ACS Catalysis

pubs.acs.org/acscatalysis

Perspective

Automated in Silico Design of Homogeneous Catalysts

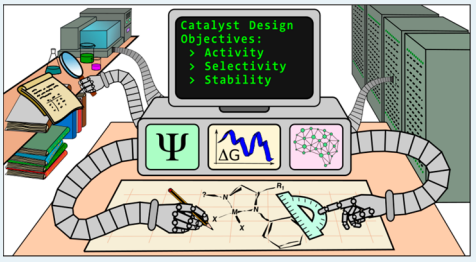
Marco Foscato* and Vidar R. Jensen*

Cite This: *ACS Catal.* 2020, 10, 2354−2377

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Catalyst discovery is increasingly relying on computational chemistry, and many of the computational tools are currently being automated. The state of this automation and the degree to which it may contribute to speeding up development of catalysts are the subject of this Perspective. We also consider the main challenges associated with automated catalyst design, in particular the generation of promising and chemically realistic candidates, the tradeoff between accuracy and cost in estimating the catalytic performance, the opportunities associated with automated generation and use of large amounts of data, and even how to define the objectives of catalyst design. Throughout the Perspective, we take a cross-disciplinary approach and evaluate the potential of methods and experiences from fields other than homogeneous catalysis. Finally, we provide an overview of software packages available for automated in silico design of homogeneous catalysts.



KEYWORDS: *automation, virtual screening, de novo design, high-throughput screening, inverse design, synthetic accessibility, machine learning, multiobjective*

1. INTRODUCTION

Catalysts make chemical transformations both faster and more selective, advantages that are vital for the sustainable production of energy, materials, and bioactive compounds.[1] The numerous important applications of catalysis have propelled rational catalyst design to becoming a "Holy Grail" of computational chemistry.[2]

Indeed, computational tools have taken on important roles in homogeneous catalysis, thanks to ever-increasing computer power and molecular modeling methods that balance cost and accuracy.[3−5] The computational tools complement the experimental tools by helping to interpret experimental results, by guiding experiments, and by predicting properties such as catalytic activity and selectivity. As illustrated in Figure 1, the predictive strategies for catalyst design may be divided into three categories: (i) manual or interactive trial and error, (ii) the use of prediction models, and (iii) automated design.

The first category pertains to the "everyday" interactive use of computational tools to test ideas and chemical intuition. Chemists of all sorts, not only the trained computational chemists, are using molecular-level computational tools in this straightforward fashion to nurture their creativity and thinking to solve problems in catalysis. Even simple visualization of three-dimensional (3D) molecular models, which now easily can be rendered also by virtual reality or even coupled with real-time simulations,[6] can provide valuable insights for catalyst design. An example is how 3D molecular models may be enhanced with measurements such as the volume and shape of the catalytic site.[7] At the more computationally demanding end of the spectrum, calculation of free energy profiles along the reaction pathways has become common practice,[8] even in light of the challenging tradeoff between the computational cost, which may be substantial for a multistep reaction catalyzed by a transition-metal complex, and desirable accuracy.[9−12] Outstanding examples of interactive catalyst design have been reviewed recently.[4] Although too few predictions are followed up by experimental verification (see refs 13 and 14 for excellent examples), the results are promising.[4]

The second category of predictive computational catalyst design, the use of prediction models, involves quantitative or qualitative models derived from statistical data analysis. Quantitative structure−activity/property relationships (QSAR/QSPR) are prime examples of such models that correlate a set of descriptors with desired properties such as catalytic activity and selectivity. Once established, the correlation can be used to quickly estimate the properties of novel compounds that are not too different from those of the data set used to build the model. In other words, the model is associated with a region of chemical space, its applicability domain, outside of which it is unreliable. Such predictive models have helped interpret experimental trends and have also been used in catalyst design.[15−28]

Received: November 15, 2019
Revised: January 17, 2020
Published: January 21, 2020

https://dx.doi.org/10.1021/acscatal.9b04952
ACS Catal. 2020, 10, 2354−2377

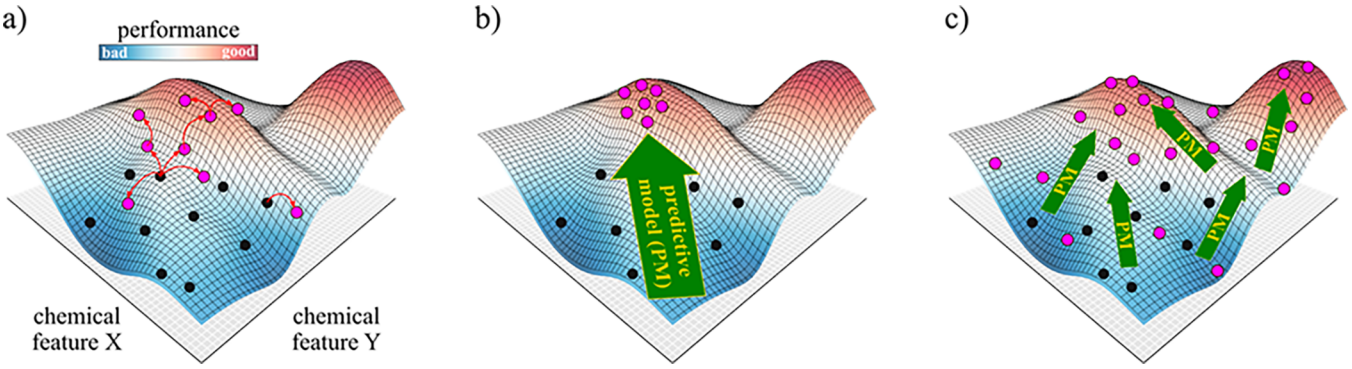


Figure 1. Three categories of computational catalyst design and how they navigate the performance landscape, here sketched as a surface resulting from the combination of two chemical features X and Y: (a) manual, trial and error based design in the vicinity of known catalysts (black points), with red arrows indicating the individual steps taken to new candidate catalysts (magenta points); (b) design based on prediction models (PMs) exploiting statistical analysis of data from known catalysts and/or calculations to indicate which direction in chemical space to follow (arrow) and to guide candidate selection; (c) automated design, which may also exploit predictive models, aiming for a more thorough exploration of the performance landscape, including the possibility to discover distant optima.

The third category, automated design, pertains to the automation of the many computational tasks associated with the identification of candidate catalysts with desired properties. This category includes the use of prediction models, albeit in an automated fashion, and importantly, automated generation of candidate molecules. We clarify right from the start that this automation does not, and probably never will, imply "black-box" use of computational techniques. Rather, it implies addressing the challenges of in silico catalyst design systematically, objectively, and automatically to maximize predictive power.

Some of the many challenges of catalyst design are intrinsic to the very nature of catalysis: a catalyst flattens the potential energy surface (PES), which thus becomes more susceptible to perturbations by factors such as solvents or additives.[29] However, these factors are often ignored by the necessarily rather approximate prediction models. Additional challenges result from the often large conformational, configurational, and reactivity landscapes, as well as from the often complex electronic structure of catalysts and their intermediates and transition states.[12,30]

Given these challenges, designing a catalyst from scratch by first principles is a formidable task that is seldomly approached. In contrast, the prediction of relative reactivity or selectivity within a relatively confined structural domain is more manageable and fruitful.[3] In addition, with confidence in the predictive methodology follows the desire to apply it systematically. The motivation behind automation is to benefit from such systematic applications without exhausting the available human resources.

While the creative and intellectual tasks are left to humans, automation may take care of the monotonous, tedious, and error-prone tasks[31] of a systematic study.[32] Moreover, via automation the available computational power and the ever-growing chemical knowledge may be exploited to an extent that is beyond human capabilities. Machines are faster, more precise, objective, and memory-rich than humans. Perhaps the most exciting of all the opportunities offered by automation is that the bias introduced by the chemist's preconceptions may be removed. This detachment could allow molecular design to go beyond our traditional and self-imposed limitations, which, in mathematical terms, can be seen as local minima instead of the global minimum representing the optimal catalyst.[4,29]

A broad range of automated techniques, with various degrees of automation, are already among the tools routinely used in catalyst design. Other techniques are blooming in closely related fields, such as in the automated exploration and mapping of reaction networks,[6,33−40] the identification of possible geometries for a given chemical composition,[41−43] extraction,[44,45] and the management of chemical information and computational or experimental data.[46,47] Nevertheless, automated molecular design rests on two main pillars: (i) workflows for prediction of molecular properties and (ii) autonomous generation of candidates: i.e., routines that build molecular structures and navigate the chemical space to regions populated by candidates displaying the desired properties.

This Perspective focuses on these pillars and on how they are shaping the development of automated in silico design methods for homogeneous catalysts. Less attention is here given to the wider range of recently reviewed[48] computational methods that contribute to the design of small-molecule catalysts by providing mechanistic insight,[4] molecular descriptors,[27] and predictive models.[26,49] Apart from inverse design,[50,51] methods for automated design have not been reviewed. Many of these methods originate from fields other than catalysis, and they have yet to be collected and compared in a single account.

For these reasons, throughout this Perspective, we take a cross-disciplinary view and describe valuable methods and approaches developed in closely related fields, such as the design of drugs, proteins, materials, and heterogeneous catalysts, that have yet to have an effect in homogeneous catalysis. We start by presenting the general design strategies and describe recent advances in methods and applications. Next, we focus on four challenges in automated in silico design: (i) the generation of realistic and novel candidates, (ii) the prediction of their properties, (iii) the definition of the objectives in catalyst design, and (iv) the management of the data generated by automated workflows. Finally, we list currently available software packages developed for automated in silico design of catalysts.

2. AUTOMATED MOLECULAR DESIGN STRATEGIES

Like any molecular design problem, catalyst design is a nonlinear optimization problem.[52] This means that changes in the properties, such as activity and selectivity, do not correlate linearly with changes in the catalytic system. The latter is defined in terms of parameters that specify the atomic composition (i.e.,

https://dx.doi.org/10.1021/acscatal.9b04952
ACS Catal. 2020, 10, 2354−2377

the number and identity of the nuclei and the number of electrons) and the relative position of the nuclei (i.e., the constitution and the stereochemistry). The catalytic system (*CS*) is connected to the catalytic performance (*P*) via a causality relation, as shown in Figure 2.

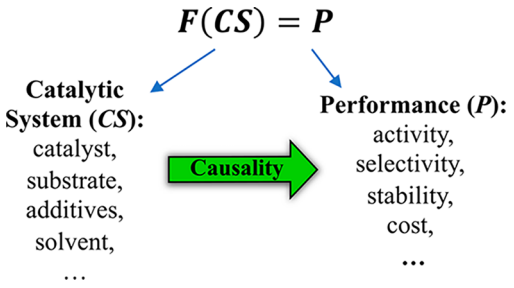


Figure 2. Graphical representation of the relation between the catalytic performance (*P*) and the catalytic system (*CS*). The *forward* operator *F* represents the generally unknown mathematical relation between *CS* and *P*.

Numerous parameters, such as those that determine the catalyst itself (e.g., the catalyst molecule in homogeneous catalysis), the substrates, the solvents, and the potential additives, are needed to completely define the catalytic system *CS*. In combination with unconstrained degrees of freedom, these parameters lead to combinatorial explosion. Thus, to obtain a tractable optimization problem, the search space is usually restricted so as to limit the degrees of freedom. In practical catalyst design, the search space usually spans only the molecular catalyst and the substrates and rarely includes additional reaction conditions such as the solvent or additives.[4] This pruning of the parameters is the first important decision in any molecular design project. This decision affects not only the complexity and the computational cost of the design problem but also how correct and useful the outcome will be. Ideally, the search space should be defined dynamically, as knowledge acquired during the design process may turn out to be relevant for determining the search space.

With the catalytic system *CS* and the search space defined, the catalytic performance *P* is given as the causal relation represented by the *forward* operator (*F* in Figure 2). This operator is typically unknown, unless the property of interest is derivable from the expectation values of quantum mechanical operators,[53] which is seldom the case in catalyst design problems. However, since the goal of molecular design is to identify the parameters (i.e., the catalytic system *CS*) that give the best performance *P*, the direction of the causality relation in Figure 2 can be used to define two molecular design strategies: *direct* and *inverse* design.

The *direct* design strategies exploit the causality relation going from the parameters to the resulting performance. These strategies use an approximate operator *F* to estimate the performance resulting from the parameters of the candidate catalysts. The latter are modified iteratively in processes mimicking the traditional "guess and check" approach to experimental catalyst development. The iterative search for the optimal catalysts usually follows heuristic techniques in *direct* design.

The *inverse* design strategies start from the optimal performance and then aim to obtain (ideal) parameters for the chemical system to reflect that performance,[54] thus inverting the causality relation defined by operator *F*. However, in general, *F* cannot be inverted.[55] Thus, the term "inverse design" is commonly applied to design strategies which add constraints that make *F* locally invertible or that include some performance-driven feedback that informs the construction of candidate catalysts and their parameters. Accordingly, performance-driven high-throughput screening as well as evolutionary-driven global optimization are often described as inverse design techniques,[54] even though these methods do not involve actual inversion of *F* and candidates are evaluated in a direct fashion.

The design techniques covered in this Perspective are organized and described below as falling into the *direct* or *indirect* category depending on which of these two philosophies the original developers intended to follow.

2.1. Direct Design. The starting point of any direct design is to create an approximate expression for the operator *F*. Such an approximation can be derived from experimentally observed trends or from a hypothesis as to the reaction mechanism based on studies of one, or a few, catalytic systems. For this reason, direct strategies usually try to refine catalysts in the vicinity of known and closely related chemical systems. Hence, structural modifications are assumed to affect the catalytic properties, which thus can be optimized, without violating the underlying assumptions, such as that of a constant reaction mechanism.[52]

2.1.1. Virtual Screening. The iterative "guess and check" nature of direct design is readily exploited in automation. In its simplest implementation, virtual screening,[56] a list of candidates are subjected to the same computational protocol to estimate their performance, often termed "scoring function", "fitness function", or "figure of merit".[57] Thus, the automated prediction workflow uses the chemical definition of the candidate and performs the calculations, some of which may be launched and managed on remote computers, needed to obtain the figure of merit. The accuracy and computational cost of the figure of merit largely determines the feasibility of the direct design.

Whereas an early example of virtual screening in molecular inorganic chemistry involved the identification, using the software package HostDesigner, of binding sites for targeted metal ions,[58] a prime illustration of the role of the figure of merit in direct catalyst design is the recent report by Munday, Wiest, Norrby, and co-workers[59] of phosphine ligands for rhodium-catalyzed asymmetric hydrogenation of enamines. Central to this screening was the fast calculation of sufficiently accurate figure of merit values via the Quantum-Guided Molecular Mechanics method (Q2MM)[60,61] for modeling the selectivity-determining transition state (TS).[62,63] Using this dedicated force field, extensive conformational searches at the diastereoisomeric TS structures could be performed, thus producing a set of conformers for each diastereoisomeric pathway and for each combination of ligand and substrate (Figure 3). The stereoselectivity was then calculated from the Boltzmann-averaged energy of the conformational ensemble. Validation of the results for two different substrates showed that, despite a suboptimal correlation between the predicted and experimental enantiomeric excess, computationally predicted ligands were experimentally verified to induce the desired selectivity, giving enantiomeric excesses above 96%.

A similar coupling of virtual screening with automated TS modeling has been reported by Wheeler and co-workers,[64−66] who have developed an automation toolkit (AARON; see section 7 for details on software packages for design) for computational protocols involving TS modeling.[67] In general, these protocols involve the construction of a TS structure guess, a conformational search, preoptimization, geometry optimiza-

https://dx.doi.org/10.1021/acscatal.9b04952
ACS Catal. 2020, 10, 2354−2377

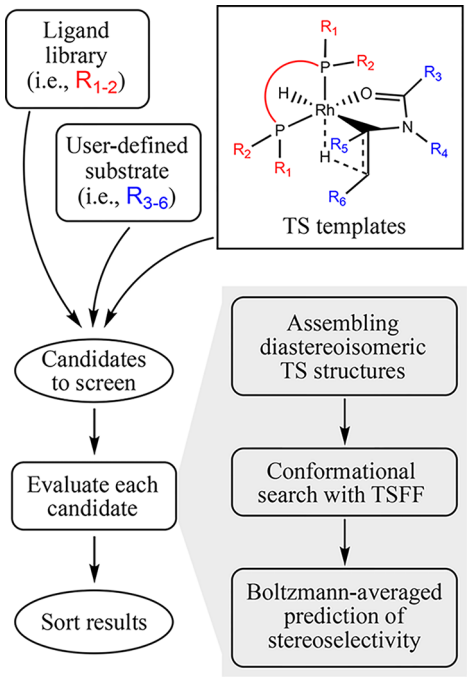


Figure 3. Flow chart for automated virtual screening of phosphine ligands for rhodium-catalyzed asymmetric hydrogenation of enamines.[59] TSFF: transition state force field.

tion with density functional theory (DFT), and final parsing and processing of the computational data. This strategy was first applied to design bipyridine *N*,*N*′-dioxide organocatalysts for asymmetric allylation and propargylation of benzaldehyde[65,66] and, shortly afterward, also to transition-metal-mediated reactions such as that of rhodium-catalyzed hydrogenation[64] and palladium-catalyzed Heck allenylation.[67]

Despite these successes, virtual screening remains a trial and error approach that is unable to navigate the search space on its own. This means that the automation is limited to looping over and evaluating a list of predefined candidates. In other words, these methods do not suggest new candidates or prioritize particularly promising regions of the search space. The latter tasks are left to the user, the chemist, who via analysis of ranked candidates from one screening may adjust the search space and launch a new, modified screening.

2.1.2. De Novo Design. A higher degree of automation can be achieved by coupling the ability to predict the performance of a candidate (i.e., the evaluation) with the ability to traverse the chemical space to optimize the candidates.[68] In order to overcome the limits of predefined libraries of candidates, new candidates are generated from scratch, in so-called de novo design,[69,70] under the guidance of global optimization algorithms. Thus, in comparison to the above virtual screening methods, de novo methods must have additional capabilities, as summarized by one of the leading developers of such methods for drug design, Gisbert Schneider:[71] "*Basically, three questions have to be addressed by a de novo design program: how to assemble the candidate compounds; how to evaluate their potential quality; and how to sample the search space effectively.*" Given the overwhelming size of the unrestricted search space,[72] Schneider's third question, the need for sampling the search space, is not addressed by searching systematically for the absolute optimum. Instead, heuristic algorithms are used to identify good candidates at a reasonable computational cost. In de novo drug design, the leading application area of automated molecular design, a variety of such optimization algorithms have been used, including evolutionary algorithms,[73] particle swarm optimization,[74,75] ant colony optimization,[76,77] and simulated annealing.[78,79]

These experiences from de novo drug design were exploited in the development of an evolutionary algorithm for the optimization of homogeneous ruthenium-based catalysts for olefin metathesis.[80] Candidate catalysts were built by connecting molecular fragments to metal-coordinating building blocks that were used to alter the properties of the metathesis-active species, the ruthenium alkylidene. Each such combination of building blocks represented the genetic material, the chromosome, of a single candidate catalyst (see Figure 4), which allowed the developers to simulate catalyst evolution following the principle of survival of the fittest. Catalysts were created from scratch or modified by mutation (modification of a single fragment) and crossover (swapping of fragments between members of the population to generate new candidates). The best-performing candidates were given high mutation and crossover probabilities so as to transmit their properties to the next generations. During the simulated evolutions, the performance of the catalyst population improved. Importantly for the validation of the method, this improvement reflected the historical transition from the so-called first-generation Grubbs catalysts (coordinated by phosphine ligands)[81,82] to the second-generation catalysts (coordinated by N-heterocyclic carbene ligands).[83,84] Although these de novo experiments suggested candidates with improved catalytic performance in comparison to the best existing catalysts at the time, the lists of optimized catalysts also demonstrated a common problem in de novo design: the poor synthetic accessibility of the automatically designed candidates, which was later addressed by controlling the kinds of bonds that are allowed to form in the automated building process.[85]

Whereas the above design of ruthenium-based olefin metathesis catalysts is an example of artificial evolution, the actual development of biocatalysts has been heavily inspired by evolutionary principles and directed evolution is extensively used in experimental catalyst design.[86,87] The same inspiration has, perhaps surprisingly, not to the same degree influenced the corresponding automated in silico design of biocatalysts, which, instead, is dominated by virtual screening.[88] Automated screening of peptide mutations was initially explored already in the 1990s.[89,90] Meanwhile, virtual screening of the biocatalytic activity of enzymes subjected to a few mutations has become reality.[91−94] Promising results have been obtained also using combinatorial backbone assembly,[95] a strategy that predominantly alters the remote parts of the enzyme rather than the active site itself, thus creating structural diversity while still retaining fundamental catalytic activity. In contrast, more conservative strategies have been followed when biocatalysts have been designed from scratch (i.e., de novo). Specifically, idealized active sites (i.e., "*theozymes*", as in theoretical enzymes)[96] have been fitted into suitable protein backbones.[97−100]

In one pioneering example of this strategy, Baker and coworkers[99] developed a method combining (i) identification of suitable protein sites capable of hosting a prebuilt TS model and (ii) optimization of TS-stabilizing interactions. These are great challenges,[101] but their method could design de novo biocatalysts for reactions not catalyzed by any natural enzyme,[102,103] an impressive achievement. However, the activities observed experimentally were low and the enzymes

https://dx.doi.org/10.1021/acscatal.9b04952
ACS Catal. 2020, 10, 2354−2377

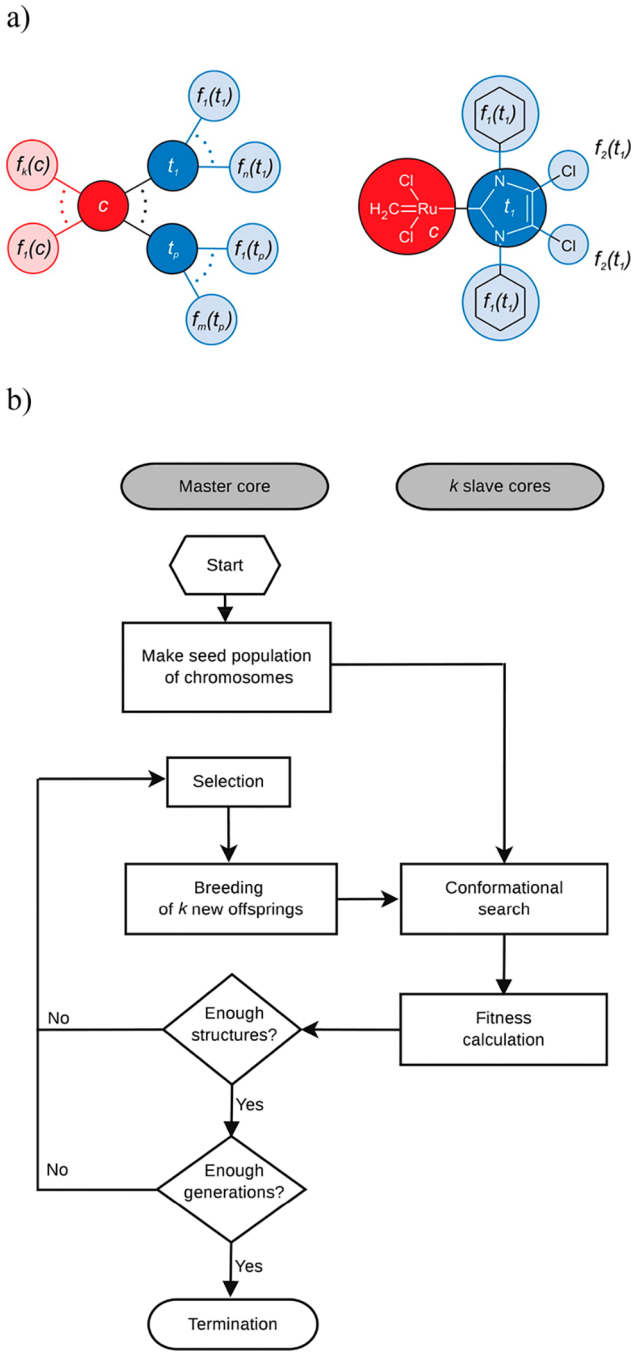


Figure 4. (a) Graph-based chromosome representing a catalyst as a collection of fragments. In the first implementation of this de novo method,[80] three kinds of fragments with different variability were used to constrain the search space: core (*c*, typically a fixed metal fragment), trial (*t*, typically one of a few possible ligand frameworks), and unconstrained (*f*, typically freely varying substituents). (b) The overall workflow of the de novo evolutionary algorithm deployed for the automated design of ruthenium-based catalysts for olefin metathesis. Reprinted with permission from ref 80. Copyright 2012 American Chemical Society.

required further refinement via directed evolution,[104−106] which can be seen as an in vitro optimization technique.[86] If the experimental directed evolution could instead be performed in silico,[107] the resulting overall method would be capable of designing useful biocatalysts from scratch in a fully automated fashion.

However, even without the additional in silico directed evolution step, the examples of de novo design of biocatalysts are remarkable given the complexity of these catalysts in comparison to the small-molecule catalysts. Presumably, this success partially originates from the modularity of the biocatalysts, which consist of chains of a limited number of building blocks, the amino acids. This modularity appears to have been chosen as Nature's preferred strategy in constructing a variety of complex systems, such as biocatalysts, with tunable functionalities at the same time as retaining practically identical synthetic pathways. Ensuring synthetic feasibility is, in fact, a major challenge in automated catalyst design, one that is addressed in section 3.

2.2. Inverse Design. Three main kinds of techniques are used in inverse design:[50,51] gradient-driven methods, alchemical transformations, and generative models.

2.2.1. Gradient-Driven Methods. These methods exploit a formulation of the design problem that expresses the figure of merit as a gradient calculated over the parameters defining the chemical composition.[50] This gradient is then used to guide the generation of new parameters to maximize the performance. An intuitive example is given by the gradient-driven molecular construction method (GdMC) proposed by Weymuth and Reiher.[108] In this method, catalytic performance is assumed to originate from certain idealized structural features, such as those of an optimal, local transition-state geometry for the catalyzed reaction of interest. This local fragment is not stable on its own, however, and is associated with internal forces (gradients). These gradients can be removed by a surrounding environment, a so-called "jacket" potential,[108] which counterbalances the forces and stabilizes the ideal, local fragment. After a proof of concept application to the design of N_2 fixation catalysts,[108] the GdMC method was coupled with a fully automated shell-wise construction algorithm and used to retrace the design of an experimentally known ruthenium-based catalyst for CO_2 activation (Figure 5).[109]

The above idea that a catalyst can be seen as a properly tuned chemical environment is a popular concept in enzyme catalysis.[110,111] Thus, other inverse design methods have also recently been developed to tune the environment or, more precisely, to optimize a simplified representation of the catalytic environment, such as a distribution of point charges that reduces the barrier of a desired reaction.[112] The question, however, is how to convert such a simplified surrounding into a chemical structure consisting of discrete atoms, a molecule that can be synthesized and tested experimentally.

Achieving this conversion is the perhaps greatest challenge in inverse design. Molecules and materials consist of discrete objects, atoms. An atom is either present or not, and its nuclear charge must be an integer. In contrast, optimization algorithms are more effective for continuous quantities, for which they take advantage of first and second derivatives. Thus, the discrete nature of chemical objects must be "smoothed out" in order to navigate the chemical space while following the property of interest.[113]

2.2.2. Alchemical Transformations. One smoothing technique is to use the coefficients of linear combinations of atomic potentials (LCAP) as the continuous parameters. Once the optimal coefficients are reached, they may be rounded to the nearest integer (0 or 1) to obtain an actual molecular representation.[113,114] Prior to rounding, the noninteger coefficients can be said to represent "alchemical" molecules: i.e., unphysical and experimentally inaccessible blends of atoms or groups. A discrete structural change from one such atom to

https://dx.doi.org/10.1021/acscatal.9b04952
ACS Catal. 2020, 10, 2354−2377

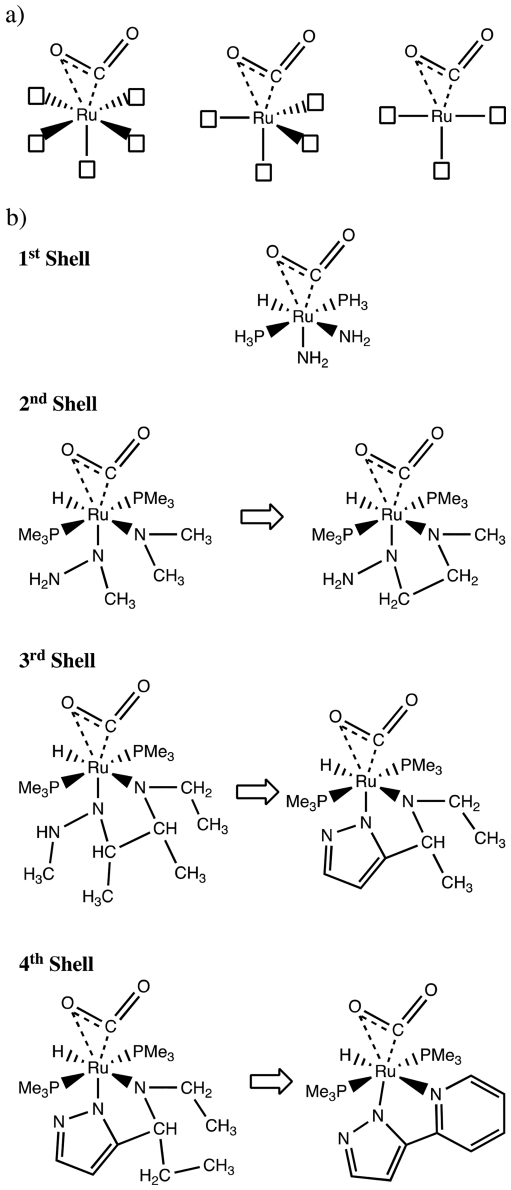


Figure 5. Exemplified gradient-driven molecule construction.[109] (a) Designs of local, idealized fragment for CO_2 activation. Squares represent open coordination sites. (b) Shellwise molecular construction and topology adaptation. Each of the illustrated structural changes reduced the atomic gradients of the idealized starting fragment.

another may, however, be depicted as a continuous path amenable to optimization.[115] For this purpose, the LCAP method was coupled with a gradient-directed Monte Carlo method that combines gradient-driven optimization with random changes that allow overcoming local barriers to reach the global optimum.[115] Recently, LCAP-based methods were used to screen synthetically viable modifications of a known catalyst, Ni^{II}-iminothiolate, for oxidation of CO to CO_2.[116]

The above transformation of unphysical alchemical objects into sensible molecules is a challenge that has been addressed also in other inverse design methods. In particular, alchemical potentials have been developed to gauge the tendency of a system to *transmutate* a given atom: i.e., to change its number of protons and electrons.[117] Central to this concept is the realization that molecular properties can be written as a functional of the proton distribution $Z(\boldsymbol{r})$ and a function of the total number of electrons N_e.[55] Considering, for example, the total electronic energy (E) as the observable of interest, the derivative of E with respect to the proton distribution $Z(\boldsymbol{r})$ is defined as the nuclear chemical potential.[55] At the position of the nuclei the nuclear chemical potential is referred to as the "alchemical" potential because it measures the tendency for each atom in the molecule to mutate its number of protons.

To make this problem amenable to optimization, a penalty function, p, has been considered as the difference between the target observable and the value of the observable obtained by a given combination of N_e and $Z(\boldsymbol{r})$: that is, a molecular system defined in terms of the nuclear charge distribution and the total number of electrons.[118] The inverse design problem thus consists of minimizing p while varying N_e and $Z(\boldsymbol{r})$,[55] the first- and second-order derivatives of which greatly improve the efficiency of the optimization. The price to pay for this efficiency is, as described above, the occurrence noninteger, "alchemical" nuclear charges. After the optimization, N_e and $Z(\boldsymbol{r})$ are therefore rounded to discrete values to give a valid chemical system.

Alchemical potentials and their derivatives offer great promise in rational and ab initio de novo design. However, the applications of these methods are so far limited to a few specific chemical systems. Proof of concept applications include the design of nonpeptidic anticancer drugs,[117] BN-doped benzene derivatives with tuned highest occupied molecular orbital (HOMO) eigenvalues,[119] and the prediction of simple energy barriers, such as that of the umbrella flipping of ammonia.[120] The alchemical derivatives could also identify where protons should be annihilated or created to enhance the interaction energy of formic acid with small, 10-proton molecules, predicting that CH_4 should be mutated to HF (a process going via H_2O and NH_3) to increase the interaction energy (Figure 6).[121]

The most recent development in the field of alchemical transformations in inverse design has been the introduction of alchemical normal modes. For an initial reference system, these modes are the eigenfunctions of a unified Hessian matrix involving second-order derivatives of the electronic energy with respect to nuclear positions, number of electrons, and number of

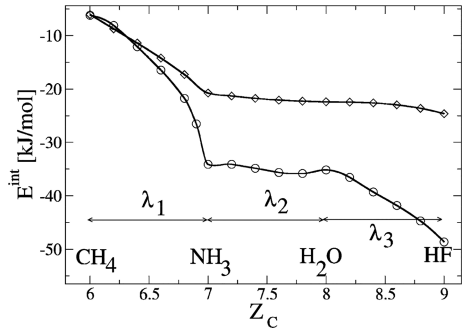


Figure 6. Potential energy of interaction (E^{int}) between a 10-proton system and formic acid along alchemical paths (λ_n) that vary the 10-proton system from CH_4 (Z_C = 6) to HF (Z_C = 9) by gradually increasing the atomic number of the central atom (Z_C) while successively decreasing (from 1.0 to 0.0) the atomic number of three neighboring hydrogen atoms. The diamonds correspond to interaction energies obtained with a frozen geometry, while the circles reflect values obtained by continuously relaxing the 10-proton system. Reprinted with permission from ref 121. Copyright 2007 American Chemical Society.

https://dx.doi.org/10.1021/acscatal.9b04952
ACS Catal. 2020, 10, 2354−2377

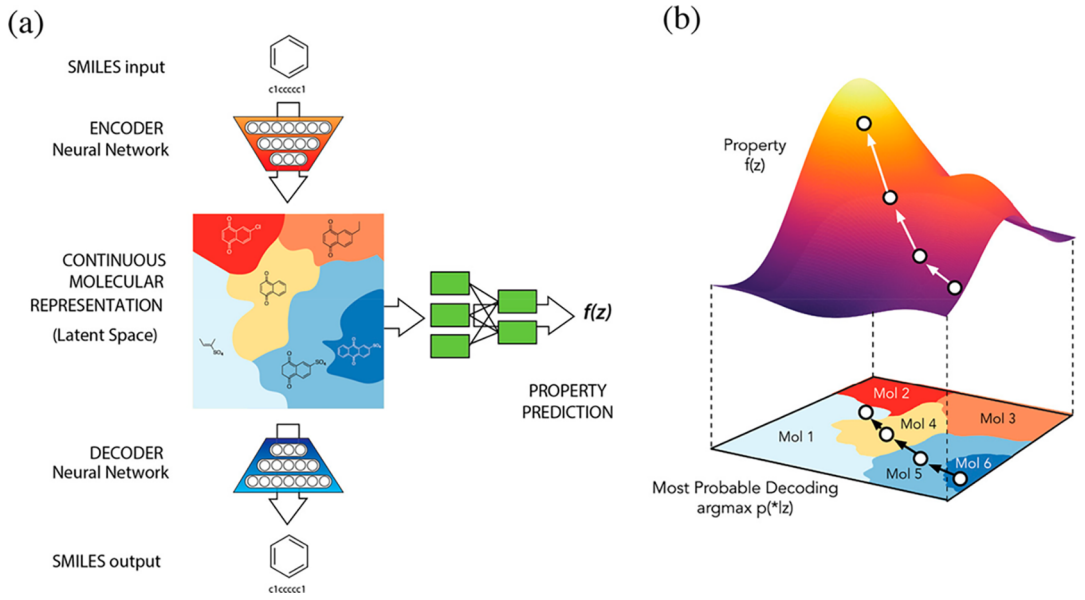


Figure 7. (a) The encoder, the latent space, the prediction model, and the decoder for automated molecular design.[133] The encoder converts the discrete SMILES-string representation into a continuous molecular representation (the latent space). The prediction model estimates the property of interest from the latent-space representation. The decoder converts the latent-space representation into a discrete SMILES string. (b) Gradient-based optimization in continuous latent space. Reprinted with permission from ref 133. Copyright 2018 American Chemical Society.

protons.[122] Thus, these modes indicate the changes in energy resulting from changes in geometry and atom identity. An analysis of these modes has been used to estimate electronic ground-state energy changes in nearly two million of B- and N-doped coronenes with encouraging accuracy, considering the negligible computational cost of these predictions in comparison to the obvious alternative: virtual screening (using standard DFT) of all the doped candidates.[122]

Although alchemical methodologies have not seen many applications in homogeneous catalysis, promising results have been achieved in heterogeneous catalysis and materials design.[123−125] In particular, linear extrapolations based on alchemical derivatives have been used to estimate the catalytic activity of palladium nanoparticles for oxygen reduction.[120] Importantly, as also seen in the examples described above, the computational cost of screening isoelectronic alchemical changes, in this case consisting of complementary changes of the identity of atoms in one or more atom pairs in the cluster, is negligible. This computational efficiency originates from the minimal cost of calculating the alchemical derivatives once an initial binding energy of oxygen with a reference palladium nanoparticle has been evaluated. The alchemical derivates can then be used to obtain fast estimates of variations in oxygen binding energy, and thus oxygen reduction, with reasonable accuracy for modified nanoparticles. The method has also been applied to other materials.[123,124]

Despite these promising results, the picture emerging from the applications of alchemical methods in inverse design so far also underlines the nonlinearity of most properties with respect to the alchemical changes.[126] The extent to which linear extrapolations based on alchemical derivatives can be used is thus limited:[127] for example, to cases where interpolation between reference compounds can be exploited.[118] As a result of these limitations, most of the applications reported so far have started from pre-existing scaffolds and have involved heavily restricted search spaces,[108,114,116,128] sometimes formally containing many compounds but having limited chemical variability.[122,129]

2.2.3. Generative Models. While alchemical methodologies start from first principles (ab initio), machine learning takes an empirical approach to inverse design. More specifically, these approaches use experimental and computed data to extract empirical rules representing either the operator F (Figure 2) or its inverted form via machine-learning models. The currently very active field of machine learning is dominated by methods for classification and correlation.[50] Most of these methods predict properties (see section 4.1.1). Instead, here we address models that generate chemical entities rather than, or in addition to, evaluate them. These so-called generative models are among the latest developments absorbed in the design of small organic drugs and aim to propose candidates without having to rely on the complex, often hard-coded, rules that otherwise must be used to restrict the generation to sensible molecules only. Although many such applications have been reported,[32,130−132] we restrict the description below to two particularly illustrative examples.

The first example combines the conversion of discrete molecular representations to and from a multidimensional continuous representation with property prediction (Figure 7).[133] The structural encoding, decoding, and property prediction are handled by models trained by neural networks (NN). The NN-trained encoding maps a string-based chemical representation (simplified molecular-input line-entry system, SMILES) into a continuous latent, vector space. As pointed out above, the continuity of the space allows for gradient-based optimization of the property of interest, which in this case is predicted from the latent-space representation by a second NN-trained model. The most promising points in the latent space are then decoded to a discrete molecular representation by a third NN model. For the optimization to work, all points in the latent space must correspond to valid molecular candidates. However, this still represents a substantial challenge, and although promising developments have been reported recently,[32,134,135]

https://dx.doi.org/10.1021/acscatal.9b04952
ACS Catal. 2020, 10, 2354−2377

including the use of semantically constrained graphs,[136] filtering of invalid candidates is often necessary.[133] Still, this contribution shows that, when enough data are available, encoders, decoders, and predictors can indeed be trained to generate candidates reflecting the property of interest, such as drug-likeness and, at the same time, ease of synthesis.[133]

Another approach, termed Reinforcement Learning for Structural Evolution (ReLeaSE) was recently developed to bypass the gradient-driven optimization. Using deep neural networks, both generative and predictive models were built.[137] The generative model produces chemically feasible molecules as SMILES, while the predictive model estimates the property of interest directly from the SMILES representation. The predicted property is used to assign a reward (or penalty) to the generated molecule, and the generative model is biased to maximize the expected reward.

Training of generative models based on machine learning requires large data sets. The examples of such models have so far been limited to design of organic, mostly druglike molecules. This is not surprising, given the amount of curated data on organic and pharmaceutically relevant compounds in comparison to, for instance, data on transition-metal catalysts. Surely, the machine-learning-based generative models should be able to produce reasonable organocatalysts, but it remains to be seen whether these tools can, for example, produce ligands for transition-metal catalysts.

3. SYNTHETIC ACCESSIBILITY: NEW VERSUS OLD DESIGNS

A new and good catalyst does not necessarily have to be based on a previously unknown compound. In fact, from an economical and practical point of view, repurposing an existing, "old", design is a better strategy, that might even include benefiting from a known and possibly cheap synthetic route to prepare the catalyst. In comparison, "new" designs, that is, unknown and not-yet-prepared compounds, may pose serious synthetic challenges that preclude their practical use. Still, the chemical space is vast and only a tiny fraction of the potentially accessible compounds have so far been made,[138] which means that effective candidates are likely to be missed if the search space contains only existing compounds. Whether or not to allow for new designs is an important decision to take in any molecular design project.

If new designs are welcome, evaluation of their synthetic accessibility and complexity, such as the number of steps required for their preparation,[139] allows for exclusion of candidates deemed to be inaccessible and for ranking the remaining ones to help select molecules for experimental followup. Therefore, measures of the synthetic accessibility become an integral part of each candidate's performance, which, as discussed in section 5, may blur or complicate the design objectives.

Computational evaluation of synthetic accessibility is a well-known challenge in drug design[140] and a key reason the screening of existing compounds is often preferred over de novo drug design.[68] To overcome the challenge and to improve the de novo methods, synthetic accessibility scores have been developed for organic molecules. Such scores may be based on measures of molecular complexity, such as the presence of rings and stereochemical features,[141−145] or on retrosynthetic analysis.[146] However, since these methods were trained on organic, druglike chemistry, little is known about their performance on, for example, organocatalysts and ligands for transition-metal compounds.

As an alternative to calculating synthetic accessibility scores, reaction-driven de novo design has been developed to only generate candidates that can, in principle, be formed by combinations of known synthetic reactions using commercially available reactants.[147−149] Thus, a synthesis route is proposed along with each new candidate.[150] However, these approaches are also best suited for standard organic chemistry and have not seen much adaptation to transition-metal and organometallic chemistry. Moreover, in contrast to the case for pharmaceuticals, where challenging synthetic pathways may be justified by the value of the final product, simple yet specific synthetic pathways are often preferred for homogeneous catalysts, for which profit margins may require catalyst recycling.

Moreover, the synthetic accessibility of a catalyst can also be interpreted as the ease with which the catalytically active species, rather than the precursor, can be provided. The reactivity of the active species requires compatibility with the functional groups of the catalytic system. This is particularly true for transition-metal catalysts, which are often incompatible with even standard, frequently occurring functional groups. To reduce the likelihood of such incompatibilities, ligands for transition-metal catalysts typically contain few functional groups, and often the only functional groups present are those, such as amines and imines, that coordinate the central metal atom. In other words, the ligand substituents are mostly carbon-based skeletons and the few functional groups present are there for a reason, typically to induce a specific electronic effect or to enhance solubility. Accordingly, search spaces are often defined as combinations of metal-coordinating groups, backbone/bridging fragments, and inert substituents (Figure 8).[151−153] The assumption behind this strategy is that the synthesis is largely modular, so that the same synthetic pathway can be applied to reactants with different carbon-based side chains.

As pointed out in section 2.1.2, modularity is a prime feature of biocatalysts, in which versatile side chains are held together by a backbone built by reiterating the same synthetic step. Even if enzymes are huge and much more complex molecules in comparison to small-molecule catalysts, their modularity allows reuse of the same biosynthetic machinery. Synthetic accessibility is thus much less of an issue than for transition-metal catalysts and ligands. While, as shown in Figure 8, modularity is a tailored feature of transition-metal ligands, this modularity is usually limited to a single class of ligand and relies on varying the reactants while preserving the synthetic pathway.[154−158] In contrast, successful attempts to exploit the modular structure of biopolymers, i.e., a constant backbone decorated by varying side chains, have led to an in vitro synthesized library of DNA-based organocatalysts for hydration of α,β-unsaturated ketones.[159] This promising approach provides catalyst variability while retaining synthetic accessibility in a modular framework. Combinations of this kind of modular synthesis with automated in silico design are still unexplored in homogeneous catalysis.

4. PREDICTION OF CATALYTIC PERFORMANCE

Predicting catalytic performance typically involves some kind of molecular modeling, to obtain energies or other molecular properties, followed by an actual prediction step, to estimate the catalytic performance on the basis of the calculated properties. In the following subsections we will briefly review the two categories of method involved in assessing the catalytic performance.

https://dx.doi.org/10.1021/acscatal.9b04952
ACS Catal. 2020, 10, 2354−2377

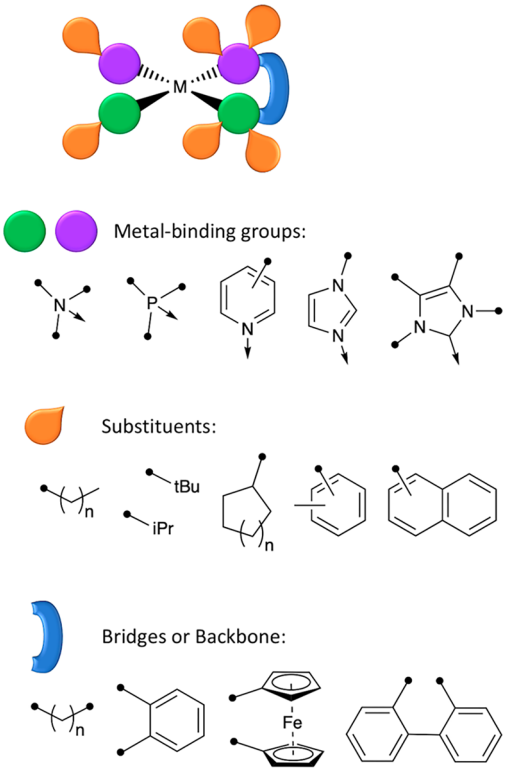


Figure 8. Commonly used definition of transition metal ligands as combinations of metal-coordinating groups, backbone/bridging fragments, and inert substituents.

4.1. Performance-Prediction Models. A prediction model is a mathematical construct that, for a given chemical representation of a candidate catalyst, estimates its performance and thus is an approximated implementation of operator F in Figure 2. Such prediction models come in many forms, but the most popular ones are collectively referred to as machine-learning models. In addition to giving an overview of the applications of the latter in catalysis, we also briefly cover the recent developments in exploiting linear free energy regression models in the design of catalysts.

4.1.1. Machine Learning and Statistical Methods. Machine learning (ML) is a family of data-driven statistical methods implementing artificial intelligence and includes models varying from single- and multivariate regression to so-called deep learning.[160,161] In order to benefit from ML methods, chemical problems must be cast so as to exploit the prime ML capabilities: correlation and classification.[50] Machine learning can, by constructing powerful correlation and classification models, greatly accelerate the discovery of catalysts and functional compounds in general.[162−164] However, correlation does not imply causation.[50] Thus, while parametrization can build predictive models, the applicability domain and uncertainty of these ML models must be evaluated carefully.[165−170]

Most ML techniques are based on the assumption that a mathematical relation exists between quantities describing intrinsic properties, such as molecular and atomic properties, of a system and some global, observable property of interest, such as the catalytic activity or selectivity.[171] A linear relationship is typically the easiest, yet often fruitful, assumption, but more complex, nonlinear models can also be constructed.

The key ingredients of regression models are the quantities that are correlated with the properties of interest. These are called descriptors, parameters, or features in ML language and should ideally encapsulate both steric and electronic properties of a candidate.[172] A plethora of such molecular descriptors have been proposed. Most of these have been developed for drug design, but descriptors are also being developed to tackle challenges in catalysis: for instance, by addressing the metal−ligand bonds.[173] Fey and co-workers have developed and surveyed a broad range of calculated descriptors for characterization of steric and electronic properties of phosphines and carbenes in transition-metal catalysts.[27] Many such descriptors are scalar values pertaining to atomic or molecular properties, such as the Tolman cone angle, geometrical features (e.g., bond distances and angles), HOMO−LUMO gaps, atomic charges, chemical shifts, and IR frequencies,[172] but even pK_a values have been used to predict catalytic activity.[174]

Predictions are also performed using vectors of such scalar descriptors and multidimensional descriptors. Three-dimensional grids of interaction energies (molecular interaction fields, MIFs)[175] are particularly useful when steric properties are dominant, such as in molecular recognition and stereoselectivity. For example, enantiomeric excesses in asymmetric catalysis (Figure 9)[17,176,177] have been predicted using such three-

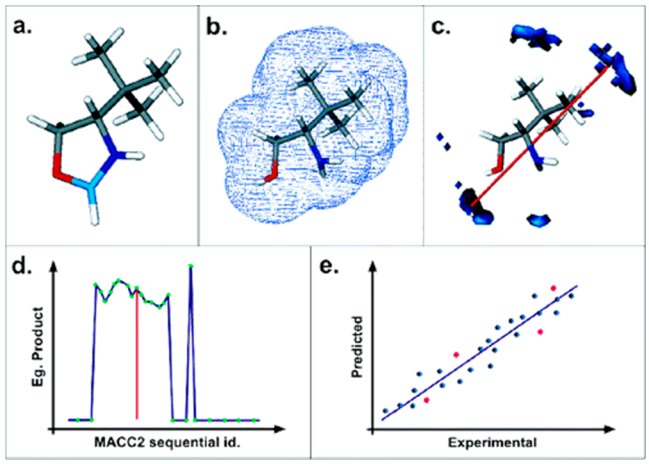


Figure 9. Procedure for constructing prediction models using alignment-independent descriptors derived from molecular interaction fields (MIFs): (a) geometry optimization; (b) MIF calculation; (c) identification of grid nodes with high interaction energy; (d) energy product vs node distance plot; (e) prediction model based on descriptors from (d). Reprinted with permission from ref 176. Copyright 2005 American Chemical Society.

dimensional maps[17] as well as alignment-independent descriptors derived from MIFs.[178] Three-dimensional maps derived from differences of MIFs have also been used to identify regions of maximum stereochemical induction around a chiral catalyst.[179]

Recently, new descriptors were developed to better account for noncovalent interactions.[180,181] Despite the weakness of individual noncovalent interactions, they may, combined, affect chemical reactivity,[182] including catalyst efficiency and selectivity,[183] and such interactions are frequently considered in catalyst design.[184,185] For instance, noncovalent interactions play a key role in the activity of molybdenum-based olefin metathesis catalysts.[186]

Overall, the picture emerging from evaluations of the applications of linear regression models to prediction of the performance of chiral catalysts is the crucial role of the

https://dx.doi.org/10.1021/acscatal.9b04952
ACS Catal. 2020, 10, 2354−2377

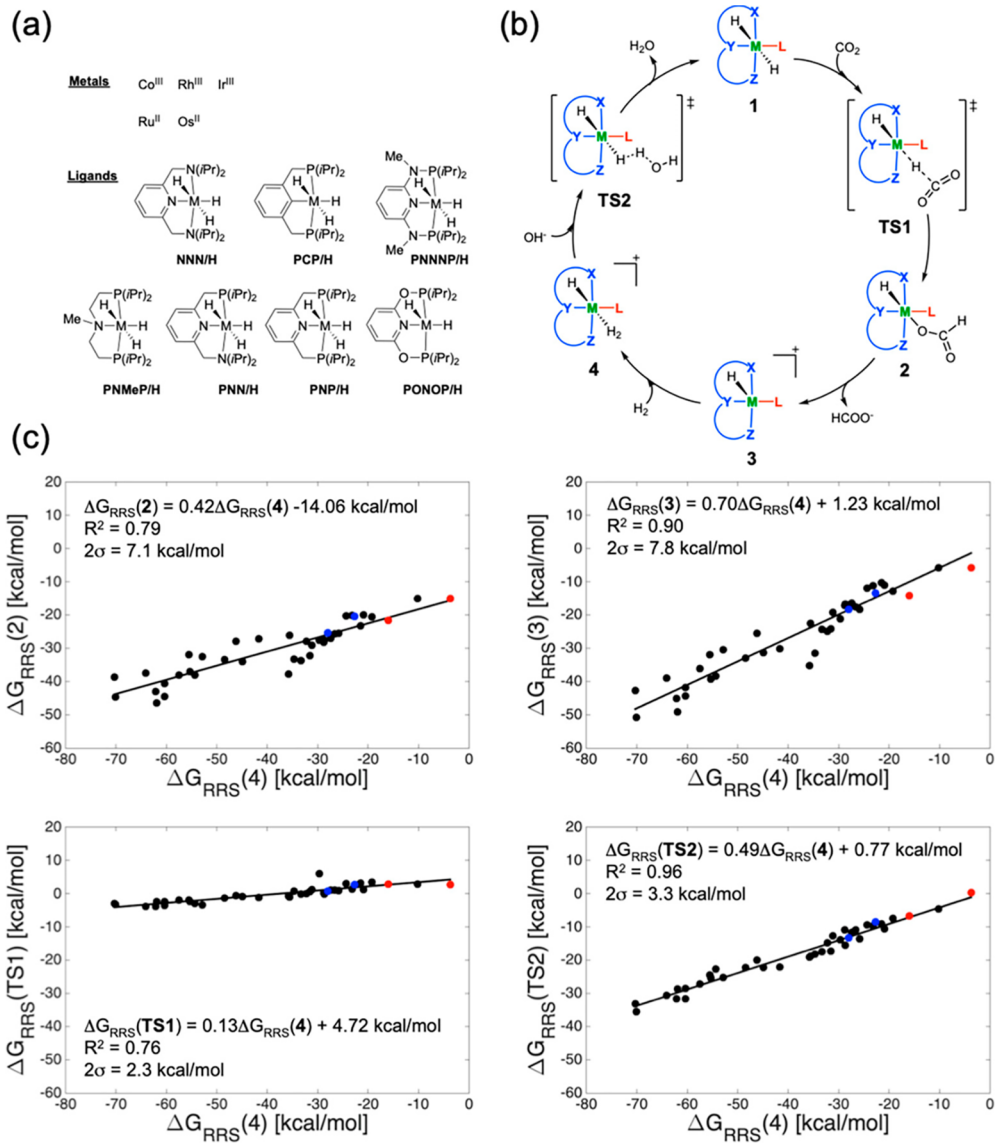


Figure 10. Identification and use of linear free energy scaling relationships (LFESRs).[208] (a) The catalysts of the training set. (b) Catalytic cycle for the conversion of CO_2 to formate. (c) Linear free energy scaling relationships of the catalytic cycle. $\Delta G_{RRS}(\mathbf{X})$ is the free energy of species **X** relative to the reference state **1**, and $\Delta G_{RRS}(\mathbf{4})$ is the descriptor variable.[206] Black points represent the training set, while red and blue points represent the validation set. Reprinted with permission from ref 208. Copyright 2019 American Chemical Society.

descriptors in determining the predictive power of the models.[26,172] The selection of descriptors may be guided by mechanistic factors, such as the interactions occurring at the rate-determining transition state.[187] Yet, one of the major advantages of ML models is that they do not necessarily rely on the reaction mechanism and can therefore be used also when the latter is unknown.[188]

Beyond linear regression, random forest models have recently been used to predict the performance of palladium-catalyzed amination.[189] Notably, the random forest models were trained using data collected from the results of high-throughput experimental testing, with more than 4000 experiments overall and with 120 molecular and atomic descriptors of metal-coordinating ligands, substrates, and additives.

Finally, so far only a few examples exist where the performance of homogeneous catalysts has been predicted by neural networks (NN) and deep learning models. As pointed out above, the training of such models requires large volumes of data. The lack of consistent, curated data and the multimolecular nature of catalytic processes have been suggested as the main challenges that impede the application of the otherwise ubiquitous deep learning models to homogeneous catalysis.[160]

Nonetheless, the recent work of Denmark and co-workers demonstrates that a deep feed-forward neural network could be trained to successfully predict the stereoselectivity of addition of a thiol to imines as catalyzed by phosphoric acid.[190] A virtual library of such catalysts was created, and the authors used sampling algorithms to identify a representative training set for their prediction models. The models, despite being trained on cases of low to medium selectivity (below 80% enantiomeric excess) only, could still predict high selectivity resulting from catalyst−substrate combinations well outside of the training set.

Notably, the NN-based ML models may be very useful when linear models fail to provide accuracy. For instance, NNs were recently trained to estimate the spin-state-dependent formation energy of metal−oxo complexes,[170] which are essential

https://dx.doi.org/10.1021/acscatal.9b04952
ACS Catal. 2020, 10, 2354−2377

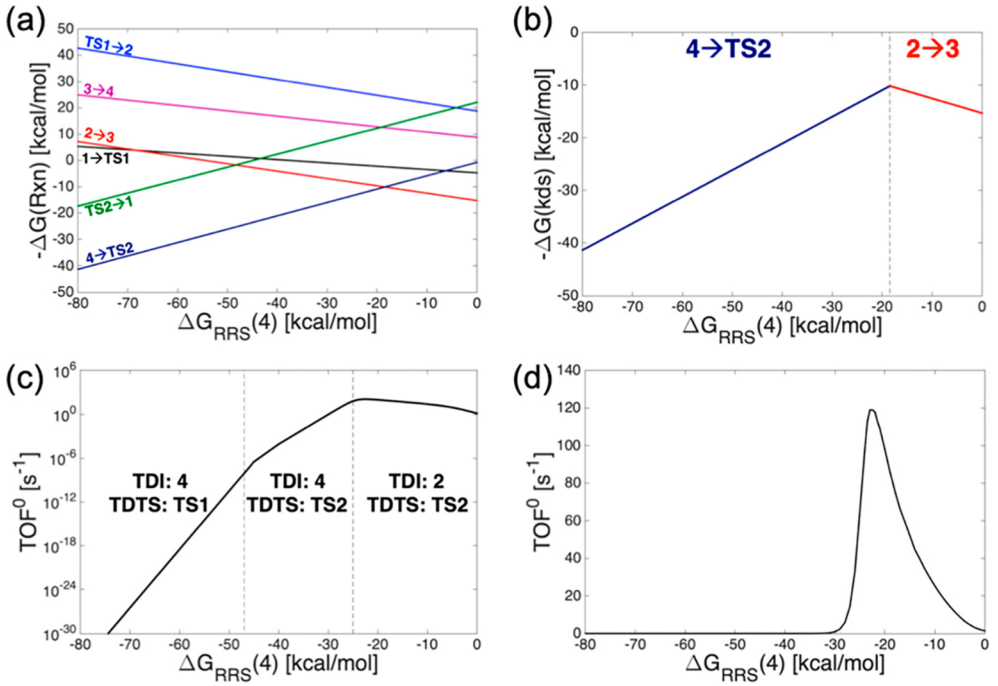


Figure 11. (a) Linear free energy scaling relationships (LFESRs) correlating ΔG along the catalytic cycle with the descriptor variable $\Delta G_{RRS}(\mathbf{4})$, which is the relative free energy of an intermediate (labeled **4** in the original publication).[208] For a given $\Delta G_{RRS}(\mathbf{4})$, the lowest line corresponds to the kinetics-determining step (kds), which thus generates the volcano plot in (b). (b) Volcano plot constructed from the lowest lines in (a). (c) TOF-based volcano plot with the ordinate given in log scale. (d) Same as (c) but with the ordinate given in linear scale. Reprinted with permission from ref 208. Copyright 2019 American Chemical Society.

intermediates in water splitting and oxidation of hydrocarbons. These formation energies correlate poorly with conventionally used electronic descriptors, thus hampering the use of descriptor-based ML models. The NN-based prediction models were used to uncover unexpected combinations of transition metal, oxidation state, and ligand set and offered promising candidate metal−oxo intermediates.[170]

Another family of flexible ML models used to model highly nonlinear functions are Gaussian process (GP) models.[191] GP models are probabilistic models that, upon training, can be used to generate predictions from unseen input. The predictions are in the form of mean values that are associated with a variance that indicates the confidence in the prediction (i.e., Bayesian nature). This allows a decision of whether the prediction is sufficiently reliable or should be discarded and possibly replaced by an explicitly calculated value that can be used to retrain the GP model. Thus, the quality of the predictions can be improved systematically as more points are added to the training set, which allows for an iterative refinement of the GP model.[192−194] Moreover, GP models are easier to train than NN models and are particularly well suited for small- to medium-sized training sets and training sets containing data of different levels of quality.[195]

So far, GP models have not, to our knowledge, been used in homogeneous catalysis. However, the applications in heterogeneous catalysis and materials design are promising.[196] For instance, GP models have proved able to predict adsorption, binding, and formation energies or enthalpies of reaction intermediates,[194,195,197] and such predicted energies have also been used to guide the exploration of reaction networks.[194,198] Importantly, even if automation in both computational and experimental chemistry (i.e., high-throughput experimentation) improves upon the situation, cases in which large, consistent, and highly accurate data sets are available are also still rare in homogeneous catalysis. Thus, the ability of GP-based prediction models to build predictive models from sparse data and small- to medium-sized data sets originating from multiple sources (experiments and computations alike) and with multiple levels of accuracy[195] holds great promise for such models, in homogeneous catalysis and beyond.

4.1.2. Linear Free Energy Scaling Relationships and the Energy Span Model. Linear free energy relationships (LFERs) have been around for nearly a century and have provided some of the most used structure−activity relationships and "rules of thumb" in chemistry, such as Brønsted's correlation of acid or base strength with catalytic activity,[199] Hammett's equation and parameters for electronic substituent and reaction effects,[200] and Taft's addition of steric effects.[201] In general, these and other LFERs provide fundamental chemical understanding by establishing linear correlations between free energies (or the logarithms of kinetic or equilibrium constants) obtained for two different reactions, as exemplified by acid strength and catalytic activity in acid catalysis.

Modern computational methods and hardware permit the exploration of more direct linear correlations involving a single reaction only. For example, calculated binding energies of reaction intermediates in series of heterogeneous catalysts have been found to correlate linearly.[202−204] More generally, the relative free energies of intermediates and transition states of catalytic reactions can often be related to one another in a linear way to achieve so-called linear free energy scaling relationships (LFESRs).[203,205] If such relationships exist and are valid over the entire set of candidate catalysts, the energy profile of a new catalyst can be estimated by computing only one relative free energy, often termed the *descriptor variable* (Figure 10).[206] Additional simplification and speedup has been achieved by

https://dx.doi.org/10.1021/acscatal.9b04952
ACS Catal. 2020, 10, 2354−2377

estimating the descriptor variable using machine-learning models.[207]

Indeed, Corminboeuf and co-workers have shown that such LFESRs exist for different homogeneous transition-metal-catalyzed reactions and hold true for a set of catalysts including different metals and ligands (Figure 10).[205−209] In contrast, the accuracy of LFESR-based models has also been reported to be limited when the changes in chemical features are substantial.[210,211] While the identification of these limitations may even be exploited to develop new design strategies,[212] improved accuracy has been obtained by constructing ligand-specific LFESR models: i.e., models specific to each type of ligand. Combinations of such models with simple ligand-specific descriptors, such as the Tolman angle, have facilitated interpretation of the results as well as derivation of ligand design criteria.[9]

LFESRs have also been proposed as a means to construct volcano plots, which are widely used in heterogeneous catalysis,[213] for the evaluation of homogeneous catalysts (Figure 11).[9,205−209,214] Volcano plots graphically represent the idea, first formulated by Sabatier,[215] that optimal catalysts should bind intermediates neither too weakly nor too strongly. Initially, volcano plots were used only for thermodynamic analysis: that is, the relationships were limited to the free energy differences between intermediates.[205] Later, the use of LFESR models has been extended to estimating activation barriers,[9,206] thus accounting for kinetics and improving the predictions of selectivity.[214]

Various flavors of LFESR and volcano plots have been used in the screening for active and selective rhodium catalysts for hydroformylation of olefins,[9,214] in the evaluation of pincer-ligand-coordinated catalysts for hydrogenation of carbon dioxide to formate,[206,208] and in the evaluation of cross-coupling catalysts.[205,207]

Although volcano plots are intuitive and ideal for visual inspection, they can also be evaluated numerically to rank candidates in an automated design framework. For example, automated analysis of catalytic cycles was recently obtained by coupling LFESRs with the energy span model.[216−219] The latter condenses the free energy profile of the catalytic cycle, including off-cycle intermediates and resting states, into a single numerical quantity representing the efficiency of the catalytic system, the turnover frequency (TOF).[220] Notably, the use of TOFs as condensed descriptors of the catalyst efficiency does not correspond to using the LFESR descriptor variable as a figure of merit. In fact, both the volcano plot and the LFESR-derived TOF define the range of values for the descriptor variable leading to the highest efficiency and thus create a nonlinear relation between the descriptor variable and the figure of merit of a candidate catalyst.

Overall, these contributions suggest that, when linear scaling relationships exist, the thermodynamic and kinetic features of a catalytic cycle can be estimated with an accuracy sufficient for high-throughput screening. Moreover, in combination with the energetic span model, LFESRs allow for quantitative evaluations of the catalytic cycle particularly suitable for automated in silico design.

4.2. Fast Molecular Modeling Techniques. While modern DFT methods still involve severe approximations and should be validated and checked against experiment and higher-level calculations in all application domains, it is nevertheless the most accurate and computationally expensive class of method that can usually be afforded for mechanistic studies and intuition-driven manual design in catalysis.[221,222] However, for high-throughput virtual screening and de novo catalyst design studies, DFT is too costly except for small chemical systems. Fortunately, comparable levels of accuracy may, in well-prepared cases, be obtained with computationally less demanding methods. These methods can be divided into the following categories: (i) specifically parametrized, empirical models, (ii) approximate and fast electronic structure methods, and (iii) machine-learned models of the potential energy surface (PES).

4.2.1. Empirical Methods: Customized Force Fields. Developing force fields is easier than ever. Data on which to train the molecular-mechanics methods are readily available, for example via quantum chemical calculations,[223] and the parametrization process may be automated.[224,225] In a catalyst design project, the challenge thus is to ensure that the parameters are broadly applicable and accurate across the corresponding search space. In addition, whereas bond rupture and formation are intrinsic to catalysis, these are phenomena traditional force fields cannot describe. In rare cases, reactive force fields, such as ReaxFF,[226] can estimate reaction barriers involving rupture and formation of bonds.[227] However, this capability comes at a price: parametrization of such force fields is still challenging,[228] albeit new force field parametrization methods might reduce this problem.[229]

When the activity- or selectivity-determining transition state is known, an alternative to reactive force fields is Quantum-Guided Molecular Mechanics (Q2MM).[60,61,230] While other methods[231,232] and force fields such as multi-configuration molecular mechanics (MCMM)[233] and the empirical valence bond model (EVB)[234] parametrize the PES in the TS region by mixing the reactant and product potential energies (Figure 12),

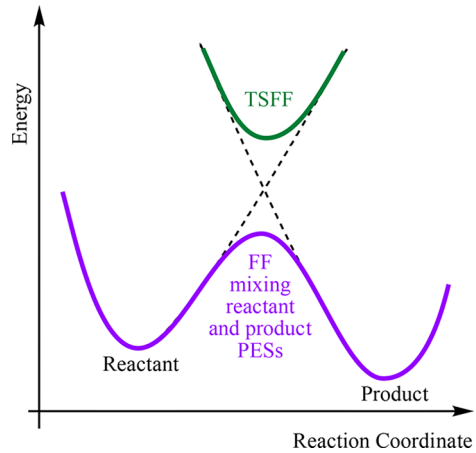


Figure 12. Comparison of force fields (FF) for modeling transition states (TS). Adapted with permission from ref 60. Copyright 2016 American Chemical Society.

the Q2MM method creates a transition state force field (TSFF).[235] A TSFF is built on standard force fields, such as MM3 and AMBER, by adding parameters for the TS, which is treated as a minimum rather than a saddle point (Figure 12). Q2MM was the engine behind the aforementioned virtual screening recently published by Munday, Wiest, Norrby, and co-workers.[59] In addition to asymmetric rhodium-catalyzed hydrogenation,[62,63] TSFFs have been developed to model asymmetric osmium-catalyzed dihydroxylation,[236−238] stereoselective addition to aldehydes,[239−241] and docking of transition-state structures into the active site of cytochrome P450.[242]

https://dx.doi.org/10.1021/acscatal.9b04952
ACS Catal. 2020, 10, 2354−2377

4.2.2. Hybrid Methods: Divide and Conquer. Hybrid methods combine the accurate, yet costly quantum mechanics (QM) methods with the empiricism and speed of classical molecular mechanics (MM) methods by dividing a chemical system into QM and MM portions.[233,243−248] The resulting QM/MM methods limit the computationally demanding QM calculation to the portion, such as a region involving bond formation or rupture or one for which a suitable empirical model is not available, of the system where an accurate, electron-aware model is needed (Figure 13).

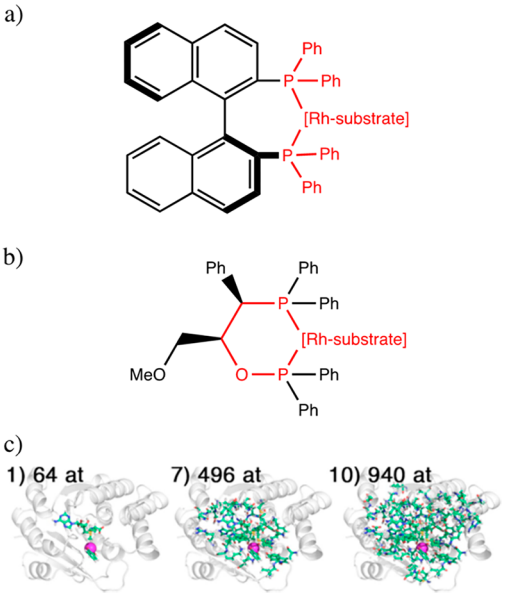


Figure 13. Examples of QM/MM partitioning of catalytic systems. (a) Partitioning used in ref 249. to model rhodium-catalyzed isomerization of allylic amines (QM region rendered in red). (b) Partitioning used in ref 250. to model rhodium-catalyzed asymmetric hydrogenation (QM region rendered in red). (c) Three QM/MM models of a biocatalyst involving successively larger QM regions, encompassing Mg^{2+} ion (rendered in magenta) and parts of the surrounding protein (in stick representation). Part c) of the figure is reprinted with permission from *J. Phys. Chem. B* **2016**, *120*, 11381−11394. Copyright 2016 American Chemical Society.

A central challenge that is addressed at different levels of sophistication by the various QM/MM methods is the interaction between the QM and MM parts of the QM/MM model.[246,251] Ideally, the two parts should influence each other continuously and completely (i.e., including all electronic and steric effects), but practical approximations have been developed to best recover most of these interactions, while still achieving low computational cost as well as compatibility with existing QM and MM software tools. In particular, the QM part is typically embedded in an effective electrostatic field resulting from the partial atomic charges in the MM part (i.e., electrostatic embedding). Moreover, the presence of covalent bonds crossing the boundary between QM and MM regions requires the saturation of the truncated system with special link atoms[252] that may require dedicated parameters in the MM model.

Despite the challenges associated with the design, setup, and validation of QM/MM models,[253] which, among other things, involve monitoring the convergence of a target property on extending the QM region,[254] these hybrid methods have been successfully applied to explore reaction mechanisms[249,255,256] and to calculate catalytic activity and selectivity.[257] QM/MM methods have proved particularly useful in systems characterized by confined catalytic "pockets" surrounded by large, mostly chemically inert regions consisting of substituents, molecules, residues, or polymeric structures, such as those found in enzymes[246] and metal−organic frameworks.[258]

4.2.3. Semiempirical Tight-Binding Methods. Most semiempirical methods are several orders of magnitude faster than ab initio calculations,[259] albeit at the price of accuracy and reliability across application domains.[260] A particularly difficult domain in this respect has been that of transition-metal and organometallic chemistry, for which the accuracy of semiempirical methods is highly dependent on the combination of transition metal and ligands. Thus, examples of useful and even surprisingly good accuracy[261] are found alongside cases for which even the geometries may be of too low a quality to be useful.[260,262] To address the accuracy problem, Grimme and coworkers[263] have recently developed a general semiempirical tight-binding (TB) method termed Geometry, Frequency, Noncovalent, eXtended TB (GFN-xTB). As the name suggests, the method is predominantly intended to provide reasonable geometries, frequencies, and noncovalent interactions. Other goals are wide applicability and numerical robustness, which is achieved[264] by relying on a few global and element-specific rather than pair-specific parameters and by including parameters up to $Z = 86$: i.e., including the lanthanoids.[265]

An improved version of this method, termed GFN2-xTB, was recently used in automated exploration of the chemical reaction space of organic, organometallic, and transition-metal compounds.[266] The method could explore both the conformational and chemical compound space (i.e., constitutional isomers) and estimate reaction pathways with high computational efficiency and sufficient accuracy. For highly accurate relative energies and properties, refinement, for example in the form of subsequent single-point energy calculations using higher-level methods, is still needed, but GFN2-xTB promises to drastically reduce the cost of identifying key geometries and reaction pathways of large reaction systems and in high-throughput and automated computational studies.

4.2.4. Neural Network Potentials. Among the ML methods that have been used to predict energies,[267,268] neural network potentials (NNP) have recently shown fast and accurate prediction of the total energy of organic molecules.[269−271] For example, in the potential termed ANI-1,[269] element-specific NNPs predict the atomic contributions to the total energy using a transferable representation of the chemical environment of each atom, in the form of a single-atom atomic environment vector built from the geometry of the input molecule. Both radial and angular features are accounted for in the representation of the atomic environment used by the NNPs. The NNPs were trained on a large quantity of data to ensure sampling of a wide variety of molecules and molecular interactions spanning both conformational and configurational degrees of freedom, and ANI-1 performs well far beyond molecules of the training set. This NNP-based method reproduces relative DFT energies, in particular near-equilibrium geometries, better than popular semiempirical methods such as AM1[272] and PM6[273] and is up to 6 orders of magnitude faster than DFT.

Although these results are very promising, ANI-1 is limited to molecules of only four elements: H, C, N, and O. Further developments, such as the recent distribution of a software package for the training of NNPs,[274] should with time enlarge the list of supported elements to including those relevant for

https://dx.doi.org/10.1021/acscatal.9b04952
ACS Catal. 2020, 10, 2354−2377

catalysis. However, the number of first-principles reference points required to achieve this goal is enormous, a challenge likely to retard the spread of NNPs across the periodic table.[275]

5. CATALYST DESIGN OBJECTIVES

In the previous sections the catalyst design objective was taken to be a generic performance sometimes exemplified in terms of catalytic activity or selectivity. However, the actual, overall performance of a catalyst must satisfy a multitude of objectives, including those that are less frequently mentioned, such as solubility, stability, robustness, initiation to catalytically active species, synthetic accessibility, cost, and toxicity. Therefore, just like the drug design problem,[276] the optimization of a catalyst is truly multiobjective and the individual objectives may represent conflicting requirements that cannot be fully satisfied simultaneously. To illustrate, high activity often comes at the price of selectivity, and a fast-initiating precatalyst might also decompose more readily than a more slowly initiating one (Figure 14).[277]

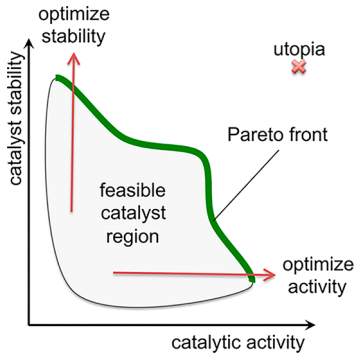


Figure 14. Different properties such as activity and stability are often conflicting objectives in catalyst design. The optimal catalysts are those that display the best possible compromises: i.e., nondominated solutions (i.e., the Pareto front).

Objectives that are known a priori and that can be estimated within the available computational resources may be included in the definition of the catalyst performance essentially in two ways.[278] One way is to specify the mathematical nature of the compromise: for example, by a weighted sum or a more complex expression combining a numerical satisfaction of each objective into an overall scalar scoring function, or fitness function.[279] This effectively converts the multiobjective problem into a single-objective one. Alternatively, the objectives can be handled independently and simultaneously by collecting the satisfaction scores for each objective in a vector of unprioritized elements. The aim of such a multiobjective approach is to discover the front of best and equally good candidates. These are termed nondominated solutions in the sense that no other known candidate has higher scores for all objectives (Figure 14).[278] Such multiobjective optimization techniques are used in drug design,[280] but, to the best of our knowledge, they have yet to be applied in the design of homogeneous catalysts.

These multiobjective techniques handle objectives that are known a priori, but there are also unknown objectives that may come to attention only after several candidates have been evaluated. For instance, the reactivity and stability of novel compounds might pose unexpected challenges that are impossible to foresee on starting the design experiment.[29] Such a change in perspective cannot be handled by a static definition of the performance in an automated design protocol. Instead, the definition of the performance should rather be a dynamic entity able to respond to new knowledge generated on the fly, either by the automated design process itself or by the user: i.e., the supervising chemist. In such a dynamic design process, the ranking of the candidates may be modified along the way: for instance, by extension of the objectives or recalculation of the desirability (figure of merit) with modified expressions. In such a scenario, the automatically generated data need to remain accessible, searchable, and automatically manipulated throughout, thus posing intriguing data management challenges.

6. HIGH-THROUGHPUT CALCULATIONS AND DATA MANAGEMENT

Perhaps the most obvious advantage of automated design is the high-throughput fashion in which candidates are evaluated. This evaluation is usually very specific to the design objective (see section 5). Still, the chemical information generated in each such catalyst evaluation, which might even include modeling of several molecular structures, is typically much more general and richer than that conveyed by a single scalar or vector alone: i.e., the desirability. This general and rich information may easily become useful in unexpected contexts.

For example, in the above sections, we have frequently pointed out the need for large amounts of consistent data for training of parametrized methods such as force fields and machine learning models. This is true also for experimental results. Reid and Sigman recently suggested that the practice of reporting only the best results must change to help develop the systematic use of statistical and ML methods for predicting catalyst performance.[26] The lagging development of information technology and community-wide approaches to contribute to data collection and accessibility, even of failed experiments,[281] has been identified as a major obstacle also for the use of artificial intelligence in materials design.[160,282] This has led to the creation of consortia that collect and share data to promote joint efforts in the design of materials,[283−286] a field where high-throughput computational screening is a more mature and frequently used tool[287,288] in comparison to the field of homogeneous catalysis.

A similar sharing and repurposing of data to boost catalyst design must involve the development and use of databases of computational and experimental data. This is true for compounds and reaction discovery across chemistry, a fact that has received attention in recent years and led to the creation of new databases.[289,290] Some of these databases and efforts seem particularly promising with respect to computationally guided catalyst design.[291−293] However, to populate these databases with enough data to train ML models on catalytically relevant chemical systems, massive efforts to reuse and repurpose computational data are needed. Such reuse and repurposing may be realized by making sure that workflow managers such as AiiDA,[294] QMflows,[295] AFlow,[296] Signac,[297] and FireWorks[298,299] prepare job summaries in standardized data formats used by the community repositories. The most detailed management control is currently offered by AiiDA,[294] which keeps track of the complete history, including information on methods, input parameters, computer, postprocessing tools, and dependencies, leading to a computational result, thereby mapping the complete data provenance necessary to ensure reproducibility and repurposing.

https://dx.doi.org/10.1021/acscatal.9b04952
ACS Catal. 2020, 10, 2354−2377

7. SOFTWARE PACKAGES FOR CATALYST DESIGN

Some of the software packages that were used in the examples reviewed above of automated catalyst design are described in the following. The description is limited to packages that may offer good starting points for readers interested in trying automated catalyst design in practice.

CatVS (Catalyst Virtual Screening) was developed for virtual screening using Q2MM force fields[59] and is a combination of Q2MM Python modules and interfaces depending on software such as Maestro[300] and MacroModel.[301] The Q2MM repository[302] contains the building blocks necessary for CatVS. The program has been used, successfully, in the screening of Rh-based catalysts for asymmetric hydrogenation of enamines.[59]

AARON (for which the most recent complete name is An Automated Reaction Optimizer for New catalysts)[65,67,185] is a collection of Python modules for the construction and analysis of molecular objects and the automation of quantum chemistry workflows. Multiple chemical species determining the performance of a catalyst, such as intermediates and transition states, may be systematically generated and subjected to calculations by AARON. Recent applications of this package include screening of both organocatalysts[66] and transition-metal catalysts.[64,67] AARON is distributed as open-source software.[303]

ACE (Asymmetric Catalyst Evaluation)[304] has been distributed as part of the Forecaster user interface[305] and can be used for the virtual screening of organocatalysts for asymmetric reactions. ACE derives the stereomeric excess of asymmetric reactions by calculating the difference between the diastereomeric TS energies. ACE automates the conformational search, geometry optimization, and energy evaluation of these transition states using an empirical molecular mechanics method in which structures in the transition region are described as linear combinations of the reactants and products. Construction of a TS-specific force field is thus avoided in ACE. The software has been tested by reproducing the selectivity of proline-catalyzed aldolizations[304] and dioxirane-catalyzed asymmetric epoxidations[232] and by constructing libraries of synthetically accessible molecules as candidate organocatalysts.[150] ACE is available, free of charge, for academic use.[306]

The Python toolkit molSimplify[307] was initially built for screening of inorganic molecules and intermolecular complexes (via intermolecular docking) and has later been extended by a genetic-algorithm optimizer[162] and possibilities for fast, ML-accelerated property prediction.[308] Applications of molSimplify range from the design of inorganic complexes with spin-crossover properties[162] or desired HOMO−LUMO gaps[309] to the virtual screening of indium carboxylate precursors for quantum dots[310] and tuning of the redox properties of the ferrocene/ferrocenium system.[311,312] The program is distributed as open-source software.[313]

DENOPTIM (DE Novo OPTimization of In/organic Molecules) is a recently released Java package for virtual screening and de novo design of functional organic, inorganic, organometallic, and supramolecular compounds.[314] This flexibility stems from the customizable molecular builder[85] that offers control of the synthetic accessibility of the candidates at the same time as allowing for the automated construction even of short-lived intermediates, transition states, ion pairs, and supramolecules.[315] These capabilities have been utilized in the de novo design of a range of different compounds, including ruthenium-based olefin metathesis catalysts,[80] azobenzenes with tailored excitation energies,[316] dyes for solar cells,[317,318] monomers for high-refractive-index organic polymers,[319] organic solvents for CO_2 capture,[320] and iron-based spin-crossover compounds.[321] One of the last iron complexes was later confirmed, in experiments, to possess spin-crossover properties and is thus, to our knowledge, the first automatically de novo designed inorganic molecule experimentally verified to reflect the intended properties.[322] DENOPTM is distributed as open-source software.[323]

Finally, a package for global optimization of catalytic environments for inverse design is under development.[112] These catalytic environments are electrostatic only, consisting of discrete point charges or multipoles, and are optimized so as to maximize the speed-up of target catalytic reactions. The optimized catalytic fields, termed GOCAT (Globally Optimal Catalysts), have so far not been transformed into well-defined chemical structures. The software, which is dependent on the OGOLEM package for global optimization,[324] is available from the authors upon request.

8. CONCLUSIONS AND OUTLOOK

The effect of computational studies on homogeneous catalysis increases rapidly. The role of these computational studies has already evolved from mainly being that of supporting interpretations of experiments to, more recently, also guiding the search for new and better catalysts. Automation of this search, that is, the prediction and generation of promising candidates, is the pillar of automated in silico design.

The perhaps most straightforward and intuitive automated strategies mimic the experimental trial and error approach and modify the catalysts in silico, step by step, toward better performance. A more elegant solution to the design problem takes the reverse approach and aims to derive candidate structures from the desired property, thereby limiting, in principle, the portions of the overwhelming chemical space that must be explored. So far, most of the examples of inverse design are not from homogeneous catalysis and still, to be practically useful, often incorporate strategies from direct design, such as screening of predefined candidates. Still, with further developments, we anticipate that inverse design will mature into a very useful tool in homogeneous catalysis.

In comparison, automated virtual screening already is a mature technique that has demonstrated the usefulness of direct design strategies. Central to these strategies is the intuitive guess−check−guess cycle that is readily automated, a key factor that has helped spread virtual screening across science, including to homogeneous catalysis. Design methods such as virtual screening depend on the performance of the underlying workflow, typically including molecular modeling and/or empirical prediction models, that estimate the catalytic properties. Thus, any development of molecular modeling and empirical prediction models that lead to faster and more accurate prediction of catalytic performance paves the way for automated screening and de novo design of catalysts. The latter method is increasingly favored as search spaces grow and the prediction workflows become faster.

Of course, chemists desire large search spaces with high chemical variability. The automated procedures may easily populate any part of these search spaces with chemical species. Still, even with these nearly unlimited possibilities, researchers are frequently limiting their searches to specific classes of compound with largely known catalytic properties, varying only noncritical substituents and groups to optimize the perform-

https://dx.doi.org/10.1021/acscatal.9b04952
ACS Catal. 2020, 10, 2354−2377

ance. This conservative strategy does not foster the discovery of truly novel candidates, a problem largely originating from the computational cost of exploring large portions of the chemical space and from the fear of generating unrealistic or synthetically challenging candidates that might not even be catalytically active. In the future, realistic, synthetically accessible compounds might still be generated even when large search spaces are used by combining modular scaffolds inspired by the modular structure of biopolymers, which consist of constant backbones and variable side chains.[159] Such modular approaches offer great promise for automated in silico design and, via automated synthesis, also for in vitro experiments in catalysis.

In addition to the desire for large and modular chemical spaces, automated design faces the tradeoff between accurate predictions and computational costs. The costs are due mostly to the molecular modeling component of the prediction models. When molecular modeling cannot, e.g., via a ML-based prediction model, be avoided, reliable and computationally efficient empirical and semiempirical methods are, due to the size of search spaces, almost always preferred over more accurate methods such as DFT. In the future, machine-learned potentials are likely to contribute to expanding the scope of empirical tools in automated catalyst discovery. In particular, we expect the development of methods that train, on the fly, empirical and machine learning models within the automated design workflow. Molecular data such as molecular geometries, energies, and frequencies, calculated during the design workflow, will thus serve two purposes: evaluation of candidates and training of an empirical model. For example, an automated design workflow could start by letting a molecular modeling method (i.e., DFT, DFTB,[325,326] or a recently developed semiempirical method such as GFN-xTB)[327] deliver the data needed both for evaluating candidates and for training an empirical model on the fly. Later, within the same automated workflow, that empirical model may be used to speed up new scoring/fitness evaluations. Such a workflow would be self-catalyzed and extract more value from computationally generated molecular data in comparison to regular approaches by producing at the same time as deploying such data.

Even with this focus on the development of cost-efficient molecular modeling methods for prediction of catalytic performance, the fundamental role of experimental data should not be forgotten.[328] In fact, prediction models trained on experimental data are frequently used in catalyst design,[19,23,172,186,189,190] and automated in silico design should build on this experience and may even benefit directly from existing prediction models based on experimental data. Integration of experiment and modeling may also boost catalyst development via early verification of computational predictions and thus feedback to the prediction routines. In addition to representing an evaluation of the accuracy of the prediction, this feedback also serves to highlight shortcomings, such as that of overlooking catalyst stability as a critical factor, in the catalyst design objectives. The prediction−experiment combination thus allows for evolving the catalyst design criteria iteratively, in cycles of computational prediction, experimental verification, and feedback to the prediction machinery.[19,329] This strategy not only ensures that predictions and design objectives are consistent with observations but also helps extract the maximum amount of information and insight possible from the molecular modeling techniques and prediction models.

In conclusion, automated catalyst design approaches used in conjunction with experimental followup may enhance our understanding of catalytic processes and speed up the discovery of new catalysts. We have so far only seen glimpses of the power of this combination, but it is, just as automation affects society at large, going to transform the way new catalysts are discovered and developed.

AUTHOR INFORMATION

Corresponding Authors

Marco Foscato − *Department of Chemistry, University of Bergen N-5007 Bergen, Norway;* orcid.org/0000-0001-7762-6931; Email: marco.foscato@uib.no

Vidar R. Jensen − *Department of Chemistry, University of Bergen N-5007 Bergen, Norway;* orcid.org/0000-0003-2444-3220; Email: vidar.jensen@uib.no

Complete contact information is available at: https://pubs.acs.org/10.1021/acscatal.9b04952

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was funded by the Research Council of Norway (RCN), via projects 262370 and 288135.

REFERENCES

(1) Wang, A.; Olsson, L. The Impact of Automotive Catalysis on the United Nations Sustainable Development Goals. *Nat. Catal.* **2019**, *2*, 566−570.

(2) Houk, K. N.; Liu, F. Holy Grails for Computational Organic Chemistry and Biochemistry. *Acc. Chem. Res.* **2017**, *50*, 539−543.

(3) Tsang, A. S. K.; Sanhueza, I. A.; Schoenebeck, F. Combining Experimental and Computational Studies to Understand and Predict Reactivities of Relevance to Homogeneous Catalysis. *Chem. - Eur. J.* **2014**, *20*, 16432−16441.

(4) Ahn, S.; Hong, M.; Sundararajan, M.; Ess, D. H.; Baik, M.-H. Design and Optimization of Catalysts Based on Mechanistic Insights Derived from Quantum Chemical Reaction Modeling. *Chem. Rev.* **2019**, *119*, 6509−6560.

(5) Houk, K. N.; Cheong, P. H.-Y. Computational Prediction of Small-Molecule Catalysts. *Nature* **2008**, *455*, 309−313.

(6) Simm, G. N.; Vaucher, A. C.; Reiher, M. Exploration of Reaction Pathways and Chemical Transformation Networks. *J. Phys. Chem. A* **2019**, *123*, 385−399.

(7) Falivene, L.; Cao, Z.; Petta, A.; Serra, L.; Poater, A.; Oliva, R.; Scarano, V.; Cavallo, L. Towards the Online Computer-Aided Design of Catalytic Pockets. *Nat. Chem.* **2019**, *11*, 872−879.

(8) Hopmann, K. H. Quantum Chemical Studies of Asymmetric Reactions: Historical Aspects and Recent Examples. *Int. J. Quantum Chem.* **2015**, *115*, 1232−1249.

(9) Wodrich, M. D.; Busch, M.; Corminboeuf, C. Accessing and Predicting the Kinetic Profiles of Homogeneous Catalysts from Volcano Plots. *Chem. Sci.* **2016**, *7*, 5723−5735.

(10) Hopmann, K. H. How Accurate Is DFT for Iridium-Mediated Chemistry? *Organometallics* **2016**, *35*, 3795−3807.

(11) Peng, Q.; Duarte, F.; Paton, R. S. Computing Organic Stereoselectivity - from Concepts to Quantitative Calculations and Predictions. *Chem. Soc. Rev.* **2016**, *45*, 6093−6107.

(12) Vogiatzis, K. D.; Polynski, M. V.; Kirkland, J. K.; Townsend, J.; Hashemi, A.; Liu, C.; Pidko, E. A. Computational Approach to

Molecular Catalysis by 3d Transition Metals: Challenges and Opportunities. *Chem. Rev.* **2019**, *119*, 2453−2523.
(13) Kwon, D.-H.; Fuller, J. T.; Kilgore, U. J.; Sydora, O. L.; Bischof, S. M.; Ess, D. H. Computational Transition-State Design Provides Experimentally Verified Cr(P,N) Catalysts for Control of Ethylene Trimerization and Tetramerization. *ACS Catal.* **2018**, *8*, 1138−1142.
(14) Chen, X.-Y.; Pu, M.; Cheng, H.-G.; Sperger, T.; Schoenebeck, F. Arylation of Axially Chiral Phosphorothioate Salts by Dinuclear Pd^{I} Catalysis. *Angew. Chem., Int. Ed.* **2019**, *58*, 11395−11399.
(15) Burello, E.; Rothenberg, G. In Silico Design in Homogeneous Catalysis Using Descriptor Modelling. *Int. J. Mol. Sci.* **2006**, *7*, 375−404.
(16) Occhipinti, G.; Bjørsvik, H.-R.; Jensen, V. R. Quantitative Structure-Activity Relationships of Ruthenium Catalysts for Olefin Metathesis. *J. Am. Chem. Soc.* **2006**, *128*, 6952−6964.
(17) Ianni, J. C.; Annamalai, V.; Phuan, P. W.; Panda, M.; Kozlowski, M. C. A Priori Theoretical Prediction of Selectivity in Asymmetric Catalysis: Design of Chiral Catalysts by Using Quantum Molecular Interaction Fields. *Angew. Chem., Int. Ed.* **2006**, *45*, 5502−5505.
(18) Drummond, M. L.; Sumpter, B. G. Use of Drug Discovery Tools in Rational Organometallic Catalyst Design. *Inorg. Chem.* **2007**, *46*, 8613−8624.
(19) Maldonado, A. G.; Rothenberg, G. Predictive Modeling in Homogeneous Catalysis: A Tutorial. *Chem. Soc. Rev.* **2010**, *39*, 1891−1902.
(20) Fey, N. The Contribution of Computational Studies to Organometallic Catalysis: Descriptors, Mechanisms and Models. *Dalton Trans.* **2010**, *39*, 296−310.
(21) Cruz, V. L.; Martinez, S.; Ramos, J.; Martinez-Salazar, J. 3D-QSAR as a Tool for Understanding and Improving Single-Site Polymerization Catalysts. A Review. *Organometallics* **2014**, *33*, 2944−2959.
(22) Pickup, O. J. S.; Khazal, I.; Smith, E. J.; Whitwood, A. C.; Lynam, J. M.; Bolaky, K.; King, T. C.; Rawe, B. W.; Fey, N. Computational Discovery of Stable Transition-Metal Vinylidene Complexes. *Organometallics* **2014**, *33*, 1751−1761.
(23) Piou, T.; Romanov-Michailidis, F.; Romanova-Michaelides, M.; Jackson, K. E.; Semakul, N.; Taggart, T. D.; Newell, B. S.; Rithner, C. D.; Paton, R. S.; Rovis, T. Correlating Reactivity and Selectivity to Cyclopentadienyl Ligand Properties in Rh(III)-Catalyzed C-H Activation Reactions: An Experimental and Computational Study. *J. Am. Chem. Soc.* **2017**, *139*, 1296−1310.
(24) Ardkhean, R.; Roth, P. M. C.; Maksymowicz, R. M.; Curran, A.; Peng, Q.; Paton, R. S.; Fletcher, S. P. Enantioselective Conjugate Addition Catalyzed by a Copper Phosphoramidite Complex: Computational and Experimental Exploration of Asymmetric Induction. *ACS Catal.* **2017**, *7*, 6729−6737.
(25) Ardkhean, R.; Mortimore, M.; Paton, R. S.; Fletcher, S. P. Formation of Quaternary Centres by Copper Catalysed Asymmetric Conjugate Addition to β-Substituted Cyclopentenones with the Aid of a Quantitative Structure-Selectivity Relationship. *Chem. Sci.* **2018**, *9*, 2628−2632.
(26) Reid, J. P.; Sigman, M. S. Comparing Quantitative Prediction Methods for the Discovery of Small-Molecule Chiral Catalysts. *Nat. Rev. Chem.* **2018**, *2*, 290−305.
(27) Durand, D. J.; Fey, N. Computational Ligand Descriptors for Catalyst Design. *Chem. Rev.* **2019**, *119*, 6561−6594.
(28) Parveen, R.; Cundari, T. R.; Younker, J. M.; Rodriguez, G.; McCullough, L. DFT and QSAR Studies of Ethylene Polymerization by Zirconocene Catalysts. *ACS Catal.* **2019**, *9*, 9339−9349.
(29) Poree, C.; Schoenebeck, F. A Holy Grail in Chemistry: Computational Catalyst Design: Feasible or Fiction? *Acc. Chem. Res.* **2017**, *50*, 605−608.
(30) Harvey, J. N.; Himo, F.; Maseras, F.; Perrin, L. Scope and Challenge of Computational Methods for Studying Mechanism and Reactivity in Homogeneous Catalysis. *ACS Catal.* **2019**, *9*, 6803−6813.
(31) Duan, C.; Janet, J. P.; Liu, F.; Nandy, A.; Kulik, H. J. Learning from Failure: Predicting Electronic Structure Calculation Outcomes with Machine Learning Models. *J. Chem. Theory Comput.* **2019**, *15*, 2331−2345.
(32) Dimitrov, T.; Kreisbeck, C.; Becker, J. S.; Aspuru-Guzik, A.; Saikin, S. K. Autonomous Molecular Design: Then and Now. *ACS Appl. Mater. Interfaces* **2019**, *11*, 24825−24836.
(33) Dewyer, A. L.; Argüelles, A. J.; Zimmerman, P. M. Methods for Exploring Reaction Space in Molecular Systems. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, No. e1354.
(34) Takayanagi, T.; Nakatomi, T. Automated Reaction Path Searches for Spin-Forbidden Reactions. *J. Comput. Chem.* **2018**, *39*, 1319−1326.
(35) Ismail, I.; Stuttaford-Fowler, H. B. V. A.; Ochan Ashok, C.; Robertson, C.; Habershon, S. Automatic Proposal of Multistep Reaction Mechanisms Using a Graph-Driven Search. *J. Phys. Chem. A* **2019**, *123*, 3407−3417.
(36) Rappoport, D.; Aspuru-Guzik, A. Predicting Feasible Organic Reaction Pathways Using Heuristically Aided Quantum Chemistry. *J. Chem. Theory Comput.* **2019**, *15*, 4099−4112.
(37) Maeda, S.; Harabuchi, Y. On Benchmarking of Automated Methods for Performing Exhaustive Reaction Path Search. *J. Chem. Theory Comput.* **2019**, *15*, 2111−2115.
(38) Martínez-Núñez, E. An Automated Method to Find Transition States Using Chemical Dynamics Simulations. *J. Comput. Chem.* **2015**, *36*, 222−234.
(39) Rodriguez, A.; Rodriguez-Fernandez, R.; Vazquez, S. A.; Barnes, G. A.; Stewart, J. J. P.; Martinez-Nunez, E. tsscds2018: A Code for Automated Discovery of Chemical Reaction Mechanisms and Solving the Kinetics. *J. Comput. Chem.* **2018**, *39*, 1922−1930.
(40) Kopec, S.; Martínez-Núñez, E.; Soto, J.; Peláez, D. vdW-TSSCDS—an Automated and Global Procedure for the Computation of Stationary Points on Intermolecular Potential Energy Surfaces. *Int. J. Quantum Chem.* **2019**, *119*, No. e26008.
(41) Avendaño-Franco, G.; Romero, A. H. Firefly Algorithm for Structural Search. *J. Chem. Theory Comput.* **2016**, *12*, 3416−3428.
(42) Payne, A.; Avendaño-Franco, G.; Bousquet, E.; Romero, A. H. Firefly Algorithm Applied to Noncollinear Magnetic Phase Materials Prediction. *J. Chem. Theory Comput.* **2018**, *14*, 4455−4466.
(43) Jimenez-Izal, E.; Alexandrova, A. N. Computational Design of Clusters for Catalysis. *Annu. Rev. Phys. Chem.* **2018**, *69*, 377−400.
(44) Swain, M. C.; Cole, J. M. ChemDataExtractor: A Toolkit for Automated Extraction of Chemical Information from the Scientific Literature. *J. Chem. Inf. Model.* **2016**, *56*, 1894−1904.
(45) Weston, L.; Tshitoyan, V.; Dagdelen, J.; Kononova, O.; Trewartha, A.; Persson, K. A.; Ceder, G.; Jain, A. Named Entity Recognition and Normalization Applied to Large-Scale Information Extraction from the Materials Science Literature. *J. Chem. Inf. Model.* **2019**, *59*, 3692−3702.
(46) Hastings, J.; Chepelev, L.; Willighagen, E.; Adams, N.; Steinbeck, C.; Dumontier, M. The Chemical Information Ontology: Provenance and Disambiguation for Chemical Data on the Biological Semantic Web. *PLoS One* **2011**, *6*, No. e25513.
(47) Bird, C. L.; Frey, J. G. Chemical Information Matters: An e-Research Perspective on Information and Data Sharing in the Chemical Sciences. *Chem. Soc. Rev.* **2013**, *42*, 6754−6776.
(48) Ess, D.; Gagliardi, L.; Hammes-Schiffer, S. Introduction: Computational Design of Catalysts from Molecules to Materials. *Chem. Rev.* **2019**, *119*, 6507−6508.
(49) Zahrt, A. F.; Athavale, S. V.; Denmark, S. E. Quantitative Structure-Selectivity Relationships in Enantioselective Catalysis: Past, Present, and Future. *Chem. Rev.* [Online], **2019**. DOI: 10.1021/acs.chemrev.9b00425 (accessed January 08, 2020).
(50) Freeze, J. G.; Kelly, H. R.; Batista, V. S. Search for Catalysts by Inverse Design: Artificial Intelligence, Mountain Climbers, and Alchemists. *Chem. Rev.* **2019**, *119*, 6595−6612.
(51) Weymuth, T.; Reiher, M. Inverse Quantum Chemistry: Concepts and Strategies for Rational Compound Design. *Int. J. Quantum Chem.* **2014**, *114*, 823−837.
(52) Kuhn, C.; Beratan, D. N. Inverse Strategies for Molecular Design. *J. Phys. Chem.* **1996**, *100*, 10595−10599.
(53) Marder, S. R.; Beratan, D. N.; Cheng, L. T. Approaches for Optimizing the First Electronic Hyperpolarizability of Conjugated Organic Molecules. *Science* **1991**, *252*, 103−106.

(54) Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse Molecular Design Using Machine Learning: Generative Models for Matter Engineering. *Science* **2018**, *361*, 360−365.
(55) von Lilienfeld, O. A.; Lins, R. D.; Röthlisberger, U. Variational Particle Number Approach for Rational Compound Design. *Phys. Rev. Lett.* **2005**, *95*, 153002.
(56) Schneider, G. Virtual Screening: An Endless Staircase? *Nat. Rev. Drug Discovery* **2010**, *9*, 273−276.
(57) The terms "scoring function", "fitness function", and "figure of merit" are often used interchangeably, although domain-specific preferences also exist. "Scoring function" has its roots in drug design, where the quality of the bioactive ligand−target binding is often evaluated as an empirical score that represents the binding affinity or binding energy. The term "fitness function" is mostly used in conjunction with genetic algorithms, where it represents the mathematical expression that produces a numerical value (i.e., the fitness) used in ranking the candidates.
(58) Hay, B. P.; Firman, T. K. Hostdesigner: A Program for the De Novo Structure-Based Design of Molecular Receptors with Binding Sites That Complement Metal Ion Guests. *Inorg. Chem.* **2002**, *41*, 5502−5512.
(59) Rosales, A. R.; Wahlers, J.; Limé, E.; Meadows, R. E.; Leslie, K. W.; Savin, R.; Bell, F.; Hansen, E.; Helquist, P.; Munday, R. H.; Wiest, O.; Norrby, P.-O. Rapid Virtual Screening of Enantioselective Catalysts Using CatVS. *Nat. Catal.* **2019**, *2*, 41−45.
(60) Hansen, E.; Rosales, A. R.; Tutkowski, B.; Norrby, P.-O.; Wiest, O. Prediction of Stereochemistry Using Q2MM. *Acc. Chem. Res.* **2016**, *49*, 996−1005.
(61) Rosales, A. R.; Quinn, T. R.; Wahlers, J.; Tomberg, A.; Zhang, X.; Helquist, P.; Wiest, O.; Norrby, P.-O. Application of Q2MM to Predictions in Stereoselective Synthesis. *Chem. Commun.* **2018**, *54*, 8294−8311.
(62) Donoghue, P. J.; Helquist, P.; Norrby, P.-O.; Wiest, O. Development of a Q2MM Force Field for the Asymmetric Rhodium Catalyzed Hydrogenation of Enamides. *J. Chem. Theory Comput.* **2008**, *4*, 1313−1323.
(63) Donoghue, P. J.; Helquist, P.; Norrby, P.-O.; Wiest, O. Prediction of Enantioselectivity in Rhodium Catalyzed Hydrogenations. *J. Am. Chem. Soc.* **2009**, *131*, 410−411.
(64) Guan, Y.; Wheeler, S. E. Automated Quantum Mechanical Predictions of Enantioselectivity in a Rhodium-Catalyzed Asymmetric Hydrogenation. *Angew. Chem., Int. Ed.* **2017**, *56*, 9101−9105.
(65) Rooks, B. J.; Haas, M. R.; Sepúlveda, D.; Lu, T.; Wheeler, S. E. Prospects for the Computational Design of Bipyridine *N,N′*-Dioxide Catalysts for Asymmetric Propargylation Reactions. *ACS Catal.* **2015**, *5*, 272−280.
(66) Doney, A. C.; Rooks, B. J.; Lu, T.; Wheeler, S. E. Design of Organocatalysts for Asymmetric Propargylations through Computational Screening. *ACS Catal.* **2016**, *6*, 7948−7955.
(67) Guan, Y.; Ingman, V. M.; Rooks, B. J.; Wheeler, S. E. AARON: An Automated Reaction Optimizer for New Catalysts. *J. Chem. Theory Comput.* **2018**, *14*, 5249−5261.
(68) Hartenfeller, M.; Zettl, H.; Walter, M.; Rupp, M.; Reisen, F.; Proschak, E.; Weggen, S.; Stark, H.; Schneider, G. DOGS: Reaction-Driven De Novo Design of Bioactive Compounds. *PLoS Comput. Biol.* **2012**, *8*, No. e1002380.
(69) Hartenfeller, M.; Schneider, G. Enabling Future Drug Discovery by De Novo Design. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 742−759.
(70) Schneider, G.; Clark, D. E. Automated De Novo Drug Design: Are We Nearly There Yet? *Angew. Chem., Int. Ed.* **2019**, *58*, 10792−10803.
(71) Schneider, G.; Fechner, U. Computer-Based De Novo Design of Drug-Like Molecules. *Nat. Rev. Drug Discovery* **2005**, *4*, 649−663.
(72) Reymond, J.-L. The Chemical Space Project. *Acc. Chem. Res.* **2015**, *48*, 722−730.
(73) Lameijer, E.-W.; Bäck, T.; Kok, J. N.; Ijzerman, A. P. Evolutionary Algorithms in Drug Design. *Nat. Comput.* **2005**, *4*, 177−243.
(74) Hartenfeller, M.; Proschak, E.; Schüller, A.; Schneider, G. Concept of Combinatorial De Novo Design of Drug-Like Molecules by Particle Swarm Optimization. *Chem. Biol. Drug Des.* **2008**, *72*, 16−26.
(75) Winter, R.; Montanari, F.; Steffen, A.; Briem, H.; Noé, F.; Clevert, D.-A. Efficient Multi-Objective Molecular Optimization in a Continuous Latent Space. *Chem. Sci.* **2019**, *10*, 8016−8024.
(76) Reutlinger, M.; Rodrigues, T.; Schneider, P.; Schneider, G. Multi-Objective Molecular De Novo Design by Adaptive Fragment Prioritization. *Angew. Chem., Int. Ed.* **2014**, *53*, 4244−4248.
(77) Hiss, J. A.; Reutlinger, M.; Koch, C. P.; Perna, A. M.; Schneider, P.; Rodrigues, T.; Haller, S.; Folkers, G.; Weber, L.; Baleeiro, R. B.; Walden, P.; Wrede, P.; Schneider, G. Combinatorial Chemistry by Ant Colony Optimization. *Future Med. Chem.* **2014**, *6*, 267−280.
(78) Bandyopadhyay, S.; Saha, S.; Maulik, U.; Deb, K. A Simulated Annealing-Based Multiobjective Optimization Algorithm: AMOSA. *IEEE Trans. Evol. Comput.* **2008**, *12*, 269−283.
(79) Pearlman, D. A.; Murcko, M. A. CONCERTS: Dynamic Connection of Fragments as an Approach to De Novo Ligand Design. *J. Med. Chem.* **1996**, *39*, 1651−1663.
(80) Chu, Y.; Heyndrickx, W.; Occhipinti, G.; Jensen, V. R.; Alsberg, B. K. An Evolutionary Algorithm for De Novo Optimization of Functional Transition Metal Compounds. *J. Am. Chem. Soc.* **2012**, *134*, 8885−8895.
(81) Schwab, P.; France, M. B.; Ziller, J. W.; Grubbs, R. H. A Series of Well-Defined Metathesis Catalysts - Synthesis of $RuCl_2(=CHR')(PR_3)_2$ and Its Reactions. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 2039−2041.
(82) Nguyen, S. T.; Johnson, L. K.; Grubbs, R. H.; Ziller, J. W. Ring Opening Metathesis Polymerization (ROMP) of Norbornene by a Group VIII Carbene Complex in Protic Media. *J. Am. Chem. Soc.* **1992**, *114*, 3974−3975.
(83) Scholl, M.; Ding, S.; Lee, C. W.; Grubbs, R. H. Synthesis and Activity of a New Generation of Ruthenium-Based Olefin Metathesis Catalysts Coordinated with 1,3-Dimesityl-4,5-Dihydroimidazol-2-ylidene Ligands. *Org. Lett.* **1999**, *1*, 953−956.
(84) Huang, J. K.; Stevens, E. D.; Nolan, S. P.; Petersen, J. L. Olefin Metathesis-Active Ruthenium Complexes Bearing a Nucleophilic Carbene Ligand. *J. Am. Chem. Soc.* **1999**, *121*, 2674−2678.
(85) Foscato, M.; Occhipinti, G.; Venkatraman, V.; Alsberg, B. K.; Jensen, V. R. Automated Design of Realistic Organometallic Molecules from Fragments. *J. Chem. Inf. Model.* **2014**, *54*, 767−780.
(86) Khersonsky, O.; Röthlisberger, D.; Wollacott, A. M.; Murphy, P.; Dym, O.; Albeck, S.; Kiss, G.; Houk, K. N.; Baker, D.; Tawfik, D. S. Optimization of the in-Silico-Designed Kemp Eliminase KE70 by Computational Design and Directed Evolution. *J. Mol. Biol.* **2011**, *407*, 391−412.
(87) Vaissier Welborn, V.; Head-Gordon, T. Computational Design of Synthetic Enzymes. *Chem. Rev.* **2019**, *119*, 6613−6630.
(88) Hilvert, D. Design of Protein Catalysts. *Annu. Rev. Biochem.* **2013**, *82*, 447−470.
(89) Dahiyat, B. I.; Mayo, S. L. Protein Design Automation. *Protein Sci.* **1996**, *5*, 895−903.
(90) Dahiyat, B. I.; Mayo, S. L. De Novo Protein Design: Fully Automated Sequence Selection. *Science* **1997**, *278*, 82−87.
(91) Hediger, M. R.; De Vico, L.; Svendsen, A.; Besenmatter, W.; Jensen, J. H. A Computational Methodology to Screen Activities of Enzyme Variants. *PLoS One* **2012**, *7*, No. e49849.
(92) Hediger, M. R.; De Vico, L.; Rannes, J. B.; Jäckel, C.; Besenmatter, W.; Svendsen, A.; Jensen, J. H. *In Silico* Screening of 393 Mutants Facilitates Enzyme Engineering of Amidase Activity in CalB. *PeerJ* **2013**, *1*, No. e145.
(93) Khersonsky, O.; Lipsh, R.; Avizemer, Z.; Ashani, Y.; Goldsmith, M.; Leader, H.; Dym, O.; Rogotner, S.; Trudeau, D. L.; Prilusky, J.; Amengual-Rigo, P.; Guallar, V.; Tawfik, D. S.; Fleishman, S. J. Automated Design of Efficient and Functionally Diverse Enzyme Repertoires. *Mol. Cell* **2018**, *72*, 178−186.
(94) Hediger, M. R.; Steinmann, C.; De Vico, L.; Jensen, J. H. A Computational Method for the Systematic Screening of Reaction Barriers in Enzymes: Searching for Bacillus Circulans Xylanase Mutants

https://dx.doi.org/10.1021/acscatal.9b04952
ACS Catal. 2020, 10, 2354−2377

with Greater Activity Towards a Synthetic Substrate. *PeerJ* **2013**, *1*, No. e111.
(95) Lapidoth, G.; Khersonsky, O.; Lipsh, R.; Dym, O.; Albeck, S.; Rogotner, S.; Fleishman, S. J. Highly Active Enzymes by Automated Combinatorial Backbone Assembly and Sequence Design. *Nat. Commun.* **2018**, *9*, 2780.
(96) Tantillo, D. J.; Jiangang, C.; Houk, K. N. Theozymes and Compuzymes: Theoretical Models for Biological Catalysis. *Curr. Opin. Chem. Biol.* **1998**, *2*, 743−750.
(97) Bolon, D. N.; Mayo, S. L. Enzyme-Like Proteins by Computational Design. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 14274.
(98) Malisi, C.; Kohlbacher, O.; Höcker, B. Automated Scaffold Selection for Enzyme Design. *Proteins: Struct., Funct., Genet.* **2009**, *77*, 74−83.
(99) Zanghellini, A.; Jiang, L.; Wollacott, A. M.; Cheng, G.; Meiler, J.; Althoff, E. A.; Röthlisberger, D.; Baker, D. New Algorithms and an in Silico Benchmark for Computational Enzyme Design. *Protein Sci.* **2006**, *15*, 2785−2794.
(100) Kiss, G.; Çelebi-Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K. N. Computational Enzyme Design. *Angew. Chem., Int. Ed.* **2013**, *52*, 5700−5725.
(101) Baker, D. An Exciting but Challenging Road Ahead for Computational Enzyme Design. *Protein Sci.* **2010**, *19*, 1817−1819.
(102) Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D. Kemp Elimination Catalysts by Computational Enzyme Design. *Nature* **2008**, *453*, 190−195.
(103) Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F.; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D. De Novo Computational Design of Retro-Aldol Enzymes. *Science* **2008**, *319*, 1387−1391.
(104) Khare, S. D.; Kipnis, Y.; Greisen, P., Jr.; Takeuchi, R.; Ashani, Y.; Goldsmith, M.; Song, Y.; Gallaher, J. L.; Silman, I.; Leader, H.; Sussman, J. L.; Stoddard, B. L.; Tawfik, D. S.; Baker, D. Computational Redesign of a Mononuclear Zinc Metalloenzyme for Organophosphate Hydrolysis. *Nat. Chem. Biol.* **2012**, *8*, 294−300.
(105) Blomberg, R.; Kries, H.; Pinkas, D. M.; Mittl, P. R. E.; Grütter, M. G.; Privett, H. K.; Mayo, S. L.; Hilvert, D. Precision Is Essential for Efficient Catalysis in an Evolved Kemp Eliminase. *Nature* **2013**, *503*, 418.
(106) Khersonsky, O.; Kiss, G.; Röthlisberger, D.; Dym, O.; Albeck, S.; Houk, K. N.; Baker, D.; Tawfik, D. S. Bridging the Gaps in Design Methodologies by Evolutionary Optimization of the Stability and Proficiency of Designed Kemp Eliminase KE59. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 10358−10363.
(107) Amrein, B. A.; Steffen-Munsberg, F.; Szeler, I.; Purg, M.; Kulkarni, Y.; Kamerlin, S. C. L. CADEE: Computer-Aided Directed Evolution of Enzymes. *IUCrJ* **2017**, *4*, 50−64.
(108) Weymuth, T.; Reiher, M. Gradient-Driven Molecule Construction: An Inverse Approach Applied to the Design of Small-Molecule Fixating Catalysts. *Int. J. Quantum Chem.* **2014**, *114*, 838−850.
(109) Krausbeck, F.; Sobez, J.-G.; Reiher, M. Stabilization of Activated Fragments by Shell-Wise Construction of an Embedding Environment. *J. Comput. Chem.* **2017**, *38*, 1023−1038.
(110) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. Electrostatic Basis for Enzyme Catalysis. *Chem. Rev.* **2006**, *106*, 3210−3235.
(111) Sokalski, W. A. Theoretical Model for Exploration of Catalytic Activity of Enzymes and Design of New Catalysts: CO_2 Hydration Reaction. *Int. J. Quantum Chem.* **1981**, *20*, 231−240.
(112) Dittner, M.; Hartke, B. Globally Optimal Catalytic Fields - Inverse Design of Abstract Embeddings for Maximum Reaction Rate Acceleration. *J. Chem. Theory Comput.* **2018**, *14*, 3547−3564.
(113) Wang, M.; Hu, X.; Beratan, D. N.; Yang, W. Designing Molecules by Optimizing Potentials. *J. Am. Chem. Soc.* **2006**, *128*, 3228−3232.
(114) Xiao, D.; Yang, W.; Beratan, D. N. Inverse Molecular Design in a Tight-Binding Framework. *J. Chem. Phys.* **2008**, *129*, 044106.
(115) Hu, X.; Beratan, D. N.; Yang, W. A Gradient-Directed Monte Carlo Approach to Molecular Design. *J. Chem. Phys.* **2008**, *129*, 064102.
(116) Chang, A. M.; Rudshteyn, B.; Warnke, I.; Batista, V. S. Inverse Design of a Catalyst for Aqueous CO/CO_2 Conversion Informed by the Ni^{II}-Iminothiolate Complex. *Inorg. Chem.* **2018**, *57*, 15474−15480.
(117) von Lilienfeld, O. A.; Tuckerman, M. E. Molecular Grand-Canonical Ensemble Density Functional Theory and Exploration of Chemical Space. *J. Chem. Phys.* **2006**, *125*, 154104.
(118) von Lilienfeld, O. A. First Principles View on Chemical Compound Space: Gaining Rigorous Atomistic Control of Molecular Properties. *Int. J. Quantum Chem.* **2013**, *113*, 1676−1689.
(119) Marcon, V.; von Lilienfeld, O. A.; Andrienko, D. Tuning Electronic Eigenvalues of Benzene Via Doping. *J. Chem. Phys.* **2007**, *127*, 064305.
(120) Sheppard, D.; Henkelman, G.; von Lilienfeld, O. A. Alchemical Derivatives of Reaction Energetics. *J. Chem. Phys.* **2010**, *133*, 084104.
(121) von Lilienfeld, O. A.; Tuckerman, M. E. Alchemical Variations of Intermolecular Energies According to Molecular Grand-Canonical Ensemble Density Functional Theory. *J. Chem. Theory Comput.* **2007**, *3*, 1083−1090.
(122) Fias, S.; Chang, K. Y. S.; von Lilienfeld, O. A. Alchemical Normal Modes Unify Chemical Space. *J. Phys. Chem. Lett.* **2019**, *10*, 30−39.
(123) Saravanan, K.; Kitchin, J. R.; von Lilienfeld, O. A.; Keith, J. A. Alchemical Predictions for Computational Catalysis: Potential and Limitations. *J. Phys. Chem. Lett.* **2017**, *8*, 5002−5007.
(124) Griego, C. D.; Saravanan, K.; Keith, J. A. Benchmarking Computational Alchemy for Carbide, Nitride, and Oxide Catalysts. *Adv. Theor. Simul.* **2019**, *2*, 1800142.
(125) Chang, K. Y. S.; von Lilienfeld, O. A. $Al_xGa_{1-x}As$ Crystals with Direct 2 eV Band Gaps from Computational Alchemy. *Phys. Rev. Mater.* **2018**, *2*, 073802.
(126) Anatole von Lilienfeld, O. Accurate Ab Initio Energy Gradients in Chemical Compound Space. *J. Chem. Phys.* **2009**, *131*, 164102.
(127) Chang, K. Y. S.; Fias, S.; Ramakrishnan, R.; von Lilienfeld, O. A. Fast and Accurate Predictions of Covalent Bonds in Chemical Space. *J. Chem. Phys.* **2016**, *144*, 174110.
(128) Xiao, D.; Martini, L. A.; Snoeberger, R. C.; Crabtree, R. H.; Batista, V. S. Inverse Design and Synthesis of acac-Coumarin Anchors for Robust TiO_2 Sensitization. *J. Am. Chem. Soc.* **2011**, *133*, 9014−9022.
(129) Keinan, S.; Therien, M. J.; Beratan, D. N.; Yang, W. Molecular Design of Porphyrin-Based Nonlinear Optical Materials. *J. Phys. Chem. A* **2008**, *112*, 12203−12207.
(130) Sanchez-Lengeling, B.; Outeiral, C.; Guimaraes, G. L.; Aspuru-Guzik, A. Optimizing Distributions over Molecular Space. An Objective-Reinforced Generative Adversarial Network for Inverse-Design Chemistry (ORGANIC). *ChemRxiv. Preprint* [Online], **2017**. DOI: 10.26434/chemrxiv.5309668.v3 (accessed November 12, 2019).
(131) Ståhl, N.; Falkman, G.; Karlsson, A.; Mathiason, G.; Boström, J. Deep Reinforcement Learning for Multiparameter Optimization in De Novo Drug Design. *J. Chem. Inf. Model.* **2019**, *59*, 3166−3176.
(132) Putin, E.; Asadulaev, A.; Ivanenkov, Y.; Aladinskiy, V.; Sanchez-Lengeling, B.; Aspuru-Guzik, A.; Zhavoronkov, A. Reinforced Adversarial Neural Computer for De Novo Molecular Design. *J. Chem. Inf. Model.* **2018**, *58*, 1194−1204.
(133) Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci.* **2018**, *4*, 268−276.
(134) Jin, W.; Barzilay, R.; Jaakkola, T. Junction Tree Variational Autoencoder for Molecular Graph Generation. *arXiv e-prints* [Online], **2018**. https://arxiv.org/abs/1802.04364 (accessed November 07, 2019).
(135) Arús-Pous, J.; Johansson, S. V.; Prykhodko, O.; Bjerrum, E. J.; Tyrchan, C.; Reymond, J.-L.; Chen, H.; Engkvist, O. Randomized

Smiles Strings Improve the Quality of Molecular Generative Models. *J. Cheminf.* **2019**, *11*, 71.
(136) Krenn, M.; Häse, F.; Nigam, A.; Friederich, P.; Aspuru-Guzik, A. SELFIES: A Robust Representation of Semantically Constrained Graphs with an Example Application in Chemistry. *arXiv e-prints* [Online], **2019**. https://arxiv.org/abs/1905.13741 (accessed May 01, 2019).
(137) Popova, M.; Isayev, O.; Tropsha, A. Deep Reinforcement Learning for De Novo Drug Design. *Sci. Adv.* **2018**, *4*, No. eaap7885.
(138) Walters, W. P. Virtual Chemical Libraries. *J. Med. Chem.* **2019**, *62*, 1116−1124.
(139) Coley, C. W.; Rogers, L.; Green, W. H.; Jensen, K. F. SCScore: Synthetic Complexity Learned from a Reaction Corpus. *J. Chem. Inf. Model.* **2018**, *58*, 252−261.
(140) Bonnet, P. Is Chemical Synthetic Accessibility Computationally Predictable for Drug and Lead-Like Molecules? A Comparative Assessment between Medicinal and Computational Chemists. *Eur. J. Med. Chem.* **2012**, *54*, 679−689.
(141) Boda, K.; Seidel, T.; Gasteiger, J. Structure and Reaction Based Evaluation of Synthetic Accessibility. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 311−325.
(142) Allu, T. K.; Oprea, T. I. Rapid Evaluation of Synthetic and Molecular Complexity for in Silico Chemistry. *J. Chem. Inf. Model.* **2005**, *45*, 1237−1243.
(143) Podolyan, Y.; Walters, M. A.; Karypis, G. Assessing Synthetic Accessibility of Chemical Compounds Using Machine Learning Methods. *J. Chem. Inf. Model.* **2010**, *50*, 979−991.
(144) Fukunishi, Y.; Kurosawa, T.; Mikami, Y.; Nakamura, H. Prediction of Synthetic Accessibility Based on Commercially Available Compound Databases. *J. Chem. Inf. Model.* **2014**, *54*, 3259−3267.
(145) Ertl, P.; Schuffenhauer, A. Estimation of Synthetic Accessibility Score of Drug-Like Molecules Based on Molecular Complexity and Fragment Contributions. *J. Cheminf.* **2009**, *1*, 8.
(146) Huang, Q.; Li, L.-L.; Yang, S.-Y. RASA: A Rapid Retrosynthesis-Based Scoring Method for the Assessment of Synthetic Accessibility of Drug-Like Molecules. *J. Chem. Inf. Model.* **2011**, *51*, 2768−2777.
(147) Patel, H.; Bodkin, M. J.; Chen, B.; Gillet, V. J. Knowledge-Based Approach to De Novo Design Using Reaction Vectors. *J. Chem. Inf. Model.* **2009**, *49*, 1163−1184.
(148) Vinkers, H. M.; de Jonge, M. R.; Daeyaert, F. F. D.; Heeres, J.; Koymans, L. M. H.; van Lenthe, J. H.; Lewi, P. J.; Timmerman, H.; Van Aken, K.; Janssen, P. A. J. SYNOPSIS: Synthesize and Optimize System in Silico. *J. Med. Chem.* **2003**, *46*, 2765−2773.
(149) Masek, B. B.; Baker, D. S.; Dorfman, R. J.; DuBrucq, K.; Francis, V. C.; Nagy, S.; Richey, B. L.; Soltanshahi, F. Multistep Reaction Based De Novo Drug Design: Generating Synthetically Feasible Design Ideas. *J. Chem. Inf. Model.* **2016**, *56*, 605−620.
(150) Pottel, J.; Moitessier, N. Customizable Generation of Synthetically Accessible, Local Chemical Subspaces. *J. Chem. Inf. Model.* **2017**, *57*, 454−467.
(151) Hageman, J. A.; Westerhuis, J. A.; Frühauf, H.-W.; Rothenberg, G. Design and Assembly of Virtual Homogeneous Catalyst Libraries - Towards *in Silico* Catalyst Optimisation. *Adv. Synth. Catal.* **2006**, *348*, 361−369.
(152) Jover, J.; Fey, N. Screening Substituent and Backbone Effects on the Properties of Bidentate P,P-Donor Ligands (LKB-PP_{screen}). *Dalton Trans.* **2013**, *42*, 172−181.
(153) Maldonado, A. G.; Hageman, J. A.; Mastroianni, S.; Rothenberg, G. Backbone Diversity Analysis in Catalyst Design. *Adv. Synth. Catal.* **2009**, *351*, 387−396.
(154) Burk, M. J. Modular Phospholane Ligands in Asymmetric Catalysis. *Acc. Chem. Res.* **2000**, *33*, 363−372.
(155) Loch, J. A.; Crabtree, R. H. Rapid Screening and Combinatorial Methods in Homogeneous Organometallic Catalysis. *Pure Appl. Chem.* **2001**, *73*, 119−128.
(156) Harper, K. C.; Sigman, M. S. Predicting and Optimizing Asymmetric Catalyst Performance Using the Principles of Experimental Design and Steric Parameters. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 2179.
(157) Mushinski, R. M.; Squires, B. M.; Sincerbox, K. A.; Hudnall, T. W. Amino-Acrylamido Carbenes: Modulating Carbene Reactivity Via Decoration with an α,β-Unsaturated Carbonyl Moiety. *Organometallics* **2012**, *31*, 4862−4870.
(158) Iwamoto, H.; Imamoto, T.; Ito, H. Computational Design of High-Performance Ligand for Enantioselective Markovnikov Hydroboration of Aliphatic Terminal Alkenes. *Nat. Commun.* **2018**, *9*, 2290.
(159) Yum, J. H.; Park, S.; Hiraga, R.; Okamura, I.; Notsu, S.; Sugiyama, H. Modular DNA-Based Hybrid Catalysts as a Toolbox for Enantioselective Hydration of α,β-Unsaturated Ketones. *Org. Biomol. Chem.* **2019**, *17*, 2548−2553.
(160) Mater, A. C.; Coote, M. L. Deep Learning in Chemistry. *J. Chem. Inf. Model.* **2019**, *59*, 2545−2559.
(161) Goh, G. B.; Hodas, N. O.; Vishnu, A. Deep Learning for Computational Chemistry. *J. Comput. Chem.* **2017**, *38*, 1291−1307.
(162) Janet, J. P.; Chan, L.; Kulik, H. J. Accelerating Chemical Discovery with Machine Learning: Simulated Evolution of Spin Crossover Complexes with an Artificial Neural Network. *J. Phys. Chem. Lett.* **2018**, *9*, 1064−1071.
(163) Gómez-Bombarelli, R.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Duvenaud, D.; Maclaurin, D.; Blood-Forsythe, M. A.; Chae, H. S.; Einzinger, M.; Ha, D.-G.; Wu, T.; Markopoulos, G.; Jeon, S.; Kang, H.; Miyazaki, H.; Numata, M.; Kim, S.; Huang, W.; Hong, S. I.; Baldo, M.; Adams, R. P.; Aspuru-Guzik, A. Design of Efficient Molecular Organic Light-Emitting Diodes by a High-Throughput Virtual Screening and Experimental Approach. *Nat. Mater.* **2016**, *15*, 1120−1127.
(164) Williams, T.; McCullough, K.; Lauterbach, J. A. Enabling Catalyst Discovery through Machine Learning and High-Throughput Experimentation. *Chem. Mater.* **2020**, *32*, 157.
(165) Liu, R.; Wang, H.; Glover, K. P.; Feasel, M. G.; Wallqvist, A. Dissecting Machine-Learning Prediction of Molecular Activity: Is an Applicability Domain Needed for Quantitative Structure-Activity Relationship Models Based on Deep Neural Networks? *J. Chem. Inf. Model.* **2019**, *59*, 117−126.
(166) Peterson, A. A.; Christensen, R.; Khorshidi, A. Addressing Uncertainty in Atomistic Machine Learning. *Phys. Chem. Chem. Phys.* **2017**, *19*, 10978−10985.
(167) Tropsha, A. Best Practices for QSAR Model Development, Validation, and Exploitation. *Mol. Inf.* **2010**, *29*, 476−488.
(168) Kulik, H. J. Making Machine Learning a Useful Tool in the Accelerated Discovery of Transition Metal Complexes. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2020**, *10*, No. e1439.
(169) Janet, J. P.; Duan, C.; Yang, T.; Nandy, A.; Kulik, H. J. A Quantitative Uncertainty Metric Controls Error in Neural Network-Driven Chemical Discovery. *Chem. Sci.* **2019**, *10*, 7913−7922.
(170) Nandy, A.; Zhu, J.; Janet, J. P.; Duan, C.; Getman, R. B.; Kulik, H. J. Machine Learning Accelerates the Discovery of Design Rules and Exceptions in Stable Metal-Oxo Intermediate Formation. *ACS Catal.* **2019**, *9*, 8243−8255.
(171) Amar, Y.; Schweidtmann, A. M.; Deutsch, P.; Cao, L.; Lapkin, A. Machine Learning and Molecular Descriptors Enable Rational Solvent Selection in Asymmetric Catalysis. *Chem. Sci.* **2019**, *10*, 6697−6706.
(172) Santiago, C. B.; Guo, J.-Y.; Sigman, M. S. Predictive and Mechanistic Multivariate Linear Regression Models for Reaction Development. *Chem. Sci.* **2018**, *9*, 2398−2412.
(173) Lakuntza, O.; Besora, M.; Maseras, F. Searching for Hidden Descriptors in the Metal-Ligand Bond through Statistical Analysis of Density Functional Theory (DFT) Results. *Inorg. Chem.* **2018**, *57*, 14660−14670.
(174) Kim, J. Y.; Kulik, H. J. When Is Ligand pK_a a Good Descriptor for Catalyst Energetics? In Search of Optimal CO_2 Hydration Catalysts. *J. Phys. Chem. A* **2018**, *122*, 4579−4590.
(175) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.
(176) Sciabola, S.; Alex, A.; Higginson, P. D.; Mitchell, J. C.; Snowden, M. J.; Morao, I. Theoretical Prediction of the Enantiomeric Excess in Asymmetric Catalysis. An Alignment-Independent Molecular Interaction Field Based Approach. *J. Org. Chem.* **2005**, *70*, 9025−9027.

https://dx.doi.org/10.1021/acscatal.9b04952
ACS Catal. 2020, 10, 2354−2377

(177) Urbano-Cuadrado, M.; Carbó, J. J.; Maldonado, A. G.; Bo, C. New Quantum Mechanics-Based Three-Dimensional Molecular Descriptors for Use in QSSR Approaches: Application to Asymmetric Catalysis. *J. Chem. Inf. Model.* **2007**, *47*, 2228−2234.
(178) Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. GRid-INdependent Descriptors (GRIND): A Novel Class of Alignment-Independent Three-Dimensional Molecular Descriptors. *J. Med. Chem.* **2000**, *43*, 3233−3243.
(179) Lipkowitz, K. B.; D'Hue, C. A.; Sakamoto, T.; Stack, J. N. Stereocartography: A Computational Mapping Technique That Can Locate Regions of Maximum Stereoinduction around Chiral Catalysts. *J. Am. Chem. Soc.* **2002**, *124*, 14255−14267.
(180) Orlandi, M.; Coelho, J. A. S.; Hilton, M. J.; Toste, F. D.; Sigman, M. S. Parametrization of Non-Covalent Interactions for Transition State Interrogation Applied to Asymmetric Catalysis. *J. Am. Chem. Soc.* **2017**, *139*, 6803−6806.
(181) Brethomé, A. V.; Fletcher, S. P.; Paton, R. S. Conformational Effects on Physical-Organic Descriptors: The Case of Sterimol Steric Parameters. *ACS Catal.* **2019**, *9*, 2313−2323.
(182) Krenske, E. H.; Houk, K. N. Aromatic Interactions as Control Elements in Stereoselective Organic Reactions. *Acc. Chem. Res.* **2013**, *46*, 979−989.
(183) Knowles, R. R.; Jacobsen, E. N. Attractive Noncovalent Interactions in Asymmetric Catalysis: Links between Enzymes and Small Molecule Catalysts. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 20678.
(184) Neel, A. J.; Hilton, M. J.; Sigman, M. S.; Toste, F. D. Exploiting Non-Covalent π Interactions for Catalyst Design. *Nature* **2017**, *543*, 637.
(185) Wheeler, S. E.; Seguin, T. J.; Guan, Y.; Doney, A. C. Noncovalent Interactions in Organocatalysis and the Prospect of Computational Catalyst Design. *Acc. Chem. Res.* **2016**, *49*, 1061−1069.
(186) Ferreira, M. A. B.; De Jesus Silva, J.; Grosslight, S.; Fedorov, A.; Sigman, M. S.; Copéret, C. Noncovalent Interactions Drive the Efficiency of Molybdenum Imido Alkylidene Catalysts for Olefin Metathesis. *J. Am. Chem. Soc.* **2019**, *141*, 10788−10800.
(187) Fang, C.; Fantin, M.; Pan, X.; de Fiebre, K.; Coote, M. L.; Matyjaszewski, K.; Liu, P. Mechanistically Guided Predictive Models for Ligand and Initiator Effects in Copper-Catalyzed Atom Transfer Radical Polymerization (Cu-ATRP). *J. Am. Chem. Soc.* **2019**, *141*, 7486−7497.
(188) Milo, A.; Neel, A. J.; Toste, F. D.; Sigman, M. S. A Data-Intensive Approach to Mechanistic Elucidation Applied to Chiral Anion Catalysis. *Science* **2015**, *347*, 737−743.
(189) Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting Reaction Performance in C-N Cross-Coupling Using Machine Learning. *Science* **2018**, *360*, 186−190.
(190) Zahrt, A. F.; Henle, J. J.; Rose, B. T.; Wang, Y.; Darrow, W. T.; Denmark, S. E. Prediction of Higher-Selectivity Catalysts by Computer-Driven Workflow and Machine Learning. *Science* **2019**, *363*, No. eaau5631.
(191) Rasmussen, C. E.; Williams, C. K. I. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*; MIT Press: 2005.
(192) Ulissi, Z. W.; Singh, A. R.; Tsai, C.; Nørskov, J. K. Automated Discovery and Construction of Surface Phase Diagrams Using Machine Learning. *J. Phys. Chem. Lett.* **2016**, *7*, 3931−3935.
(193) Kolb, B.; Marshall, P.; Zhao, B.; Jiang, B.; Guo, H. Representing Global Reactive Potential Energy Surfaces Using Gaussian Processes. *J. Phys. Chem. A* **2017**, *121*, 2552−2557.
(194) Ulissi, Z. W.; Medford, A. J.; Bligaard, T.; Nørskov, J. K. To Address Surface Reaction Network Complexity Using Scaling Relations Machine Learning and DFT Calculations. *Nat. Commun.* **2017**, *8*, 14621.
(195) Tian, H.; Rangarajan, S. Predicting Adsorption Energies Using Multifidelity Data. *J. Chem. Theory Comput.* **2019**, *15*, 5588−5600.
(196) Gu, G. H.; Noh, J.; Kim, I.; Jung, Y. Machine Learning for Renewable Energy Materials. *J. Mater. Chem. A* **2019**, *7*, 17096−17117.
(197) Mamun, O.; Winther, K. T.; Boes, J. R.; Bligaard, T. A Bayesian Framework for Adsorption Energy Prediction on Bimetallic Alloy Catalysts. *ChemRxiv. Preprint* [Online], **2019**. DOI: 10.26434/chemrxiv.10295129.v1 (accessed January 08, 2020).
(198) Simm, G. N.; Reiher, M. Error-Controlled Exploration of Chemical Reaction Networks with Gaussian Processes. *J. Chem. Theory Comput.* **2018**, *14*, 5238−5248.
(199) Brønsted, J. N.; Pedersen, K. J. Die Katalytische Zersetzung Des Nitramids Und Ihre Physikalisch-Chemische Bedeutung. *Z. Phys. Chem.* **1924**, *108U*, 185−235.
(200) Hammett, L. P. Some Relations between Reaction Rates and Equilibrium Constants. *Chem. Rev.* **1935**, *17*, 125−136.
(201) Taft, R. W. Linear Free Energy Relationships from Rates of Esterification and Hydrolysis of Aliphatic and Ortho-Substituted Benzoate Esters. *J. Am. Chem. Soc.* **1952**, *74*, 2729−2732.
(202) Abild-Pedersen, F.; Greeley, J.; Studt, F.; Rossmeisl, J.; Munter, T. R.; Moses, P. G.; Skúlason, E.; Bligaard, T.; Nørskov, J. K. Scaling Properties of Adsorption Energies for Hydrogen-Containing Molecules on Transition-Metal Surfaces. *Phys. Rev. Lett.* **2007**, *99*, 016105.
(203) Man, I. C.; Su, H.-Y.; Calle-Vallejo, F.; Hansen, H. A.; Martínez, J. I.; Inoglu, N. G.; Kitchin, J.; Jaramillo, T. F.; Nørskov, J. K.; Rossmeisl, J. Universality in Oxygen Evolution Electrocatalysis on Oxide Surfaces. *ChemCatChem* **2011**, *3*, 1159−1165.
(204) Calle-Vallejo, F.; Martínez, J. I.; García-Lastra, J. M.; Rossmeisl, J.; Koper, M. T. M. Physical and Chemical Nature of the Scaling Relations between Adsorption Energies of Atoms on Metal Surfaces. *Phys. Rev. Lett.* **2012**, *108*, 116103.
(205) Busch, M.; Wodrich, M. D.; Corminboeuf, C. Linear Scaling Relationships and Volcano Plots in Homogeneous Catalysis - Revisiting the Suzuki Reaction. *Chem. Sci.* **2015**, *6*, 6754−6761.
(206) Sawatlon, B.; Wodrich, M. D.; Corminboeuf, C. Unraveling Metal/Pincer Ligand Effects in the Catalytic Hydrogenation of Carbon Dioxide to Formate. *Organometallics* **2018**, *37*, 4568−4575.
(207) Meyer, B.; Sawatlon, B.; Heinen, S.; von Lilienfeld, O. A.; Corminboeuf, C. Machine Learning Meets Volcano Plots: Computational Discovery of Cross-Coupling Catalysts. *Chem. Sci.* **2018**, *9*, 7069−7077.
(208) Wodrich, M. D.; Sawatlon, B.; Solel, E.; Kozuch, S.; Corminboeuf, C. Activity-Based Screening of Homogeneous Catalysts through the Rapid Assessment of Theoretically Derived Turnover Frequencies. *ACS Catal.* **2019**, *9*, 5716−5725.
(209) Wodrich, M. D.; Sawatlon, B.; Busch, M.; Corminboeuf, C. On the Generality of Molecular Volcano Plots. *ChemCatChem* **2018**, *10*, 1586−1591.
(210) Zaffran, J.; Michel, C.; Delbecq, F.; Sautet, P. Trade-Off between Accuracy and Universality in Linear Energy Relations for Alcohol Dehydrogenation on Transition Metals. *J. Phys. Chem. C* **2015**, *119*, 12988−12998.
(211) Gani, T. Z. H.; Kulik, H. J. Understanding and Breaking Scaling Relations in Single-Site Catalysis: Methane to Methanol Conversion by Fe^{IV}═O. *ACS Catal.* **2018**, *8*, 975−986.
(212) Pérez-Ramírez, J.; López, N. Strategies to Break Linear Scaling Relationships. *Nat. Catal.* **2019**, *2*, 971−976.
(213) Bligaard, T.; Nørskov, J. K.; Dahl, S.; Matthiesen, J.; Christensen, C. H.; Sehested, J. The Brønsted-Evans-Polanyi Relation and the Volcano Curve in Heterogeneous Catalysis. *J. Catal.* **2004**, *224*, 206−217.
(214) Wodrich, M. D.; Busch, M.; Corminboeuf, C. Expedited Screening of Active and Regioselective Catalysts for the Hydroformylation Reaction. *Helv. Chim. Acta* **2018**, *101*, No. e1800107.
(215) Sabatier, P. *La Catalyse En Chimie Organique*; Librarie Polytechnique: Paris, 1913.
(216) Kozuch, S.; Shaik, S. How to Conceptualize Catalytic Cycles? The Energetic Span Model. *Acc. Chem. Res.* **2011**, *44*, 101−110.
(217) Kozuch, S. A Refinement of Everyday Thinking: The Energetic Span Model for Kinetic Assessment of Catalytic Cycles. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 795−815.
(218) Kozuch, S.; Martin, J. M. L. Turning over" Definitions in Catalytic Cycles. *ACS Catal.* **2012**, *2*, 2787−2794.

https://dx.doi.org/10.1021/acscatal.9b04952
ACS Catal. 2020, 10, 2354−2377

(219) Kozuch, S. Steady State Kinetics of Any Catalytic Network: Graph Theory, the Energy Span Model, the Analogy between Catalysis and Electrical Circuits, and the Meaning of "Mechanism. *ACS Catal.* **2015**, *5*, 5242−5255.
(220) Uhe, A.; Kozuch, S.; Shaik, S. Automatic Analysis of Computed Catalytic Cycles. *J. Comput. Chem.* **2011**, *32*, 978−985.
(221) Huber, R.; Passera, A.; Gubler, E.; Mezzetti, A. P-Stereogenic PN(H)P Iron(II) Catalysts for the Asymmetric Hydrogenation of Ketones: The Importance of Non-Covalent Interactions in Rational Ligand Design by Computation. *Adv. Synth. Catal.* **2018**, *360*, 2900−2913.
(222) Gaggioli, C. A.; Stoneburner, S. J.; Cramer, C. J.; Gagliardi, L. Beyond Density Functional Theory: The Multiconfigurational Approach to Model Heterogeneous Catalysis. *ACS Catal.* **2019**, *9*, 8481−8502.
(223) Grimme, S. A General Quantum Mechanically Derived Force Field (QMDFF) for Molecules and Condensed Phase Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 4497−4514.
(224) Horton, J. T.; Allen, A. E. A.; Dodda, L. S.; Cole, D. J. QUBEKit: Automating the Derivation of Force Field Parameters from Quantum Mechanics. *J. Chem. Inf. Model.* **2019**, *59*, 1366−1381.
(225) Prampolini, G.; Campetella, M.; De Mitri, N.; Livotto, P. R.; Cacelli, I. Systematic and Automated Development of Quantum Mechanically Derived Force Fields: The Challenging Case of Halogenated Hydrocarbons. *J. Chem. Theory Comput.* **2016**, *12*, 5525−5540.
(226) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A* **2001**, *105*, 9396−9409.
(227) Senftle, T. P.; Hong, S.; Islam, M. M.; Kylasa, S. B.; Zheng, Y.; Shin, Y. K.; Junkermeier, C.; Engel-Herbert, R.; Janik, M. J.; Aktulga, H. M.; Verstraelen, T.; Grama, A.; van Duin, A. C. T. The ReaxFF Reactive Force-Field: Development, Applications and Future Directions. *Npj Comput. Mater.* **2016**, *2*, 15011.
(228) Nakata, H.; Bai, S. Development of a New Parameter Optimization Scheme for a Reactive Force Field Based on a Machine Learning Approach. *J. Comput. Chem.* **2019**, *40*, 2000−2012.
(229) Furman, D.; Carmeli, B.; Zeiri, Y.; Kosloff, R. Enhanced Particle Swarm Optimization Algorithm: Efficient Training of ReaxFF Reactive Force Fields. *J. Chem. Theory Comput.* **2018**, *14*, 3100−3112.
(230) Norrby, P.-O.; Liljefors, T. Automated Molecular Mechanics Parameterization with Simultaneous Utilization of Experimental and Quantum Mechanical Data. *J. Comput. Chem.* **1998**, *19*, 1146−1166.
(231) Jensen, F. Locating Minima on Seams of Intersecting Potential Energy Surfaces. An Application to Transition Structure Modeling. *J. Am. Chem. Soc.* **1992**, *114*, 1596−1603.
(232) Weill, N.; Corbeil, C. R.; De Schutter, J. W.; Moitessier, N. Toward a Computational Tool Predicting the Stereochemical Outcome of Asymmetric Reactions: Development of the Molecular Mechanics-Based Program ACE and Application to Asymmetric Epoxidation Reactions. *J. Comput. Chem.* **2011**, *32*, 2878−2889.
(233) Lin, H.; Zhao, Y.; Tishchenko, O.; Truhlar, D. G. Multi-configuration Molecular Mechanics Based on Combined Quantum Mechanical and Molecular Mechanical Calculations. *J. Chem. Theory Comput.* **2006**, *2*, 1237−1254.
(234) Åqvist, J.; Warshel, A. Simulation of Enzyme Reactions Using Valence Bond Force Fields and Other Hybrid Quantum/Classical Approaches. *Chem. Rev.* **1993**, *93*, 2523−2544.
(235) Eksterowicz, J. E.; Houk, K. N. Transition-State Modeling with Empirical Force Fields. *Chem. Rev.* **1993**, *93*, 2439−2461.
(236) Norrby, P.-O.; Rasmussen, T.; Haller, J.; Strassner, T.; Houk, K. N. Rationalizing the Stereoselectivity of Osmium Tetroxide Asymmetric Dihydroxylations with Transition State Modeling Using Quantum Mechanics-Guided Molecular Mechanics. *J. Am. Chem. Soc.* **1999**, *121*, 10186−10192.
(237) Fristrup, P.; Tanner, D.; Norrby, P.-O. Updating the Asymmetric Osmium-Catalyzed Dihydroxylation (AD) Mnemonic: Q2MM Modeling and New Kinetic Measurements. *Chirality* **2003**, *15*, 360−368.
(238) Fristrup, P.; Jensen, G. H.; Andersen, M. L. N.; Tanner, D.; Norrby, P.-O. Combining Q2MM Modeling and Kinetic Studies for Refinement of the Osmium-Catalyzed Asymmetric Dihydroxylation (AD) Mnemonic. *J. Organomet. Chem.* **2006**, *691*, 2182−2198.
(239) Norrby, P.-O.; Brandt, P.; Rein, T. Rationalization of Product Selectivities in Asymmetric Horner−Wadsworth−Emmons Reactions by Use of a New Method for Transition-State Modeling. *J. Org. Chem.* **1999**, *64*, 5845−5852.
(240) Rasmussen, T.; Norrby, P.-O. Modeling the Stereoselectivity of the β-Amino Alcohol-Promoted Addition of Dialkylzinc to Aldehydes. *J. Am. Chem. Soc.* **2003**, *125*, 5130−5138.
(241) Lee, J. M.; Zhang, X.; Norrby, P.-O.; Helquist, P.; Wiest, O. Stereoselectivity in (Acyloxy)Borane-Catalyzed Mukaiyama Aldol Reactions. *J. Org. Chem.* **2016**, *81*, 5314−5321.
(242) Rydberg, P.; Hansen, S. M.; Kongsted, J.; Norrby, P.-O.; Olsen, L.; Ryde, U. Transition-State Docking of Flunitrazepam and Progesterone in Cytochrome P450. *J. Chem. Theory Comput.* **2008**, *4*, 673−681.
(243) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227−249.
(244) Lin, H.; Truhlar, D. G. QM/MM: What Have We Learned, Where Are We, and Where Do We Go from Here? *Theor. Chem. Acc.* **2007**, *117*, 185.
(245) Bo, C.; Maseras, F. QM/MM Methods in Inorganic Chemistry. *Dalton Trans.* **2008**, 2911−2919.
(246) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.
(247) van der Kamp, M. W.; Mulholland, A. J. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry* **2013**, *52*, 2708−2728.
(248) Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H.-B.; Ding, L.; Morokuma, K. The ONIOM Method and Its Applications. *Chem. Rev.* **2015**, *115*, 5678−5796.
(249) Yoshimura, T.; Maeda, S.; Taketsugu, T.; Sawamura, M.; Morokuma, K.; Mori, S. Exploring the Full Catalytic Cycle of Rhodium(I)-BINAP-Catalysed Isomerisation of Allylic Amines: A Graph Theory Approach for Path Optimisation. *Chem. Sci.* **2017**, *8*, 4475−4488.
(250) Donald, S. M. A.; Vidal-Ferran, A.; Maseras, F. A DFT/MM Analysis of the Effect of Ligand Substituents on Asymmetric Hydrogenation Catalyzed by Rhodium Complexes with Phosphine-Phosphinite Ligands. *Can. J. Chem.* **2009**, *87*, 1273−1279.
(251) Cao, L.; Ryde, U. On the Difference between Additive and Subtractive QM/MM Calculations. *Front. Chem.* **2018**, *6*, 1.
(252) Maseras, F.; Morokuma, K. IMOMM: A New Integrated Ab Initio + Molecular Mechanics Geometry Optimization Scheme of Equilibrium Structures and Transition States. *J. Comput. Chem.* **1995**, *16*, 1170−1179.
(253) Quesne, M. G.; Borowski, T.; de Visser, S. P. Quantum Mechanics/Molecular Mechanics Modeling of Enzymatic Processes: Caveats and Breakthroughs. *Chem. - Eur. J.* **2016**, *22*, 2562−2581.
(254) Karelina, M.; Kulik, H. J. Systematic Quantum Mechanical Region Determination in QM/MM Simulation. *J. Chem. Theory Comput.* **2017**, *13*, 563−576.
(255) Maeda, S.; Ohno, K. Lowest Transition State for the Chirality-Determining Step in Ru((R)-BINAP)-Catalyzed Asymmetric Hydrogenation of Methyl-3-Oxobutanoate. *J. Am. Chem. Soc.* **2008**, *130*, 17228−17229.
(256) Ho, M.-H.; Rousseau, R.; Roberts, J. A. S.; Wiedner, E. S.; Dupuis, M.; DuBois, D. L.; Bullock, R. M.; Raugei, S. Ab Initio-Based Kinetic Modeling for the Design of Molecular Catalysts: The Case of H_2 Production Electrocatalysts. *ACS Catal.* **2015**, *5*, 5436−5452.
(257) Jover, J.; Maseras, F., QM/MM Calculations on Selectivity in Homogeneous Catalysis. In *Computational Studies in Organometallic Chemistry*; Macgregor, S. A., Eisenstein, O., Eds.; Springer International Publishing: Cham, 2016; pp 59−79.

https://dx.doi.org/10.1021/acscatal.9b04952
ACS Catal. 2020, 10, 2354−2377

(258) Wu, X.-P.; Gagliardi, L.; Truhlar, D. G. Multilink F* Method for Combined Quantum Mechanical and Molecular Mechanical Calculations of Complex Systems. *J. Chem. Theory Comput.* **2019**, *15*, 4208−4217.
(259) Christensen, A. S.; Kubař, T.; Cui, Q.; Elstner, M. Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications. *Chem. Rev.* **2016**, *116*, 5301−5337.
(260) Minenkov, Y.; Sharapa, D. I.; Cavallo, L. Application of Semiempirical Methods to Transition Metal Complexes: Fast Results but Hard-to-Predict Accuracy. *J. Chem. Theory Comput.* **2018**, *14*, 3428−3439.
(261) Varela, J. A.; Vázquez, S. A.; Martínez-Núñez, E. An Automated Method to Find Reaction Mechanisms and Solve the Kinetics in Organometallic Catalysis. *Chem. Sci.* **2017**, *8*, 3843−3851.
(262) Børve, K. J.; Jensen, V. R.; Karlsen, T.; Støvneng, J. A.; Swang, O. Evaluation of PM3(tm) as a Geometry Generator in Theoretical Studies of Transition-Metal-Based Catalysts for Polymerizing Olefins. *J. Mol. Model.* **1997**, *3*, 193−202.
(263) Grimme, S.; Bannwarth, C.; Shushkov, P. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1−86). *J. Chem. Theory Comput.* **2017**, *13*, 1989−2009.
(264) Bursch, M.; Neugebauer, H.; Grimme, S. Structure Optimisation of Large Transition-Metal Complexes with Extended Tight-Binding Methods. *Angew. Chem., Int. Ed.* **2019**, *58*, 11078−11087.
(265) Bursch, M.; Hansen, A.; Grimme, S. Fast and Reasonable Geometry Optimization of Lanthanoid Complexes with an Extended Tight Binding Quantum Chemical Method. *Inorg. Chem.* **2017**, *56*, 12485−12491.
(266) Grimme, S. Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *J. Chem. Theory Comput.* **2019**, *15*, 2847−2862.
(267) Behler, J. Perspective: Machine Learning Potentials for Atomistic Simulations. *J. Chem. Phys.* **2016**, *145*, 170901.
(268) Yao, K.; Herr, J. E.; Toth, D. W.; McKintyre, R.; Parkhill, J. The TensorMol-0.1 Model Chemistry: A Neural Network Augmented with Long-Range Physics. *Chem. Sci.* **2018**, *9*, 2261−2269.
(269) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem. Sci.* **2017**, *8*, 3192−3203.
(270) Unke, O. T.; Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *J. Chem. Theory Comput.* **2019**, *15*, 3678−3693.
(271) Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. Approaching Coupled Cluster Accuracy with a General-Purpose Neural Network Potential through Transfer Learning. *Nat. Commun.* **2019**, *10*, 2903.
(272) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and Use of Quantum Mechanical Molecular Models. 76. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(273) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. *J. Mol. Model.* **2007**, *13*, 1173−1213.
(274) Lee, K.; Yoo, D.; Jeong, W.; Han, S. SIMPLE-NN: An Efficient Package for Training and Executing Neural-Network Interatomic Potentials. *Comput. Phys. Commun.* **2019**, *242*, 95−103.
(275) Behler, J. Constructing High-Dimensional Neural Network Potentials: A Tutorial Review. *Int. J. Quantum Chem.* **2015**, *115*, 1032−1050.
(276) Nicolaou, C. A.; Brown, N. Multi-Objective Optimization Methods in Drug Design. *Drug Discovery Today: Technol.* **2013**, *10*, No. e427.
(277) Bailey, G. A.; Foscato, M.; Higman, C. S.; Day, C. S.; Jensen, V. R.; Fogg, D. E. Bimolecular Coupling as a Vector for Decomposition of Fast-Initiating Olefin Metathesis Catalysts. *J. Am. Chem. Soc.* **2018**, *140*, 6931−6944.
(278) Daeyaert, F.; Deem, M. W. A Pareto Algorithm for Efficient De Novo Design of Multi-Functional Molecules. *Mol. Inf.* **2017**, *36*, 1600044.
(279) Cummins, D. J.; Bell, M. A. Integrating Everything: The Molecule Selection Toolkit, a System for Compound Prioritization in Drug Discovery. *J. Med. Chem.* **2016**, *59*, 6999−7010.
(280) Nicolaou, C. A.; Kannas, C.; Loizidou, E. Multi-Objective Optimization Methods in De Novo Drug Design. *Mini-Rev. Med. Chem.* **2012**, *12*, 979−987.
(281) Raccuglia, P.; Elbert, K. C.; Adler, P. D. F.; Falk, C.; Wenny, M. B.; Mollo, A.; Zeller, M.; Friedler, S. A.; Schrier, J.; Norquist, A. J. Machine-Learning-Assisted Materials Discovery Using Failed Experiments. *Nature* **2016**, *533*, 73.
(282) Tabor, D. P.; Roch, L. M.; Saikin, S. K.; Kreisbeck, C.; Sheberla, D.; Montoya, J. H.; Dwaraknath, S.; Aykol, M.; Ortiz, C.; Tribukait, H.; Amador-Bedolla, C.; Brabec, C. J.; Maruyama, B.; Persson, K. A.; Aspuru-Guzik, A. Accelerating the Discovery of Materials for Clean Energy in the Era of Smart Automation. *Nat. Rev. Mater.* **2018**, *3*, 5−20.
(283) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation. *APL Mater.* **2013**, *1*, 011002.
(284) Curtarolo, S.; Setyawan, W.; Wang, S.; Xue, J.; Yang, K.; Taylor, R. H.; Nelson, L. J.; Hart, G. L. W.; Sanvito, S.; Buongiorno-Nardelli, M.; Mingo, N.; Levy, O. AFLOWLIB.ORG: A Distributed Materials Properties Repository from High-Throughput *Ab Initio* Calculations. *Comput. Mater. Sci.* **2012**, *58*, 227−235.
(285) Materials Cloud. https://www.materialscloud.org (accessed November 12, 2019).
(286) Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM* **2013**, *65*, 1501−1509.
(287) Curtarolo, S.; Hart, G. L. W.; Nardelli, M. B.; Mingo, N.; Sanvito, S.; Levy, O. The High-Throughput Highway to Computational Materials Design. *Nat. Mater.* **2013**, *12*, 191.
(288) Montoya, J. H.; Persson, K. A. A High-Throughput Framework for Determining Adsorption Energies on Solid Surfaces. *Npj Comput. Mater.* **2017**, *3*, 14.
(289) Morgante, P.; Peverati, R. ACCDB: A Collection of Chemistry Databases for Broad Computational Purposes. *J. Comput. Chem.* **2019**, *40*, 839−848.
(290) Gómez-Bombarelli, R.; Aspuru-Guzik, A., Machine Learning and Big-Data in Computational Chemistry. In *Handbook of Materials Modeling: Methods: Theory and Modeling*; Andreoni, W., Yip, S., Eds.; Springer International Publishing: Cham, 2018; pp 1−24.
(291) Bo, C.; Maseras, F.; López, N. The Role of Computational Results Databases in Accelerating the Discovery of Catalysts. *Nat. Catal.* **2018**, *1*, 809−810.
(292) Álvarez-Moreno, M.; de Graaf, C.; López, N.; Maseras, F.; Poblet, J. M.; Bo, C. Managing the Computational Chemistry Big Data Problem: The ioChem-BD Platform. *J. Chem. Inf. Model.* **2015**, *55*, 95−103.
(293) Winther, K. T.; Hoffmann, M. J.; Boes, J. R.; Mamun, O.; Bajdich, M.; Bligaard, T. Catalysis-Hub.org, an Open Electronic Structure Database for Surface Reactions. *Sci. Data* **2019**, *6*, 75.
(294) Pizzi, G.; Cepellotti, A.; Sabatini, R.; Marzari, N.; Kozinsky, B. AiiDA: Automated Interactive Infrastructure and Database for Computational Science. *Comput. Mater. Sci.* **2016**, *111*, 218−230.
(295) Zapata, F.; Ridder, L.; Hidding, J.; Jacob, C. R.; Infante, I.; Visscher, L. QMflows: A Tool Kit for Interoperable Parallel Workflows in Quantum Chemistry. *J. Chem. Inf. Model.* **2019**, *59*, 3191−3197.
(296) Curtarolo, S.; Setyawan, W.; Hart, G. L. W.; Jahnatek, M.; Chepulskii, R. V.; Taylor, R. H.; Wang, S.; Xue, J.; Yang, K.; Levy, O.; Mehl, M. J.; Stokes, H. T.; Demchenko, D. O.; Morgan, D. AFLOW: An

https://dx.doi.org/10.1021/acscatal.9b04952
ACS Catal. 2020, 10, 2354−2377

Automatic Framework for High-Throughput Materials Discovery. *Comput. Mater. Sci.* **2012**, *58*, 218−226.
(297) Adorf, C. S.; Dodd, P. M.; Ramasubramani, V.; Glotzer, S. C. Simple Data and Workflow Management with the Signac Framework. *Comput. Mater. Sci.* **2018**, *146*, 220−229.
(298) Jain, A.; Ong, S. P.; Chen, W.; Medasani, B.; Qu, X.; Kocher, M.; Brafman, M.; Petretto, G.; Rignanese, G.-M.; Hautier, G.; Gunter, D.; Persson, K. A. FireWorks: A Dynamic Workflow System Designed for High-Throughput Applications. *Concurrency Comput. Pract. Ex.* **2015**, *27*, 5037−5059.
(299) Mathew, K.; Montoya, J. H.; Faghaninia, A.; Dwarakanath, S.; Aykol, M.; Tang, H.; Chu, I.-h.; Smidt, T.; Bocklund, B.; Horton, M.; Dagdelen, J.; Wood, B.; Liu, Z.-K.; Neaton, J.; Ong, S. P.; Persson, K.; Jain, A. Atomate: A High-Level Interface to Generate, Execute, and Analyze Computational Materials Science Workflows. *Comput. Mater. Sci.* **2017**, *139*, 140−152.
(300) *Maestro*; Schrödinger, LLC: New York, NY, 2019.
(301) *MacroModel*; Schrödinger, LLC: New York, NY, 2019.
(302) Q2MM. https://github.com/Q2MM/q2mm (accessed November 12, 2019).
(303) AARON. https://github.com/QChASM/Aaron (accessed November 12, 2019).
(304) Corbeil, C. R.; Thielges, S.; Schwartzentruber, J. A.; Moitessier, N. Toward a Computational Tool Predicting the Stereochemical Outcome of Asymmetric Reactions: Development and Application of a Rapid and Accurate Program Based on Organic Principles. *Angew. Chem., Int. Ed.* **2008**, *47*, 2635−2638.
(305) Therrien, E.; Englebienne, P.; Arrowsmith, A. G.; Mendoza-Sanchez, R.; Corbeil, C. R.; Weill, N.; Campagna-Slater, V.; Moitessier, N. Integrating Medicinal Chemistry, Organic/Combinatorial Chemistry, and Computational Chemistry for the Discovery of Selective Estrogen Receptor Modulators with Forecaster, a Novel Platform for Drug Discovery. *J. Chem. Inf. Model.* **2012**, *52*, 210−224.
(306) Molecular Forecaster. https://www.molecularforecaster.com/ (accessed November 12, 2019).
(307) Ioannidis, E. I.; Gani, T. Z. H.; Kulik, H. J. molSimplify: A Toolkit for Automating Discovery in Inorganic Chemistry. *J. Comput. Chem.* **2016**, *37*, 2106−2117.
(308) Janet, J. P.; Liu, F.; Nandy, A.; Duan, C.; Yang, T.; Lin, S.; Kulik, H. J. Designing in the Face of Uncertainty: Exploiting Electronic Structure and Machine Learning Models for Discovery in Inorganic Chemistry. *Inorg. Chem.* **2019**, *58*, 10592−10606.
(309) Nandy, A.; Duan, C.; Janet, J. P.; Gugler, S.; Kulik, H. J. Strategies and Software for Machine Learning Accelerated Discovery in Transition Metal Chemistry. *Ind. Eng. Chem. Res.* **2018**, *57*, 13973−13986.
(310) Kim, J. Y.; Steeves, A. H.; Kulik, H. J. Harnessing Organic Ligand Libraries for First-Principles Inorganic Discovery: Indium Phosphide Quantum Dot Precursor Design Strategies. *Chem. Mater.* **2017**, *29*, 3632−3643.
(311) Gani, T. Z. H.; Ioannidis, E. I.; Kulik, H. J. Computational Discovery of Hydrogen Bond Design Rules for Electrochemical Ion Separation. *Chem. Mater.* **2016**, *28*, 6207−6218.
(312) Janet, J. P.; Gani, T. Z. H.; Steeves, A. H.; Ioannidis, E. I.; Kulik, H. J. Leveraging Cheminformatics Strategies for Inorganic Discovery: Application to Redox Potential Design. *Ind. Eng. Chem. Res.* **2017**, *56*, 4898−4910.
(313) molSimplify. https://github.com/hjkgrp/molSimplify (accessed November 12, 2019).
(314) Foscato, M.; Venkatraman, V.; Jensen, V. R. DENOPTIM: Software for Computational De Novo Design of Organic and Inorganic Molecules. *J. Chem. Inf. Model.* **2019**, *59*, 4077−4082.
(315) Foscato, M.; Venkatraman, V.; Occhipinti, G.; Alsberg, B. K.; Jensen, V. R. Automated Building of Organometallic Complexes from 3D Fragments. *J. Chem. Inf. Model.* **2014**, *54*, 1919−31.
(316) Abburu, S.; Venkatraman, V.; Alsberg, B. K. TD-DFT Based Fine-Tuning of Molecular Excitation Energies Using Evolutionary Algorithms. *RSC Adv.* **2016**, *6*, 3661−3670.
(317) Venkatraman, V.; Foscato, M.; Jensen, V. R.; Alsberg, B. K. Evolutionary De Novo Design of Phenothiazine Derivatives for Dye-Sensitized Solar Cells. *J. Mater. Chem. A* **2015**, *3*, 9851−9860.
(318) Venkatraman, V.; Abburu, S.; Alsberg, B. K. Artificial Evolution of Coumarin Dyes for Dye Sensitized Solar Cells. *Phys. Chem. Chem. Phys.* **2015**, *17*, 27672−27682.
(319) Venkatraman, V.; Alsberg, B. K. Designing High-Refractive Index Polymers Using Materials Informatics. *Polymers* **2018**, *10*, 103.
(320) Venkatraman, V.; Gupta, M.; Foscato, M.; Svendsen, H. F.; Jensen, V. R.; Alsberg, B. K. Computer-Aided Molecular Design of Imidazole-Based Absorbents for CO_2 Capture. *Int. J. Greenhouse Gas Control* **2016**, *49*, 55−63.
(321) Foscato, M.; Houghton, B. J.; Occhipinti, G.; Deeth, R. J.; Jensen, V. R. Ring Closure to Form Metal Chelates in 3D Fragment-Based De Novo Design. *J. Chem. Inf. Model.* **2015**, *55*, 1844−1856.
(322) Bernhardt, P. V.; Bilyj, J. K.; Brosius, V.; Chernyshov, D.; Deeth, R. J.; Foscato, M.; Jensen, V. R.; Mertes, N.; Riley, M. J.; Törnroos, K. W. Spin Crossover in a Hexaamineiron(II) Complex: Experimental Confirmation of a Computational Prediction. *Chem. - Eur. J.* **2018**, *24*, 5082−5085.
(323) DENOPTIM. https://github.com/denoptim-project/DENOPTIM (accessed November 12, 2019).
(324) Dieterich, J. M.; Hartke, B. OGOLEM: Global Cluster Structure Optimisation for Arbitrary Mixtures of Flexible Molecules. A Multiscaling, Object-Oriented Approach. *Mol. Phys.* **2010**, *108*, 279−291.
(325) Seifert, G. Tight-Binding Density Functional Theory: An Approximate Kohn−Sham DFT Scheme. *J. Phys. Chem. A* **2007**, *111*, 5609−5613.
(326) Oliveira, A. F.; Seifert, G.; Heine, T.; Duarte, H. A. Density-Functional Based Tight-Binding: An Approximate DFT Method. *J. Braz. Chem. Soc.* **2009**, *20*, 1193−1205.
(327) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—an Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652−1671.
(328) Sperger, T.; Sanhueza, I. A.; Schoenebeck, F. Computation and Experiment: A Powerful Combination to Understand and Predict Reactivities. *Acc. Chem. Res.* **2016**, *49*, 1311−1319.
(329) Greenaway, R. L.; Santolini, V.; Bennison, M. J.; Alston, B. M.; Pugh, C. J.; Little, M. A.; Miklitz, M.; Eden-Rump, E. G. B.; Clowes, R.; Shakil, A.; Cuthbertson, H. J.; Armstrong, H.; Briggs, M. E.; Jelfs, K. E.; Cooper, A. I. High-Throughput Discovery of Organic Cages and Catenanes Using Computational Screening Fused with Robotic Synthesis. *Nat. Commun.* **2018**, *9*, 2849.
(330) Patrascu, M. B.; Pottel, J.; Pinus, S.; Bezanson, M.; Norrby, P.-O.; Moitessier, N. From Desktop to Benchtop − A Paradigm Shift in Asymmetric Synthesis. **2019**, *ChemRxiv* Preprint (Online). DOI: 10.26434/chemrxiv.9758558.v2. (accessed January 27, 2020).

NOTE ADDED IN PROOF

Very recently, the Forecaster user interface[305] has been expanded to include a new platform, Virtual Chemist,[330] of which Asymmetric Catalyst Evaluation (ACE)[304] is now part.

https://dx.doi.org/10.1021/acscatal.9b04952
ACS Catal. 2020, 10, 2354−2377