ACS Catalysis

Research Article

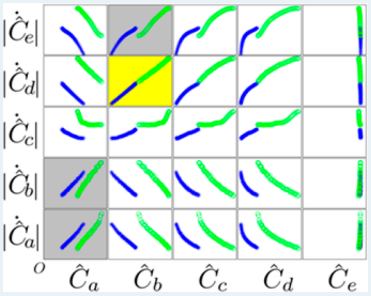
pubs.acs.org/acscatalysis

Two-Dimensional (2D) Correlation Analysis and the Search for Intermediates: A Strictly Mathematical Approach to an Important Mechanistic Question

Qisong Xu, LiangFeng Guo, Tung Nguyen Dinh, Angie Cheong, and Marc Garland*

Institute of Chemical and Engineering Sciences, Agency for Science, Technology and Research, 1 Pesek Road, Jurong Island 627833, Singapore

ABSTRACT: In situ spectroscopic studies of metal-mediated syntheses of new and previously unstudied systems are being increasingly used to better understand speciation and mechanistic aspects. These types of experiments give rise to an interesting question: namely, can one deduce from in situ data alone, and with no a priori chemical knowledge (i.e. chemical assignments), which pure component spectral estimates correspond to intermediates? In the present contribution, a statistical 2D correlation analysis is introduced to solve this problem for unicyclic catalytic systems. Such a methodological development achieves two goals: (1) it allows the experimentalist to concentrate on the most meaningful information at the outset of a new exploratory study (focus on the species directly associated with the catalysis), and (2) it helps to free the experimentalist from chemical bias and prejudice, i.e. believing that a specific organometallic species has to be an intermediate due to one or more chemical arguments, when in fact it may be just a side product or spectator species in the metal-mediated synthesis. The 2D correlation analysis is first tested with a numerically simulated data set and then with a real in situ FTIR data set from an unmodified rhodium-catalyzed hydroformylation. The resulting statistical 2D correlation analysis provides a clear and correct answer.



KEYWORDS: *hydroformylation, homogeneous catalysis, in situ FTIR, BTEM, intermediates*

1. INTRODUCTION

In the past few decades, in situ spectroscopic studies of metal-mediated organic syntheses have become considerably more common,[1] and here in situ FTIR[2] and in situ NMR[3] have played a central role. A few groups that contributed extensively to the early developments include those from (1) Monsanto,[4] ICI,[5] and Zurich,[6] focusing on understanding the fundamentals of organometallic transformations, and those from (2) Marburg,[7] Twente,[8] York,[9] Oxford,[10] Zurich,[11] Sheffield/BP,[12] Liverpool,[13] Amsterdam,[14] and Singapore,[15] focusing more on the catalysis. Catalysis science has benefited enormously from these early developments, since it was clearly demonstrated that means could be implemented to better understand the bracket and arrow (or "black box") of metal-mediated homogeneous organic transformations, which are essential to today's bulk chemical, fine chemical, and pharmaceutical industries (Figure 1).[16,17]

Although these earlier studies established the usefulness of in situ spectroscopy, it became rather clear that there is often a large gap between the observations and a confident understanding of the catalysis. This situation existed in large part due to the numerous new spectroscopic bands which arose in many studies. It became important to find a means of untangling the spectroscopic information so that proper chemical assignments could be made and subsequently to get an accurate estimation of concentrations and rates.

In the past decade, advanced signal processing has been applied on a large scale, to hundreds if not thousands of in situ

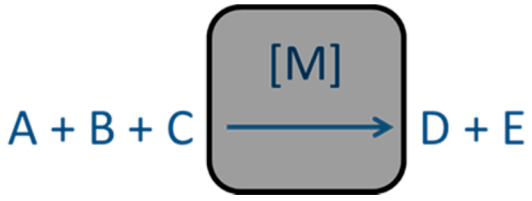


Figure 1. Commonly used and generic representation of a liquid-phase metal-mediated homogeneous organic transformation, where the bracket and arrow emphasize the need for a metal and the catalytic nature of the transformation. A black box has been added for visual and conceptual emphasis. In situ methods such as FTIR and NMR have targeted a better understanding (speciation) of the black box.

spectra from a system. The main goal of such numerical work has been to obtain sets of pure component spectral estimates.[18] Examples of such algorithms used in catalysis and based on entropy minimization include minimization of entropy and spectral similarity (MESS)[19] and pure component decomposition (PCD),[20] which generate the set simultaneously, and band-target entropy minimization (BTEM), which generates the pure component spectral estimates one at a time.[21]

When such pure component spectral estimates are first generated by the above algorithms, they are all normalized

Received: December 31, 2014
Revised: April 9, 2015
Published: May 14, 2015

DOI: 10.1021/cs502127y
ACS Catal. 2015, 5, 3588−3599

(usually scaled to unity). Moreover, if the reaction is truly homogeneous and the species are all distinct small molecules (in contrast to distributions of colloids etc.), one can introduce the initial mass balance(s) and solve simultaneously the stoichiometries corresponding to each spectral estimate.[22] By fitting the normalized pure component spectral estimates onto the original data, "relative concentration" profiles can be obtained, and by fitting the calibrated pure component spectral estimates onto the original data, "real concentration" profiles can be obtained. If a synthesis is not "clean", i.e. either colloids or precipitates form, the initial mass balance is not valid for the liquid phase for the entire reaction period. Therefore, for many systems, the absorptivities cannot be calibrated (analogously, the stoichiometries cannot be determined) and the "real concentration" profiles cannot be obtained. Thus, the take-home message is clear; in exploratory studies, the experimentalist might be forced to be content with just normalized pure component spectral estimates and "relative concentration" profiles. This situation is certainly unfortunate, but it often represents real-world synthetic situations.

The identification of intermediates is often the primary goal of an exploratory study.[23] This being the case, the first question is often as follows: what can be further done with just the normalized pure component spectral estimates and "relative concentration" profiles? This initial question then gives rise to a much more specific and very useful question: can one deduce from in situ data alone, and with no a priori chemical knowledge, which pure component spectral estimates correspond to intermediates? If such a methodological development could be achieved, then two problems are overcome: (1) it allows the experimentalist to concentrate on the most meaningful information (signals) at the outset of a new exploratory study (the species directly involved in the catalysis), and (2) it frees the experimentalist from chemical bias and prejudice, i.e. believing that an organometallic species has to be an intermediate when in fact it may be just a side product or a spectator species in the metal-mediated synthesis.

In the present contribution, we introduce a set of purely statistical tests in order to determine if a normalized pure component spectrum represents an intermediate in the metal-mediated synthesis or if it is just a side product or a spectator species. In order to convey the statistical test results in a clear manner, the results are assembled in a visual 2D correlation format. In the first instance, this approach is applied to a known numerically simulated model system, and in the second instance, it is applied to the unmodified rhodium-catalyzed hydroformylation of cyclopentene to form cyclopentanecarboxaldehyde using in situ FTIR as the experimental spectroscopy. The resulting statistical 2D correlation analysis provides a clear and correct answer in both cases.[24]

2. EXPERIMENTAL ASPECTS

2.1. General Considerations. In the following, the well-studied unmodified rhodium-catalyzed hydroformylation reaction is used as a model system.[25] Specifically, cyclopentene is used as the substrate and $Rh_4(CO)_{12}$ as the catalyst precursor. The catalytic system is homogeneous, with no detectable light scattering. In situ FTIR normally shows the presence of only $Rh_4(CO)_{12}$, $C_5H_9CORh(CO)_4$, and $Rh_6(CO)_{16}$ as observable organometallics and only cyclopentene and cyclopentanecarboxaldehyde as observable organics in the active system. It has been shown repeatedly that there is a 1:1 relationship between the instantaneous concentration of the intermediate $C_5H_9CORh(CO)_4$ and the rate of aldehyde formation. In other words, $C_5H_9CORh(CO)_4$ is a true intermediate.[26]

2.2. Chemicals and Preparations. All solution preparations and transfers were carried out under a purified argon (99.9995%, Soxal, Singapore) atmosphere using standard Schlenk techniques or conducted in a glovebox (MBRAUN UNIlab). Carbon monoxide (research grade, 99.97%, Soxal, Singapore) and hydrogen (99.9995%, Soxal, Singapore) were further purified through columns packed with activated 4 Å molecular sieves, BTS catalyst, Faujesite (CBV 780 CY (1.6)), and β-zeolite (CP 811C-300 CY (1.6)) to eliminate trace amounts of water and $Ni(CO)_4$ from the CO. *n*-Hexane (99.5%, Sigma-Aldrich puriss) was purified by distillation from NaK under argon. $Rh_4(CO)_{12}$ (98%, Strem) was used without further purification. Cyclopentene (99%, Fluka) was purified by distillation from maleic anhydride to remove dienes formed by Diels–Alder cycloaddition, followed by distillation from CaH_2 under argon.

2.3. Equipment. The in situ spectroscopic studies were conducted with the use of a hermetically sealed recycling reaction system, as shown in Figure 2. The system included (a) an in-house-designed and -fabricated SS316 stirred tank reactor with an internal volume of 100 mL, (b) 1/16 in. SS316 transfer lines, (c) a hermetically sealed magnetically coupled gear pump (Model GAH-X21, Micro pump, USA), (d) an SS316 high-pressure (HP) flow-through infrared cell with CaF_2 windows separated by a Teflon spacer (path length) of 200 μm in thickness, (e) high-pressure (HP) SS316 syringe pumps for injection of solutions (Model PHD 4400, Harvard Instruments, USA), and (f) a cryostat (Poly-Science, USA). A rather complete description of the system is provided elsewhere.[27]

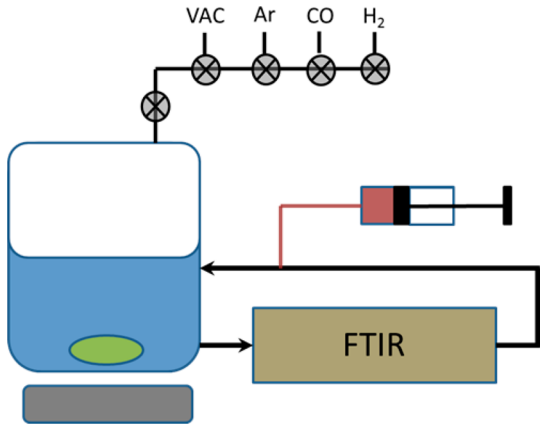


Figure 2. Schematic of the recycling reaction system used for the present in situ FTIR spectroscopic studies. The HP syringes facilitate batch–feed operation with perturbations which are especially important for subsequent BTEM analysis.

Transport effects in the present system were minimal. The liquid phase became saturated with dissolved gas on the order of 6 min, and the recycle time was on the order of τ = 2 min. Thus, in ca. 6 min (three recycle times), the system is, to a very good approximation, thoroughly mixed. The rate of conversion of substrate to product was such that concentrations of substrate, dissolved CO, and dissolved H_2 did not vary more than 5% over the length of the recycle loop. The present reaction system belongs to category H of the Hatta reaction regimes: i.e., an infinitely slow reaction in comparison to mass transfer.[28]

DOI: 10.1021/cs502127y
ACS Catal. 2015, 5, 3588–3599

Measurements were performed on a Bruker Vertex 70 FTIR spectrometer equipped with a DTGS ambient detector and KBr beam splitter. Spectra were acquired at 2 cm^{-1} resolution, from 600 to 4000 cm^{-1}, using 16 coadds per spectrum at intervals of ca. 2 min. Zero filling was used in order to provide data points with ca. 0.1 cm^{-1} separation. The other IR measurement parameters are as follows: order of zero filling, 16; aperture setting, 6 mm; preamp gain, A; scanning mirror velocity, 10 kHz; use of optical filter, open.

2.4. Experimental Design. An appropriate experimental design for exploratory studies in combination with BTEM analysis should involve a few well-chosen perturbations in reagents. The purpose of these perturbations is to obtain more variation in the signals present (particularly of the organometallics present). By performing a perturbation, one upsets the monotonic and asymptotic trajectories of the reactant concentrations and hence signal ratios. This results in better pure component spectra estimates.

Moreover, in the present study for 2D correlation analysis, it will be imperative to aim for very small substrate changes and low yields of product, since both initial rate and pseudo-steady-state conditions for the organics are desirable (even though a batch reactor is being used). Taken together, the two aforementioned prerequisites must be balanced in a semi-batch experiment. Table 1 shows such an experimental design. Evidence to support the proposition that both prerequisties are satisfied is provided in the sections 4 and 5.

Table 1. Experimental Design for the Present Study[a]

perturbation step	action	timeline
1	infuse initial solution[b]	spectral acquisition[e]
2	add CO (20 bar) then H_2 (20 bar)	spectral acquisition[f]
3	add $Rh_4(CO)_{12}$[c]	t = 0 h
4	add $Rh_4(CO)_{12}$[d]	t = 2 h
5	add $Rh_4(CO)_{12}$[d]	t = 3.8 h
6	end of experiment	t = 4.5 h, end of acquisition

[a]The stock solution was prepared as 11.9 mg of $Rh_4(CO)_{12}$ in 8 mL of *n*-hexane. The reaction temperature was 293 K. [b]Conditions: cyclopentene (1 mL) in *n*-hexane solution (32 mL) and then addition of CO (20 bar) and H_2 (20 bar). [c]Conditions: 3 mL of stock solution. [d]Conditions: 2.5 mL of stock solution. [e]Initial spectra before catalytic reaction. [f]Continuous acquisition.

The data from the last two rows (entries 5,6) in Table 1 were not used in the final 2D correlation analysis (see section 3.5). It was found that the conversion of alkene and the yield of aldehyde were too high and hence the assumption of pseudo-first-order kinetics is no longer valid. Thus, only data from times 0−3.8 h (perturbation steps 1−4) were used.

2.5. Calibrations. The solvent intensity was used as an internal standard. Calibrating or not calibrating the solvent would not change the final interpretation of the 2D correlation matrix.[29]

Spectra of the solvent *n*-hexane were acquired using a special variable path length cell (Model SP07500, with ZnSe window, Specac U.K.). Seven path lengths (25, 50, 75, 100, 125. 150, 175 μm) were used, and spectra were recorded on a Bruker Vertex 70 FTIR spectrometer equipped with a KBr beam splitter and DTGS detector. The system was purged with dry air at a rate of 10 L/min. Spectra were acquired at 2 cm^{-1} resolution, from 600 to 4000 cm^{-1}, using 16 coadds per spectrum at intervals of ca. 2 min. Zero filling was used in order to provide data points with ca. 0.1 cm^{-1} separation. The *n*-hexane calibration gave a molar absorptivity of 2.977 L/(mol cm) at ν = 1136 cm^{-1}.

3. COMPUTATIONAL ASPECTS

3.1. Spectral Processing. Raw spectra were truncated for analysis of specific spectral windows. The truncated spectra were renormalized by using the respective molar absorptivity of the solvent at its characteristic band, resulting in a set of dimensionless spectra for further quantitative analysis.[30]

Equation 1 is the main equation used for initial processing of the experimental spectra, where $\mathbf{A}^{\mathrm{exp}}_{k\times\nu}$ is the raw experimental

$$\mathbf{A}^{\mathrm{renorm}}_{k\times\nu} = \Gamma_{k\times k}\mathbf{A}^{\mathrm{exp}}_{k\times\nu} \quad (1)$$

reaction mixture matrix, $\Gamma_{k\times k}$ is a diagonal renormalization factor, and $\mathbf{A}^{\mathrm{renorm}}_{k\times\nu}$ is the dimensionless absorbance matrix. The index k is the number of experimental spectra, and the index ν is the number of channels of data in each spectrum. The dimensionless absorbance matrix is needed in order to account for changes (i) in reaction volumes either due to thermal/pressure changes or the addition of additional components during the reaction (semi-batch operation, perturbations) and (ii) in the cell path length—again due to pressure or temperature or swelling effects of the O-rings.

3.2. BTEM Analysis and Pure Component Spectral Estimation. The band−target entropy minimization (BTEM) algorithm[21] was used in order to obtain pure component spectral estimates of the solutes present in the system, by analyzing appropriate sets of reaction spectra. BTEM has been repeatedly used to study complex metal-mediated homogeneous catalytic organic syntheses.[18,21,23,30]

Equation 2 is the main equation used for BTEM analysis, where $\hat{a}_{1\times\nu}$ is a BTEM pure component spectral estimate, $\mathbf{T}_{1\times z}$

$$\hat{a}_{1\times\nu} = \mathbf{T}_{1\times z}\mathbf{V}^{\mathrm{T}}_{z\times\nu} \quad (2)$$

is an optimized transformation vector, and $\mathbf{V}^{\mathrm{T}}_{z\times\nu}$ is the matrix of right singular vectors (obtained from a singular value decomposition (SVD) of the spectroscopic data $\mathbf{A}$). The indices are ν for the total number of channels of data in each spectrum and z for the number of vectors used, where normally $k \gg z \gg s$, where s is the number of observable species present. During BTEM analysis, a global search is performed to find the simplest spectral estimate with the lowest signal entropy.

After an exhaustive one by one spectral search, the relative concentration matrix $\hat{c}_{k\times s}$ can be calculated as in eq 3, where $\mathbf{A}^{\mathrm{renorm}}_{k\times\nu}$ is the dimensionless absorbance matrix and $(\hat{a}_{s\times\nu})^{+}$ is the generalized inverse of the matrix of normalized BTEM spectral estimates.

$$\hat{c}_{k\times s} = \mathbf{A}^{\mathrm{renorm}}_{k\times\nu}(\hat{a}_{s\times\nu})^{+} \quad (3)$$

3.3. Pseudo-Steady-State Single-Product Multipath-way Unicyclic Catalysis. An accurate working definition for the catalytic system is needed in order to place the 2D correlation into the proper perspective. Accordingly, an abstract block-diagram relationship among precursor(s), degradation product(s), and a generalized unicyclic catalytic cycle is shown in Figure 3. The precursor(s) may be metal salt(s) or

DOI: 10.1021/cs502127y
ACS Catal. 2015, 5, 3588−3599

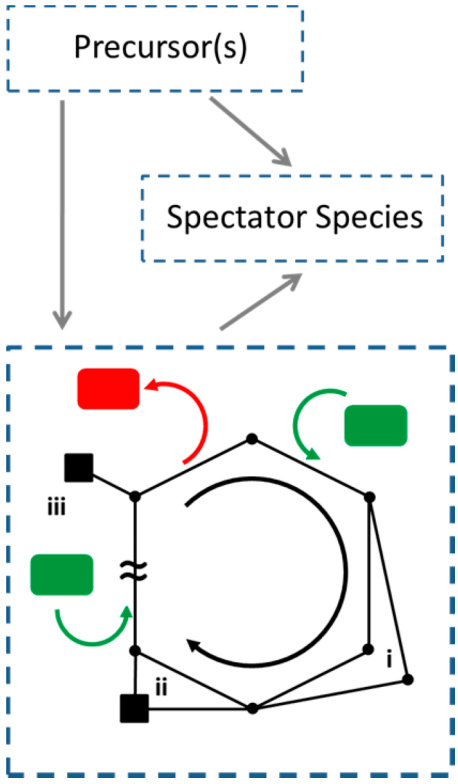


Figure 3. Abstract representation of the three distinct pools of metal-containing species in single-product multipathway unicyclic catalysis. See text for further explanation.

organometallic(s), and modifiers such as phosphorus- or nitrogen-containing ligands may be included. The spectator species are considered to be metal-containing inorganic or organometallic complexes that are neither precursors nor intermediates. In other words, they do not participate in product formation. They may be formed by reaction of the precursor or an intermediate with impurities, the reactor wall, and/or reactants. The spectator species may also arise due to thermal degradation. The gray arrows in Figure 3 are only an indication of the usual direction for the net flux of metal from one block to another and in no way are meant to suggest the reactions are irreversible. Indeed, measurable equilibria between precursors and intermediates have been observed at least twice (so-called equilibrium-controlled precursor conversion),[26c,31] and selectivities for the conversion of precursors to spectator species between 0% and 100% have been observed.[32] The green boxes and arrows in the "catalytic block" are used to indicate the addition of reagents (from the overall reaction, Figure 1) to the cycle, and the red box and arrow are used to indicate the singular product of the overall reaction.

The most common graph representation for a catalytic cycle will be used: namely, the nodes represent intermediates and the edges represent reactions.[33] Moreover, the following features have been added to the single-product multipathway unicyclic catalysis. (1) An arrow in the center of the diagram will represent the direction of next product formation, (2) The dots represent intermediates having the lowest concentrations. (3) Black boxes represent intermediates with the highest concentrations and, therefore, may be observable if appropriate in situ spectroscopy is used. (4) The branch (i) indicates that two pathways exist, in this case, due to the presence of isomers. (5) The branch (ii) indicates that a transformation can be performed either in a single concerted manner or as two sequential elementary reactions. (6) The object (iii) represents a classic example of a reservoir—a species in equilibrium exchange with an intermediate on a cycle. (7) the symbol ≈ adds further emphasis to the fact that this diagram is general—there may be many more elementary steps present.

Figure 3 was drawn with a number of important points in mind. One point is that, in general, a single product synthesis will have multiple, independent reaction pathways, even if it has an "isolated unicycle mechanism". This issue can be visualized more readily by hypothetically "labeling" individual substrate molecules. In this manner, one can trace each pathway taken. Thus, suppose for the moment a hypothetically labeled molecule takes the most probable pathway through the mechanism. This might be pathway a in Figure 4. The next

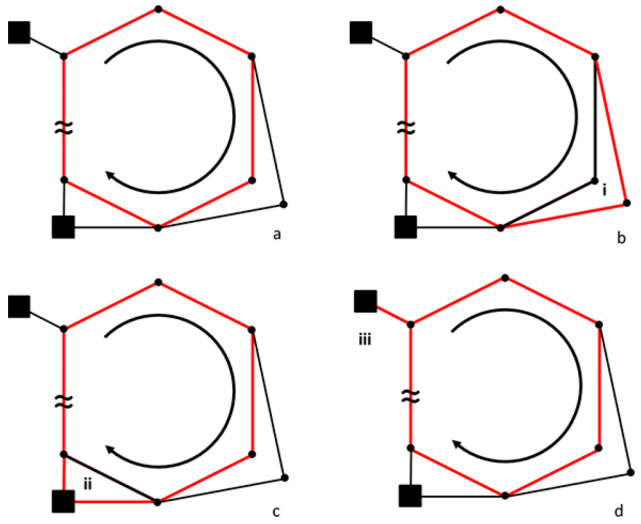


Figure 4. Four possible pathways highlighted in red, for a "hypothetically labelled substrate molecule" through the single-product multipathway unicyclic catalytic cycle: (a) most direct and probable pathway through cycle; (b) a pathway through an isomer; (c) a pathway through two-step transformation versus concerted transformation; (d) a pathway involving exchange with the reservoir at least one time. Additional pathways can be generated as permutations/combinations of the above four examples.

substrate molecule might take the least likely isomer pathway at the branch point (i) and complete the pathway (Figure 4b). The third substrate molecule might take the two-step pathway at the branch point (ii) and complete the pathway (Figure 4c). The fourth substrate molecule substrate molecule might take pathway iii, thereby populating a reservoir at least once during its journey to product formation (Figure 4d). Then, of course, there are a number of permutations of the above scenarios that can be envisaged. In conclusion, Figure 4 shows that, depending on the pathway taken, there are different intermediates.

Taking the whole picture into consideration, all nodes must be considered to be intermediates. Thus, a sufficiently robust single-product multipathway unicyclic catalytic cycle must accommodate numerous branching issues and recognize that all nodes are potential intermediates—dependent on the specific pathway taken. Another important purpose of Figures 3 and 4 is to bring self-consistency between a mathematically rigorous description of single-product multipathway unicyclic catalysis and features which are familiar to the catalytic chemist.

DOI: 10.1021/cs502127y
ACS Catal. 2015, 5, 3588−3599

King and Altman derived a general matrix solution for enzyme catalysis,[34] and this description was expanded by others.[35] Upon close inspection it is evident that enzyme mathematical results extend directly to the case of homogeneous metal-mediated catalysis, given one important prerequisite; each and every intermediate has the same nuclearity, and therefore there are no bimolecular reactions between intermediates. Subsequently, one observes that the features introduced in Figures 3 and 4, i.e. branch points, reservoirs, etc., are admissible within the model. The mathematical description of the activity of such a unicyclic system, in the approximation of a steady-state or pseudo-steady-state product formation rate r_p, can be expressed in terms of either the concentration of each and every ith intermediate I_i (eq 4) or the total sum of all intermediate concentrations (eq 5), where the terms k_i and k_{TOF} contain only constants and terms involving reactant concentrations (but not concentrations of organometallics):

$$r_p = k_i[I_i] \ \forall \ i \tag{4}$$

$$r_p = k_{TOF} \sum [I_i] \tag{5}$$

Specifically, eq 5 is useful as the formal definition for the well-known turnover frequency (TOF) of a reaction cycle, especially when it is evaluated in terms of instantaneous reaction rates and intermediate concentrations at any time t.

In the context of the present contribution, the most central and important concept concerning unicyclic mechanisms is that the resulting system activity is linear in each intermediate concentration (eq 4). Consequently, in the regression of real rate data (criterion I) the intercept should be in the close vicinity of 0 and (criterion II) the slope should be a close approximation of a straight line, with little or no hint of curvature. Clearly, the two aforementioned criteria can be implemented in practice as statistical tests. Criteria I and II will serve as the starting points for the following 2D correlation analysis.

3.4. Initial Catalysis Startup. In most real-world homogeneous catalytic experimentation, the experimentalist will use a batch reactor, although the past decade has seen more examples of continuous-flow systems.[36] At the moment that all reactants are brought together, but the fluid elements are not yet intimately mixed, there are no intermediates yet, and the catalytic cycle is unpopulated. A finite time is required in order to start populating the cycle. This finite delay time is also seen as a delay between the time a substrate goes into a cycle and a product comes out, and this delay is on the order of TOF^{-1}, where TOF is the turnover frequency. There are also delay times associated with mass transfer from the gas phase into the liquid phase, mixing of the liquid phase, and mixing of all of the liquid throughout the experimental apparatus shown in Figure 2. Therefore, the earliest part (first few minutes) of initial rate data will not be appropriate for 2D correlation analysis.

3.5. Scatter Plots and Statistical Analysis. As mentioned above, careful design of the experiment is essential for meaningful statistical analysis. Experimental conditions were chosen with an excess of reactants, and small extents of reaction, in order to ensure that catalytic system follows pseudo-first-order kinetics as closely as possible.

The statistical analysis proceeds in the following manner. First, BTEM analysis is applied in order to get the pure component spectral estimates and the associated relative concentrations. Next, a trajectory for each rate of change $|\dot{C}_y|$ (for any arbitrary yth species) is generated against each concentration C_x (for any arbitrary xth species). This produces a scatterplot for each and every combination of observable species x and y. Since $\max\{x\} = \max\{y\} = s_{obs}$, there are a total of s_{obs}^2 scatterplots to be evaluated. These s_{obs}^2 scatterplots will form the basis for the 2D correlation analysis.

For any particular scatterplot, a set of unit tangent vectors can be calculated from the successive data points (Figure 5).

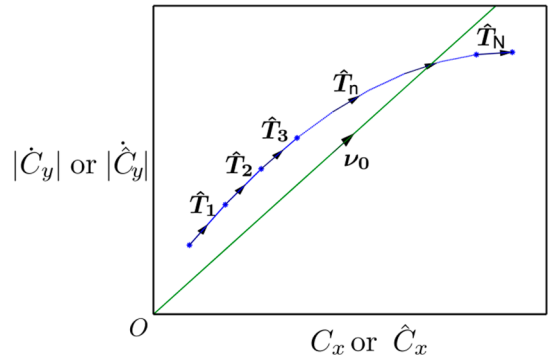


Figure 5. Representation of a scatterplot where ν_0 is a unit vector and $\hat{T}_n$ is the successive unit tangent vectors. The x axis can be either real concentration C_x or relative concentration $\hat{C}_x$. The y axis can be either absolute real concentration rate $|\dot{C}_y|$ or absolute relative concentration rate $|\hat{\dot{C}}_y|$, respectively.

The scalar product of each tangent vector with a unit vector (directed exactly through the origin) gives a series of similarity coefficients. This sample of coefficients was tested statistically to assess if the data points satisfy two criteria: (1) they pass through the point of origin, and (2) they have a general linear trend.

As mentioned previously in section 3.3, a true intermediate will possess two necessary characteristics: namely, criterion I, which is that the intercept is 0, and criterion II, which is that the slope is a constant. Therefore, restating these criteria in the framework of Figure 5 provides testable hypotheses.

Criterion I: data points that are directed toward the vicinity of the origin should have similarity coefficients with values approaching unity. Thus, the first criterion was assessed by a hypothesis test—the Sign test. The Sign test is a nonparametric procedure which tests hypothesis on the median of any continuous distribution.[37] The null hypothesis is that the sample median of similarity coefficients is close to a specified median $\tilde{\mu}_0$ while the alternative is that the sample median is lower than $\tilde{\mu}_0$. The one-tailed test was performed using a lower tail at a significance level of 0.05.

Criterion II: the slopes of the data points are generally straight and should have mean stationary similarity coefficients. Thus, the second criterion was assessed by a hypothesis test—the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test. The KPSS procedure parametrizes a time series for the hypothesis test of stationarity.[38] The null hypothesis is that the sample of similarity coefficients is mean stationary while the alternative is that the sample is a nonstationary unit root series. The one-tailed test was performed using an upper tail at a significance level of 0.05.

If both criteria I and II are fulfilled, then we will say that the variables are "linearly correlated". This statistical hypothesis testing can be applied to the respective scatterplots from different experiments. The final result is a two-dimensional

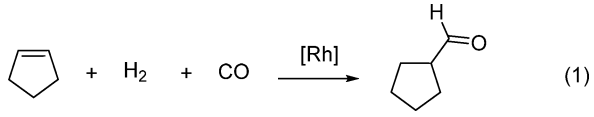
DOI: 10.1021/cs502127y
ACS Catal. 2015, 5, 3588–3599

correlation matrix which identifies possibly high correlation between $|\dot{C}_y|$ and C_x or $|\hat{C}_y|$ and $\hat{C}_x$.

3.6. Computational Platforms. All of the simulations and signal processing were performed with in-house-developed algorithms (BTEM), using Matlab 7.12.[39] Numerical calculations were performed on an Intel Pentium CPU G640 2.80 GHz processor with 4 GB of RAM.

3.7. Numerical Simulations of Catalytic System. A semi-batch reaction for the rhodium-catalyzed hydroformylation of alkene was numerically simulated so that all characteristics of the system could be controlled. The precursor $Rh_4(CO)_{12}$ reacts with CO, H_2, and alkene to produce an acylrhodium species—the intermediate, the side product $Rh_6(CO)_{16}$, and a product aldehyde. The simulated reaction was assumed to be performed at constant temperature and pressure. The overall reaction can be represented as Scheme 1.

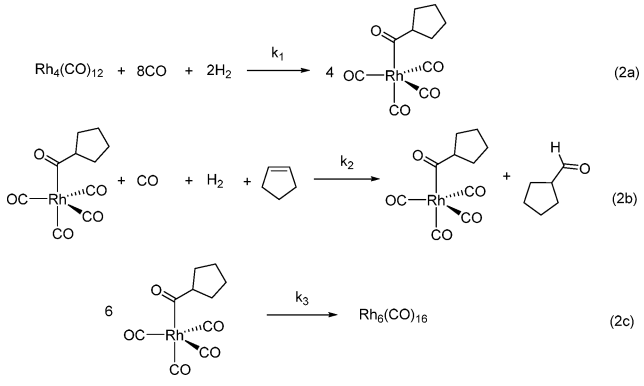
Scheme 1. Overall Hydroformylation of an Alkene



(1)

A simplified reaction model for unmodified hydroformylation was constructed for simulating the overall reaction and thus testing the 2D correlation analysis. The details of this catalytic mechanism are provided in Scheme 2. Reaction 2a is the

Scheme 2. Conversion of the Organometallic Precursor to the Acyl Intermediate (a), Formation of the Product Aldehyde and Conversion of the Intermediate Back to Itself (b), and Unbalanced Reaction for Degradation of the Intermediate (c)



conversion of the organometallic precursor $Rh_4(CO)_{12}$ to the observable intermediate $RCORh(CO)_4$, where R is an alkyl moiety. Reaction 2b is an overall reaction from the observable intermediate back to itself and the formation of the product aldehyde (note that the delay in the catalytic cycle cannot be fully represented by this equation alone). Reaction 2c is an unbalanced reaction (vide infra).

Reactions 2a and 2b in Scheme 2 are based on actual detailed modeling of hydroformylation reactions. Thus, at constant partial pressures of CO and H_2, the rates of formation or disappearance for the catalyst precursor $Rh_4(CO)_{12}$, the intermediate $RCORh(CO)_4$, the substrate alkene, and the product aldehyde are simplified—only apparent rate constants are needed. In this model the respective apparent rate constants are set to be $k_1 = 2.3 \times 10^{-4}$ s^{-1} and $k_2 = 2.5 \times 10^{-3}$ s^{-1}, consistent with real experimental data.[26] These rate constants are approximately the apparent rate constants previously observed in real unmodified hydroformylations at ca. 20 bar of CO and 20 bar of H_2.

In order to keep the test system interesting, a rate expression is temporarily and artificially imposed in the numerical simulation for the formation of $Rh_6(CO)_{16}$ (reaction 2c in Scheme 2). In this proposed expression, it is assumed that the observable intermediate $RCORh(CO)_4$ degrades to form $Rh_6(CO)_{16}$ through an unknown but first-order rate process, where $k_3 = 10^{-6}$ s^{-1}. Indeed, $Rh_6(CO)_{16}$ is frequently observed from in situ unmodified rhodium-catalyzed hydroformylations, and the mechanisms by which it forms have not been clearly identified. Some routes clearly involve reactions with impurities (water, oxygen); thus, the process is definitely complex. Thus, for the purpose of our numerical simulations, we assume only one mechanism and it is first order.

The specific rate expressions for each of the five species are given respectively in eqs 6−10, where the exponent is $a = 0.15$, a value observed experimentally.[26]

$$\frac{d[Rh_4(CO)_{12}]}{dt} = -\frac{k_1}{4}[Rh_4(CO)_{12}][alkene]^a \quad (6)$$

$$\frac{d[RCORh(CO)_4]}{dt} = +k_1[Rh_4(CO)_{12}][alkene]^a \quad (7)$$

$$\frac{d[RCHO]}{dt} = +k_2[RCORh(CO)_4] \quad (8)$$

$$\frac{d[alkene]}{dt} = -k_2[RCORh(CO)_4] \quad (9)$$

$$\frac{d[Rh_6(CO)_{16}]}{dt} = +k_3[RCORh(CO)_4] \quad (10)$$

Given an initial concentration of the each species, this set of ordinary differential equations (ODEs) was solved numerically to generate the time-dependent concentration profiles of every species. A Matlab inbuilt function, ode45 (Runge−Kutta method), was used to perform the numerical integration procedure. The amounts of $Rh_4(CO)_{12}$ added in each perturbation for the simulated batch reaction are given in Table 2.

Table 2. "Experimental Design" for Numerical Simulation

perturbation step	action	timeline
0	start of numerical experiment with 0.03 mol of alkene	before start of reaction
1 (frame 1)	add $Rh_4(CO)_{12}$ to 4.10×10^{-5} mol/L	$t = 0$ h
2 (frame 2)	add $Rh_4(CO)_{12}$ to 4.10×10^{-5} mol/L	$t = 2$ h
3	end of numerical experiment	$t = 4$ h

The simulation provided concentration profiles for each solute. To these profiles random noise was added to reflect the stochastic fluctuations in the experimental conditions. The simulated system was made to undergo a perturbation by means of addition of precursor $Rh_4(CO)_{12}$ alone without the introduction of additional solvent.

DOI: 10.1021/cs502127y
ACS Catal. 2015, 5, 3588−3599

4. RESULTS

4.1. Numerical Simulation. The test system outlined in section 3.7 was numerically simulated. The concentration profiles are shown in Figure 6. The first frame (frame 1) occurs between reaction times of 0 and 2 h. The figure also clearly shows the perturbation introduced at t = 2 h in the 4 h experiment. The perturbation with the addition of precursor $Rh_4(CO)_{12}$ at the end of the second hour introduces a second frame (frame 2). More importantly, it is seen that the concentrations of the organic species (alkene and aldehyde) vary only a few percent. This is important, since it fulfils the prerequisite of pseudo-steady-state conditions, and hence enforces the likelihood of pseudo-first-order kinetics for true intermediates (section 3.3).

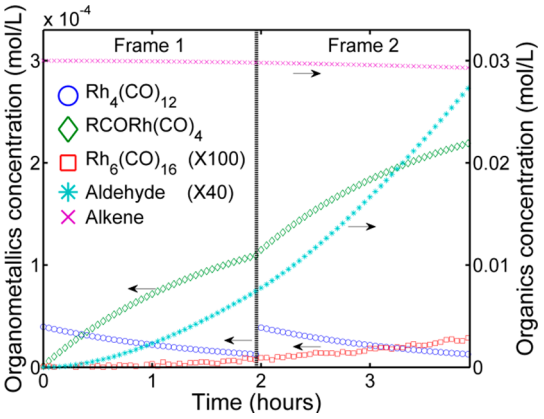


Figure 6. Time-dependent concentration profiles of the solutes (with perturbation at the second hour) for the numerically simulated hydroformylation reaction. For clarity in presentation, the figure is plotted with every third data point.

The solute concentration profiles from Figure 6 were used to assemble an associated set of instantaneous reaction rates for each species. Therefore, there is a correspondence between any species concentration C_x in time and an associated set of reaction rates for each species rate $|\dot{C}_y|$. Taking one species profile concentration C_x in time and any individual species rate in reaction time rate $|\dot{C}_y|$ produces a scatterplot.

All possible scatterplots from both reaction frames were assembled into a two-dimensional (2D) representation as shown in Figure 7. This 2D representation acts as a graphical tool that depicts the relationship among all possible scatterplots. Each individual scatterplot has a concentration along the x axis and a rate along the y axis (the absolute value of the rate is used to keep representations simple). The origin of each individual scatterplot lies at 0 for both axes, and the maximum value of each axis corresponds to the maximum value of the entry observed during the numerical simulation. Thus, the trends in variables are easily visualized in each individual scatterplot. A linearly correlated scatterplot has (1) data points that form generally straight lines, (2) each straight line, or its extrapolation has an intercept in the vicinity of the origin, (3) one straight line for frame 1, and (4) one straight line for frame 2. In Figure 7, scatterplots that contain linearly correlated variables are highlighted in yellow for easy identification. It is clear that the remaining scatterplots do not satisfy all the criteria for linearity stated above.

In a more distilled form, it can be said that the data points of each frame were assessed with respect to only two main criteria:

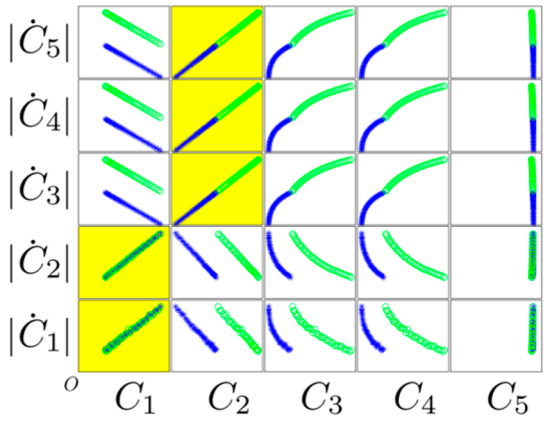


Figure 7. Scatterplot matrix produced from numerically simulated data set. Each entry of the matrix gives a visual display of the possible relationship between rate $|\dot{C}_y|$ and concentration C_x. Legend: (1) $Rh_4(CO)_{12}$; (2) $RCORh(CO)_4$; (3) $Rh_6(CO)_{16}$; (4) aldehyde; (5) alkene. Blue data points belong to frame 1, and green data points belong to frame 2. The x axis domain for each entry has the range of [0, max{C_x}], and the y axis domain for each entry has the range of [0, max {$|\dot{C}_y|$}].

criterion I (pass through the point of origin) and criterion II (have a general linear trend), as described in detail in section 3.5. Both hypothesis tests were applied to all possible scatterplots in all reaction frames, in order to reinforce conclusions about scatterplots which are truly linear. The final result of the hypothesis testing on the 2D representation of the scatterplot matrix was then mapped onto a 2D correlation matrix, as shown in Figure 8.

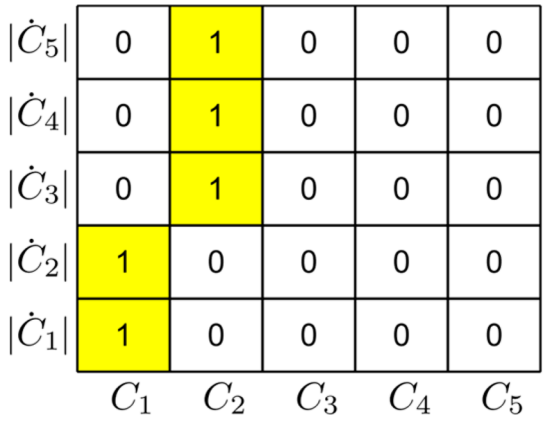


Figure 8. Final 2D correlation matrix of the numerically simulated catalytic system after application of statistical tests to the scatterplot matrix. Each entry of the matrix expresses a yes−no answer to the statistical tests and hence the linearity of rate $|\dot{C}_y|$ versus concentration C_x. The value 1 will be assigned to positive results for both the Sign test and KPSS test, while the value 0 represents otherwise. Legend: (1) $Rh_4(CO)_{12}$; (2) $RCORh(CO)_4$; (3) $Rh_6(CO)_{16}$; (4) aldehyde; (5) alkene.

If and only if both tests are positive, a matrix entry of unity is assigned (yellow); otherwise, a matrix entry of 0 is assigned (white). On the basis of the 2D correlation matrix shown in Figure 8, the following can be concluded. (1) The concentration of species 1 is linearly related to both the rates of species 1 and species 2. This implies that the concentration of $Rh_4(CO)_{12}$ is linearly related to the rates of $Rh_4(CO)_{12}$ and

DOI: 10.1021/cs502127y
ACS Catal. 2015, 5, 3588−3599

$RCORh(CO)_4$. (2) The concentration of species 2 is linearly related to rates of species 3, species 4, and species 5. This implies that the concentration of $RCORh(CO)_4$ is linearly related to the rates of $Rh_6(CO)_{16}$, aldehyde, and alkene. Reflecting back on the original model (section 2.3), it is verified that the 2D correlation analysis correctly identifies the linearly correlated scatterplots. In the numerically simulated system (Scheme 2) there are three linear rate constants and these rates involve all five solutes. $Rh_4(CO)_{12}$ and $RCORh(CO)_4$ are connected by Scheme 2a, $RCORh(CO)_4$ is connected to alkene and aldehyde by Scheme 2b, and $RCORh(CO)_4$ and $Rh_6(CO)_{16}$ are connected by Scheme 2c.

4.2. Experimental Test. The aforementioned 2D correlation analysis was then applied to the experimental rhodium-catalyzed hydroformylation of cyclopentene to cyclopentanecarboxaldehyde using in situ FTIR as the experimental spectroscopy. The hydroformylation of alkene was performed with the experimental conditions specified in section 2.4. The reaction duration was 3.8 h, with the addition of additional precursor $Rh_4(CO)_{12}$ and solvent at t = 2 h. A total of 110 raw spectra were collected. The changes in the intensities of the absorption bands in the wavenumber range of 1550−2300 cm^{-1} can be observed in Figure 9.

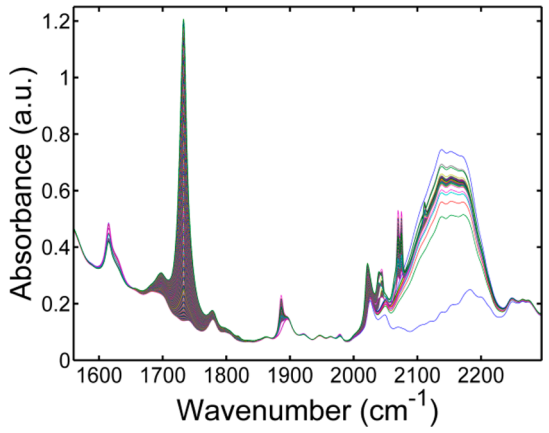


Figure 9. Raw in situ spectra from the experimental rhodium-catalyzed hydroformylation of cyclopentene.

The BTEM algorithm was used to deconvolute the spectral set in Figure 9 to obtain the set of normalized (uncalibrated) pure components underlying the raw spectra. In the current experiment, five normalized pure component spectral estimates were obtained (excluding the solvent and dissolved CO),as shown in Figure 10. At this point, the identity of the chemical species associated with the pure component spectra is immaterial for the following 2D correlation analysis. The 2D correlation analysis presented in this contribution is deliberately constructed to avoid introducing a priori chemical knowledge on speciation into the analysis of the system.

At the same time, the relative concentration profiles for the respective components (indices a−e) are obtained from the BTEM algorithm. These relative uncalibrated concentration profiles are shown in Figure 11. In this figure, the perturbation is clearly seen at t = ca. 2 h. Most importantly, we see that the concentration of the organic reagent e starts at a more or less maximum value and decreases only a few percent, in a regular monotonic fashion, over the entire experimental period. Therefore, we can conclude that (i) e is most likely our

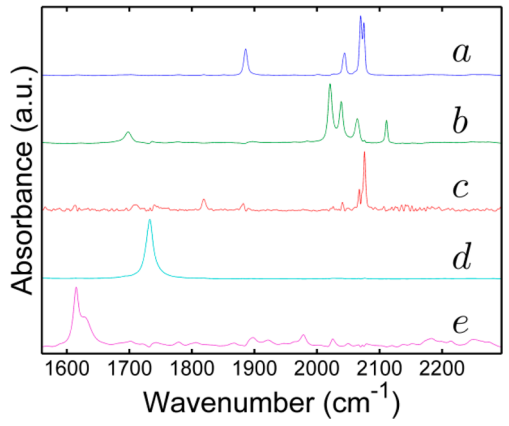


Figure 10. Five individual BTEM spectral estimates a−e obtained from the raw data. No assignments have been made at this point. However, one can note (i) the narrow spectral features of a−c, which suggest that they may be organometallics, and (ii) the broad spectral features of d and e, which suggest that they may be organics.

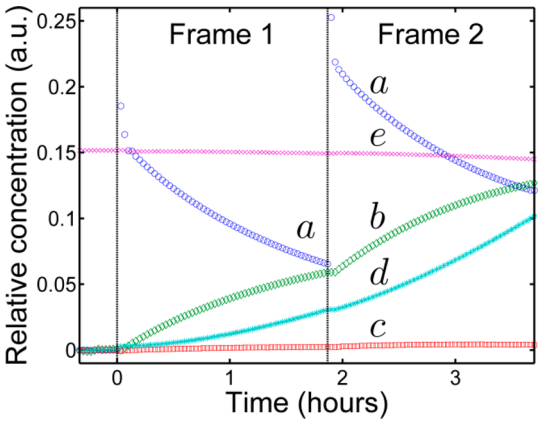


Figure 11. Time-dependent relative concentration profiles of the five components in the experimental unmodified rhodium hydroformylation. The time-dependent curves a−e correspond with the individual BTEM spectra a−e shown in Figure 10.

substrate and that (ii) pseudo-first-order kinetics are probably valid for the system, since so little conversion has occurred.

Due to mixing issues in the recycle system (section 3.4) the first few data points (about three data points corresponding to ca. 6 min) in frames 1 and 2 were discarded from further analysis. Therefore, there is a very small discontinuity in time for the data points between different frames.

Scatterplots were generated from the relative concentrations data according to section 3.5. The collection of all scatterplots from both reaction frames were assembled into a scatterplot matrix, as shown in Figure 12. From a visual inspection of the scatterplot matrix alone, one entry appears to clearly possess linearity (in yellow) and three entries might possess linearity (in gray). All other scatterplots clearly do not fulfill criteria I and II, as discussed in detail in section 3.5. Both hypothesis tests were applied to all frames in scatterplots. The final 2D correlation matrix is shown in Figure 13.

It is seen from Figure 13 that only one entry satisfies both the Sign test and the KPSS test, and this entry is highlighted in yellow. This entry is associated with the concentration $\hat{C}_b$ versus the rate $|\dot{\hat{C}}_d|$. Therefore, solely on the basis of the scatterplots, and without recourse to chemical assignments, it

DOI: 10.1021/cs502127y
ACS Catal. 2015, 5, 3588−3599

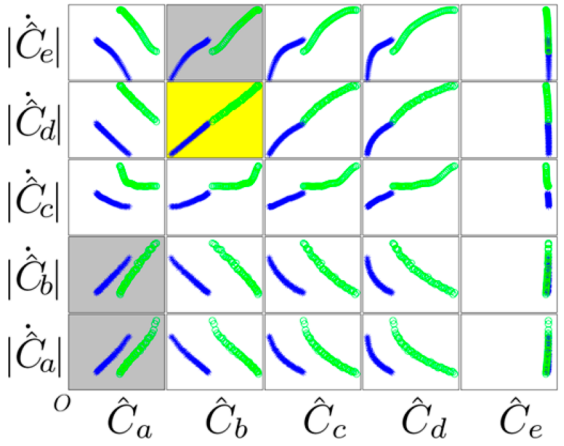


Figure 12. Scatterplot matrix of the real data set. Each entry of the matrix gives a visual display of the possible relationship between rate $|\dot{\hat{C}}_y|$ and concentration $\hat{C}_x$. Blue data points belong to frame 1, and green data points belong to frame 2. The yellow entry clearly possesses linearity, and the three gray entries might possess linearity. The x axis domain for each entry has the range of [0, max $\{\hat{C}_x\}$], and the y axis domain for each entry has the range of [0, max $\{|\dot{\hat{C}}_y|\}$].

	$\hat{C}_a$	$\hat{C}_b$	$\hat{C}_c$	$\hat{C}_d$	$\hat{C}_e$
$\|\dot{\hat{C}}_e\|$	0	0	0	0	0
$\|\dot{\hat{C}}_d\|$	0	1	0	0	0
$\|\dot{\hat{C}}_c\|$	0	0	0	0	0
$\|\dot{\hat{C}}_b\|$	0	0	0	0	0
$\|\dot{\hat{C}}_a\|$	0	0	0	0	0

Figure 13. Resulting 2D correlation matrix of the real experimental hydroformylation system after application of statistical tests. Each entry of the matrix expresses a yes−no answer to the statistical tests and hence the linearity of rate $|\dot{\hat{C}}_y|$ versus concentration $\hat{C}_x$. The value 1 will be assigned to positive results for both the Sign test and KPSS test, while the value 0 represents otherwise.

can be concluded that there is only one observable intermediate in the system.

At this point, for the purpose of the present contribution, it can be revealed that the index b corresponds to $RCORh(CO)_4$ and the index d corresponds to the organic product cyclopentanecarboxaldehyde. This is indeed the correct overall and expected conclusion. Only the species $RCORh(CO)_4$ is an intermediate in the overall catalytic transformation. Neither index a ($Rh_4(CO)_{12}$) nor index c ($Rh_6(CO)_{16}$) is an intermediate in the unicyclic catalysis. Finally, there is no positive cross correlation between the indices d and e (aldehyde and alkene), since these two species are separated by an entire cycle of intermediates, and moreover, there is an intrinsic delay time of ca. 10 min along the cycle (TOF ≈ 0.1 min^{-1}).

5. DISCUSSION

5.1. Numerical Simulation. The most important issue here is the details of the model used. It was assumed that there are only three reactions, and these were enough to set up a model system for the numerical simulation. In other words, a detailed catalytic system, with multiple intermediates, was not modeled. Moreover, the finite delay time in a catalytic cycle was deliberately not captured. The reasons for this decision are straightforward. We simply wanted to verify the validity of the statistical tests and the 2D correlation as it relates to the identification of first-order kinetics. In the same manner, the introduction of a "presumed" first-order process for the formation of $Rh_6(CO)_{16}$ from intermediates was based on the same reasoning. When the model system is kept extremely simple, it is much easier to understand the meaning of the final 2D correlation.

The resulting 2D correlation matrix possessed five yellow entries (Figure 7). These entries relate (1) the disappearance of the precursor $Rh_4(CO)_{12}$ and the formation of the intermediate $RCORh(CO)_4$, (2) the loss of alkene and the formation of aldehyde due to the intermediate $RCORh(CO)_4$, and finally (3) the formation of $Rh_6(CO)_{16}$ from the observable $RCORh(CO)_4$. The first pair of reactions in (1) are easily understood by inspection of Scheme 2a and the first-order kinetic expressions given by eqs 6 and 7. The second pair of reactions in (2) is easily understood in terms of Scheme 2b and the first-order kinetic expressions given by eqs 8 and 9. Finally, the reaction in (3) is understood in terms of Scheme 2c and the first-order kinetic expression given by eq 10—which was deliberately and artificially introduced. Taken together, the above 2D correlation results are consistent with the model used. All first-order reactions provide positive results for the Sign test, as well as the KPSS test.

5.2. Experimental Test. The experimental system introduces an entire set of new physical phenomena, and these affect the kinetics and hence the statistical tests. The significant issues to consider in the experimental system are (i) the fact that the system has a finite delay time in the cycle associated with the turnover frequency, (ii) solvent is actually introduced with the precursor during the perturbation (this changes the mass balance in the system, and the concentrations in a significant way), and (iii) the organometallics are labile and there are equilibria present. These issues will influence the final way that the experimental scatterplots look as well as the results of the statistical tests.

Figure 14 is an expanded view of the gray and yellow entries found previously in Figure 12, which help to clarify further many of these aforementioned remarks.

The entry $[\hat{C}_b, |\dot{\hat{C}}_e|]$ in Figure 14a contains some subtle but noticeable peculiarities. Close inspection shows that (i) there is significant curvature at the end of frame 1 and the beginning of frame 2 and (ii) upon dilution of the system between frames 1 and 2, a discontinuity in the curves occurs. Both of these observations can be traced back to the fact that alkene (species e) is involved in two overall reactions: namely, transformation of precursor to $RCORh(CO)_4$ and catalysis. Therefore, if alkene is involved in two overall reactions, a plot of rate versus just $RCORh(CO)_4$ must show some curvatures. In contrast, frames 1 and 2 in $[\hat{C}_b, |\dot{\hat{C}}_d|]$ are both very straight. Thus, entry $[\hat{C}_b, |\dot{\hat{C}}_e|]$ almost passes the statistical tests, but entry $[\hat{C}_b, |\dot{\hat{C}}_d|]$ certainly passes the statistical tests.

DOI: 10.1021/cs502127y
ACS Catal. 2015, 5, 3588−3599

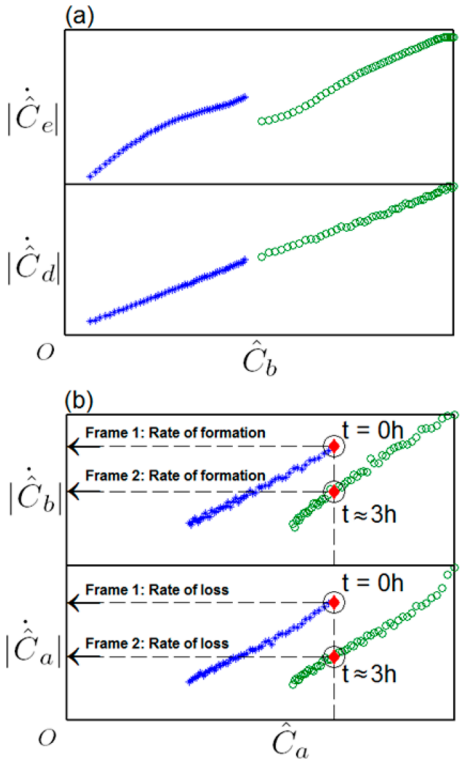


Figure 14. Expanded views of gray and yellow entries from Figure 12: (a) entries $[\hat{C}_b,|\dot{\hat{C}}_d|]$ and $[\hat{C}_b,|\dot{\hat{C}}_e|]$; (b) entries $[\hat{C}_a,|\dot{\hat{C}}_a|]$ and $[\hat{C}_a,|\dot{\hat{C}}_b|]$. Blue symbols are given for frame 1, and green symbols are given for frame 2. In (b), the red filled diamonds are data points where $\hat{C}_a$ values are the same. The corresponding values of $|\dot{\hat{C}}_a|$ and $|\dot{\hat{C}}_b|$ in frames 1 and 2 are obtained for a given $\hat{C}_a$ value, as shown in the directed dashed line. See text for explanation.

The issues of lability and reversibility of the organometallic transformations is seen by the "offset" in curves for entries $[\hat{C}_a,|\dot{\hat{C}}_a|]$ and $[\hat{C}_a,|\dot{\hat{C}}_b|]$, as shown in Figure 14b. Each frame alone gives a good linear plot for the precursor $Rh_4(CO)_{12}$, but upon combination of both frames, there is a distinct mismatch. As observed in entry $[\hat{C}_a,|\dot{\hat{C}}_a|]$, there was a decrease in the rate of loss of species *a* ($Rh_4(CO)_{12}$) from frame 1 to frame 2 for a given $\hat{C}_a$. At the beginning of frame 1, when $Rh_4(CO)_{12}$ is introduced, there is no $RCORh(CO)_4$ present. However, at the beginning of frame 2, when $Rh_4(CO)_{12}$ is introduced, there is quite a lot of $RCORh(CO)_4$ already present. Therefore, for frame 2, the rate of loss of $Rh_4(CO)_{12}$ is markedly lower. Similar observation was noticed in entry $[\hat{C}_a,|\dot{\hat{C}}_b|]$: a decrease in the rate of formation of species *b* ($RCORh(CO)_4$) from frame 1 to frame 2 for a given $\hat{C}_a$ value. Thus, the entries $[\hat{C}_a,|\dot{\hat{C}}_a|]$ and $[\hat{C}_a,|\dot{\hat{C}}_b|]$ almost pass the linearity tests (gray shading) but not entirely.

This demonstrates that the analysis of multiple frames seems to be a particularly useful tool in the 2D correlation in order to reject potential positive tests. The final outcome is that one and only one organometallic correlates experimentally with product formation, and that is $RCORh(CO)_4$.

5.3. Model Reduction. In many or perhaps most exploratory studies of new metal-mediated syntheses, detailed spectral analysis will results in the detection of numerous organometallics (far more than in the present study). In the general case, some of these organometallics will be intermediates in the synthesis and others will be spectator species/side products/degradation products. It is quite possible that the identity of all of these organometallics will prove too difficult or impossible to determine—even if enormous efforts using density functional theory (DFT) are performed. The explanation is simple enough; some of the spectator species/side products/degradation products will be structurally very complex large clusters and there are just too many variations. Moreover, and probably of primary importance, some of the spectator species/side products/degradation products will be the result of reactions with unknown impurities.

The present 2D correlation analysis helps to focus attention on the most important spectroscopic information: namely, the spectra associated with the intermediates directly involved in the overall organic synthesis. It is these spectra which are most important to consider in DFT calculations, and it is these intermediates which are most important from a stoichiometric and structural viewpoint. In other words, the present 2D correlation analysis provides a rational basis for model reduction. If the experimentalist can deduce the stoichiometry and the structure of the intermediates, a large part of the central mechanistic issues is clarified. If the experimentalist can also determine the identity of most spectator species/side products/degradation products, that is of course a bonus. However, if some of the latter determinations are not possible, one still has a rather clear picture of the salient features of the catalytic system.

5.4. Postanalysis Data Reconciliation and Criteria III and IV. For a well-designed experimental data set, and at least in the present case, criteria I and II appear necessary and sufficient to ensure a proper result.

Close inspection of the King and Altman formalism shows that there are actually two additional criteria which are implicit in the analytical expressions. Continuing from eq 4, we arrive at the following.

Criterion III: as implied by eq 4, $\dot{C}_p = k_i C_i$ and hence an intermediate i is related to the product p where $i \neq p$.[40] Therefore, all diagonal entries $[C_x,|\dot{C}_x|]$ or $[\hat{C}_x,|\dot{\hat{C}}_x|]$ (for any species x) can be confidently set identical with 0 in the 2D correlation matrix for the identification of intermediates.

Criterion IV: as implied by eq 4, $\dot{C}_p = k_i C_i$ and hence the rate constant k is implicitly positive ($k > 0$) for the case relating intermediate i to product p. Therefore, for all constants k found to be nonpositive ($k \leq 0$), the respective entry $[C_x,|\dot{C}_y|]$ or $[\hat{C}_x,|\dot{\hat{C}}_y|]$ (for any pair of species x and y) may be confidently set to 0 in the 2D correlation matrix for the identification of intermediates.

In the context of the real catalytic system and the 2D correlation matrices/entries of Figures 12 and 14, we see that application of criterion III applies to $[\hat{C}_a,|\dot{\hat{C}}_a|]$. Thus, this entry may be confidently set to 0. Presently, it appears that Criteria I and II are sufficiently strong criteria to evaluate well-designed experimental data sets with good signal to noise ratios. Accordingly, it appears that criteria III and IV are, at the moment, not needed for real data unless difficulties with the quality of the data is encountered: i.e., insufficient signal to noise ratios.

5.5. Consequence and Future Studies. The present 2D correlation analysis was carried out under an explicit assumption that there is one isolated catalytic cycle. A soluble and isolated catalytic cycle will have one and only one route for

DOI: 10.1021/cs502127y
ACS Catal. 2015, 5, 3588−3599

product formation, and there will be only one product. Clearly, this is a specific case for a much broader and more general situation. Many catalytic systems possess chemoselectivity, regioselectivity, or stereoselectivity, and therefore, they have by definition more than one product and more than one cycle of intermediates leading to product formation. Such systems will have coupled catalytic cycles, where some but not all intermediates are common to both cycles. This situation requires much more consideration with respect to setting up a properly posed 2D correlation analysis.

More difficult situations could also arise, at least in principle. Indeed, there are systems that possess more than one completely independent catalytic cycle leading to product formation (think of two metals that can perform a hydrogenation and add them to the same synthesis). Another complication arises when a catalytic system possesses intermediates of more than one nuclearity in the same complex reaction network. A number of synergistic systems have been shown to function in this manner.[41]

Moreover, and perhaps the most difficult situation to consider, is a system which has additional reaction pathways above and beyond the coupled systems mentioned above. These more complex systems might have contributions to product formation due to (1) uncatalyzed thermally induced product formation, (2) surface-catalyzed reactions at the wall (perhaps the reactor is not as clean as originally thought), and (3) surface-catalyzed product formation due to colloids generated during the reaction. These types of situations, at least at the moment, appear to be considerably more difficult to treat using the approach we have introduced. The main reason is that it is difficult to quantify some of the time-dependent quantities involved in points (1)−(3) above.

Efforts are underway to generalize the formulation of the present 2D correlation analysis in order to address as many of the aforementioned scenarios as possible.

6. CONCLUSION

In the present contribution, a strictly mathematical 2D correlation analysis has been developed in order to identify the intermediates participating in a unicyclic catalytic cycle. The analysis has been successfully tested on a numerical model and then on a set of experimental data for a rhodium-catalyzed hydroformylation of alkene. Without any prior chemical knowledge of the reaction, but with proper experimental design, the present approach allows an experimentalist to focus on the most meaningful spectroscopic information at the exploratory stage of a research program. Efforts are underway in our laboratory to extend the concept of 2D correlation to more complex catalytic systems.

AUTHOR INFORMATION

Corresponding Author

*E-mail for M.G.: marc_garland@ices.a-star.edu.sg.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Support from the Agency for Science, Technology and Research (A*STAR) of Singapore is gratefully acknowledged.

REFERENCES

(1) *Mechanisms in Homogeneous Catalysis: A Spectroscopic Approach*; Heaton, B., Ed.; Wiley-VCH: Weinheim, Germany, 2005.
(2) Haynes, A. The Use of High Pressure Infrared Spectroscopy to Study Catalytic Mechanisms. In *Mechanisms in Homogeneous Catalysis. A Spectroscopic Approach*; Heaton, B., Ed.; Wiley-VCH: Weinheim, Germany, 2005; pp 107−150.
(3) Viviente, M.; Pregosin, P.; Schott, D. NMR Spectroscopy and Homogeneous Catalysis. In *Mechanisms in Homogeneous Catalysis: A Spectroscopic Approach*; Heaton, B., Ed.; Wiley-VCH: Weinheim, Germany, 2005; pp 1−80.
(4) Tinker, H.; Morris, D. *Rev. Sci. Instrum.* **1972**, *43*, 1024−1028.
(5) Whyman, R. *J. Chem. Soc. D* **1970**, *23*, 1194−1195.
(6) Bor, G.; Dietler, U.; Noack, K. *J. Chem. Soc., Chem. Commun.* **1976**, *22*, 914−916.
(7) Mirbach, M. *J. Organomet. Chem.* **1984**, *265*, 205−213.
(8) Alemdaroğlu, N.; Penninger, M.; Oltay, E. *Monatsh. Chem.* **1976**, *107*, 1153−1165.
(9) Duckett, S.; Newell, C.; Eisenberg, R. *J. Am. Chem. Soc.* **1994**, *116*, 10548−10556.
(10) Brown, J.; Kent, A. *J. Chem. Soc., Chem. Commun.* **1982**, 723−725.
(11) Garland, M.; Bor, G. *Inorg. Chem.* **1989**, *28*, 410−413.
(12) Haynes, A.; Maitlis, P.; Morris, G.; Sunley, G.; Adams, H.; Badger, P.; Bowers, C.; Cook, D.; Elliott, P.; Ghaffar, T.; Green, H.; Griffin, T.; Payne, M.; Pearson, J.; Taylor, J.; Vickers, P.; Watt, R. *J. Am. Chem. Soc.* **2004**, *126*, 2847−2861.
(13) Heaton, B.; Jacob, C.; Monks, G.; Hursthouse, M.; Ghatak, I.; Somerville, R.; Heggie, W.; Page, P.; Villax, I. *J. Chem. Soc., Dalton Trans.* **1996**, *1*, 61−67.
(14) van Rooy, A.; Orij, E.; Kamer, P.; van Leeuwen, P. *Organometallics* **1995**, *14*, 34−43.
(15) Li, C.; Guo, L.; Garland, M. *Organometallics* **2004**, *23*, 2201−2204.
(16) More recently, important in situ FTIR contributions have been made by a number of groups. Particularly noteworthy is the large contribution from Rostock, Germany (LIKAT): (a) Christiansen, A.; Li, C.; Garland, M.; Selent, D.; Ludwig, R.; Franke, R.; Börner, A. *ChemCatChem* **2010**, *2*, 1278−1285. (b) Christiansen, A.; Li, C.; Garland, M.; Selent, D.; Ludwig, R.; Spannenberg, A.; Baumann, W.; Franke, R.; Börner, A. *Eur. J. Org. Chem.* **2010**, *14*, 2733−2741. (c) Reference 20.
(17) For the purpose of the present contribution, we will exclude consideration of at-line mass spectrometry work, although it has provided a wealth of new information on homogeneous catalysis: (a) Chen, P. *Angew. Chem., Int. Ed.* **2003**, *42*, 2832−2847. (b) Henderson, W.; McIndoe, J. C. *Mass Spectrometry of Inorganic and Organometallic Compounds: Tools-Techniques-Tips*; Wiley-Interscience: New York, 2005.
(18) Garland, M., Processing Spectroscopic Data. In *Mechanisms in Homogeneous Catalysis: A Spectroscopic Approach*; Heaton, B., Ed.; Wiley-VCH: Weinheim, Germany, 2005; pp 151−193.
(19) Li, C.; Chew, W.; Garland, M. *Appl. Spectrosc.* **2003**, *57*, 491−498.
(20) (a) Kubis, C.; Ludwig, R.; Sawall, M.; Neymeyr, K.; Börner, A.; Wiese, K. D.; Hess, D.; Franke, R.; Selent, D. *ChemCatChem* **2010**, *2*, 287−295. (b) Kubis, C.; Selent, D.; Sawall, M.; Ludwig, R.; Neymeyr, K.; Baumann, W.; Franke, R.; Börner, A. *Chem. - Eur. J.* **2012**, *18*, 8780−8794.
(21) (a) Chew, W.; Widjaja, E.; Garland, M. *Organometallics* **2002**, *21*, 1982−1990. (b) Widjaja, E.; Li, C.; Garland, M. *Organometallics* **2002**, *21*, 1991−1997. (c) Widjaja, E.; Li, C.; Chew, W.; Garland, M. *Anal. Chem.* **2003**, *75*, 4499−4507.
(22) Widjaja, E.; Li, C.; Garland, M. *J. Catal.* **2004**, *223*, 278−289.
(23) Garland, M.; Li, C.; Guo, L. *ACS Catal.* **2012**, *2*, 2327−2334.
(24) In order to avoid unnecessary confusion, we note that the term "2D correlation analysis" arises in quite a few areas of the mathematical and physical sciences. In the area of chemistry and spectroscopy, one of the most well-known forms is the Noda transform, which arises in

DOI: 10.1021/cs502127y
ACS Catal. 2015, 5, 3588−3599

the application of the Hilbert transform to raw spectroscopic data sets: Noda, I.; Ozaki, Y. *Two-Dimensional Correlation Spectroscopy-Applications in Vibrational and Optical Spectroscopy*; Wiley: New York, 2005. Furthermore, BTEM has been used to untangle spectra prior to further application of the Noda transform: Widjaja, E.; Tan, B. H.; Garland, M. *Appl. Spectrosc.* **2006**, *60*, 294−303. The type of 2D correlation analysis introduced in the present contribution does not use any of the mathematical constructs of the Noda Transform, the Hilbert Transform, nor synchronous/asynchronous maps in any way.

(25) van Leeuwen, P., Claver, C. *Rhodium Catalyzed Hydroformylation*; Springer, Heidelberg, Germany, 2002.

(26) (a) Garland, M.; Pino, P. *Organometallics* **1991**, *10*, 1693−1704. (b) Feng, J.; Garland, M. *Organometallics* **1999**, *18*, 417−427. (c) Fyhr, C.; Garland, M. *Organometallics* **1993**, *12*, 1753−1764.

(27) Gao, F.; Ng, K.; Li, C.; Krummel, K.; Allian, A.; Garland, M. *J. Catal.* **2006**, *237*, 49−57.

(28) Garland, M. Transport Effects in Homogeneous Catalysis. In *Encyclopedia of Catalysis*; Horvath, I., Ed.; Wiley: New York, 2002; pp 550−577.

(29) In the present case, we have chosen to calibrate the solvent. However, it should be noted that there are two choices for 2D correlation: namely, (1) no calibration of the solvent and (2) calibration of the solvent to obtain an absorptivity. If (1) is chosen, then any arbitrary nonzero absorptivity can be assigned: for example, setting the absorptivity to unity.

(30) Chew, W.; Widjaja, E.; Garland, M. *Appl. Spectrosc.* **2007**, *61*, 734−746.

(31) Liu, G.; Li, C.; Guo, L.; Garland, M. *J. Catal.* **2006**, *237*, 67−78.

(32) Liu, G.; Hakimifard, M.; Garland, M. *J. Mol. Catal. A: Chem.* **2001**, *168*, 33−37.

(33) (a) Masters, C. *Homogeneous Transition-metal Catalysis: A Gentle Art*; Springer: Heidelberg, Germany, 2013. (b) Temkin, O.; Zeiggarnik, A.; Bonchev, D. *Chemical Reaction Networks: A Graph−Theoretical Approach*; CRC Press: Boca Raton, FL, 1996.

(34) King, E.; Altman, C. *J. Phys. Chem.* **1956**, *60*, 1375−1378.

(35) Chou, K. *J. Biol. Chem.* **1989**, *264*, 12074−12079.

(36) (a) Zagajewski, M.; Behr, A.; Sasse, P.; Wittman, J. *Chem. Eng. Sci.* **2014**, *115*, 88−94. (b) Zagajewski, M.; Dreimann, J.; Behr, A. *Chem. Ing. Tech.* **2014**, *86*, 449−457. (c) Behr, A.; Witte, H.; Kämper, A.; Haßelberg, J.; Nickel, M. *Chem. Ing. Tech.* **2014**, *86*, 458−466. (d) Roda−Monsalvez, N.; Tran, D.; Battilocchio, C.; Labes, R.; Ingham, R.; Hawkins, M.; Ley, S. *Org. Biomol. Chem.* **2015**, *13*, 2550−2554. (e) Brzozowski, M.; O'Brien, M.; Ley, S.; Polyzos, A. *Acc. Chem. Res.* **2015**, *48*, 349−362.

(37) Montgomery, D.; Runger, G. *Applied Statistics and Probability for Engineers*, 2nd ed.; Wiley: New York, 1999.

(38) Kwiatkowski, D.; Phillips, P.; Schmidt, P.; Shin, Y. *Journal of Econometrics.* **1992**, *54*, 159−178.

(39) *MATLAB version 7.12.0 (R2011a)*; The MathWorks Inc., Natick, MA, 2011.

(40) Note that in simple catalytic systems, which are the focus of the present contribution, the possibility of autocatalysis will not be considered.

(41) (a) Li, C.; Widjaja, E.; Garland, M. *J. Am. Chem. Soc.* **2003**, *125*, 5540−5548. (b) Li, C.; Chen, L.; Garland, M. *J. Am. Chem. Soc.* **2007**, *129*, 13327−13334. (c) Li, C.; Cheng, S.; Tjahjono, M.; Schreyer, M.; Garland, M. *J. Am. Chem. Soc.* **2010**, *132*, 4589−4599.

DOI: 10.1021/cs502127y
ACS Catal. 2015, 5, 3588−3599