

MIT Open Access Articles

When Is Ligand pK_a a Good Descriptor for Catalyst Energetics? In Search of Optimal CO_2 Hydration Catalysts

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Kim, Jeong Yun and Heather J. Kulik. "When Is Ligand pK_a a Good Descriptor for Catalyst Energetics? In Search of Optimal CO_2 Hydration Catalysts." *Journal of Physical Chemistry A* 122, 18 (April 2018): 4579-4590 © 2018 American Chemical Society

As Published: <http://dx.doi.org/10.1021/acs.jpca.8b03301>

Publisher: American Chemical Society (ACS)

Persistent URL: <https://hdl.handle.net/1721.1/123833>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



When is Ligand pK_a a Good Descriptor for Catalyst Energetics? In Search of Optimal CO_2 Hydration Catalysts

Jeong Yun Kim¹ and Heather J. Kulik^{1,*}

¹*Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA*

02139

ABSTRACT: We present a detailed study of nearly 70 Zn molecular catalysts for CO_2 hydration from four diverse ligand classes ranging from well-studied carbonic anhydrase mimics (e.g., cyclen) to new structures we obtain by leveraging diverse hits from large organic libraries. Using kinetic analysis and establishing linear free energy relationships, we confirm turnover to be sensitive to relative thermodynamic stability of reactive hydroxyl and bound bicarbonate moieties. We observe a wide range of thermodynamic stabilities for these intermediates, showing up to 6 kcal/mol improvement over well-studied cyclen catalysts. We observe a good correlation between the pK_a of the Zn-OH₂ moiety and the resulting relative stability of hydroxyl moieties over bicarbonate, which may be rationalized by the dominant effect of higher Zn-OH bond order over weak bonding in bicarbonate and water. A direct relationship is identified between isolated organic ligand pK_a and the pK_a of a bound water molecule on the catalyst. Thus, organic ligand pK_a , which is intuitive, easy to compute or tabulate, and much less sensitive to electronic structure method choice than whole-catalyst properties, is a good quantitative descriptor for predicting the effect of through-bond electronic effects on relative CO_2 hydration energetics. We expect this to be applicable to other reactions where hydroxyl moieties are essential to stabilize as turnover-determining species with respect to more weakly-bound moieties. Finally, we note exceptions for rigid ligands (e.g., porphyrins) that are observed to preferentially stabilize hydroxyl over bicarbonate without reducing pK_a values as substantially. We expect the strategy outlined here, to i) curate diverse ligands from large organic libraries and ii) identify when ligand-only properties can determine catalyst energetics, to be broadly useful for both experimental and computational catalyst design.

1. Introduction

Catalyst design through high-throughput computational¹⁻⁷ or experimental⁸⁻¹⁰ approaches has emerged as an essential tool to accelerate discovery of paths to efficient resource and energy utilization. One prominent target for catalyst design, the selective capture and sequestration of CO₂, e.g., to aqueous phase as bicarbonate (HCO₃⁻), is recognized as an important means by which atmospheric CO₂ could be limited. In nature, human carbonic anhydrase (HCAII) is one of the fastest known enzymes with a CO₂ hydration turnover frequency of $\sim 10^6$ s⁻¹ at pH 9 and 298 K that approaches the diffusion limit.¹¹ The active site of HCAII is composed of a tetrahedral Zn(II) center bound to three histidine (His) residues leaving open a fourth coordination site where a water molecule can bind. Water molecules bound to the Zn(II) center in HCAII have reduced pK_a values of around 7 (vs. 15.7 in bulk water) due to the through-bond Zn-His interactions as well as a through-space interaction with a hydroxyl of an adjacent threonine. This produces reactive Zn-OH at physiological pH, which is believed to be the rate-limiting step in HCAII.

Several small-molecule Zn(II) molecular catalysts have been developed and studied as HCAII mimics. Example mimics¹²⁻¹⁴ include Zn(II) catalysts with imidazoles¹⁵ and macrocyclic amine ligands¹⁶⁻¹⁸. Although the rate determining step in mimics is not believed to be water deprotonation as it is in HCAII¹⁹, a comparable mechanism has been proposed for both mimics and the enzyme. Deprotonation of a Zn-bound water and bicarbonate release¹⁸ have both been proposed to be rate-determining depending on the catalyst. Microkinetic models have been developed to study the effect of pH on representative mimics²⁰. Few mimics have approached the impressive turnover frequencies achieved by HCAII. The wide gap between enzymatic and mimic activities has not yet been bridged by the experimental^{18, 21-22} and computational studies¹⁸,

1
2
3 20, 23-29 of mimics or enzymes³⁰⁻³⁵, and few exhaustive catalyst screening efforts have been carried
4
5 out for CO₂ hydration.
6

7
8 Computational catalyst screening¹⁻⁷ can provide valuable insight into the design rules for
9
10 highly active catalysts. Computational heterogeneous catalyst design is greatly accelerated by
11
12 well-established linear free energy relationships (LFERs), i.e., linear scaling relations between
13
14 intermediates³⁶⁻³⁸ and Brønsted-Evans-Polanyi (BEP) relations³⁹⁻⁴⁰ between activation and
15
16 reaction energies, to predict reactivity trends and relate overall catalytic activity to the energies
17
18 of key intermediates^{5, 41-43}. The validity of LFERs in single-site catalysts is less clear, where
19
20 geometric variation, through-bond, and through-space interactions can all tune the chemical
21
22 environment differently. LFERs have been explored most in single-site electrocatalysts⁴⁴⁻⁴⁷
23
24 motivated by the strong parallels between single-site and surface mechanisms.^{41, 48-49} Within
25
26 chemical catalysis, LFERs have occasionally been disrupted⁵⁰ even by modest changes to ligand
27
28 geometry or character. Across the subsets of complexes where LFERs have been identified to
29
30 hold, a focus has been on screening thermodynamic quantities readily characterized through
31
32 experiments (e.g., acidity constants⁵¹⁻⁵², hydricities⁵³⁻⁵⁶, and redox potentials⁵⁷⁻⁵⁹). Generally,
33
34 correlations to quantities⁶⁰⁻⁶⁷ associated with the organic ligand present the most promising
35
36 avenue for rapid experimental or computational catalyst design. When ligand-only properties can
37
38 be correlated to expected activities, both experiment and simulation are greatly simplified. The
39
40 CO₂ hydration reaction presents a useful target for establishing LFERs with experimentally
41
42 observable quantities in design screens. The divergent rate-determining step in the molecular
43
44 mimics provides a challenging test for the applicability of a single LFER. At the same time, the
45
46 role of pK_a in HCAII efficiency motivates it as a potential descriptor for catalytic activity.
47
48
49
50
51
52

53
54 In this work, we present a detailed study of nearly 70 Zn molecular catalysts from four
55
56
57
58
59
60

1
2
3 diverse ligand classes. We identify and rationalize relationships between pK_a and reaction
4 kinetics/thermodynamics, providing guidance for when pK_a can be a good descriptor for reaction
5 energetics, ultimately reconciling differences between HCAII and some widely studied
6 derivatives (i.e., porphyrins or cyclens). The rest of this article is outlined as follows. In section
7
8 2, we provide the Computational Details of the simulations in this work. In section 3, we present
9
10 Results and Discussion starting with the curation of four diverse ligand sets, followed by details
11
12 of the kinetic and thermodynamic relationships that govern turnover in Zn CO₂ hydration
13
14 catalysts, and we determine the limits of universal descriptors, such as metal-bound water pK_a , to
15
16 predict changes in reaction kinetics across each class of Zn catalysts. Finally, in section 4, we
17
18 provide our Conclusions.
19
20
21
22
23
24

25 26 **2. Computational Details**

27
28 *First-principles calculations.* Single point energies and geometry optimizations with density
29 functional theory (DFT) were carried out using Q-Chem 5.0⁶⁸. These DFT calculations employed
30 B3LYP⁶⁹⁻⁷¹ with its default definition in in Q-Chem, which corresponds to the VWN1-RPA form
31 for the LDA VWN⁷² component of LYP⁶⁹ correlation. We selected B3LYP from a number of
32 exchange-correlation functionals after verifying it reproduces the relative experimental CO₂
33 hydration turnover frequencies of a family of Zn(II) aza-macrocyclic catalysts (Supporting
34 Information Figure S1). All geometry optimizations and frequency calculations were performed
35 with the composite LACVP* basis set, which corresponds to treating Zn atoms with the
36 LANL2DZ⁷³ effective core potential and the 6-31G* basis set for the remaining atoms. Reported
37 energies in the main text correspond to single point energies calculated with the larger def2-
38 TZVP basis set⁷⁴ on LACVP* geometries. The def2-TZVP basis set⁷⁴ was selected after
39 extensive evaluation of basis set dependence (Supporting Information Figure S2). Including
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 explicit dispersion by augmenting B3LYP with the empirical DFT-D3 correction⁷⁵ preserves
4 relative energetics on representative test cases (Supporting Information Figure S3).
5
6

7 Aqueous solvent effects were modeled using an implicit polarizable continuum model (PCM)
8 with the conductor-like solvation model (COSMO⁷⁶) and $\epsilon = 78.39$. The solute cavity was built
9 with the conductor-like solvation model (COSMO⁷⁶) and $\epsilon = 78.39$. The solute cavity was built
10 using Bondi's van der Waals radii⁷⁷ scaled by a factor of 1.2 for all elements. For all stationary
11 points, vibrational calculations were carried out to obtain zero-point energies (ZPE) and thermal
12 enthalpy and entropy corrections at 298.15 K. Since the overall translational and rotational
13 motions make negligible contributions to molecular association and dissociation in aqueous
14 solutions at 1 M standard state,⁷⁸ we only included vibrational contributions in thermal
15 corrections (Supporting Information Figure S4). Basis set superposition error (BSSE) in
16 supramolecular complexes of small molecules (CO₂, H₂O, HCO₃⁻, and H₂CO₃) with the catalyst
17 was corrected using the Boys-Bernardi counterpoise correction.⁷⁹
18
19
20
21
22
23
24
25
26
27
28
29
30

31 *Transition States.* Transition state (TS) searches were carried out through: i) obtaining an
32 approximate interpolated path from a converged nudged elastic band (NEB) calculation⁸⁰⁻⁸¹ in
33 solvent environment (i.e., with COSMO^{76, 82}) using TeraChem⁸³⁻⁸⁴ at the B3LYP/LACVP* level
34 of theory followed by ii) optimizing the highest energy NEB image to a true saddle point using
35 partitioned rational function optimization (P-RFO)⁸⁵ with Q-Chem 5.0⁶⁸. This procedure used
36 multiple codes because full path-based MEP search is not implemented in Q-Chem, whereas
37 analytic Hessians and Hessian-based TS optimization are not available in TeraChem. All TSs
38 were characterized with vibrational frequency analysis to confirm a single imaginary frequency
39 corresponding to the expected nuclear motion across the barrier.
40
41
42
43
44
45
46
47
48
49
50

51 Possible CO₂ hydration mechanistic steps were first carried out on representative Zn
52 molecular catalysts, including four Zn(II) aza-macrocyclic catalysts and an imidazole (Zn(imi)₃)
53
54
55
56
57
58
59
60

complex (Supporting Information Figures S1 and S5). We then generated initial guesses for intermediates and transition states of the remaining Zn catalysts using the custom core functionality in the molSimplify toolkit.⁸⁶ We preserved the central core reacting small molecules (CO₂, H₂O, HCO₃⁻, and H₂CO₃), Zn(II), and surrounding nitrogen atoms, which acted as a template for the placement of functionalized ligands, as outlined in our previous discovery work.⁸⁷⁻⁸⁸

Analysis. The pK_a of axially-bound water (M-H₂O) on Zn molecular catalysts was computed with a thermodynamic cycle suggested by Liptak and Shields⁸⁹⁻⁹⁰:

$$\Delta G_{aq}^{\square} - \Delta G_{gas}^{\square} - \Delta G_{sol}^{\square}(\text{M-OH}^{\square}) - \Delta G_{sol}^{\square}(\text{H}^{\square}) - \Delta G_{sol}^{\square}(\text{M-H}_2\text{O}) \quad (1)$$

where the aqueous free energy of water deprotonation is determined from solvation free energy differences evaluated on gas phase structures along with gas phase free energy differences of the intermediates. We used the suggested⁸⁹⁻⁹⁰ empirical values for $G_{gas}(\text{H}^+) = -6.28$ kcal/mol and $\Delta G_{sol}(\text{H}^+) = -264.61$ kcal/mol to obtain $pK_a = \Delta G_{aq}^{\square} / 2.303RT$ at 298.15 K. Since the calculation of ΔG_{gas} uses a reference state of 1 atm and the calculation of ΔG_{sol} uses a reference state of 1 M, we convert ΔG_{gas} to a reference state of 1 M using the relationship⁹¹:

$$\Delta G_{gas}^{\square}(1\text{ M}) - \Delta G_{gas}^{\square}(1\text{ atm}) - RT \ln(24.46) \quad (2)$$

The same procedure was carried out for first-principles ligand pK_a calculations or using semi-empirical calculations in the MarvinSketch⁹² code (**Set 3** ligands only), as detailed in the Supporting Information. Natural bond orbital (NBO) partial charges were obtained as implemented in the Q-Chem interface to the NBO package.⁹³

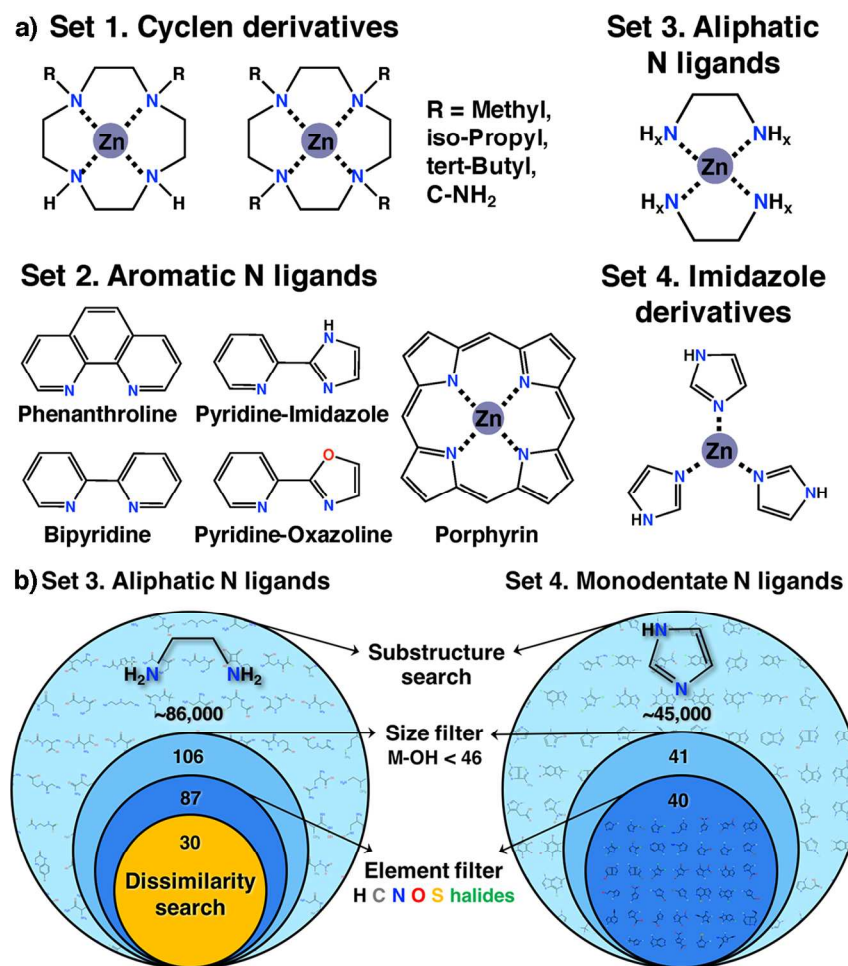
Feature selection. We used regularized multiple linear regression (i.e., LASSO⁹⁴), as implemented in the glmnet⁹⁵ package in R3.4.0 to evaluate candidate descriptor sets and identify

1
2
3 the best-correlating features. We normalized descriptors and centered them to have zero mean.
4
5 We evaluated the minimum leave-one-out cross-validation (LOOCV) error⁹⁴ to judge feature
6
7 sets.
8
9

10 11 12 **3. Results and Discussion**

13 14 15 **3a. Data Set Curation**

16
17
18 To bridge the range of ligands common to both HCAII and its synthetic mimics for the
19
20 CO₂ hydration reaction (CHR), we study both three- or fourfold-coordinated metal–nitrogen
21
22 complexes with Zn(II) metal centers in tetrahedral or square-planar geometries, respectively. We
23
24 curate four distinct sets of nitrogen ligands to also vary the nature of coordinating nitrogen atoms
25
26 from sp^2 or aromatic to sp^3 as well as the denticity of the ligand (Figure 1). **Set 1** includes eight
27
28 1,4,7,10-tetraazacyclododecane (cyclen) derivatives that chelate Zn(II) with four sp^3 N atoms,
29
30 which were obtained from available experimentally characterized cyclens in the Cambridge
31
32 Structural Database (CSD).⁹⁶ These derivatives differ in the number of functionalized N atoms
33
34 (i.e., 2 and 4) and nature of functional groups (i.e., methyl, methylamine, isopropyl, and
35
36 tertbutyl). **Set 2** consists of representative fourfold-coordinated Zn(II) complexes with aromatic
37
38 nitrogen compounds. In addition to tetradentate porphyrin, **Set 2** includes four common bidentate
39
40 ligands characterized by two nitrogen atoms that are separated by two carbon atoms, and at least
41
42 one nitrogen is in a six-membered ring, as is common for many polycyclic chelating ligands⁹⁷
43
44 (Figure 1). **Sets 3** and **4** were obtained through a previously outlined diversity-oriented ligand
45
46 discovery procedure^{59, 88} to obtain: i) fourfold-coordinating bidentate aliphatic sp^2 and sp^3 N
47
48 atoms and ii) threefold-coordinated Zn(II) with functionalized imidazole sp^2 N ligands for **Set 3**
49
50 and **Set 4**, respectively.
51
52
53
54
55
56
57
58
59
60



34
35
36
37
38
39
40
41
42
43
44
45

Figure 1. (a) Schematic of four Zn model catalyst nitrogen ligand sets: **Set 1**, experimentally-characterized functionalized, cyclen derivatives; **Set 2**, bidentate and tetradentate aromatic ligands; **Set 3**, bidentate aliphatic ligands; and **Set 4**, functionalized monodentate imidazole ligands. **Sets 1-3** form four-coordinate Zn complexes as shown, whereas **Set 4** is used to generate three-coordinate Zn complexes. (b) Schematic of nitrogen ligand candidate chemical database search for **Set 3** (left) and **Set 4** (right) with sequential filtering steps indicated by concentric circles with the number of species identified indicated as in inset: i) substructure search, ii) size filter, iii) element filter, iv) and dissimilarity search for **Set 3** only (left).

46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For **Set 3** and **Set 4**, we employed a database search^{59, 88} to discover new ligand candidates from ChEMBL-22⁹⁸ using the molSimplify toolkit.⁸⁶ In both cases, we began by matching a target functional group pattern (i.e., substructure). For **Set 3**, this target pattern was undercoordinated nitrogen atoms (i.e., connected to at most 1 or 2 other species) separated by two aliphatic (i.e., rotatable) carbons (Figure 1). This search alone returns over 86,000

1
2
3 molecules, which requires a reduction in scope. We limited the size of Zn-OH complexes to 45
4 atoms (i.e., 21 atoms for each bidentate ligand) to focus on direct metal-ligand environment
5 effects^{59, 88}, which reduces our ligand pool to 106 molecules. Further filtering to remove
6 counterions, stereoisomers, duplicates, and only allowing organic (H, C, N, O, S) or halide (F,
7 Cl, Br, I) elements produces 87 candidate ligands. Finally after maximizing organic diversity⁹⁹⁻
8 ¹⁰⁰ as judged by differences in molecular fingerprint¹⁰¹ of each ligand to ensure a broad search of
9 chemical space^{59, 88}, we obtain the 30 molecules in **Set 3** (Supporting Information Figure S6). For
10 **Set 4**, we repeat the procedure, with a substructure match for imidazole (Figure 1). The nearly
11 45,000 imidazole-containing molecules are reduced by filtering for size (here, 14 atoms) along
12 with an elemental filter to produce 40 molecules in **Set 4** (Supporting Information Figure S7).

3b. Mechanism of CO₂ hydration reaction for Zn catalysts

13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31 Although the rate determining step may differ, in both HCAII¹⁰²⁻¹⁰³ and its mimics^{18, 20}, a
32 consistent CO₂ hydration reaction (CHR) mechanism has been proposed. We now quantify the
33 key features of the CHR reaction coordinate with Zn(im₃) ligands (Figure 2). Deprotonation of
34 Zn-bound water (**0**) generates the reactive Zn-OH moiety (**1**). Addition of CO₂ produces a
35 weakly bound precursor complex (**2**) characterized by some bending of the CO₂ molecule
36 (Figure 2). Nucleophilic attack (**TS23**) by Zn-OH on CO₂ leads to C-O bond formation and a Zn-
37 bound bicarbonate (**3**) product. Two prominent mechanisms have been proposed to explain the
38 subsequent bicarbonate rearrangement to produce a suitable leaving group (**4**):^{31, 104-106} i) an
39 intramolecular proton transfer or ii) an internal rotation of the bicarbonate (**TS34**). For the
40 HCAII mimic studied here, the internal rotation is more favorable, with a lower activation free
41 energy than that for the proton transfer mechanism (see Supporting Information Figure S8).

Therefore, for this screening effort we consider only the rotation pathway. During bicarbonate release, a solvent water molecule hydrogen bonds with the bicarbonate structure (**5**) before exchanging with bicarbonate (**6**). The removal of bicarbonate leads to the return to the beginning of the catalytic cycle with a Zn-OH₂ species (**0**) that must be deprotonated again. Inclusion of pH-dependent mechanism effects could produce an exothermic rather than endothermic ($\Delta G_r = 13$ kcal/mol) profile observed here, e.g. by producing carbonic acid rather than bicarbonate, but our convention is chosen for consistency with prior computational modeling work.¹⁸ In HCAII¹⁰³, proton transfer between solvent and the metal center^{102, 107} is the diffusion-limited rate determining step, but in slower molecular catalyst mimics, the rate-limiting step is expected to be bicarbonate (HCO₃⁻) release¹⁸ (i.e., **6** to **0**) or CO₂ addition²⁶ (i.e., **2** to **3**).

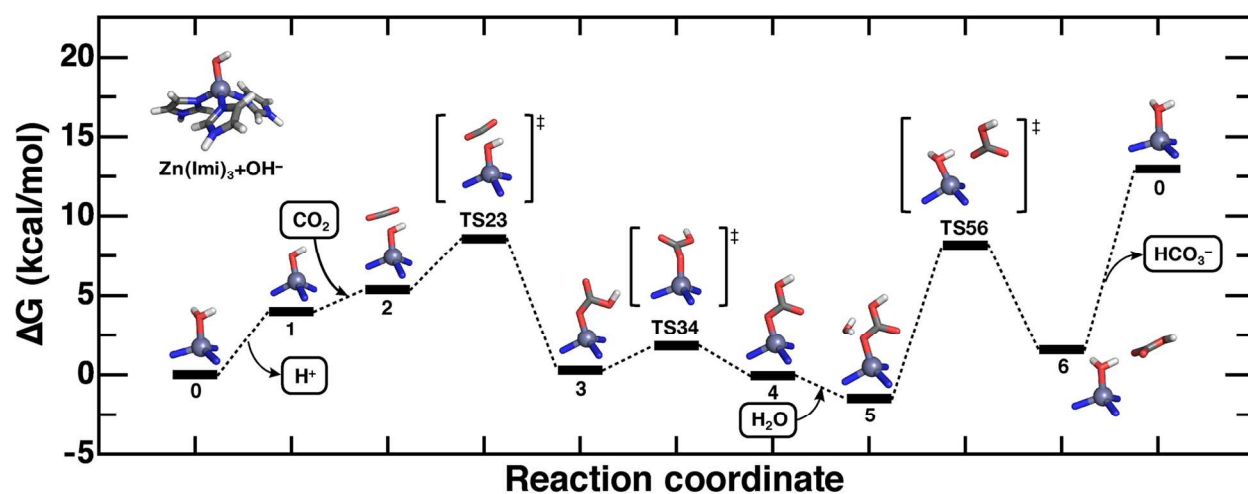


Figure 2. CO₂ hydration reaction relative free energy profile (ΔG , in kcal/mol) with Zn(imi)₃ model catalyst (structure shown in inset at top left) and the free energies aligned to zero at intermediate **0**. Partial TS and intermediate structures are shown in inset with Zn shown as a sphere and coordinating atoms or intermediates are shown as sticks.

For catalyst design, we focus on trends that optimize relative turnover frequencies (TOFs), which are more reliable due to greater cancellation of errors than in absolute TOF prediction. A typical simplifying assumption is that a single turnover determining (i.e., most

stable) intermediate (TDI) and turnover determining (i.e., least stable) transition state (TDTS) may be identified. In that limit, the TOF can be estimated from the free energy difference between the TDTS and TDI, ΔE :¹⁰⁸⁻¹⁰⁹

$$\Delta E = \begin{cases} G_{\text{TDTS}} - G_{\text{TDI}} & \text{if TDTS after TDI} \\ G_{\text{TDTS}} - G_{\text{TDI}} - G_r & \text{if TDTS before TDI} \end{cases} \quad (3)$$

where:

$$\text{TOF} = \frac{k_B T}{h} e^{-\Delta E/RT} \quad (4)$$

This approximation will fail if numerous transition states (TSes) or intermediates contribute equivalently to the reaction kinetics, as judged through the degree of rate control, x_{TOF} , in the complete expression of TOF from all elementary steps, as developed by Campbell¹¹⁰⁻¹¹¹ or Kozuch and coworkers¹⁰⁸:

$$x_{\text{TOF},i} = \left| \frac{1}{\text{TOF}} \frac{\partial \text{TOF}}{\partial E_i} \right| \quad (5)$$

where E_i is a free energy of transition state or intermediate. For the CHR cycle with $\text{Zn}(\text{imi})_3$, **TS23** is the unambiguous TDTS and **5** is the predominant TDI with some contribution from **4** (see Supporting Information Table S1). Although we have limited ourselves to a single TDTS and single TDI, the relaxation of the requirement that they belong to a single elementary step allows us to treat both CO_2 addition (i.e., the TDTS) and bicarbonate release (i.e., the TDI) to be turnover-determining. Thus, our catalyst design screen is focused on variations that reduce the energetic difference between the nucleophilic attack TS (**TS23**) and bound bicarbonate (**4** or **5**) to increase the TOF.

For catalyst screening, however, an additional concern is whether functionalization alters the nature of the TDTS and TDI. As a preliminary investigation and to verify that a range of

energies are achieved for the CHR reaction profile, we compared $\text{Zn}(\text{imi})_3$ energetics to three catalysts with imidazole-derivative ligands functionalized with CH_3 , F, and Cl (Figure 3). Electron-withdrawing groups (EWG: F and Cl) stabilize the Zn-OH bond in relevant intermediates (i.e., **1**, **2**, and **TS23**), reducing the overall reaction barrier, whereas electron-donating groups (EDG: CH_3) do the opposite (Figure 3). Although energetics shift substantially, the nature of the TDTS and TDI is unchanged with this functionalization. Although within a single catalytic cycle for EWGs the TS56 energy is now below TS23, the energy span model (eq. 3) requires a positive shift of TS23 by $\Delta G_r = 13$ kcal/mol, meaning TS23 remains the TDTS regardless of functionalization. Furthermore, a good linear free energy relationship (LFER) is observed between the energy difference of intermediates **2** and **4** (i.e., $\Delta E'$) in comparison to the full ΔE ($R^2 = 1.00$) with a slope of 1, simplifying screening to thermodynamic quantities (Figure 3). Thus, for the remainder of this work we focus on the effect of ligand structure and properties on tuning $\Delta E'$, which should remain a good proxy for the overall TOF of the CO_2 hydration reaction across all ligand sets (Supporting Information Figure S5).

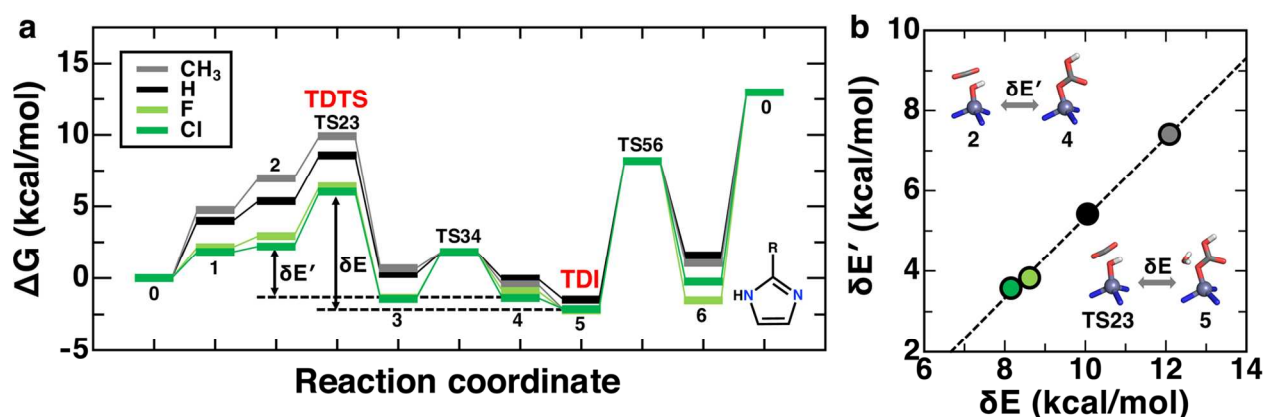


Figure 3. (a) CO_2 hydration reaction coordinate for $\text{Zn}(\text{imi})_3$ ($R = \text{H}$, black line) and three derivatives ($R = \text{CH}_3$ in gray, F in light green, and Cl in green, as indicated in inset). The TDTS and TDI for all four complexes is the same, as labeled, and the TDI-TDTS energetic difference, ΔE , is labeled along with a related thermodynamic quantity, $\Delta E'$. (b) The linear free energy relationship between ΔE and $\Delta E'$ for the catalysts at left, with a best fit line (black, dashed) and

1
2
3 partial structures (Zn sphere and reactive moiety or N atoms shown as sticks) corresponding to
4 the two quantities are shown in inset.
5

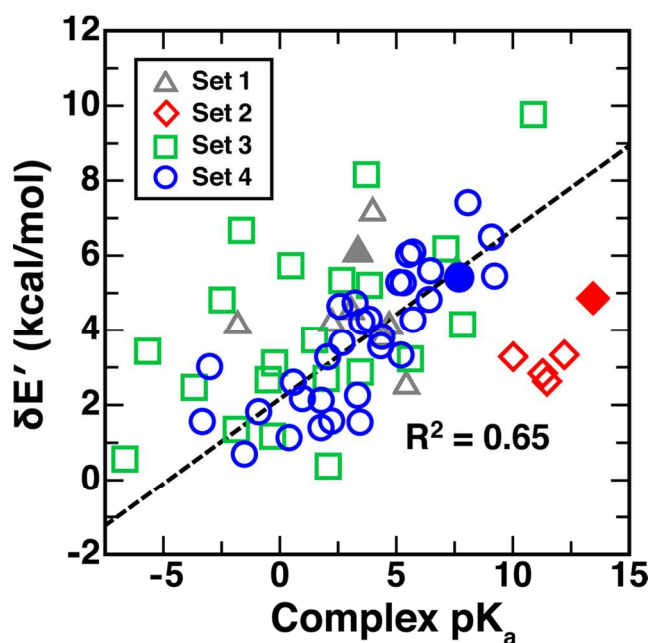
6 **3c. Tuning and Rationalizing Catalytic Efficiency.**

7

8
9 After exclusion of select compounds that form unphysical geometries, we characterize 67
10 catalysts from **Sets 1-4** and find that they span a wide range of the thermodynamic descriptor,
11 $\Delta E'$, from 0 to 10 kcal/mol (Supporting Information Text S1). In comparison to reference
12 compounds such as cyclen (**Set 1**) or imidazole (**Set 4**) that have intermediate $\Delta E'$ values around
13 5-6 kcal/mol, the majority of ligands have substantially reduced $\Delta E'$. The most functionally
14 diverse **Set 3**, which contains both sp^2 and sp^3 nitrogen connecting atoms, spans the widest $\Delta E'$
15 range, but the imidazole derivatives in **Set 4** have a nearly as large ($\Delta E' \sim 0-8$ kcal/mol) variation.
16 Overall, there is no clear division in the energetics sampled by the nature of the chelating
17 nitrogen (Supporting Information Tables S2-S5). The energetic range observed over **Sets 3** and **4**
18 far exceeds the range of more well-known cyclen derivatives (**Set 1**) or other common chelating
19 ligands (**Set 2**), motivating the extraction of design rules from these sets (**3** and **4**) that include
20 more diverse chemical variations.
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38

39 Our subset of functionalized imidazole ligands revealed that changes in $\Delta E'$ arose mainly
40 from stabilization of intermediate **2** (i.e., Zn-OH) with functionalization whereas intermediates **4**
41 (i.e., Zn-OCOOH) and **0** (i.e., Zn-H₂O) were less sensitive to ligand chemistry (see Figure 3).
42 This observation holds over the broader data set (see Supporting Information). The formal bond
43 order of a Zn-OH bond in intermediate **2** corresponds to a stronger interaction with the catalyst
44 than the predominantly electrostatically bound water or bicarbonate. Thus, we hypothesize that it
45 should be possible to relate favorable CHR energetics through intermediate **2** or **TS23**
46 stabilization relative to **4** or **5** to the Zn-OH₂ pK_a (i.e., **0** vs. **1**) which is an intuitive,
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 experimentally-observable measure of relative Zn-OH stability. Indeed, pK_a of the Zn-bound
4 water for diverse nitrogen ligands reveals a wide range of pK_a values of around 20 units (Figure
5
6
7
8 4). Within each ligand set, a good correlation is observed between $\Delta E'$ and pK_a of Zn-bound
9
10 water, with reduction in pK_a corresponding to improved CHR energetics (Figure 4 and
11
12 Supporting Information Tables S2-S5). Differences between trends among the sets will be
13
14 revisited in Sec. 3e. The good correlation of low pK_a to favorable $\Delta E'$ suggests that relative
15
16 stabilization of Zn-OH with respect to Zn-OCOOH⁻ is similar to the former species' stabilization
17
18 with respect to Zn-OH₂. This result is somewhat surprising due to differences in charge of
19
20 bicarbonate versus water.
21
22
23



45
46 **Figure 4.** Modified energetic span ($\Delta E'$, in kcal/mol) for CO₂ hydration plotted against the pK_a of
47 Zn-bound water (complex pK_a) for 67 catalysts with nitrogen ligands classified by set type with
48 color and symbol as indicated in inset legend. Representative **Set 1** (cyclen), **Set 2** (porphyrin),
49 and **Set 4** (imidazole) ligands are indicated as filled symbols. A trendline (dashed, black) is fit to
50 **Set 4** with correlation coefficient indicated in inset.
51

52 Although a general trend is observed, it is most straightforward to compare the
53
54 relationship between $\Delta E'$ and pK_a for individual sets. Of all ligand subsets, the 32 functionalized
55
56

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 imidazole ligands in **Set 4** exhibit the best correlation ($R^2 = 0.65$) between $\Delta E'$ and the pK_a of the
4 Zn-H₂O (Figure 4 and Supporting Information Text S1). The relationship for **Set 4** ligands
5 suggests that a unit reduction in pK_a with respect to the unfunctionalized imidazole reference
6 reduces $\Delta E'$ by nearly 0.5 kcal/mol (Supporting Information Figure S9). Over the 32 cases,
7 imidazole ligands functionalized with strong electron-withdrawing groups (e.g., nitro and cyano),
8 adjacent to the metal-coordinating nitrogen atom generally have the lowest pK_a values and most
9 favorable CHR energetics. In contrast, electron-donating groups (e.g., primary amines and alkyl
10 groups) give rise to elevated pK_a values and less favorable $\Delta E'$ than the unfunctionalized case.
11 We will revisit shortly the relative importance of through-bond, electronic origins for CHR
12 tuning in which weaker chelating N atoms strengthen the Zn-OH interaction preferentially in **Set**
13 **4**.

14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30 For the 21 aliphatic nitrogen compounds in **Set 3**, a correlation is observed between $\Delta E'$
31 and the pK_a of Zn-H₂O, but it is somewhat weaker ($R^2 = 0.30$, see Figure 4 and Supporting
32 Information Figure S10). This weaker correlation also corresponds to a shallower slope of
33 around 0.3 kcal/mol reduction in $\Delta E'$ with unit reduction in pK_a . Some of the broader variation in
34 **Set 3** can be attributed to differences in sp^2 and sp^3 N chelating atoms in the ligands. Of the 21
35 molecules, 4 sp^2 N-only complexes have the most favorable $\Delta E'$ and lowest pK_a , generally lying
36 below the overall trendline (Supporting Information Figure S10). Conversely, the 4 sp^3 N-only
37 complexes predominantly lie above the trendline, with higher $\Delta E'$ for comparable pK_a to the sp^2
38 N cases. Overall, the remaining 13 mixed ligands sample a broad range of both quantities and lie
39 along the overall trendline.

40
41
42
43
44
45
46
47
48
49
50
51
52
53
54 A smaller chemical space is spanned by the **Set 1** cyclen derivatives since they were
55
56
57
58
59
60

1
2
3 selected from experimentally characterized species, and as a result these compounds span pK_a
4 and $\Delta E'$ ranges roughly half of those observed in the other subsets. The main variation in
5 energetics for **Set 1** is due to the bulk (e.g., methyl vs. t-butyl) of the ligands (see Supporting
6 Information Figure S11 and Text S2). The bulkiest t-butyl groups can weaken or disrupt the Zn-
7 N bonds due to steric clashing. In contrast to **Sets 3** and **4**, the **Set 1** pK_a and $\Delta E'$ values are
8 observed to decrease slightly for bulky functionalizations due to through-space hydrogen
9 bonding between the Zn-bound moieties (i.e., H_2O , OH , and HCO_3^-) and bulky functional groups
10 (see Supporting Information). Therefore, the correlations observed in relationships of pK_a to $\Delta E'$
11 from through-bond electronic effects for **Sets 3** and **4** are absent in this subset, although all
12 complexes possess both moderate $\Delta E'$ and reduced pK_a for Zn- OH_2 (Supporting Information
13 Figure S11).

14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30 The five complexes in **Set 2** include rigid aromatic chelating nitrogen atoms from
31 macrocycles or bidentate ligands that yield the smallest variations in pK_a and $\Delta E'$ (Supporting
32 Information Figure S12). In contrast to **Sets 1** and **3**, which lie roughly along the **Set 4** trendline,
33 **Set 2** complexes are characterized by high Zn- H_2O pK_a but more favorable CHR energetics. The
34 most comparable **Set 3** complexes with sp^2 chelating N atoms also had lower $\Delta E'$ but not
35 alongside raised Zn- H_2O pK_a values, suggesting that the difference here is due to ligand rigidity.
36
37
38
39
40
41
42
43
44

45 Focusing on the **Set 3** and **4** cases where pK_a correlations to energetics are clear, it
46 becomes apparent that functionalizing classes of ligand structures can tune metal-ligand bond
47 strength in a specific manner due to changes in through-bond electronic structure. Here, complex
48 pK_a acts as a good proxy to identify strengthening metal-ligand axial interactions that stabilize
49 strongly binding hydroxyl moieties over water (i.e., **1** and **2**) that also serve to preferentially
50
51
52
53
54
55
56
57
58
59
60

1
2
3 stabilize hydroxyl moieties over bicarbonate (i.e., **4**). Across different classes of catalysts, i.e.,
4 cyclen vs. porphyrin vs. imidazole, this correlation no longer holds, with higher pK_a observed for
5
6 cyclen vs. porphyrin vs. imidazole, this correlation no longer holds, with higher pK_a observed for
7
8 some catalysts (i.e., porphyrin) with favorable energetics, likely due to greater differences in
9
10 through-space interactions in addition to through-bond interactions, thus reconciling these
11
12 observations with earlier studies on small-molecule Zn(II) catalysts^{16, 18} that indicated increasing
13
14 activity with increasing pK_a of the Zn-coordinated water. Thus, pK_a is a good proxy for through-
15
16 bond interaction strength tuning-based catalyst design where the differences in hydroxyl/water
17
18 binding are representative of the key component of a catalytic cycle. This scope at first appears
19
20 limited but could apply to a number of key reactions where water and more strongly-binding
21
22 oxygen containing species are relevant intermediates.
23
24
25
26

27 **3d. Rationalizing pK_a -energy Relationships in Imidazole Derivatives.**

28
29

30 In order to simplify and extract design rules of CHR catalysts, we applied feature
31
32 selection techniques (LASSO⁹⁴) to identify key descriptors that are well correlated to CHR
33
34 energetics on the **Set 4** imidazole ligands. If ligand-only properties can be identified to determine
35
36 whole-catalyst energetics, their calculation either with DFT or tabulation from heuristic models
37
38 is a dramatic simplification in comparison to direct predictions of inorganic complex
39
40 properties⁶⁰⁻⁶⁶. We considered properties of intermediates **2** and **4** including partial charges of
41
42 atoms in the first coordination sphere (i.e., Zn, N, and O), bond lengths, and the complex pK_a .
43
44 We also considered ligand-only properties such as pK_a , chelating N atom partial charges, and
45
46 topological descriptors for inorganic chemistry (Supporting Information Tables S6-S7).¹¹²
47
48
49
50

51 Feature selection confirms complex pK_a as the dominant contributor, although the single
52
53 descriptor model was improved upon in the LASSO selected features by incorporating the effect
54
55
56
57
58
59
60

1
2
3 that shorter Zn-N bonds in **4** or longer Zn-O bonds in **2** are associated with less favorable
4 energetics (Supporting Information Text S3). Importantly, although we noted through-bond
5 electronic effects from electron-donating or -withdrawing substituents to be important for
6 explaining the range of energetics and pK_a values sampled, partial charges from NBO are not a
7 uniquely good proxy for this effect because the variations are small or correlated strongly to the
8 LASSO-selected features (see Supporting Information Table S9 and Figure S13). For example,
9 increasingly negative oxygen partial charges in the bicarbonate intermediate **4** have a positive
10 correlation with lower complex pK_a values (Supporting Information Figure S13).
11
12
13
14
15
16
17
18
19
20
21

22 Although whole-complex descriptors were selected exclusively with LASSO, we return
23 to ligand-only features to identify if they are suitable for predicting catalyst energetics and
24 properties. We observe a good correlation ($R^2 = 0.73$) between the ligand and complex pK_a
25 values (Figure 5). Generally, a reduction of the computed ligand pK_a by 2 units decreases the
26 complex pK_a by 1 unit. Imidazole ligands that are functionalized with strong electron-
27 withdrawing groups adjacent to the metal-coordinating nitrogen atom generally have the lowest
28 ligand and complex pK_a values, whereas electron-donating groups, such as primary amines and
29 alkyl groups have increased ligand and complex pK_a values with respect to imidazole. That is,
30 electronic effects that stabilize the deprotonated isolated nitrogen atom in the ligand also
31 stabilize Zn-OH in the catalyst intermediate **2**. Deviations from the best-fit line for this
32 correlation can be rationalized by the effect of through-space interactions. For example,
33 functional groups with oxygen atoms (e.g., nitro, carboxylic acid, and hydroxyl groups) form
34 additional, covalent bonds with the Zn center. Other functional groups (e.g., cyano,
35 trifluoromethyl, and hydroxyl) form HBs in the catalyst that occur preferentially with either Zn-
36 OH or Zn-OH₂.
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

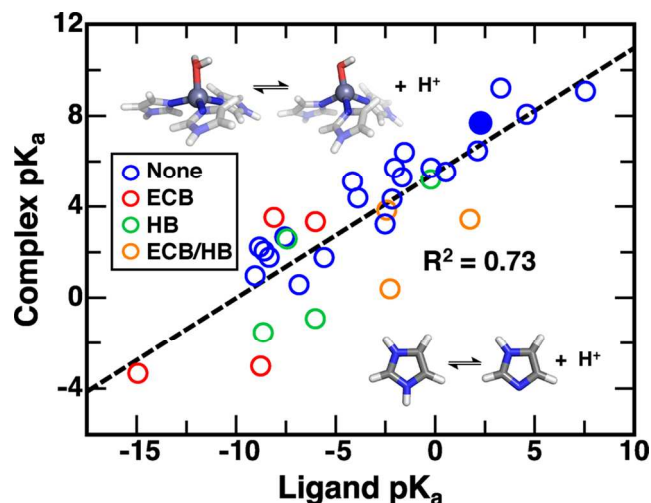


Figure 5. Complex pK_a versus ligand pK_a for **Set 4** with linear best-fit line (black, dashed) and correlation coefficient indicated in inset. Symbols are colored according to the presence of through-space interactions in the complex: ECB (extra covalent bonding around Zn, red), HB (hydrogen bonding around bound water or hydroxide, green), both types of interactions (ECB/HB orange), or no additional interactions (none, blue).

Returning to the fact that geometric quantities were selected in our LASSO model for predicting $\Delta E'$, we now consider the relationship between energetics and geometric properties of intermediates **2** and **4**. For ligands with reduced pK_a , we observe weak, long Zn-N bonds and concomitant strengthening and shortening of a Zn-O bond in intermediates **2** and **4** (Figure 6). Comparing intermediates **2** and **4**, the stronger Zn-OH bond in **2** has a smaller range and longer overall Zn-N bond lengths, as might be expected from the higher strength of the Zn-OH bond with respect to Zn-OCOOH. Reviewing relationships between ligand pK_a and bond lengths also reveals some cases where Zn-N bond lengths are short despite low ligand pK_a or long despite high ligand pK_a , especially in intermediate **4**, highlighting why feature selection identified Zn-N bond lengths to be informative (Figure 6). These geometric- pK_a relationship outliers are typically the same complexes that we identified to form additional non-covalent interactions with Zn (see Figure 5).

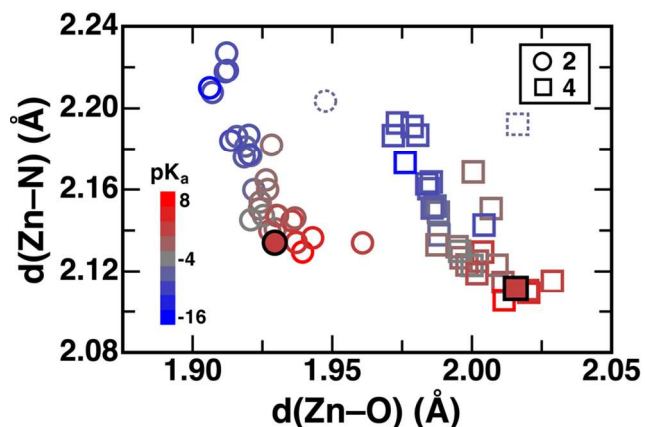


Figure 6. Zn–O bond distance ($d(\text{Zn–O})$, in Å) versus the average of three Zn–N bond distances ($d(\text{Zn–N})$, in Å) for intermediates **2** (circle) and **4** (square) for **Set 4** complexes. Symbols are colored according to the corresponding ligand pK_a , as indicated in inset. One outlier that has long Zn–O bonds due to hydrogen bonding with adjacent functional groups is indicated with a dashed symbol.

The geometric push-pull of shorter Zn–N bonds and longer Zn–O bonds holds for all intermediate structures (i.e., **0**, **1**, **2** and **4**) in the CHR reaction coordinate. If the energetic effect of any functionalization is comparable on all intermediates, a constant shift in energetics across all intermediates should produce no change in turnover frequency. However, the fact that we observe energetic differences suggests that relative binding strength of intermediates plays an important role. For instance, comparable structural changes between Zn and reactive moieties should correspond to smaller energetic changes for shallower potential wells (e.g., HCO_3^- or H_2O instead of OH^-). To confirm this idea, we computed an activation strain model (ASM)¹¹³⁻¹¹⁴-inspired interaction energy (E_{int}) for the Zn–O and Zn–N bonds in **2** and **4** by rigidly removing the reactive moiety alone or Zn with the reactive moiety, respectively (see Supporting Information Text S4 and Table S10). The E_{int} then corresponds to these respective rigid binding energies. As ligand pK_a decreases, $E_{\text{int}}(\text{Zn–N})$ for both intermediates increases. Ligand pK_a strongly influences $E_{\text{int}}(\text{Zn–N})$: a 2 kcal/mol increase occurs with a unit reduction in pK_a . However, the effect of reduced pK_a on $E_{\text{int}}(\text{Zn–N})$ of each intermediate is nearly equal,

1
2
3 producing a negligible influence on the CHR energetic span (see Supporting Information Figure
4 S14). In contrast, $E_{\text{int}}(\text{Zn-O})$ in both intermediates decreases as ligand pK_a decreases, although
5
6 the effect with a unit reduction in pK_a is smaller: 0.6 and 0.3 kcal/mol for **2** and **4**, respectively.
7
8 Because the effect of functionalization on each intermediate differs (i.e., $E_{\text{int}}(\text{Zn-O})$ for **2**
9 decreases 2x as fast as for **4**), relative energetics are tuned (Figure 7). Here, weakly bound
10 bicarbonate is less influenced by ligand functionalization than strongly bound hydroxyl, and
11 strengthening overall binding therefore reduces $\Delta E'$ and complex pK_a values. Although these
12 observations focus on the imidazole ligands of **Set 4** alone, the large 7-8 kcal/mol range of CHR
13 energetics we observe over all ligand sets can be traced back to electronic effects that strengthen
14 the Zn-O bond by weakening the Zn-N bond in a manner that selectively stabilizes the least
15 stable reactive intermediates in the cycle. Comparison of trends across **Set 3** confirms similar
16 observations, albeit with the complication of variable chelating nitrogen ligand pK_a for
17 asymmetric ligands (Supporting Information Figure S15). Thus, overall, reduced ligand pK_a (i.e.,
18 less nucleophilic metal-coordinating N) produces a more nucleophilic Zn-OH for attack on CO_2
19 by increasing the Zn-O bond strength and weakening Zn-N bonding.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

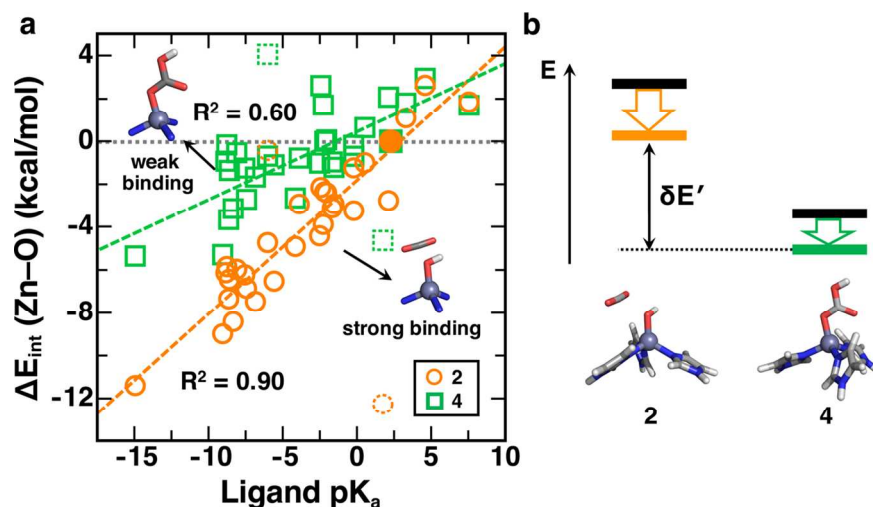


Figure 7. (a) Zn-O bond ASM interaction energies ($\Delta E_{\text{int}}(\text{Zn-O})$, in kcal/mol) relative to an imidazole reference (filled symbols) as a function of ligand pK_a for intermediates **2** (orange circles) and **4** (green squares) for **Set 4** complexes. The dashed lines indicate best-fit lines with correlation coefficients as shown. Two outliers excluded from the fit are distinguished by a dotted symbol outline. (b) Qualitative illustration of how decreasing the imidazole ligand pK_a tunes the relative energetics of **2** and **4**, reducing $\delta E'$.

3e. Dependence of Catalyst Design Principles on Ligand Class.

For the functionalized imidazole ligands in tetrahedral geometries in **Set 4**, we have arrived at a unified picture that weakening the Zn-N bond strengthens interactions with the axial moieties. This axial bond strengthening has a more significant effect on the Zn-OH intermediate **2** than less strongly bound bicarbonate (**4**) or water (**0**) intermediates, leading to overall improved CHR energetics. Although the **Set 4** class of functionalized imidazoles have revealed compelling relationships between the pK_a of the isolated ligand and favorable CHR reaction energetics, some differences were observed across the distinct ligand classes. Now, we return to the other ligand sets to identify if similar geometric and energetic relationships hold across the other ligand sets.

1
2
3 Overall, a clear relationship between shorter Zn-OH bonds in intermediate **2** with longer
4 average Zn-N bonds is apparent for all sets except for the **Set 1** cyclen derivatives (Figure 8 and
5 Supporting Information Tables S7 and S11-S13). Standard cyclen samples a longer average Zn-
6 N bond lengths versus Zn(imi)₃ or the **Set 2** porphyrin reference. Within the **Set 1** cyclen
7 derivatives, a wide range of average Zn-N bond lengths is sampled, but the Zn-N bond
8 elongation arises from steric interactions due to bulky functional groups that also will prevent
9 shortening of the Zn-O bond. The near constant Zn-O bonds in **Set 1** are observed for both Zn-
10 OH and Zn-OCOOH intermediates (Figure 8 and Supporting Information Figure S16).
11 Therefore, the earlier modest variations in pK_a and $\Delta E'$ range can indeed be attributed to varied
12 through-space hydrogen bonding interactions with intermediates and not a significant change of
13 the electronic structure around Zn (see Section 3c). Despite the lack of variation across the **Set 1**
14 group, it is apparent that the standard cyclen affords an already long Zn-N bond length compared
15 to other four-coordinate Zn-N catalyst ligand sets and modest pK_a, consistent with the
16 observation that cyclen already has favorable $\Delta E'$ and pK_a values not likely amenable to further
17 enhancement.
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

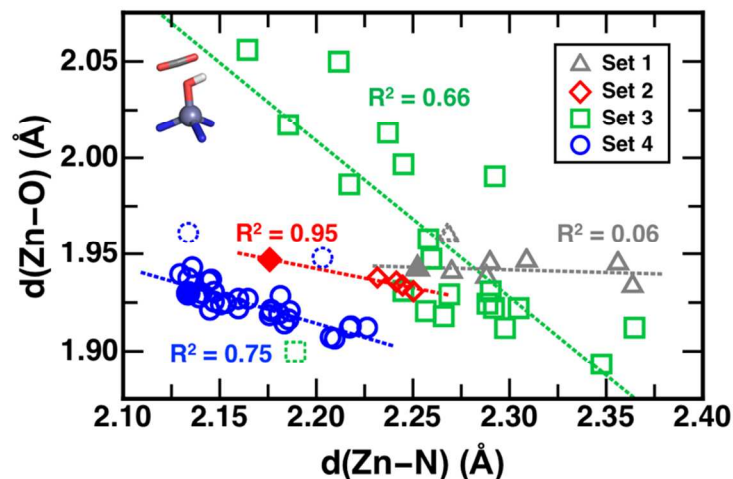


Figure 8. Intermediate **2** average Zn–N bond distances ($d(\text{Zn–N})$, in Å) versus Zn–O bond distance ($d(\text{Zn–O})$, in Å). The **Set 1** (gray triangles), **Set 2** (red diamonds), **Set 3** (green squares), and **Set 4** (blue circles) complexes are shown. Trend lines and correlation coefficients are indicated in inset with outliers due to additional hydrogen bonding or bond cleavage excluded from the trendline, as indicated by dotted symbols, here: two for **Set 1** and **Set 4** and one from **Set 3**.

In the three other ligand sets, there is a clear relationship is apparent between functional group choice and tuning of Zn–N and Zn–O bonds. In the **Set 4** imidazoles, each unit of Zn–N bond elongation reduces the proportionately Zn–O bond in both intermediate **2** and **4**. This effect is not due solely to the flexibility of the three-fold coordination environment in the **Set 4** ligands: fourfold-coordinated **Set 3** ligands in fact span the widest average Zn–N distance and show the strongest inverse relationship to Zn–O in intermediate **2** (Figure 8). Therefore, geometrically, four-fold coordinated bidentates in **Set 3** actually show greater range of electronically driven effects apparent in geometric shifts than the imidazoles, and dependence on N atom chelating character (i.e., sp^2 vs sp^3) is not substantial (Supporting Information Table S13).

Overall comparison of the Zn–O bond in intermediates **2** and **4** confirms a consistent relationship across all ligand sets observed previously for **Set 4** alone. The longer Zn–OCOOH[−] bond is tuned in a roughly 1:1 ratio with the shorter Zn–OH bond, but differences arise because

1
2
3 these consistent geometric changes correspond to differing energetic effects (Supporting
4 Information Figure S17). Outliers in this relationship for **Set 3** are apparent though, with
5 occasionally more comparable Zn-O bond lengths in both intermediates or unusually long Zn-
6 OCOOH- bonds, likely rationalizing the large variation in $\Delta E'$ and pK_a (see Sec. 3c and
7 Supporting Information Figure S17).
8
9

10
11
12
13
14
15 The small set of aromatic ligands in **Set 2** exhibits Zn-N bond lengths intermediate
16 between the long Zn-N bonds in cyclen derivatives and short Zn-N bonds in imidazoles (Figure
17 8). Therefore, a remaining question is why, given the geometric and energetic consistency of **Set**
18 **2** with the other ligand types, is the pK_a range observed in these compounds distinct from all
19 other structures? Computing ASM-inspired relative $E_{int}(Zn-O)$ of Zn-OH vs. Zn-H₂O moieties
20 reveals reduced stabilization of Zn-OH in **Set 2** ligands in comparison to cyclen or imidazole
21 references (Supporting Information Figure S18). Conversely, the relative interaction strength of
22 intermediates **2** and **4** (i.e., OH vs. bicarbonate) remains comparable or strengthened (Supporting
23 Information Figure S18). Returning to the relationship observed for **Set 4** ligands between ligand
24 pK_a and $\Delta E'$, we identify that for **Set 2**, ligand pK_a s are low and in the range that would be
25 expected based on the favorable $\Delta E'$ for these ligands even though complex pK_a s are high
26 (Supporting Information Table S14).
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43

44 We can identify the source of the anomalously high Zn-OH₂ complex pK_a by
45 decomposing the contributions to the complex pK_a in terms of i) 0 K electronic energy
46 differences, ii) enthalpic and entropic differences, and iii) solvation free energy differences
47 (Supporting Information Table S15). Doing so, we determine that high metal-bound pK_a arises in
48 **Set 2** ligands from differences in the 0 K electronic energies and secondarily from solvation free
49 energy differences. From this analysis, we can conclude that charge delocalization throughout
50
51
52
53
54
55
56
57
58
59
60

1
2
3 the rigid, aromatic macrocycles in **Set 2** stabilizes H₂O intermediates with respect to OH⁻ without
4 also stabilizing the HCO₃⁻. We computed Mayer bond orders of representative **Set 1** and **Set 4**
5
6 complexes to compare to our **Set 2** ligands to identify any differences in the bonding between
7
8 intermediates that correlate to pK_a and intermediates that correlate to the turnover-determining
9
10 step. This analysis reveals that **Set 2** ligands preferentially form a nearly bidentate structure
11
12 (average Mayer bond orders: 0.38 and 0.18) in intermediate **4**, whereas the other complexes
13
14 prefer a nearly monodentate bicarbonate (Supporting Information Table S16). The result is a
15
16 higher overall bond order for the Zn-O interactions in intermediate **4** than for the weak bonding
17
18 in Zn-OH₂ complexes (average Mayer bond order: 0.28) for **Set 2** ligands than the other ligand
19
20 classes. Thus, Zn-moiety interactions in **Set 2** that arise due to geometric differences of the rigid,
21
22 planar ligands can produce favorable reaction energetics for CO₂ hydration without reducing the
23
24 complex pK_a.
25
26
27
28
29
30

31 **4. Conclusions**

32
33 We have carried out a detailed study of nearly 70 Zn molecular catalysts for CO₂
34
35 hydration from four diverse ligand classes ranging from well-studied carbonic anhydrase mimics
36
37 (e.g., cyclen) to new structures obtained by leveraging diverse hits from large organic libraries.
38
39 We determined the turnover-determining transition state and intermediates to correspond to CO₂
40
41 addition and bicarbonate release, respectively. We observed these turnover-determining species,
42
43 not necessarily connected by a single rate-determining step, to be preserved over significant
44
45 energetic tuning through catalyst functionalization. To streamline our screen, we thus focused on
46
47 relative energies between likely turnover-determining intermediates (i.e., hydroxyl vs.
48
49 bicarbonate energies) that were related to key transition states through a good linear free energy
50
51 relationship.
52
53
54
55
56
57
58
59
60

1
2
3 Over our molecular catalyst test set, we observed a wide range of relative thermodynamic
4 stabilities of hydroxyl and bicarbonate intermediates, with functionalized ligands demonstrating
5 up to a 6 kcal/mol relative stabilization of the reactive hydroxyl moiety over well known cyclen
6 or imidazole references. Much like in HCAII, across three of the four ligand sets, we observe a
7 good correlation between the pK_a of the water molecule bound to the Zn catalyst and the
8 resulting relative stability of the hydroxyl moiety over bicarbonate. The same bonding
9 relationships that stabilize hydroxyl over water to lower the complex pK_a are therefore similar to
10 the needed stabilization of hydroxyl over bicarbonate. In agreement with prior studies, we
11 observed that porphyrins and other aromatic bidentate ligands represent an exception to the rule
12 we observe over other more flexible ligand structures due to differences in electronic structure
13 for this ligand class. We showed that for the imidazole cases where individual ligand pK_a s may
14 be unambiguously calculated, a good relationship is found between the isolated ligand pK_a and
15 the pK_a of a bound water molecule on the catalyst. Thus, organic ligand pK_a , which is intuitive,
16 easy to compute or tabulate, and much less sensitive to method choice than whole-catalyst
17 properties, is a good descriptor for predicting relative CO_2 hydration energetics. This correlation
18 is likely observed because pK_a tuning occurs through through-bond electronic effects that also
19 correspond to disproportionate stabilization of hydroxyl moieties over bicarbonate.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41

42 Overall, our results suggest ligand pK_a as a promising descriptor for quantitative
43 estimation of tuning catalyst energetics with through-bond electronic effects in ligands. We
44 expect these results obtained for CO_2 hydration to be transferable to other reactions where the
45 turnover-determining transition state or intermediate involves a hydroxyl species and the
46 remaining species has a weaker metal-moiety bond more comparable to water (e.g., in water
47 splitting or hydroxylation reactions). The strategy outlined here to i) curate diverse ligands from
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 large organic libraries and ii) identify when ligand-only properties can determine catalyst
4 energetics is expected to be broadly useful for both experimental and computational catalyst
5 design.
6
7
8
9

10 11 12 ASSOCIATED CONTENT 13

14
15 **Supporting Information.** CHR energetics with different exchange-correlation functional and
16 basis set choices; CHR energy profile with/without DFT-D3 correction; thermal correction from
17 translational, rotational, and vibrational motion; details and chemdraws of ligand library
18 screening set (for **Set 3** and **4**); comparison of ΔG^\ddagger for two bicarbonate rearrangement pathways;
19 degree of TOF control in CHR reaction profile; details of $\Delta E'$ and complex pK_a values; subgroup
20 $\Delta E'$ -complex pK_a correlations; details of descriptor set, model and correlation analysis; bond
21 interaction energies for intermediate **2** and **4** (PDF)
22
23
24
25
26
27
28
29

30
31
32
33 Structures of intermediates (**0**, **1**, **2** and **4**) (ZIP)
34
35

36 This material is available free of charge via the Internet at <http://pubs.acs.org>.
37
38

39 40 AUTHOR INFORMATION 41

42 43 **Corresponding Author** 44

45 *email: hjkulik@mit.edu
46
47

48 49 **Notes** 50

51 The authors declare no competing financial interest.
52
53
54

55 56 ACKNOWLEDGMENT 57

1
2
3 This work was supported in part by the Office of Naval Research under grant number N00014-
4 17-1-2956, the National Science Foundation under grant numbers CBET-1704266 and ECCS-
5 1449291, and MIT Energy Initiative seed grants (2014, 2017). H.J.K. holds a Career Award at
6 the Scientific Interface from the Burroughs Wellcome Fund. This work was carried out in part
7 using computational resources from the Extreme Science and Engineering Discovery
8 Environment (XSEDE), which is supported by National Science Foundation grant number ACI-
9 1548562. This work used the XStream computational resource, supported by the National
10 Science Foundation Major Research Instrumentation program (ACI-1429830). The authors thank
11 Adam H. Steeves for providing a critical reading of the manuscript.
12
13
14
15
16
17
18
19
20
21
22
23

24 REFERENCES

- 25 1. Thiel, W., Computational Catalysis—Past, Present, and Future. *Angew. Chem., Int. Ed.* **2014**, *53*, 8605-8613.
- 26 2. Raugei, S.; DuBois, D. L.; Rousseau, R.; Chen, S.; Ho, M.-H.; Bullock, R. M.; Dupuis, M., Toward Molecular Catalysts by Computer. *Acc. Chem. Res.* **2015**, *48*, 248-255.
- 27 3. Nørskov, J. K.; Bligaard, T., The Catalyst Genome. *Angew. Chem., Int. Ed.* **2013**, *52*, 776-777.
- 28 4. Greeley, J.; Jaramillo, T. F.; Bonde, J.; Chorkendorff, I.; Nørskov, J. K., Computational High-Throughput Screening of Electrocatalytic Materials for Hydrogen Evolution. *Nat. Mater.* **2006**, *5*, 909-913.
- 29 5. Nørskov, J. K.; Bligaard, T.; Rossmeisl, J.; Christensen, C. H., Towards the Computational Design of Solid Catalysts. *Nat. Chem.* **2009**, *1*, 37-46.
- 30 6. Vilé, G.; Albani, D.; Almora Barrios, N.; López, N.; Pérez Ramírez, J., Advances in the Design of Nanostructured Catalysts for Selective Hydrogenation. *ChemCatChem* **2016**, *8*, 21-33.
- 31 7. Bridier, B.; López, N.; Pérez-Ramírez, J., Molecular Understanding of Alkyne Hydrogenation for the Design of Selective Catalysts. *Dalton Trans.* **2010**, *39*, 8412-8419.
- 32 8. Caruthers, J.; Lauterbach, J. A.; Thomson, K.; Venkatasubramanian, V.; Snively, C.; Bhan, A.; Katare, S.; Oskarsdottir, G., Catalyst Design: Knowledge Extraction from High-Throughput Experimentation. *J. Catal.* **2003**, *216*, 98-109.
- 33 9. Katare, S.; Caruthers, J. M.; Delgass, W. N.; Venkatasubramanian, V., An Intelligent System for Reaction Kinetic Modeling and Catalyst Design. *Ind. Eng. Chem. Res.* **2004**, *43*, 3484-3512.
- 34 10. Corma, A.; Díaz-Cabanas, M. J.; Moliner, M.; Martínez, C., Discovery of a New Catalytically Active and Selective Zeolite (Itq-30) by High-Throughput Synthesis Techniques. *J. Catal.* **2006**, *241*, 312-318.

11. Khalifah, R. G., The Carbon Dioxide Hydration Activity of Carbonic Anhydrase. I. Stop-Flow Kinetic Studies on the Native Human Isoenzymes B and C. *J. Biol. Chem.* **1971**, *246*, 2561-73.
12. Brown, R.; Salmon, D.; Curtis, N.; Kusuma, S., Carbonic Anhydrase Models. 4.[Tris [(4, 5-Dimethyl-2-Imidazolyl) Methyl] Phosphine Oxide] Cobalt (2+), a Small-Molecule Mimic of the Spectroscopic Properties of Cobalt (II) Carbonic Anhydrase. *J. Am. Chem. Soc.* **1982**, *104*, 3188-3194.
13. Brown, R.; Curtis, N.; Huguet, J., Tris (4, 5-Diisopropylimidazol-2-Yl) Phosphine: Zinc (2+). A Catalytically Active Model for Carbonic Anhydrase. *J. Am. Chem. Soc.* **1981**, *103*, 6953-6959.
14. Slebocka-Tilk, H.; Cocho, J.; Frackman, Z.; Brown, R., Carbonic Anhydrase Models. 5. Tris (4, 5-Di-N-Propyl-2-Imidazolyl) Phosphine-Zinc (2+) and Bis (4, 5-Di-Isopropyl-2-Imidazolyl)-2-Imidazolylphosphine-Zinc (2+). Catalysts Facilitating Hydrogen Carbonate Carbon Dioxide Interconversion. *J. Am. Chem. Soc.* **1984**, *106*, 2421-2431.
15. Parkin, G., Synthetic Analogues Relevant to the Structure and Function of Zinc Enzymes. *Chem. Rev.* **2004**, *104*, 699-767.
16. Zhang, X. P.; Vaneldik, R., A Functional-Model for Carbonic-Anhydrase - Thermodynamic and Kinetic-Study of a Tetraazacyclododecane Complex of Zinc(II). *Inorg. Chem.* **1995**, *34*, 5606-5614.
17. Zhang, X. P.; Vaneldik, R.; Koike, T.; Kimura, E., Kinetics and Mechanism of the Hydration of Co₂ and Dehydration of Hco₃⁻ Catalyzed by a Zn(II) Complex of 1,5,9-Triazacyclododecane as a Model for Carbonic-Anhydrase. *Inorg. Chem.* **1993**, *32*, 5749-5755.
18. Koziol, L.; Valdez, C. A.; Baker, S. E.; Lau, E. Y.; Floyd, W. C.; Wong, S. E.; Satcher, J. H.; Lightstone, F. C.; Aines, R. D., Toward a Small Molecule, Biomimetic Carbonic Anhydrase Model: Theoretical and Experimental Investigations of a Panel of Zinc(II) Aza-Macrocyclic Catalysts. *Inorg. Chem.* **2012**, *51*, 6803-6812.
19. Silverman, D. N.; Lindskog, S., The Catalytic Mechanism of Carbonic-Anhydrase - Implications of a Rate-Limiting Protolysis of Water. *Acc. Chem. Res.* **1988**, *21*, 30-36.
20. Ma, R. H.; Schuette, G. F.; Broadbelt, L. J., Microkinetic Modeling of CO₂ Hydrolysis over Zn-(1,4,7, 10-Tetraazacyclododecane) Catalyst Based on First Principles: Revelation of Rate-Determining Step. *J. Catal.* **2014**, *317*, 176-184.
21. Satcher Jr., J. H.; Baker, S. E.; Kulik, H. J.; Valdez, C. A.; Krueger, R. L.; Lightstone, F. C.; Aines, R. D., Modeling, Synthesis and Characterization of Zinc Containing Carbonic Anhydrase Active Site Mimics. *Energy Procedia* **2011**, *4*, 2090-2095.
22. Looney, A.; Saleh, A.; Zhang, Y.; Parkin, G., Tris (Pyrazolyl) Hydroborato Complexes of Cadmium: A Bidentate Nitrate Derivative and Its Relevance to Carbonic Anhydrase Activity. *Inorg. Chem.* **1994**, *33*, 1158-1164.
23. Wong, S. E.; Lau, E. Y.; Kulik, H. J.; Jr., J. H. S.; Valdez, C. A.; Worsely, M.; Lightstone, F. C.; Aines, R. D., Designing Small-Molecule Catalysts for CO₂ Capture. *Energy Procedia* **2011**, *4*, 817-823.
24. Lau, E. Y.; Wong, S. E.; Baker, S. E.; Beringer, J. P.; Koziol, L.; Valdez, C. A.; Satcher, J. H.; Aines, R. D.; Lightstone, F. C., Comparison and Analysis of Zinc and Cobalt-Based Systems as Catalytic Entities for the Hydration of Carbon Dioxide. *PLoS ONE* **2013**, *8*.
25. Kulik, H. J.; Wong, S. E.; Baker, S. E.; Valdez, C. A.; Satcher, J.; Aines, R. D.; Lightstone, F. C., Developing an Approach for First-Principles Catalyst Design: Application to Carbon-Capture Catalysis. *Acta Crystallogr., Sect. C: Struct. Chem.* **2014**, *70*, 123-131.

- 1
2
3 26. Mauksch, M.; Brauer, M.; Weston, J.; Anders, E., New Insights into the Mechanistic
4 Details of the Carbonic Anhydrase Cycle as Derived from the Model System
5 [(NH₃)₃Zn(OH)](+)/CO₂: How Does the H₂O/Hco(3)(-)Replacement Step Occur?
6 *ChemBioChem* **2001**, *2*, 190-198.
7
8 27. Ma, R.; Schuette, G. F.; Broadbelt, L. J., Insights into the Relationship of Catalytic
9 Activity and Structure: A Comparison Study of Three Carbonic Anhydrase Mimics. *Int. J. Chem.*
10 *Kinet.* **2014**, *46*, 683-700.
11
12 28. Ma, R.; Schuette, G. F.; Broadbelt, L. J., Toward Understanding the Activity of Cobalt
13 Carbonic Anhydrase: A Comparative Study of Zinc-and Cobalt-Cyclen. *Appl. Catal., A* **2015**,
14 *492*, 151-159.
15
16 29. Merz, K. M.; Banci, L., Binding of Bicarbonate to Human Carbonic Anhydrase II: A
17 Continuum of Binding States. *J. Am. Chem. Soc.* **1997**, *119*, 863-871.
18
19 30. Piazzetta, P.; Marino, T.; Russo, N.; Salahub, D., Direct Hydrogenation of Carbon
20 Dioxide by an Artificial Reductase Obtained by Substituting Rhodium for Zinc in the Carbonic
21 Anhydrase Catalytic Center. A Mechanistic Study. *ACS Catal.* **2015**, *5*, 5397-5409.
22
23 31. Bottoni, A.; Lanza, C. Z.; Miscione, G. P.; Spinelli, D., New Model for a Theoretical
24 Density Functional Theory Investigation of the Mechanism of the Carbonic Anhydrase: How
25 Does the Internal Bicarbonate Rearrangement Occur? *J. Am. Chem. Soc.* **2004**, *126*, 1542-1550.
26
27 32. Miscione, G. P.; Stenta, M.; Spinelli, D.; Anders, E.; Bottoni, A., New Computational
28 Evidence for the Catalytic Mechanism of Carbonic Anhydrase. *Theor. Chem. Acc.* **2007**, *118*,
29 193-201.
30
31 33. Jing, Q.; Okrasa, K.; Kazlauskas, R. J., Stereoselective Hydrogenation of Olefins Using
32 Rhodium Substituted Carbonic Anhydrase—a New Reductase. *Chem. - Eur. J.* **2009**, *15*, 1370-
33 1376.
34
35 34. Marino, T.; Russo, N.; Toscano, M., A Comparative Study of the Catalytic Mechanisms
36 of the Zinc and Cadmium Containing Carbonic Anhydrase. *J. Am. Chem. Soc.* **2005**, *127*, 4242-
37 4253.
38
39 35. Piazzetta, P.; Marino, T.; Russo, N.; Salahub, D. R., Explicit Water Molecules Play a Key
40 Role in the Mechanism of Rhodium Substituted Human Carbonic Anhydrase. *ChemCatChem*
41 **2017**, *9*, 1047-1053.
42
43 36. Abild-Pedersen, F.; Greeley, J.; Studt, F.; Rossmeisl, J.; Munter, T.; Moses, P. G.;
44 Skulason, E.; Bligaard, T.; Nørskov, J. K., Scaling Properties of Adsorption Energies for
45 Hydrogen-Containing Molecules on Transition-Metal Surfaces. *Phys. Rev. Lett.* **2007**, *99*,
46 016105.
47
48 37. Montemore, M. M.; Medlin, J. W., Scaling Relations between Adsorption Energies for
49 Computational Screening and Design of Catalysts. *Catal. Sci. Technol.* **2014**, *4*, 3748-3761.
50
51 38. Greeley, J., Theoretical Heterogeneous Catalysis: Scaling Relationships and
52 Computational Catalyst Design. *Annu. Rev. Chem. Biomol. Eng.* **2016**, *7*, 605-635.
53
54 39. Bligaard, T.; Nørskov, J. K.; Dahl, S.; Matthiesen, J.; Christensen, C. H.; Sehested, J.,
55 The Brønsted–Evans–Polanyi Relation and the Volcano Curve in Heterogeneous Catalysis. *J.*
56 *Catal.* **2004**, *224*, 206-217.
57
58 40. van Santen, R. A.; Neurock, M.; Shetty, S. G., Reactivity Theory of Transition-Metal
59 Surfaces: A Brønsted- Evans- Polanyi Linear Activation Energy- Free-Energy Analysis. *Chem.*
60 *Rev.* **2009**, *110*, 2005-2048.

- 1
2
3 41. Man, I. C.; Su, H. Y.; Calle-Vallejo, F.; Hansen, H. A.; Martínez, J. I.; Inoglu, N. G.;
4 Kitchin, J.; Jaramillo, T. F.; Nørskov, J. K.; Rossmeisl, J., Universality in Oxygen Evolution
5 Electrocatalysis on Oxide Surfaces. *ChemCatChem* **2011**, *3*, 1159-1165.
6
7 42. Peterson, A. A.; Nørskov, J. K., Activity Descriptors for CO₂ Electroreduction to
8 Methane on Transition-Metal Catalysts. *J. Phys. Chem. Lett.* **2012**, *3*, 251-258.
9
10 43. Latimer, A. A.; Kulkarni, A. R.; Aljama, H.; Montoya, J. H.; Yoo, J. S.; Tsai, C.; Abild-
11 Pedersen, F.; Studt, F.; Nørskov, J. K., Understanding Trends in C-H Bond Activation in
12 Heterogeneous Catalysis. *Nat. Mater.* **2017**, *16*, 225-229.
13
14 44. Baran, J. D.; Grönbeck, H.; Hellman, A., Analysis of Porphyrines as Catalysts for
15 Electrochemical Reduction of O₂ and Oxidation of H₂O. *J. Am. Chem. Soc.* **2014**, *136*, 1320-
16 1326.
17
18 45. Pegis, M. L.; McKeown, B. A.; Kumar, N.; Lang, K.; Wasylenko, D. J.; Zhang, X. P.;
19 Raugei, S.; Mayer, J. M., Homogenous Electrocatalytic Oxygen Reduction Rates Correlate with
20 Reaction Overpotential in Acidic Organic Solutions. *ACS Cent. Sci.* **2016**, *2*, 850-856.
21
22 46. Calle-Vallejo, F.; Krabbe, A.; García-Lastra, J. M., How Covalence Breaks Adsorption-
23 Energy Scaling Relations and Solvation Restores Them. *Chem. Sci.* **2017**, *8*, 124-130.
24
25 47. Pegis, M. L.; Wise, C. F.; Koronkiewicz, B.; Mayer, J. M., Identifying and Breaking
26 Scaling Relations in Molecular Catalysis of Electrochemical Reactions. *J. Am. Chem. Soc.* **2017**,
27 *139*, 11000-11003.
28
29 48. Nørskov, J. K.; Rossmeisl, J.; Logadottir, A.; Lindqvist, L.; Kitchin, J. R.; Bligaard, T.;
30 Jonsson, H., Origin of the Overpotential for Oxygen Reduction at a Fuel-Cell Cathode. *J. Phys.*
31 *Chem. B* **2004**, *108*, 17886-17892.
32
33 49. Rossmeisl, J.; Dimitrievski, K.; Siegbahn, P.; Nørskov, J. K., Comparing Electrochemical
34 and Biological Water Splitting. *J. Phys. Chem. C* **2007**, *111*, 18821-18823.
35
36 50. Gani, T. Z. H.; Kulik, H. J., Understanding and Breaking Scaling Relations in Single-Site
37 Catalysis: Methane-to-Methanol Conversion by Fe(IV)=O. *ACS Catal.* **2018**, *8*, 975-986.
38
39 51. Borovik, A. S., Role of Metal-Oxo Complexes in the Cleavage of C-H Bonds. *Chem.*
40 *Soc. Rev.* **2011**, *40*, 1870-1874.
41
42 52. Chen, S.; Ho, M.-H.; Bullock, R. M.; DuBois, D. L.; Dupuis, M.; Rousseau, R.; Raugei,
43 S., Computing Free Energy Landscapes: Application to Ni-Based Electrocatalysts with Pendant
44 Amines for H₂ Production and Oxidation. *ACS Catal.* **2013**, *4*, 229-242.
45
46 53. Mondal, B.; Neese, F.; Ye, S., Control in the Rate-Determining Step Provides a
47 Promising Strategy to Develop New Catalysts for CO₂ Hydrogenation: A Local Pair Natural
48 Orbital Coupled Cluster Theory Study. *Inorg. Chem.* **2015**, *54*, 7192-7198.
49
50 54. Mondal, B.; Neese, F.; Ye, S., Toward Rational Design of 3d Transition Metal Catalysts
51 for CO₂ Hydrogenation Based on Insights into Hydricity-Controlled Rate-Determining Steps.
52 *Inorg. Chem.* **2016**, *55*, 5438-5444.
53
54 55. Waldie, K. M.; Ostericher, A. L.; Reineke, M. H.; Sasayama, A. F.; Kubiak, C. P.,
55 Hydricity of Transition-Metal Hydrides: Thermodynamic Considerations for CO₂ Reduction.
56 *ACS Catal.* **2017**, 1313-1324.
57
58 56. Tsay, C.; Livesay, B. N.; Ruelas, S.; Yang, J. Y., Solvation Effects on Transition Metal
59 Hydricity. *J. Am. Chem. Soc.* **2015**, *137*, 14114-14121.
60
61 57. Mavros, M. G.; Tsuchimochi, T.; Kowalczyk, T.; McIsaac, A.; Wang, L.-P.; Van
62 Voorhis, T., What Can Density Functional Theory Tell Us About Artificial Catalytic Water
63 Splitting? *Inorg. Chem.* **2014**, *53*, 6386-6397.

- 1
2
3 58. Zagal, J. H.; Koper, M., Reactivity Descriptors for the Activity of Molecular Mn⁴
4 Catalysts for the Oxygen Reduction Reaction. *Angew. Chem., Int. Ed.* **2016**, *55*, 14510-14521.
- 5 59. Janet, J. P.; Gani, T. Z. H.; Steeves, A. H.; Ioannidis, E. I.; Kulik, H. J., Leveraging
6 Cheminformatics Strategies for Inorganic Discovery: Application to Redox Potential Design.
7 *Ind. Eng. Chem. Res.* **2017**, *56*, 4898-4910.
- 8 60. Phan, H.; Hrudka, J. J.; Igimbayeva, D.; Lawson Daku, L. v. M.; Shatruk, M., A Simple
9 Approach for Predicting the Spin State of Homoleptic Fe (II) Tris-Diimine Complexes. *J. Am.*
10 *Chem. Soc.* **2017**, *139*, 6437-6447.
- 11 61. Rodríguez-Jiménez, S.; Yang, M.; Stewart, I.; Garden, A. L.; Brooker, S., A Simple
12 Method of Predicting Spin State in Solution. *J. Am. Chem. Soc.* **2017**, *139*, 18392-18396.
- 13 62. Miller, J. J.; Sigman, M. S., Quantitatively Correlating the Effect of Ligand-Substituent
14 Size in Asymmetric Catalysis Using Linear Free Energy Relationships. *Angew. Chem., Int. Ed.*
15 **2008**, *120*, 783-786.
- 16 63. Harper, K. C.; Sigman, M. S., Using Physical Organic Parameters to Correlate
17 Asymmetric Catalyst Performance. *J. Org. Chem.* **2013**, *78*, 2813-2818.
- 18 64. Niemeyer, Z. L.; Milo, A.; Hickey, D. P.; Sigman, M. S., Parameterization of Phosphine
19 Ligands Reveals Mechanistic Pathways and Predicts Reaction Outcomes. *Nat. Chem.* **2016**, *8*,
20 610.
- 21 65. Santiago, C. B.; Milo, A.; Sigman, M. S., Developing a Modern Approach to Account for
22 Steric Effects in Hammett-Type Correlations. *J. Am. Chem. Soc.* **2016**, *138*, 13424-13430.
- 23 66. Sigman, M. S.; Harper, K. C.; Bess, E. N.; Milo, A., The Development of
24 Multidimensional Analysis Tools for Asymmetric Catalysis and Beyond. *Acc. Chem. Res.* **2016**,
25 *49*, 1292-1301.
- 26 67. Santiago, C. B.; Guo, J.-Y.; Sigman, M. S., Predictive and Mechanistic Multivariate
27 Linear Regression Models for Reaction Development. *Chem. Sci.* **2018**.
- 28 68. Shao, Y. H., et al., Advances in Molecular Quantum Chemistry Contained in the Q-Chem
29 4 Program Package. *Mol. Phys.* **2015**, *113*, 184-215.
- 30 69. Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti Correlation-Energy
31 Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785-789.
- 32 70. Becke, A. D., Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J.*
33 *Chem. Phys.* **1993**, *98*, 5648-5652.
- 34 71. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J., Ab Initio Calculation of
35 Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields.
36 *J. Phys. Chem.* **1994**, *98*, 11623-11627.
- 37 72. Vosko, S. H.; Wilk, L.; Nusair, M., Accurate Spin-Dependent Electron Liquid
38 Correlation Energies for Local Spin Density Calculations: A Critical Analysis. *Can J. Phys.*
39 **1980**, *58*, 1200-1211.
- 40 73. Hay, P. J.; Wadt, W. R., Abinitio Effective Core Potentials for Molecular Calculations -
41 Potentials for the Transition-Metal Atoms Sc to Hg. *J. Chem. Phys.* **1985**, *82*, 270-283.
- 42 74. Weigend, F.; Ahlrichs, R., Balanced Basis Sets of Split Valence, Triple Zeta Valence and
43 Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem.*
44 *Chem. Phys.* **2005**, *7*, 3297-3305.
- 45 75. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., A Consistent and Accurate Ab Initio
46 Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu.
47 *J. Chem. Phys.* **2010**, *132*.
- 48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 76. Klamt, A.; Schuurmann, G., Cosmo - a New Approach to Dielectric Screening in
4 Solvents with Explicit Expressions for the Screening Energy and Its Gradient. *J. Chem. Soc.,*
5 *Perkin Trans. 2* **1993**, 799-805.
- 6 77. Bondi, A., Van Der Waals Volumes + Radii. *J. Phys. Chem.* **1964**, 68, 441-&.
- 7 78. Yu, Y. B.; Privalov, P. L.; Hodges, R. S., Contribution of Translational and Rotational
8 Motions to Molecular Association in Aqueous Solution. *Biophys. J.* **2001**, 81, 1632-1642.
- 9 79. Simon, S.; Duran, M.; Dannenberg, J. J., How Does Basis Set Superposition Error
10 Change the Potential Surfaces for Hydrogen Bonded Dimers? *J. Chem. Phys.* **1996**, 105, 11024-
11 11031.
- 12 80. Henkelman, G.; Jonsson, H., Improved Tangent Estimate in the Nudged Elastic Band
13 Method for Finding Minimum Energy Paths and Saddle Points. *J. Chem. Phys.* **2000**, 113, 9978-
14 9985.
- 15 81. Henkelman, G.; Uberuaga, B. P.; Jonsson, H., A Climbing Image Nudged Elastic Band
16 Method for Finding Saddle Points and Minimum Energy Paths. *J. Chem. Phys.* **2000**, 113, 9901-
17 9904.
- 18 82. Liu, F.; Luehr, N.; Kulik, H. J.; Martinez, T. J., Quantum Chemistry for Solvated
19 Molecules on Graphical Processing Units Using Polarizable Continuum Models. *J. Chem.*
20 *Theory Comput.* **2015**, 11, 3131-3144.
- 21 83. Ufimtsev, I. S.; Martinez, T. J., Quantum Chemistry on Graphical Processing Units. 3.
22 Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics.
23 *J. Chem. Theory Comput.* **2009**, 5, 2619-2628.
- 24 84. Petachem. <http://www.petachem.com>. (accessed December 9, 2017).
- 25 85. Baker, J., An Algorithm for the Location of Transition-States. *J. Comput. Chem.* **1986**, 7,
26 385-395.
- 27 86. Ioannidis, E. I.; Gani, T. Z.; Kulik, H. J., Molsimplify: A Toolkit for Automating
28 Discovery in Inorganic Chemistry. *J. Comput. Chem.* **2016**, 37, 2106-2117.
- 29 87. Gani, T. Z. H.; Ioannidis, E. I.; Kulik, H. J., Computational Discovery of Hydrogen Bond
30 Design Rules for Electrochemical Ion Separation. *Chem. Mater.* **2016**, 28, 6207-6218.
- 31 88. Kim, J. Y.; Steeves, A. H.; Kulik, H. J., Harnessing Organic Ligand Libraries for First-
32 Principles Inorganic Discovery: Indium Phosphide Quantum Dot Precursor Design Strategies.
33 *Chem. Mater.* **2017**, 29, 3632-3643.
- 34 89. Liptak, M. D.; Shields, G. C., Accurate pK(a) Calculations for Carboxylic Acids Using
35 Complete Basis Set and Gaussian-N Models Combined with Cpcm Continuum Solvation
36 Methods. *J. Am. Chem. Soc.* **2001**, 123, 7314-7319.
- 37 90. Liptak, M. D.; Shields, G. C., Experimentation with Different Thermodynamic Cycles
38 Used for pK(a) Calculations on Carboxylic Acids Using Complete Basis Set and Gaussian-N
39 Models Combined with Cpcm Continuum Solvation Methods. *Int. J. Quantum Chem.* **2001**, 85,
40 727-741.
- 41 91. Cramer, C. J., *Essentials of Computational Chemistry : Theories and Models*, 2nd ed.;
42 Wiley: Chichester, West Sussex, England ; Hoboken, NJ, 2004, p xx, 596 p.
- 43 92. MarvinSketch (17.1.19.0), 2016, Chemaxon. <http://www.chemaxon.com> (accessed
44 February 2, 2017).
- 45 93. NBO6.0., E. D. Glendening, J. K. Badenhoop, A. E. Reed, J. E. Carpenter, J. A.
46 Bohmann, C. M. Morales, C. R. Landis, and F. Weinhold, Theoretical Chemistry Institute,
47 University of Wisconsin, Madison. 2013.
- 48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 94. Hastie, T.; Tibshirani, R.; Friedman, J. H., *The Elements of Statistical Learning : Data*
4 *Mining, Inference, and Prediction*, 2nd ed.; Springer: New York, 2009, p xxii, 745 p.
- 5 95. Friedman, J.; Hastie, T.; Tibshirani, R., Regularization Paths for Generalized Linear
6 Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1-22.
- 7
8 96. Allen, F. H., The Cambridge Structural Database: A Quarter of a Million Crystal
9 Structures and Rising. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 380-388.
- 10 97. Hansen, E. C.; Pedro, D. J.; Wotal, A. C.; Gower, N. J.; Nelson, J. D.; Caron, S.; Weix,
11 D. J., New Ligands for Nickel Catalysis from Diverse Pharmaceutical Heterocycle Libraries.
12 *Nat. Chem.* **2016**, *8*, 1126-1130.
- 13 98. Bento, A. P., et al., The ChEMBL Bioactivity Database: An Update. *Nucleic Acids Res.*
14 **2014**, *42*, D1083-D1090.
- 15 99. Bender, A.; Glen, R. C., Molecular Similarity: A Key Technique in Molecular
16 Informatics. *Org. Biomol. Chem.* **2004**, *2*, 3204-3218.
- 17 100. Eckert, H.; Bojorath, J., Molecular Similarity Analysis in Virtual Screening: Foundations,
18 Limitations and Novel Approaches. *Drug Discovery Today* **2007**, *12*, 225-233.
- 19 101. Butina, D., Unsupervised Data Base Clustering Based on Daylight's Fingerprint and
20 Tanimoto Similarity: A Fast and Automated Way to Cluster Small and Large Data Sets. *J. Chem.*
21 *Inf. Comput. Sci.* **1999**, *39*, 747-750.
- 22 102. Lindskog, S., Structure and Mechanism of Carbonic Anhydrase. *Pharmacol. Ther.* **1997**,
23 *74*, 1-20.
- 24 103. Kimura, E., Model Studies for Molecular Recognition of Carbonic Anhydrase and
25 Carboxypeptidase. *Acc. Chem. Res.* **2001**, *34*, 171-179.
- 26 104. Tautermann, C. S.; Loferer, M. J.; Voegelé, A. F.; Liedl, K. R., About the Kinetic
27 Feasibility of the Lipscomb Mechanism in Human Carbonic Anhydrase II. *J. Phys. Chem. B*
28 **2003**, *107*, 12013-12020.
- 29 105. Liang, J. Y.; Lipscomb, W. N., Hydration of CO₂ by Carbonic-Anhydrase -
30 Intramolecular Proton-Transfer between Zn²⁺-Bound H₂O and Histidine-64 in Human Carbonic
31 Anhydrase-II. *Biochemistry* **1988**, *27*, 8676-8682.
- 32 106. Lindskog, S., The Catalytic Mechanism of Carbonic-Anhydrase. *Inorg. Chim. Acta* **1983**,
33 *79*, 36-37.
- 34 107. Mikulski, R. L.; Silverman, D. N., Proton Transfer in Catalysis and the Role of Proton
35 Shuttles in Carbonic Anhydrase. *Biochim. Biophys. Acta, Proteins Proteomics* **2010**, *1804*, 422-
36 426.
- 37 108. Kozuch, S.; Shaik, S., How to Conceptualize Catalytic Cycles? The Energetic Span
38 Model. *Acc. Chem. Res.* **2011**, *44*, 101-110.
- 39 109. Kozuch, S., A Refinement of Everyday Thinking: The Energetic Span Model for Kinetic
40 Assessment of Catalytic Cycles. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 795-815.
- 41 110. Campbell, C. T., Finding the Rate-Determining Step in a Mechanism - Comparing
42 Dedonder Relations with the "Degree of Rate Control". *J. Catal.* **2001**, *204*, 520-524.
- 43 111. Campbell, C. T., Micro- and Macro-Kinetics: Their Relationship in Heterogeneous
44 Catalysis. *Top. Catal.* **1994**, *1*, 353-366.
- 45 112. Janet, J. P.; Kulik, H. J., Resolving Transition Metal Chemical Space: Feature Selection
46 for Machine Learning and Structure-Property Relationships. *J. Phys. Chem. A* **2017**, *121*, 8939-
47 8954.
- 48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 113. Fernandez, I.; Bickelhaupt, F. M., The Activation Strain Model and Molecular Orbital
4 Theory: Understanding and Designing Chemical Reactions. *Chem. Soc. Rev.* **2014**, *43*, 4953-
5 4967.

6 114. van Zeist, W. J.; Bickelhaupt, F. M., The Activation Strain Model of Chemical
7 Reactivity. *Org. Biomol. Chem.* **2010**, *8*, 3118-3127.
8
9
10
11
12
13
14
15

