Supplementary information

Constructing and interpreting volcano plots and activity maps to navigate homogeneous catalyst landscapes

In the format provided by the authors and unedited

Supporting Information for Constructing and Interpreting Volcano Plots and Activity Maps to Navigate Homogeneous Catalyst Landscapes

Rubén Laplaza^{1,2}, Shubhajit Das¹, Matthew D. Wodrich^{1,2} and Clémence Corminbœuf^{1,2*}

¹Laboratory for Computational Molecular Design (LCMD), Institute of Chemical Sciences and Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, 1015, Switzerland.
²National Center for Competence in Research-Catalysis (NCCR-Catalysis), École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland.

*Corresponding author(s). E-mail(s): clemence.corminboeuf@epfl.ch; Contributing authors: ruben.laplazasolanas@epfl.ch; shubhajit.das@epfl.ch; matthew.wodrich@epfl.ch;

Keywords: Volcano plot, Homogeneous catalysis, DFT, Activity map

Example procedure for CO_2 hydrogenation cast into volcano plots

Below a fully worked out step by step procedure following the protocol is provided.

Mechanism Identification

1. The catalytic cycle mechanism was identified from a previously reported DFT study on transition metal pincer complex-catalyzed hydrogenation of CO_2 to formate (see the reaction below).[1]

$$1 + H_2 + CO_2 \xrightarrow{\operatorname{Cat.}} HCOO^- + H_2O$$

2. Skipped (known mechanism).

Catalyst Pool Selection

- 3. The metal center and the coordinated pincer ligands were identified from the structure of the catalyst.
- 4. Five metals (Co(III), Rh(III), Ir(III), Ru(II), Os(II)) were chosen by starting from the PNP-pincer iridium(III) catalyst reported by Nozaki *et al.*[2] and considering metal centers in the same group as well as in the adjacent one in order to vary the properties of the active site. Seven pincer ligands were chosen on the basis of their connecting atom (N, P), the linker atoms which control the biting angle (C, N, O)[3], and the rigidity of the pincer backbone (Figure S1). By combining the five metals with seven pincer ligands a diverse pool of 35 catalysts was generated.

Reaction Data Acquisition

The computations in this step were performed with Gaussian09[4] (steps 6-8) and ADF[5] (steps 9,12).

- 5. Gaussview was used to to build the structures of each catalytic cycle intermediate/TS for all the catalysts in the pool.
- 6. Full geometry optimizations were performed for each catalytic cycle intermediate at the M06/def2-SVPD level of theory in implicit water solvent using the SMD solvation model[6] and the "ultrafine" integration grid.
- 7. Full transition state optimizations were performed for each catalytic cycle transition state at the M06/def2-SVPD level of theory in implicit water solvent using the SMD solvation model[6] and the "ultrafine" integration grid.
- 8. Following convergence of the intermediate/transition state structures, frequency calculations (at the same theoretical level used for the previous step) were performed for all species.



Figure S1: a) Catalytic cycle for the transition metal pincer complex-catalyzed hydrogenation of CO_2 to formate. b) Chemical composition of the catalyst pool used to construct the LFESRs and the volcano plots.

- Single point computations were performed on the optimized geometries of all species at the PBE0-dDsC/ATZ2P level of theory as implemented in ADF.[5, 7–10]
- 10. It was confirmed that all the calculations have terminated normally. The total energy for each species was obtained from the single point calculations at the PBE0-dDsC/ATZ2P level of theory.
- 11. Thermochemical corrections were obtained from frequency computations at the M06/def2-SVPD level of theory in implicit water solvent using the SMD solvation model[6] and the "ultrafine" integration grid at 298 K and 1 atm pressure.
- 12. Solvation (water) corrections were obtained using the COSMO-RS model[11] at the PBE0-dDsC/ATZ2P level of theory.
- 13. Correction for the translational entropies in solution were obtained employing the GoodVibes program by Paton and Funes-Ardoiz.[12, 13]
- 14. The solvation corrected free energy values were calculated by adding the electronic energy from the single point, the thermal correction to the Gibbs free energy from the frequency computations and the solvent correction.
- 15. The relative stabilities, $\Delta G_{RRS}(\mathbf{Int}/\mathbf{TS n})$, of each intermediate and transition state in the catalytic cycle were estimated relative to the catalyst, **1** and the substrates H₂ and CO₂.
- 16. The final free energy profiles for each catalyst was tabulated in a spreadsheet where each row corresponds to a different catalyst and each column to a different catalytic cycle stationary point.

Determination of Linear Scaling Relationships

- 17. A plotting tool was launched.
- 18. From the previously created spreadsheet, the values in each intermediate column were plotted against each other intermediate and transition state column.
- 19. An ordinary least-squares fitting of the energetic values was performed to obtain a linear free energy scaling relationship for each intermediate/intermediate or intermediate/transition state pair.



Figure S2: Regression diagnostic heatmap for the least-squares fitted linear equations of all stationary point pairs.

20. For each $\Delta G_{RRS}(\mathbf{Int/TS n})$, the average R^2 , MAE and MAPE from all the $\Delta G_{RRS}(\mathbf{Int/TS n})$ vs. $\Delta G_{RRS}(\mathbf{Int m})$ (for $n \neq m$) linear fits were calculated. The individual values are shown in Figure S2. Note that considering both a diagnostic of correlation (R^2) and of regression error (MAE) is recommended.

Descriptor Selection

21. Based on the quality of the linear fits, $\Delta G_{RRS}(\mathbf{4})$, was chosen as the descriptor variable to build the volcano plot. Note that $\Delta G_{RRS}(\mathbf{3})$, which appears as a viable option in terms of R^2 values, leads to a significant MAE when regressing $\Delta G_{RRS}(\mathbf{TS2})$ and is therefore discarded.

22. After selecting the descriptor variable, the relative stabilities of each intermediate and transition state were expressed as a linear function of $\Delta G_{RRS}(\mathbf{4})$. The set of LFESRs (see Figure S3) are given below:

 $\Delta G_{RRS}(\mathbf{2}) = 0.4 \Delta G_{RRS}(\mathbf{4}) - 13.6 \text{ kcal/mol}, \\ \Delta G_{RRS}(\mathbf{3}) = 0.7 \Delta G_{RRS}(\mathbf{4}) + 1.0 \text{ kcal/mol}, \\ \Delta G_{RRS}(\mathbf{TS1}) = 0.1 \Delta G_{RRS}(\mathbf{4}) + 5.2 \text{ kcal/mol}, \\ \Delta G_{RRS}(\mathbf{TS2}) = 0.5 \Delta G_{RRS}(\mathbf{4}) + 1.1 \text{ kcal/mol}$

- 23. Skipped (one descriptor chosen)
- 24. Since one descriptor was chosen, steps 25 to 39 were followed to construct and analyze all three variants (thermodynamic, kinetic, and energy span) of the volcano plots. Steps pertaining to activity map construction are not shown.



Figure S3: LFESRs with descriptor variable set to $\Delta G_{RRS}(\mathbf{4})$, plotted against a) $\Delta G_{RRS}(\mathbf{2})$ b) $\Delta G_{RRS}(\mathbf{3})$ c) $\Delta G_{RRS}(\mathbf{TS1})$ d) $\Delta G_{RRS}(\mathbf{TS2})$

Simulated Reaction Profile for one descriptor

25. Estimate the relative free energies associated with moving between two (non)-linked intermediates/transition states as a function of $\Delta G_{RRS}(\mathbf{4})$



Figure S4: Simulated reaction profiles that lead to a) thermodyanamic, b)

kinetic, and c) energy span volcano plots as derived from LFESRs.

using the previously established LFESRs. Depending on the desired volcano plot variant (thermodynamic, kinetic, or energy span) follow step 26, 27, or 28, respectively. All free energy values are reported in kcal/mol. The total reaction energy is obtained from,

$$\Delta G_{rxn} = G(\mathbf{HCOO}^{-}) + G(\mathbf{H_2O}) - G(\mathbf{CO_2}) - G(\mathbf{H_2}) - G(\mathbf{OH}^{-}) = -21.3 \ kcal/mol$$
(1)

The relative free energy of the separated reactants is zero, by definition. i.e. $G(1) = G(CO_2) = G(H_2) = G(OH^-) = 0$. Additionally, $G(HCOO^-)$ is also considered to be zero and $G(H_2O)$ is assigned to the value of ΔG_{rxn} .

26. Thermodynamic volcano

Estimate the free energy change associated with moving between intermediates directly linked in the catalytic cycle (Figure S4a). In all cases, the ΔG_{RRS} values on the right-side of the equations should be replaced with the LFESRs. The new equations describing the free energy associated with moving between linked intermediates derived below should each be cast in terms of the descriptor variable. Recall that, by definition, the entry point of the catalytic cycle has a relative free energy of zero (*i.e.*, $\Delta G(\mathbf{1}) = 0$), and that for the reaction step closing the cycle the reaction free energy (*i.e.*, ΔG_{rxn}) must be added. For this cycle consisting of four

6

Supporting Information

intermediates (Int), the following equations would define the $-\Delta G$ of the individual reaction steps.

For the first step:

$$-\Delta G(\mathbf{1} \to \mathbf{2}) = -\left[\Delta G_{RRS}(\mathbf{2}) - \Delta G_{RRS}(\mathbf{1}) - G(\mathbf{CO}_{\mathbf{2}})\right]$$

= -0.4\Delta G_{RRS}(\mathbf{4}) + 13.6 kcal/mol (2)

For the intermediate steps :

$$-\Delta G(\mathbf{2} \to \mathbf{3}) = -\left[\Delta G_{RRS}(\mathbf{3}) + G(\mathbf{HCOO}^{-}) - \Delta G_{RRS}(\mathbf{2})\right]$$

= -[(-0.7\Delta G_{RRS}(\mathbf{4}) + 1.0) - (0.4\Delta G_{RRS}(\mathbf{4}) - 13.6)]
= -0.3\Delta G_{RRS}(\mathbf{4}) - 14.6 \ kcal/mol
(3)

$$-\Delta G(\mathbf{3} \to \mathbf{4}) = -\left[\Delta G_{RRS}(\mathbf{4}) - \Delta G_{RRS}(\mathbf{3}) - G(\mathbf{H_2})\right]$$

=
$$-\left[\left(\Delta G_{RRS}(\mathbf{4}) - (0.7\Delta G_{RRS}(\mathbf{4}) + 1.0)\right]$$
(4)
=
$$-0.3\Delta G_{RRS}(\mathbf{4}) - 1.0 \ kcal/mol$$

For the final step to close the catalytic cycle:

$$-\Delta G(\mathbf{4} \to \mathbf{1}) = -\left[\Delta G_{RRS}(\mathbf{1}) + \Delta G_{H_2O} - \Delta G_{RRS}(\mathbf{4})\right]$$
$$= -\left[-21.3 - \Delta G_{RRS}(\mathbf{4})\right]$$
$$= \Delta G_{RRS}(\mathbf{4}) + 21.3 \ kcal/mol$$
(5)

27. Kinetic volcano

Estimate the free energy change associated with moving between directly linked intermediates and transition states (Figure S4b). In all cases, the ΔG_{RRS} values on the right-side of the equations should be replaced with the LFESRs. The new equations describing the free energy associated with moving between linked intermediates/transition states derived below should each be cast in terms of the descriptor variable. Recall that, by definition, the entry point of the catalytic cycle has a relative free energy of zero (*i.e.*, $\Delta G(\mathbf{1}) = 0$), and that for the reaction step closing the cycle the reaction free energy (*i.e.*, ΔG_{rxn}) must be added. For a cycle consisting of four intermediates (**Int**) and two transition states (**TS**), the following equations would define the $-\Delta G$ of the individual reaction steps.

For the first step:

$$-\Delta G(\mathbf{1} \to \mathbf{TS1}) = -\left[\Delta G_{RRS}(\mathbf{TS1}) - \Delta G_{RRS}(\mathbf{1}) - G(\mathbf{CO_2})\right]$$

= -0.1\Delta G_{RRS}(\mathbf{4}) - 5.2 kcal/mol (6)

For the intermediate steps :

$$-\Delta G(\mathbf{TS1} \rightarrow \mathbf{2}) = -\left[\Delta G_{RRS}(\mathbf{2}) - \Delta G_{RRS}(\mathbf{TS1})\right]$$

= -[(0.4\Delta G_{RRS}(\mathbf{4}) - 13.6) - (0.1\Delta G_{RRS}(\mathbf{4}) + 5.2)]
= -0.3\Delta G_{RRS}(\mathbf{4}) + 18.8 \ kcal/mol
(7)

$$-\Delta G(\mathbf{2} \to \mathbf{3}) = -\left[\Delta G_{RRS}(\mathbf{3}) + G(\mathbf{HCOO}^{-}) - \Delta G_{RRS}(\mathbf{2})\right]$$

= -[(-0.7\Delta G_{RRS}(\mathbf{4}) + 1.0) - (0.4\Delta G_{RRS}(\mathbf{4}) - 13.6)]
= -0.3\Delta G_{RRS}(\mathbf{4}) - 14.6 \ kcal/mol
(8)

$$-\Delta G(\mathbf{3} \to \mathbf{4}) = -\left[\Delta G_{RRS}(\mathbf{4}) - \Delta G_{RRS}(\mathbf{3}) - G(\mathbf{H_2})\right]$$

= -\left[\left(\Delta G_{RRS}(\mathbf{4}) - (0.7\Delta G_{RRS}(\mathbf{4}) + 1.0)\right] = -0.3\Delta G_{RRS}(\mathbf{4}) - 1.0 \ kcal/mol \text{(9)}

$$-\Delta G(\mathbf{4} \to \mathbf{TS2}) = -\left[\Delta G_{RRS}(\mathbf{TS2}) + G(\mathbf{OH}^{-}) - \Delta G_{RRS}(\mathbf{3}) - G(\mathbf{H_2})\right]$$
$$= -\left[(0.5\Delta G_{RRS}(\mathbf{4}) - 1.1) - \Delta G_{RRS}(\mathbf{4})\right]$$
$$= 0.5\Delta G_{RRS}(\mathbf{4}) - 1.1 \ kcal/mol$$
(10)

For the final step to close the catalytic cycle:

$$-\Delta G(\mathbf{TS2} \to \mathbf{1}) = -\left[\Delta G_{RRS}(\mathbf{1}) + \Delta G_{H_2O} - \Delta G_{RRS}(\mathbf{TS2})\right]$$
$$= -\left[-21.3 - (0.5\Delta G_{RRS}(\mathbf{4}) + 1.1)\right]$$
$$= 0.5\Delta G_{RRS}(\mathbf{4}) + 22.4 \ kcal/mol$$
(11)

28. Energy span volcano

Estimate the free energy difference associated with moving between each linked and non-linked intermediate/transition state pair (Figure S4c). In all cases, the ΔG_{RRS} values on the right-side of the equations should be replaced with the LFESRs. The new equations describing the free energy associated with moving between linked and non-linked intermediate/transition state pairs derived below should each be cast in terms of the descriptor variable. Recall that, by definition, the entry point of the catalytic cycle has a relative free energy of zero (*i.e.*, $\Delta G(\mathbf{Int1}) = 0$). For a cycle consisting of four intermediates (**Int**) and two transition states (**TS**), the following equations would define the $-\Delta G$ of the individual reaction steps

$$-\Delta G(\mathbf{1} \to \mathbf{TS1}) = -\left[\Delta G_{RRS}(\mathbf{TS1}) - \Delta G_{RRS}(\mathbf{1})\right]$$

= -0.1\Delta G_{RRS}(\mathbf{4}) - 5.2 kcal/mol (12)

$$-\Delta G(\mathbf{1} \to \mathbf{TS2}) = -\left[\Delta G_{RRS}(\mathbf{TS2}) - \Delta G_{RRS}(\mathbf{1})\right]$$

$$= -0.5\Delta G_{RRS}(\mathbf{4}) - 1.1 \ kcal/mol$$

$$-\Delta G(\mathbf{2} \to \mathbf{TS1}) = -\left[\Delta G_{RRS}(\mathbf{TS1}) - \Delta G_{RRS}(\mathbf{2}) + \Delta G_{rxn}\right]$$

$$= -\left[(0.1\Delta G_{RRS}(\mathbf{4}) + 5.2) - (0.4\Delta G_{RRS}(\mathbf{4}) - 13.6) - 21.3\right]$$

$$= 0.3\Delta G_{RRS}(\mathbf{4}) + 2.5 \ kcal/mol$$
(14)

$$-\Delta G(\mathbf{2} \to \mathbf{TS2}) = -\left[\Delta G_{RRS}(\mathbf{TS2}) - \Delta G_{RRS}(\mathbf{2})\right]$$

= -[(0.5\Delta G_{RRS}(\mathbf{4}) + 1.1) - (0.4\Delta G_{RRS}(\mathbf{4}) - 13.6)]
= -0.1\Delta G_{RRS}(\mathbf{4}) - 14.7 \ kcal/mol
(15)

$$-\Delta G(\mathbf{3} \to \mathbf{TS1}) = -\left[\Delta G_{RRS}(\mathbf{TS1}) - \Delta G_{RRS}(\mathbf{3}) + \Delta G_{rxn}\right]$$

= -[(0.1\Delta G_{RRS}(\mathbf{4}) + 5.2) - (0.7\Delta G_{RRS}(\mathbf{4}) + 1.0) - 21.3]
= 0.6\Delta G_{RRS}(\mathbf{4}) + 17.1 \ kcal/mol
(16)

$$-\Delta G(\mathbf{3} \to \mathbf{TS2}) = -\left[\Delta G_{RRS}(\mathbf{TS2}) - \Delta G_{RRS}(\mathbf{3})\right]$$

= -\left[(0.5\Delta G_{RRS}(\mathbf{4}) + 1.1) - (0.7\Delta G_{RRS}(\mathbf{4}) + 1.0)\right]
= 0.2\Delta G_{RRS}(\mathbf{4}) - 0.1 kcal/mol (17)

$$-\Delta G(\mathbf{4} \to \mathbf{TS1}) = -\left[\Delta G_{RRS}(\mathbf{TS1}) - \Delta G_{RRS}(\mathbf{4}) + +\Delta G_{rxn}\right]$$

$$= -\left[(0.1\Delta G_{RRS}(\mathbf{4}) + 5.2) - \Delta G_{RRS}(\mathbf{4}) - 21.3\right] \quad (18)$$

$$= 0.9\Delta G_{RRS}(\mathbf{4}) + 16.1 \ kcal/mol$$

$$-\Delta G(\mathbf{4} \to \mathbf{TS2}) = -\left[\Delta G_{RRS}(\mathbf{TS2}) - \Delta G_{RRS}(\mathbf{4})\right]$$

$$= -\left[(0.5\Delta G_{RRS}(4) + 1.1) - \Delta G_{RRS}(4) \right]$$
(19)
= $0.5\Delta G_{RRS}(4) - 1.1 \ kcal/mol$
Note that the correction term ΔG_{rxn} is added to account for combina-

Note that the correction term ΔG_{rxn} is added to account for combinations in which the transition state appears before the intermediate in the catalytic cycle.

29. Using a plotting tool, a single plot was created where each of linear equations defining the negative of reaction energy (obtained from the proceeding steps) are plotted as a function of $\Delta G_{RRS}(4)$. For a thermodynamic volcano these lines are equations 2-5, for a kinetic volcano equations 6-11 and for an energy span volcano equations 12-19. Plotting each of these respective groups of equations leads to the "simulated reaction profile" (Figure S4) that can predict the energetics associated with each step of the catalytic cycle through knowledge of the value of the descriptor variable.

Construction of the volcano plot

30. Volcano plots define the most energetically costly reaction step [*i.e.*, the potential determining step (pds) for thermodynamic, the kinetic determining step (kds) for kinetic, and the energy span (δE) for energy span volcanoes] of the catalytic cycle for a given value of descriptor variable. To obtain the volcano curve from the simulated reaction profile, the reaction line segments with the smallest value amongst all possibilities over the particular range of the $\Delta G_{RRS}(4)$ were taken. The three variants, along with the TOF volcano, are shown in Figure S5a-d. Mathematically, this is accomplished using the following expressions for a thermodynamic, kinetic, and energy span volcanoes, respectively.

$$-\Delta G(pds) = min.[\{-\Delta G(\mathbf{1} \to \mathbf{2}), -\Delta G(\mathbf{2} \to \mathbf{3}), -\Delta G(\mathbf{3} \to \mathbf{4}), -\Delta G(\mathbf{4} \to \mathbf{1})\}]$$
(20)

$$-\Delta G(kds) = min.[\{-\Delta G(\mathbf{1} \to \mathbf{TS1}), -\Delta G(\mathbf{TS1} \to \mathbf{2}), -\Delta G(\mathbf{2} \to \mathbf{3}), -\Delta G(\mathbf{3} \to \mathbf{4}), -\Delta G(\mathbf{4} \to \mathbf{TS1}), -\Delta G(\mathbf{TS2} \to \mathbf{1})\}]$$
(21)

$$-\delta E = min.[\{-\Delta G(\mathbf{1} \to \mathbf{TS1}), -\Delta G(\mathbf{1} \to \mathbf{TS2}), -\Delta G(\mathbf{2} \to \mathbf{TS1}), \\ -\Delta G(\mathbf{2} \to \mathbf{TS2}), -\Delta G(\mathbf{3} \to \mathbf{TS1}), -\Delta G(\mathbf{3} \to \mathbf{TS2}), \\ -\Delta G(\mathbf{4} \to \mathbf{TS1}), -\Delta G(\mathbf{4} \to \mathbf{TS2})\}]$$
(22)

The final volcano shapes consist of multiple reaction lines as the range of descriptor values is traversed. In practice creating the volcano plot simply involves taking the lowest line seen in the simulated reaction profile for each value of the descriptor variable. Retrieval of the volcano shaped curve marks the final step in volcano construction for the thermodynamic and kinetic volcano. In the present case, the energy span volcano is further post-processed to generate a turnover frequency (TOF) volcano (see steps 29 to 31).

- 31. To construct a TOF volcano, the hypothetical free energy profile from the LFESRs for values of $\Delta G_{RRS}(4)$ within a range 0-80 kcal/mol (in 0.3 kcal/mol interval).
- 32. The TOF corresponding for each desired descriptor variable were calculated using the hypothetical free energy profiles from volcanic.py. The following TOF formula was used.

$$TOF = \frac{k_B T}{h} \frac{e^{-\Delta G_{rxn}/RT} - 1}{\sum_{n,m=1}^{N} e^{(\Delta G(\operatorname{Int} \mathbf{n} \to \mathbf{TS} \mathbf{m}))/RT}}$$
(23)

where $\Delta G(\operatorname{Int} \mathbf{n} \to \mathbf{TS} \mathbf{m})$ is the free energy change associated with going from the *n*th intermediate to the *m*th transition state as defined in step 28; ΔG_{rxn} is the Gibbs free energy of the reaction, k_B is the Boltzmann constant, *h* is the Planck constant, *R* is the ideal gas constant and *T* is the temperature.

33. A plotting tool was launched. $\Delta G_{RRS}(4)$ and the corresponding TOF values (the relative activity of different catalysts are easier to see in the y-axis is plotted in log scale) are plotted to obtain the TOF volcano curve.



Figure S5: Different variants of the volcano plots illustrating the anticipated activity for CO_2 hydrogenation to formate a) Thermodynamic volcano plot b) Kinetic volcano plot c) Energy Span volcano plot d) TOF volcano plot

Analysis and interpretation of volcano plots

- 34. Points for the individual catalysts were added to the volcano plot using the value of $\Delta G_{RRS}(4)$ and $-\Delta G(pds/kds)$ or $-\delta E$ as sets of Cartesian coordinates.
- 35. The relative positioning of the catalysts were examined in all three variants of the volcano plots. The anticipated most active candidates are found highest on the plot, on or near the volcano summit or plateau. Less active catalysts lie lower on the plot, often along the left and right slopes.

- 36. Splitting the volcano plot into sections (by drawing vertical lines) based on the most difficult process in the catalytic cycle can help understand what limits catalytic activity. Both the thermodynamic (Figure S 5a) and kinetic (Figure S 5a) volcanoes can be divided into two regions. The right region in both variant is limited by $2 \rightarrow 3$ while the left region is governed by $4 \rightarrow 1$ and $4 \rightarrow TS2$ for thermodynamic and kinetic volcanoes, respectively. The energy span volcano (Figure S 5c) is divided into three regions with the TDI/TDTS combinations being 4/TS1, 4/TS2, and 2/TS2, respectively.
- 37. The TOF volcano in Figure S 5d is a continuous curve and information about the TDI/TDTS pair that control the TOF can be obtained from splitting the related energy span volcano into three sections.
- 38. Analyzing these point reveals that group 9 metals combined with π -acidic ligands have the best profiles, with activity increasing when switching from cobalt to rhodium to iridium. Note that similar conclusions regarding the underlying chemistry can be reached from both the kinetic and the energy span/TOF volcano plot variant, albeit the latter two lead to a more accurate picture. For a more detailed analysis of the volcanoes, the readers are referred to the original publications[14, 15].
- 39. For screening new catalysts, only the descriptor value, $\Delta G_{RRS}(4)$, would need to be computed. The location of each point on the volcano plot then corresponds to the value of the descriptor value (x-coordinate) and the negative of the pds, kds, or δE (y-coordinate). The value of the y-coordinate is obtained from the equation of the volcano curve line associated with the region of the volcano where the catalyst is located, which is determined from the value of the descriptor variable.

Generating volcano plots/activity maps using volcanic

We provide the reader with command line instructions to reproduce all the volcano plots/activity maps of the main text using volcanic. All required data files and the volcanic code can be obtained from https://github.com/lcmd-epfl/volcanic within the examples directory.

1. For Figure 3a:

```
volcanic.py -i hydroformylation.xlsx --thermo -pm 0 -refill -is simple
```

2. For Figure 3b:

volcanic.py -i hydrogenation_test.xlsx --k -pm 0 -refill -is simple

3. For Figure 3c:

volcanic.py -i aryl_ether_cleavage.xlsx --es -pm 0

4. For Figure 3d:

```
volcanic.py -i aryl_ether_cleavage.xlsx --tof -pm 0
5. For Figure 4c:
  volcanic.py -i formate.xlsx --es -lm 5 -rm 5 -pm 1
6. For Figure 4d:
  volcanic.py -i formate.xlsx --tof -lm 5 -rm 5 -pm 1
7. For Figure 5b:
  volcanic.py -i suzuki_s.xlsx --thermo -pm 0 -rm 0 -ic 2 -fc 4
8. For Figure 5d:
  volcanic.py -i cc_genome.xlsx --thermo -nd 2 -pm 0
```

Example input for volcanic

volcanic expects headers for all columns. The first column must contain names/identifiers. The program expects a number of columns with relative free energies for the species in the catalytic cycle (in order of appearance), whose headers must contain "TS" if the species is a transition state, and a final column whose header is "Product" containing the reaction energy. Additional descriptors (both energy and non-energy based) can be input as a separate file using the -df flag or as extra columns whose headers contain the word "Descriptor".

An example is given in Table 1. Real examples in machine readable file formats can be found in https://github.com/lcmd-epfl/volcanic within the examples directory.

	T	able 1: Examp	ole input file	for volcani	.c. All energi	es in kcal/mol.	
Catalyst	$\Delta GRRS(1)$	$\Delta GRRS(TS1)$	$\Delta GRRS(2)$	$\Delta GRRS(3)$	$\Delta GRRS(4)$	$\Delta GRRS(TS2)$	Product
Ir(PNP/-H),	0.00 ,	1.42,	-20.10 ,	-11.13,	-23.19 ,	-9.65 ,	-21.31
Ir(NNN/-H),	0.00,	-0.03 ,	-27.99 ,	-24.88 ,	-32.25,	-14.76 ,	-21.31
Ir(PCP-H),	0.00,	-1.08 ,	-26.18,	-22.47 ,	-35.58 ,	-18.83 ,	-21.31
Ir(PNMeP/-H),	0.00,	3.01,	-20.26 ,	-11.95,	-24.32 ,	-9.36 ,	-21.31
Ir(PNN/-H),	0.00,	1.01,	-26.97 ,	-16.48,	-27.43 ,	-11.63 ,	-21.31
Ir(PNNNP/-H),	0.00,	1.21,	-19.97 ,	-11.10,	-20.89 ,	-10.57 ,	-21.31
Ir(PONOP/-H),	0.00 ,	2.94,	-15.02 ,	-5.81 ,	-10.22,	-4.59,	-21.31
Rh(PNP/-H),	0.00,	1.05,	-25.65 ,	-17.66 ,	-26.65,	-10.93 ,	-21.31
Rh(NNN/-H),	0.00,	0.77,	-33.36 ,	-31.57,	-34.67,	-18.15 ,	-21.31
Rh(PCP/-H),	0.00,	-1.20 ,	-27.22 ,	-30.18,	-41.66,	-22.11,	-21.31
Rh(PNMeP/-H),	0.00,	6.01,	-27.70 ,	-21.26,	-29.68,	-13.92 ,	-21.31
Rh(PNN/-H),	0.00 ,	0.67,	-32.33 ,	-24.16,	-31.50,	-17.30 ,	-21.31
Rh(PNNP/-H),	0.00,	0.89,	-25.60 ,	-18.29 ,	-25.86,	-13.53 ,	-21.31
Rh(PONOP/-H),	0.00,	3.54,	-20.58 ,	-12.77 ,	-19.26,	-7.45 ,	-21.31
Co(PNP/-H),	0.00,	0.12,	-28.15 ,	-17.13 ,	-28.78 ,	-10.81 ,	-21.31
Co(NNN/-H),	0.00,	-0.78 ,	-37.77 ,	-35.18,	-35.79 ,	-19.05,	-21.31
Co(PCP/-H),	0.00,	-0.90 ,	-34.02,	-31.34,	-44.96,	-22.29 ,	-21.31
Co(PNMeP/-H),	0.00,	1.20,	-29.11,	-19.22,	-31.18,	-12.65 ,	-21.31
Co(PNN/-H),	0.00,	-0.17 ,	-33.73 ,	-24.37,	-33.30 ,	-17.41,	-21.31
Co(PNNNP/-H),	0.00,	-0.06 ,	-27.62 ,	-16.70,	-28.72 ,	-15.45,	-21.31
Co(PONOP/-H),	0.00,	3.13,	-23.26 ,	-10.33,	-21.47,	-9.05 ,	-21.31
Ru(PNP-H),	0.00,	-2.33 ,	-37.04 ,	-39.35 ,	-55.33 ,	-25.44,	-21.31
Ru(NNN/-H),	0.00,	-2.33 ,	-46.52,	-49.12,	-61.98,	-28.69 ,	-21.31
Ru(PCP/-H),	0.00,	-3.31 ,	-44.69,	-50.85 ,	-70.07 ,	-35.62 ,	-21.31
Ru(PNMeP/-H),	0.00,	-3.86 ,	-43.07 ,	-45.06,	-62.16,	-31.60,	-21.31
Ru(PNN/-H),	0.00,	-3.53 ,	-44.45,	-44.41,	-60.35 ,	-31.60,	-21.31
Ru(PNNNP/-H),	0.00 ,	-2.96,	-38.05 ,	-38.32 ,	-54.32,	-22.75,	-21.31
Ru(PONOP/-H),	0.00,	-1.26 ,	-33.51,	-33.04 ,	-48.50,	-22.24 ,	-21.31

References

- Yang, X.: Hydrogenation of carbon dioxide catalyzed by pnp pincer iridium, iron, and cobalt complexes: a computational design of base metal catalysts. ACS Catalysis 1(8), 849–854 (2011)
- [2] Tanaka, R., Yamashita, M., Nozaki, K.: Catalytic hydrogenation of carbon dioxide using ir(III)-pincer complexes. Journal of the American Chemical Society 131(40), 14168–14169 (2009). https://doi.org/10.1021/ja903574e
- [3] Leis, W., Mayer, H., Kaska, W.: Cycloheptatrienyl, alkyl and aryl PCP-pincer complexes: Ligand backbone effects and metal reactivity. Coordination Chemistry Reviews 252(15-17), 1787–1797 (2008). https: //doi.org/10.1016/j.ccr.2008.02.002
- [4] Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G.A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H.P., Izmaylov, A.F., Bloino, J., Zheng, G., Sonnenberg, J.L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J.A. Jr., Peralta, J.E., Ogliaro, F., Bearpark, M., Heyd, J.J., Brothers, E., Kudin, K.N., Staroverov, V.N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J.C., Iyengar, S.S., Tomasi, J., Cossi, M., Rega, N., Millam, J.M., Klene, M., Knox, J.E., Cross, J.B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R.E., Yazyev, O., Austin, A.J., Cammi, R., Pomelli, C., Ochterski, J.W., Martin, R.L., Morokuma, K., Zakrzewski, V.G., Voth, G.A., Salvador, P., Dannenberg, J.J., Dapprich, S., Daniels, A.D., Farkas, [U+FFFD], Foresman, J.B., Ortiz, J.V., Cioslowski, J., Fox, D.J.: Gaussian~09 Revision D.01. Gaussian Inc. Wallingford CT 2009
- [5] te Velde, G., Bickelhaupt, F.M., Baerends, E.J., Guerra, C.F., van Gisbergen, S.J.A., Snijders, J.G., Ziegler, T.: Chemistry with ADF. Journal of Computational Chemistry 22(9), 931–967 (2001). https://doi.org/10. 1002/jcc.1056
- [6] Marenich, A.V., Cramer, C.J., Truhlar, D.G.: Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. The Journal of Physical Chemistry B 113(18), 6378–6396 (2009). https://doi. org/10.1021/jp810292n
- [7] Steinmann, S.N., Corminboeuf, C.: A system-dependent density-based dispersion correction. Journal of Chemical Theory and Computation 6(7), 1990–2001 (2010). https://doi.org/10.1021/ct1001494

- [8] Steinmann, S.N., Corminboeuf, C.: A density dependent dispersion correction. CHIMIA International Journal for Chemistry 65(4), 240–244 (2011). https://doi.org/10.2533/chimia.2011.240
- [9] Steinmann, S.N., Corminboeuf, C.: A generalized-gradient approximation exchange hole model for dispersion coefficients. The Journal of Chemical Physics 134(4), 044117 (2011). https://doi.org/10.1063/1.3545985
- [10] Steinmann, S.N., Corminboeuf, C.: Comprehensive benchmarking of a density-dependent dispersion correction. Journal of Chemical Theory and Computation 7(11), 3567–3577 (2011). https://doi.org/10.1021/ ct200602x
- Klamt, A.: The COSMO and COSMO-RS solvation models. WIREs Computational Molecular Science 8(1) (2017). https://doi.org/10.1002/wcms. 1338
- [12] Grimme, S.: Supramolecular binding thermodynamics by dispersioncorrected density functional theory. Chemistry - A European Journal 18(32), 9955–9964 (2012). https://doi.org/10.1002/chem.201200497
- [13] Luchini, G., Alegre-Requena, J.V., Funes-Ardoiz, I., Paton, R.S.: Good-Vibes: automated thermochemistry for heterogeneous computational chemistry data. F1000Research 9, 291 (2020). https://doi.org/10.12688/ f1000research.22758.1
- [14] Sawatlon, B., Wodrich, M.D., Corminboeuf, C.: Unraveling metal/pincer ligand effects in the catalytic hydrogenation of carbon dioxide to formate. Organometallics 37(24), 4568–4575 (2018). https://doi.org/10.1021/acs. organomet.8b00490
- [15] Wodrich, M.D., Sawatlon, B., Solel, E., Kozuch, S., Corminboeuf, C.: Activity-based screening of homogeneous catalysts through the rapid assessment of theoretically derived turnover frequencies. ACS Catalysis 9(6), 5716–5725 (2019). https://doi.org/10.1021/acscatal.9b00717