ACS Catalysis

Perspective

pubs.acs.org/acscatalysis

Toward Benchmarking in Catalysis Science: Best Practices, Challenges, and Opportunities

Thomas Bligaard,[†] R. Morris Bullock,[‡] Charles T. Campbell,[§] Jingguang G. Chen,[*,∥,⊥] Bruce C. Gates,[#] Raymond J. Gorte,[g] Christopher W. Jones,[h] William D. Jones,[i] John R. Kitchin,[j] and Susannah L. Scott[*,k]

[†]SUNCAT - Center for Interface Science and Catalysis, SLAC National Accelerator Laboratory, Menlo Park, California 94025, United States

[‡]Center for Molecular Electrocatalysis, Pacific Northwest National Laboratory, Richland, Washington 99352, United States

[§]Department of Chemistry, University of Washington, Box 351700, Seattle, Washington 98195-1700, United States

[∥]Department of Chemical Engineering, Columbia University, New York, New York 10027, United States

[⊥]Chemistry Department, Brookhaven National Laboratory, Upton, New York 11973, United States

[#]Department of Chemical Engineering & Materials Science, University of California, Davis, California 95616, United States

[g]Department of Chemical & Biomolecular Engineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104, United States

[h]School of Chemical & Biomolecular Engineering, Georgia Institute of Technology, 311 Ferst Drive NW, Atlanta, Georgia 30332, United States

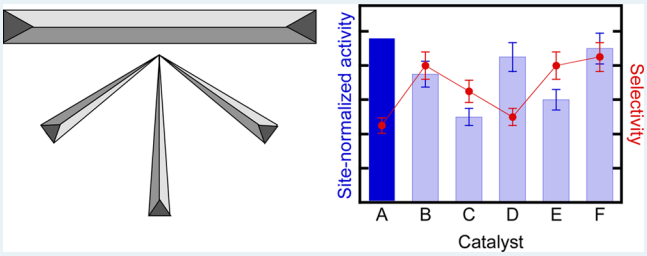
[i]Department of Chemistry, University of Rochester, Rochester, New York 14627, United States

[j]Department of Chemical Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213, United States

[k]Department of Chemical Engineering, University of California, Santa Barbara, California 93106, United States

Supporting Information

ABSTRACT: Benchmarking is a community-based and (preferably) community-driven activity involving consensus-based decisions on how to make reproducible, fair, and relevant assessments. In catalysis science, important catalyst performance metrics include activity, selectivity, and the deactivation profile, which enable comparisons between new and standard catalysts. Benchmarking also requires careful documentation, archiving, and sharing of methods and measurements, to ensure that the full value of research data can be realized. Beyond these goals, benchmarking presents unique opportunities to advance and accelerate understanding of complex reaction systems by combining and comparing experimental information from multiple, *in situ* and *operando* techniques with theoretical insights derived from calculations characterizing model systems. This Perspective describes the origins and uses of benchmarking and its applications in computational catalysis, heterogeneous catalysis, molecular catalysis, and electrocatalysis. It also discusses opportunities and challenges for future developments in these fields.



KEYWORDS: *benchmarking, catalytic performance, computational catalysis, heterogeneous catalysis, molecular catalysis, electrocatalysis*

1. INTRODUCTION

1.1. Place for Benchmarking in Catalysis Research.

Recent advances in the controlled synthesis of soluble and solid catalysts, their detailed characterization at the atomic and/or molecular scales under realistic reaction conditions, and sophisticated computational modeling of moderately large systems, are converging to create a new paradigm for catalysis research based on unprecedented insight into structure–activity/selectivity relationships at the level of the active site. At the same time, catalysis research has become a truly global activity, engaging many thousands of researchers worldwide.[1] This effort generates ever-larger amounts of information, which poses challenges for appropriate documentation, archiving, and data sharing. Concerted efforts toward benchmarking in and among subfields of catalysis (molecular, heterogeneous, photocatalysis, electrocatalysis, biocatalysis, computational modeling, etc.) will play a critical role in accelerating discovery, refining understanding, and promoting the application of better

Received: January 19, 2016
Revised: March 6, 2016
Published: March 7, 2016

DOI: 10.1021/acscatal.6b00183
ACS Catal. 2016, 6, 2590–2602

catalysts. The goal of this Perspective is to point out key issues and discuss guidelines for the judicious implementation of benchmarking in catalysis science.

1.2. Terms and Definitions. *Repurposing an Old Word.* It is helpful to begin from a mutually agreed definition: *to benchmark* is to evaluate by comparison with an established standard.[2] The verb stems from the name for a surveyor's mark, typically a horizontal slot above an arrow, permanently chiseled into a rock or wall to indicate its exact height above sea level (see Figure 1). Computer scientists were innovators in using

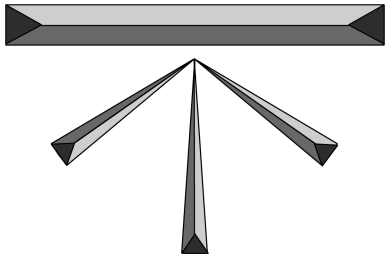


Figure 1. Representation of the original Ordnance Survey benchmark, as inscribed on many buildings and permanent geological features in Great Britain to indicate the height above sea level. When an angle-iron is placed in the horizontal slot, it creates a "bench" in which a leveling rod can be positioned reproducibly.

the term in a scientific sense, to compare hardware and software component performance. More recently, benchmarking has been appropriated to assess and drive improvements in education, business, technology, and science. In catalysis science, explicit benchmarking has been reported only infrequently in the literature (less than 500 mentions in ca. 1×10^6 articles describing catalytic phenomena in the past decade, according to a Chemical Abstracts Service keyword search), but it is increasing rapidly, as Figure 2 shows.

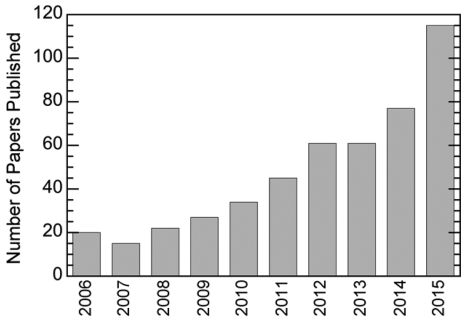


Figure 2. Frequency of research papers (journal articles, book sections, and conference proceedings) describing benchmarking of catalysis results, according to the Chemical Abstracts Service, November, 2015.

Modern Definition. Benchmarking can be defined as the translation into concrete terms of an abstract scientific paradigm;[3] furthermore, it requires a community effort to evaluate results using the scientific method.[4] The benchmarks themselves have both technical components (e.g., the recognition of central research questions) and social components (e.g., the rules for probing these questions). In other words, benchmarking requires communication and collaboration within a community to establish consensus about which questions are valid and how to evaluate their answers. Users of benchmarks must select problems and methods of measurement that are realistic, to avoid biasing their research toward results that are not broadly applicable.[5] A successful benchmark will be accessible, affordable, clear, relevant, solvable, portable, reproducible, and scalable.[3]

1.3. Potential Benefits. Benchmarking discussions in research are usually motivated by efforts to improve reproducibility and effectiveness, as well as the need for better documentation of both methods and results in scientific reporting.[6] The statisticians Buckheit and Donoho, paraphrasing Claerbout, famously wrote on this topic: "An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures."[7] According to Donoho, reproducible research confers the following advantages: (1) improved work and work habits; (2) better teamwork; (3) greater impact, resulting from less inadvertent competition and more acknowledgment; and (4) greater continuity and cumulative impact.[8] Ready accessibility of research data facilitates replication and makes old data available for new research purposes. As added benefits of more effective documentation, we anticipate better stewardship of and better access to public goods arising from taxpayer-funded research.

In addition to these useful observations, it is clear that benchmarking can offer advantages beyond performance comparisons, validation, and reproducibility. Notably, benchmarking can clarify the problems to be solved and lead to more rigorous assessment of research results, as well as better assessment methods. Thus, when a scientific discipline embraces benchmarking, it triggers more rapid scientific progress. Because evaluation by benchmarking ideally is transparent and replicable, it promotes research conduct that is collaborative, open, and ethical, thereby promoting community-building. Furthermore, there is an additional incentive for individual researchers: studies show a positive effect of data sharing on citation rates.[9]

From a global point of view and from the perspective of catalysis researchers, some of our most pressing concerns turn on the intertwined futures of energy science and technology and the health of our environment. Catalysis is central to a wide range of energy technologies that could play important roles in the storage, interconversion, and use of chemical and electrical energy. Catalytic control of the making and breaking of chemical bonds is a core scientific challenge in the effort to build a foundation for more efficient and sustainable energy technologies. Improved benchmarking for catalysts and catalytic processes of all kinds has the potential to boost the translation of fundamental catalysis science into the energy technology sector and others, including sustainable manufacturing.

1.4. Early Adopters. *Catalyst Discovery and Technological Applications.* A form of benchmarking that has played a central role in identifying new opportunities for catalyst discovery and optimization is the use, by industrial practitioners, of protocols for reproducible large-scale catalyst preparation, characterization (e.g., catalyst composition and surface area/pore volume properties), and performance under standard conditions, often measured in banks of nearly identical bench-scale reactors and in pilot-scale reactors. For decades, catalyst manufacturers have paid close attention to key (and often proprietary) criteria to ensure reproducibility in performance and thereby justify guarantees provided to catalyst users.

DOI: 10.1021/acscatal.6b00183
ACS Catal. 2016, 6, 2590−2602

Catalyst users are correspondingly constrained by agreements with manufacturers to use catalysts within specified operating ranges in order to maintain the guarantees. However, the criteria used, the protocols themselves, and the performance measurement results were generally not widely shared.

Experimental Assessment of New Catalysts. Heterogeneous catalysis researchers have long recognized the need to ensure reliability in measurements of catalytic properties and performance. The desire to standardize and ensure the reproducibility of catalyst performance data led to work in multiple European laboratories beginning in the 1980s to conduct rigorous comparisons of data via the use of ample batches of standard catalysts made available to numerous researchers. A material containing 5 wt % Pt supported on SiO_2 was produced by Johnson-Matthey and named EuroPt-1; 5 wt % Ni/SiO_2 was produced by EUROCAT and named EuroNi-1.[10] These early examples were followed by the naphtha reforming catalysts EuroPt-3 (0.3 wt % Pt/Al_2O_3) and EuroPt-4 (0.3 wt % Pt−0.3 wt % Re/Al_2O_3), as well as two EUROCAT oxides (V_2O_5−WO_3/TiO_2) and a zeolite, EuroTS-1.[11] In 2002, the World Gold Council commissioned the preparation of standard catalysts consisting of Au nanoparticles supported on TiO_2, Fe_2O_3, and C, allowing researchers in different laboratories to compare the performance of emerging supported gold catalysts with the standards.[12]

In the examples above, benchmark catalysts were not generally associated with benchmark reactions, and the experimental conditions under which testing should be conducted were not prescribed, presumably because these can vary widely depending on the goals of the study (comparison of activity, selectivity, stability, etc.). Further benchmarking specifications may be appropriate when the primary objective is comparison within a narrow class of catalytic materials (e.g., metal nanoparticles of different sizes, or on different solid supports). However, they become problematic for comparing disparate kinds of catalytic materials (e.g., molecular catalysts vs supported metal nanoparticles) because of their vastly different operating conditions, or for optimizing more than one performance metric.

Catalyst characterization protocols can also be benchmarked. For example, there are standard materials and methods for measuring surface areas and pore size distributions in porous solids;[13] users have a responsibility to be mindful of subtleties in data interpretation, for which experienced researchers have provided guidance.[14] In zeolite research, there are standard methods for assessing catalyst properties, including pore volume measurements by uptake of liquid hydrocarbons, determination of the degree of crystallinity by X-ray diffraction crystallography, and (approximate) evaluation of the content of amorphous material by NMR spectroscopy. For research on supported metals, methods for determining metal surface areas include H_2 chemisorption, H_2 titration of adsorbed oxygen, CO chemisorption, and EXAFS spectroscopy.[15]

In homogeneous catalysis research, well-defined molecular complexes that are readily synthesized and purified have been selected by community consensus for use as standards in benchmarking for a particular desired outcome (such as turnover frequency, turnover number, isolated yield, stereoselectivity, etc.), sometimes in combination with standard reactions. For example, zirconocene dichloride activated with methylaluminoxane has long been used as a benchmark catalyst for single-site ethylene polymerization catalysts,[16] while stereorigid *rac*-[ethylene(1-indenyl)$_2$]$ZrCl_2$ is used to benchmark stereospecificity in catalysts for isospecific propylene polymerization.[17] $Rh_2(O_3CCH_3)_4$ catalyzes the stereospecific cyclopropanation of styrene by ethyl diazoacetate and is used to benchmark enantiomeric excess for asymmetric catalysts.[18] For experimental benchmarking of oxidation catalysts, the rate of cyclooctene epoxidation with H_2O_2 is often used.[19] In the electrocatalysis community, Pt supported on high-surface-area carbon is widely considered to be the benchmark catalyst for overpotential in H_2 evolution and oxidation. The special challenges and complexity of reporting activities for both homogeneous[20] and heterogeneous[21] electrocatalysts have prompted recent efforts to articulate best practices and pitfalls for both measurement and reporting of data.[22]

Computational Approaches. Benchmarking involves comparing one level of theory against others, as well as comparing the results of theoretical calculations with experiments. Benchmarking has been used to assess many of the theoretical methods commonly applied to catalysis, most notably the popular and powerful density functional theory (DFT) methods. Examples include the assessment of dispersion corrections[23] and hydrogen-bonding interactions in DFT,[24] comparison of DFT with coupled-cluster calculations,[25] and the development of unbiased methods for benchmarking itself.[26] Recent examples of benchmarking applications include computational studies of homogeneous catalysts,[27] molecular electrocatalysts,[28] ligand-protected metal clusters,[29] and organic reactions.[30]

Benchmarking of DFT functionals is needed to accelerate progress toward the chemical accuracy needed in catalysis research, for example, in the computation of bond energies and barrier heights.[31] The availability of careful experimental measurements is critical for such comparisons. Benchmarked measurements of adsorption energies exist for some metal oxide clusters,[32] supported metal clusters,[33] and their adsorbates,[34] as well as for the barriers[35] of some elementary steps relevant to surface-catalyzed reactions. In molecular catalysis, benchmarking can be used to provide critical mechanistic insights. For example, in the homogeneous epoxidation of cyclohexene by H_2O_2 catalyzed by CH_3ReO_3, careful comparison of DFT-predicted equilibrium constants and rates with experimental results revealed the critical participation of the solvent.[36] This led to the discovery of a new reaction mechanism involving solvent-mediated proton transfer to generate (and regenerate) the active sites, thereby resolving discrepancies of 7−9 orders of magnitude or more between computed and experimentally measured rate constants.[37]

There are still major gaps in our ability to predict mechanisms *in silico*.[38] Nevertheless, computation is also starting to play a role in identifying new opportunities for catalyst discovery, and it is now of practical importance due to the possibility that computations can facilitate rapid, inexpensive screening of candidate catalysts.

1.5. Using Benchmarking to Advance Catalysis Science. Benchmarking should promote fair, relevant, quantitative comparisons of catalysts and, *ceteris paribus,* their performance. A community discussion about how to use benchmarking to accelerate progress and deepen understanding in catalysis science will also help move the field toward shared standards for how to compute, measure, report, and compare catalytic properties. Wider use of such standards should facilitate the reuse of published information in studies conducted by different research groups, both within each

DOI: 10.1021/acscatal.6b00183
ACS Catal. 2016, 6, 2590−2602

subfield of catalysis, and eventually between subfields. Three important questions will frame our discussion:

(1) Which practices will best enable benchmarking and data sharing between research groups, as well as among the different subfields of catalysis?
(2) How can benchmarking be used to identify new opportunities and to accelerate progress in catalysis science?
(3) How can the benchmarking tools themselves be assessed and improved?

Because reaction mechanisms are at the heart of our fundamental understanding of catalysis, it is a grand challenge to examine all the elementary steps of a reaction and to determine how the rate of each correlates with the structure of the catalyst. This is especially difficult when the latter is a solid presenting nonuniform surface sites. The most important task is identifying key intermediates and transition states with high "degrees of rate control".[39] Major progress will require the study of model catalysts to systematically simplify and control catalyst structure. This is sometimes possible in investigations of molecular and supported molecular catalysts, single crystal surfaces in ultrahigh vacuum, single-faceted nanomaterials, and some highly ordered porous crystalline materials, such as zeolites, metal−organic frameworks (MOFs), and related solids. Close collaborations between experimentalists and theorists will also be needed. Obviously, benchmarking requires practitioners to know and understand relevant literature precedents.

In this Perspective, we discuss benchmarking issues as they apply to four major catalysis subfields: computational catalysis, heterogeneous catalysis, molecular (and supported molecular) catalysis, and electrocatalysis. In each subfield, we suggest best practices for benchmarking, and we identify some of the challenges and opportunities for using benchmarking to advance catalysis science. Enzyme catalysis and photocatalysis are beyond the expertise of the authors and are not discussed, although we expect similar issues to be relevant in those subfields as well. Our suggestions and conclusions are the responsibility of the signing authors and are derived from free, open, and nonconsensus-seeking discussions among a subset of the investigators supported by the Catalysis Science program, Office of Science, Basic Energy Sciences, of the U.S. Department of Energy.

2. BEST PRACTICES AND OPPORTUNITIES FOR BENCHMARKING

2.1. Computational Catalysis. A grand challenge for theory and computation is to establish a coherent conceptual framework for predicting the functionality of catalytic molecules and materials. This challenge can be addressed by (1) establishing the catalysis-relevant computational results needed to accelerate the experimental search for new catalysts, in both known and novel catalytic reactions, and (2) more effectively sharing the results so that they can be used by others. As mentioned above, benchmarking involves comparing one level of theory against others, as well as comparing the results of theoretical calculations with experiments. To facilitate comparisons, some standardization is necessary. Thus, the field of theoretical modeling in catalysis would benefit greatly from a set of common conventions and perhaps even a uniform infrastructure for sharing results, methods, models, and codes. Such standardization in reporting results and conventions should aim to establish fast and exact reproducibility of scientific conclusions from computational studies, to the greatest extent practical.

2.1.1. Best Practices. Sharing Existing Results. The computational results that are available now could be shared much more effectively between theory groups but at the cost of some additional effort. It is clear, for example, that the structure of an adsorbed molecule does not provide nearly enough information to recreate an electronic structure calculation in order to obtain a quantitative adsorption energy. Furthermore, at present, the results reported in a paper or its Supporting Information are rarely sufficient to allow a detailed understanding of how the results in the paper were generated. Making the computational input and output files available online (in addition to the output results) would significantly enhance the reader's ability to reproduce the published results. *ACS Catalysis* now places hyperlinks to Supporting Information files near the end of each published paper, and a recently published *Viewpoint* pointed out how much more powerful this feature could be if such files contained editable information in multiple formats, rather than just conventional text.[40] As one example of the possibilities, Elsevier's *Article of the Future* project allows authors to include interactive graphs, executable code, and links to data repositories in their scientific articles.[41]

For quantities derived from electronic structure calculations, such as adsorption energies, vibrational frequencies, reaction barriers, entropies, and so on, authors should explain how these quantities were calculated from the original results at a level of detail that allows the calculations to be repeated. Sharing of spreadsheets or the codes in which values were calculated can help to achieve this goal. Since the output of electronic structure simulations is typically used as input for secondary complex models, such as microkinetic models, sharing of the input files, as well as the kinetics tool itself, would assist others significantly when they try to reproduce or improve on the original study.

Identifying Common Conventions and Tools. Establishing common tools for use in accessing and analyzing results could yield significant benefits. In identifying such tools, as well as common conventions for presenting results, it is important to remember that an approach best-suited to one researcher may not be relevant to another, even within the same research area. Various theory subcommunities will need to converge on their own best practices and use their own judgments to decide what information and tools are important enough to share and how to share them.

Establishing Benchmarking Methods. Sharing models and methods for use in benchmarking computational catalysis is an even greater challenge than the sharing of results. Since the codes that implement methods and models evolve, a key need in methodology sharing at this high level is a system for version control of implementations. Some enabling resources are provided in the Supporting Information.

2.1.2. Challenges and Opportunities. It is essential that, when possible, theoretical results for small and simple model systems be compared with results of more accurate theoretical methods or with experimental results from carefully synthesized and well-characterized materials, such as molecular and single-crystal catalysts. An example of high-level theoretical benchmarks and high-quality experimental measurements on atomically well-characterized transition metal surfaces is shown in Figure 3.[34b] In contrast, there is not much high-quality theoretical/experimental information available for the surfaces

DOI: 10.1021/acscatal.6b00183
ACS Catal. 2016, 6, 2590−2602

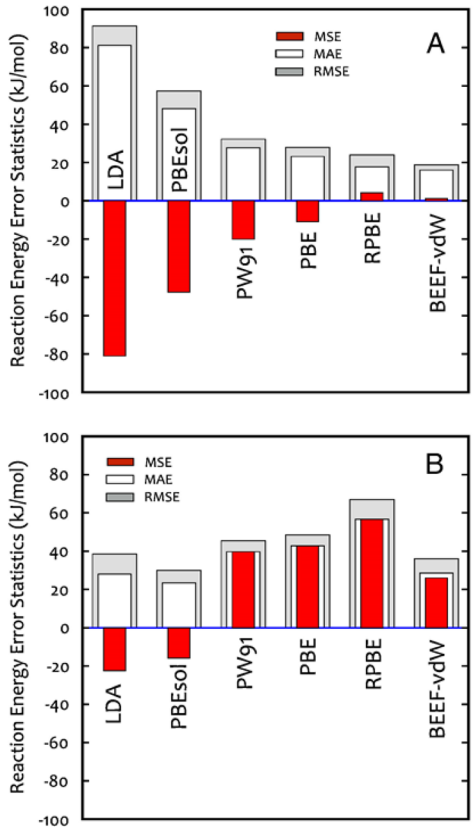


Figure 3. Comparison of the accuracy of density functionals relative to experimental adsorption energies for a range of small molecules on single-crystal late transition metal surfaces, for (A) purely chemisorbed systems (CO, NO, O, H, etc.) and (B) systems in which van der Waals interactions contribute substantially to the adsorption energy (NH_3, CH_3OH, CH_4, C_6H_6, D_2O, etc.). Reproduced with permission from ref 34b. Copyright 2015 Elsevier.

of transition-metal compounds (e.g., metal oxides, carbides, and nitrides), although these (especially the metal oxides) are very important in catalysis. This deficiency is primarily a reflection of the much larger system sizes of metal oxides, which make higher-level quantum chemical simulations prohibitively expensive at the moment, and which make it difficult to achieve experimental characterization with atomic-scale accuracy.

It remains difficult to use theoretical simulations in making accurate predictions for values characterizing essential catalytic properties (such as reaction rates or selectivities) with chemical accuracy (e.g., $\leq$ 1 kcal/mol for activation energies).[42] Molecular systems may be better defined and therefore easier to model than heterogeneous catalysts, as illustrated in a recent study of olefin hydrogenation catalyzed by ruthenium–xantphos complexes.[43] However, such systems usually involve reactions in solution, for which calculations can be complicated by difficulties in describing solvation,[44] as well as speciation and concentration effects. Theoretical studies are often much better at predicting or explaining trends that appear when the catalyst or the reaction conditions are varied. Consequently, experimental investigations that explore and report trends are key to enabling meaningful comparisons between theory and experiments in both homogeneous and heterogeneous catalysis.

A recent example involves theoretical calculations by Nørskov and co-workers, which show that for similar materials, exemplified by single crystal facets of late transition metals and their alloys, the energies of adsorbates as well as the energies of transition states in the catalytic reactions of small molecules scale linearly with a few simple descriptors, such as the adsorption energies of atomic O, C, and N.[45] This scaling behavior, illustrated in Figure 4, defines a set of benchmarks

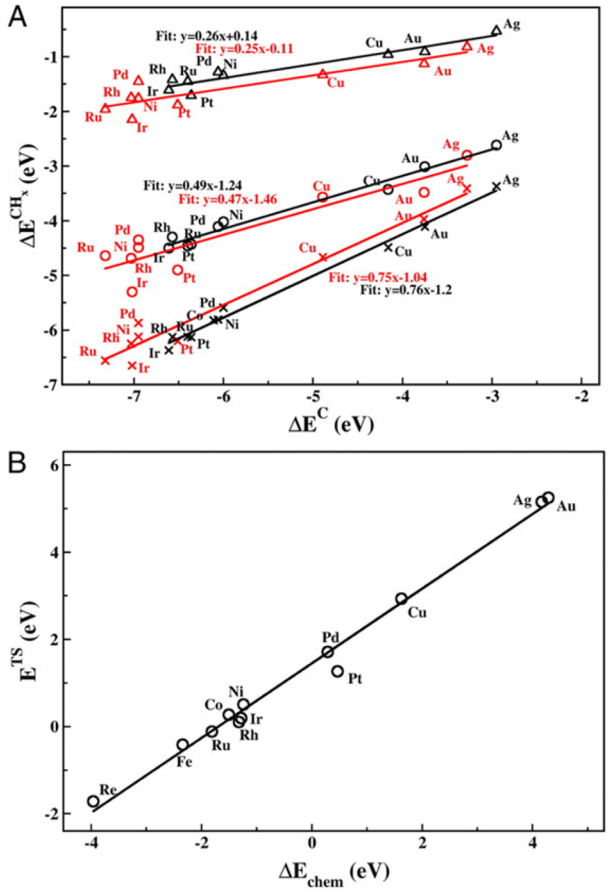


Figure 4. (A) Typical scaling relationships showing the linear correlation between calculated CH_x and C adsorption energies for various late transition metal surfaces (close-packed: black; stepped: red). (B) Linear relationship between the activation energy for CO dissociation and the energy for dissociative CO chemisorption on stepped metal surfaces. Reproduced with permission from ref 46. Copyright 2011 National Academy of Sciences.

that have greatly improved our understanding of structure–function relationships for catalysis by metals,[46] and has enabled the prediction of metal alloy compositions whose electronic structures are optimized for a particular reaction.[47] Moreover, it led to the important insight that the mere existence of linear-scaling relationships within a certain class of materials severely limits our ability to find better catalysts within that class. The activity of even the best material is inevitably constrained, as dictated by the scaling relationships.

The original scaling relationships can generate insights for nonincremental catalyst improvement, via assessment of the limitations that constrain the performance of even optimized systems, and can suggest how to design new types of catalysts to purposefully circumvent a particular linear-scaling limitation. They have inspired refinements, such as incorporating structure-sensitivity via the coordination number of the metal at the adsorption site.[48] For example, a system may involve adsorbates interacting with multiple parts of the catalyst, whose

DOI: 10.1021/acscatal.6b00183
ACS Catal. 2016, 6, 2590–2602

components have been designed to stabilize or destabilize other adsorbates or transition states. Such multisite interactions are inherent to many homogeneous and enzyme catalysts, and some solid catalysts already incorporate structural characteristics that facilitate multisite interactions with adsorbates (e.g., the narrow pores found in zeolites), setting them apart from the single crystals that have provided so much insight as model systems. In one example, Rodriguez and co-workers reported that the water−gas shift activity of gold decorated with CeO_x or TiO_x involves cooperativity between the metal and the oxide at interface sites.[49] Cargnello et al. correlated CO oxidation activity with TEM images of monodisperse ceria-supported metal nanoparticles to demonstrate that the reaction occurs at the metal−support interface, and predominantly at nanoparticle corner sites (Figure 5).[50] These systems provide incentives to

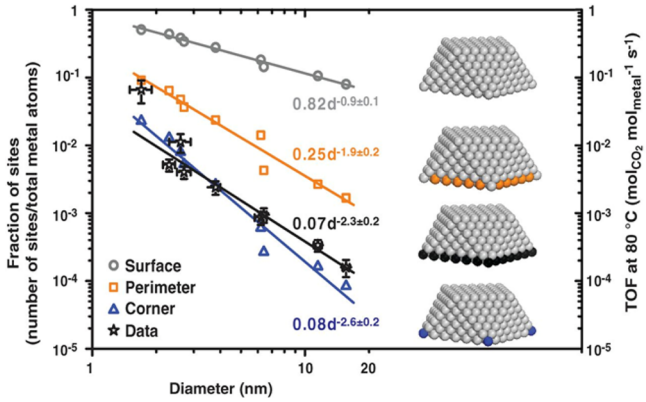


Figure 5. Calculated number of sites with a particular geometry and measured catalytic activity (TOF) for CO oxidation at 80 °C, both as a function of average particle diameter for group VIII metals. The best agreement between geometrical prediction and experiment involves the corner sites at the metal−support interface, suggesting that such sites dominate the catalytic activity. Reproduced with permission from ref 50. Copyright 2013 American Association for the Advancement of Science.

extend theoretical methods to incorporate greater chemical complexity within molecular-scale proximity of adsorbates. More importantly, for our purposes here, this type of thinking exemplifies how benchmarking can be used to create new opportunities in catalyst design and thus to accelerate progress in catalyst development.

2.2. Heterogeneous Catalysis. In experimental heterogeneous catalysis, benchmarking should involve a comparison of the performance of a new catalyst formulation against that of an accepted standard catalyst, under standard reaction conditions. However, practical implementation of this definition is complicated by two commonly encountered issues: (1) optimum reaction conditions for one catalyst may be markedly suboptimal for another; and (2) the nature and number of the active sites may evolve during the test, complicating normalization of parameters such as rates and yields. Nevertheless, our ability to assess the performance of heterogeneous catalysts is expanding rapidly as researchers apply powerful new techniques to determine the surface structures of functioning catalysts while at the same time measuring catalytic performance, even at elevated temperatures and pressures. Emerging techniques are capable of pinpointing the structures of surfaces in action and identifying catalytically active sites. IR, Raman, UV−vis, NMR, X-ray absorption, and other spectra recorded in *operando* mode are illuminating these sites as never before and also revealing new challenges in the integration of characterization techniques and in benchmarking. Coupling between experiment and theory in catalysis research is becoming ever closer. It is essential that all calculations, spectra, and images represent a catalyst in the same state, insofar as possible. There is a strong incentive to record atomic-resolution images of catalysts under reaction conditions as well, although capabilities for such measurements are still in an early stage of development.[51]

2.2.1. Best Practices. Accurate Reaction Kinetics. A key challenge for benchmarking in heterogeneous catalysis is to ensure that performance data truly represent intrinsic rates and selectivities. A set of "standard catalytic reactions", with accepted area- or site-normalized rates for well-characterized standard catalytic materials performing under well-documented conditions, should be used in comparing the performances of new and known catalysts. It is essential that measurements of turnover frequencies (TOFs, or site-normalized catalytic reaction rates), turnover numbers (TONs, or numbers of turns of a catalytic cycle, equivalent to site-normalized catalytic productivities), and selectivities be accurate, reproducible, uncorrupted by mass or heat transport limitations, and cover wide ranges of conversions. Benchmark protocols for accurate rate and TOF/TON measurements in the absence of mass, heat, transport, and mixing effects should be followed, as described in detail in the catalysis literature[52] and in general terms in textbooks.[53]

Determination of reaction rates is facilitated by the use of ideal, isothermal reactors, especially when the reaction occurs with no significant change in volume. Ideal reactors include perfectly mixed batch reactors, plug-flow reactors, and perfectly mixed tank reactors, known as CSTRs or backmix reactors. Initial rates can be determined at low conversions and require a demonstration of linear dependence of conversion on time in a batch reactor or on inverse-space velocity in a plug-flow reactor; in a backmix reactor, the rate can be determined directly at any nonequilibrium conversion. Because the TOF is, in general, concentration-dependent, reported TOF values should be based on rate measurements rather than on conversion at a single time point, and the values should be checked by systematically varying the concentrations of all species present in the reactor.

Standard Materials. Cooperation between research groups in a blind, round-robin series of measurements of reaction rates, site densities, and TOF values would be valuable to characterize simple benchmark catalytic materials and reactions, to identify subtleties in such measurements, and to determine realistic error bars. Such concerted efforts will require the participation of many established researchers, and there are historical precedents for implementing such strategies (as described in section 1.4 above).[10−12] Important questions remain about the availability of resources to conduct such work, as well as how to incentivize and recognize efforts that may be regarded as more routine than innovative.

Catalyst Deactivation. Rates of catalyst deactivation should be quantified where possible. The kinetics of deactivation are often complex, reflecting a variety of causes, such as sintering and loss of surface area, poisoning, coking, and/or formation of other deposits that cover catalytic sites, block pores, and/or block interstices in beds of catalyst particles. There is no simple figure of merit for deactivation of a catalyst. The industrial approach to systematic catalyst testing involves long-term operation under standard conditions in flow reactors to

DOI: 10.1021/acscatal.6b00183
ACS Catal. 2016, 6, 2590−2602

compare candidate catalysts with each other in terms of loss of activity and selectivity. Accelerated catalyst aging tests can be expeditious, but these tests are risky unless the causes of deactivation are known and extrapolations justified on the basis of experiments.

Benchmarking Adsorption Energies and Activation Energies for Elementary Reaction Steps. Accurate measurements of adsorption energies of surface catalytic reaction intermediates and activation energies of elementary reaction steps on surfaces are essential benchmarks for validating the accuracy of evolving computational approaches in fundamental catalysis (*vide supra*). Although surface science has facilitated many such measurements for transition-metal surfaces, a similar level of rigor has yet to be achieved for reactions on carbons and on metal oxides, sulfides, nitrides, phosphides, and carbides. Consequently, there is a need for basic surface science research on materials with high degrees of site homogeneity (e.g., single crystals, single-faceted nanomaterials, zeolites, MOFs). This type of work will provide experimental values to serve as benchmarks for validating computational methods.

2.2.2. Challenges and Opportunities. It is critical to establish benchmarking materials and protocols that enable us to extend such practices beyond early work on monometallic surfaces and supported nanoparticles, to include important practical catalysts with complex architectures. This includes challenging structure types such as multimetallic nanoparticles, mixed-oxides, and so on. It will require concerted efforts to create well-characterized systems and to make correct and detailed measurements of catalytic reaction rates and TOFs. The recent explosion in methods for creating tailored nanomaterials generates many exciting prospects for advanced benchmarks with more uniform active sites and control of their densities. Single-faceted nanoparticles, isolated transition metal atoms on metal oxide supports, and MOFs, for example, may facilitate tuning of specific surface sites for such investigations. Consequently, such materials are promising for advancing basic understanding, as well as for linking heterogeneous catalysts with molecular and supported molecular catalysts. Many tailored nanomaterials have high surface-to-volume ratios and therefore facilitate types of benchmarking measurements (and applications) that are infeasible with single crystals.

Caution must be exercised, however, because the dominant contribution to the catalytic activity may come from hard-to-identify minority sites such as defects. Even nominally highly uniform materials such as zeolites, MOFs, and metal and metal-oxide single crystals contain defects. Such sites are themselves worthy of research. Informative experiments may include titration with poisons, adsorption of spectroscopic probe molecules, observation by high-resolution TEM, and computational modeling, all of which may be employed to identify and count such sites in order to quantify their contributions to reactivity.[54]

Advances in the availability and accessibility of both *ex situ* and *in situ* characterization tools will continue to accelerate progress in benchmarking catalysis research. The structures of active sites on the surfaces of solid catalysts are key to understanding relationships between different catalyst formulations and therefore to guiding rational catalyst improvement. Better tools for *operando* structural characterization, or for *ex situ* structural analysis that is closely correlated with reaction rate measurements, will markedly improve our ability to identify structure−function relationships in catalysis with sufficient depth to enable meaningful catalyst comparisons. It is therefore important to continue to improve the tools, to make them more broadly accessible to the catalysis community, and to help new users become proficient.

Experimental work will increasingly benefit from computational analysis to interpret reactivity differences. Close feedback between theory and experiment is crucial for realizing the full benefits.[38] For example, experimental benchmarks will drive improvements in computational accuracy, which will in turn provide deeper understanding that suggests new experiments and more reliable predictions of better catalysts to be tested by experiment. Remarkable developments in the linking of computation and experiment have appeared in recent years, and coordinated benchmarking of theory and experiment promises further advances.

2.3. Catalysis by Molecular and Supported Molecular Species. Molecular catalysts are often viewed as being well-defined, relatively easy to characterize fully, and even single-sited, making the benchmarking of their catalytic performance under standard conditions more straightforward than for heterogeneous catalysts. However, researchers in this subfield recognize this view as naïve: issues of reproducibility occur frequently. Agreement among investigators is often elusive, even for catalytic properties that have long been considered well-established and clearly defined.[55] For example, a laboratory that tries to repeat a catalytic reaction described in the literature may find much less than the reported activity or none at all. Discrepancies may be caused by differences in reagents and protocols that are incompletely documented. Details of syntheses are crucial and often not reported in sufficient detail to allow them to be reproduced. These issues apply equally to hybrid materials created by immobilizing molecular catalysts on solid supports. Benchmarking is important to ensure more reliable and complete reports and more robust comparisons between catalysts in individual investigations, as well as between different laboratories.

2.3.1. Best Practices. Purity of Components. Most journals require that the purity of a catalyst or catalyst precursor be established with high-quality analytical data. Elemental analysis and NMR spectroscopy can be used for this purpose, for example, but they may be insufficient if they do not reflect the material actually used for catalysis (e.g., when a "purified" sample is used for analysis, but a crude material is used for catalysis). When possible, these samples should be demonstrably the same. This suggestion may be difficult to apply to catalysts generated *in situ* or to air/moisture-sensitive catalysts, for which chemical modification can occur prior to the catalytic reaction. In such cases, spectroscopic methods should be used to follow changes that occur as catalyst components are transformed. When catalysts are supported, it is essential to apply high standards of purity to the support (especially its surface) in addition to the species anchored to it.

Kinetic Measurements. Reaction kinetics provide critical information not only about how fast an overall catalytic reaction or a key step in a cycle proceeds but also where bottlenecks exist and how precatalysts are transformed into active catalysts. Whenever possible, a rate law should be established and appropriate rate parameters determined.[56] In batch reactor experiments, rates (and hence turnover frequencies) change continuously with time and are usually not a clear measure of catalytic activity. However, a TOF measurement may be possible if a large excess of one reactant results in a nearly constant reaction rate.[55b,57]

DOI: 10.1021/acscatal.6b00183
ACS Catal. 2016, 6, 2590−2602

Using a single, high-conversion measurement to report an average reaction rate is not generally useful, because it provides little insight into true rates. If this is the only kind of kinetic information available, "site-time yield" or a related term should be used instead of TOF. In batch reactions, initial rates should (at least) be obtained. Selectivity must be reported with the corresponding conversion; optimally, selectivity should be observed at both low and high conversions (the latter being of practical relevance). To avoid reporting nonintrinsic kinetics, the absence of mass/heat transfer and mixing effects must be verified. TON values should be demonstrably greater than 1 to ensure that the data correspond to a catalytic rather than a stoichiometric reaction and to turnover rather than the induction period during which the catalytic cycle is established. Mass balance should be verified, and if the measurement is made by NMR spectroscopy, an internal integration standard should be used.

Specification of Methodology. When reporting quantities such as the TON or TOF, they should be accompanied by a clear statement about how values were obtained from the raw data. This statement must include a description of how the number of active sites was counted or estimated, using some transparent and reproducible method. This point is especially important for immobilized molecular catalysts, because the immobilization procedure can create significant variability in site activity that may be difficult to detect spectroscopically.

Although the total number of sites of a particular type (e.g., a particular metal complex) may be quantified readily, it is desirable (and more challenging) to determine how many of these sites participate directly in the catalytic reaction. When the sites are titrated with a poison, it is often assumed that each poison molecule removes a certain number of sites from the catalytic cycle, but such assumptions must be checked. Results showing how the catalytic activity declines with the number of added poison molecules are not sufficient; it is also necessary to identify any assumptions made about the selectivity of bonding between the active sites and the poison and to compare that with the selectivity of reactants interacting with the same sites.

Selectivity. When the selectivities of several catalysts are compared, values should be measured at the same conversion. The overall mass balance (including carbon balance) should be verified. Efforts should be made to identify minor products; because these are often reaction-specific, they can provide clues about the reaction mechanism and sometimes about the causes of catalyst deactivation.

Catalyst Deactivation. Most catalysts lose their activity and/or selectivity over time, and measurements of catalyst lifetime are valuable. Analysis of the mechanisms of catalyst deactivation and characterization can be used to prepare more robust next-generation catalysts, as well as to indicate whether and how the catalyst can be reactivated. Catalyst regeneration also benefits from standard protocols and characterization methods. In particular, recycling studies in which product yield is measured solely at high conversion do not demonstrate catalyst stability.[58] Instead, it is preferable to compare initial rates after catalyst recycle in a batch reactor or to monitor steady-state conversion for an extended period in a continuous flow reactor.[59] The latter is particularly helpful for assessing catalyst leaching from a support, which may be obscured under batch conditions.

Comparisons and Controls. When reporting data for a new catalyst, it is important to compare the performance with well-established catalysts and/or state-of-the-art catalysts measured under the same experimental conditions. Naturally, not all quantities are relevant in every study, but the comparison should at least include those metrics identified as important goals by the authors (e.g., reactant scope, rate, yield, selectivity, reaction mechanism, deactivation behavior). An example is shown in Figure 6, where the total number of turnovers for H/D exchange in benzene was determined after 2 h for a series of Pt complexes. Site-time yields were measured early in the reaction, allowing a simple performance comparison of five different catalysts at four different temperatures, even though detailed kinetics were not determined.[60]

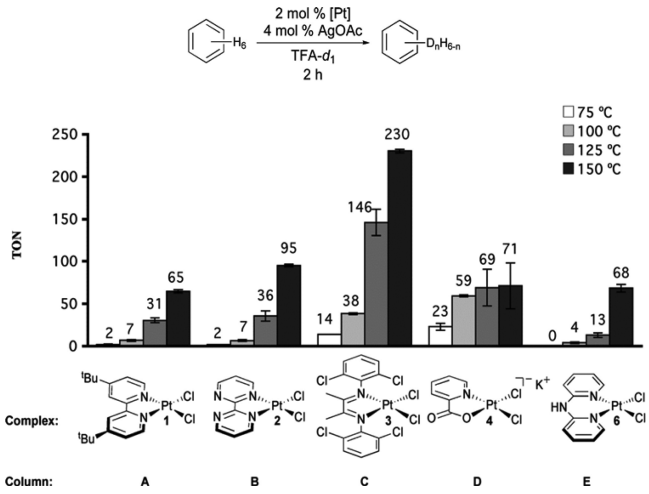


Figure 6. Productivity comparison of molecular Pt catalysts: turnover numbers (reported as TON ± standard deviation), achieved after 2 h for H/D exchange between C_6H_6 and CF_3CO_2D, as a function of temperature. Conditions: Pt^{II} catalyst (2 mol %, 5 μmol), benzene (23.2 μL, 0.26 mmol), AgOAc (1.7 mg, 10 μmol) in CF_3CO_2D (0.5 mL, 6.5 mmol, 25 equiv relative to benzene). Reprinted with permission from ref 60.

It is essential to perform appropriate control experiments without catalyst present to establish the significance of uncatalyzed reactions and to establish that the molecule/material claimed to be the catalyst is actually doing the work instead of being just a spectator to the reaction. Careful experimentation may require purification of reagents and testing of reagents from different suppliers to guard against catalytically active impurities, as well as repeat tests to demonstrate the absence of inadvertent reactions. Purification methods should be described fully. One should be aware of contamination when reactants and/or products contact catalytically active materials, such as $Fe(CO)_5$ formed in high-pressure CO cylinders, or adventitious metals deposited on reactor walls.

2.3.2. Challenges and Opportunities. Although straightforward ways to improve the quality of reports of molecular and supported molecular catalysts are described above, broad acceptance and adoption by the community has yet to be achieved and will require standardization of protocols. It is becoming feasible to monitor catalysts spectroscopically during catalysis, and specialized experimental capabilities (e.g., high-pressure/high-temperature NMR, IR, Raman, or XAFS spectroscopies) may be used, provided reactors that allow measurements of reaction rates simultaneously with catalyst spectra are available. Access to exotic materials may be required, such as unusual isotopes, to aid in the identification of catalytic

DOI: 10.1021/acscatal.6b00183
ACS Catal. 2016, 6, 2590−2602

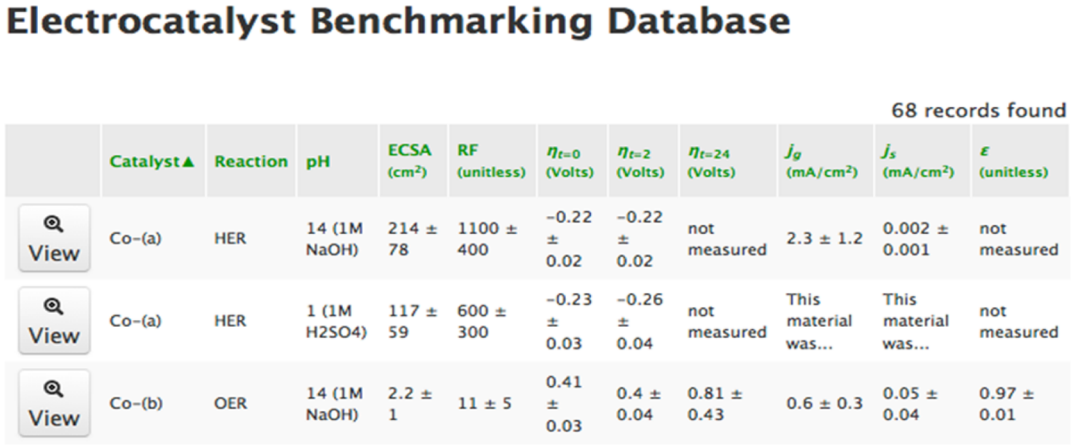


Figure 7. Sample of benchmarking information provided for HER and OER catalysts in the JCAP online database.[61]

intermediates. For example, use of ^{57}Fe would allow the *in situ* recording of Mössbauer spectra in solution. The extra work and expense required to obtain this type of benchmarking data could greatly increase the value of catalytic reports and may enable more direct comparisons of different catalysts and catalytic procedures.

2.4. Electrocatalysis. Given their high efficiency and other appealing environmental benefits, technologies based on electrocatalytic processes are particularly promising in the quest for more sustainable energy. The benchmarking of both homogeneous and heterogeneous electrocatalysts involves quantitative comparisons of appropriately normalized electrocatalytic activity and selectivity.

2.4.1. Best Practices. Standardized Reporting. It is important to report catalyst performance for both molecular and heterogeneous systems using common standards. Each electrocatalysis community should standardize how it reports catalytic performance (e.g., activity, selectivity) in a manner that allows for ready comparison among different catalyst materials for a particular reaction and type of catalyst (e.g., molecular or surface-confined). Importantly, currents (or current densities) should be reported as a function of overpotential and normalized to several different catalyst parameters so that all researchers can better understand the catalyst performance. For example, in electrocatalysis on surfaces, catalytic currents should be normalized to electrode geometric areas (e.g., with dimensions of A/cm$^2_{geo}$), catalyst electrochemically active surface areas (ECSA, A/cm$^2_{ECSA}$), and catalyst loadings (A/g). An example is available in the JCAP database of heterogeneous systems for the hydrogen evolution reaction, HER, and the oxygen evolution reaction, OER, as shown in Figure 7.[61]

Complete Descriptions of Experimental Protocol and Methodology. A full description of both experimental protocol and methodology is essential in electrocatalysis reports. The level of detail should be sufficient for comparing and benchmarking catalysts and to allow different research groups to reproduce the results. The information that should be included will vary according to the catalyst type, but it should generally include concentrations of catalyst, reactants and electrolyte, scan rate for cyclic voltammograms, Faradaic efficiency, TOF (with a clear description of how it was determined), dependence of current on scan rate, pH for aqueous solutions, temperature, cell material, cell geometry, treatment or polishing of the electrode, long-term stability, overpotential (including clear information about how the thermodynamic potential was determined, estimated, or calculated), particle size, loading, mass activity, specific activity, and normalized electrochemical surface area activity.

Chemical and Physical Properties of Electrocatalysts. Researchers investigating electrocatalysis by metal particles/alloys, oxides, or complex materials should provide experimental evidence for both the chemical composition and the physical characteristics of catalysts with interesting properties. This information allows researchers to make fair and informed comparisons of kinetics data, and ensures that proposed reaction mechanisms are based on the clearly stated assumptions regarding which sites are active. Accordingly, authors should provide as much of the following data as possible, obtained using the techniques listed in parentheses: chemical composition (XPS or energy-dispersive X-ray spectroscopy, EDX); phase and crystallinity (XRD or selected area electron diffraction, SAED); morphology, particle size, and size distribution (a histogram of SEM- and/or TEM-imaged particles, including multiple dimensions for rods and other anisotropic shapes); preferred orientation, exposed facets, defect structure, and analysis of edge or vertex sites (HR-TEM, STEM, or high-angle annular dark field, HAADF, imaging).

Similarly, researchers characterizing molecular electrocatalysts should provide diffusion coefficients of all reactants and products (determined, for example, by NMR spectroscopy), as well as details of the electrode geometry and area, among others, using the techniques mentioned above, in addition to routine characterization of catalyst composition, structure, and purity as is common to molecular chemistry. These parameters will help accelerate the translation of laboratory-scale reactivity to the process scale as researchers across the community are better able to repeat, reproduce, and assess mechanistic claims, thereby freeing time to explore, modify, optimize, and ultimately improve new electrocatalysts.

Researchers in both the molecular and heterogeneous electrocatalysis communities should report catalyst loading (mass and/or concentration, and reaction volume). Besides characterization of the catalysts, characterization of reaction intermediates and products is also needed, for example, by use of *in situ* techniques such as surface enhanced infrared absorption spectroscopy (SEIRAS) and differential electrochemical mass spectrometry (DEMS), under both stationary and flow conditions.

DOI: 10.1021/acscatal.6b00183
ACS Catal. 2016, 6, 2590−2602

Time-Dependent Behavior of Electrode Performance. Reporting the approach to a steady-state current (if any) at an applied potential is critical for understanding electrocatalytic performance. This time-dependent behavior is affected by the appearance of reactant and product concentration gradients in the electrolyte, development of electrical double layers, and adsorption transients of reactants or spectator ion species. The characteristic times of these processes are often longer than residence times during a voltage sweep. The time dependence of the morphology or surface structure of the electrode must also be carefully examined and reported.

2.4.2. Challenges and Opportunities. It is critical to examine various catalyst compositions, sizes, shapes, and environments, particularly under reaction conditions. The ideal benchmark for nanoparticle catalysts is a single, structurally well-defined particle that yields identical electrochemical reaction rates under a standard set of conditions. Ideally, a library of these materials having different sizes, shapes, and compositions would be created and made widely available to researchers. The use of individual nanoparticles would eliminate ensemble effects (i.e., arising from polydispersity) for more reliable structure−function correlations. In concept, this resembles the use of well-defined chemical standards purchased from a commercial laboratory to benchmark a particular assay.

Because Pt-containing materials have played a dominant role in energy-related electrocatalysis for decades, there are many reliable methods to determine the ECSA for Pt-based electrocatalysts, including CO adsorption, H_2 underpotential deposition (in acidic media), and metal underpotential deposition (e.g., of Cu or Pb). However, the recent emergence of non-Pt group metal catalysts (metals, metal oxides, carbides, etc.) with interesting electrocatalytic properties has revealed that traditional Pt ECSA methods are insufficient, because of fundamental differences between the surface chemistry of Pt and other materials. Electrochemical double-layer capacitance-based approaches require unproven assumptions and are known to be inaccurate. The lack of a reliable method to quantify the ECSA of non-Pt electrocatalysts has limited our ability to benchmark activities for these catalysts, because the ECSA is a fundamental electrochemical parameter needed to define the specific activity and exchange current density of each catalyst. The development of chemistry-specific methods would be helpful to the community, but the discovery of a universal method to determine the electrochemically active area of known and unknown materials in their reacting environments would be even more valuable, representing a step change in our understanding and benchmarking of catalysts for many different energy applications.

Advances in the field will require the development of and ready access to *operando* tools to characterize speciation in complex electrochemical environments. There is a profound need, as well as a significant opportunity, to develop new *operando* characterization methods that can fully define the atomic, compositional, and molecular speciation of electrochemical systems with high temporal and spatial resolution. At present, the atomic-scale properties and structural dynamics of materials mediating electrochemical catalysis, and the molecular adjuncts through which such transformations may be coupled, remain poorly understood. Progress will require new approaches to fully document structural features that evolve during catalytic transformations mediated by potential-driven electron transfer processes. No single methodology will suffice, and new multiprobe methodologies and supporting theory will be needed to meet this challenge.

The development of computational tools will provide atomic-level understanding of the chemistry and electrostatics at solid−electrolyte interfaces, including more complete characterization of the double-layer and elucidation of how heterogeneity of the electrocatalyst and/or electrode affects reactive transformations. The electrocatalyst/electrode−electrolyte interface is extremely complex, because interactions between the phases are both chemical (short-range) and electrostatic (long-range). Our current understanding of the properties of this interface is rudimentary, being based mainly on mean-field theory (Gouy−Chapman−Stern) that lacks chemical content. Thus, spatially large-scale, long-time simulations that combine molecular or Newtonian dynamics with *ab initio* quantum mechanical calculations (e.g., DFT) should be developed to provide atomic-level understanding. The purpose of these calculations is to help interpret electrochemical and spectroscopic measurements, determine mechanisms, predict the behavior of new materials and phases to guide synthesis, and ultimately discover more efficient electrocatalytic systems.

Another critical area is the development and benchmarking of electrocatalysts for selective activation of strong bonds (e.g., N−N, C−C, C−O) close to their thermodynamic potentials. Complex, multifunctional catalysts will undoubtedly be necessary. Designing such catalysts represents a grand challenge for electrocatalysis because of the great difficulties associated with these processes and their significant technological potential. For example, efficient splitting of the C−C bond in ethanol, an energy-dense, renewable, and readily available fuel, could facilitate widespread use of fuel cells in transportation and as residential and portable power sources. Efficient splitting of N−N bonds could allow ammonia synthesis via an electrochemical variant of the Haber−Bosch process.

3. GENERAL CONCLUSIONS

Adopting Community-Defined Standards. Properly conducted and documented catalyst performance measurements are critical to promoting synergistic activities within and between different catalysis communities, as well as to advancing the field of catalysis as a whole. New researchers could benefit from readily accessible guides about how to best make such measurements and compare new catalysts with standard catalysts in terms of activity, selectivity, and deactivation. The journal *Organometallics* recently instituted a new practice of publishing tutorials on how to conduct and assess particular types of experiments.[62] A similar initiative would be useful for the broad and diverse catalysis community and would complement recommendations on how to perform and report the results of catalytic experiments to maximize their impact.[57]

However, comparisons of catalytic performance between different communities remain constrained by major variations in methods and even terminology. For example, differences in rate-limiting processes in electrocatalysis, such as reactant transport to the electrode in heterogeneous systems vs reactant and catalyst transport in molecular systems, have resulted in the use of very different methods, metrics, and even units to define catalyst performance. Similarly, TOFs are often not reported in heterogeneous electrocatalysis, partly as a consequence of challenges in determining the number of active sites. In the biochemistry community, the term "turnover number" is still used to describe the activity (TOF) of enzymatic systems, in

DOI: 10.1021/acscatal.6b00183
ACS Catal. 2016, 6, 2590−2602

contrast to the current usage of TON to describe catalyst productivity in other catalysis subfields. The use of TOF is appropriate to describe the steady-state activity of a heterogeneous catalyst in a flow reactor but is generally inappropriate for the nonsteady-state batch reactor conditions commonly used with homogeneous catalysts.

Incentivizing Reproducibility. In the broader scientific landscape, disclosures of poor reproducibility and/or generalizability in the results of biomedical and behavioral[63] research have generated ample attention from the popular press,[64] with attendant concerns about costs to society resulting from wasted resources and loss of public trust.[65] These issues have stimulated broad discussion and inquiry regarding the barriers to and incentives for documenting and archiving research results in all fields.[66] The challenge can be exacerbated when journals, funding agencies, and institutional reward systems emphasize novelty without sufficiently valuing openness and reproducibility. A recent editorial in *Analytical Chemistry* suggested that researchers and their institutions need to accept more responsibility when they conduct irreproducible research.[67]

However, there are promising signs of change. Best practices were presented in a 2009 U.S. National Academies report entitled "Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age".[68] The Open Knowledge Foundation has an Open Data in Science working group that encourages access to scientific data and promotes the freedom to use, reuse, and redistribute data with appropriate attribution.[69] The journal *Nature* has instituted a checklist for its authors, requiring complete descriptions of methods, better characterization of reagents, and precise descriptions of statistical methods.[70] Authors are encouraged to provide the source data used to make their figures and to upload detailed step-by-step experimental protocols in open data repositories. The journal *Science* recently announced new publication guidelines named TOP (Transparency and Openness Promotion), and called for more clearly defined rules on the sharing of data and methods.[71] For example, one recommendation calls for extending citations to data, code, and research materials and for making such information publicly available in trusted repositories.

Several funding agencies, including the Department of Energy's Office of Science, the National Science Foundation, and the National Institutes of Health in the U.S.A., the Wellcome Trust and the United Kingdom Research Council, as well as the European Union's Horizon 2020 Research and Innovation Project, are starting to require more extensive data management and data sharing from their grantees. The issue has even appeared on government agendas, including the U.S. House of Representatives Committee on Science, Space, and Technology, Subcommittee on Research, and the U.K. House of Commons Select Committee on Science and Technology.[72] There is legitimate concern in the scientific community about the additional time commitment and the need to depend on essentially volunteer curation efforts that are difficult to sustain over the long-term, or on commercial operations that may attempt to monetize scientific information created with public funds. These concerns could be allayed by free and open source infrastructure to record laboratory work, manage scientific workflow, track the history of documents and data sets, record the provenance of project materials, create web-based computing notebooks, manage data-intensive research, and archive the results.[73]

4. PROSPECTS

We are entering a new era in catalysis, stimulated by recent advances in both experimental and theoretical methods. Complementary spectra of working catalysts, measured quasi-simultaneously using multitechnique instruments, are providing unprecedented atomic-scale understanding of catalytic sites and reaction mechanisms. In parallel, rapid advances in theoretical methods are helping to unify and allow better interpretation of experimental results. Making full use of these opportunities will require that benchmarking, in combination with the "best practices" such as those described in this Perspective, be embraced by both new and established catalysis researchers. Although some activities may be difficult for individuals to implement in the scope of small projects, initiatives funded on a much larger scale are providing researchers with opportunities to show leadership in creating and curating benchmarked databases,[61,74] to which smaller groups may be encouraged to contribute. We anticipate that such concerted efforts, both within and among the catalysis subfields, will enable and accelerate continued advances in catalysis science.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscatal.6b00183.

Code-sharing, data sharing, and archiving tools (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail for J.G.C.: jgchen@columbia.edu.
*E-mail for S.L.S.: sscott@engineering.ucsb.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors are grateful to the Catalysis Science Program, Office of Science, Basic Energy Sciences, of the U.S. Department of Energy, for support of their research programs, and to Catalysis Program Managers Raul Miranda, Viviane Schwartz, and Charles Peden for helpful advice and discussions.

■ REFERENCES

(1) Estimated on the basis of 1 × 10^6 catalysis papers published in archival journals in the preceding decade. With an average output of 5 papers per group per year, there must be 20 000 groups working in the field. If each group has on average 5 members, there are ca. 100 000 catalysis researchers publishing papers. Industrial researchers represent another important catalysis community, not counted here.

(2) "benchmark, v.". In *OED Online* [Online] December 2015 ed.; Oxford University Press. http://www.oed.com/view/Entry/251392?result=2&rskey=fbkKBY& (accessed January 17, 2016).

(3) Sim, S. E.; Easterbrook, S.; Holt, R. C. *Using Benchmarking to Advance Research: A Challenge to Software Engineering*, 25th International Conference on Software Engineering (ICSE '03), Washington, DC, IEEE Computer Society: Washington, DC, 2003; pp 74−83.

(4) Venema, V. Variable Variability. http://variable-variability.blogspot.com/2012/01/what-distinguishes-benchmark.html (accessed January 17, 2016).

(5) Farlin, J.; Majewsky, M. *Environ. Sci. Technol.* **2015**, *49*, 2604−2605.

(6) Schwab, M.; Karrenbach, N.; Claerbout, J. *Comput. Sci. Eng.* **2000**, *2*, 61−67.

DOI: 10.1021/acscatal.6b00183
ACS Catal. 2016, 6, 2590−2602

(7) Buckheit, J.; Donoho, D. L. In *Wavelets and Statistics*, Antoniadis, A., Ed.; Springer: New York, 1995; pp 55−81.
(8) Donoho, D. L. *Biostatistics* **2010**, *11*, 385−388.
(9) Piwowar, H. A.; Day, R. S.; Fridsma, D. B. *PLoS One* **2007**, *2*, e308.
(10) Wells, P. B. *Platinum Met. Rev.* **1985**, *29*, 168−174.
(11) Bond, G. C. In *Handbook of Heterogeneous Catalysis*, Ertl, G., Knözinger, H., Schüth, F., Weitkamp, J., Eds. Wiley-VCH: Weinheim, Germany, 2008; Vol. *1*, pp 700−715.
(12) World Gold Council. *Gold Bull.* **2003**, *36*, 24.
(13) Bond, G. C.; Sing, K. S. W.; Everett, D. H.; Haul, R. A. W.; Moscou, L.; Pierotti, R. A.; Rouquerol, J.; Siemieniewska, T. In *Handbook of Heterogeneous Catalysis, Part A*, Ertl, G., Knözinger, H., Weitkamp, J., Eds.; Wiley-VCH Verlag GmbH: Weinheim, Germany, 1997; pp 1489−1558.
(14) Hammond, K. D.; Conner, W. C., Jr. *Adv. Catal.* **2013**, *56*, 1−101.
(15) Alexeev, O.; Kim, D.-W.; Graham, G. W.; Shelef, M.; Gates, B. C. *J. Catal.* **1999**, *185*, 170−181.
(16) Kaminsky, W. *J. Chem. Soc., Dalton Trans.* **1998**, 1413−1418.
(17) Resconi, L.; Cavallo, L.; Fait, A.; Piemontesi, F. *Chem. Rev.* **2000**, *100*, 1253−1345.
(18) Doyle, M. P.; Forbes, D. C. *Chem. Rev.* **1998**, *98*, 911−936.
(19) Alsters, P. L.; Witte, P. T.; Neumann, R.; Sloboda Rozner, D.; Adam, W.; Zhang, R.; Reedijk, J.; Gamez, P.; ten Elshof, J. E.; Chowdhury, S. R. In *Mechanisms in Homogeneous and Heterogeneous Epoxidation Catalysis*; Oyama, S. T., Ed.; Elsevier: Amsterdam, 2008; pp 415−428.
(20) (a) Artero, V.; Savéant, J.-M. *Energy Environ. Sci.* **2014**, *7*, 3808−3814. (b) Costentin, C.; Savéant, J.-M. *ChemElectroChem* **2014**, *1*, 1226−1236. (c) Costentin, C.; Passard, G.; Savéant, J.-M. *J. Am. Chem. Soc.* **2015**, *137*, 5461−6467.
(21) (a) Gasteiger, H. A.; Kocha, S. S.; Sompalli, B.; Wagner, F. T. *Appl. Catal., B* **2005**, *56*, 9−35. (b) Frydendal, R.; Paoli, E. A.; Knudsen, B. P.; Wickman, B.; Malacrida, P.; Stephens, I. E. L.; Chorkendorff, I. *ChemElectroChem* **2014**, *1*, 2075−2081.
(22) (a) Colic, V.; Tymoczko, J.; Maljusch, A.; Ganassin, A.; Schuhmann, W.; Bandarenka, A. S. *ChemElectroChem* **2015**, *2*, 143−149. (b) McCrory, C. C. L.; Jung, S.; Ferrer, I. M.; Chatman, S. M.; Peters, J. C.; Jaramillo, T. F. *J. Am. Chem. Soc.* **2015**, *137*, 4347−4357. (c) McCrory, C. C. L.; Jung, S.; Peters, J. C.; Jaramillo, T. F. *J. Am. Chem. Soc.* **2013**, *135*, 16977−16987.
(23) (a) Steinmann, S. N.; Corminboeuf, C. *J. Chem. Theory Comput.* **2011**, *7*, 3567−3577. (b) Marom, N.; Tkatchenko, A.; Rossi, M.; Gobre, V. V.; Hod, O.; Scheffler, M.; Kronik, L. *J. Chem. Theory Comput.* **2011**, *7*, 3944−3951.
(24) Santra, B.; Michaelidis, A.; Scheffler, M. *J. Chem. Phys.* **2007**, *127*, 184104.
(25) Xu, X.; Zhang, W.; Tang, M.; Truhlar, D. G. *J. Chem. Theory Comput.* **2015**, *11*, 2036−2052.
(26) Korth, M.; Grimme, S. *J. Chem. Theory Comput.* **2009**, *5*, 993−1003.
(27) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009**, *5*, 324−333.
(28) Chen, S.; Raugei, S.; Rousseau, R.; Dupuis, M.; Bullock, R. M. *J. Phys. Chem. A* **2010**, *114*, 12716−12724.
(29) Mollenhauer, D.; Gaston, N. *J. Comput. Chem.* **2014**, *35*, 986−997.
(30) (a) Guner, V.; Khuong, K. S.; Leach, A. G.; Lee, P. S.; Bartberger, M. D.; Houk, K. N. *J. Phys. Chem. A* **2003**, *107*, 11445−11459. (b) Keith, J. A.; Carter, E. A. *J. Chem. Theory Comput.* **2012**, *8*, 3187−3206.
(31) Yang, K.; Zheng, J.; Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2010**, *132*, 164117.
(32) Bhattacharya, S.; Levchenko, S. V.; Ghiringhelli, L. M.; Scheffler, M. *Phys. Rev. Lett.* **2013**, *111*, 135501.
(33) Paz-Borbón, L. O.; Barcaro, G.; Fortunelli, A.; Levchenko, S. V. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2012**, *85*, 155409.
(34) (a) Ghiringhelli, L. M.; Gruene, P.; Lyon, J. T.; Rayner, D. M.; Meijer, G.; Fielicke, A.; Scheffler, M. *New J. Phys.* **2013**, *15*, 083003. (b) Wellendorff, J.; Silbaugh, T. L.; Garcia-Pintos, D.; Nørskov, J. K.; Bligaard, T.; Studt, F.; Campbell, C. T. *Surf. Sci.* **2015**, *640*, 36−44.
(35) Kroes, G.-J. *J. Phys. Chem. Lett.* **2015**, *6*, 4106−4114.
(36) Hwang, T.; Goldsmith, B. R.; Peters, B.; Scott, S. L. *Inorg. Chem.* **2013**, *52*, 13904−13917.
(37) Goldsmith, B. R.; Hwang, T.; Seritan, S.; Peters, B.; Scott, S. L. *J. Am. Chem. Soc.* **2015**, *137*, 9604−9616.
(38) Plata, R. E.; Singleton, D. A. *J. Am. Chem. Soc.* **2015**, *137*, 3811−3826.
(39) (a) Campbell, C. T. *J. Catal.* **2001**, *204*, 520−524. (b) Stegelmann, C.; Andreasen, A.; Campbell, C. T. *J. Am. Chem. Soc.* **2009**, *131*, 8077−8082.
(40) Kitchin, J. R. *ACS Catal.* **2015**, *5*, 3894−3899.
(41) Aalbersberg, I. J.; Atzeni, S.; Koers, H.; Specker, B.; Zudilova-Seinstra, E. *Liber Quart.* **2014**, *23*, 274−299.
(42) Peterson, K. A.; Feller, D.; Dixon, D. A. *Theor. Chem. Acc.* **2012**, *131*, 1079.
(43) Rohmann, K.; Hölscher, M.; Leitner, W. *J. Am. Chem. Soc.* **2016**, *138*, 433−443.
(44) Jensen, J. H. *Phys. Chem. Chem. Phys.* **2015**, *17*, 12441−12451.
(45) Jones, G.; Bligaard, T.; Abild-Pedersen, F.; Nørskov, J. K. *J. Phys.: Condens. Matter* **2008**, *20*, 064239.
(46) Norskøv, J. K.; Abild-Pedersen, F.; Studt, F.; Bligaard, T. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 937−943.
(47) Nørskov, J. K.; Bligaard, T.; Rossmeisl, J.; Christensen, C. H. *Nat. Chem.* **2009**, *1*, 37−46.
(48) Calle-Vallejo, F.; Loffreda, D.; Koper, M. T. M.; Sautet, P. *Nat. Chem.* **2015**, *7*, 403−410.
(49) Rodriguez, J. A.; Ma, S.; Liu, P.; Hrbek, J.; Evans, J.; Perez, M. *Science* **2007**, *318*, 1757−1760.
(50) Cargnello, M.; Doan-Nguyen, V.; Gordon, T. R.; Diaz, R. E.; Stach, E. A.; Gorte, R. J.; Fornasiero, P.; Murray, C. B. *Science* **2013**, *341*, 771−773.
(51) (a) Hansen, T. W.; Wagner, J. B. *ACS Catal.* **2014**, *4*, 1673−1685. (b) Su, D. S.; Zhang, B.; Schlögl, R. *Chem. Rev.* **2015**, *115*, 2818−2882.
(52) Madon, R. J.; Boudart, M. *Ind. Eng. Chem. Fundam.* **1982**, *21*, 438−447.
(53) Davis, M. E.; Davis, R. J. *Fundamentals of Chemical Reaction Engineering*; McGraw-Hill: Boston, 2003; pp 184−232.
(54) Canivet, J.; Vandichel, M.; Farrusseng, D. *Dalton Trans.* **2016**, *45*, 4090−4099.
(55) (a) Kozuch, S.; Martin, J. M. L. *ACS Catal.* **2012**, *2*, 2787−2794. (b) Lente, G. *ACS Catal.* **2013**, *3*, 381−382.
(56) Blackmond, D. G. *Angew. Chem., Int. Ed.* **2005**, *44*, 4302−4320.
(57) Jones, C. W.; Smith, D. J.; Gunnoe, T. B.; Zhao, H.; Sautet, P.; Scott, S. L.; Xu, B.-Q. *ACS Catal.* **2014**, *4*, 2827−2828.
(58) Jones, C. W. *Top. Catal.* **2010**, *53*, 942−952.
(59) Hintermair, U.; Franció, G.; Leitner, W. *Chem. Commun.* **2011**, *47*, 3691−3701.
(60) Hickman, A. J.; Villalobos, J. M.; Sanford, M. S. *Organometallics* **2009**, *28*, 5316−5322.
(61) Joint Center for Artificial Photosynthesis (JCAP). Electrocatalyst Benchmarking Database. https://internal.solarfuelshub.org/jcapresources/benchmarking/catalysts/ (accessed January 17, 2016).
(62) Chirik, P. J. *Organometallics* **2015**, *34*, 4783.
(63) (a) Prinz, F.; Schlange, T.; Asadullah, K. *Nat. Rev. Drug Discovery* **2011**, *10*, 712−713. (b) Begley, C. G.; Ellis, L. M. *Nature* **2012**, *483*, 531−533. (c) Ioannidis, J. P. A. *PLoS Med.* **2005**, *2*, e124. (d) Shanks, D. R.; Newell, B. R.; Lee, E. H.; Balakrishnan, D.; Ekelund, L.; Cenac, Z.; Kavvadia, F.; Moore, C. *PLoS One* **2013**, *8*, e56515. (e) Open Science Collaboration. *Science* **2015**, *349*, 943.
(64) (a) Anonymous Unreliable Research: Trouble at the Lab. *The Economist*; October 19, **2013**; pp 26−30. (b) Lehrer, J. The Truth Wears Off: Is There Something Wrong with the Scientific Method? *New Yorker*; December 13, **2010**; pp 52−57;. (c) Freedman, D. H. Lies, Damned Lies, and Medical Science. *Atlantic Monthly*; November,

DOI: 10.1021/acscatal.6b00183
ACS Catal. 2016, 6, 2590−2602

2010. (d) Carey, B. Science, Now Under Scrutiny Itself. *New York Times*; June 15, **2015**.
(65) Freedman, L. P.; Cockburn, I. M.; Simcoe, T. S. *PLoS Biol.* **2015**, *13*, e1002165.
(66) Fecher, B.; Friesike, S.; Hebing, M. *PLoS One* **2015**, *10*, e0118053.
(67) Sweedler, J. V. *Anal. Chem.* **2015**, *87*, 11603−11604.
(68) Sharp, P. A.; Kleppner, D. *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age*; Institute of Medicine and National Academy of Sciences. The National Academies Press: Washington, D.C., 2009; http://www.nap.edu/catalog/12615/ensuring-the-integrity-accessibility-and-stewardship-of-research-data-in-the-digital-age (accessed January 17, 2016).
(69) Molloy, J. C. *PLoS Biol.* **2011**, *9*, e1001195.
(70) Anonymous. *Nat. Methods* **2013**, *10*, 367.
(71) (a) Transparency and Openness Promotion Committee. https://cos.io/top/ (accessed January 17, 2016). (b) Nosek, B. A.; Alter, G.; Banks, G. C.; Borsboom, D.; Bowman, S. D.; Breckler, S. J.; Buck, S.; Chambers, C. D.; Chin, G.; Christensen, G.; Contestabile, M.; Dafoe, A.; Eich, E.; Freese, J.; Glennerster, R.; Goroff, D.; Green, D. P.; Hesse, B.; Humphreys, M.; Ishiyama, J.; Karlan, D.; Kraut, A.; Lupia, A.; Mabry, P.; Madon, T.; Malhotra, N.; Mayo-Wilson, E.; McNutt, M.; Miguel, E.; Paluck, E. L.; Simonsohn, U.; Soderberg, C.; Spellman, B. A.; Turitto, J.; VandenBos, G.; Vazire, S.; Wagenmakers, E. J.; Wilson, R.; Yarkoni, T. *Science* **2015**, *348*, 1422−1425. (c) Alberts, B.; Cicerone, R. J.; Fienberg, S. E.; Kamb, A.; McNutt, M.; Nerem, R. M.; Schekman, R.; Shiffrin, R.; Stodden, V.; Suresh, S.; Zuber, M. T.; Pope, B. K.; Jamieson, K. H. *Science* **2015**, *348*, 1420−1422.
(72) (a) *Scientific Integrity and Transparency*; 113-10; Committee on Science, Space, and Technology, Subcommittee on Research, U.S. Government Printing Office: Washington, D.C., March 5, 2013; http://www.gpo.gov/fdsys/pkg/CHRG-113hhrg79929/pdf/CHRG-113hhrg79929.pdf (accessed January 17, 2016). (b) *Peer review in scientific publications: Eighth Report of Session 2010−2012; HC-856*; House of Commons Science and Technology Committee, The Stationary Office, Ltd.: London, U.K., 2011; http://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/inquiries/peer-review (accessed January 17, 2016).
(73) Buck, S. *Science* **2015**, *348*, 1403.
(74) (a) NIST Materials Genome Initiative. https://mgi.nist.gov/density-functional-theory-dft-informatics-and-repositories (accessed January 17, 2016). (b) NoMaD (Novel Materials Discovery) Repository. http://nomad-repository.eu/cms/ (accessed January 17, 2016). (c) Benchmark Energy & Geometry Database (BEGDB). http://www.begdb.com/ (accessed January 17, 2016).

DOI: 10.1021/acscatal.6b00183
ACS Catal. 2016, 6, 2590−2602