epitaxy and the natural preference for close packed structures undergoes a crossover at a certain size. Goniakowski *et al.* used BH and PEW for noble and coinage metal clusters of selected sizes up to 500 atoms on MgO(100).[106–108] A size-dependent transition from cube-on-cube (100) structures to fcc (111) motifs was found. The onset size of the transition was observed to be smaller for particles with larger lattice mismatch with the surface.

For multicomponent particles, the different strength of surface adsorption for the component elements can drive surface-induced segregation, and even affect the preferred structure of the cluster. For example, Ismail *et al.* used a basin hopping algorithm with a two phase EP/DFT-GO approach to investigate the segregation of Pd to the surface, for adsorbed AuPd clusters on MgO(100).[105] Exchange moves were employed at a frequency of 10% to speed up the search for the wide permutational isomer space. This is a crucial consideration for the efficiency of GO methods on multicomponent systems. In both Monte Carlo and nature-inspired methods, the choice of swap move frequency is made, based on the balance between the structure and permutational isomer optimization.

Subnanometre-scale metal particles on oxide supports have been intensively investigated, especially since the discovery of enhanced intrinsic catalytic activity at ultrasmall sizes. Subnanometre sized metal particles have been shown to exhibit even higher activities in heterogeneous catalysis (Cu for CO-to-methanol conversion,[109] Ag for propylene epoxidation,[110] Au for

alkyne hydration,[74] Pd for electrocatalytic water splitting[87] and Pt for propane oxidative dehydrogenation[111]). As in the case of ultrasmall isolated particles, the GO of the very smallest particles upon surfaces requires the accurate capture of electronic properties, such as the "metal-on-top effect"[112] and spin effects. Davis *et al.* used the recently developed "pool EA" to examine the stabilisation of Au subnanometre particles on MgO(100), by alloying with iridium. The authors found significant stabilisation of the particles as the iridium content was increased, in agreement with experiments, which suggested an enhanced sintering resistance.[113] Vilhelmsen and Hammer developed and applied a direct surface DFT-EA for subnanometre clusters of gold upon MgO,[114] for which the upper surface layers were free to locally relax during local optimization. They found a number of Au_8 structures lower in energy than previously reported, despite the great number of studies performed without unbiased GO methods. This is a clear indication of the need for an open-ended GO search. More recently, the same group has developed and benchmarked their EA, which is built into the atomic simulation environment (ASE), to more efficiently perform direct DFT-GO.[26] The authors introduced a series of new moveclasses, specific to adsorbates upon a surface. One such moveclass is the rotation mutation, which moves the cluster around the surface normal, so as to locate the optimal overlap between the cluster and surface.

Another is the symmetrisation operator, which reflects half of the cluster across a mirror plane drawn through its centre. These moves yielded only small improvements to the search efficiency, but it should be noted that the GM was not difficult to locate, with a success rate of around 98%. Hence the increase in diversity introduced by the new steps was probably not necessary in the tested case. One may expect that for a more difficult case, these moves would have a more significant effect. Extended to a (100)-oriented Au nanorod upon TiO_2(110), the authors predict an unusual interfacial layer of oxygen, which was unlikely to be predicted from a biased search.[115] The thermodynamics and kinetics of CO oxidation on this system were subsequently examined, as depicted in Fig. 4. DFT-BH has also been applied to subnanometre gold clusters on hydroxide supports[116,117] which are proposed to exhibit good low temperature CO oxidation activity.[118] The activity is explained by charge transfer between the surface and the gold cluster, which activates the O–O bond of adsorbed oxygen.

2.5.2. Charge transfer, sintering and migration. Particle sintering is a major deactivation route for industrial catalysts. Defects at surfaces play a large role in trapping catalytically active species, and are important in understanding the processes of growth and coalescence. Defects are often unavoidable consequences of surface preparation, but can also be caused by the adsorption of catalytic particles, through charge transfer and surface abstraction. However important, the inclusion of defects adds no conceptual complication to GO methods, and so only some interesting results will be discussed here.

Some of the first full DFT-GO studies of catalytic particles on defective oxides were performed using a DFT-BH method to investigate Ag and AgPd clusters on MgO(100). On the neutral double vacancy, which is common in MgO preparation, the recovery of the gas phase Ag_8 magic number cluster was predicted.[82] This cluster exhibits large HOMO–LUMO gaps and stability with respect to nearby sizes. By contrast, a DFT-BH investigation by the same authors on Ag_n ($n \leq 11$) on the F_s vacancy of the same oxide shows the complete loss of the magic number.[119] The frustration between the metal–metal bonding

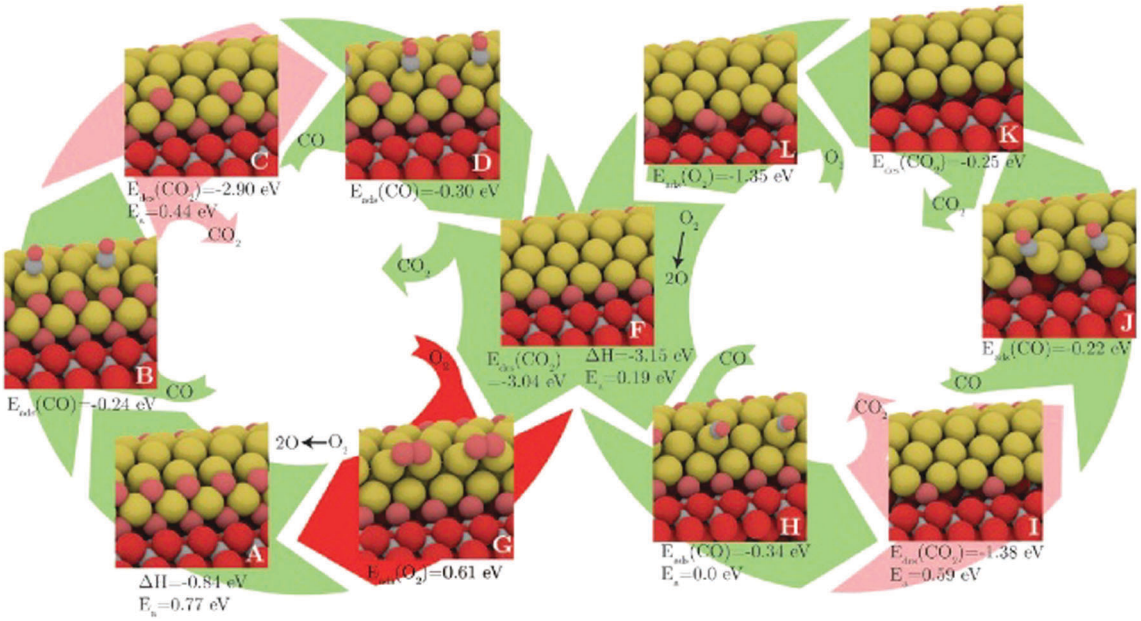

Fig. 4 The structures and energetics of CO adsorption, CO_2 production and subsequent reoxidation of the (100)-oriented Au nanorod edge on TiO_2(110). Barriers for CO_2 formation are prohibitively high, while the reoxidation of the reduced gold nanorod is facile. Reprinted with permission from ref. 115. Copyright 2013 American Institute of Physics.

and the distortion necessary to maximise the Ag-vacancy interaction leads to a global minimum structure which is distorted and not particularly stable. The apparently complex interaction between catalytic particles and the defective support is based on the balance between satisfying the closed-shell stable structures found in the vacuum phase and maximising bonding to the TiO_2 (110) surface. This balance is difficult to predict without a full GO investigation. Another important catalytic system is that of Pt upon ceria surfaces, which are used in automotive catalysis and have been recently proposed as a potential high power density PEM fuel cell anode.[120,121] Paz-Borbòn and colleagues recently employed a DFT surface GO method to determine the low energy structures of Pt_n on a CeO_2(111) surface ($n \leq 11$).[122] Basin hopping coupled to a plane-wave DFT code allowed the authors to locate putative global minima across the size range. It was found that 2D Pt structures are preferred up to $n = 8$, owing to the strong interaction between the metal and substrate. Charge transfer occurs from Pt to surface oxygen, along with reduction of Ce^{4+} to Ce^{3+}. This is in good agreement with studies which show the reducibility of ceria to pin particles to the surface through charge transfer.[123] Experimental data suggest that charge transfer goes through a per-atom maximum for particles of around 50 atoms.

The development of ultrahigh vacuum deposition techniques and the sophisticated surface science characterisation methods have raised the questions of trapping, migration and coalescence of catalytically relevant deposited particles.[124] The diffusion pathways of ultrasmall particles have been studied computationally mostly with pathway search methods as described in Section 2.1. Upon the model metal oxide MgO(100), Xu *et al.* calculated the diffusion pathways and rates according to harmonic transition state theory.[125] Interestingly, they observed that tetramers were even more mobile than monomers, while single atoms are not especially attracted to Pd_1/F_s sites. Hence, the sintering mechanism was predicted to be one of Pd atoms trapped at defects, while small clusters grow and freely migrate around the surface until coalescence with the Pd_1/F_s centres. This differs from the previous model of single atoms combining with Pd_1/F_s centres and growing in a stepwise fashion. The predictions were further tested with kinetic Monte Carlo simulations over the 200–800 K temperature range, finding excellent agreement with experiment.[126] Similar findings have been made for coinage metal clusters,[127] suggesting the importance of small particles in coalescence processes. The vast configuration space available to particles larger than 1 nm makes global searches for migration and sintering pathways prohibitively expensive, though studies have aimed to describe the kinetics of Ostwald ripening processes by fitting DFT energetics to sintering rate equations.[128]

2.5.3. Adsorbates and reactions. As is clear from the range of elements, surfaces and applications of the above studies, the use of global optimization techniques for sub-nanometre sized catalysts on supports has become relatively standard, both as an unconstrained investigative tool and for supporting experimental characterisation of complex systems. The methods are robust, and the application of the appropriate model is becoming ever more important. The model should ideally include all present species, both reactive and inactive, to represent the correct environment of the real catalyst.

It was found by Wang and Hammer that under the reduced conditions relevant for surface science experiments, the Au_7 cluster is only weakly adsorbed to the surface.[129] By contrast, under the oxidising conditions of the real catalyst there is strong adsorption of partially cationic gold, which leads to low CO oxidation barriers. Hence, the problem of transferring results between different experimental techniques is also a concern for the choice of model in GO studies. The adsorption and diffusion of Pt clusters on TiO_2(110) surfaces[130] has also been studied with a DFT-EA approach. Oxidised (O atoms on top), reduced (hydrogen adatoms) and surfaces containing oxygen vacancies were considered in the surface model. These models gave differing results, in good agreement with experiment, that O defects trap small particles, reducing diffusivity and maintaining small particle sizes, while hydrogen has little effect, allowing migration and sintering. While not strictly a catalytic system, the role of solvating water in the stabilisation of particular surface terminations has been probed with a direct DFT-MH investigation for CaF_2.[131] It was found that the polar (100) termination becomes preferred over (111) in the presence of water, while facile reconstructions between nearly-isoenergetic local minima lead to a fluxional surface structure. In this case, the presence of solvating water is completely responsible for the structure, and the consequent chemistry of the surface.

As stated at the start of the section, the energetics derived from global optimization studies are primarily potential energies, which represent low temperature behaviour, from which thermodynamic approximations may be made. However, catalysis is often a kinetics-driven process. As such, the 0 K approximation tells only part of the story. A method which combines GO, KMC and path sampling was developed by Fortunelli and coworkers with the intention of application directly to catalysis, and is denoted as Reactive Global Optimization (RGO).[132,133] This method aims to globally seek the combined energy landscape of a catalytic particle, the surface and all gas phase reactants of the catalytic reaction in question. The search is based on the calculation of kinetic prefactors and internal energy barriers of elementary steps. The accessible region of configuration space is deliberately limited by choosing cutoffs in the maximum barrier height, which is analogous to a defined experimental temperature. In brief, the RGO process consists of a cycle in which, (i) a structure is identified, (ii) all neighbouring minima are located by following each eigenvector of the Hessian matrix, (iii) barriers are determined for the connection between adjacent minima, (iv) unfeasible steps are purged, (v) the next structure is selected based on a KMC simulation and (vi) adsorbates are added to the structure. One benefit of the method is that it is inherently parallelisable, as multiple walkers may explore disconnected regions of configuration space concurrently, and share structural information so as to avoid repetition. In this way, RGO is similar to the parallel excitable walkers method. Additionally, by virtue of being a kinetic, rather than a thermodynamic process, it is suited to probing kinetically controlled reactions. Furthermore, complex ligand

effects, such as the adsorbate-induced decomposition of small particles may naturally be taken into account. For CO oxidation over $Ag_{3-x}Au_x$, Ag_2Au_1 was found to have the best balance of reactivity and stability.[133] The limitation of the method is the great cost of searching even a realistically truncated potential energy surface in a reasonable time. The selection of eigenvector following methods and the accuracy of the saddle point convergence allow for some tunability, but the problem remains that hundreds of relevant local minima may be present under the conditions relevant for experiment.

2.5.4. Encapsulated particles. The encapsulation of clusters into a host matrix represents a natural way to control their size distribution and improve their sintering resistance.[134–138] Besides imposing steric constraints, encapsulation provides an additional handle to tune the properties of clusters by varying the dielectric and charge properties of the confining environment. However, direct cluster GO in confinement is computationally expensive and thus extensive manual construction of possible clusters in confinement followed by local optimization is still a popular method of choice.[139–141] Admittedly, manual construction is well-justified in situations when formation of only ultrasmall clusters (couple of atoms) is possible due to an extreme confinement, which significantly limits the number of isomers to be tested.

A more involved approach is to utilize a two phase EP-GO/DFT approach, employing GO for gas-phase clusters at the EP-level with subsequent local DFT reoptimization of embedded low energy gas-phase isomers. The applicability of this procedure relies on the assumption that confinement neither causes significant reordering of the gas-phase isomer stabilities nor creates entirely new isomers which are not local minima in the gas-phase. This two-phase procedure was used for Pt_{13} clusters in Y zeolite, using iterative metadynamics[142] calculations to obtain low energy gas-phase Pt_{13} isomers, and for small copper clusters embedded in the ERI zeolite optimized using an EA.[143]

Finally, a few studies have attempted direct cluster GO in confinement. The most prominent examples are the works of Vilhelmsen *et al.*[144,145] who employed their EA-based GO for clusters on surfaces[114,146] to find global minima structures of Pd clusters embedded in UiO-66 MOF and Pd, Au and PdAu clusters in MOF-74. In these works a cluster is subjected to a GO process inside the flexible MOF nanopore. The only change made to the original EA for supported clusters[114] (see Section 2.5.1) is the way the starting population was generated; employing either a cylindrical coordinate system in the MOF-74 pore or Monte Carlo technique with insertions, deletions, and displacements for UiO-66. The authors show that interaction with the walls of the nanopores, which are composed of aromatic rings and Zn open metal sites, results in significant deformation of the putative gas-phase GMs of Au_8, Pd_8 and Au_4Pd_4 clusters, with deformation energies above 0.6 eV, supporting the need for an unbiased GO process. The authors also note that as a by-product of generating a large number of candidate structures distributed through the MOF unit cell during the EA run, a diffusion path from one unit cell to the next can be established through the identified structures. In the follow-up paper on Pd clusters in UiO-66 MOF, much larger clusters were considered (up to Pd_{32}). The chosen MOF structure contains pores defined by cages separated by relatively narrow windows (about 3.9 Å), which was hypothesized to stabilize isolated Pd clusters preventing their agglomeration. However, their calculations show that the Pd cluster would not only grow to fill up the cages in the MOF, but interconnect with Pd clusters in neighboring cages to form thermodynamically stable aggregates.

The GA-based GO has also been recently used to obtain the structures of subnanometer $(PbS)_n$ ($n < 6$) quantum dots confined in a sodalite cage,[147] which is a building unit of a number of industrially relevant zeolites. The sodalite cage with different compositions (pure silica, H-, Li, and Na-exchanged) was considered and a modified cut-and-splice operator was proposed, which also included parts of the confining environment (extra-framework cations) in the crossover operation. The author reported stability reordering of the isomers with respect to the gas-phase, with changes even to the global minimum. Moreover, these changes were dependent on the type of extra-framework cation present in the SOD cage. Results for encapsulated $(PbS)_2$ are shown in Fig. 5. These results hint at the possibility of fine-tuning the structure and properties of embedded clusters with a suitably chosen confining environment. The possibility of tuning the cluster structure by adjusting the environment composition (Al/Si ratio) was investigated also by Palagin *et al.*[148] in a BH-GO study on subnanometer copper oxide clusters in MOR zeolites. As the copper oxide clusters act as cations that charge-compensate the negative charge of the zeolite framework, the interaction with the zeolite framework is strong, substantiating the need for an appropriate *ab initio*

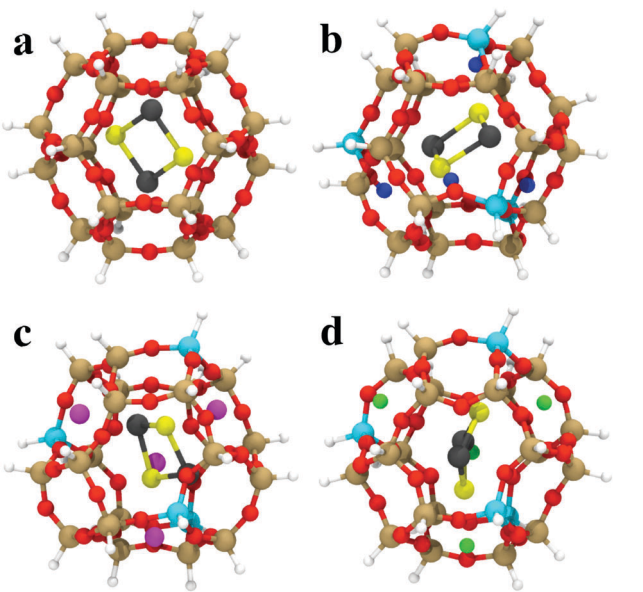

Fig. 5 The most stable $(PbS)_2$ adsorption configurations in (a) SiSOD, (b) HSOD, (c) LiSOD and (d) NaSOD cages, described by an H-terminated cluster model at the PBE0/TZVP level of DFT. O is red, Si is brown, Al is cyan, Li is magenta, Na is green, Pb is black, S is yellow, terminating-H is white, charge-compensating H^+ is blue. Reprinted with permission from ref. 147. Copyright 2016, American Chemical Society.

treatment of both the cluster and environment during the GO process. We note that both implementations, the GA-based GO used by Vilhelmsen *et al.*[144] and BH-GO used in the study by Palagin *et al.*,[148] are now publicly available through the open source project "Atomic Simulation Environment" (ASE).[149]

Common to all GO methods discussed in this section are the limitations of the harmonic approximation, from which kinetics are derived, and the lack of real temperature effects. The free energy surface, which is the true landscape to be sought, is the subject of techniques such as *ab initio* molecular dynamics and metadynamics, which are considered in other sections of the current review. What these methods gain in accuracy for the calculation of chemical properties, they lack in exploration scope, due to their computational cost. Hence, the use of GO methods for qualitative information, screening of structures and broad comparison to experiment remains valuable in the field of heterogeneous catalysis.

3. *Ab initio* constrained thermodynamics

3.1. Basic principles

The effect of the finite partial pressures of surrounding gases on the catalyst properties can be taken into account by constrained *ab initio* thermodynamics (AITD) as has been formulated by Reuter and Scheffler.[150,151] The thermodynamic formalism is briefly described below, a more comprehensive review can be found, *e.g.* in ref. 152 and 153 The catalyst environment is described by pressure p and temperature T. The Gibbs free energy $G(T,p,N_iN_j)$ describing the system depends on the number of i and j atoms (*e.g.*, metal and oxygen in the case of surface metal oxides) in the system, in addition to p and T. The most stable system geometry and composition is determined by the minimum surface free energy:

$$\gamma(T,p) = \frac{1}{A}\left[G\left(T,p,N_iN_j\right) - N_i\mu_i(T,p) - N_j\mu_j(T,p)\right] \quad (1)$$

where μ_i and μ_j are the chemical potentials of individual components. It is assumed that there are separate reservoirs for each of the components. The surface free energy defined above represents the cost to form the particular surface structure (configuration) from the corresponding reservoirs. Finding the thermodynamically most stable configuration at given T and p is thus realized by finding the surface configuration that minimizes $\gamma(T,p)$. Therefore, it is sufficient to calculate just the excess surface free energy with respect to a suitably chosen reference system:[152]

$$\gamma(T,p) - \gamma_0(T,p) = \frac{1}{A}\Big[G\left(T,p,N_iN_j\right) - G_0\left(T,p,N_i'N_j'\right) - \Delta N_i\mu_i(T,p) - \Delta N_j\mu_j(T,p)\Big] \quad (2)$$

where γ_0 and G_0 are the surface free energy and Gibbs free energy of the reference system containing N_i' and N_j' atoms.

To avoid demanding calculations of Gibbs free energies it has been shown that it can be safely approximated by total DFT energies $E^{\mathrm{total}}(V,N_i,N_j)$ calculated for volume V.[150] First, the $pV(T,p,N_i,N_j)$ term required for the transition from Gibbs to Helmholtz free energy was shown to be negligible (on the order of 10^{-3} meV Å^{-2}). Second, the vibrational contribution to surface free energy $F^{\mathrm{vib}}(T,p,N_{\mathrm{M}},N_{\mathrm{O}})$ was analyzed and the upper limit was estimated to be $\pm$10 meV Å^{-2}. Accepting these approximations makes the search for the thermodynamically stable surface computationally reduced, to the evaluation of energy difference:

$$\Delta E^{\mathrm{tot}} = E^{\mathrm{tot}}(N_i,N_j) - E_0^{\mathrm{tot}}(N_i',N_j') - \Delta N_iE^{\mathrm{tot}}(i) - \Delta N_jE^{\mathrm{tot}}(j) \quad (3)$$

where individual terms on the right-hand side are DFT total energies calculated for the surface structure under investigation, reference system structure, and for reservoirs of components i and j. Taking surface metal oxides as an example, $E^{\mathrm{tot}}(i)$ is the DFT energy of the O_2 molecule in the gas phase and $E^{\mathrm{tot}}(j)$ is the DFT energy of the bulk metal. All the temperature and pressure dependences in eqn (2) are contained in the remaining part of the excess surface energy $\Delta N_i\Delta\mu_i(T,p)$.

3.2. Structure of catalysts in a reaction environment

Within the approximations outlined above the constrained *ab initio* thermodynamics is computationally affordable and, thus, it becomes more and more popular in the catalysis community. The AITD approach has originally been introduced as a method for analyzing the chemical composition of the open catalyst surfaces under varying reaction conditions and it is currently routinely employed for predicting the most stable surface termination of complex multicomponent systems[155–158] and for studying active site speciation in confined space of microporous crystalline materials.[159–162]

An illustrative example for this method is a comprehensive study by Scheffler and co-workers on the composition of the Pd(100) model catalyst in a reactive environment corresponding to CO oxidation.[154] The thermodynamic stability of various (relevant) surface structures formed under an O_2 and CO atmosphere on the Pd(100) surface was investigated as a function of temperature and chemical potential of individual components. A combinatorially representative set of 119 ordered adsorption phases of O and CO on metal Pd(100) and oxide PdO(101)-$\left(\sqrt{5}\times\sqrt{5}\right)R27^\circ$ surfaces were considered. Only 11 of them were found to be thermodynamically most stable structures for particular windows in (T, μ_{CO}, $\mu_{\mathrm{O_2}}$) space. The calculated surface phase diagram is depicted in Fig. 6. The bottom left corner corresponds to vanishing pressure of both CO and O_2 gases and the clean Pd(100) surface is the most stable phase. Moving from left to right corresponds to increasing O_2 pressure (and no increase of CO concentration). First, a p(2 × 2)-O/Pd(100) surface is formed where O atoms occupy the hollow sites on the Pd(100) surface with a corresponding oxygen coverage of θ = 0.25 ML. With increasing O_2 pressure the $\left(\sqrt{5}\times\sqrt{5}\right)R27^\circ$ surface is formed. At even higher oxygen concentration the PdO bulk becomes the most stable phase. Similarly moving from the bottom left corner upwards the thermodynamically stable ordered CO adsorption structures on Pd(100) are apparent. When both O_2 and CO partial pressures increase there are three stable phases where O or CO

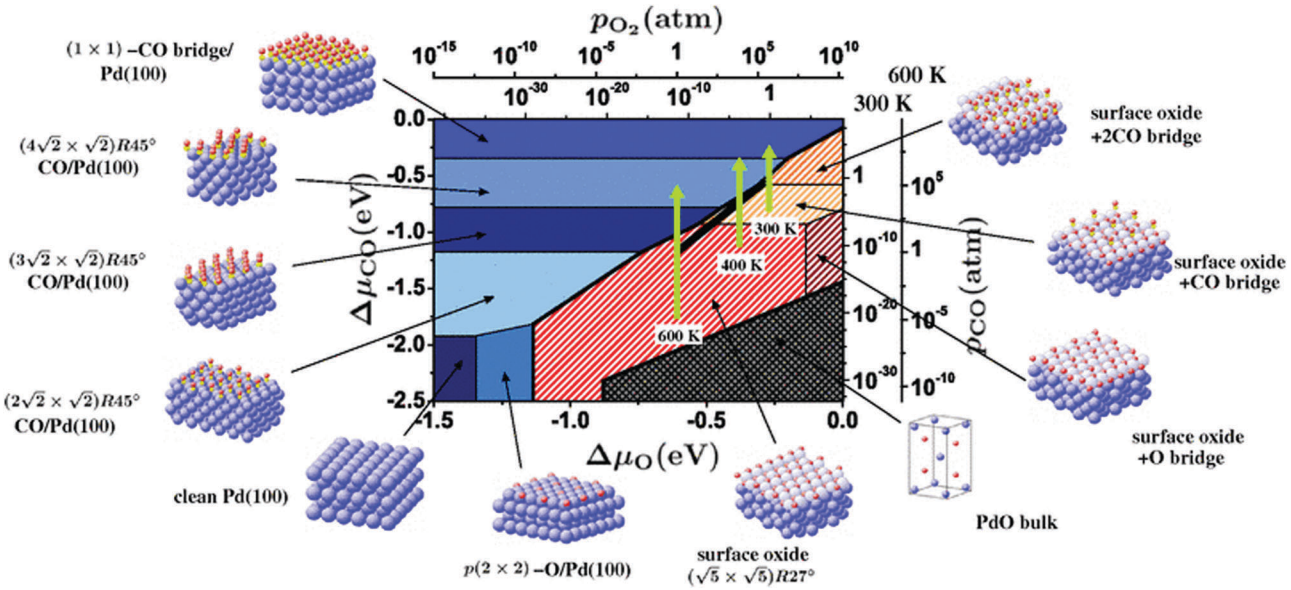

Fig. 6 Surface phase diagram for the Pd(100) surface in "constrained thermodynamic equilibrium" with an environment consisting of O_2 and CO. The atomic structures underlying the various stable (co)-adsorption phases on Pd(100) and the surface oxide are crosshatched, phases involving the surface oxide are hatched. The dependence on the chemical potentials of O_2 and CO in the gasp phase is translated into pressure scales at 300 and 600 K. The thick black line marks gas phase conditions representative of technological CO oxidation catalysis, *i.e.* partial pressures of 1 atm and temperatures between 300–600 K. Reprinted with permission from ref. 154. Copyright 2007 American Physical Society.

is adsorbed on the $(\sqrt{5} \times \sqrt{5})R27^\circ$ surface oxide. However, no phase with both O and CO co-adsorbed on the $(\sqrt{5} \times \sqrt{5})R27^\circ$ surface was found (in agreement with experimental data[163]). Experimental conditions relevant to the technological CO oxidation catalysts were found to fall close to the boundary between phases derived from gas adsorption on $(\sqrt{5} \times \sqrt{5})R27^\circ$ surface oxide and from gas adsorption on the Pd(100) surface. Two important conclusions were drawn from this phase diagram: (i) thick bulk-like PdO oxide is not formed on the surface under technologically relevant conditions and (ii) the $(\sqrt{5} \times \sqrt{5})R27^\circ$ surface oxide must be taken into consideration with respect to the catalytic activity of Pd under "*operando*" conditions.

The stability of the phase diagram with respect to changed exchange–correlation functional was also investigated.[154] Comparing the results obtained at the LDA, PBE and rPBE levels, the topology of the phase diagram remained unchanged. However, individual phase boundaries are shifted one way or the other, depending on the functional employed. The differences were partly tracked down to the description of the gas-phase O_2 and CO molecules. The constrained *ab initio* thermodynamics results reported above helped in identifying the relevant phases that must be considered for the assessment of the catalytic activity of the catalyst. They were subsequently used in first principles kinetic Monte Carlo simulations. Results of this study helped in understanding the relevant experimental data and they clearly showed that the catalyst surface (active state) can be dramatically changed moving from the UHV region to the "*operando*" conditions.

As a second example we discuss a relatively simple but straightforward application of *ab initio* thermodynamics on a NH_3-SCR process catalyzed by a Cu-CHA catalyst: phase of Cu^I under the reaction conditions.[164] The catalytic reduction of NO_x is an environmentally important process increasingly enforced by legislation. The current approach for NO_x to N_2 conversion is the selective catalytic reduction with NH_3 as the reducing agent (SCR-NH_3). Chen *et al.* used a combination of *ab initio* thermodynamics and *ab initio* molecular dynamics to elucidate the structure of the active species and mechanism of O_2 activation in the Cu-CHA zeolite catalyst (see Fig. 7). The key question addressed by *ab initio* thermodynamics was the location and coordination of Cu^I ions in CHA under the reaction conditions (temperature and NH_3 partial pressure). Previous experimental studies indicated that at temperatures below 523 K and above 623 K the SCR-NH_3 activity showed second and first order dependence on the Cu loading.[165]

The state of the copper ions under SCR conditions was investigated by a thermodynamic analysis, constructing the phase diagram of Cu^I coordinated to increasing number of NH_3 molecules. Following the strategy of Reuter and Scheffler[150] the free energy difference between solvated and framework-coordinated Cu^I ions was obtained from:

$$\Delta G = G_{Cu(NH_3)_x} - G_{Cu} - x\mu_{NH_3}(T,p) \quad (4)$$

where $G_{Cu(NH_3)_x}$ and G_{Cu} were free energies of Cu^I ion solvated by x molecules of NH_3 and Cu^I ions coordinated to the zeolite framework, respectively, and μ_{NH_3} was the ammonium chemical potential. Free energies were calculated with explicitly included zero-point energy corrections and vibrational entropies, within the harmonic approximation.

The phase diagram shown in Fig. 7 shows that at low temperature and high ammonia partial pressure the Cu^I ions

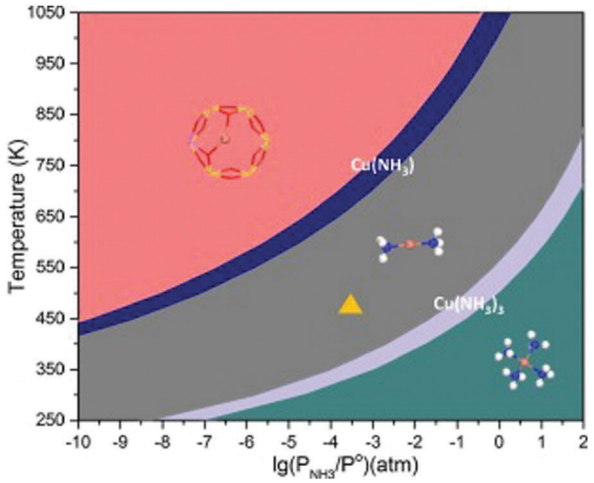

Fig. 7 Phase diagram for $Cu(NH_3)_{x+}$ in CHA with varying NH_3 pressure and temperature. The yellow triangle indicates typical operating conditions, *i.e.* a temperature of 473 K and an NH_3 concentration of 300 ppm. The phase-diagram is constructed with NH_3 in the gas-phase as reference. Reprinted with permission from ref. 164. Copyright 2018 The Royal Society of Chemistry.

are solvated by 4 NH_3 molecules while at high temperature and low ammonia partial pressure they are coordinated to the framework oxygen atoms in the vicinity of framework Al. Under the conditions typical for the SCR-NH_3 process the Cu^I ions are linearly coordinated to just two NH_3 molecules and these $Cu^I(NH_3)_2$ species are catalytically active species in O_2 activation. Chen *et al.* also considered the fact that $Cu^I(NH_3)_x$ species formed inside the zeolite channels and in the external gas phase may have different stabilities. The phase diagram constructed with respect to such a fluid phase (shown in ESI of ref. 164) is shifted in favor of Cu^I extra-framework cations; however, the conclusions about the character of the catalytically active $Cu^I(NH_3)_2$ species remained unchanged.

Various applications of *ab initio* constrained thermodynamics have appeared in the literature during the last decade. The structures and stabilities of various nanoparticles used in catalysis were investigated as a function of synthesis conditions or "*operando*" environment, *e.g.* Mo_2C, ZnO and Ni_2P,[166–169] surface coverage as a function of T and p_i,[161,170–172] doping effects,[173] metal alloying,[174,175] surfaces under electrochemical conditions,[176,177] water in the interlayer space,[178] or controlled growth conditions,[179] to name just some. Several recent reviews and perspectives are also available.[152,153,180–184]

In summary, the *ab initio* constrained thermodynamics approach in the modeling of heterogeneous catalysts is a clear success of present-day computational investigation of catalysis. It brings a major step forward from 0 K/UHV conditions towards much needed "*operando*" modeling. As discussed in the introduction, such a step is inherently connected with increased complexity in modeling. The predictive power of *ab initio* constrained thermodynamics depends on the selection of particular surface configurations explicitly considered in the investigation. The use of global optimization techniques (described in the previous section) for the configuration selection with the *ab initio* thermodynamics is a major step towards minimizing the risk connected with missing important configurations. The accuracy of *ab initio* thermodynamics depends on the accuracy of the approximations involved. The choice of exchange–correlation functional is important for quantitative predictions; however, it has been shown that the qualitative picture remains unchanged regardless of the exchange correlation functional (see above).[154] The approximation of Gibbs free energy by DFT total energies discussed above has been shown to be adequate (*e.g.*, ref. 150 or 168), giving an error of a similar or smaller size as the precision of the underlying exchange–correlation functionals. However, this assumption may not necessarily hold for all systems.

4. Free energy techniques

The transition from the potential-energy surface (PES) to the free-energy surface (FES) is essential for understanding catalysis (or chemical reactivity in general). Two classes of methods allowing a transition from PES to FES are discussed below. A conceptually and computationally (relatively) simple approach based on calculations of the Hessian matrix is covered briefly in Section 4.1 while computationally much more demanding biased molecular dynamics techniques are discussed in detail in Section 4.2.

4.1. Hessian-based thermal corrections

The transition from PES to FES can be simplified by construction of approximate partition functions for relevant stationary points on the PES. Since this is a well-established technique that has been used in computational chemistry (and catalysis) for decades we will only briefly outline its advantages and limitations.

Within the Hessian-based thermal corrections approach, partition functions are constructed for stationary points. The ideal gas expressions are typically used for translational and rotational contributions of free molecules while the harmonic approximation is used for vibrational partition functions. Thus, the evaluation of the matrix of second derivatives of energy with respect to atomic coordinates becomes the computationally most demanding part within this approximation. Since such calculations are expensive for systems with hundreds or more atoms, suitable approximations have been established. Partial Hessian vibrational analysis[185] can greatly reduce the computational requirements with only a small error in calculated characteristics with respect to reference full Hessian vibrational analysis.

Gibbs or Helmholtz free energies can be obtained from partition functions in principle for any temperature; however, it should be noted that the validity of such extrapolation decreases with increasing temperature. An obvious problem is in the harmonic approximation used in the construction of vibrational partition function, in particular, for low energy modes. It has been recently shown that partition functions based on anharmonic vibrational frequencies lead to improved rate constants and other characteristics of catalytic systems.[186–188] However, evaluation of vibrational frequencies beyond the

harmonic approximation is associated with significant computational expense. A problem of low energy modes can be partially overcome with the mobile adsorbate method within which the low energy vibrational mode is treated as translational or rotational degree of freedom.[189]

The Hessian-based thermal corrections model allows accounting for realistic temperature, leaving the system composition and complexity unchanged; it is thus placed vertically above the 0 K/UHV model in Fig. 1. It follows that this method is suitable for the description of systems where (i) the structure of the catalyst active sites does not depend on the temperature and reaction environment, (ii) concentration of reactants is low (gaseous reactions) and (iii) reaction temperature is moderate. An important class of catalysts where Hessian-based thermal corrections have been successfully employed are zeolites; they are thermally and chemically stable in the catalytic systems in which they are used, and due to their microporous character the concentration of reactants at the active sites is limited. For more details see the recent review by Van Speybroeck *et al.*[190]

4.2. Biased molecular dynamics

With increasing temperature and complexity of reacting systems (*e.g.*, reactions at the liquid/solid interface), one needs to switch from descriptions based on a few individual configurations to a statistical description over ensembles representing reactant, product and transition state configurations. In other words, one moves from the potential energy to the free energy surface, which is the true landscape to be investigated. The free energy F is defined

$$F = -k_B T \ln Z \tag{5}$$

$$Z = \int e^{-\frac{H(q,p)}{k_B T}} dq dp \tag{6}$$

where Z is the partition function, k_B is the Boltzmann constant and $H(q,p)$ is the Hamiltonian of the system. However, absolute free energies are typically not calculated since, with the exception of very simple systems, accurate evaluation of the phase space integration in eqn (5) is infeasible. Rather, the free-energy differences between (macro)states (reactant state, product state, transition state – separatrix, *etc.*) are calculated. Various approaches to sample representative ensemble of configurations and to accurately estimate free energy differences have been devised and are discussed in detail below.

The temperature-dependent configuration ensembles can be generated by molecular dynamics (MD) or Monte Carlo methods.[191] However, within the context of first-principles description of reactive systems, Monte Carlo methods were used very little[192,193] due to their typically low efficiency (low acceptance rates, highly correlated data series) and inherent inability to provide temporal characteristics (mechanism, kinetics). Therefore, *ab initio* molecular dynamics (AIMD) is nowadays a standard tool to study reactions on the free energy surface. Using plain AIMD for this purpose is, however, problematic, since most of the chemical reactions occur on very long time scales compared to elementary molecular motions which need to be described with AIMD. Thus, statistically significant sampling of these rare events in a plain AIMD is, with current computational resources and technology, largely intractable. To overcome this disparity in time scales and enhance sampling of these highly activated regions, a number of methods have been proposed and we refer the reader to topical reviews[194–196] for an in-depth discussion. Here, the focus will be on two classes of enhanced AIMD methods which have been used extensively to obtain free energy reaction profiles within the field of catalyzed heterogeneous reactions: (i) methods using a biasing potential such as umbrella sampling[197] or metadynamics,[198] and (ii) thermodynamic integration.[199]

The metadynamics (MTD) is the most popular choice in catalytic applications not only from the first class of methods but also in general.[155,200–231] The most relevant MTD applications will be discussed below. In MTD, an adaptive biasing potential is added on-the-fly during the simulation. The biasing potential is gradually accumulated from small repulsive Gaussian-shaped hills. The Gaussian hill width, height and frequency of deposition are, in the original formulation,[198] fixed during the simulation and act as free parameters that need to be tested for the system at hand. The free energy profile can be estimated directly from the negative of the biasing potential. Numerous modifications of the original MTD have been developed[232] since. In well-tempered MTD,[233] the bias deposition rate automatically decreases with time, leaving the user with a single free parameter to test. In addition, the well-tempered formulation diminishes both the problem of when to stop the simulation and the problem of ''hill surfing'', *i.e.* the behavior whereby the biasing potential overfills the underlying free energy surfaces and pushes the system into high-energy regions. This modification has already found application within the heterogeneous catalysis field in the study by Ghoussoub *et al.*[231] on CO_2 reduction *via* surface frustrated Lewis pairs of hydroxylated indium oxide. An important technical improvement of MTD is the multiple walkers implementation[234] which enables running a number of MTD simulations in parallel on the same free energy surface, which all contribute to the overall history-dependent biasing potential. This technique ported to a supercomputer infrastructure has been successfully used to unravel the highly complex network of reaction pathways leading to the synthesis of methanol on ZnO[206,211] and Cu/ZnO[222] catalysts, accumulating about 2 ns of AIMD simulation time. Lately, a metadynamics-based collective variable-driven hyperdynamics[235] has been employed to study plasma-induced surface charging effects on CO_2 activation on supported M/Al_2O_3 (M = Ti, Ni, Cu) single atom catalysts.[236] Importantly, most of the MTD-type schemes proposed are implemented and made easily available *via* PLUMED,[237] an external plugin that can be interfaced with MD codes through a simple patching procedure.

Besides MTD, other enhanced AIMD schemes that use a biasing potential, including umbrella sampling[197] or integrated tempering sampling (ITS),[238] have also been employed in the field,[239–244] albeit only scarcely. In a typical umbrella sampling simulation, a set of fixed biasing potentials is introduced spanning the entire region of interest in the order parameter. A common choice of potentials is a set of uniformly distributed

harmonic functions with some overlap. For each overlapping region, or 'window', an AIMD simulation is performed yielding a set of partially overlapping histograms for the biased system. The unbiased free energy profile is recovered from biased histograms using schemes such as the weighted histogram analysis method[245] or the dynamic histogram analysis method.[246] The most notable examples of umbrella sampling applications are the QM/MM study on benzene hydrogenation on molybdenum carbide nanoparticles in benzene solvent[242] and Car–Parrinello molecular dynamics (CPMD) based determination of free energies of methanol and water dissociation over TiO_2 surfaces.[240] Lastly, in ITS simulations the particles move in the effective potential corresponding to generalized distribution composed of a weighted sum of normal Boltzmann distributions at a series of temperatures around the target temperature. The free energy profiles can be extracted from ITS simulations by a proper re-weighting scheme.[247] This approach, thus, mimics the replica-exchange[248] type of simulations, in which the enhanced sampling is achieved by running a number of parallel trajectories at different temperatures with a Metropolis-type criterion for exchanging configurations. The appealing property of ITS, unlike other methods such as MTD or umbrella sampling, is the fact that it enhances sampling of all degrees of freedom, doing away with the non-trivial task of choosing a representative order parameter (see below), or so-called collective variable (CV). Combination of replica-exchange with MTD has also been proposed.[249] However, the downside is the limited height of barriers that can be crossed using temperature as a switching parameter. Nevertheless, it was shown to be a suitable approach for low-energy activated processes such as carbene decomposition on a Ni(111) surface[244] or CO diffusion on a Ru(0001) surface.[243]

Thermodynamic integration (TI) has also been extensively used for description of catalytically relevant systems,[250–265] although not as much as MTD. It relies on calculating and subsequently integrating the derivatives of free energy with respect to a reaction coordinate along a reaction path. It can be shown that free energy derivatives are equal to a restorative force acting on the reaction coordinate, hence the alternative name for TI – Potential of Mean Force method. In practice, the integral is typically approximated by a quadrature with quadrature points regularly spaced along the reaction path. The free-energy derivative at each quadrature point is obtained from a constrained AIMD simulation, the blue-moon ensemble method,[266] with the reaction coordinate value fixed.

Choosing between MTD-type methods and TI is not a simple task, as both have their weaknesses and strengths. TI is free of tunable parameters, with the possibility of systematically decreasing the statistical errors along the whole reaction path just by adding more sampling points and/or prolonging the simulation time. In addition, extraction of both kinetic and entropic information is rather straightforward[264] in TI, which is not the case for MTD.[232] However, MTD is much better suited to explore higher dimensional free energy surfaces characterized by multiple collective variables (typically only 2-D or maximally 3-D surfaces[206,222] are explored due to increasing computational costs). As a result, the MTD and TI were sometimes[155,208,212,217,225,258,259] used side by side with TI applied for simple reactions well-described by a single CV, while MTD was employed for more complex ones better described by two CVs.

Notwithstanding the exact dimension of still a very low-dimensional CV space that can be sampled in the enhanced AIMD schemes discussed so far, a good choice of the CV (or a small set of CVs) is essential for obtaining meaningful insight into the reaction. However, a choice of CV properly describing the true reaction process is not simple[267] and no general purpose formula exists to obtain it. The problem is exacerbated for reactions, in which (i) solvent degrees of freedom are expected to play a role in the reaction mechanism, and (ii) a number of competing mechanisms are envisioned without a clear *a priori* preference for a specific one. One of the most often used solutions for both problems is to employ the atomic coordination numbers (CN) as the CVs, which are often flexible enough to accommodate various reaction scenarios. For example, using the CN of the carbon atom in methanol with oxygens of the surrounding water or methanol molecules as CVs, De Wispelaere *et al.*[228] studied the role of solvent in the methanol-to-olefin process over a H-SAPO-34 microporous material (see Fig. 8). Similarly, Martínez-Suárez *et al.*[222] employed three CNs (CN[C–O], CN[O–H], CN[C–H]), to investigate the complex reaction network of methanol synthesis over a Cu/ZnO nanocatalyst characterized by numerous competing mechanisms with a number of distinct C_1 species identified. There are also a couple of methods to study catalytic reaction dynamics in an unbiased way, circumventing the problem of the correct choice of reaction coordinates, namely the quasiclassical trajectory (QCT) simulations[268] and transition path sampling[269] (TPS). In QCT, starting from the previously identified transition state a set of MD trajectories are propagated in an unbiased way, with the initial velocities chosen using quantum-mechanical population of vibrational states at a chosen temperature. Such simulations are particularly important for cases where the zero point vibrational energy (ZPVE) is large, since ZPVE is neglected in AIMD simulations where nuclei are treated as point charges moving on the electronic potential energy surface. Similarly, TPS creates an ensemble of unbiased reactive trajectories starting from the initial reactive trajectory, performing an important sampling in the trajectory space. The reactive trajectories corresponding to different reaction mechanisms are represented in the ensemble in proportion to the relative likelihood of the system to choose the particular mechanism, taking into account the effects of entropy and temperature. Both methods are capable of providing both the product selectivities (corresponding to differences in free energies of the products) and kinetic reaction rates for complex reaction networks. However, these methods also incur significant computational costs associated with a need to generate thousands of AIMD trajectories of a few ps in length to obtain converged results. Hence, they are mostly used as a tool for a qualitative understanding of complex reaction networks, possibly guiding another more quantitative investigation using, *e.g.*, enhanced sampling AIMD methods such as in the case of propane cracking over acidic chabazite by Bučko *et al.*[255] The application of both methods in the heterogeneous catalysis field has focused so far

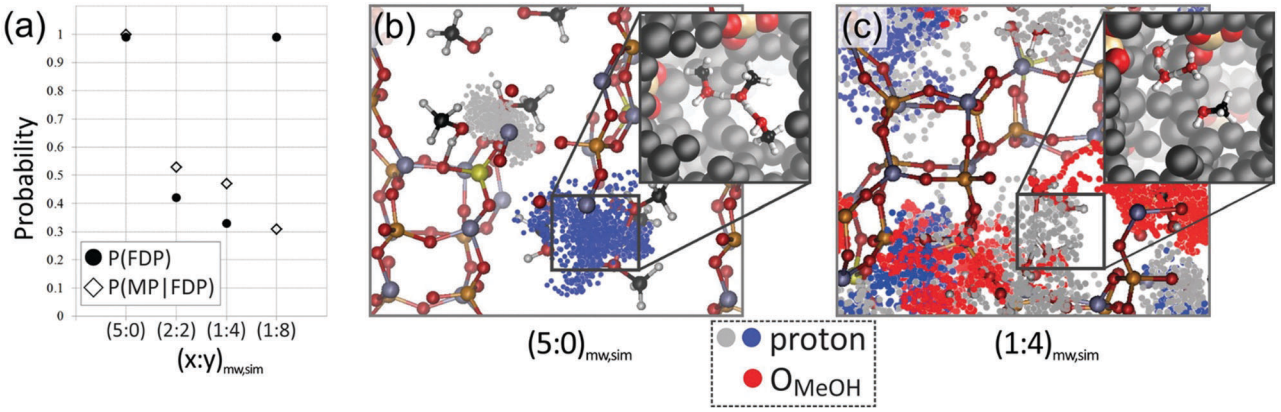

Fig. 8 (a) Probability for framework deprotonation (FDP) and probability of methanol protonation when the framework is deprotonated (MP|FDP) during 50 ps MD simulations of different methanol–water mixtures adsorbed in H-SAPO-34 at 330 °C and around ambient pressure. (b and c) H-SAPO-34 loaded with a (5 : 0)mw,sim and (1 : 4)mw,sim methanol–water mixture. The gray, blue, and red dots represent the positions of the two acid protons and methanol oxygen atoms, respectively. The insets show snapshots of the MD run with highlighted acid sites. (*x*:*y*)mw,sim stands for a simulation with *x*MeOH and *y*H_2O molecules per Brønsted acid site. Reprinted with permission from ref. 228. Copyright 2016 American Chemical Society.

exclusively on reactions in acidic zeolites, investigating linear hydrocarbon cracking in H-MFI[270,271] and in H-CHA,[255] alkane dehydrogenation[252] and methanol coupling[272] in H-CHA, and alkene methylation by methanol in H-MFI.[273]

4.2.1. Physisorption and chemisorption. The catalytic process typically starts with the formation of a reactant/adsorption complex. With increasing temperature, the characterization of this complex using a single configuration becomes problematic. A need for statistical treatment is particularly important for molecules with rather weak and non-specific interactions with the catalyst as was first exemplified in the study of Bučko *et al.*[255] The authors quantified the temperature effects on physisorption of propane in H-CHA zeolites using only equilibrium MD, which is sufficient for physisorbed systems. They showed that at low temperatures (100 K) the propane is bound to a Brønsted acid site but at 800 K, a typical catalytic cracking temperature, propane is mostly detached from the catalytic site with its movement in the zeolite being much less restricted. This temperature-dependent change in adsorption behavior is associated with a significant decrease of adsorption energy of about 20 kJ mol^{-1}. The follow-up study by Göltl *et al.*,[274] extended to other alkanes and employing higher levels of theory, confirmed the previous findings of Bučko *et al.* In addition, Göltl *et al.* proposed a simplified approach to obtain dynamically averaged adsorption energies from shorter equilibrium MD simulations. A cheaper way to include the temperature corrections to adsorption enthalpy and entropy was used by Tranca *et al.*[271] who complemented their static DFT calculation with temperature corrections derived from Monte Carlo simulations using empirical force fields. In line with experimental data and previous studies by Bučko and coworkers[255,274] a decrease of adsorption enthalpies with increasing temperature and temperature dependence of adsorption entropies were reported. A similar approach to that of Tranca *et al.*[271] has been recently used by Van der Mynsbrugge *et al.*[193] to investigate the influence of pore geometry on monomolecular cracking and dehydrogenation of *n*-butane in various acidic zeolites. The effect of spatial constraints on the free energy of adsorption was also studied by Bučko *et al.*[257] who probed propane adsorption in two pores of different dimensions in H-MOR zeolites. Their free-energy profiles from TI using the blue-moon ensemble clearly showed that with increasing temperature from 0 to 800 K the entropy shifts the balance towards propane occupying the less confined pore, despite the larger adsorption enthalpy in the smaller pore.

Moving beyond physisorbed complexes but staying in the zeolite hydrocracking field, Hajek *et al.*[229] showed how inclusion of temperature effects beyond the harmonic approximation changes the relative stabilities of four types of pentene adsorption complexes in acid zeolite H-ZSM-5 including both physisorbed and chemisorbed species. Using only the harmonic approximation, the so-called π-complex, a complex bound to Brønsted acid sites *via* a C=C double bond, is found to be the most stable species at 323 K (see Fig. 9). However, upon inclusion of dynamical effects from equilibrium MD, stability of chemisorbed species, an alkoxide, becomes basically equal to that of the π-complex. The MTD-based free energy profiles for the π-complex → alkoxide transformation confirmed this finding, showing also that formation of the chemisorbed complex is an activated process with a barrier of approximately 40 kJ mol^{-1}. In the following year, the same group[230] extended their (biased) AIMD investigation to other C_4–C_5 alkenes focusing on dynamical effects at operating temperatures of catalytic alkene cracking of about 800 K. They found that another chemisorbed species, an ion-pair called the carbenium ion, is the prevalent species in the zeolite channels under these conditions, in stark contrast to predictions from static calculations favoring the π-complex. The change in stabilities of various species was attributed to entropy effects which may disfavor the formation of tightly bound physisorbed or chemisorbed complexes such as the π-complex or alkoxide. This is in line with the observation that the stabilization of the carbenium ion relative to other species increased from primary to branched C_4–C_5 alkenes. Recently, a general approach for estimation of adsorption free energies in zeolites based on enhanced AIMD simulations using TI has been

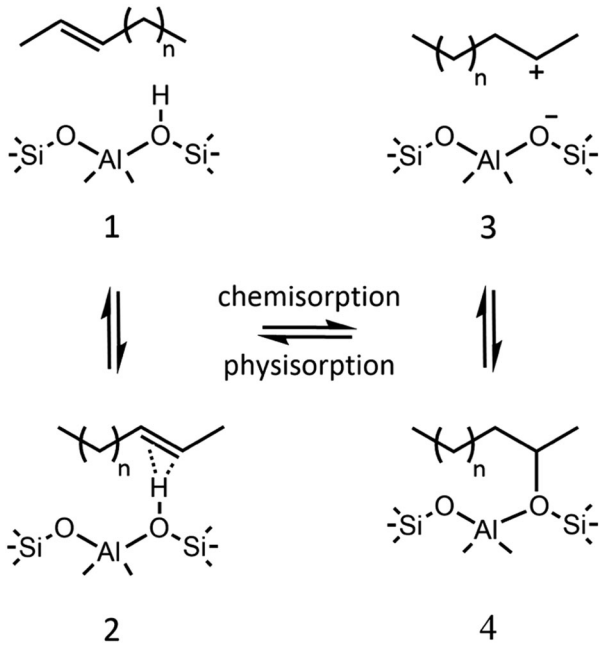

Fig. 9 Illustration of the different intermediates upon alkene (2-pentene) adsorption in the presence of a Brønsted acid site (BAS): (1) physisorbed van der Waals complex, (2) alkane π-complex, (3) chemisorbed carbenium ion and (4) chemisorbed alkoxide. Reprinted with permission from ref. 230. Copyright 2017 Elsevier.

proposed[263] and applied for adsorption of small-molecules at Cu sites in chabazite.

Another aspect of a catalytic process, co-adsorption of reacting species, was studied by De Wispelaere *et al.*[228] providing insight into water–methanol and water–propene competition for catalytic sites in a H-SAPO-34 microporous material, an important catalytic system for the methanol-to-olefin (MTO) process. Despite larger adsorption enthalpies of methanol than water, obtained both from static and dynamic calculations at 330 °C, the equilibrium MD runs with methanol–water mixtures showed that water and methanol have nearly equal probability to occupy the Brønsted acid site. As a result, the mixed methanol–water clusters compete for the acidic proton and exhibit lower apparent proton affinity than either of the pure systems (see Fig. 8). This means that methanol protonation, an elementary activation step for the MTO process, is slowed down in water. Similarly, the water is reported to displace propene from Brønsted acid sites in the co-adsorption scenario, decreasing the probability that propene becomes activated for further reaction toward the formation of cyclic hydrocarbon pool species.

4.2.2. Quantifying entropy effects in simple reactions. To understand the role of entropy and to assess the limitations of harmonic transition state theory (HTST) in heterogeneous catalysis it is instructive first to look at a series of simple reactions of hydrocarbons in acid zeolites studied by Bučko and coworkers.[251–253,257] In their first study,[251] Bučko *et al.* tried to understand the origin of experimentally observed regioselectivity in proton exchange of isobutane in acid zeolites, *i.e.* why the methine (CH) group in contrast to the methyl group of isobutane is completely inactive. While the activation free energies from HTST at 800 K were basically the same for both groups, the barriers from TI-based AIMD differed by 55 kJ mol^{-1} in favor of proton exchange *via* the methyl group. This effect was shown to originate in the entropic contributions and was clearly related to different steric restrictions for proton access to two types of carbon groups in the early stages of the proton transfer. The differences in entropy contributions correlated well with relative probabilities of methine and methyl group to form adsorption complexes with Brønsted acid sites, which could be obtained already from the unbiased molecular dynamics of the reactant state. Hence, the failure of HTST to account for the regioselectivity can be again (see Section 4.2.1) traced back to inadequate representation of the reactant state. In 2010, Bučko *et al.*[253] briefly analyzed entropy effects for propane cracking both using TI-based AIMD simulations and the static harmonic approximation. The discrepancy between the two approaches for entropy estimation amounted to as much as 80 kJ mol^{-1} at 800 K. Moreover, the activation entropies from the two approaches differed even qualitatively, with HTST associating larger entropy with the transition state, a pentavalent carbonium cation, while AIMD showing that the entropy of the reactant state, a loosely bound propane in the cavity, is larger than that of the transition state. Again, the shortcomings of a static approach were related to an improper description of the loosely bound reactant state. Another shortcoming of the static approach, inability to account for reaction intermediates which are not potential energy stationary points, has been highlighted in the case of propane dehydrogenation in H-CHA zeolites.[252] The TI-based AIMD study complemented by TPS simulations revealed a more complex mechanism than expected based on static TS search, which originated in entropic stabilization of the propyl cation, a non-stationary point on the potential energy surface. TPS simulations showed that the propyl cation is an important branching point in the dehydrogenation mechanism, undergoing various transformations (internal rearrangement, rotations, translation in the cavity) during its lifetime, eventually collapsing directly to various stable products such as propene, the main experimentally observed product. As a result, creation of the main HTST-based intermediate, the alkoxide, can be, and in most dynamical trajectories is, avoided. Lastly, the role of spatial constraints on the reactivity of propane in the zeolite catalyzed cracking was investigated[257] in the model system of acid mordenite containing larger and smaller cavities. The activation free energies, derived from TI-based AIMD simulation, were lower for cracking in stronger confinement in the smaller cavity mostly due to different entropies of activation. In both cavities, the entropies of the TS ensemble are lower than those of the reactant ensemble, however, the reactant in the narrower pore is rather confined to start with, so the relative entropy loss is smaller for a narrower pore. This explanation was justified by the significantly higher collision probability between propane and the active site in the smaller pore. However, the reactant can access the small pore only from the larger one, which is an activated process that tilts the balance toward cracking in the larger pore, in line with the experimental observations. This model study nicely illustrates the intricacies of even rather

simple reactions in confined environments and a complex interplay of entropic and enthalpic effects that need to be considered under the working conditions of a catalytic process.

The entropy effects and HTST limitations in simple reactions were analyzed and discussed also outside the field of zeolite catalyzed hydrocarbon conversions, albeit only to a limited extent. Sun *et al.*[244] compared reaction entropies, free energies and reaction rates for carbene (CH_2) decomposition on an Ni(111) surface obtained from harmonic approximation and enhanced AIMD simulations using integrated tempering sampling (ITS). The HTST values were qualitatively consistent with the results from ITS-AIMD simulations; however, rates were about an order of magnitude larger for HTST and HTST reaction free energies were about 0.15 eV larger. The authors proposed *ad hoc* generalization of HTST including multiple configurations in the partition functions of the reactant and product states, which improved the agreement with ITS-AIMD reaction free energies and entropies. In addition, Sun *et al.* tested some of the general assumptions underlying the generalized TST (non-recrossing, quasi-equilibrium between reactant and transition states) by comparing generalized TST rates obtained from biased ITS simulations and true reaction rates from a TPS-inspired unbiased approach,[275] which directly samples the (un)reactive trajectories. Rates from both approaches were in very good agreement. Hence, their general conclusion was that, for such a simple surface reaction, the basic assumptions of generalized TST theory are valid but the harmonic approximation is an over-simplification even in this simple case. The importance of entropic effects was further highlighted in the umbrella sampling CPMD simulation of methanol and water dissociation on TiO_2 surfaces.[240] The authors reported that with increasing temperature, the dissociation of water on the anatase surface becomes more favorable than that of methanol due to entropic effects. The large entropy loss on the side of methanol was associated with hindered rotation of the methyl group after dissociation. As a last example, a conceptually nice case study of temperature effects was presented by Schnur *et al.*[241] for H_2 dissociation on water-covered Pt(111), Ru(0001) and Pd/Au(111) surfaces. They reported how distortions of initially ice-like hexagonal water structure over metal surfaces at room temperature lead to an increase in the free energy barrier for H_2 dissociation by 0.15 eV. The increase in free-energy barrier is related to an irregular shape of the hexagonal water rings under thermal conditions which makes the propagation of the spherical H_2 molecule through the water layer much harder.

4.2.3. Complex reaction mechanisms and reaction networks. Many of the mechanistic studies using AIMD simulations focused on rather simple reactions, aiming primarily at properly quantifying the temperature effects for well-known reaction mechanisms, often with an assumption of a unique reactant/TS/product sequence. However, in many industrially relevant heterogeneous catalytic processes, complex reaction mechanisms with multiple reaction channels and side reactions are at play.

The first step on the way to simulate more realistic reaction processes is to allow for multiple transition states connecting the reactant/product pair. One can either: (i) consider more reaction channels chosen based on previous reports and/or chemical intuition, tailor mechanism-specific collective variables, evaluate free-energy profiles and compare, such as in the case of single atom catalysis of O_2 activation and CO oxidation over Rh_1/γ-Al_2O_3,[213] or (ii) use a more bias-free approach such as TPS simulations to create unbiased reactive pathways which connect the reactant and product basins without a need to constrain the transformation mechanism search using collective variables. The latter approach has been used by Bučko *et al.*[255] to choose between possible realizations of the first reaction step of protolytic cracking of propane using acid chabazite as a catalyst at realistic reaction temperature (T = 800 K). The free-energy profile of the dominant mechanism determined in TPS simulations was later refined by TI-based AIMD.

The possibility of forming multiple products from a single transition state represents an additional layer of complexity in the realistic reaction mechanisms. An approach specifically constructed to meet the challenge is quasiclassical trajectory (QCT) simulations shooting a set of unbiased MD trajectories from the TS, followed by the analysis of the end products of the simulations, which provides an estimate of the product distributions at operating temperature. The QCT has been used in pioneering studies of Bell and coworkers on product selectivities for alkane cracking[270] and alkene methylation by methanol[273] over acid zeolite H-ZSM-5. In both studies, the authors reported qualitative discrepancy between product distributions obtained by static and dynamic reaction pathways obtained from QST suggesting that the high temperature pathways, *i.e.* the free-energy pathways, differ significantly from 0 K potential surfaces. Besides QST, the TPS simulations, with a judicious choice of the order parameter that accommodates multiple product scenarios, may also be used to estimate product selectivities as shown for the propane dehydrogenation mechanism in H-CHA zeolites[252] (see Section 4.2.2 for a more detailed discussion).

The first truly complex reaction network investigated using dynamical *ab initio* methods was the methanol formation from CO on a defective hydroxylated ZnO(000$\bar{1}$) surface.[206] The reaction network was investigated in a two-stage procedure starting with a multiple-walker metadynamics simulation with constraints on carbon diffusion into or away from the surface and a coarser biasing setup to speed up the exploration of the vast free surface. The gross free energy surface obtained from the exploratory mapping already contained the molecular species considered in previous studies and also yielded additional subspecies. In the second stage, the most important transformations were refined by individual one- to three-dimensional metadynamics runs with collective variables tailored for a particular transformation. Altogether, about ten stable intermediate species were identified being interconnected *via* five distinct reaction channels leading eventually to full hydrogenation of the CO molecule. In the follow-up study by the same group,[222] the complexity of the catalytic system as well as its relevance for experiment was increased further by switching the focus to methanol synthesis from CO_2 over a Cu/ZnO nanocatalyst, which was modelled using a Cu_8 cluster deposited on an O-terminated and partially hydroxylated ZnO(000$\bar{1}$) surface. The relevance of this catalyst model under conditions of industrial

methanol synthesis was established from the surface phase diagram constructed using *ab initio* thermodynamics[276] (see Section 3). Rather than providing an accurate free energy surface with converged barriers and reaction energies, the authors aimed at exploring the breadth of the reaction network identifying all possible types of C_1 species (more than 20) and reaction channels present over the $Cu_8/ZnO(000\bar{1})$ catalyst (see Fig. 10). Their almost 2 ns long exploratory CPMD-based metadynamics run in three-dimensional collective variable space also included several well-known side reactions, such as coking, methanation and water-gas shift reactions. In addition, the study highlighted the need to systematically include the surface region at the interface of the catalyst and gas phase as an active reaction space since (i) the Cu cluster is highly dynamic at 500 K, changing from 2D planar structures lying flat on the ZnO surface to "spherical" 3D morphologies, with Cu atoms migrating across the catalytic system, (ii) there are strong-metal support interactions manifested in spontaneous creation of O vacancies, which migrate from atop the $ZnO(000\bar{1})$ surface layer onto the Cu_8 cluster, thus giving rise to O adatoms or to OH adspecies after a subsequent reaction of these O atoms with H adspecies from Cu_8, and (iii) Cu atoms interact strongly with C_1 species which may cause a spatial redistribution of some of these Cu atoms over the support. All these facts give rise to various active morphologies and new putative active sites created *in situ* that can stabilize reactant, intermediate, and product states of the involved C_1 species. Admittedly, the degrees of freedom responsible for these "surface-reconstruction" processes are not accelerated by the metadynamics but occur on the time scale that is accessible only to nonbiased AIMD dynamics, which limits the configurational space of the catalyst transformations that could be accessed. Also, the AIMD setup did not allow for an on-the-fly insertion (or removal) of reactants in the sense of grand canonical equilibrium with suitable reservoirs for these molecules, which might be needed for faithful and automated description of the industrial process. Nevertheless, along with the reactive global optimization approach discussed in Section 2.5.3, this work, which couples *ab initio* thermodynamics with extensive biased AIMD simulations to faithfully map a complex reactive network under working conditions, presents one of the most comprehensive examples of using *ab initio* simulations to study catalytic processes.

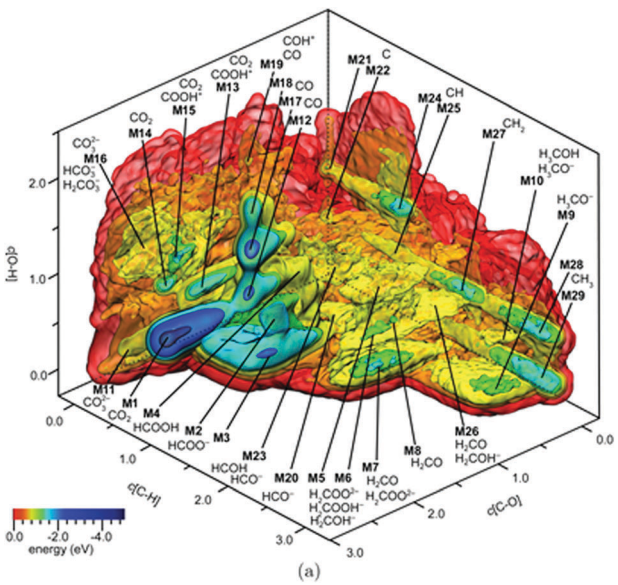

Fig. 10 Free energy landscape from the metadynamics sampling of methanol synthesis based on CO_2 over the reduced $Cu_8/ZnO(000\bar{1})$ catalyst surface model (a). The coordination numbers c[C–O], c[C–H], and c[O–H] (see text) were employed as collective variables (CVs) to describe the interaction of the carbon atom with the oxygen atoms of the top layer of ZnO and the two oxygen atoms of the reactant CO_2, the carbon atom with all hydrogen atoms in the system, and all hydrogen atoms with the two oxygen atoms of the reactant CO_2, respectively. Relative free energies ΔF are reported according to the shown color scale. Bold capital M plus a number labels distinct free energy minima. Reprinted with permission from ref. 222. Copyright 2015 American Chemical Society.

5. Simulating reaction kinetics for mechanistic analysis and catalyst optimization

5.1. Basic principles

Previous sections have illustrated the power of modern computational approaches for unraveling the nature of catalytic ensembles and studying individual chemical transformations under realistic *operando* conditions. Yet, most of the mechanistic studies of practical catalytic reactions and detailed analysis of complex reaction networks commonly encountered in heterogeneous catalysis is still limited to electronic structure calculations on the 0 K/UHV models. Such an approach has been proven over the last two decades to be extremely powerful in unraveling the fine mechanistic details of the chemical transformations underlying heterogeneously catalyzed reactions.

In practice, even the simplest of such processes are represented by complex networks of competing and parallel elementary reaction steps involving different sites. Quantum chemical calculations provide direct access to the rate constants for each of these steps. However, the resulting information has often only a limited value by itself as the means to provide direct guidance for the optimization and design of an improved catalytic process. The next step in this direction requires the reduction of the mechanistic complexity and a bridge between the microscopic insights into surface reactions and the macroscopic kinetics of the catalytic process. This step can be readily accomplished through microkinetic modeling, which is currently one of the most popular and powerful approaches to analyze reaction mechanisms and reaction kinetics in both experimental and computational catalysis.[277–280] Microkinetic models can be directly used to identify which intermediates or specific reaction paths dominate the formation of a particular product, providing a practical tool to directly optimize process conditions and even to guide the *in silico* design of improved catalysts. The theory and applications of first principles kinetic modeling have been extensively discussed in a number of excellent recent reviews.[281–286] In this section, we therefore limit ourselves to only briefly outlining the fundamental

approximations made in such methods and highlighting their power by discussing selected relevant examples from recent literature.

In microkinetic modeling, rate constants of elementary reaction steps are used in the mean-field differential equations that describe the kinetics of the reaction.[287] The output of such a simulation is production rates and surface concentrations. The mean-field approximation implies that the adsorbates are not correlated spatially and therefore their mutual interactions are neglected. Although the development of more advanced kinetic modeling approaches accounting for such correlation effects is an active field of research,[288–290] the conceptual simplicity and ease of their practical implementation determine the success and widespread utilization of the mean-field phenomenological kinetic modeling approaches.

The formulation of a microkinetic model begins by expressing all the rates of elementary reaction steps in a catalytic reaction network *via*

$$r_n = k_n \prod_i c_i^{\nu_{i,n}} \tag{7}$$

where k_n stands for the rate constant of the elementary reaction step n, c_i is the concentration of the component i and $\nu_{i,n}$ is the stoichiometric coefficient for species i in step n. The time-dependent concentration of the component or surface coverage i is calculated by

$$\left(\frac{\mathrm{d}\theta_i}{\mathrm{d}t} = \sum_j \nu_{ij} r_j f_j(\theta_1, \ldots, \theta_N)\right)_{i=1-N} \tag{8}$$

where θ_i is the surface coverage of species i at time t, ν_{ij} is the stoichiometric coefficient for species i in step j, r_j is the rate of the reaction j and f_i is a function of several coverages involved in step j. This system of differential equations describes effectively all chemical processes taking place in the catalytic system. In practice, one solves this system of equations numerically until a steady state is reached for the overall reaction system. In these equations, the rate constants are commonly computed in the framework of the transition state theory (TST) as

$$k = A \exp^{-\frac{E_a}{k_B T}} \tag{9}$$

where E_a is the intrinsic activation barrier for a particular elementary reaction step readily accessible from DFT calculations, while the pre-exponential factor A can be represented as

$$A = \frac{k_B T}{h} \frac{Q^{TS}}{Q} \tag{10}$$

where k_B, h and T are, respectively, the Boltzmann constant, Planck constant and the temperature, and Q^{TS} and Q are the partition functions of the transition state and reactant state, respectively.

These partition functions reflect the entropic effects associated with the chemical transformation and they can also be estimated from the results of DFT calculations by, for example, treating each degree of freedom in the reactive system by a frustrated vibration that is in turn treated as a harmonic function. Despite other more advanced approaches involving the explicit consideration of the translation degrees of freedom for the adsorbates and accurate sampling procedures allowing an increase in the computational accuracy manifold,[185,291–293] semiempirical and phenomenological approaches for estimating pre-exponential factors provide useful practical solutions for constructing microkinetic models with a high predictive power.[294–298]

When the overall description of the catalytic system is constructed, the apparent activation energy (E_a^{app}) can be computed from the model as

$$E_a^{app} = k_B T^2 \left(\frac{\partial \ln(r_t)}{\partial T}\right)_P \tag{11}$$

And the rate-determining step can be identified by using the degree of rate control (DRC) parameter[299,300] that effectively describes how the overall kinetics is influenced by a particular elementary step i considered in a microkinetic model:

$$\chi_{RC,i} = \left[\frac{\partial \ln(r_t)}{\partial \ln(k_i)}\right]_{k_{j\neq i}, K_i} \tag{12}$$

The DRC can be viewed as a weighing factor that allows directly relating such macroscopic rate parameters as the apparent activation energy and reaction orders and the microscopic characteristics as the elementary reaction rates and activation barriers for elementary steps.[27] In principle, the DRC analysis method can be directly employed for studying and optimizing catalytic reactions. The important advantage of this methodology is that it does not require the complete derivation of the complete catalytic mechanism for its successful application.[302]

To summarize, microkinetic modelling is a powerful tool for analyzing complex reaction mechanisms and constructing predictive models capable of connecting the microscopic description of the reactive systems with measurable macroscopic activity descriptors. The mean-field approximation underlying these methodologies not only facilitates the analysis of the results, but also ensures the high efficiency of the associated calculations as well as its straightforward implementation in a working code. This has given rise to a number of programs for advanced microkinetic simulations and analysis of their results which are currently available to the scientific community.

5.2. Microkinetic modeling and linear energy relations for catalyst design

If the kinetic parameters for a large enough selection of catalyst candidates are available, microkinetic modeling becomes a practical computational tool for catalyst design and optimization. However, the explicit calculation of the activation barriers for different reactions and catalyst formulations using accurate electronic structure methods is a highly resource- and time-consuming task. The theory-guided catalyst design can be greatly assisted through so-called linear scaling relationships, which establish correlations between the adsorption energies of specific intermediates and the activation barriers for the related chemical transformations.[303–305] The existence of such linear scaling relations has been demonstrated for a wide range

of reactions over different catalysts.[306–308] When such relations hold, they allow for a significant reduction of the number of independent parameters which determine the catalyst activity. They therefore facilitate enormously the *in silico* search for an optimal catalyst,[309–311] but at the same time place fundamental limits on the maximum achievable activity or selectivity. The search for ways to break these scaling relations is currently an active research topic in computational catalysis.[301,312,313]

As an illustrative example of the power of the integration of DFT modeling and microkinetic simulations for catalysis design, let us discuss one of the most classical and important catalytic processes – ammonia synthesis ($N_2 + H_2 \rightarrow NH_3$), where the catalyst performance is actually limited by such scaling laws.[303] According to the Sabatier principle, an ideal catalyst for this process should be active enough to promote the cleavage of the strong bond in molecular N_2 and, at the same time, bind various NH_x species rather weakly so that they can be removed from the surface by the hydrogenation as the NH_3 product. However, because the adsorption energies for these intermediates and the activation energies for the elementary steps are correlated with each other, one cannot independently adjust them to maximize the performance. Microkinetic simulations based on the results of DFT simulations for a wide range of catalytic materials have revealed clear volcano-type relations between the catalytic performance and binding energy of atomic nitrogen (Fig. 11, "plasma-off"), which has been identified as a suitable reactivity descriptor for this process.[314,315]

In a recent work, Go, Hicks, Schneider and co-workers proposed a way to overcome the fundamental limitations imposed by such linear relations by coupling the conventional catalysis with non-thermal plasma.[301] Indeed, the correlation between the adsorption and activation energies is directly related to the intrinsic chemistry of the catalytic materials. The thermocatalytic limit for ammonia synthesis was assessed through a microkinetic model based on the DFT-computed energetics for the ammonia synthesis reaction intermediates treated in the frozen-adsorbate limit and tabulated standard entropies for the gaseous reactants. This model was then adjusted to include the influence of the N_2 vibrational excitation on the elementary reaction rates. It was proposed that the vibrational excitation of the gaseous N_2 through the interaction with the non-thermal plasma would increase the energy of the initial state by the energy of vibration, resulting in an effective lowering of the associated transition state. This new model provided the initial evidence that the optimal catalysts and active sites in plasma catalysis may differ from those in thermal catalysis (Fig. 11 – "plasma-on"). Importantly, besides enhancing the overall rate of the catalytic reaction over the open terrace sites (Fig. 11b), the selective excitation of the N_2 vibrational states of the reactant was shown to shift substantially the maximum of the volcano curve computed for the more reactive step-sites from the expensive Rh and Ru to the cheap and earth-abundant Ni and Co catalysts. These theoretical predictions were found to coincide very well with the experimentally determined reaction rates under plasma-induced catalysis conditions.[301]

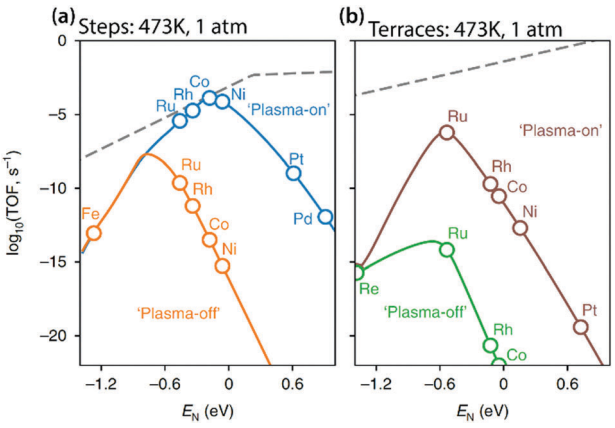

Fig. 11 The calculated rates of ammonia synthesis on (a) step and (b) terrace sites under thermal ("Plasma-off") and plasma-induced ("Plasma-on") conditions. Reprinted with permission from ref. 301 Copyright 2018 Springer Nature.

5.3. Microkinetic modeling for deep mechanistic analysis and process optimization

The derivation of straightforward activity relations can be complicated by the high complexity of the chemical ensembles acting as the active sites. Effects such as substrate pre-activation and active site relaxation can induce deviations from the expected activity trends. Furthermore, processes such as the multiple-site activation, active site cooperativity and confinement-induced reactivity make the definition of simple reactivity descriptors suitable for the large-scale computational screening of different catalysts very difficult if not impossible.[317]

In such cases, MKM can be used to reduce greatly the mechanistic complexity of the DFT-computed reaction networks and identify the optimal reaction conditions so that the desirable reaction path is enabled resulting in the enhanced selectivity of the overall catalytic process. A recent study by Liu *et al.* on the mechanism of isobutene–propane alkylation by faujasite-type zeolite catalysts illustrates such an approach.[316] A detailed mechanistic DFT analysis of the extended catalytic network underlying the isobutene–propene alkylation process using realistic models of La-containing low-silica faujasite-type zeolite catalysts has been carried out. A particular focus was laid on enhancing the selectivity to the desirable alkylate product, while suppressing the paths which resulted in the deactivation of the zeolite catalysts. Microkinetic models were constructed, based on the DFT-computed energetics of the elementary reaction steps and augmented by configuration-bias Monte Carlo simulations to more accurately account for the relative concentrations of reactants at the reaction centers. The simulations clearly showed that the production of the desirable C7 alkylate over the La-FAU catalyst is favored when operating the reaction at a high pressure and low-temperature (Fig. 12). Furthermore, the mechanistic insights obtained through the MKM simulations pointed to the fundamental requirement of the micropore structure of the hypothetic optimal alkylation catalyst. Given the large size of the hydride transfer complex of isobutane and carbenium ions, a zeolite structure with large pore size should be beneficial.

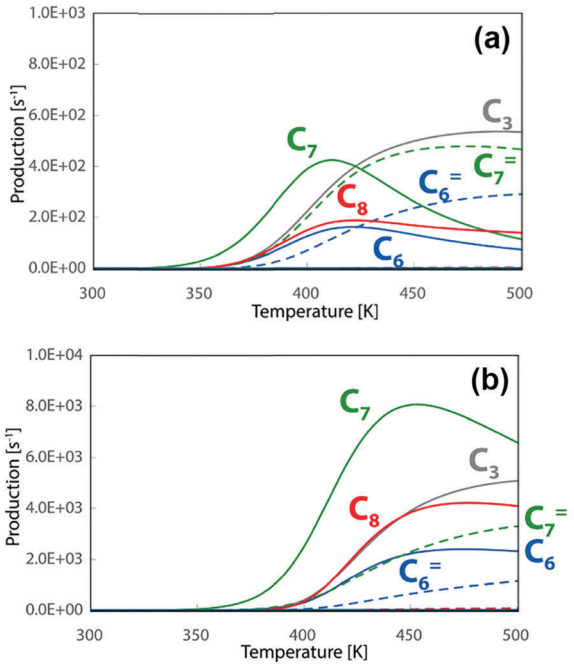

Fig. 12 Microkinetics simulated production rates of the alkylation reaction by the La-FAU model at the total pressure of (a) 3.2 bar and (b) 32 bar as a function of temperature. The desirable reaction product is the C7 alkylate dominated by 2,2-dimethylpentane (which rapidly isomerizes to 2,3-dimethylpentane by a secondary reaction that was not included in the model) produced by the reaction of *tert*-butyl cation and propene stabilized inside the La-FAU pores. The second major C8 product is formed *via* the self-alkylation path. The formation of unsaturated hydrocarbons at elevated temperatures gives rise to oligomerization reactions inside the pores resulting in the pore blockage and rapid catalyst deactivation. Reprinted with permission from ref. 316. Copyright 2017, American Chemical Society.

On the other hand, the microkinetic simulations highlighted the importance of high isobutane occupation in the zeolite micropores that would be optimally realized for the small-pore zeolites. Based on these conflicting requirements on pore size dimensions, the authors proposed that a bimodal channel structure should be looked for in an optimal catalyst.[316]

6. Boosting catalyst design with machine learning approaches

6.1. Machine learning in chemistry

Over the past few decades chemical science has produced gigantic amounts of data. This, aligned with the maturing of practical data science approaches and the enormous growth in computational power, has begun to render new Big Data strategies efficient for discovering new correlations, developing models, and making profound predictions.[318] One of the most powerful strategies of this sort is machine learning (ML), the utilization of which for data analysis has spread rapidly in computational chemistry.[319] The possibility of making fast predictions with accuracy comparable to conventional computational chemistry makes ML methods especially appealing, not only as a research tool but as a vital ingredient of the catalysis-by-design strategy. There is a noticeable expansion of ML approaches to data analysis in different areas of chemistry. ML models are currently employed for studying chemical reactions, predicting properties of different chemical substances, materials design and the development and testing of new continuous and discrete descriptors.[320–332] Some of these approaches allow for the optimization of existing chemical reactions and even the discovery of yet unknown ones.[333,334] Important applications of the ML techniques have also been witnessed in recent years in different branches of catalysis sciences.[305,329,335–338]

The availability of the vast, machine-readable and readily accessible scientific data, together with the enormous available computational power of modern CPUs and GPUs, which are capable of carrying out fast and cheap calculations, gives rise to a situation whereby ML approaches are becoming an important ingredient of rational design strategies for catalysis.[339,340] So far, ML techniques have already been employed as an enabling technique to achieve breakthroughs in the optimization of chemical processes.[325,334] In this section, we present a concise overview of the most important recent applications of ML in catalysis with a special focus on the specific difficulties and challenges in the field.

6.1.1. Machine learning approaches in chemistry. Machine learning is a rapidly growing field of computer science, where algorithms are trained to find empirical correlations in data. The concept of ML is conventionally attributed to the work by Arthur Samuel, dating back to 1959, who described an approach to teach machines to play checkers.[341] However, one could track the seminal studies in machine learning to as early as the 1940s, when the first artificial neural networks (NN, see below) were introduced by McCulloch and Pitts.[342] Although the first academic ML studies have appeared already more than 60 years ago, the peak in ML tools as a practical technology has been reached only in 2016 according to the research by Gartner Inc.[343]

ML approaches can be classified in a variety of ways based on the particular approach employed for solving practical problems. One of the widely employed classifications distinguishes supervised and unsupervised learning. For the latter, the learning is performed on a training dataset, in which only input values are provided to the algorithm without the corresponding output values. By contrast, the supervised learning is carried out on examples of input–output pairs with the possibility of generalization of the output prediction to an expanded or even different input dataset of a similar type. Another approach is to distinguish the ML strategies based on the type of predictions that the model delivers, which are classification, regression, clustering, dimensionality reduction and density estimation. In this case, we differentiate the type of task which a machine needs to accomplish, that is, to classify types of data, calculate some output from input values, separate data into different classes, reduce data descriptors or estimate distributions.[344]

The basic ML approaches such as linear and logistic regression, support vector machine (SVM), decision trees, and random forest are well established in applied data science.[344] The basic strategies underlying the most commonly employed methodologies are

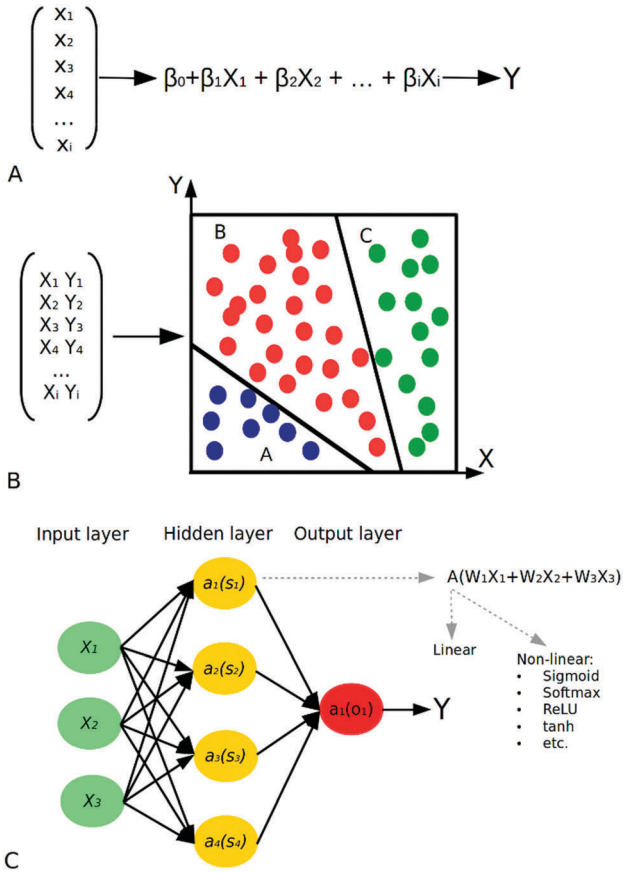

Fig. 13 A schematic representation of the three main ML approaches often employed for modeling of catalytic processes: (A) linear regression, (B) support vector machine and (C) Neural Network approaches. In the linear regression (A) a linear relation is sought between a set of descriptors X_i and the activity measure Y. The support vector machine (B) processes the input descriptors X and Y and separates them into classes (A, B, C). The Neural networks (C) usually contain three types of layers: input, hidden and output. The model accepts the input parameters in the input layer, while the neurons of the hidden layers reevaluate these input parameters. Every neuron is a composition of an activation function a_n and a summatory s_k. The output layer recalculates the results from the inputs and produces the final output of the network. The activation functions $a_n(s_k)$ may be either linear or non-linear depending on the formulation of the network.

schematically illustrated in Fig. 13. In chemistry and catalysis, the selection of a particular ML strategy is commonly based on the type of task to be carried out. Linear regression (Fig. 13A) is a common strategy for QSAR/QSPR studies in drug design and QSPR studies in materials science and it is based on the idea that a specific descriptor – a specific measurable parameter – can be identified as an activity measure. The SVM model (Fig. 13B) operates in a multidimensional space that is built from various descriptors potentially reflecting the target characteristics. The separation of the input values in the multidimensional space can be achieved by linear, polynomial or hyperbolic functions (depending on which particular one is used as a kernel in the model).[345] SVM models are used in data screening for the identification of efficient catalysts or porous materials with optimal adsorption characteristics.[345–347] ML procedures are commonly applied to sufficiently large training datasets, making statistical procedures indispensable for estimating model performance and accuracy, and ultimately for determining the ML fitting properties. For example, principal component analysis (PCA) was employed to identify suitable descriptors for organometallic complexes.[348,349]

The most recent emergence of the Deep Learning concept resulted in widespread acclaim for ML technologies.[343] Within this concept, artificial neural networks (NNs) are used to find patterns and correlations in the data (Fig. 13C). These networks are built from interconnected layers of neurons, which may formally resemble linear regression functions if the so-called linear activation functions are employed for their construction. However, nowadays a commonly accepted standard in the field necessitates the use of non-linear activation functions in the neurons. The output values from the neurons serve as the input for the next hidden layer or the output layer generating the final output of the NN. The NN is trained basically by providing the input data and reevaluating weight coefficients, which are (re)determined by the backpropagation algorithm initiated from the last hidden layer of the NN.

Despite a variety of different ML approaches available nowadays, they all share some common challenges. The key one is actually shared with the more conventional modeling approaches. It is associated with the notion of "The Black Box" and can be illustrated by the "Garbage In–Garbage Out" problem (Fig. 14). Machine learning requires consistent input data for developing an adequate predictive model.[350,351] Thus, the preparation of the datasets and the so-called feature engineering are the necessary steps in the construction of proper models for the ML studies. An intrinsic challenge for ML approaches is that the transparency in how the machine learns patterns or predicts properties depends on a chosen approach and cannot be fully achieved when, for example, neural networks are employed.[352]

6.1.2. Descriptors for machine learning in chemistry. The identification of a digital parameter – descriptor – that reflects the target measurable property of a chemical system is the corner-stone of all ML approaches. Descriptor is a very general term and it can take a form of a digital representation of a molecule, material or any chemical system, its properties, structure (geometric or electronic) as well as that of any parameter of the environment. The examples of molecular or material descriptors are conventional chemical reactivity parameters such as the HOMO–LUMO gap, electron affinity and the

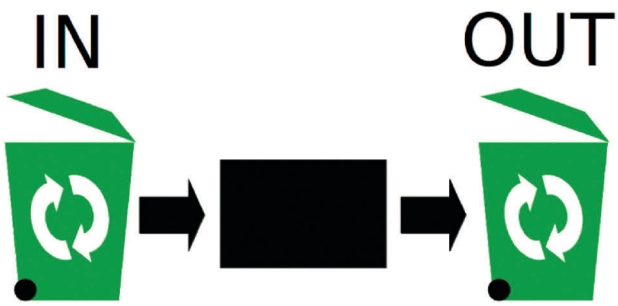

Fig. 14 "Garbage In–Garbage out" problem in ML studies. If input values are inadequate, the model produces inadequate results.

d-band center, or a mathematical structural representation such as a connectivity matrix encoding chemical bonding in a molecule.[353] Basically, any scalar or tensor that encodes in some manner the relevant property of a chemical compound, chemical system, or chemical environment may be used as a descriptor.

The mathematical representation of the geometric structure of an organic molecule ready to serve as an input for ML studies does not pose an issue nowadays. The presence of a carbon framework connected by strong covalent bonds in organic molecules makes it particularly attractive to employ the graph-based notations that can be backdated to Morgan's original idea of molecule representation in mathematical graphs.[354–356] To date, several linear representations have gained a particular importance for the description of organic compounds. These are the so-called SMILES and InChI as well as the connectivity-based encoding formats implemented in MDL and XML files.[355,357–359] These notations have become truly widespread, especially SMILES, with its various modifications in different areas of chemoinformatics.[360,361]

Although the utilization of these approaches can in principle be extended to inorganic and organometallic molecules, their direct representation faces a number of problems mostly related to the ambiguity of the bonding representation and consistent algorithms generally applicable to such chemical systems are substantially under-represented.[358,362] Nevertheless, there are several ML studies where the successful utilization of the graph-based notation such as SMILES for the representation of simple organometallic molecules has been demonstrated.[357,363–365] Catalytic systems based on inorganic and organometallic compounds are commonly characterized by the diversity of coordination polyhedra, stereoisomerism of the transition metal complexes, and the variability of the electronic nature of organometallic bonds. These are regarded as the key hurdles for the graph representation of the respective chemical systems and they have to be accounted for when applying the ML approaches to catalytic problems.

Previous sections clearly illustrate the key roles of electronic effects for the catalytic reactivity, rendering the associated parameters reflecting the electronic structure particularly important for the ML studies on catalytic systems. Besides the conventional chemical descriptors such as the atomic charges, Tolman's χ-factor, *etc.*, several more comprehensive mathematical representations have been introduced so far. The simplest example of the related descriptors is the so-called Coulomb matrix that encodes the information about the electrostatic forces in a molecule. The diagonal elements of such matrices are a polynomial fit to the energy of free atoms constituting a given molecule. The off-diagonal elements correspond to the Coulomb repulsion energy values for all pairs of nuclei. This type of electronic structure representation was used, for example, for the ML prediction of the atomization energies of organic molecules.[327,330] An alternative electronic structure descriptor was developed on the basis of the Fourier series of atomic radial distribution functions as an alternative to Coulomb matrices.[366]

The combination of steric and electronic descriptors is common in QSPR studies of organometallic compounds.[339,348,363,367–369] Sigman and co-workers demonstrated that a combination of steric descriptors such as STERIMOL and Tolman cone angles together with the metal NBO charges can be used within ML approaches for predicting yields in homogeneous catalytic reactions.[369,370] Density Functional Theory (DFT) methods may be effectively combined with the QSPR models to estimate the catalytic performance.[371] The application of ML approaches to catalytic systems is commonly coupled with DFT calculations, as the DFT-computed molecular properties provide additional useful descriptors that can directly be employed in the training datasets.[305,329,335]

New computational approaches for the correct representation of transition metal complexes have been introduced recently. A notable example is the molSimplify open-source code developed by the group of Kulik.[372] This computational toolkit has been designed to facilitate the generation of relevant structures and calculate properties of transition metal complexes. The program is based on the ''divide and conquer'' strategy, in which the organic ligand and the metal center are described separately. The implementation of an artificial NN in molSimplify allowed for the prediction of geometrical structures without the need for expensive geometry optimizations with conventional methods (such as DFT), while the combination of steric and electronic descriptors implemented in the code allowed for the prediction of electronic structure-related properties.[373] This computational tool may be efficiently employed to aid in the design of new inorganic materials and transition metal-based catalysts.[362,373]

Chemical reactions and, even more so, the catalytic reaction mechanisms are intrinsically much more complex in representations compared to individual compounds when their representation as computer-processed data is considered. Most studies on chemical transformations employ combinations of descriptors corresponding to reactants, reaction conditions, catalysts, and efficiency metrics as conversion and yield (Fig. 15).[374–379] There are special notations and file formats designed for the representation of organic chemical reactions such as the so-called SMARTS, which is the straightforward extension of the SMILES approach.[380–382] Another useful method for the representation of chemical reactions is by encoding in an extended chemical data format such as an MDL RXN file.[383] Besides this, it is worth mentioning the methods based on matrix transformations, such as the Dugundji-Ugi model, which is a formalism for representing chemical reactions based on BE- (Bond and Electron matrix) and R-matrices. The diagonal elements of the BE-matrix represent free valence electrons and off-diagonal ones correspond to bond orders between atoms. The R-matrix represents electron redistributions in the reactions; particularly, positive element values indicate bond formation and negative values indicate bond cleavage.[383,384] Nevertheless, all of these digital formats are limited to organic reactions[385] and they need to be substantially adjusted for catalytic applications commonly involving organometallic and inorganic components.

The comprehensive sets of descriptors that provide a sufficient representation of the specific chemical system are aligned with the measurable target characteristics, forming the datasets

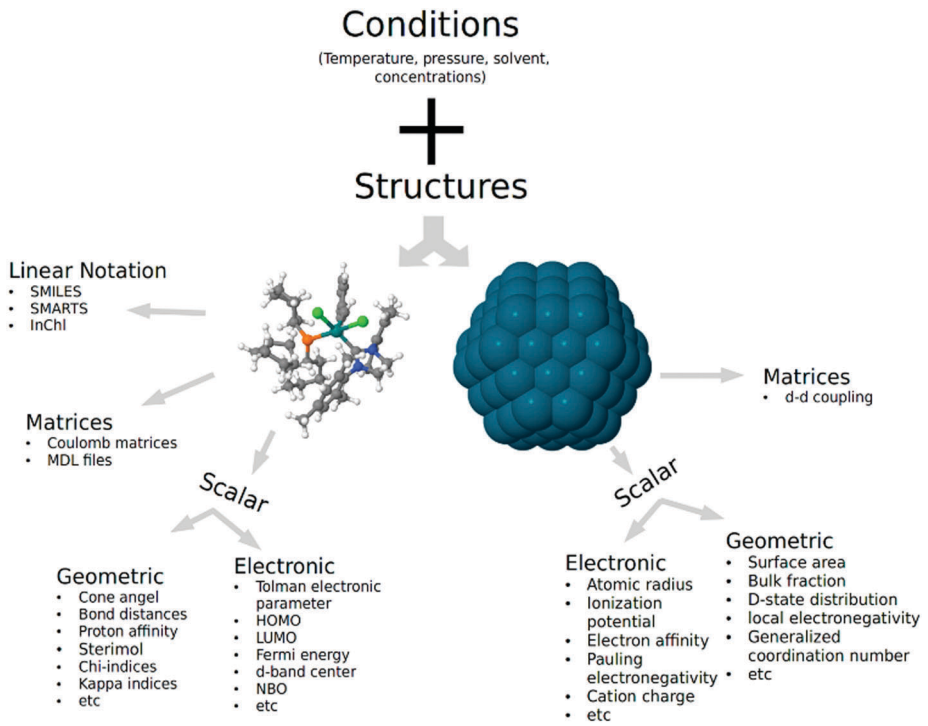

Fig. 15 Chemical reactions can be represented by combinations of electronic and geometric structure descriptors with descriptors that encode the reaction conditions. One commonly distinguishes the descriptors designed for organometallic or organic molecular species and for bulk or nanoparticulate materials.

used in ML approaches. In practical applications, some crucial parameters of the chemical reactions may be omitted from the training datasets either because of the limited nature of the available information or for the sake of simplicity. It is important to realize that despite the growing volume of the available digitalized data on various catalytic systems reported in the scholarly literature, the construction of reliable datasets from the reported experimental data and proprietary databases is a general challenge.[385] The mechanistic aspects of catalytic reactions, such as information about the transition states, intermediates and the respective energetics of elementary reaction steps, which were the focal point in the previous sections of this review, are rarely considered in ML. There are only a few studies accounting for transition states upon constructing the ML approaches for the prediction of chemical reactions.[322,383,385] In experimental catalysis, catalytic tests commonly employ catalyst precursors, whereas the nature of the actual catalytic species is often unknown. Furthermore, the nature of the catalytic species may strongly depend on the activation procedure employed and/or evolve in the course of the catalytic reaction. Similar to computational modeling of catalytic reactions, the construction of adequate ML models and the selection of representative sets of descriptors still require a substantial human mechanistic insight into the fine details of chemical reactions.[370] This results in the apparently conflicting requirements of availability of extended versatile training datasets, and detailed understanding of the crucial mechanistic parameters and their influence on the reactivity of different catalyst classes. This conflict is one of the most important conceptual challenges in the field.

6.2. Machine learning in catalysis

Many attempts to integrate ML methods into heterogeneous catalysis have been made in the last 3 decades. The first NN-based catalytic studies date back to the mid-1990s.[374,380] Earlier applications of related methodologies could be found among the heterogeneous catalysis literature from the late 80s, which coincides with the peak of popularity of so-called expert systems.[386] The essential properties of heterogeneous catalysts are surface area, elemental composition, and surface morphology – making these easily accessible characteristics attractive candidates for the descriptors. Most studies reported so far construct the training datasets by parsing the scholarly literature or by using in-house laboratory data.[374,376–379,387] In general, the major part of ML studies in heterogeneous catalysis target two main challenges: (i) the direct prediction of catalytic activity (at the molecular level) or (ii) modeling of the chemical reaction efficiency, that is, to estimate indirectly the activity by building a model that relates the set of descriptors to the reaction yield or the reaction rate. The assessment of the intrinsic catalytic activity at the molecular or nanoscale is normally carried out by computing the adsorption energies or investigating the elementary reaction steps on surfaces. Here, the ML method is commonly used in concert with DFT modeling used to provide molecular-level insight and the necessary descriptors for the datasets. In this case, the ML provides the necessary

predictive mechanism that effectively enables the transition from the microscopic DFT modeling to DFT-guided catalyst design. The indirect modeling of the catalytic efficiency mostly involves training of the ML models on the experimental datasets that combine descriptors related to the reaction conditions, nature, composition and physico-chemical characteristics of the catalysts, and those of the reagents. These two conceptually different approaches will be considered in more detail below.

6.2.1. Modeling of catalytic reaction efficiency. The ''popularity'' of particular ML approaches in catalytic studies has evolved over time. Aligned with the general development of the artificial intelligence approaches, the first catalytic applications of ML employed expert systems, which attempted to emulate the decision making processes by ''reasoning'' through the available set of knowledge, following standard rule-based systems. For example, the so-called INCAP expert system was employed to design a promoted SnO_2-based catalyst for the oxidative dehydrogenation of ethylbenzene.[386] A later decline of expert systems, followed by the reintroduction of NN, shifted the paradigm. Notably, the same catalytic process was studied with the NN-based methods several years later and the great potential of this methodology for predicting reaction selectivities for related catalyst compositions has been demonstrated.[379] The particular effectiveness of NNs has been shown for the applications in combinatorial catalysis. A representative example has been reported by Corma and co-workers, who employed an NN methodology to optimize a transition metal catalyst for the oxidative dehydrogenation of ethane.[388] In practical applications, the representative training datasets used contain sets of descriptors related to the catalyst structure and reaction conditions (temperature, reaction time, and concentrations of reagents and products). In principle, the NN approach can be successfully employed for the identification of the optimal reaction conditions for a given catalytic system. A representative example is the earlier work by Sasaki *et al.* on NO decomposition over Cu-containing ZSM-5 zeolites.[374] A conceptually similar approach has been employed for the NN-driven optimization of the alkene epoxidation by polymer-supported Mo(VI) complexes.[375] In this work, the particle size and pore diameters were included as the key catalyst structure descriptors.

An NN-based or any other ML approach allows for the construction of predictive models that are suitable for optimizing catalyst formulation and conditions (the parameters included in the descriptor set) for highly complex processes without direct insight into mechanisms or knowledge of the specific atomistic details of the catalyst or the catalytic process under certain conditions. These conditions, in accordance with the basic principles of chemical engineering, are that a chosen combination of the descriptors that account for the catalyst structure, reaction conditions, and reaction efficiency (for example the reaction yield) adequately captures the key factors crucial for the particular catalytic process. For example, an NN based model has been successfully employed for analyzing the photocatalytic activity of TiO_2 in the oxidative degradation of 17 α-ethynylestradiol.[387] The NN model in this case accounted for the reaction conditions (catalyst and substrate concentrations) as well as the environment conditions such as water matrix conductivity and impurity (*e.g.* organic carbon) concentrations. This study has produced quite an intriguing and counter-intuitive insight that the impurity concentration had a comparable significance to the substrate concentration in the final ML model. The water matrix conductivity was found to be even more significant.

During the 2000s the focus of the applied catalysis community was significantly shifted towards the SVM approach, which has been widely employed to achieve accurate predictions of catalytic reaction efficiency for many systems. For example, the SVM and NN approaches in combination with genetic algorithms were used to predict the yield and selectivity of benzene isopropylation over H-beta zeolite catalysts. The results corresponded well with experimental data.[389] A different SVM setup was employed for studying olefin epoxidation over a titanium silicate mesoporous catalyst.[345] Markedly, the related ML models constructed based on the SVM approach to predict the outcome of the hydrothermal synthesis of hybrid organic–inorganic materials (such as MOFs) have been recently shown to yield results substantially outperforming conventional human experience-based strategies.[325]

Modeling of the efficiency of catalytic heterogeneous reactions based on experimental datasets is a developed field. In some cases such studies may have straightforward practical implications as, for example, the ML study by Akcayol and Cinar on the efficiency of a heated catalytic converter.[390] The phenomenological approach that does not require the detailed mechanistic and structural information regarding the nature of the catalytic centers substantially facilitates the construction of very large training datasets containing macroscopic descriptors commonly employed in chemical engineering. The ML catalysis models constructed in this manner capture only the essential and (mostly) macroscopic physics and allow for a fast answer to a specific question. A thorough account of the mechanistic details of catalytic reactions for extended datasets of sizes large enough for ML studies was barely possible until very recently. This is because the respective information and the associated sets of descriptors could not be obtained with experimental techniques, given the requirement for the extremely resource- and time-consuming experiments, and the lack of broadly available open databases. Neither could this information be gained *via* computational modeling, as quantum chemical analysis of catalytic processes for extended catalyst libraries was well beyond the computing power of CPUs. However, such an approach fundamentally limits the predictive power of the ML outside the pre-defined classes of the catalytic systems. One naturally misses many possible catalyst design insights when using such ML models in line with the missing detailed mechanistic and structural information in the training datasets.

The solution to this natural drawback of the indirect ML approaches has emerged *via* their integration with the atomistic DFT modeling. Indeed the widespread of fast and semi-quantitatively

accurate DFT methods together with the enormous progress in computational hardware witnessed in the last decade created a basis for this qualitative shift in catalytic applications of ML. This powerful combination, discussed below in more detail, holds promise as a practical approach for theory-guided ad-hoc catalyst design.

6.2.2. Boosting DFT modeling of heterogeneous catalysts with machine learning approaches. The direct DFT modeling of complete reaction networks for different reactions over varied catalysts in a single study remains well beyond the current capacity of computational and human resources.[305,336,391] Indeed, even geometry optimization of realistic catalyst models consisting of hundreds of atoms (slabs, nanoparticles, or supported metal clusters) may take considerable time to compute. Moreover, the mechanistic analysis of the competing reaction channels still requires the manual construction of model systems as well as the starting configurations of the reaction intermediates, and an initial guess for the transition states and/or the actual reaction steps. This is tedious work, and also the efficiency and even outcome often depend on the experience and skills of the researcher. Conventionally, DFT studies in catalysis consider rather limited sets of model systems and states and rather focus on formulating a conceptual understanding of chemical reactivity than generating data for the subsequent processing. Another limiting factor for the integration of quantum chemical modeling with ML was related to the limited accuracy of practical computational methods, which are usually hard to estimate *a priori*.[392–394]

There is no doubt that the big data analytics has the potential to provide technologies to greatly boost catalyst design. There is a clear demand for developing strategies alternative to the conventional quantum chemical modeling for providing the mechanistic details in an inexpensive and human bias-free manner to be employed in ML-based catalyst design approaches. The workaround to direct quantum chemical modeling in catalysis may be to employ ML methodologies also for generating the necessary mechanistic and microscopic information by training an ML model on the DFT-computed results. For example, the DFT calculations of binding energies of N, O, and NO species to various sites in Au–Rh-nanoparticles, clusters, and surfaces allowed for the construction of a linear regression model of the nanoparticle activity in the NO decomposition reaction with the so-called local structural similarity kernel as a key descriptor. In other words, the model in this study was trained on the descriptors constructed under the assumption of similar activity of sites with a similar local structure. The resulting ML model predicted activity of the bimetallic nanoparticles with variable size and composition and allowed for kinetic modeling of the direct NO decomposition process.[329,395]

Nørskov and co-workers developed an ML-based surrogate model that allows for a reduction in the number of necessary DFT calculations by an order of magnitude, while at the same time providing a means to model complex networks of surface reactions with sufficient accuracy. It has been applied, for example, to processes such as CO_2 electroreduction on Ni–Ga-bimetallic nanoparticles and syngas conversion over a Rh(111) surface.[305,335] Modeling of catalytic processes with bimetallic nanoparticles and exhaustive sampling of the attainable

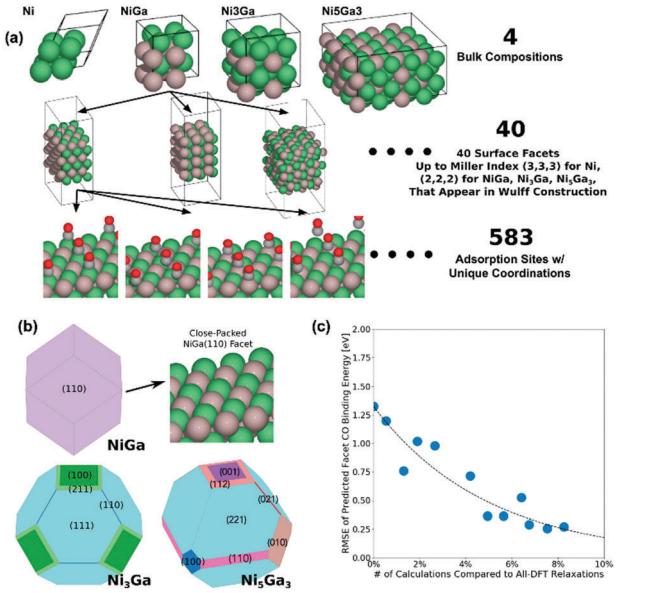

Fig. 16 The application of ML approaches to predictive modeling of CO adsorption on bimetallic Ni–Ga nanoparticles. The associated model complexity stems from the large number of possible configurations emerging from (a) the wide variety of adsorption configurations and the configurations of the surface models along with (b) the variety of surface terminations available for the adsorption. By treating such a complex system using ML approaches, a substantial decrease (c) in the required CPU time of the adsorption energy predictions could be achieved compared to the conventional all-DFT geometry relaxation methods. Reprinted with permission from ref. 335. Copyright 2017 American Chemical Society.

configurations may become extremely demanding in computational resources even if adsorption of a single simple intermediate such as the CO molecule is considered (Fig. 16a and b). The calculation of the adsorption energies of gaseous intermediates can be facilitated through the utilization of ML methods, which can reduce the number of necessary DFT computations to sample all relevant reaction pathways (Fig. 16c).

No significant accuracy deterioration should be expected when constructing ML models based on the DFT-computed energetics of heterogeneously catalyzed reactions. The Brønsted–Evans–Polanyi principle provides a practical means for estimating catalyst activity from the computed adsorption energies of key reaction intermediates. The study on the electrochemical reduction of CO_2 over metal alloys employed an NN-based model that was trained on a dataset obtained from periodic DFT calculations. Such a model predicted the adsorption energies for reaction intermediates with an error of *ca.* 0.1 eV. Notably, by combining simple geometric and electronic structure descriptors such as local electronegativity, the effective coordination number of an adsorption site, ionic potential, electron affinity, and the Pauling electronegativity, a sufficiently high accuracy of the predicted energetics could be obtained.[396,397]

The use of a combined ML-DFT approach in computational catalysis achieves the accuracy of conventional DFT methods with a substantially lower computational demand, if a trained ML model is available. Recent studies demonstrate that well-trained ML models outperform hybrid DFT methods in the prediction of properties of organic molecules such as

enthalpies and free energies of atomization, HOMO–LUMO energies, dipole moments, polarizabilities, zero point vibrational energies, heat capacities, *etc.*[323] This suggests that it may become soon possible to develop an ML-only based computational procedure providing access to molecular-level information about chemical transformations that is cheaper and at the same time more accurate than the conventional quantum chemical methods. The key prerequisite for this is the availability of reliable experimental datasets to ensure the exhaustive training of such an ML model.

Datasets in heterogeneous catalysis are conventionally built from continuous operation that allows varying a limited number of parameters and obtain coherent data making it relatively straightforward to obtain large datasets from kinetic experiments. One can therefore anticipate the upcoming breakthroughs in heterogeneous catalyst design driven by the recent methodological progress in ML. Homogeneous catalysis studies conventionally deal with small scale batch experiments, where an individual reaction entry is a separate experiment that is not directly related to the other entries in a target dataset making it particularly challenging to generate larger datasets from the kinetic studies. The implementation of flow chemistry approaches to homogeneous catalysis studies may be viewed as one of the crucial ingredients towards the successful implementation of the big data strategy in this field. More important in our opinion is the development of broadly available open-access databases containing well-structured machine-readable catalytic activity datasets (that is every entry contains data necessary for training of ML models).

6.3. Open datasets as the basis for the catalyst design with machine learning techniques

Despite the many successful examples of the use of ML for addressing different scientific and technological problems reported in the past few years, there are several general problems that substantially limit the power and general applicability of ML-based approaches. The most important and the most generic problem that is particularly relevant for catalysis is the absence of comprehensive, large and publically-available datasets for training of efficient ML models. Despite the emergence of Big Data and the availability of large databases containing millions of chemical reactions, the majority of chemistry domains of catalysis still lack open datasets of appropriate size.[319,398]

There are several problems related to the proprietary nature of databases, difficulties with obtaining data from scientific articles and often-encountered cherry-picking (mis)practices.[385,399] Addressing these problems is a general scientific challenge that spans well beyond the current subject of the development of ML approaches for catalysis design.[400] Fortunately, practical measures are becoming available with the implementation of state-of-the-art software and methods for database organization. For instance, data nowadays can be organized with NoSQL technologies,[401,402] which provide a straightforward method for the construction of databases that will include chemical reaction data on chemical environment parameters, reaction conditions, reactants, and products in a digital form. Combining the Open Access policy and user-friendly application programming interface to a database would enable sharing scientific data in a semi-automatic, fast and simple manner. Data in scholarly publications can be organized in a way facilitating the training of the ML models, by providing it in machine-ready table or CSV formats in the supplementary information. Such datasets should contain structural information on the employed catalysts and chemical compounds uniformly described through encoding in MDL files or using linear notations like SMILES and InChi.[403] Furthermore, organizing the data according to the R Markdown format would make it ready for an algorithmic data analysis.[397,404–406] Such a format facilitates reproduction of the results and verification of the correctness of statistical analysis.[407] Fig. 17 illustrates how the advanced information technologies such as Git workflow and R Markdown boosted the Ocean Health Index monitoring.[397] We believe that algorithmic analysis-ready data science approaches could become transformative factors in catalysis design.

The scientific community steadily progresses towards the widespread implementation of the open access publishing policies driven by the clear socio-economic benefits and strong ethical implications.[408,409] Publicly available data and instruments are transparent and can easily be controlled by the scientific community. The under-representation of negative scientific results in the scholarly literature is a well-recognized general problem[410,411] and it is well known in all fields of catalysis sciences. The lack of negative data entries is an important problem for the development of practical ML instruments as they are the necessary component of a balanced training dataset and therefore are necessary to construct predictive models.[325]

6.4. Towards rational catalyst design with machine learning techniques

The pursuit of the general theory of catalysis that will give researchers a deep understanding of catalytic processes and a theoretical framework to anticipate new catalytic events has a nearly century-old history[412] that has emerged in recent years in the general concept of the rational catalyst design. The rational

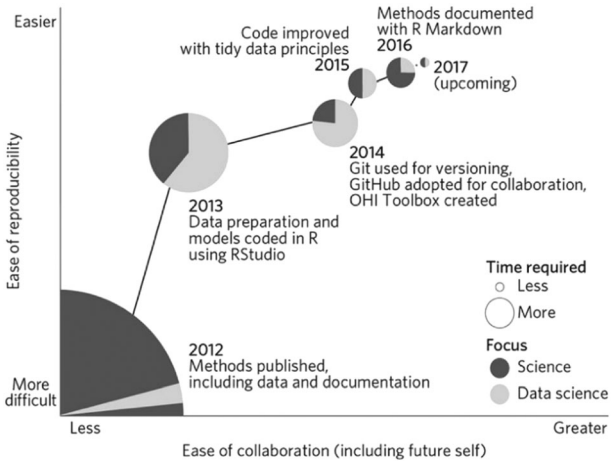

Fig. 17 The influence of the open source software on the quality of the research results and its accelerating effect on the overall research progress within the Ocean Health Index (OHI) research, reprinted with permission from Macmillan Publishers Limited, 2017. Adapted from ref. 397.

design is a strategy for creating new structures with specific functions and properties based on a deep understanding of the fundamental factors that define the properties of interest. It is generally believed that the realization of this strategy in catalysis requires the understanding of how the catalyst structure and key chemical phenomena in catalytic processes are related to the observed activity.

While the progress in computational chemistry methodologies together with the availability of more and more powerful computational resources gradually enables modeling of any imaginable catalytic process, one still needs to construct a model system as well as to consider the relevant underlying chemical phenomena in the model. Therefore, any conventional computational modeling method by itself gives no deep understanding of the relation between the structure and underlying chemistry to the activity, and ultimately it is up to the skills of the researcher to select which predictions to make *via* the modeling. Machine learning-based modeling of catalytic reactions, in contrast, offers a way to enable a truly predictive modeling by virtue of its formalism.

As has been discussed above, a great number of descriptors are already available to model chemical processes involving transition metal catalysts. Theoretically, there is no fundamental limitation to compute these descriptors in an automatic way using already available cheminformatics software. By selecting a proper set of descriptors (either based on heuristic guess or *via* a trial-and-error approach), this would allow one to determine the key geometric and electronic properties of reactants and catalysts as well as the important environment parameters together with the optimal reaction conditions for a chemical process in question. The activity parameters such as TOF values, reaction yields, activation barriers for elementary steps could then be correlated with these descriptor-encoded key properties through the use of machine learning techniques. The accuracy of a trained ML model depends on the reliability and completeness of the data in the training dataset while the model generality depends on the size of the dataset.

The realization of the rational catalyst design strategy requires the descriptor representation that is detailed enough to be coupled with a training dataset that is sufficiently large and well-balanced. Indeed, the ideal case for rational design of, for example, electrocatalytic water splitting, the Fischer–Tropsch process, or CO_2 hydrogenation catalysts would need to include the descriptor representation of all catalysts known to date and the corresponding activity data in the dataset. Moreover, the data on inactive catalysts have to be included explicitly to make the dataset balanced and the model predictive. This, in turn, would allow estimating the activity of a catalyst that is not yet synthesized or to determine descriptor values that maximize the activity. The latter would enable the targeted synthesis of catalysts having the desired properties. Proper accounting for the catalyst poisoning and degradation while training ML models will allow *a priori* tuning of the optimal reaction conditions.

Therefore, the availability of consistent open-access databases on catalyst activity is of paramount importance for the rational catalyst design with ML approaches. The promotion of open-access policies in scientific data publishing and active use of new tools for digital data representation enabling automated computer data analysis are thus the crucial ingredients towards the realization of such an ML-based rational design strategy. The proliferation of data science technologies to catalysis research is therefore much anticipated. Because of the recent successes in prediction of organic reactions,[326,385,413,414] properties of materials,[321,331,336] and the well-known utility of the QSAR/QSPR approach in drug design, there is a strong indication that the ML-based rational catalyst design is about to emerge.

7. Conclusions

Rational catalyst design – the concept by which a successful catalytic system could be forecasted based on the results of only computations – has long been and still remains the ''Holy Grail'' of heterogeneous catalysis. An opportunity to avoid or even to just minimize tedious and costly experimental search for the optimal composition of multicomponent catalysts and reaction conditions for a given chemical transformation by replacing it with some computer-based algorithm capable of directing this search is very attractive and if practically realized holds a promise of revolutionising chemical science and technology. Despite great progress witnessed during the last two decades in the development of new approaches for theory-guided catalyst development, substantial new methodological advances are still necessary to enable their widespread implementation in the daily lives of catalysis researchers.

One of the most important shortcomings of the established computational strategies is the dominance of the basic 0 K/UHV approximation commonly employed for the development of mechanistic concepts in catalysis. Despite being capable of providing a satisfactory mechanistic description of a catalytic phenomenon, they often lack sufficient predictive power mostly due to the inability to adequately account for the crucial physical effects encountered under the conditions of actual catalytic processes. The understanding of such phenomena and their impact on the molecular-level processes underlying the performance of catalytic systems is one of the key challenges to realization of the catalysis by design approach.

In this review, we have discussed some important recent methodological developments enabling the transition from the 0 K/UHV to *operando* computational modelling. The importance of this transition was highlighted by discussing how the molecular level picture of the catalytic sites and the associated reaction mechanisms evolve drastically when a correct account for chemical environment, pressure and temperature effects is given in the molecular simulations.

An important challenge in modern heterogeneous catalysis is to reveal the nature of the active sites and to understand how their structure evolves when exposed to the realistic catalytic environments and temperatures. In the first sections of this review, we discussed computational approaches allowing comprehensive sampling of the complex chemical space to

Table 1 The computational approaches discussed in the review, with appropriate types of problems, examples of relevant system classes and general remarks on key issues and limitations

	Application guidelines	Example of catalytic system classes	Remarks
Global optimization	Unknown catalyst structure (known composition, low T, low P).	Subnanometre metallic clusters on inert substrates/metal oxide surfaces for low T oxidation catalysis.	Number of structures scale exponentially with system size. MC methods preferred for maintaining local structural information during search (pathways).
	Unknown catalyst structure (unknown composition, any T or P).	Dynamically restructuring reducible oxide surface catalysts in an O_2 atmosphere.	Couple with AITD for unknown system stoichiometry.
Ab initio thermodynamics	Unknown structure of catalyst surface.	Phase diagram of metal oxide catalyst surfaces.	Only modest computational requirements when vibrational contributions to surface free energies are neglected.
	Morphology/growth direction (using Wulff construction).	Nanoparticle shape (nanoalloys/metal carbides).	
Biased molecular dynamics	Reactions on liquid/solid interface.	Electrochemical reduction on metal surfaces.	Typically on the order of 10^5 MD steps (force evaluations) per elementary reaction.
	High reactant concentration.	Heavy oil hydrogenation on metal carbide nanoparticles.	Choice of proper collective variable is an issue.
	High reaction temperature.	NOx reduction over metal-exchanged zeolites.	Two MD approaches to choose from: Born–Oppenheimer or Car–Parrinello.
	Competing species in reactive mixture.	Methanol-to-olefin process in nanoporous solid acids.	Kinetic information is accessible.
	Loosely-bound complexes of reactants/TS/products with catalyst.	Hydrocarbon cracking in acid zeolites.	Only fixed composition runs in the *NVT* ensemble done so far.
Microkinetic modelling	Complex reaction networks.	Partial (de)hydrogenation reactions of unsaturated hydrocarbons on metal surfaces.	Lateral interactions are rarely accounted for due to the mean-field approximation.
	Translation of molecular-level mechanistic data into directly measurable macroscopic kinetic parameters.	Alkylation over FAU zeolites/NH_3 synthesis over metal catalysts.	Algorithms to include secondary processes (*e.g.* catalyst surface reconstruction, long-term deactivation, *etc.*) are not available.
Machine learning	Availability of experimental dataset on catalytic activity. Easy construction of the training dataset *via* DFT computations and the system too complex for DFT-only modeling.	NO decomposition over Cu-containing zeolites. Binary metal alloy catalysts (*e.g.*, Au–Rh or Ni–Ga).	Straightforward account for reaction conditions is possible. Adequacy of the training data determined by the accuracy of the DFT computations (garbage in-garbage out principle). Lack of open-access comprehensive databases on catalytic activity. Lack of data related to negative catalytic results.

determine the potential active site candidates and construct representative active site models as well as to assess their thermodynamic stabilities as a function of the reaction conditions. We discussed how global optimization (GO) techniques can be used to address the basic structural problem in heterogeneous catalysis. These methods can be efficiently used to screen candidate structures and automatically search for stable active site formulations. This was followed by the discussion of the constrained *ab initio* thermodynamic analysis (AITD) approaches for assessing the thermodynamic stabilities of different active site ensembles under the varying reaction conditions. Indeed, the integration of GO and AITD methods has proven useful in isolating relevant structures under realistic conditions for a number of systems, from gas phase particles to oxide surfaces, and the combination of the two techniques is becoming commonplace. Obtaining relevant structures of the active site is, however, just a prerequisite for a reliable description of the catalytic process, with reactant concentration, temperature, pressure or presence of solvent to be accounted for. This situation corresponds to studying catalytic processes on the free energy landscape, which is a generic problem of computational chemistry and was tackled in the next section by discussing the use of Hessian-based as well as advanced *ab initio* molecular dynamics (AIMD) approaches. These techniques are already quite routinely applied for studying mechanisms of catalytic reactions, either by open-ended searches of the free energy surface, for example with metadynamics, or by integration with GO methods through free energy path search techniques. It can be seen that the state of the art in both GO and AIMD are converging towards *operando* descriptions of complex reactions on multicomponent systems. In GO, the reactive global optimization (RGO) method allows for a kinetics-based description of the reaction network for supported catalysts along with adsorbates, under pre-defined conditions. In AIMD, complex, multidimensional collective variables allow for a broad sweep of the reaction network for catalysts of similar complexity to that of RGO, and the discovery of new mechanisms. In both cases, the limitations are the computational expense of the calculation methods. One possible route to alleviate this problem is, in our opinion, the development of machine learning-based potentials, which are more robust, transferable and can adequately handle chemical transformations.[391,415,416]

These endeavours inevitably lead to the generation of large volumes of mechanistic data and insights, which can become so complex and heavy that it is no longer possible to rely solely on the human ability to analyse and rationalise them. New approaches that would limit the human bias in analysis and at the same time provide with the means to extract the experimentally verifiable parameters from the microscopic data are becoming crucial. In this context, the

conventional chemical engineering reductionist approach in the form of microkinetic modelling or kinetic Monte Carlo becomes instrumental to reduce the mechanistic complexity to a tangible number of experimentally verifiable parameters suitable for guiding the experimental catalyst development and process optimization efforts. Kinetic modelling is naturally well integrated with any method which provides energetic data about minima and transitions states, such as GO, AITD or AIMD. One current challenge for the increased adoption of engineering approaches by computational chemists is to increase the sophistication of the models. Moving beyond mean-field descriptions to models with proper adsorbate interactions, coverage dependences and substrates which change under the conditions of the reaction is important to match the complexity of *operando* catalysis. An alternative approach that has gained importance and attention recently in basically all areas of human activities including chemistry and catalysis relies on a machine to not only generate the numerical data but also analyse and guide the research and development efforts. Ideally, machine learning would allow removing completely the human bias from the model formulation, data analysis and expanding the scale at which the analysis is carried out to drive innovation in catalysis research. However, these idealistic views are still very far from coming true. The success of machine learning approaches in catalysis sciences still heavily relies on the definition of suitable descriptors and on the quality of the available datasets. Human interference is still often a necessity for pre-processing the data to assess its quality and to adjust it for subsequent construction of the ML algorithms. Having said this, we are confident that ML approaches for the data analysis will gain importance in computational catalysis research and will find many applications not only in identification of trends enabling the search for improved catalysts outside the conventional scopes, but also as the means to facilitate the very basics of the computational catalysis that is electronic structure calculations and statistical thermodynamic analysis.

The main features of individual methods discussed in this review are summarized in Table 1. It has been already mentioned that subsequent or even simultaneous application of all methods summarized in Fig. 1 would be computationally prohibitive. However, for many catalytic systems it is not critical to apply all extensions: for example, (i) catalysts with known structure do not require GO methods, (ii) catalysts with an inert structure and low concentration of active sites do not require AITD methods, (iii) catalysis at a gas phase interface with low reactant concentrations could be treated without biased MD methods. For individual extensions beyond the reference 0 K/UHV model, Table 1 shows under which conditions each method should be applied, gives a few examples and some comments with respect to their practical use. It is our hope that this review helps readers to better understand the principles and applicability of methods that extend beyond the 0 K/UHV model towards computational *operando*.

Abbreviations

UHV	Ultrahigh vacuum
PES	Potential energy surface
MD	Molecular dynamics
GO	Global optimization
BH	Basin hopping
MH	Minima hopping
EA	Evolutionary algorithm
NEB	Nudged elastic band
DFT	Density functional theory
GM	Global minimum
MP2	Second order Moller–Plesset perturbation theory
HAGA	Hybrid *ab initio* genetic algorithm
STM	Scanning tunneling microscopy
KMC	Kinetic Monte Carlo
RGO	Reactive global optimization
AITD	*Ab initio* thermodynamics
LDA	Local density approximation
PBE	Perdew–Burke–Ernzerhof (exchange correlation functional)
SCR	Selective catalytic reduction
AIMD	*Ab initio* molecular dynamics
MTD	Metadynamics
ITS	Integrated tempering sampling
CPMD	Car–Parinello molecular dynamics
TI	Thermodynamic Integration
QCT	Quasiclassical trajectory
TPS	Transition path sampling
ZPVE	Zero-point vibrational energy
MTO	Metal-to-olefin
HTST	Harmonic transition state theory
MKM	Microkinetic modelling
TST	Transition state theory
DRC	Degree of rate control
MLM	Machine learning
NN	Neural networks
SVM	Support vector machine
QSAR	Quantitative structure–activity relationships
QSPR	Quantitative structure–property relationships
CPU	Central processing unit
GPU	Graphics processing unit
PCA	Principal component analysis
SMILES	Simplified molecular-input line-entry system (molecular representation format)
InChI	International Chemical Identifier (molecular representation format)
MDL	MDL Information Systems, Inc.
XML	Extensible markup language

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

E. A. P. acknowledges the support from the European Research Council (ERC) under the European Union's Horizon 2020

research and innovation programme (grant agreement no. 725686). A. B. thanks the Government of the Russian Federation (Grant 08-08) and M. P. thanks the Ministry of Education and Science of Russian Federation (Project 11.1706.2017/4.6) for financial support. J. M. acknowledges financial support through the Royal Thai Government Scholarship. Work at Charles University was supported by Charles University Centre of Advanced Materials (CUCAM) (OP VVV Excellent Research Teams, project number CZ.02.1.01/0.0/0.0/15_003/0000417) and by the Czech Science Foundation (project No. P106/12/G015). L. G. additionally acknowledges support from the Czech Science Foundation Grant No. 17-01440S.

Notes and references

1 H. Topsoe, *J. Catal.*, 2003, **216**, 155–164.
2 M. A. Banares, *Catal. Today*, 2005, **100**, 71–77.
3 A. Chakrabarti, M. E. Ford, D. Gregory, R. R. Hu, C. J. Keturakis, S. Lwin, Y. D. Tang, Z. Yang, M. H. Zhu, M. A. Banares and I. E. Wachs, *Catal. Today*, 2017, **283**, 27–53.
4 C. W. Jones, F. Tao and M. V. Garland, *ACS Catal.*, 2012, **2**, 2444–2445.
5 B. M. Weckhuysen, *Natl. Sci. Rev.*, 2015, **2**, 147–149.
6 K. F. Kalz, R. Kraehnert, M. Dvoyashkin, R. Dittmeyer, R. Gläser, U. Krewer, K. Reuter and J.-D. Grunwaldt, *ChemCatChem*, 2017, **9**, 17–29.
7 K. Reuter, C. P. Plaisance, H. Oberhofer and M. Andersen, *J. Chem. Phys.*, 2017, **146**, 040901.
8 M. A. van Spronsen, J. W. M. Frenken and I. M. N. Groot, *Chem. Soc. Rev.*, 2017, **46**, 4347–4374.
9 H. B. Schlegel, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2011, **1**, 790–809.
10 R. Jin, C. Zeng, M. Zhou and Y. Chen, *Chem. Rev.*, 2016, **116**, 10346–10413.
11 M. S. Jørgensen, U. F. Larsen, K. W. Jacobsen and B. Hammer, *J. Phys. Chem. A*, 2018, **122**, 1504–1509.
12 S. Heiles and R. L. Johnston, *Int. J. Quantum Chem.*, 2013, **113**, 2091–2109.
13 D. J. Wales, *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*, Cambridge University Press, Cambridge, UK, 2003.
14 D. J. Wales, *J. Phys. Chem. A*, 1997, **101**, 5111–5116.
15 S. Goedecker, *J. Chem. Phys.*, 2004, **120**, 9911–9917.
16 K. Bao, S. Goedecker, K. Koga, F. Lançon and A. Neelov, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2009, **79**, 041405R.
17 H. A. Eivari, S. A. Ghasemi, H. Tahmasbi, S. Rostami, S. Faraji, R. Rasoulkhani, S. Goedecker and M. Amsler, *Chem. Mater.*, 2017, **29**, 8594–8603.
18 J. A. Gauthier, C. F. Dickens, L. D. Chen, A. D. Doyle and J. K. Nørskov, *J. Phys. Chem. C*, 2017, **121**, 11455–11463.
19 M. Sicher, S. Mohr and S. Goedecker, *J. Chem. Phys.*, 2011, **134**, 044106.
20 B. Schaefer, S. Mohr, M. Amsler and S. Goedecker, *J. Chem. Phys.*, 2014, **140**, 214102.
21 G. Rossi and R. Ferrando, *Chem. Phys. Lett.*, 2006, **423**, 17–22.
22 G. Barcaro, A. Fortunelli, G. Rossi, F. Nita and R. Ferrando, *J. Phys. Chem. B*, 2006, **110**, 23197–23203.
23 R. L. Johnston, *Dalton Trans.*, 2003, 4193–4207.
24 Y. Ge and J. D. Head, *J. Phys. Chem. B*, 2004, **108**, 6025–6034.
25 L. Gell, A. Kulesza, J. Petersen, M. I. S. Röhr, R. Mitrić and V. Bonačić-Koutecký, *J. Phys. Chem. C*, 2013, **117**, 14824–14831.
26 L. B. Vilhelmsen and B. Hammer, *J. Chem. Phys.*, 2014, **141**, 044711.
27 M. S. Jørgensen, M. N. Groves and B. Hammer, *J. Chem. Theory Comput.*, 2017, **13**, 1486–1493.
28 R. Ferrando, A. Fortunelli and R. L. Johnston, *Phys. Chem. Chem. Phys.*, 2008, **10**, 640–649.
29 H. G. Kim, S. K. Choi and H. M. Lee, *J. Chem. Phys.*, 2008, **128**, 144702.
30 H. Zabodsky, S. Peleg and D. Avnir, *J. Am. Chem. Soc.*, 1993, **115**, 8278–8289.
31 M. T. Oakley, R. L. Johnston and D. J. Wales, *Phys. Chem. Chem. Phys.*, 2013, **15**, 3965–3976.
32 S. E. Schönborn, S. Goedecker, S. Roy and A. R. Oganov, *J. Chem. Phys.*, 2009, **130**, 144108.
33 Z. Kang, C. H. A. Tsang, N. B. Wong, Z. Zhang and S. T. Lee, *J. Am. Chem. Soc.*, 2007, **129**, 12090–12091.
34 C. Gonzalez and H. B. Schlegel, *J. Chem. Phys.*, 1989, **90**, 2154–2161.
35 D. Sheppard, R. Terrell and G. Henkelman, *J. Chem. Phys.*, 2008, **128**, 134106.
36 E. Weinan, W. Ren and E. Vanden-Eijnden, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2002, **66**, 052301.
37 L. Maragliano, A. Fischer, E. Vanden-Eijnden and G. Ciccotti, *J. Chem. Phys.*, 2006, **125**, 024106.
38 D. Branduardi, F. L. Gervasio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 054103.
39 S. A. Trygubenko and D. J. Wales, *J. Chem. Phys.*, 2004, **120**, 2082–2094.
40 I. D. Seymour, S. Chakraborty, D. S. Middlemiss, D. J. Wales and C. P. Grey, *Chem. Mater.*, 2015, **27**, 5550–5561.
41 F. C. Chuang, C. V. Ciobanu, V. B. Shenoy, C. Z. Wang and K. M. Ho, *Surf. Sci.*, 2004, **573**, L375–L381.
42 X. J. Zhang, C. Shang and Z. P. Liu, *J. Chem. Phys.*, 2017, **147**, 152706.
43 C. Massen, T. V. Mortimer-Jones and R. L. Johnston, *J. Chem. Soc., Dalton Trans.*, 2002, 4375, DOI: 10.1039/b207847c.
44 A. Rapallo, G. Rossi, R. Ferrando, A. Fortunelli, B. C. Curley, L. D. Lloyd, G. M. Tarbuck and R. L. Johnston, *J. Chem. Phys.*, 2005, **122**, 194308.
45 J. M. Dieterich and B. Hartke, *J. Comput. Chem.*, 2011, **32**, 1377–1385.
46 G. Rossi, R. Ferrando, A. Rapallo, A. Fortunelli, B. C. Curley, L. D. Lloyd and R. L. Johnston, *J. Chem. Phys.*, 2005, **122**, 194309.
47 D. Bochicchio and R. Ferrando, *Nano Lett.*, 2010, **10**, 4211–4216.
48 R. Ferrando, A. Fortunelli and G. Rossi, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2005, **72**, 085449.

49 R. Ismail and R. L. Johnston, *Phys. Chem. Chem. Phys.*, 2010, **12**, 8607–8619.
50 V. S. Baturin, S. V. Lepeshkin, N. L. Matsko, A. R. Oganov and Y. A. Uspenskii, *EPL*, 2014, **106**, 37002.
51 P. Pyykko, *Angew. Chem., Int. Ed.*, 2004, **43**, 4412–4456.
52 E. Aprà, R. Ferrando and A. Fortunelli, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2006, **73**, 205414.
53 S. A. Serapian, M. J. Bearpark and F. Bresme, *Nanoscale*, 2013, **5**, 6445–6457.
54 Y. Gao, N. Shao, S. Bulusu and X. C. Zeng, *J. Phys. Chem. C*, 2008, **112**, 8234–8238.
55 A. Shayeghi, C. J. Heard, R. L. Johnston and R. Schafer, *J. Chem. Phys.*, 2014, **140**, 054312.
56 R. Fournier, *Can. J. Chem.*, 2010, **88**, 1071–1078.
57 S. Heiles, R. L. Johnston and R. Schafer, *J. Phys. Chem. A*, 2012, **116**, 7756–7764.
58 R. D. Adams, D. A. Blom, B. Captain, R. Raja, J. Meurig Thomas and E. Trufan, *Langmuir*, 2008, **24**, 9223–9226.
59 L. O. Paz-Borbon, A. Hellman, J. M. Thomas and H. Gronbeck, *Phys. Chem. Chem. Phys.*, 2013, **15**, 9694–9700.
60 S. Bhattacharya, S. V. Levchenko., L. M. Ghiringhelli and M. Scheffler, *Phys. Rev. Lett.*, 2013, **111**, 135501.
61 E. C. Beret, L. C. Ghiringhelli and M. Scheffler, *Faraday Discuss.*, 2011, **152**, 153–167.
62 H. Dhillon and R. Fournier, *Comput. Theor. Chem.*, 2013, **1021**, 26–34.
63 Y. Pei, Y. Gao, N. Shao and X. C. Zeng, *J. Am. Chem. Soc.*, 2009, **131**, 13619–13621.
64 Y. Pei, R. Pal, C. Liu, Y. Gao, Z. Zhang and X. C. Zeng, *J. Am. Chem. Soc.*, 2012, **134**, 3015–3024.
65 Y. Liu, Z. Tian and L. Cheng, *RSC Adv.*, 2016, **6**, 4705–4712.
66 H. Xiang, S.-H. Wei and X. Gong, *J. Am. Chem. Soc.*, 2010, **132**, 7355–7360.
67 F. Bertorelle, R. Hamouda, D. Rayane, M. Broyer, R. Antoine, P. Dugourd, L. Gell, A. Kulesza, R. Mitrić and V. Bonačić-Koutecký, *Nanoscale*, 2013, **5**, 5637.
68 B. Bellina, R. Antoine, M. Broyer, L. Gell, Ž. Sanader, R. Mitrić, V. Bonačić-Koutecký and P. Dugourd, *Dalton Trans.*, 2013, **42**, 8328.
69 Y. Ge and J. D. Head, *J. Phys. Chem. B*, 2002, **106**, 6997–7004.
70 Y. Ge and J. D. Head, *Int. J. Quantum Chem.*, 2003, **95**, 617–626.
71 Y. Ge and J. D. Head, *Chem. Phys. Lett.*, 2004, **398**, 107–112.
72 P. Biswas, R. Atta-Fynn and S. R. Elliott, *Phys. Rev. B*, 2016, **93**, 1–14.
73 P. Biswas, D. Paudel, R. Atta-Fynn, D. A. Drabold and S. R. Elliott, *Phys. Rev. Appl.*, 2017, **7**, 024013.
74 N. L. Rosi and C. A. Mirkin, *Chem. Rev.*, 2005, **105**, 1547–1562.
75 M. Zhu, E. Lanni, N. Garg, M. E. Bier and R. Jin, *J. Am. Chem. Soc.*, 2008, **130**, 1138–1139.
76 Y. Negishi, K. Nobusada and T. Tsukuda, *J. Am. Chem. Soc.*, 2005, **127**, 5261–5270.
77 P. D. Jadzinsky, G. Calero, C. J. Ackerson, D. A. Bushnell and R. D. Kornberg, *Science*, 2007, **318**, 430–433.
78 M. Zhu, C. M. Aikens, F. J. Hollander and G. C. Schatz, *J. Am. Chem. Soc.*, 2008, **130**, 5883–5885.
79 H. Häkkinen, M. Walter and H. Grönbeck, *J. Phys. Chem. B*, 2006, **110**, 9927–9931.
80 Z. Kang, C. H. A. Tsang, Z. Zhang, M. Zhang, N. B. Wong, J. A. Zapien, Y. Shan and S. T. Lee, *J. Am. Chem. Soc.*, 2007, **129**, 5326–5327.
81 Z. H. Kang, Y. Liu and S. T. Lee, *Nanoscale*, 2011, **3**, 777–791.
82 M. Rupp, *Int. J. Quantum Chem.*, 2015, **115**, 1003–1004.
83 A. O. Lyakhov, A. R. Oganov, H. T. Stokes and Q. Zhu, *Comput. Phys. Commun.*, 2013, **184**, 1172–1182.
84 J. Harris, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1985, **31**, 1770–1779.
85 K. E. Jelfs, E. Flikkema and S. T. Bromley, *Phys. Chem. Chem. Phys.*, 2013, **15**, 20438–20443.
86 A. Cuko, A. Macià, M. Calatayud and S. T. Bromley, *Comput. Theor. Chem.*, 2017, **1102**, 38–43.
87 N. L. Abraham and M. I. J. Probert, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2006, **73**, 224104.
88 M. Sierka, T. K. Todorova, J. Sauer, S. Kaya, D. Stacchiola, J. Weissenrieder, S. Shaikhutdinov and H. J. Freund, *J. Chem. Phys.*, 2007, **126**, 234710.
89 M. Sierka, *Prog. Surf. Sci.*, 2010, **85**, 398–434.
90 S. Kaya, J. Weissenrieder, D. Stacchiola, T. K. Todorova, M. Sierka, J. Sauer, S. Shaikhutdinov and H. J. Freund, *Surf. Sci.*, 2008, **602**, 3338–3342.
91 M. Batzill and U. Diebold, *Prog. Surf. Sci.*, 2005, **79**, 47–154.
92 L. R. Merte, M. S. Jorgensen, K. Pussi, J. Gustafson, M. Shipilin, A. Schaefer, C. Zhang, J. Rawle, C. Nicklin, G. Thornton, R. Lindsay, B. Hammer and E. Lundgren, *Phys. Rev. Lett.*, 2017, **119**, 096102.
93 A. Fujishima, X. Zhang and D. Tryk, *Surf. Sci. Rep.*, 2008, **63**, 515–582.
94 U. Martinez, L. B. Vilhelmsen, H. H. Kristoffersen, J. Stausholm-Møller and B. Hammer, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2011, **84**, 205434.
95 U. Martinez, J. O. Hansen, E. Lira, H. H. Kristoffersen, P. Huo, R. Bechstein, E. Laegsgaard, F. Besenbacher, B. Hammer and S. Wendt, *Phys. Rev. Lett.*, 2012, **109**, 155501.
96 R. Bechstein, H. H. Kristoffersen, L. B. Vilhelmsen, F. Rieboldt, J. Stausholm-Moller, S. Wendt, B. Hammer and F. Besenbacher, *Phys. Rev. Lett.*, 2012, **108**, 236103.
97 C. L. Freeman, F. Claeyssens, N. L. Allan and J. H. Harding, *Phys. Rev. Lett.*, 2006, **96**, 066102.
98 S. M. Kozlov, I. Demiroglu, K. M. Neyman and S. T. Bromley, *Nanoscale*, 2015, **7**, 4361–4366.
99 N. Krainara, J. Limtrakul, F. Illas and S. T. Bromley, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2011, **83**, 233305.
100 N. Krainara, J. Limtrakul, F. Illas and S. T. Bromley, *J. Phys. Chem. C*, 2013, **117**, 22908–22914.
101 G. A. Ferguson, F. Mehmood, R. B. Rankin, J. P. Greeley, S. Vajda and L. A. Curtiss, *Top. Catal.*, 2012, **55**, 353–365.
102 K. Miyazaki and T. Inoue, *Surf. Sci.*, 2002, **501**, 93–101.
103 J. Zhuang, T. Kojima, W. Zhang, L. Liu, L. Zhao and Y. Li, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2002, **65**, 045411.

104 M. Eckhoff, D. Schebarchov and D. J. Wales, *J. Phys. Chem. Lett.*, 2017, **8**, 5402–5407.
105 R. Ismail, R. Ferrando and R. L. Johnston, *J. Phys. Chem. C*, 2012, **117**, 293–301.
106 R. Ferrando, G. Rossi, A. C. Levi, Z. Kuntova, F. Nita, A. Jelea, C. Mottet, G. Barcaro, A. Fortunelli and J. Goniakowski, *J. Chem. Phys.*, 2009, **130**, 174702.
107 J. Goniakowski, A. Jelea, C. Mottet, G. Barcaro, A. Fortunelli, Z. Kuntova, F. Nita, A. C. Levi, G. Rossi and R. Ferrando, *J. Chem. Phys.*, 2009, **130**, 174703.
108 S. M. Kozlov, H. A. Aleksandrov, J. Goniakowski and K. M. Neyman, *J. Chem. Phys.*, 2013, **139**, 084701.
109 B. Yang, C. Liu, A. Halder, E. C. Tyo, A. B. F. Martinson, S. Seifert, P. Zapol, L. A. Curtiss and S. Vajda, *J. Phys. Chem. C*, 2017, **121**, 10406–10412.
110 Y. Lei, F. Mehmood, S. Lee, J. P. Greeley, B. Lee, S. Seifert, R. E. Winans, J. W. Elam, R. J. Meyer, P. C. Redfern, D. Teschner, R. Schlogl, M. J. Pellin, L. A. Curtiss and S. Vajda, *Science*, 2010, **328**, 224–228.
111 S. Vajda, M. J. Pellin, J. P. Greeley, C. L. Marshall, L. A. Curtiss, G. A. Ballentine, J. W. Elam, S. Catillon-Mucherie, P. C. Redfern, F. Mehmood and P. Zapol, *Nat. Mater.*, 2009, **8**, 213–216.
112 G. Barcaro and A. Fortunelli, *J. Chem. Theory Comput.*, 2005, **1**, 972–985.
113 J. B. A. Davis, S. L. Horswell and R. L. Johnston, *J. Phys. Chem. C*, 2016, **120**, 3759–3765.
114 L. B. Vilhelmsen and B. Hammer, *Phys. Rev. Lett.*, 2012, **108**, 126101.
115 L. B. Vilhelmsen and B. Hammer, *J. Chem. Phys.*, 2013, **139**, 204701.
116 D. E. Jiang, S. H. Overbury and S. Dai, *J. Phys. Chem. Lett.*, 2011, **2**, 1211–1215.
117 C. Jia and W. Fan, *Phys. Chem. Chem. Phys.*, 2015, **17**, 30736–30743.
118 D. A. H. Cunningham, W. Vogel, H. Kageyama, S. Tsubota and M. Haruta, *J. Catal.*, 1998, **177**, 1–10.
119 G. Barcaro, E. Apra and A. Fortunelli, *Chemistry*, 2007, **13**, 6408–6418.
120 R. Fiala, A. Figueroba, A. Bruix, M. Vaclavu, A. Rednyk, I. Khalakhan, M. Vorokhta, J. Lavkova, F. Illas, V. Potin, I. Matolinova, K. M. Neyman and V. Matolin, *Appl. Catal., B*, 2016, **197**, 262–270.
121 A. Figueroba, G. Kovács, A. Bruix and K. M. Neyman, *Catal. Sci. Technol.*, 2016, **6**, 6806–6813.
122 L. O. Paz-Borbòn, A. Lopez-Martinez, I. L. Garzon, A. Posada-Amarillas and H. Grönbeck, *Phys. Chem. Chem. Phys.*, 2017, **19**, 17845–17855.
123 Y. Lykhach, S. M. Kozlov, T. Skala, A. Tovt, V. Stetsovych, N. Tsud, F. Dvorak, V. Johanek, A. Neitzel, J. Myslivecek, S. Fabris, V. Matolin, K. M. Neyman and J. Libuda, *Nat. Mater.*, 2016, **15**, 284–288.
124 C. T. Campbell, *Surf. Sci. Rep.*, 1997, **27**, 1–111.
125 L. Xu, G. Henkelman, C. T. Campbell and H. Jonsson, *Phys. Rev. Lett.*, 2005, **95**, 146103.
126 L. Xu, C. T. Campbell, H. Jónsson and G. Henkelman, *Surf. Sci.*, 2007, **601**, 3133–3142.
127 G. Barcaro and A. Fortunelli, *New J. Phys.*, 2007, **9**, 22.
128 R. Ouyang, J. X. Liu and W. X. Li, *J. Am. Chem. Soc.*, 2013, **135**, 1760–1771.
129 J. G. Wang and B. Hammer, *Phys. Rev. Lett.*, 2006, **97**, 136107.
130 F. Rieboldt, L. B. Vilhelmsen, S. Koust, J. V. Lauritsen, S. Helveg, L. Lammich, F. Besenbacher, B. Hammer and S. Wendt, *J. Chem. Phys.*, 2014, **141**, 214702.
131 G. Fisicaro, M. Sicher, M. Amsler, S. Saha, L. Genovese and S. Goedecker, *Phys. Rev. Mater.*, 2017, **1**, 033609.
132 F. R. Negreiros, E. Apra, G. Barcaro, L. Sementa, S. Vajda and A. Fortunelli, *Nanoscale*, 2012, **4**, 1208–1219.
133 F. R. Negreiros, L. Sementa, G. Barcaro, S. Vajda, E. Aprá and A. Fortunelli, *ACS Catal.*, 2012, **2**, 1860–1864.
134 R. Qin, P. Liu, G. Fu and N. Zheng, *Small Methods*, 2018, **2**, 1700286.
135 L. Liu, U. Díaz, R. Arenal, G. Agostini, P. Concepción and A. Corma, *Nat. Mater.*, 2016, **16**, 132–138.
136 A. Goldbach and M. Saboungi, in *Encyclopedia of Inorganic and Bioinorganic Chemistry*, ed. R. A. Scott, John Wiley & Sons, Ltd, Chichester, UK, 2011, DOI: 10.1002/9781119951438.eibc0339.
137 A. S. Kuznetsov, V. K. Tikhomirov, M. V. Shestakov and V. V. Moshchalkov, *Nanoscale*, 2013, **5**, 10065–10075.
138 G. Lu, S. Li, Z. Guo, O. K. Farha, B. G. Hauser, X. Qi, Y. Wang, X. Wang, S. Han, X. Liu, J. S. DuChene, H. Zhang, Q. Zhang, X. Chen, J. Ma, S. C. J. Loo, W. D. Wei, Y. Yang, J. T. Hupp and F. Huo, *Nat. Chem.*, 2012, **4**, 310–316.
139 A. Uzun, D. A. Dixon and B. C. Gates, *ChemCatChem*, 2011, **3**, 95–107.
140 V. K. Markova, G. N. Vayssilov, A. Genest and N. Rosch, *Catal. Sci. Technol.*, 2016, **6**, 1726–1736.
141 S. G. Chiodo and T. Mineva, *J. Phys. Chem. C*, 2016, **120**, 4471–4480.
142 C. Di Paola, L. Pavan, R. D'Agosta and F. Baletto, *Nanoscale*, 2017, **9**, 15658–15665.
143 J. Antúnez-García, D. H. Galván, A. Posada-Amarillas and V. Petranovskii, *J. Mol. Struct.*, 2014, **1059**, 232–238.
144 L. B. Vilhelmsen, K. S. Walton and D. S. Sholl, *J. Am. Chem. Soc.*, 2012, **134**, 12807–12816.
145 L. B. Vilhelmsen and D. S. Sholl, *J. Phys. Lett.*, 2012, **3**, 3702–3706.
146 L. B. Vilhelmsen and B. Hammer, *J. Chem. Phys.*, 2014, **141**, 044711.
147 L. Grajciar, *J. Phys. Chem. C*, 2016, **120**, 27050–27065.
148 D. Palagin, A. J. Knorpp, A. B. Pinar, M. Ranocchiari and J. A. van Bokhoven, *Nanoscale*, 2017, **9**, 1144–1153.
149 A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. Bjerre Jensen, J. Kermode, J. R. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng and K. W. Jacobsen, *J. Phys.: Condens. Matter*, 2017, **29**, 273002.

150 K. Reuter and M. Scheffler, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2002, **65**, 035406.
151 K. Reuter and M. Scheffler, *Phys. Rev. Lett.*, 2003, **90**, 046103.
152 K. Reuter, *Catal. Lett.*, 2016, **146**, 541–563.
153 K. Reuter, C. P. Plaisance, H. Oberhofer and M. Andersen, *J. Chem. Phys.*, 2017, **146**, 040901.
154 J. Rogal, K. Reuter and M. Scheffler, *Phys. Rev. Lett.*, 2007, **98**, 046101.
155 X. Huang, J. W. Bennett, M. N. Hang, E. D. Laudadio, R. J. Hamers and S. E. Mason, *J. Phys. Chem. C*, 2017, **121**, 5069–5080.
156 A. S. M. Jonayat, A. C. T. van Duin and M. J. Janik, *J. Phys. Chem. C*, 2017, **121**, 21439–21448.
157 W. A. Saidi, M. Lee, L. Li, G. Zhou and A. J. H. McGaughey, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2012, **86**, 245429.
158 Y. Zhen and R. Karsten, *ChemCatChem*, 2018, **10**, 465–469.
159 G. Li, E. A. Pidko, R. A. van Santen, C. Li and E. J. M. Hensen, *J. Phys. Chem. C*, 2013, **117**, 413–426.
160 S. Grundner, M. A. C. Markovits, G. Li, M. Tromp, E. A. Pidko, E. J. M. Hensen, A. Jentys, M. Sanchez-Sanchez and J. A. Lercher, *Nat. Commun.*, 2015, **6**, 7546.
161 T. Wang, X. X. Tian, Y. Yang, Y. W. Li, J. G. Wang, M. Beller and H. J. Jiao, *Surf. Sci.*, 2016, **651**, 195–202.
162 C. Liu, G. Li, E. J. M. Hensen and E. A. Pidko, *ACS Catal.*, 2015, **5**, 7024–7033.
163 E. M. Stuve, R. J. Madix and C. R. Brundle, *Surf. Sci.*, 1984, **146**, 155–178.
164 L. Chen, H. Falsig, T. V. W. Janssens and H. Gronbeck, *J. Catal.*, 2018, **358**, 179–186.
165 C. Paolucci, A. A. Parekh, I. Khurana, J. R. Di Iorio, H. Li, J. D. Albarracin Caballero, A. J. Shih, T. Anggara, W. N. Delgass, J. T. Miller, F. H. Ribeiro, R. Gounder and W. F. Schneider, *J. Am. Chem. Soc.*, 2016, **138**, 6028–6048.
166 J. Engelhardt, P. B. Lyu, P. Nachtigall, F. Schuth and A. M. Garcia, *ChemCatChem*, 2017, **9**, 1985–1991.
167 S. Posada-Perez, F. Vines, R. Valero, J. A. Rodriguez and F. Illas, *Surf. Sci.*, 2017, **656**, 24–32.
168 J. He, A. Morales-Garcia, O. Bludsky and P. Nachtigall, *CrystEngComm*, 2016, **18**, 3808–3818.
169 S. Kenmoe and P. U. Biedermann, *J. Chem. Phys.*, 2018, **148**, 054701.
170 R. G. Zhang, X. B. Hao, T. Duan and B. J. Wang, *Fuel Process. Technol.*, 2017, **156**, 253–264.
171 J. M. Lorenzi, S. Matera and K. Reuter, *ACS Catal.*, 2016, **6**, 5191–5197.
172 G. A. Fergusson, V. Vorotnikov, N. Wunder, J. Clark, K. Gruchalla, T. Bartholomew, D. J. Robichaud and G. T. Beckham, *J. Phys. Chem. C*, 2016, **120**, 26249–26258.
173 Z. Yao and K. Reuter, *ChemCatChem*, 2018, **10**, 465–469.
174 M. Vandichel, A. Moscu and H. Gronbeck, *ACS Catal.*, 2017, **7**, 7431–7441.
175 J. S. Kim, B. K. Kim and Y. C. Kim, *J. Nanosci. Nanotechnol.*, 2015, **15**, 8205–8210.
176 A. Farkas, D. Fantauzzi, J. E. Mueller, T. W. Zhu, C. Papp, H. P. Steinruck and T. Jacob, *J. Electron Spectrosc. Relat. Phenom.*, 2017, **221**, 44–57.
177 K. S. Exner and H. Over, *Acc. Chem. Res.*, 2017, **50**, 1240–1247.
178 K. Emmerich, F. Koeniger, H. Kaden and P. Thissen, *J. Colloid Interface Sci.*, 2015, **448**, 24–31.
179 T. Lee, Y. Lee, S. Piccinin and A. Soon, *J. Phys. Chem. C*, 2017, **121**, 2228–2233.
180 M. Maestri, *Chem. Commun.*, 2017, **53**, 10244–10254.
181 K. F. Kalz, R. Kraehnert, M. Dvoyashkin, R. Dittmeyer, R. Glaser, U. Krewer, K. Reuter and J. D. Grunwaldt, *ChemCatChem*, 2017, **9**, 17–29.
182 M. K. Sabbe, M.-F. Reyniers and K. Reuter, *Catal. Sci. Technol.*, 2012, **2**, 2010–2024.
183 E. A. Carter, *Science*, 2008, **321**, 800–803.
184 J. K. Norskov, T. Bligaard, J. Rossmeisl and C. H. Christensen, *Nat. Chem.*, 2009, **1**, 37–46.
185 B. A. De Moor, A. Ghysels, M.-F. Reyniers, V. Van Speybroeck, M. Waroquier and G. B. Marin, *J. Chem. Theory Comput.*, 2011, **7**, 1090–1101.
186 G. Piccini, M. Alessio and J. Sauer, *Angew. Chem., Int. Ed.*, 2016, **55**, 5235–5237.
187 G. Piccini and J. Sauer, *J. Chem. Theory Comput.*, 2013, **9**, 5038–5045.
188 G. Piccini and J. Sauer, *J. Chem. Theory Comput.*, 2014, **10**, 2479–2487.
189 B. A. De Moor, M. F. Reyniers and G. B. Marin, *Phys. Chem. Chem. Phys.*, 2009, **11**, 2939–2958.
190 V. Van Speybroeck, K. Hemelsoet, L. Joos, M. Waroquier, R. G. Bell and C. R. A. Catlow, *Chem. Soc. Rev.*, 2015, **44**, 7044–7111.
191 D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*, Academic Press, San Diego, 2nd edn, 2002.
192 J. Leiding and J. D. Coe, *J. Chem. Phys.*, 2016, **144**, 174109.
193 J. Van Der Mynsbrugge, A. Janda, S. Mallikarjun Sharada, L. C. Lin, V. Van Speybroeck, M. Head-Gordon and A. T. Bell, *ACS Catal.*, 2017, **7**, 2685–2697.
194 C. Chipot and A. Pohorille, *Free energy calculations: theory and applications in chemistry and biology*, Springer, New York, Study edn, 2007.
195 C. D. Christ, A. E. Mark and W. F. van Gunsteren, *J. Comput. Chem.*, 2009, **31**, 1569–1582.
196 N. Hansen and W. F. Van Gunsteren, *J. Chem. Theory Comput.*, 2014, **10**, 2632–2647.
197 G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187–199.
198 A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 12562.
199 J. G. Kirkwood, *J. Chem. Phys.*, 1935, **3**, 300–313.
200 A. Rodríguez-Fortea, M. Iannuzzi and M. Parrinello, *J. Phys. Chem. B*, 2005, **110**, 3477–3484.
201 M. K. Kostov, E. E. Santiso, A. M. George, K. E. Gubbins and M. B. Nardelli, *Phys. Rev. Lett.*, 2005, **95**, 1–4.
202 A. Rodríguez-Fortea, M. Iannuzzi and M. Parrinello, *J. Phys. Chem. C*, 2007, **111**, 2251–2258.
203 A. Rodríguez-Fortea and M. Iannuzzi, *J. Phys. Chem. C*, 2008, **112**, 19642–19648.

204 E. Molina-Montes, D. Donadio, A. Hernández-Laguna, C. I. Sainz-Díaz and M. Parrinello, *J. Phys. Chem. B*, 2008, **112**, 7051–7060.
205 M. Ceriotti and M. Bernasconi, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2007, **76**, 245309.
206 J. Kiss, J. Frenzel, N. N. Nair, B. Meyer and D. Marx, *J. Chem. Phys.*, 2011, **134**, 0–14.
207 E. Schreiner, N. N. Nair, C. Wittekindt and D. Marx, *J. Am. Chem. Soc.*, 2011, **133**, 8216–8226.
208 K. Koizumi, M. Boero, Y. Shigeta and A. Oshiyama, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2012, **85**, 1–4.
209 I. F. W. Kuo, C. D. Grant, R. H. Gee, S. C. Chinn and A. H. Love, *J. Phys. Chem. C*, 2012, **116**, 9631–9635.
210 G. Santarossa, K. Hahn and A. Baiker, *Langmuir*, 2013, **29**, 5487–5499.
211 J. Frenzel, J. Kiss, N. N. Nair, B. Meyer and D. Marx, *Phys. Status Solidi B*, 2013, **250**, 1174–1190.
212 K. Koizumi, M. Boero, Y. Shigeta and A. Oshiyama, *J. Phys. Chem. Lett.*, 2013, **4**, 1592–1596.
213 T. K. Ghosh and N. N. Nair, *ChemCatChem*, 2013, **5**, 1811–1821.
214 S. L. C. Moors, K. De Wispelaere, J. Van Der Mynsbrugge, M. Waroquier and V. Van Speybroeck, *ACS Catal.*, 2013, **3**, 2556–2567.
215 T. Laino and A. Curioni, *New J. Phys.*, 2013, **15**, 095009.
216 J. Van Der Mynsbrugge, S. L. C. Moors, K. De Wispelaere and V. Van Speybroeck, *ChemCatChem*, 2014, **6**, 1906–1918.
217 K. Koizumi, K. Nobusada and M. Boero, *J. Phys. Chem. C*, 2015, **119**, 15421–15427.
218 S. H. Mushrif, J. J. Varghese and C. B. Krishnamurthy, *Phys. Chem. Chem. Phys.*, 2015, **17**, 4961–4969.
219 K. K. Ghuman, S. Yadav and C. V. Singh, *J. Phys. Chem. C*, 2015, **119**, 6518–6529.
220 T. K. Ghosh and N. N. Nair, *Surf. Sci.*, 2015, **632**, 20–27.
221 F. R. Negreiros, M. F. Camellone and S. Fabris, *J. Phys. Chem. C*, 2015, **119**, 21567–21573.
222 L. Martínez-Suárez, N. Siemer, J. Frenzel and D. Marx, *ACS Catal.*, 2015, **5**, 4201–4218.
223 K. Dewispelaere, B. Ensing, A. Ghysels, E. J. Meijer and V. Vanspeybroeck, *Chem. – Eur. J.*, 2015, **21**, 9385–9396.
224 V. Haigis, F. X. Coudert, R. Vuilleumier, A. Boutin and A. H. Fuchs, *J. Phys. Chem. Lett.*, 2015, **6**, 4365–4370.
225 K. Koizumi, K. Nobusada and M. Boero, *Chem. – Eur. J.*, 2016, **22**, 5181–5188.
226 D. Munoz-Santiburcio, A. Hernandez-Laguna and C. I. Sainz-Díaz, *J. Phys. Chem. C*, 2016, **120**, 28186–28192.
227 K. De Wispelaere, S. Bailleul and V. Van Speybroeck, *Catal. Sci. Technol.*, 2016, **6**, 2686–2705.
228 K. De Wispelaere, C. S. Wondergem, B. Ensing, K. Hemelsoet, E. J. Meijer, B. M. Weckhuysen, V. Van Speybroeck and J. Ruiz-Martínez, *ACS Catal.*, 2016, **6**, 1991–2002.
229 J. Hajek, J. Van Der Mynsbrugge, K. De Wispelaere, P. Cnudde, L. Vanduyfhuys, M. Waroquier and V. Van Speybroeck, *J. Catal.*, 2016, **340**, 227–235.
230 P. Cnudde, K. De Wispelaere, J. Van der Mynsbrugge, M. Waroquier and V. Van Speybroeck, *J. Catal.*, 2017, **345**, 53–69.
231 M. Ghoussoub, S. Yadav, K. K. Ghuman, G. A. Ozin and C. V. Singh, *ACS Catal.*, 2016, **6**, 7109–7117.
232 O. Valsson, P. Tiwary and M. Parrinello, *Annu. Rev. Phys. Chem.*, 2016, **67**, 159–184.
233 A. Barducci, G. Bussi and M. Parrinello, *Phys. Rev. Lett.*, 2008, **100**, 020603.
234 P. Raiteri, A. Laio, F. L. Gervasio, C. Micheletti and M. Parrinello, *J. Phys. Chem. B*, 2006, **110**, 3533–3539.
235 K. M. Bal and E. C. Neyts, *J. Chem. Theory Comput.*, 2015, **11**, 4545–4554.
236 K. M. Bal, S. Huygh, A. Bogaerts and E. C. Neyts, *Plasma Sources Sci. Technol.*, 2018, **27**, 024001.
237 G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Comput. Phys. Commun.*, 2014, **185**, 604–613.
238 Y. Q. Gao, *J. Chem. Phys.*, 2008, **128**, 64105.
239 K. Leung, I. Nielsen and L. Criscenti, *J. Am. Chem. Soc.*, 2009, **131**, 18358–18365.
240 V. M. Sánchez, J. A. Cojulun and D. A. Scherlis, *J. Phys. Chem. C*, 2010, **114**, 11522–11526.
241 S. Schnur and A. Groß, *Catal. Today*, 2011, **165**, 129–137.
242 X. Liu and D. R. Salahub, *J. Am. Chem. Soc.*, 2015, **137**, 4249–4259.
243 Z. N. Chen, L. Shen, M. Yang, G. Fu and H. Hu, *J. Phys. Chem. C*, 2015, **119**, 26422–26428.
244 G. Sun and H. Jiang, *J. Chem. Phys.*, 2015, **143**, 234706.
245 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 1011–1021.
246 E. Rosta and G. Hummer, *J. Chem. Theory Comput.*, 2015, **11**, 276–285.
247 Y. Q. Gao and L. Yang, *J. Chem. Phys.*, 2006, **125**, 114103.
248 Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.*, 1999, **314**, 141–151.
249 G. Bussi, F. L. Gervasio, A. Laio and M. Parrinello, *J. Am. Chem. Soc.*, 2006, **128**, 13435–13441.
250 M. Boero, M. Parrinello and K. Terakura, *J. Am. Chem. Soc.*, 1998, **120**, 2746–2752.
251 T. Bučko, L. Benco, J. Hafner and J. G. Ángyán, *J. Catal.*, 2007, **250**, 171–183.
252 T. Bučko, L. Benco, O. Dubay, C. Dellago and J. Hafner, *J. Chem. Phys.*, 2009, **131**, 214508.
253 T. Bučko and J. Hafner, *J. Phys.: Condens. Matter*, 2010, **22**, 384201.
254 F. Zipoli, R. Car, M. H. Cohen and A. Selloni, *J. Am. Chem. Soc.*, 2010, **132**, 8593–8601.
255 T. Bučko, L. Benco, J. Hafner and J. G. Ángyán, *J. Catal.*, 2011, **279**, 220–228.
256 L. Benco, *J. Catal.*, 2013, **298**, 122–129.
257 T. Bučko and J. Hafner, *J. Catal.*, 2015, **329**, 32–48.
258 T. Cheng, H. Xiao and W. A. Goddard, *J. Phys. Chem. Lett.*, 2015, **6**, 4767–4773.
259 T. Cheng, H. Xiao and W. A. Goddard, *J. Am. Chem. Soc.*, 2016, **138**, 13802–13805.
260 T. Sheng, D. Wang, W. F. Lin, P. Hu and S. G. Sun, *Electrochim. Acta*, 2016, **190**, 446–454.
261 T. Sheng, J.-Y. Ye, W.-F. Lin and S.-G. Sun, *Phys. Chem. Chem. Phys.*, 2017, **19**, 7476–7480.

262 Y. Ming, N. Kumar and D. J. Siegel, *ACS Omega*, 2017, **2**, 4921–4928.
263 H. Li, C. Paolucci and W. F. Schneider, *J. Chem. Theory Comput.*, 2018, **14**, 929–932.
264 T. Bučko, S. Chibani, J.-F. Paul, L. Cantrel and M. Badawi, *Phys. Chem. Chem. Phys.*, 2017, **19**, 27530–27543.
265 T. Sheng and S. G. Sun, *Appl. Surf. Sci.*, 2018, **428**, 514–519.
266 E. A. Carter, G. Ciccotti, J. T. Hynes and R. Kapral, *Chem. Phys. Lett.*, 1989, **156**, 472–477.
267 S. Zheng and J. Pfaendtner, *Mol. Simul.*, 2015, **41**, 55–72.
268 J. J. P. Stewart, L. P. Davis and L. W. Burggraf, *J. Comput. Chem.*, 1987, **8**, 1117–1123.
269 C. Dellago, P. G. Bolhuis and P. L. Geissler, *Adv. Chem. Phys.*, 2002, **123**, 1–78.
270 P. M. Zimmerman, D. C. Tranca, J. Gomes, D. S. Lambrecht, M. Head-Gordon and A. T. Bell, *J. Am. Chem. Soc.*, 2012, **134**, 19468–19476.
271 D. C. Tranca, N. Hansen, J. A. Swisher, B. Smit and F. J. Keil, *J. Phys. Chem. C*, 2012, **116**, 23408–23417.
272 C. S. Lo, R. Radhakrishnan and B. L. Trout, *Catal. Today*, 2005, **105**, 93–105.
273 J. Gomes, M. Head-Gordon and A. T. Bell, *J. Phys. Chem. C*, 2014, **118**, 21409–21419.
274 F. Göltl, A. Grüneis, T. Bučko and J. Hafner, *J. Chem. Phys.*, 2012, **137**, 114111.
275 X. Fu, L. Yang and Y. Q. Gao, *J. Chem. Phys.*, 2007, **127**, 154106.
276 L. Martínez-Suárez, J. Frenzel, D. Marx and B. Meyer, *Phys. Rev. Lett.*, 2013, **110**, 086108.
277 M.-F. Reyniers and G. B. Marin, *Annu. Rev. Chem. Biomol. Eng.*, 2014, **5**, 563–594.
278 M. Salciccioli, M. Stamatakis, S. Caratzoulas and D. G. Vlachos, *Chem. Eng. Sci.*, 2011, **66**, 4319–4355.
279 F. J. Keil, in *Multiscale Molecular Methods in Applied Chemistry*, ed. B. Kirchner and J. Vrabec, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 69–107, DOI: 10.1007/128_2011_128.
280 S. Michail, *J. Phys.: Condens. Matter*, 2015, **27**, 013001.
281 A. T. Marshall, *Curr. Opin. Electrochem.*, 2018, **7**, 75–80.
282 K. Reuter, in *Operando Research in Heterogeneous Catalysis*, ed. J. Frenken and I. Groot, Springer International Publishing, Cham, 2017, pp. 151–188, DOI: 10.1007/978-3-319-44439-0_7.
283 A. Kulkarni, S. Siahrostami, A. Patel and J. K. Nørskov, *Chem. Rev.*, 2018, **118**, 2302–2312.
284 Y. Mao, H. F. Wang and P. Hu, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2017, **7**, e1321.
285 A. H. Motagamwala, M. R. Ball and J. A. Dumesic, *Annu. Rev. Chem. Biomol. Eng.*, 2018, **9**, 413–450.
286 H. Prats, F. Illas and R. Sayós, *Int. J. Quantum Chem.*, 2018, **118**, e25518.
287 J. A. Dumesic, D. F. Rudd, L. M. Aparicio, J. E. Rekoske and A. A. Trevino, *The Microkinetics of Heterogeneous Catalysis*, American Chemical Society, 1992.
288 M. Pineda and M. Stamatakis, *J. Chem. Phys.*, 2017, **147**, 024105.
289 Z. Chen, H. Wang, N. Q. Su, S. Duan, T. Shen and X. Xu, *ACS Catal.*, 2018, **8**, 859–868.
290 D.-J. Liu, F. Zahariev, M. S. Gordon and J. W. Evans, *J. Phys. Chem. C*, 2016, **120**, 28639–28653.
291 M. Jørgensen and H. Grönbeck, *J. Phys. Chem. C*, 2017, **121**, 7199–7207.
292 N. Nikbin, S. Caratzoulas and D. G. Vlachos, *ChemCatChem*, 2012, **4**, 504–511.
293 V. Van Speybroeck, K. De Wispelaere, J. Van der Mynsbrugge, M. Vandichel, K. Hemelsoet and M. Waroquier, *Chem. Soc. Rev.*, 2014, **43**, 7326–7357.
294 X. Zhang, J.-X. Liu, B. Zijlstra, I. A. W. Filot, Z. Zhou, S. Sun and E. J. M. Hensen, *Nano Energy*, 2018, **43**, 200–209.
295 I. A. W. Filot, R. A. v. Santen and E. J. M. Hensen, *Angew. Chem., Int. Ed.*, 2014, **53**, 12746–12750.
296 R. Y. Rohling, E. Uslamin, B. Zijlstra, I. C. Tranca, I. A. W. Filot, E. J. M. Hensen and E. A. Pidko, *ACS Catal.*, 2018, **8**, 760–769.
297 G. Li, E. A. Pidko, I. A. W. Filot, R. A. van Santen, C. Li and E. J. M. Hensen, *J. Catal.*, 2013, **308**, 386–397.
298 R. Y. Brogaard, C.-M. Wang and F. Studt, *ACS Catal.*, 2014, **4**, 4504–4509.
299 C. A. Wolcott, A. J. Medford, F. Studt and C. T. Campbell, *J. Catal.*, 2015, **330**, 197–207.
300 C. Stegelmann, A. Andreasen and C. T. Campbell, *J. Am. Chem. Soc.*, 2009, **131**, 8077–8082.
301 P. Mehta, P. Barboun, F. A. Herrera, J. Kim, P. Rumbach, D. B. Go, J. C. Hicks and W. F. Schneider, *Nat. Catal.*, 2018, **1**, 269–275.
302 C. T. Campbell, *ACS Catal.*, 2017, **7**, 2770–2779.
303 A. Vojvodic and J. K. Nørskov, *Natl. Sci. Rev.*, 2015, **2**, 140–143.
304 J. Cheng, P. Hu, P. Ellis, S. French, G. Kelly and C. M. Lok, *J. Phys. Chem. C*, 2008, **112**, 1308–1311.
305 Z. W. Ulissi, A. J. Medford, T. Bligaard and J. K. Nørskov, *Nat. Commun.*, 2017, **8**, 14621.
306 J.-X. Liu, Y. Su, I. A. W. Filot and E. J. M. Hensen, *J. Am. Chem. Soc.*, 2018, **140**, 4580–4587.
307 C. Liu, I. Tranca, R. A. van Santen, E. J. M. Hensen and E. A. Pidko, *J. Phys. Chem. C*, 2017, **121**, 23520–23530.
308 F. Abild-Pedersen, J. Greeley, F. Studt, J. Rossmeisl, T. R. Munter, P. G. Moses, E. Skúlason, T. Bligaard and J. K. Nørskov, *Phys. Rev. Lett.*, 2007, **99**, 016105.
309 F. Calle-Vallejo, D. Loffreda, M. T. M. Koper and P. Sautet, *Nat. Chem.*, 2015, **7**, 403.
310 L. Yu, L. Vilella and F. Abild-Pedersen, *Commun. Chem.*, 2018, **1**, 2.
311 A. A. Latimer, A. R. Kulkarni, H. Aljama, J. H. Montoya, J. S. Yoo, C. Tsai, F. Abild-Pedersen, F. Studt and J. K. Nørskov, *Nat. Mater.*, 2016, **16**, 225.
312 M. L. Pegis, C. F. Wise, B. Koronkiewicz and J. M. Mayer, *J. Am. Chem. Soc.*, 2017, **139**, 11000–11003.
313 T. Z. H. Gani and H. J. Kulik, *ACS Catal.*, 2018, **8**, 975–986.
314 A. Logadottir, T. H. Rod, J. K. Nørskov, B. Hammer, S. Dahl and C. J. H. Jacobsen, *J. Catal.*, 2001, **197**, 229–231.
315 A. Vojvodic, A. J. Medford, F. Studt, F. Abild-Pedersen, T. S. Khan, T. Bligaard and J. K. Nørskov, *Chem. Phys. Lett.*, 2014, **598**, 108–112.

316 C. Liu, R. A. van Santen, A. Poursaeidesfahani, T. J. H. Vlugt, E. A. Pidko and E. J. M. Hensen, *ACS Catal.*, 2017, **7**, 8613–8627.
317 E. A. Pidko, *ACS Catal.*, 2017, **7**, 4230–4234.
318 L. Chiang, B. Lu and I. Castillo, *Annu. Rev. Chem. Biomol. Eng.*, 2017, **8**, 63–85.
319 G. B. Goh, N. O. Hodas and A. Vishnu, *J. Comput. Chem.*, 2017, **38**, 1291–1307.
320 D. Fooshee, A. Mood, E. Gutman, M. Tavakoli, G. Urban, F. Liu, N. Huynh, D. Van Vranken and P. Baldi, *Mol. Syst. Des. Eng.*, 2018, **3**, 442–452.
321 J. D. Evans and F.-X. Coudert, *Chem. Mater.*, 2017, **29**, 7833–7839.
322 C. W. Coley, R. Barzilay, T. S. Jaakkola, W. H. Green and K. F. Jensen, *ACS Cent. Sci.*, 2017, **3**, 434–443.
323 F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley and O. A. von Lilienfeld, *J. Chem. Theory Comput.*, 2017, **13**, 5255–5264.
324 G. Hautier, C. C. Fischer, A. Jain, T. Mueller and G. Ceder, *Chem. Mater.*, 2010, **22**, 3762–3767.
325 P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier and A. J. Norquist, *Nature*, 2016, **533**, 73–76.
326 P. Schwaller, T. Gaudin, D. Lanyi, C. Bekas and T. Laino, 2017, arXiv:1711.04810.
327 K. Hansen, G. Montavon, F. Biegler, S. Fazli, M. Rupp, M. Scheffler, O. A. von Lilienfeld, A. Tkatchenko and K.-R. Müller, *J. Chem. Theory Comput.*, 2013, **9**, 3404–3419.
328 B. Meredig, A. Agrawal, S. Kirklin, J. E. Saal, J. W. Doak, A. Thompson, K. Zhang, A. Choudhary and C. Wolverton, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2014, **89**, 1–7.
329 R. Jinnouchi and R. Asahi, *J. Phys. Chem. Lett.*, 2017, **8**, 4279–4283.
330 G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K.-R. Müller and O. Anatole von Lilienfeld, *New J. Phys.*, 2013, **15**, 095003.
331 L. Ward, A. Agrawal, A. Choudhary and C. Wolverton, *npj Comput. Mater.*, 2016, **2**, 16028.
332 R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams and A. Aspuru-Guzik, *ACS Cent. Sci.*, 2018, **4**, 268–276.
333 C. Houben and A. A. Lapkin, *Curr. Opin. Chem. Eng.*, 2015, **9**, 1–7.
334 Z. Zhou, X. Li and R. N. Zare, *ACS Cent. Sci.*, 2017, **3**, 1337–1344.
335 Z. W. Ulissi, M. T. Tang, J. Xiao, X. Liu, D. A. Torelli, M. Karamad, K. Cummins, C. Hahn, N. S. Lewis, T. F. Jaramillo, K. Chan and J. K. Nørskov, *ACS Catal.*, 2017, **7**, 6600–6608.
336 G. Pilania, C. Wang, X. Jiang, S. Rajasekaran and R. Ramprasad, *Sci. Rep.*, 2013, **3**, 1–6.
337 H. Li, Z. Zhang and Z. Liu, *Catalysts*, 2017, **7**, 306.
338 J. R. Kitchin, *Nat. Catal.*, 2018, **1**, 230–232.
339 E. Burello, D. Farrusseng and G. Rothenberg, *Adv. Synth. Catal.*, 2004, **346**, 1844–1853.
340 J. K. Nørskov and T. Bligaard, *Angew. Chem., Int. Ed.*, 2013, **52**, 776–777.
341 A. L. Samuel, *IBM J. Res. Dev.*, 1959, **3**, 210–229.
342 W. S. McCulloch and W. Pitts, *Bull. Math. Biophys.*, 1943, **5**, 115–133.
343 Gartner's 2016 Hype Cycle for Emerging Technologies Identifies Three Key Trends That Organizations Must Track to Gain Competitive Advantage, https://www.gartner.com/newsroom/id/3412017, (accessed 14 August 2018).
344 S. B. Kotsiantis, ed. I. Maglogiannis, K. Karpouzis, M. Wallace and J. Soldatos, IOS Press, 2007, pp. 3–24.
345 L. A. Baumes, J. M. Serra, P. Serna and A. Corma, *J. Comb. Chem.*, 2006, **8**, 583–596.
346 M. Fernandez and A. S. Barnard, *ACS Comb. Sci.*, 2016, **18**, 243–252.
347 M. Fernandez, P. G. Boyd, T. D. Daff, M. Z. Aghaji and T. K. Woo, *J. Phys. Lett.*, 2014, **5**, 3056–3060.
348 N. Fey, A. G. Orpen and J. N. Harvey, *Coord. Chem. Rev.*, 2009, **253**, 704–722.
349 J. Jover and N. Fey, *Dalton Trans.*, 2013, **42**, 172–181.
350 L. M. Ghiringhelli, J. Vybiral, S. V. Levchenko, C. Draxl and M. Scheffler, *Phys. Rev. Lett.*, 2015, **114**, 1–5.
351 D. Hand, *ACM SIGKDD Explor. Newsl.*, 1999, **1**, 16–19.
352 J. E. Herr, K. Yao, R. McIntyre, D. Toth and J. Parkhill, *J. Chem. Phys.*, 2018, **148**, 241710.
353 L. Spialter, *J. Chem. Doc.*, 1964, **4**, 261–269.
354 D. Weininger, *J. Chem. Inf. Comput. Sci.*, 1988, **28**, 31–36.
355 S. Heller, A. McNaught, S. Stein, D. Tchekhovskoi and I. Pletnev, *J. Cheminf.*, 2014, **6**, P4.
356 H. L. Morgan, *J. Chem. Doc.*, 1965, **5**, 107–113.
357 S. P. Mbue and K.-H. Cho, *Bull. Korean Chem. Soc.*, 2015, **36**, 1569–1574.
358 A. M. Clark, *J. Chem. Inf. Model.*, 2011, **51**, 3149–3157.
359 P. Vinoth and P. Sankar, *J. Mol. Graphics Modell.*, 2017, **76**, 242–259.
360 R. G. A. Bone, M. A. Firth and R. A. Sykes, *J. Chem. Inf. Comput. Sci.*, 1999, **39**, 846–860.
361 N. M. O'Boyle, *J. Cheminf.*, 2012, **4**, 22.
362 J. P. Janet, L. Chan and H. J. Kulik, *J. Phys. Chem. Lett.*, 2018, **9**, 1064–1071.
363 A. A. Toropov, A. P. Toropova and E. Benfenati, *Chem. Phys. Lett.*, 2008, **461**, 343–347.
364 A. A. Toropov, A. P. Toropova and E. Benfenati, *Mol. Diversity*, 2010, **14**, 183–192.
365 I. V. Tetko, H. P. Varbanov, M. Galanski, M. Talmaciu, J. A. Platts, M. Ravera and E. Gabano, *J. Inorg. Biochem.*, 2016, **156**, 1–13.
366 O. A. Von Lilienfeld, R. Ramakrishnan, M. Rupp and A. Knoll, *Int. J. Quantum Chem.*, 2015, **115**, 1084–1093.
367 M. Karelson, V. S. Lobanov and A. R. Katritzky, *Chem. Rev.*, 1996, **96**, 1027–1044.
368 F. Dioury, A. Duprat, G. Dreyfus, C. Ferroud and J. Cossy, *J. Chem. Inf. Model.*, 2014, **54**, 2718–2731.
369 J. Y. Guo, Y. Minko, C. B. Santiago and M. S. Sigman, *ACS Catal.*, 2017, **7**, 4144–4151.

370 Z. L. Niemeyer, A. Milo, D. P. Hickey and M. S. Sigman, *Nat. Chem.*, 2016, **8**, 610–617.
371 S. Tang, Z. Liu, X. Zhan, R. Cheng, X. He and B. Liu, *J. Mol. Model.*, 2014, **20**, 2129.
372 E. I. Ioannidis, T. Z. H. Gani and H. J. Kulik, *J. Comput. Chem.*, 2016, **37**, 2106–2117.
373 J. P. Janet, T. Z. H. Gani, A. H. Steeves, E. I. Ioannidis and H. J. Kulik, *Ind. Eng. Chem. Res.*, 2017, **56**, 4898–4910.
374 M. Sasaki, H. Hamada, Y. Kintaichi and T. Ito, *Appl. Catal., A*, 1995, **132**, 261–270.
375 M. L. Mohammed, D. Patel, R. Mbeleck, D. Niyogi, D. C. Sherrington and B. Saha, *Appl. Catal., A*, 2013, **466**, 142–152.
376 M. Baysal, M. E. Günay and R. Yıldırım, *Int. J. Hydrogen Energy*, 2017, **42**, 243–254.
377 E. Avşar, *Int. J. Hydrogen Energy*, 2017, **42**, 23326–23333.
378 Ç. Odabaşi, M. E. Günay and R. Yildirim, *Int. J. Hydrogen Energy*, 2014, **39**, 5733–5746.
379 S. Kite, T. Hattori and Y. Murakami, *Appl. Catal., A*, 1994, **114**, L173–L178.
380 D. K. Agrafiotis, D. Bandyopadhyay, J. K. Wegner and H. Van Vlijmen, *J. Chem. Inf. Model.*, 2007, **47**, 1279–1293.
381 P. C. D. Hawkins, *J. Chem. Inf. Model.*, 2017, **57**, 1747–1756.
382 T. Engel, *J. Chem. Inf. Model.*, 2006, **46**, 2267–2277.
383 W. A. Warr, *Mol. Inf.*, 2014, **33**, 469–476.
384 A. Dietz, *J. Chem. Inf. Comput. Sci.*, 1995, **35**, 787–802.
385 M. A. Kayala, C.-A. Azencott, J. H. Chen and P. Baldi, *J. Chem. Inf. Model.*, 2011, **51**, 2209–2222.
386 H. Tadashi, N. Hideyuki, S. Atsushi, K. Shigeharu and M. Yuichi, *Appl. Catal.*, 1989, **50**, L11–L15.
387 Z. Frontistis, V. M. Daskalaki, E. Hapeshi, C. Drosou, D. Fatta-Kassinos, N. P. Xekoukoulotakis and D. Mantzavinos, *J. Photochem. Photobiol., A*, 2012, **240**, 33–41.
388 A. Corma, J. M. Serra, E. Argente, V. Botti and S. Valero, *ChemPhysChem*, 2002, **3**, 939–945.
389 S. Nandi, Y. Badhe, J. Lonari, U. Sridevi, B. S. Rao, S. S. Tambe and B. D. Kulkarni, *Chem. Eng. J.*, 2004, **97**, 115–129.
390 M. A. Akcayol and C. Cinar, *Appl. Therm. Eng.*, 2005, **25**, 2341–2350.
391 Z. Li, J. R. Kermode and A. De Vita, *Phys. Rev. Lett.*, 2015, **114**, 1–5.
392 L. H. Hu, X. J. Wang, L. H. Wong and G. H. Chen, *J. Chem. Phys.*, 2003, **119**, 11501–11507.
393 C. J. Cramer and D. G. Truhlar, *Phys. Chem. Chem. Phys.*, 2009, **11**, 10757.
394 J. Wellendorff, T. L. Silbaugh, D. Garcia-Pintos, J. K. Nørskov, T. Bligaard, F. Studt and C. T. Campbell, *Surf. Sci.*, 2015, **640**, 36–44.
395 R. Jinnouchi, H. Hirata and R. Asahi, *J. Phys. Chem. C*, 2017, **121**, 26397–26405.
396 X. Ma, Z. Li, L. E. K. Achenie and H. Xin, *J. Phys. Chem. Lett.*, 2015, **6**, 3528–3533.
397 J. S. S. Lowndes, B. D. Best, C. Scarborough, J. C. Afflerbach, M. R. Frazier, C. C. O'Hara, N. Jiang and B. S. Halpern, *Nat. Ecol. Evol.*, 2017, **1**, 0160.
398 A. J. Lawson, J. Swienty-Busch, T. Géoui and D. Evans, *ACS Symp. Ser.*, 2014, **1164**, 127–148.
399 M. Baker and D. Penny, *Nature*, 2016, **533**, 452–454.
400 M. Björnmalm and F. Caruso, *Angew. Chem., Int. Ed.*, 2018, **57**, 1122–1123.
401 A. Williams and V. Tkachenko, *J. Comput.-Aided Mol. Des.*, 2014, **28**, 1023–1030.
402 M. Nakata and T. Shimazaki, *J. Chem. Inf. Model.*, 2017, **57**, 1300–1308.
403 S. J. Chalk, *J. Cheminf.*, 2016, **8**, 1–24.
404 Rmarkdown in a scientific workflow, http://predictiveecology.org/2016/10/21/Rmarkdownscience-workflow.html, (accessed 14 August 2018).
405 K. Ram, *Source Code Biol. Med.*, 2013, **8**, 1–8.
406 B. S. Lerner and E. R. Boose, *RDataTracker: Collecting Provenance in an Interactive Scripting Environment*, USENIX Association, Cologne, 2014.
407 B. Baumer, M. Cetinkaya-Rundel, A. Bray, L. Loi and N. J. Horton, 2014, arXiv:1402.1894.
408 J. C. Molloy, *PLoS Biol.*, 2011, **9**, 1–4.
409 P. F. Uhlir and P. Schröder, *Data Sci. J.*, 2007, **6**, OD36–OD53.
410 A. Doerr, *Nat. Methods*, 2010, **7**, 10–11.
411 Positively Negative: A New PLOS ONE Collection focusing on Negative, Null and Inconclusive Results, http://blogs.plos.org/everyone/2015/02/25/positively-negative-new-plosone-collection-focusing-negative-null-inconclusive-results/, (accessed 14 August 2018).
412 V. P. Ananikov and I. P. Beletskaya, *Organometallics*, 2012, **31**, 1595–1604.
413 G. Skoraczyñski, P. Dittwald, B. Miasojedow, S. Szymkuc, E. P. Gajewska, B. A. Grzybowski and A. Gambin, *Sci. Rep.*, 2017, **7**, 1–9.
414 M. A. Kayala and P. F. Baldi, *A Machine Learning Approach to Predict Chemical Reactions*, Curran Associates, Inc., Granada, 2011.
415 J. Behler, *Angew. Chem., Int. Ed.*, 2017, **56**, 12828–12840.
416 S. Chmiela, A. Tkatchenko, H. E. Sauceda and I. Poltavsky, *Sci. Adv.*, 2017, **3**, e1603015.

