Deep Reaction Network Exploration at a Heterogeneous Catalytic Interface

Qiyuan Zhao[†], Yinan Xu[†], Jeffrey P. Greeley,[*] and Brett M. Savoie[*]

Davidson School of Chemical Engineering, Purdue University, West Lafayette, IN, 47906

E-mail: jgreeley@purdue.edu; bsavoie@purdue.edu

Abstract

Characterizing the reaction energies and reaction barriers of complex reaction networks is central to catalyst development and optimization. Nevertheless, heterogeneous catalytic surfaces pose several unique challenges to automatic reaction network characterization, including large system sizes and open-ended reactant lists, that make *ad hoc* network construction and characterization the current state-of-the-art. Here we show how automated algorithms for exploring and characterizing reaction networks can be adapted to the constraints of heterogeneous systems using ethylene oligomerization on silica-supported single site Ga^{3+} catalysts as a model system. Using only graph-based rules for exploring the network and elementary constraints based on activation energy and system size for identifying network terminations, a comprehensive reaction network was generated for this system and validated against standard methods. The automated algorithm (re)discovers the classic Cossee-Arlman mechanism for this system that is hypothesized to drive major product formation while remarkably also predicting several new pathways for producing alkanes and coke precursors. This demonstration represents the largest heterogeneous catalyst (more than 50 atoms, with an open-ended

pool of reactants) to be characterized using a quantum chemistry-based automated reaction method.

1 Introduction

Establishing the kinetic details of complex reaction networks is central to understanding heterogeneous catalytic surfaces.[1–3] The development of such networks for new systems is often painstaking, even when good hypotheses exist for the governing reactions and cycles.[4,5] Nevertheless, this sort of domain knowledge is often outpaced by the increased synthetic and experimental throughput that are driving exploratory catalyst development. The time and cost of generating reaction data for new systems are thus impediments to interpreting catalyst performance, rationalizing structure-function relationships, and leveraging burgeoning (and data demanding) machine learning approaches to catalyst development. For these reasons it is urgent to develop computational methods to accelerate and automate the exploration, characterization, and refinement of complex reaction networks at surfaces.

In the context of heterogeneous catalysis, computational methods are relatively mature for characterizing the transition states of targeted reactions,[6–12] performing microkinetic modeling on established reaction networks, and using descriptor-based methods for improving catalysts.[13–16] However, a central challenge in characterizing new catalytic interfaces lies in establishing the kinetically relevant reaction network, which is often based on intuition and can be time-consuming and error prone to characterize *ad hoc*.[4,5] Indeed, even seemingly simple heterogeneous reactions, like methane activation on metal oxide surfaces, can be decomposed into multiple elementary steps.[17,18] Furthermore, catalytic cycles can involve many intermediates or even open ended reactant lists such that brute force enumeration and characterization are infeasible. Such examples include the oxidative coupling of methane and olefin oligomerization, each involving the forma-

tion/dissociation of long carbon backbones as intermediates and an open set of olefins as potentially adsorbed reactants.[19,20]

For these reasons, the recent advent of automated reaction prediction approaches is potentially promising for elucidating reaction networks involving heterogeneous interfaces.[21–23] These methods can be categorized on the basis of whether the potential energy surface (PES) is explored in detail to locate transition states or whether the reaction networks are enumerated using a closed set of reaction templates. The latter class includes packages like Network Generation (NetGen),[24] and Reaction Mechanism Generator (RMG);[25] however, due to the reliance on established reaction templates, this strategy is less relevant to characterizing exploratory catalysts where such data is typically absent. In contrast, methods that directly explore the PES circumvent this limitation, at least in principle. This class includes several approaches that are under active development, including the artificial force induced reaction (AFIR) method,[26] stochastic surface walking reaction sampling (SSW),[27] the ZStucture method from the Zimmerman group,[28] and Yet Another Reaction Program (YARP),[29] our recently developed methodology. All of these approaches are intrinsically more expensive than template-based methods because they sample the PES (e.g., using quantum chemistry calculations), which has been a major obstacle to applying them to inhomogeneous systems in an exploratory context. For example, both SSW and AFIR have been applied to successfully (re)discover the relatively simple heterogeneous water-gas shift reaction occurring at a copper surface.[30,31] Nevertheless, this required millions of density functional theory (DFT) gradient calls, despite the small reactive system sizes. As the number of elementary steps grows and the complexity increases, a highly efficient reaction exploration scheme becomes even more indispensable to mitigate computational costs. Moreover, heterogeneous applications have several other technical obstacles to applying automated approaches that were designed for molecular systems. These include the larger system sizes that are typical of surface models; the occurrence of spectator atoms that do not participate in reactions but nevertheless play important non-covalent

or structural roles in the reaction pathways; and the use of periodic versus molecular models of the reacting systems. The optimal manner of addressing these obstacles are all outstanding research questions.

Here, we show how these problems can be addressed by combining YARP, a graph-based reaction exploration scheme, with a cluster model of a reactive interface. Ethylene oligomerization on silica-supported single site Ga^{3+} catalysts is used as a benchmark system for this approach based on the fact that some reaction data exists for this system while it still exhibits several unaccounted for product pathways. In particular, it has been previously observed that single site Ga^{3+} performs oligomerization chemistry via the classic Cossee-Arlman mechanism with reasonably high selectivity to short linear alpha-olefins; whereas, side products, such as light alkanes and coke, have also been detected.[32] Here, YARP not only (re)discovers the 1-butene-centered Cossee-Arlman catalytic cycle, but also discovers TSs responsible for side reactions, such as the formation of isomers of 1-butene, odd-number oligomers, alkanes, and coke. Moreover, this represents the largest heterogeneous cataytic system (more than 50 atoms, with an open-ended pool of reactants) to be characterized using a quantum chemistry based automated reaction method. The reactions that are discovered by YARP are generalized based on mechanisms, including carbon-backbone lengthening, oligomer liberation, and hydrogen transfer to form alkanes. The kinetic significance of the TSs are further analyzed based on the energy surfaces of three representative catalytic cycles comprising 36 elementary steps.

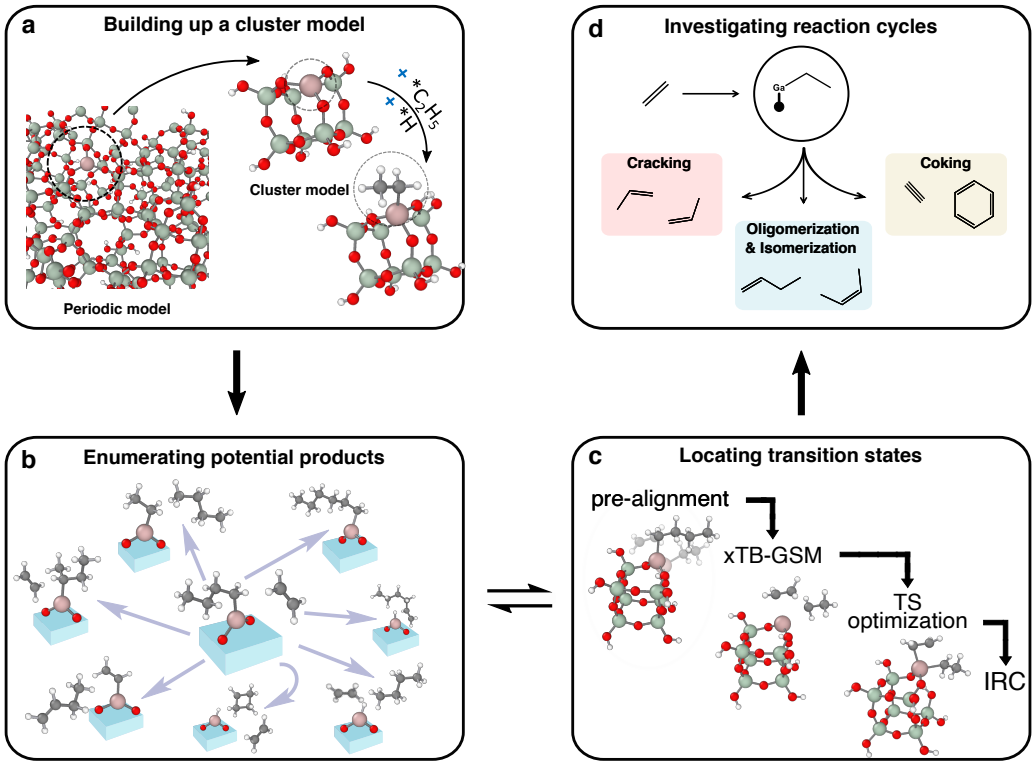

Figure 1: Overview of automated reaction network characterizing applied to ethylene oligomerization on single site Ga^{3+} catalysts supported on silica. (a) A cluster model of a Ga^{3+} single site is built from a conventional periodic model. (b) Possible products are recursively enumerated from reactants/intermediates following the elementary reaction steps on the cluster model. (c) A series characterizations is applied to each enumerated reaction to locate and characterize transition states. (d) Once the network exploration recursion terminates, detailed reaction mechanisms and relevant reaction cycles are summarized.

2 Method

2.1 Cluster Model Construction

Ethylene oligomerization on single-site Ga^{3+}/SiO_2 was modeled based on a $Si_8O_{12}(OH)_8$ cluster that was adapted from Ugliengo et al.[33] A Ga^{3+} single-site was created by substitution of a Si-OH moiety with a Ga atom. The cluster model can be viewed as a finite portion of the solid silica, with

the dangling oxygen atoms passivated by hydrogen atoms. The localized nature of oxides and the Ga^{3+} center make the cluster model a credible approximation for assessing surface reactivity.[33,34] Comparisons between the energies and barrier heights calculated on periodic surfaces and the cluster model were used to validate this assumption. Based on an earlier study, initiation of a Ga-ethyl site from a bare single site and gaseous ethylene was expected to be facile, and a low barrier Ga-ethyl-centered Cossee-Arlman mechanism has been observed. Such active intermediates could easily form when catalysts are treated with hydrogen or ethylene gas.[32] Thus, here the activated Ga-ethyl site with excess ethylene was treated as the starting reactant for network exploration. The Ga-ethyl site was created in the model cluster by adding an ethyl group to the Ga site and a proton to the adjacent oxygen atom to maintain charge balance (Fig. 1a).

2.2 Reaction Network Characterization

The recently developed YARP methodology was used to enumerate the reactions and characterize the transitions states associated with the Ga-ethyl species modeled in the presence of excess ethylene. For a more detailed description of the YARP methodology we direct readers to our previous publication.[29] In the following sections we focus on the modifications that were implemented to the reaction enumeration and reaction pathway construction steps to adapt YARP to explore ethylene oligomerization on single-site Ga^{3+}/SiO_2.

Product enumeration

The YARP methodology consists of recursively applying graph-based elementary reaction steps (ERS) of the form break m bonds and form n bonds (bmfn). These rules are sufficiently generic to recapitulate many reactions without relying on explicit reaction templates and they define reaction spaces that can be comprehensively explored (e.g., all b2f2 pathways of a given set of reactants is a well-defined set). For neutral closed-shell systems, the simplest reaction that yields

non-trivial closed-shell products is b2f2 (e.g., an E2 reaction); however, single-step b3f3 reactions might also be both thermodynamically and kinetically accessible (e.g. Diels-Alder reaction and Claisen rearrangement). Here we applied a compromise scheme, including all b2f2 reactions and the subset of b3f3 reactions involving at least one double bond breaking. These ERSs were applied to the gallium, carbon, and hydrogen atoms attached to carbon in the cluster model (shown as pink, gray, and white balls in Fig. 1b) to enumerate all products for each reactant in the network. Reactions that did not involve Ga (e.g., non-catalytic reactions between ethane and other alkyl products) and reactions that yielded species with C>5 were discarded from consideration.

Transition state localization

After product enumeration, YARP attempts to localize transition states (TSs) for each reaction. This consists of initializing a reaction geometry, estimating the transition state at the semi-empirical GFN2-xTB[35] level using the growing string method (GSM),[36] transition state optimization at the DFT level using Berny optimization, and intrinsic reaction coordinate (IRC) calculations to classify the resulting transition states (Fig. 1c). For the geometry initialization, the joint-optimization algorithm reported in the original YARP publication was retained, with the exception that the position of silica atoms (except the two oxygen atoms attached to gallium) were fixed to preserve the initial DFT-level cluster structure. These structures were then used as the fixed endpoints for GSM calculations, and after convergence, the highest energy node along the reaction pathway was selected as the initial guess for an unconstrained DFT level Berny transition state optimization. The final TS (i.e. after successful convergence of previous steps with a structure exhibiting a single imaginary frequency) was characterized by an IRC calculation to ensure its correspondence to the attempted reaction. When the two end nodes obtained by the IRC calculation matched the input reactant and product, the attempted reaction was classified as an "intended" reaction and was included in the reaction network.

Reaction network construction

To construct the reaction network, interleaved product enumeration and transition state localization was performed as described in the previous sections until exhausting the discovery of new reactions. At each stage of this iteration, the Ga-products of the previous iteration served as potential reactants for the next iteration subject to conditions that were designed to manage the size of the reaction network while being relatively permissive in terms of exploring new reactivities. Specifically, Ga-species were only included as potential reactants at the next iteration if they were connected to the rest of the network by an intended reaction with an activation energy less than 3 eV ($\sim$ 70 kcal/mol). Additionally, the size of the reactant species attached to the gallium site was limited to butyl and smaller to avoid the trivial growth of the network due to lengthening of the carbon backbone. All of the Ga intermediates obtained without violating these constraints were included as species capable of participating in reactions in the next iteration. At each iteration, the set of explored reactants consisted of all combinations of the active Ga-species and any free olefins that were produced as products during previous iterations of exploration. Thus, a newly generated Ga-species would participate in up to $n+1$ separate reactant sets, where n is the collection of free olefins discovered up until that point of exploration and the additional one corresponds to considering unimolecular reactions involving the Ga-species. Reactant combinations involving more than six carbons were discarded to avoid trivial growth of the network. For each set of reactants, the ERS generated reactions were characterized and the recursion ended after no new reactions were discovered. We note that it is possible for a Ga-species to fail the activation energy constraint at an early iteration, but then to be included later if an alternative pathway is discovered.

2.3 Periodic DFT calculations

The reaction energies and reaction barriers for a subset of pathways were recalculated on a amorphous silica slab model with a large unit cell (21.6 Å× 21.6 Å× 34.5 Å) and compared with cluster model results for validation. These calculations were performed on the amorphous structure reported by Floryan, generated from an annealing process using classical molecular dynamics and multiple dehydration processes, which results in siloxane rings with different sizes.[37] The Ga-ethyl moiety was created using the same approach as was employed in cluster models. Previous studies indicate that the less-constrained, three-coordinated Ga sites are responsible for the oligomerization chemistry, whereas the constrained four-coordinated sites are relatively inactive due to steric hindrance effects.[32] Therefore, the periodic slab calculations focused on the less-constrained Ga site.

2.4 Computational Details

In the present study, YARP used Gaussian 16 as the reference quantum chemistry engine for the DFT calculations associated with the Berny optimizations and IRC calculations.[38] Calculations were performed at the B3LYP/6-31G level of theory during network exploration, while reaction energies and reaction barriers were refined at the B3LYP-D3/6-311G(d,p) level of theory for validation and comparison with periodic calculations. The GSM calculations were performed by the pyGSM package using eleven images, fixed reactant and product geometries, and other default hyperparameters.[39] All GFN2-xTB calculations were performed with the xTB program (version 6.2.3).[35]

Periodic DFT calculations were performed using Vienna Ab-initio Simulation Package (VASP, 5.4.1), where plane-wave basis sets expanded the Kohn-Sham orbitals, and the Kohn-Sham equations were solved self-consistently.[40–44] The BEEF-VdW exchange-correlation functional with pro-

jector augmented wave (PAW) pseudopotentials was employed.[44–46] A Monkhorst-Pack k-sampling was used, and a k point grid of $2 \times 2 \times 1$ was applied. A cutoff energy of 400 eV and a force-convergence criterion of 20 meV Å^{-1} for energy local minima were used. The climbing image nudged-elastic-band (CI-NEB) method was used as a first step to locate transition states.[47,48] Seven images were used in each NEB calculation as generated by the Image Dependent Pair Potential (IDPP) tool.[49] Following each NEB calculation, Lanczos diagonalization was used to identify the transition state with a greater accuracy.[50] The force-convergence criterion of a transition state optimization was 20 meV Å^{-1}. All energies are reported with respect to the ground state energy of Ga-ethyl plus a gaseous ethylene molecule.

3 Results and Discussion

3.1 Deep reaction network constructed by YARP

The overall reaction network that was generated by YARP for ethylene oligomerization on silica-supported single site Ga^{3+} is shown in Figure 2. Network exploration was initialized with the Ga-ethyl species (node 0 in Fig. 2), which has been proposed as a key intermediate in Cossee-Arlman ethylene oligomerization cycle.[32] After a single-step of reaction enumeration and TS characterization, Ga-n-butyl, Ga-vinyl + ethane, and Ga-hydride + 1-butene, were identified as intended products of a reactions between Ga-ethyl and ethylene. The free energies of activation ($\Delta G^{\dagger}$) of forming Ga-n-butyl, Ga-vinyl, and Ga-hydride are 44.1, 59.8, and 93.5 kcal/mol, respectively. Based on its high activation energy, YARP excluded Ga-hydride from further exploration, whereas Ga-n-butyl and Ga-vinyl were included as active nodes for further reaction exploration. The high activation energy of β-hydrogen elimination forming Ga-hydride has also been observed in our previous studies using conventional periodic DFT.[32] The second step of exploration identifies Ga-

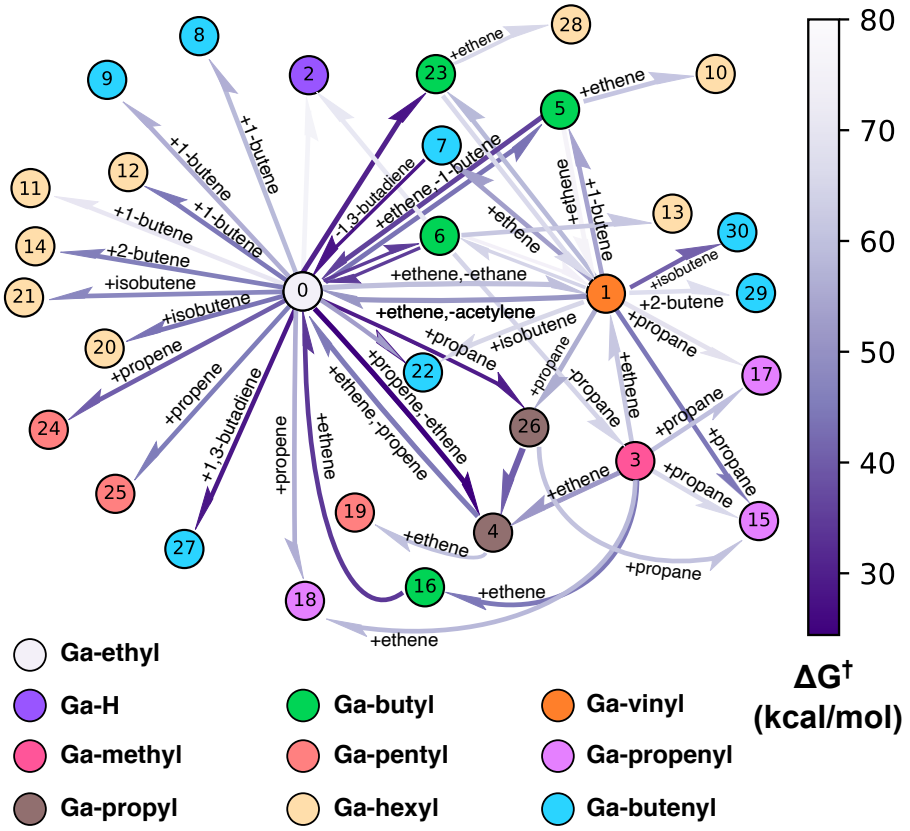

Figure 2: Gallium catalyzed olefin oligomerization reaction network obtained from YARP exploration. Free energies of activation are presented by the edge colors that represent kinetic accessibility. Intermediate types are classified based on the alkyl and alkenyl attached to Ga and are denoted by different node colors.

n-butenyl (from Ga-vinyl, $\Delta G^{\dagger} = 53.2$ kcal/mol), acetylene (formed with Ga-ethyl from Ga-vinyl, $\Delta G^{\dagger} = 51.4$ kcal/mol), Ga-hexyl (from Ga-butyl, $\Delta G^{\dagger} = 61.3$ kcal/mol), 1-butene (formed with Ga-ethyl from Ga-butyl, $\Delta G^{\dagger} = 36.0$ kcal/mol) and butane (formed with Ga-vinyl from Ga-butyl, $\Delta G^{\dagger} = 76.4$ kcal/mol) as intended products. Notably, the lowest barrier step yielding 1-butene constitutes a rediscovery by the algorithm of the classic Cossee-Arlman mechanism that has previously been studied as the likely pathway for major product formation in this system. Based on the activation energies of the reactions at this iteration, Ga-n-butenyl was included as a new active node for further exploration (node 7), Ga-n-hexyl was classified as a terminal node (node 13)

due to its size, and 1-butene was added to the free-olefin list as a candidate for further reactions with the active nodes, Ga-ethyl (node 0) and Ga-vinyl (node 1). YARP recursively explored the reaction space via the same approach that was employed in the first and second iteration until all reactions within the prescribed constraints had been explored. All reactions explored with $\Delta G^{\dagger} < 80$ kcal/mol are presented in Figure. 2 and detailed geometries of each node can be found in the SI.

3.2 Three key reaction types occurring on Ga^{3+}

Three distinct types of reactions were discovered during the network exploration that are distinguished by their reactions with the adsorbed carbon species. All instances of each class exhibit $\Delta G^{\dagger} < 70$ kcal/mol. The first types is responsible for lengthening (or breaking as the reverse reaction) the carbon backbone (Type I in Fig.3). The TS of the Type I reaction involves a "C=C" moiety bonding to the catalyst to form a four-coordinate Ga intermediate that precedes bond formation with an adsorbed alkyl species. The second type of reaction is β-hydride transfer that enables liberation of an oligomer and closes an oligomerization cycle (Type II in Fig.3). In the TS, the β-hydrogen of the adsorbed alkyl species transfers to an incoming olefin, which binds to the Ga center and becomes a new adsorbate. An oligomerization cycle can also be completed by a β-hydride elimination step to form Ga-hydride, but YARP predicts a much higher activation energy for this pathway. The third type of reaction produces an alkane, leaving a hydrogen-deficient adsorbed species, like Ga-vinyl (Type III in Fig.3). The TS of the Type III reaction resembles that of Type II, except that the hydrogen transfers to the α-carbon. Alkane formation has been reported in multiple olefin oligomerization experiments,[32,51–53] which may be explained by moderate-barrier Type III pathways. Further, we hypothesize that the products of type III reactions may undergo additional type I and type II steps. The combination of type I-III reactions may eventually liberate alkynes and aromatics that are commonly considered coke precursors.

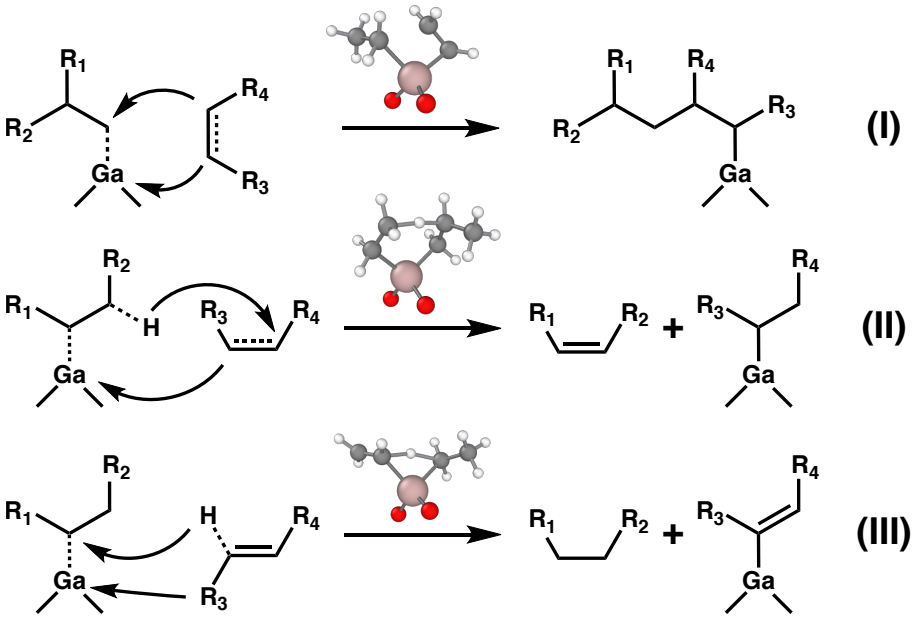

Figure 3: Three elementary reaction types identified during reaction network exploration. (I) olefin insertion; (II) β-hydride transfer; (III) α-hydride transfer.

In addition to recapitulating the expected Cossee-Arlman oligomerization cycle, these elementary reaction types can also participate in several catalytic cycles for olefin isomerization and chain cracking (Fig. 4). Interestingly, following the formation of 1-butene and the recovery of the Ga-ethyl intermediate (species (5) in Fig.4), where a Cossee-Arlman oligomerization cycle is about to finish, the 1-butene molecule can be re-adsorbed with a simple rotation and react with Ga-ethyl again through another β-hydride transfer step (type II), producing Ga-2-butyl (species (6)). This newly reported intermediate can undergo a facile type II reaction, forming cis- or trans-2-butene (only cis-2-butene formation is considered here, species (14)). Alternatively, Ga-2-butyl can undergo additional type I and II reactions to form Ga-methyl with physisorbed propylene (species (8)). Intriguingly, there can be another re-adsorption step of propylene on Ga-methyl, resulting in a Ga-isobutyl species (species (15)), which eventually leads to isobutene (species (17)). Throughout the isomerization and cracking pathways, the type III step can occur on each Ga-alkyl species.

For example, a plausible pathway involving the type III reaction is outlined in the green circle of Figure 4, where the resulting Ga-vinyl intermediate undergoes additional β-hydride transfer, leading to the formation of acetylene (a coke precursor).

3.3 Kinetic significance of types 1-3 transition states

Focusing on the proposed oligomerization, isomerization, cracking, and coking pathways, the reaction energies and reaction barriers predicted by the cluster model were compared with the results of periodic DFT and NEB-Lanczos TS characterizations using the slab model. The kinetic relevance of three reaction cycles, which determine the selectivity of producing various gaseous products and coke precursors, were compared using potential energy diagrams (Fig. 5):

(1) $2C_2H_4 \longrightarrow C_4H_8$

(2) $3C_2H_4 \longrightarrow 2C_3H_6$

(3) $\frac{n+2}{2}C_2H_4 \longrightarrow C_2H_2 + C_nH_{2n+2}$, where n = 1, 2, and 3.

Cycle (1) involves multiple ethylene dimerization products (Fig. 4a-c). One catalytic cycle simply closes through an ethylene insertion (denoted as type I) and a β-hydride transfer (denoted as type II). Following further type I-II steps occurring on Ga-ethyl with an adsorbed 1-butene (species (5)), cis-2-butene and isobutene can also form (Fig. 4b-c). To complete the catalytic cycle of cycle (2), one propylene molecule can be obtained through C-C bond breaking of a Ga-2-butyl species (reverse type I). The production of a second propylene molecule occurs via the same Cossee-Arlman oligomerization cycle initiated by the Ga-methyl intermediate (species (9), Fig. 4d). In cycle (3), the type III elementary step generates an alkane, which may occur for all Ga-alkyl intermediates, and an alkyne, like acetylene, is formed that balances the stoichiometry. A relatively facile acetylene formation pathway occurs through a type II step occurring on the Ga-vinyl species from the type III reaction (species (3_a), Fig. 4e). Many other relatively low barrier pathways ($\leq$70 kcal/mol) are discovered by YARP, including the formation of various

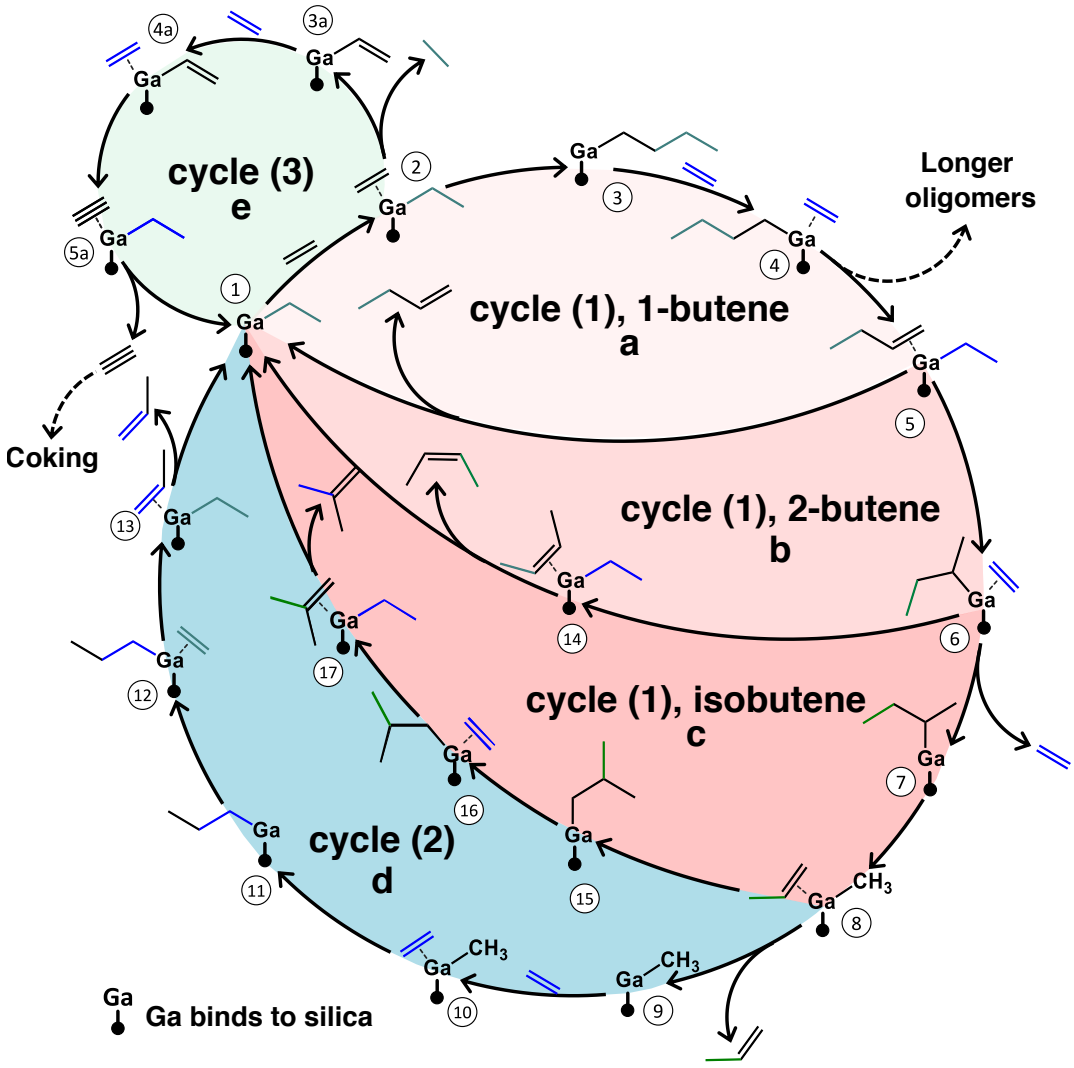

Figure 4: Proposed pathways of oligomerization, isomerization, cracking and coking reactions. (a) Oligomerization pathway of formation of 1-butene. (b-c. isomerization pathways of forming 2-butene (b) and isobutene (c). (d) Cracking pathway for propylene formation. (e) An example of a pathway for the formation of alkanes and hydrogen-deficient products starting from Ga-ethyl. Similar alkane formation cycles can also occur for species ②, ④, ⑥, and ⑩.

C_nH_{2n} species, and C_nH_{2n-2} isomers. The combination of these three elementary catalytic cycles thus results in a broad diversity of possible products.

To validate the accuracy of the cluster model results, they were benchmarked against conventional periodic DFT results on reaction cycles (1) and (2) (Fig. 5a). For this comparison, the cluster model results were further refined at the B3LYP-D3/6-311G(d,p) level of theory to minimize the DFT errors as a confounding factor when comparing the cluster and slab results (see the Supporting Information for additional details). Overall, the conventional periodic DFT and the cluster model results generate similar binding energies, reaction energies, and reaction barriers. However, consistently higher activation energies are observed for the cluster model. For example, for the ethylene insertion step (species (2) to species (3)), the cluster and periodic catalyst site models give activation energies of 1.8 and 1.5 eV, respectively. The difference may be attributed to long-range order in the silica support, which may cause a lower activation energy, but is absent in the cluster model. Moreover, the models are evaluated with distinct functionals due to their differing availability in the reference molecular and periodic quantum chemistry packages being used. Nevertheless, the two approaches predict similar relative barriers for all of the TS under each elementary step type. The mean difference between activation energies for type I versus type II reactions are 0.8 and 1.0 eV, calculated by the cluster model and the periodic model, respectively. Further, all type I transition states are nearly equally accessible in the energy landscape, and both models predict type II reactions to be consistently lower barrier. The overall excellent qualitative and quantitative agreement between the cluster and periodic models with respect to relative barrier heights validates the usefulness of the cluster models for performing reaction network exploration.

Figures 5b-c outline the energy landscape comparison between the overall reaction cycles (1)-(3) using the cluster results. In cycle (1), where the carbon chain length doubles and 1-butene is formed (species (1) - (5)), the ethylene insertion involves an activation energy (1.76 eV) higher than

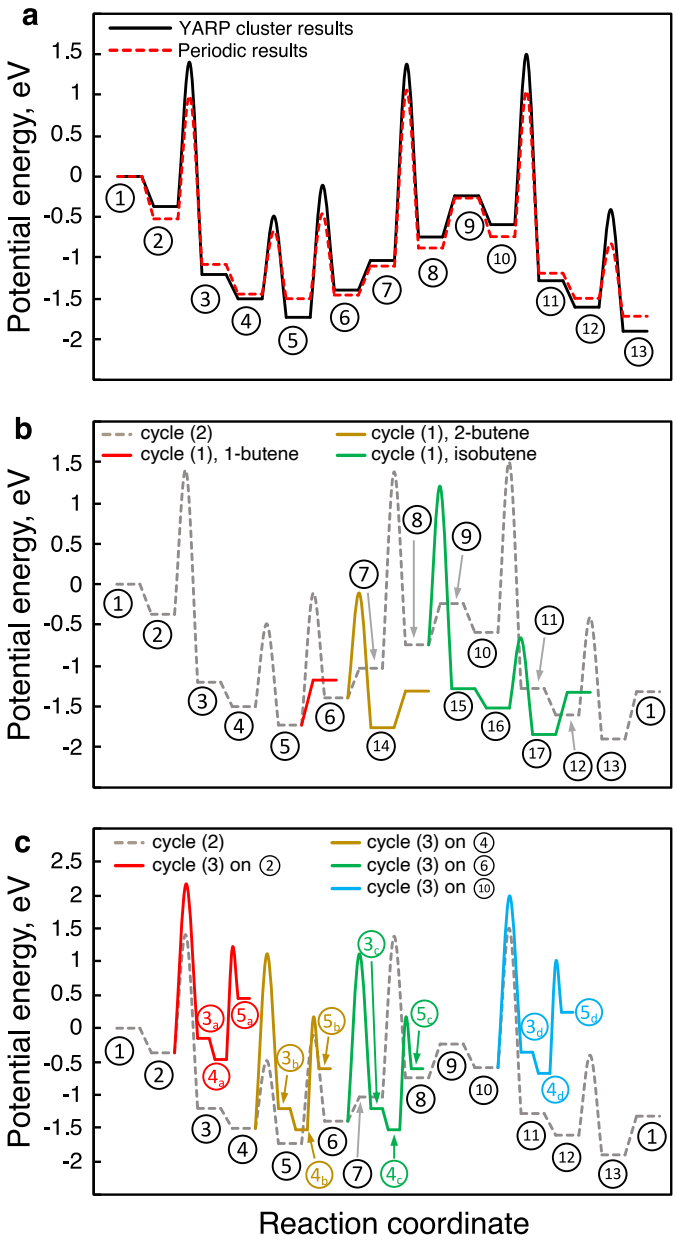

Figure 5: Energy diagrams of three kinetically relevant reaction cycles discovered within the reaction network. (a) Comparison of the energy landscape for cycle (2) using the cluster model and periodic slab. (b) Comparison of competing olefin formation pathways (colored) and cracking pathways (gray). (c) Comparison of competing acetylene formation pathways (colored) and cracking pathways (gray). The species are numbered based on the pathway diagram in Figure 4.

the olefin liberation step (1.02 eV). Indeed, previous studies have shown that ethylene insertion is rate-determining in this system,[32] and the cluster calculations predict that the energy of a type I TS is typically higher than a type II. Throughout the energy landscape of cycle (2), three type I elementary steps have distinctly high activation energies: the ethylene insertion shared by cycle (1), the cracking of Ga-2-butyl (species (7), 2.40 eV), and the step forming Ga-1-propyl (species (10)) from Ga-methyl and ethylene (2.08 eV). Particularly, the cracking of Ga-2-butyl, forming Ga-methyl and propylene, involves the highest activation energy since it is a reversed type I step. Both periodic and YARP-cluster results predict that type I reactions are exothermic. Therefore, cycle (2) may not dominate the reaction network. Indeed, previous experimental results of Ga single sites show a strong selectivity to olefin oligomerization at 250 °C and 1 atm, forming short linear oligomers.[32]) However, the cracking activation energy becomes lower as the temperature increases due to a favorable entropic effect (more degrees of freedom) on the reverse type I step versus the formation of longer Ga-alkyl carbon chain, thus narrowing the energy difference between Ga-2-butyl (species (7)) and the cracking TS. Entropy can also shift the equilibrium since cycle (2) produces a higher number of gas molecules than cycle (1), suggesting that cycle (2) becomes more favorable at higher temperatures. Finally, the high barrier of the reverse type I step provides a basis for the competition between type I and III reactions starting from the Ga-2-butyl species. In particular, the formation of 1-butane (species (6) - (3_c)) can be competitive with cracking reactions (species (7) - (8)). Subsequently, acetylene formation can occur via facile type II reactions (species (4_c) - (5_c),1.68 eV). Therefore, our pathway analysis suggests that type III reactions are kinetically less favorable, but nevertheless represent side-reaction channels that becomes accessible as they compete with the reverse of type I step. With the formation of alkynes, other side reactions, such as aromatization and coking, may occur as subsequent thermodynamic products.

4 Conclusion

The maturation of reaction exploration algorithms will provide access to chemical network kinetics during the chemical discovery and design phases, rather than retrospectively. This presents several exciting possibilities for catalyst design, including optimizing catalysts with respect to specific products rather than specific mechanisms, predicting off-target pathways, and generating hypotheses for novel catalytic cycles. Here we have demonstrated how automatic exploration can be applied to heterogeneous catalytic networks using ethylene oligomerization catalyzed by a silica-supported Ga single site as a benchmark. The method (re)discovered the classic Cossee-Arlman oligomerization cycle and several side-product pathways with minimal user intervention. Given the generic reaction rules and size constraints that were used to generate this network, there are relatively few obstacles to applying this approach to other heterogeneous systems. Among the salient details of the implementation to consider moving forward are the use of a cluster model as a surrogate for a periodic slab and the major speedup provided by semi-empirical quantum chemistry. Neither detail is intrinsic to applying YARP. The cluster assumption was validated here and adopted out of convenience, since the useful GFN2-XTB semi-empirical method is at present non-periodic. There are no obstacles to applying YARP using a periodic code, outside of cost. Notably, cluster models used to be much more prevalent before the adoption of periodic DFT, and as demonstrated here, the assumption is robust for establishing relative barriers in localized catalysts, like oxide surfaces, that can be refined later at the periodic level as required. The applicability of this approach to other heterogeneous surfaces is therefore anticipated and is currently under investigation.

5 Data Availability

The authors declare that the data supporting the findings of this study are available within the paper and its supplementary information files.

6 Code Availability

The version of YARP used in this study and a guide to reproducing the results is available through GitHub under the GNU GPL-3.0 License [repository doi will be inserted upon acceptance].

References

(1) Ertl, G. Elementary Steps in Heterogeneous Catalysis. *Angew. Chem., Int. Ed.* **1990**, *29*, 1219–1227.

(2) Chorkendorff, I.; Niemantsverdriet, J. W. *Concepts of modern catalysis and kinetics*; John Wiley & Sons, 2003.

(3) Medford, A. J.; Vojvodic, A.; Hummelshøj, J. S.; Voss, J.; Abild-Pedersen, F.; Studt, F.; Bligaard, T.; Nilsson, A.; Nørskov, J. K. From the Sabatier principle to a predictive theory of transition-metal heterogeneous catalysis. *J. Catal.* **2015**, *328*, 36–42.

(4) Mazeau, E. J.; Satpute, P.; Blöndal, K.; Goldsmith, C. F.; West, R. H. Automated Mechanism Generation Using Linear Scaling Relationships and Sensitivity Analyses Applied to Catalytic Partial Oxidation of Methane. *ACS Catal.* **2021**, *11*, 7114–7125.

(5) Xu, J.; Cao, X.-M.; Hu, P. Perspective on computational reaction prediction using machine

learning methods in heterogeneous catalysis. *Phys. Chem. Chem. Phys.* **2021**, *23*, 11155–11179.

(6) 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.

(7) Nørskov, J.; Bligaard, T.; Logadottir, A.; Bahn, S.; Hansen, L.; Bollinger, M.; Bengaard, H.; Hammer, B.; Sljivancanin, Z.; Mavrikakis, M.; Xu, Y.; Dahl, S.; Jacobsen, C. Universality in Heterogeneous Catalysis. *J. Catal.* **2002**, *209*, 275–278.

(8) Greeley, J. Theoretical Heterogeneous Catalysis: Scaling Relationships and Computational Catalyst Design. *Annu. Rev. Chem. Biomol. Eng.* **2016**, *7*, 605–635.

(9) Bligaard, T.; Nørskov, J.; Dahl, S.; Matthiesen, J.; Christensen, C.; Sehested, J. The Brønsted–Evans–Polanyi relation and the volcano curve in heterogeneous catalysis. *J. Catal.* **2004**, *224*, 206–217.

(10) Kropp, T.; Mavrikakis, M. Brønsted–Evans–Polanyi relation for CO oxidation on metal oxides following the Mars–van Krevelen mechanism. *J. Catal.* **2019**, *377*, 577–581.

(11) Vojvodic, A.; Calle-Vallejo, F.; Guo, W.; Wang, S.; Toftelund, A.; Studt, F.; Martínez, J. I.; Shen, J.; Man, I. C.; Rossmeisl, J.; Bligaard, T.; Nørskov, J. K.; Abild-Pedersen, F. On the behavior of Brønsted-Evans-Polanyi relations for transition metal oxides. *J. Chem. Phys.* **2011**, *134*, 244509.

(12) Loffreda, D.; Delbecq, F.; Vigné, F.; Sautet, P. Fast Prediction of Selectivity in Heterogeneous

Catalysis from Extended Brønsted-Evans-Polanyi Relations: A Theoretical Insight. *Angew. Chem.* **2009**, *121*, 9140–9142.

(13) Greeley, J.; Mavrikakis, M. Alloy catalysts designed from first principles. *Nat. Mater.* **2004**, *3*, 810–815.

(14) Greeley, J.; Jaramillo, T. F.; Bonde, J.; Chorkendorff, I.; Nørskov, J. K. Computational high-throughput screening of electrocatalytic materials for hydrogen evolution. *Nat. Mater.* **2006**, *5*, 909–913.

(15) Greeley, J.; Stephens, I. E. L.; Bondarenko, A. S.; Johansson, T. P.; Hansen, H. A.; Jaramillo, T. F.; Rossmeisl, J.; Chorkendorff, I.; Nørskov, J. K. Alloys of platinum and early transition metals as oxygen reduction electrocatalysts. *Nat. Chem* **2009**, *1*, 552–556.

(16) Studt, F.; Sharafutdinov, I.; Abild-Pedersen, F.; Elkjær, C. F.; Hummelshøj, J. S.; Dahl, S.; Chorkendorff, I.; Nørskov, J. K. Discovery of a Ni-Ga catalyst for carbon dioxide reduction to methanol. *Nat. Chem* **2014**, *6*, 320–324.

(17) Nilsson, A.; Pettersson, L. G. *Chemical Bonding at Surfaces and Interfaces*; Elsevier, 2008; pp 57–142.

(18) Li, J.; Zhou, S.; Zhang, J.; Schlangen, M.; Usharani, D.; Shaik, S.; Schwarz, H. Mechanistic Variants in Gas-Phase Metal-Oxide Mediated Activation of Methane at Ambient Conditions. *J. Am. Chem. Soc.* **2016**, *138*, 11368–11377.

(19) Mleczko, L.; Baerns, M. Catalytic oxidative coupling of methane—reaction engineering aspects and process schemes. *Fuel Process. Technol.* **1995**, *42*, 217–248.

(20) Vernuccio, S.; Bickel, E. E.; Gounder, R.; Broadbelt, L. J. Microkinetic Model of Propylene

Oligomerization on Brønsted Acidic Zeolites at Low Conversion. *ACS Catal.* **2019**, *9*, 8996–9008.

(21) Goldsmith, C. F.; West, R. H. Automatic generation of microkinetic mechanisms for heterogeneous catalysis. *J. Phys. Chem. C* **2017**, *121*, 9970–9981.

(22) Ismail, I.; Stuttaford-Fowler, H. B.; 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.

(23) Blondal, K.; Jelic, J.; Mazeau, E.; Studt, F.; West, R. H.; Goldsmith, C. F. Computer-Generated Kinetics for Coupled Heterogeneous/Homogeneous Systems: A Case Study in Catalytic Combustion of Methane on Platinum. *Ind. Eng. Chem. Res.* **2019**, *58*, 17682–17691.

(24) Broadbelt, L. J.; Stark, S. M.; Klein, M. T. Computer generated pyrolysis modeling: on-the-fly generation of species, reactions, and rates. *Ind. Eng. Chem. Res.* **1994**, *33*, 790–799.

(25) Gao, C. W.; Allen, J. W.; Green, W. H.; West, R. H. Reaction Mechanism Generator: Automatic construction of chemical kinetic mechanisms. *Comput. Phys. Commun.* **2016**, *203*, 212–225.

(26) Maeda, S.; Taketsugu, T.; Morokuma, K. Exploring transition state structures for intramolecular pathways by the artificial force induced reaction method. *J. Comput. Chem.* **2014**, *35*, 166–173.

(27) Shang, C.; Liu, Z. P. Stochastic surface walking method for structure prediction and pathway searching. *J. Chem. Theory Comput.* **2013**, *9*, 1838–1845.

(28) Zimmerman, P. M. Automated discovery of chemically reasonable elementary reaction steps. *J. Comput. Chem.* **2013**, *34*, 1385–1392.

(29) Zhao, Q.; Savoie, B. M. Simultaneously improving reaction coverage and computational cost in automated reaction prediction tasks. *Nat. Comput. Sci.* **2021**, *1*, 479–490.

(30) Zhang, X.-J.; Shang, C.; Liu, Z.-P. Stochastic surface walking reaction sampling for resolving heterogeneous catalytic reaction network: A revisit to the mechanism of water-gas shift reaction on Cu. *J. Chem. Phys.* **2017**, *147*, 152706.

(31) Iwasa, T.; Sato, T.; Takagi, M.; Gao, M.; Lyalin, A.; Kobayashi, M.; Shimizu, K.-i.; Maeda, S.; Taketsugu, T. Combined Automated Reaction Pathway Searches and Sparse Modeling Analysis for Catalytic Properties of Lowest Energy Twins of Cu13. *J. Phys. Chem. A* **2018**, *123*, 210–217.

(32) LiBretto, N. J.; Xu, Y.; Quigley, A.; Edwards, E.; Nargund, R.; Vega-Vila, J. C.; Caulkins, R.; Saxena, A.; Gounder, R.; Greeley, J.; Zhang, G.; Miller, J. T. Olefin oligomerization by main group Ga^{3+} and Zn^{2+} single site catalysts on SiO2. *Nat. Commun.* **2021**, *12*.

(33) Roggero, I.; Civalleri, B.; Ugliengo, P. Modeling physisorption with the ONIOM method: the case of NH3 at the isolated hydroxyl group of the silica surface. *Chem. Phys. Lett.* **2001**, *341*, 625–632.

(34) Tielens, F.; Gierada, M.; Handzlik, J.; Calatayud, M. Characterization of amorphous silica based catalysts using DFT computational methods. *Catal. Today* **2020**, *354*, 3–18.

(35) 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.

(36) Zimmerman, P. M. Growing string method with interpolation and optimization in internal coordinates: Method and examples. *J. Chem. Phys.* **2013**, *138*, 184102.

(37) Floryan, L.; Borosy, A. P.; Núñez-Zarur, F.; Comas-Vives, A.; Copéret, C. Strain effect and dual initiation pathway in CrIII/SiO2 polymerization catalysts from amorphous periodic models. *J. Catal.* **2017**, *346*, 50–56.

(38) Frisch, M. J. et al. Gaussian 16 Revision C.01. 2016; Gaussian Inc. Wallingford CT.

(39) Aldaz, C.; Kammeraad, J. A.; Zimmerman, P. M. Discovery of conical intersection mediated photochemistry with growing string methods. *Phys. Chem. Chem. Phys.* **2018**, *20*, 27394–27405.

(40) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **1996**, *54*, 11169–11186.

(41) Kresse, G.; Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **1996**, *6*, 15–50.

(42) Kresse, G.; Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **1993**, *47*, 558–561.

(43) Kresse, G.; Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Phys. Rev. B* **1994**, *49*, 14251–14269.

(44) Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **1999**, *59*, 1758–1775.

(45) Wellendorff, J.; Lundgaard, K. T.; Møgelhøj, A.; Petzold, V.; Landis, D. D.; Nørskov, J. K.; Bligaard, T.; Jacobsen, K. W. Density functionals for surface science: Exchange-correlation model development with Bayesian error estimation. *Phys. Rev. B* **2012**, *85*, 235149.

(46) Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **1994**, *50*, 17953–17979.

(47) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **2000**, *113*, 9901–9904.

(48) Henkelman, G.; Jónsson, H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.* **2000**, *113*, 9978–9985.

(49) Smidstrup, S.; Pedersen, A.; Stokbro, K.; Jónsson, H. Improved initial guess for minimum energy path calculations. *J. Chem. Phys.* **2014**, *140*, 214106.

(50) Olsen, R. A.; Kroes, G. J.; Henkelman, G.; Arnaldsson, A.; Jónsson, H. Comparison of methods for finding saddle points without knowledge of the final states. *J. Chem. Phys.* **2004**, *121*, 9776–9792.

(51) Sydora, O. L. Selective Ethylene Oligomerization. *ACS Catal.* **2019**, *38*, 997–1010.

(52) Moussa, S.; Concepción, P.; Arribas, M. A.; Martínez, A. Nature of Active Nickel Sites and Initiation Mechanism for Ethylene Oligomerization on Heterogeneous Ni-beta Catalysts. *ACS Catal.* **2018**, *8*, 3903–3912.

(53) Dagle, V. L.; Lopez, J. S.; Cooper, A.; Luecke, J.; Swita, M.; Dagle, R. A.; Gaspar, D. Production and fuel properties of iso-olefins with controlled molecular structure and obtained from butene oligomerization. *Fuel* **2020**, *277*, 118147.