Revisiting Catalytic Cycles: A Broader View Through the Energy Span Model

Diego Garay-Ruiz*,† and Carles Bo*,†,‡

†Institute of Chemical Research of Catalonia (ICIQ), The Barcelona Institute of Science and Technology (BIST), Av. Països Catalans, 16, 43007 Tarragona, Spain

‡Departament de Química Física i Inorgànica, Universitat Rovira i Virgili, Marcel·lí Domingo s/n, 43007 Tarragona, Spain

E-mail: cbo@iciq.cat

Abstract

The computational study of catalytic processes allows discovering really intricate and detailed reaction mechanisms that involve many species and transformations. This increasing level of detail can even result detrimental when drawing conclusions from the computed mechanism, as many co-existing reaction pathways can be in close competence. Here we present a reaction network-based implementation of the energy span model in the form of a computational code, gTOFfee, capable of dealing with any user-specified reaction network. This approach, compared to microkinetic simulations, enables a much easier and straightforward analysis of the performance of any catalytic reaction network. In this communication, we will go through the foundations and applicability of the underlying model, and will tackle the application to two relevant catalytic systems: homogeneous Co-mediated propene hydroformylation and heterogeneous CO_2 hydrogenation over Cu(111).

Introduction

Chemical reaction networks (CRNs)[1–3] are a quintessential concept in chemistry and especially in catalysis. A reaction network encodes all the chemical entities participating in a chemical transformation process and the relationships between them. From there, additional information, such as relative stability of the different species and the kinetic rate constants may also be included in the network to complete the description with *quantitative* information.

Furthermore, CRNs can also be immediately assimilated to *graphs*.[4,5] From a chemical perspective, it is natural to identify reaction intermediates as the *nodes* of the graph and the transition states that allow for the interconversion of these intermediates as the *edges*. This framework is a way to formalize what a reaction network is, and opens the door to a plethora of analysis and manipulation techniques that are available for this kind of mathematical objects. This formalism is nowadays being used to automatize mechanistic searches[6–10], simplify the setup of microkinetic simulations[11–14] for heterogeneous, homogeneous and biochemical catalysts[15,16], and it has been even implemented in chemical computing.[17]

When representing reaction networks as graphs, it is common practice to consider *directed* edges for the forward and reverse directions of every elemental step of the network. This directed nature supposes a clear and compact framework for the kinetics of the system: we would just need to map each forward and reverse edge to the corresponding reaction rate constant. Through this approach, the setup of microkinetic simulations can be integrated in more complex workflows, and calculation automatization and result analysis are improved. Microkinetic simulations require to solve systems of differential equations numerically, which scale rapidly with the size of the network. Applying numerical methods to solve the complex systems of equations that arise from complex networks can be slow and cumbersome, and likely to encounter problems related to numerical instabilities or equation stiffness.

In this way, we propose a distinct and more simple approach for the general analysis of complex catalytic cycles, a novel way of reading reaction networks following the energy

span model (ESM) introduced by Kozuch and Shaik.[18,19] The original formulation of the ESM considers linear free energy profiles, but not reaction networks. The requirement of linearity is a major drawback on the applicability of the model to real catalytic systems, which are often much more complex and can only be described as entangled networks. The latest developments on the model[20,21] address this critical issue and propose a fully general reaction network-based formulation. However, no further applications of this variant of the ESM have yet explored its full potential to deal with catalytic cycles as graphs. In contrast to microkinetic simulations, the ESM only requires to solve simple algebraic expressions that are, not only much faster to compute, but also less likely to encounter mathematical issues.

Obviously, our approach does not intend to be a *substitute* for microkinetic simulations, it just provides a framework to compute the *turnover frequency* (henceforth TOF) of complex catalytic cycles. Having a way to compute the TOF from computational data supplies an inexpensive additional descriptor to compare experimental data and theoretical predictions, or even for the inter-comparison of different catalytic profiles. The model also accounts for the possibility of including the effect that the concentrations of reactants and products in the reaction medium have on catalytic performance, which improves the description of the system. This feature, together with the simplicity of the overall model, makes the ESM a very interesting routine tool for a more in-depth analysis of catalytic networks: a niche where microkinetics might be too tricky and time-consuming to set up.

Therefore, what we propose in this communication is the first implementation of the energy span model based on the most general network-based expression[20,21] to compute the TOF for an arbitrarily complex catalytic cycle, avoiding any kind of *ad-hoc* simplification of the network so as to maintain all the chemical information about the system. To achieve this, we developed a Python code, gTOFfee[22], capable of handling reaction networks in the form of graphs to compute the corresponding turnover frequency applying the ESM.

The manuscript is organized as follows: first, we revisit Kozuch's model and discuss the basics of our implementation. A more detailed description can be found in the Supporting

Information as well as a simple catalytic cycle as example. Then, we discuss how we apply the new algorithms to treat two well-known catalytic transformations: the homogeneous cobalt-catalyzed propene hydroformylation, and the heterogeneous CO_2 hydrogenation over Cu(111).

The model

One key aspect of the network-based TOF formulation proposed by Kozuch[20] is the formal definition of what a catalytic reaction mechanism is. In this sense, a mechanism is defined as a *subgraph* of the original network, including all the nodes (chemically, all stable intermediates, or minima of the potential energy surface), but in which only a single cycle leading to any of the possible reaction products is closed. Although many cycles may be intertwined in the original network, a valid mechanism would only include one closed cycle. However, the other possible channels are not disregarded, as every possible intermediate is still taken into account, in the form of an off-cycle branch that will have an effect into the overall reaction. Given a determined reaction network, then, we could define a set of N mechanisms following this definition, considering all different possible cycles and branching patterns.

The main equation for the calculation of the turnover frequency from an arbitrary reaction network is:[20]

$$TOF = \frac{k_B T}{h} \sum_n \frac{\mu_n (1 - e^{\Delta G_r / RT})}{(\sum_k \tau_k) \sum_j (e^{(-I_j + \delta G_{ij})/RT})} \tag{1}$$

In this expression, the TOF for the whole reaction network is expressed as a sum including a term for every of the **n** possible mechanisms that can be defined from the network. The overall TOF for the complete system would then be the summation of all the individual TOF_n terms. At this point, it makes sense to go back to the much simpler TOF expression for linear profiles[19,23], to see how it compares with eq. 1.

4

$$TOF = \frac{k_B T}{h} \frac{(1 - e^{\Delta G_r / RT})}{\sum_{ij} e^{(T_i - I_j + \delta_{ij})/RT}} \qquad (2)$$

The numerator in equation 2 corresponds to the thermodynamic driving force of the reaction, which must be *exergonic* for this term to be positive. The ESM will not work for endergonic, non-spontaneous processes, neither in the linear nor in the network-based implementation. The denominator of the equation accounts for the kinetic resistance including every possible intermediate/TS combination along the cycle. In this term, T_i corresponds to the energy of the i-th TS, I_j to the energy of the j-th intermediate, while δ_{ij} considers the relative position of every intermediate/TS pair. This δ_{ij} takes the value 0 when i < j (I_j is *before* T_i) and ΔG_{reac} when i > j: I_j is *after* T_i).

If we compare eq. 1 with eq. 2, we can observe that eq. 1 also includes the driving force in the numerator and has a similar underlying structure, with the main difference of a couple of graph-related terms: μ_n and $\sum_k \tau_k$. μ_n corresponds to the n-th mechanism derived from the graph, while the latter ($\sum_k \tau_k$) is a summation extended to all the *spanning trees* derived from the main reaction network. These spanning trees are *acyclic* subgraphs which still must connect every single node in the network. Therefore, two kinds of subgraphs can be derived from the original graph to apply equation 1: **n** single-cycle mechanisms and **k** acyclic trees.

Mathematically, both μ_n and τ_i terms are simple exponentials of the form $e^{(\sum -T_i)/RT]}$, where the sum is extended to all the edge energies that appear in the corresponding subgraph (mechanism or spanning tree). Therefore, once all necessary subgraphs are defined, the application of equation 1 only requires a trivial calculation of a series of simple exponential terms. It can also be demonstrated[20] that equation 1 reduces to eq. 2 in the case of simple cycles, which confirms the consistency of the network-based expression.

One important question that should be raised at this point is how to define the δ_{ij} term in equation 1. In the linear variant, there was a clear sequential relationship between T_i and I_j terms to determine whether a TS was located before or after an intermediate, but that is

missing in this new framework.

Fortunately a sense of sequentiality can still be derived for a given spanning tree, forging a simple rule for defining δ_{ij}. Therefore, we should seek whether a **node** in a tree τ_k lies before or after the first **edge** that would close that tree to a valid mechanism. In other words, we need to find an edge that will form a single cycle (valid mechanism) from the purely acyclic tree structure. In this sense, once this edge is identified, δ will be defined as:

$$\delta_{ij} = \begin{cases} \Delta G_r & \text{if int. after selected edge} \\ 0 & \text{if int. before selected edge} \end{cases} \tag{3}$$

As the mechanism-closing edge that we use as reference will be different for every tree, the $-I_j + \delta_{ij}$ terms will also vary across trees, and shall be computed accordingly.

Defining the "energy span" δE is also important. In the linear version, this magnitude is defined as the energy difference between the Turnover Determining Intermediate (TDI) and the Turnover Determining Transition State (TDTS), corresponding to the two states that control the overall reactivity of the process. To define these states properly, the model proposes another measurement, the degree of TOF control (x_{TOF}) of every state (going from 0, no control, to 1, total control, and computed separately for intermediates and transition states). The intermediate and the TS with the largest x_{TOF} would be the TDI and the TDTS, respectively. If $x_{TOF}(TDI) \approx 1$ and $x_{TOF}(TDTS) \approx 1$, the TOF calculation can be hugely simplified, depending only on the energy span through an Eyring-like expression.

$$TOF \approx \frac{k_B T}{h} e^{-\delta E/RT} \tag{4}$$

The utility of eq. 4 should not be overestimated: as a proper determination of δE requires knowing the degree of TOF control for the whole profile, it does not actually provide a way to estimate the TOF *a priori*. Mostly, the energy span is useful as a descriptor in energy units that can be directly related to the TOF, but fails when the profiles get more complicated

and cannot be summarized in a single TDI/TDTS pair.

Extending the definition to more complex reaction networks is not straightforward: on the one hand, the expressions for the degree of TOF control are not easy to reformulate for arbitrarily complex graphs. On the other, if equation 4 is already limited for simple profiles, it will be even worse for complex networks. Nevertheless, we can somehow overcome this limitation by defining what we have named the *effective energy span* , δE_{eff}, considering equation 4 but with the exact TOF value obtained with equation 1.

$$\delta E_{eff} = -RT \log \left[\frac{h}{k_B T} \cdot TOF \right] \tag{5}$$

This effective energy span still allows us to gauge the feasibility of a catalytic system (either for the whole network or an individual mechanism) in energy units, without losing accuracy throughout the process.

To finish, we should mention that the ESM can account for the effect of the concentrations of the entering reactants and the formed products, as one of its most interesting features for practical work. This extension is pretty straightforward over either the linear or the network-based versions of the model, as it only requires replacing the free energies of nodes and edges by the corresponding semi-standard Gibbs free energies[21]. This "semi-standard" correction still considers all catalyst-bearing species in a standard state (for solutions, 1.0 M concentration), while only correcting reactants and products to their actual concentrations. This might seem unphysical, as the catalyst-containing species will be in very low concentrations and therefore farther from the assumed standard state than reactants or products. However, as the usual procedure is to assume *all* species to be in a standard state, even this simple approximation supposes an improvement in the description of the system to better match experimental conditions. These concentration-dependent energies would be directly introduced as the T_i and I_j terms in equation 1, without any further overhead compared to using non-concentration-corrected values.

Practical Examples

Homogeneous Catalysis: Hydroformylation

The hydroformylation of alkenes with syngas ($CO + H_2$) catalyzed by organometallic complexes bearing either Co or Rh is one of the most prevalent homogeneous catalytic reactions in industry.[24–26] Because of this, plenty of experimental and computational studies have targeted this reaction along the years, in order to understand its mechanistic intricacies and tune the catalyst. This abundant knowledge makes hydroformylation a great benchmark for novel methods that address issues related to catalytic performance, like the one that we present in this communication.

Our starting point is the study from Rush, Pringle and Harvey[27], based on the phosphine-free **HCo(CO)$_4$** carbonyl complex, where the authors report CCSD(T) Gibbs free energies for both the *hydroformylation* and the competitive *hydrogenation* pathways (with propene as substrate). Those results provided very good agreement with experiment.
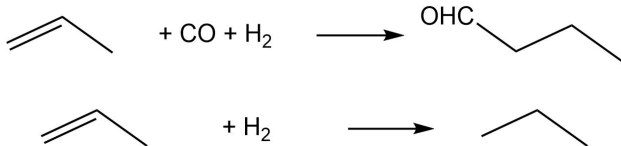

Figure 1: Hydroformylation and hydrogenation reactions.

The input Gibbs free energies must be referred to a 1.0 M reference state (standard in solution) to use the semi-standard approach for concentration effects. As the values from Rush et al.[27] consider a gas phase 1 atm reference state, we shall apply this state correction beforehand. Fortunately, the correction does only require a straightforward shifting of all Gibbs free energies as long as the temperature is constant between the two states.

$$G = G^0 + RT \log \frac{P_2}{P_1} = G^0 + RT \log \frac{cRT}{1 \text{ atm}} \tag{6}$$

As c = 1.0 M, the correction factor will only depend on temperature: 1.89 kcal·mol^{-1} at 298.15 K, and 2.98 kcal·mol^{-1} at 423 K, the value considered for Co-based hydroformylation.

Moreover, this standard state correction does also account for some degree of entropic correction. As it must be applied to every single species participating in the cycle, the relative energies of the system will be more or less shifted depending on the number of molecular entities that are involved in a given stage of the cycle. Therefore, situations with a high degree of association (with few entities) will be less destabilized than very disassociated (separated reactants) states. This compensates the overestimation of entropy loss upon association made by usual relative Gibbs free energies, in analogy with other corrections tailored for this specific issue (Martin[28], Wertz-Ziegler[29–31].). Previous studies[32] already highlight the importance of using entropic corrections to improve the match between experimental and computed TOFs, and even more so when several mechanisms with different degrees of association have to be compared. For the specific case of hydroformylation, indeed, the formation of butyraldehyde from propene would be *endergonic* with the 1 atm-based free energies, but *exergonic* in the 1M-state, while the competitive hydrogenation to propane (less associative) is exergonic in both descriptions. Thus, standard state correction is an essential asset for a proper comparison of both routes. Values extracted from literature and those used in this work can be found in the Supporting Information.

Figure 2 shows the translation from a catalytic cycle (adapted from the proposal of Rush et al.[27]) to a graph that can be fed to gTOFfee to proceed with the TOF calculation. It contains **13** nodes and **14** edges, which define two cycles, entering H2, CO and alkene as reactants, and exiting aldehyde and alkane as products. From there, **12** unique mechanisms can be generated, from which other **41** unique trees are derived. We must be careful with this mechanism assignation: depending on which cycle is closed (through **5X** to **2** or through **9X** to **2**) either alkane or aldehyde will be formed, involving different reaction energies. Even another complication arises: interestingly, some of the generated mechanisms correspond to aldehyde decarbonylation, forming CO and alkane. The reaction energy is then G(decarbonylation) = G(hydrogenation) - G (hydroformylation). gTOFfee is already prepared to handle this type of cases, however, it is always advisable to check the proposed
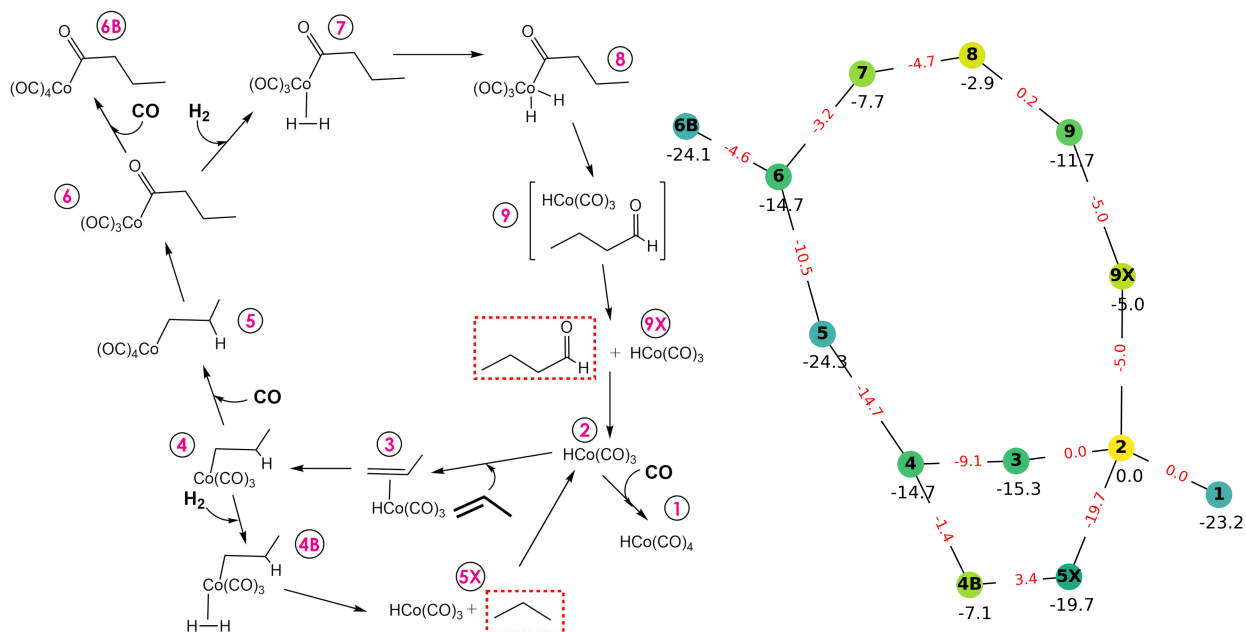
9

Figure 2: Reaction scheme for Co-catalyzed hydroformylation catalytic cycle (adapted from Rush[27]) (left) and full reaction network mapping chemical structures to nodes and edges (right).

mechanisms before using them for further calculations. A mechanism example for each of the three reaction routes (propene hydrogenation (M3), aldehyde decarbonylation (M4) and propene hydroformylation (M7), respectively) is provided in Figure 3.

Regarding the selectivity of the process to yield either aldehyde (via hydroformylation) or alkane (via hydrogenation or via decarbonylation), we should group the mechanisms by the formed product and add the corresponding turnover frequencies accordingly. As the alkane is thermodynamically favored, but the aldehyde is the major product in the experiments, the selectivity insights from turnover frequencies are undoubtedly relevant to improve the analysis of the computational data.

To begin with, we will compute TOFs for three different situations: i) semi-standard state, ii) high reactant concentrations C_{high}: [CO] = 1.5 M, [H_2] = 1.0 M, [Alkene] = 2.0 M, [Aldehyde] = [Alkane] = 0.01 M and iii) low reactant concentrations C_{low}: [CO] = 0.2 M, [H_2] = 0.1 M, [Alkene] = 0.5 M, [Aldehyde] = [Alkane] = 0.01 M.

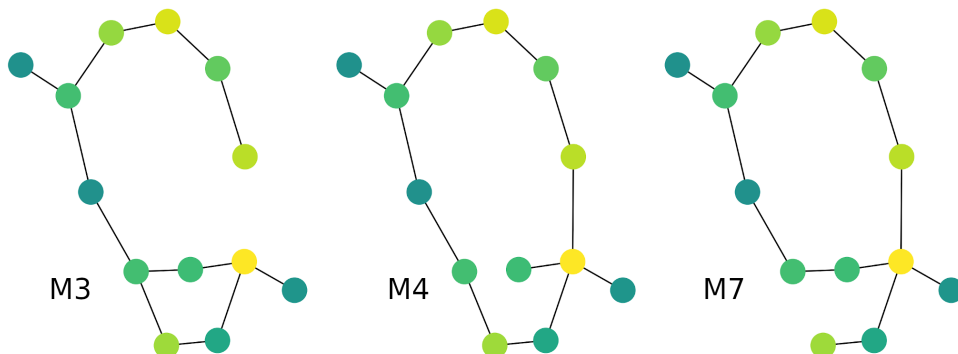For the most simple situation (semi-standard state) hydroformylation is shown to be

Figure 3: Mechanism examples for Co-catalyzed hydroformylation. From left to right: propene hydrogenation (M3), aldehyde decarbonylation (M4) and propene hydroformylation (M7).

Table 1: Derived valid mechanisms for the Co-catalyzed alkene hydroformylation. Mechanism index, edge removed from the main graph, G_{reac} and δE_{eff} values, in kcal mol^{-1}, for the three cases: semi-standard state (c^0), at high (c_{high}) and at low (c_{low}) concentrations, respectively. Below, conjoined δE_{eff} for aldehyde and alkane-producing routes.

Mech.	Edge	G_r	δE_{eff}^0	δE_{eff}^{high}	δE_{eff}^{low}	Product
1	2-3	-14.7	28.6	29.5	23.4	Decarbonylation
2	2-5X	-5.0	48.3	24.9	39.4	Hydroformylation
3	2-9X	-19.7	33.6	29.6	44.0	Hydrogenation
4	3-4	-14.7	37.7	38.2	35.7	Decarbonylation
5	4B-5X	-5.0	25.2	28.9	43.4	Hydroformylation
6	9X-9	-19.7	33.6	37.3	51.8	Hydrogenation
7	4-4B	-5.0	30.0	29.8	48.2	Hydroformylation
8	4-5	-19.7	43.3	42.8	55.1	Hydrogenation
9	5-6	-19.7	39.1	39.3	48.2	Hydrogenation
10	6-7	-19.7	31.8	31.6	53.8	Hydrogenation
11	7-8	-19.7	33.3	33.1	47.6	Hydrogenation
12	8-9	-19.7	28.4	28.2	42.7	Hydrogenation
Aldehyde			25.2	24.9	39.4	
Alkane			27.9	27.9	23.4	

kinetically preferred to hydrogenation. A more thorough analysis shows that only mechanism M5 contributes to aldehyde formation, while for alkane production both M1 and M12 have similar effective TOFs and both influence the global reactivity. If we analyze the preferred mechanism, M5, we observe that it lacks the **4B-5X** edge, which blocks the interconversion between these two nodes and therefore hinders the formation of the alkane.

A completely different picture emerges when we bring concentrations into play. At low reactant quantities (c_{low} set) the hydroformylation is completely shut down, and only hydrogenation is feasible: aldehyde will not be formed and the catalyst will only produce propane. However, when we consider high concentrations we recover the hydroformylation-favoring trend, with values (25.0 vs 25.2) that are very close to the semi-standard state concentration results. All of this is in line with known experimental behavior and with computational kinetic studies[27] that point at the need for using large H_2 and CO pressures to drive the reaction to the aldehyde.

These trends can be explored in much more depth through our new methodology straightforwardly. Instead of testing individual concentration value sets, we can automate changes in the concentration of some of the species and see how the overall TOF for each process (hydrogenation or hydroformylation) changes, and then build maps of reactivity depending on initial concentrations. As a more compact measure of selectivity, we can consider the quotient $TOF_{aldehyde}/TOF_{alkane}$[33] as the descriptor for our mapping. As CO and H_2 concentrations are known to be crucial to this selectivity, they were chosen as the mapping variables, while setting [Alkene]=3.0 M and [Aldehyde] = [Alkane] = 0.25 M (assuming some production of both species).

Figure 4 shows how quite large concentrations of both CO and hydrogen are required to drive the reaction to butyraldehyde formation: the selectivity for this product in the upper left corner ([CO]↓,[H_2]↓) is really poor. This matches very well with the known experimental trends, where large pressures of both reagents have to be used.

Once we enter the hydroformylation-enhancing region (in dark red in Figure 4), the ef-
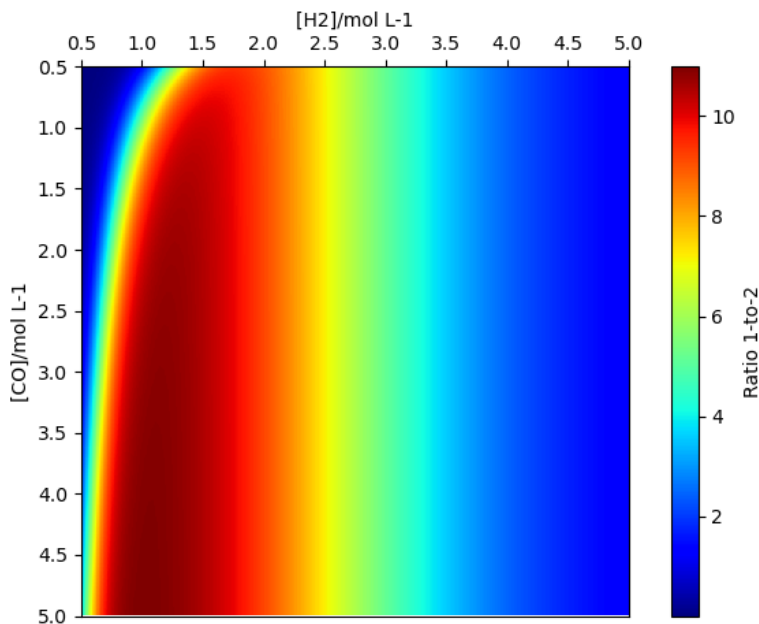
Figure 4: Selectivity of the reaction regarding [CO] and [H_2]. Concentrations in mol·L^{-1}

fect of the excess of one or the other reactant is drastically different. If the concentration of carbon monoxide rises (going down in the y-axis), the selectivity for butyraldehyde remains high. However, an excess of H_2 has the opposite effect: selectivity is progressively lost and at the right side of the map the hydrogenation route will be preferred. These two behaviors align well with the trends that we may expect through chemical intuition. If we want to insert CO into the propene scaffold, it is coherent that larger quantities of this species keep the selectivity high. In contrast, when more H_2 is added to the medium, the favored process is the insertion of hydrogen, which results on shutting down hydroformylation to form more propane. If we look at experimental data, the agreement of our approach remains excellent. The reaction is carried out with excess of carbon monoxide against hydrogen, so as to maximize the aldehyde selectivity. The experimental and microkinetic-based values for this selectivity[27] are around a 90-92% percent, in perfect agreement with the 10:1 ratio obtained with our model. Similar maps can be obtained for other variables (see Supporting Information), demonstrating that reaction selectivity is almost insensitive to the concentrations of propene and butyraldehyde, but very dependent on the quantity of the propane byproduct:

the presence of alkane blocks its own production route (hydrogenation). The conjunction of all of these factors explains the robustness of the catalytic system once the appropriate quantities of CO and H_2 have been selected. It is mostly because of this robustness that this homogeneous reaction is employed effectively in the chemical industry. The current application of the ESM, despite its relative simplicity, is able to extract and highlight this crucial feature from a standard homogeneous catalysis reaction network.

Heterogeneous Catalysis: CO_2 Hydrogenation

Although the ESM was developed with *homogeneous* catalysis in mind, once the step to consider any reaction network under this framework has been taken, the framework could also be applied to heterogeneous catalytic systems in which reaction networks are actually more commonly used.

Nevertheless, some practical considerations limit the applicability of the energy span model to this field of catalysis. First of all, the semi-standard approximation is designed to include *concentrations* in solution, and thus to deal only with homogeneous systems. Therefore, our test application to heterogeneous systems will only consider the 'bare' TOF.

Second, we should acknowledge the differences in the way that reaction networks are reported in homogeneous and heterogeneous catalysis. While in solution the focus is on the different forms adopted by the catalyst, and how it transforms along the process, in the solid phase this focus is on the individual elemental reactions undergone by reactants and products. In this sense, an heterogeneous network must be "translated" to a more homogeneous-like form in which all energies refer to the current state of the metal catalyst and how it should return to its original, all-free-site form.

Finally, heterogeneous catalysis usually describes relatively simple reactions in a very detailed manner, while homogeneous studies aim for more complicated transformations in a coarser way. Therefore, heterogeneous reaction networks might probably include more intermediates, edges and intertwined cycles, thus producing a much larger number of mech-

anisms and trees. Because of all these issues, we are only aiming for the basic possibilities that gTOFfee may offer for heterogeneous catalysis, but reckoning the limitations of the approach: a full development of a ESM/gTOFfee variant capable of treating large heterogeneous networks thoroughly will be a totally new branch to look at in the future.

For this qualitative depiction, we have chosen the CO_2 hydrogenation over Cu(111) on the basis of the reaction network proposed by Zhao et al.[34], which is relatively simple but still remains a clearly different challenge than the homogeneous hydroformylation reaction. The corresponding simplified network is shown in Figure 5. It contains **28** nodes and **31** edges, which define three interconnected cycles that produce methanol as main product, and formaldehyde as intermediate byproduct.
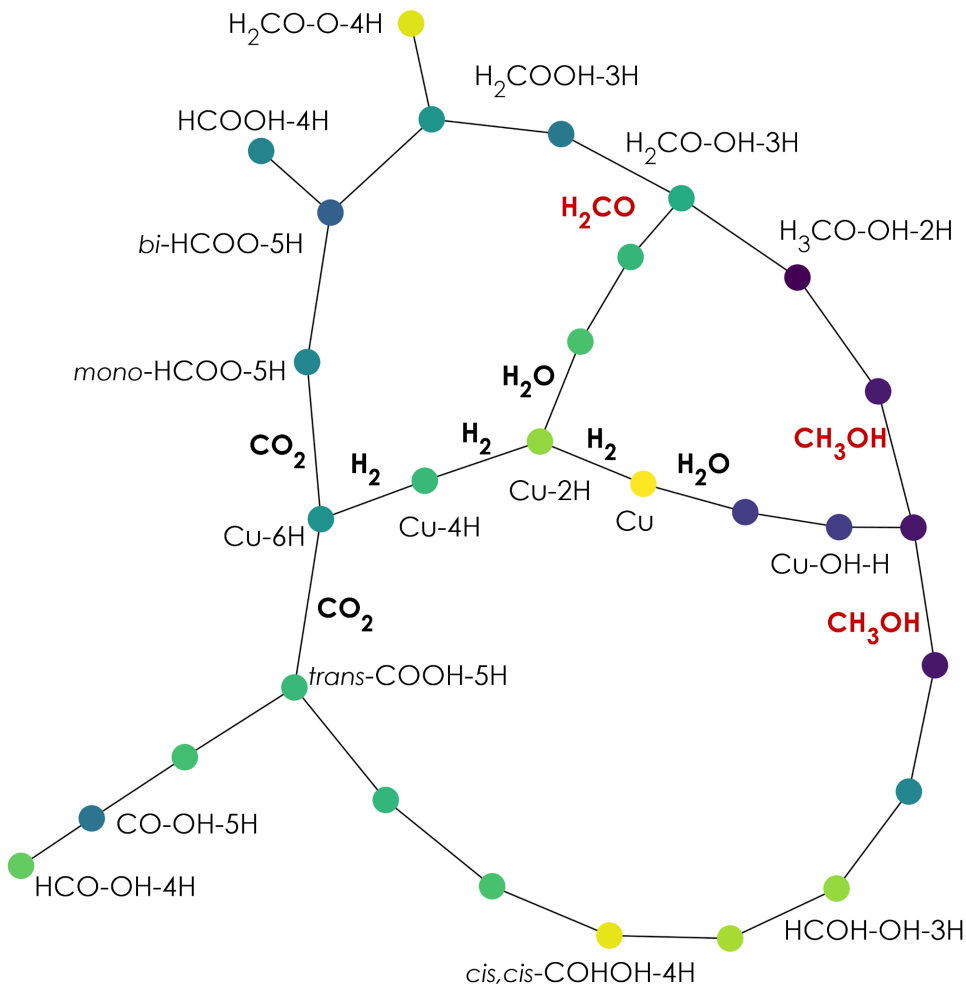

Figure 5: Reaction network for the Cu(111) catalyzed reaction between CO_2 and H_2

Despite the apparent simplicity of the network, up to 240 mechanisms and 1117 spanning trees are generated from it. This large number of subgraphs is due to combinatorics: when multiple individual cycles are present on the network, the number of possible subgraphs with different branching patterns grows quickly. Although the exact number of valid mechanisms or trees is not predictable, the number of tested subgraphs depends on the combinatorial number $\binom{N_{edges}}{N_{removal}}$. Here, N_{edges} is the total number of edges in the network, while $N_{removal}$ refers to the number of edges that need to be removed to form the required subgraph (mechanism or edge). This $N_{removal}$ value is directly related to the minimal number of cycles required to form the network: the more cycles we have, the more edges we need to remove to obtain one-cycle or acyclic subgraphs. In this specific case, we have **31** edges and the minimal cycle basis comprises **3** cycles. Therefore, there will be $\binom{31}{3-1} = 465$ mechanism proposals (from which 240 are valid mechanisms) and $\binom{31}{3} = 4495$ tree candidates (accepting 1117 of them). However, as we actually generate trees from mechanisms and not from the main network (see Supporting Information for details), the number of acyclic subgraphs to be tested will be much smaller.

The volume of feasible mechanisms prevents us from carrying out an individual analysis of each one as we did for hydroformylation (where only 12 possibilities had to be treated): we should instead try to implement a way to let the code summarize the most important information about the system.

First of all, we should categorize these 240 mechanisms. The reaction network by Zhao et al. considers the formation reactions of methanol or formaldehyde from the hydrogenation of carbon dioxide, which are exergonic processes. However, referring to our previous analysis above, mixed routes can appear throughout the network: e.g. pre-formed formaldehyde could be hydrogenated to form methanol. To identify all possibilities, we can focus on the *closed cycle* step of each mechanism, ignoring the branch patterns, and obtain all unique typologies.

Six mechanism types, as shown in Figure 6, are much more manageable than the 240
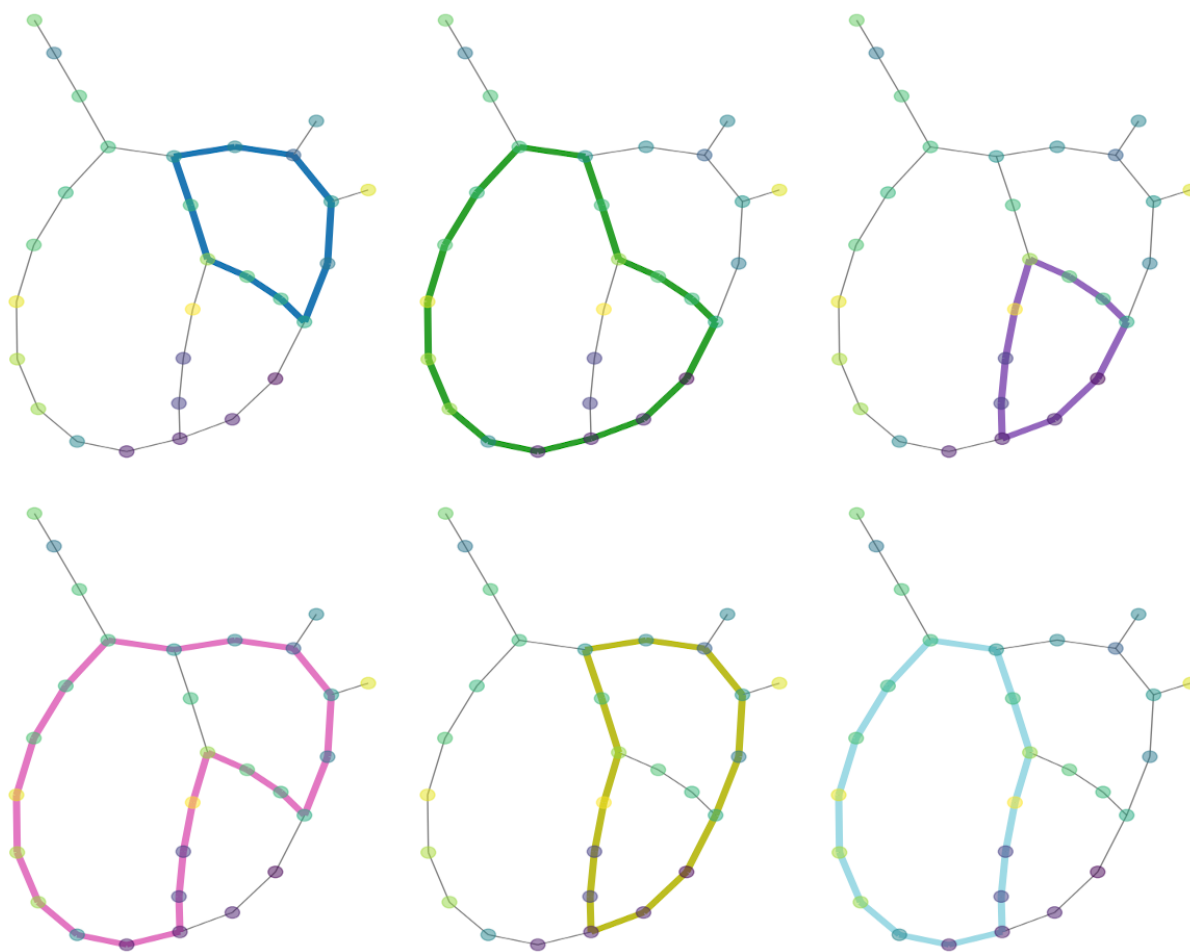
Figure 6: Graph depiction of the six unique closed-cycle typologies arising from Cu(111)-catalyzed CO_2 hydrogenation network

.

individual mechanisms. We can then proceed to inspect the details of these six typologies and get the corresponding effective TOF values for each one.

Table 2: Summary of mechanism typologies for Cu(111)-catalyzed CO_2 hydrogenation network. All energies in eV. From left to right, typology tag, formed product, number of associated mechanisms, reaction energy, effective energy span of the best mechanism, and effective energy span for the sum of all TOFs.

Type	Prod.	N	G_r	δE_{eff}^*	δE_{eff}
R1	H_2CO	75	-0.57	1.51	1.51
R2	H_2CO	20	-0.57	1.99	1.99
R3	CH_3OH	73	-1.04	1.51	1.48
R4	CH_3OH	6	-1.04	2.61	2.58
R5	CH_3OH	27	-1.61	1.73	1.73
R6	CH_3OH	39	-1.61	2.21	2.21

Table 2 shows that from the six possible catalytic cycle types, two routes will be clearly favored: **R1**, forming formaldehyde from CO_2, and **R3**, forming methanol from pre-existing H_2CO. The best δE^* values are 1.51 eV for both cases, but R3 becomes slightly favored when all of its mechanisms are added up, which lowers the effective energy span $\delta E_{eff}(R3)$ to 1.48 eV. Therefore, the catalytic system would be adequate for methanol production: although formaldehyde generation is competitive, the resulting H_2CO will be transformed onto methanol. However, the barriers are high, and require large temperatures (around 500 K).

To get a better grasp into the contribution of the different mechanisms assigned to each typology, we also plotted the corresponding histograms for the effective energy span values per mechanism group (Figure 7), which reveal that R3-type mechanisms show indeed lower effective energy spans than those of R1-type.

While every mechanism type in table 2 is mostly reduced to the single, best-performing mechanism, the histogram view gives an idea on how many accessible mechanisms do actually exist for each typology. For example, when comparing R1 and R3, the curve for R3 is centered at lower effective energy span values, which means that the average R3 mechanism is more accessible than the average R1 mechanism. In this sense, even if some individual mechanisms
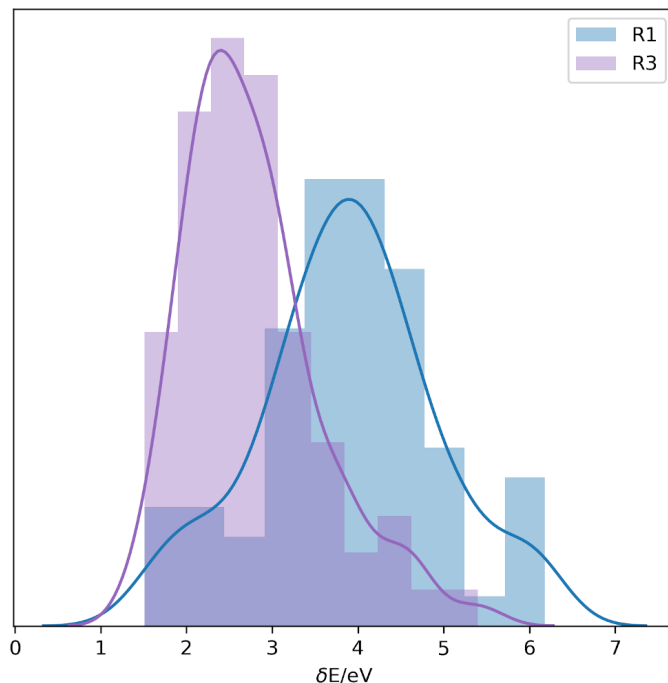
Figure 7: Density plots for the two major contributing typologies R1 and R3. Energies in eV.

were to be blocked due to pressure, coverage or multi-site catalysis effects that we currently cannot predict, we can still have a good qualitative picture of how the overall reactivity of the system should behave, instead of limiting our view to a single numerical value. Therefore, the network analysis capabilities that gTOFfee provides show the strengths of our approach even in cases that might seem far from the original reach of the energy span model.

Final remarks

We have developed a suitable computational implementation, gTOFfee, for the graph-based TOF calculation scheme originally proposed by Kozuch, using the minimum number of approximations in order to tackle the full complexity of catalytic cycles. In line with the original spirit of the energy span model, our new tool allows for a computationally inexpensive and versatile analysis of catalytic cycles, including the effects of reaction conditions such as temperature or reactant/product concentrations.

Apart from obtaining raw values for the turnover frequencies, more complex analyses of the reaction network are possible such as mapping product selectivities or changes in the reaction performance to different initial catalytic conditions. Moreover, the study of the formally defined individual reaction mechanisms provides yet another layer of information to work with in complex systems.

The new analysis of chemical reaction networks that we present is able to reproduce experimental trends both in homogeneous and heterogeneous catalysis, through very simple mathematical expressions, providing a quick, easy-to-setup tool for a profound analysis of computed catalytic cycles. We think that this combination of simplicity and completeness makes gTOFfee an interesting addition to the computational chemist's toolbox.

Acknowledgement

Supporting Information Available

Detailed description of gTOFfee implementation, details on reaction network construction, examples, additional graphs for hydroformylation mechanisms, tables of energy values for nodes and edges.

References

(1) Oster, G.; Perelson, A. Chemical Reaction Networks. *IEEE Transactions on Circuits and Systems* **1974**, *21*, 709–721.

(2) Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry* **1977**, *81*, 2340–2361.

(3) Othmer, H. A graph-theoretic analysis of chemical reaction networks. *Lecture Notes, Rutgers University* **1979**, 1–36.

(4) Craciun, G.; Feinberg, M. Multiple equilibria in complex chemical reaction networks: Extensions to entrapped species models. *IEE Proceedings: Systems Biology* **2006**, *153*, 179–186.

(5) Shinar, G.; Feinberg, M. Concordant chemical reaction networks and the Species-Reaction Graph. *Mathematical Biosciences* **2013**, *241*, 1–23.

(6) Broadbelt, L. J.; Stark, S. M.; Klein, M. T. Computer Generated Pyrolysis Modeling: On-the-Fly Generation of Species, Reactions, and Rates. *Industrial and Engineering Chemistry Research* **1994**, *33*, 790–799.

(7) Rappoport, D.; Galvin, C. J.; Zubarev, D. Y.; Aspuru-Guzik, A. Complex chemical reaction networks from heuristics-aided quantum chemistry. *Journal of Chemical Theory and Computation* **2014**, *10*, 897–907.

(8) Gao, C. W.; Allen, J. W.; Green, W. H.; West, R. H. Reaction Mechanism Generator: Automatic construction of chemical kinetic mechanisms. *Computer Physics Communications* **2016**, *203*, 212–225.

(9) Dewyer, A. L.; Argüelles, A. J.; Zimmerman, P. M. Methods for exploring reaction space in molecular systems. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2018**, *8*, 1–20.

(10) Proppe, J.; Reiher, M. Mechanism Deduction from Noisy Chemical Reaction Networks. *Journal of Chemical Theory and Computation* **2019**, *15*, 357–370.

(11) Hoops, S.; Gauges, R.; Lee, C.; Pahle, J.; Simus, N.; Singhal, M.; Xu, L.; Mendes, P.; Kummer, U. COPASI - A COmplex PAthway SImulator. *Bioinformatics* **2006**, *22*, 3067–3074.

(12) Kee, R. J.; Rupley, F. M.; Meeks, E.; Miller, J. a. *Chemkin-III: a fortran chemical kinetics package for the analysis of gas- phase chemical and plasma kinetics*; 1996; pp 3–164.

(13) Glowacki, D. R.; Liang, C. H.; Morley, C.; Pilling, M. J.; Robertson, S. H. MESMER: An open-source master equation solver for Multi-Energy well reactions. *Journal of Physical Chemistry A* **2012**, *116*, 9545–9560.

(14) Besora, M.; Maseras, F. Microkinetic modeling in homogeneous catalysis. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2018**, *8*, 1–13.

(15) Erhard, F.; Friedel, C. C.; Zimmer, R. FERN - A Java framework for stochastic simulation and evaluation of reaction networks. *BMC Bioinformatics* **2008**, *9*, 1–12.

(16) Walther, G. R.; Hartley, M.; Mincheva, M. GraTeLPy: Graph-theoretic linear stability analysis. *BMC Systems Biology* **2014**, *8*.

(17) Salehi, S. A.; Parhi, K. K.; Riedel, M. D. Chemical reaction networks for computing polynomials. *ACS Synthetic Biology* **2017**, *6*, 76–83.

(18) Kozuch, S.; Shaik, S. How to Conceptualize Catalytic Cycles? The Energetic Span Model. *Accounts of chemical research* **2011**, *44*, 101–110.

(19) Uhe, A.; Kozuch, S.; Shaik, S. Automatic analysis of computed catalytic cycles. *Journal of Computational Chemistry* **2011**, *32*, 978–985.

(20) 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 Catalysis* **2015**, *5*, 5242–5255.

(21) Solel, E.; Tarannam, N.; Kozuch, S. Catalysis: Energy is the measure of all things. *Chemical Communications* **2019**, *55*, 5306–5322.

(22) Garay-Ruiz, D. gTOFfee. `https://gitlab.com/dgarayr/gtoffee`, May, 2020.

(23) Kozuch, S. A refinement of everyday thinking: The energetic span model for kinetic assessment of catalytic cycles. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2*, 795–815.

(24) Kamer, P. C.; Reek, J. N.; van Leeuwen, P. W. In *Mechanisms in Homogeneous Catalysis: A Spectroscopic Approach*; Van Leeuwen, P. W. N. M., Claver, C., Eds.; Catalysis by Metal Complexes; Springer Netherlands: Dordrecht, 2005; Vol. 22; pp 231–269.

(25) Stanley, G. G. *Kirk-Othmer Encyclopedia of Chemical Technology*; Wiley, 2017; pp 1–19.

(26) Hood, D. M.; Johnson, R. A.; Carpenter, A. E.; Younker, J. M.; Vinyard, D. J.; Stanley, G. G. Highly active cationic cobalt(II) hydroformylation catalysts. *Science* **2020**, *367*, 542–548.

(27) Rush, L. E.; Pringle, P. G.; Harvey, J. N. Computational kinetics of cobalt-catalyzed alkene hydroformylation. *Angewandte Chemie - International Edition* **2014**, *53*, 8672–8676.

(28) Martin, R. L.; Hay, P. J.; Pratt, L. R. Hydrolysis of ferric ion in water and conformational equilibrium. *Journal of Physical Chemistry A* **1998**, *102*, 3565–3573.

(29) Wertz, D. H. Relationship between the Gas-Phase Entropies of Molecules and Their

Entropies of Solvation in Water and 1-Octanol. *Journal of the American Chemical Society* **1980**, *102*, 5316–5322.

(30) Cooper, J.; Ziegler, T. A density functional study of SN2 substitution at square-planar platinum(II) complexes. *Inorganic Chemistry* **2002**, *41*, 6614–6622.

(31) Morales, G.; Martínez, R.; Ziegler, T. Theoretical comparison of ketene dimerization in the gas and liquid phase. *Journal of Physical Chemistry A* **2008**, *112*, 3192–3200.

(32) González-Fabra, J.; Castro-Gómez, F.; Sameera, W. M.; Nyman, G.; Kleij, A. W.; Bo, C. Entropic corrections for the evaluation of the catalytic activity in the Al(iii) catalysed formation of cyclic carbonates from CO2 and epoxides. *Catalysis Science and Technology* **2019**, *9*, 5433–5440.

(33) Meir, R.; Kozuch, S.; Uhe, A.; Shaik, S. How can theory predict the selectivity of palladium-catalyzed cross-coupling of pristine aromatic molecules? *Chemistry - A European Journal* **2011**, *17*, 7623–7631.

(34) Zhao, Y.-F.; Yang, Y.; Mims, C.; Peden, C. H.; Li, J.; Mei, D. Insight into methanol synthesis from CO2 hydrogenation on Cu(111): Complex reaction network and the effects of H2O. *Journal of Catalysis* **2011**, *281*, 199–211.

Graphical TOC Entry