Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Engineering

Accelerating the structure search of catalysts with machine learning

Eric Musa[1,2], Francis Doherty[1,2] and Bryan R Goldsmith[1,2]

Identifying the structure of heterogeneous catalysts is a critical step to model and understand catalytic reactions and structure-property relations. Computational predictions of catalyst structures, which thus far have been dominated by computationally expensive quantum mechanical methods such as density functional theory, are now being augmented by machine learning (ML) to accelerate the structure-search of catalysts. Further development of structure-search workflows using ML will enable modeling of larger, more complex catalyst structures, aiding catalyst understanding and design. Herein, we highlight applications of ML to accelerate the search of catalyst structures and discuss on-going challenges.

Addresses
[1] Department of Chemical Engineering, University of Michigan, Ann Arbor, MI 48109-2136, USA

[2] Catalysis Science and Technology Institute, University of Michigan, Ann Arbor, MI 48109-2136, USA

Corresponding author:

Current Opinion in Chemical Engineering 2022, **35**:xx–yy

This review comes from a themed issue on **Reaction engineering and catalysis**

Edited by **Lars Grabow** and **Ryan Hartman**

https://doi.org/10.1016/j.coche.2021.100771

Introduction

Heterogeneous catalysts are widely used to catalyze reactions in chemical conversion, emissions control, and energy generation [1,2]. Catalytic properties (e.g. activity, selectivity, and stability) are inherently linked to catalyst structure, including composition, morphology, and size [3]. Many atomistic modeling efforts have focused on understanding reactions and identifying active sites and structure-property relationships of heterogeneous catalysts [4,5]. However, catalyst structures are difficult to know *a priori*. Catalyst morphology can change under reaction conditions, particularly with a support and in the presence of adsorbates, which complicates identification of the active site and structure-property relations.

Finding representative structures of catalysts is an important first step to study and predict catalytic performance and to interpret experiments. As a result, developing structure-search algorithms has been an important area of research. For example, genetic algorithms (GAs) [6,7] are often used in conjunction with quantum mechanical (QM) methods to predict stable catalyst structures, represented by local minima and the global minimum on the potential energy surface (PES). Structure-search algorithms can combine with *ab initio* thermodynamics [8] or Monte Carlo simulations [9,10] to study catalysts under reaction conditions. Yet, accurately predicting the thermodynamically favored structures of catalysts is difficult due to the relatively high computational expense of accurate QM methods and the high dimensionality of the PES.

To address these challenges, researchers are utilizing fast, inexpensive machine learning (ML) calculations with structure-search workflows to complement QM calculations and expedite the identification of relevant catalytic structures [11,12,13•,14••]. For example, Figure 1 shows a representative ML-assisted structure search workflow. Structure-search algorithms (Figure 1a) can combine with surrogate ML models (Figure 1b) trained on data generated by high-fidelity QM calculations to rapidly predict energies and forces of structures with relatively high accuracy (Figure 1c), enabling researchers to model larger, more complex catalytic systems. Once catalyst structures are identified, reaction mechanisms can be predicted to test mechanistic hypotheses, examine plausible active sites, and compare with experiment (Figure 1d).

Herein, we discuss applications of using ML to aid modeling of catalyst structures. We focus on surrogate ML models such as High Dimensional Neural Network Potentials (HDNNPs) [15,16•] and Gaussian Approximation Potentials (GAPs) [17,18] used with structure-search algorithms to identify the structures of catalytic systems. We also discuss challenges in using ML for structure-search and optimization, specifically creating high-quality training sets, controlling ML model prediction error, and the high computational cost of modeling multi-component systems. State-of-the-art techniques such as transfer learning [19,20] and active learning [21,22] are discussed as promising approaches to resolve some of these challenges. Transition state searches [23,24] and the determination of structures from spectroscopic data have also been assisted by ML [25], but these topics are beyond the