Divergence Between Activity Metrics and Mechanistic Interpretability in Anchored Molecular Electrocatalysts

Akash Philip Nedumthuruthiyil,†[a] Akhil Rajendran,[a] Dhruv Dhiman,[a] Pavithra B,[a,b] Tapas Ghatak,[a] Biswajit Saha,[c] Kuldeep Singh,[d] Sreetama Ghosh[b*] and Abir Sarbajna[a*]

[a]Department of Chemistry, School of Advanced Sciences, Vellore Institute of Technology, Vellore 632014, Tamil Nadu, India.

[b]CO_2 Research and Green Technologies Centre, Vellore Institute of Technology, Vellore 632014, Tamil Nadu, India.

[c]Materials Science Group, Coal Energy and Materials Sciences Division, CSIR-North East Institute of Science and Technology, Jorhat, Assam 785006, India

[d]Makromolekulare Chemie, Universität Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany
*Address for correspondence: abir.sarbajna@vit.ac.in, sreetama.ghosh@vit.ac.in

Contents

S1. Identity and Anchoring Census **S4**

S2. Electrochemical and solvent medium reporting **S5**

S3. Assignment of Mechanistic weights **S8**

S4. Utility Descriptor Potentials and Rate and Stability Determining Limitations **S10**

S5. Integration of Design Logic and Evidence Completeness **S14**

S6. Experimental Regime Descriptors **S17**

S7. Experimental Regime Organization of the Mechanistic Dataset **S19**

S8. Quadrant-Level Enrichment and Ordering Statistics in the U–R Space **S25**

S9. Robustness of Order Index **S27**

S10. Forward Validation and Boundary-Distance Analysis **S29**

S11. Statistical analysis reported in the manuscript **S32**

S12. Performance–Interpretability Divergence **S34**

S13. Circularity sensitivity check **S35**

References **S35**

Data curation

All information were taken directly from the literature. When any information could not be unambiguously extracted without interpretation, the corresponding field was left blank.

All inputs were finalized before downstream analysis in later sections.

Assignment of descriptor values

The descriptors were assigned manually by applying fixed, reaction specific extraction rules from the reports. These assignments represent qualitative values which arise from structured chemical judgment guided by extracted rules. Statistical fitting or optimization were not used in any of the assignments.

Descriptor definitions and assignment rules are provided in later sections or in the supplementary dataset.

Derived quantities introduced in later sections (e.g., summed descriptors, composite indices, or regime classifications) were computed using predefined formulas.

Role of python scripts and verification files

Scripts included in the Supplementary Information were used to compute derived quantities from stored values, verify numeric ranges, labels, and maintain the data composition. All such data summaries that build progressively from S1 onwards are also provided in the supplementary dataset.

Catalyst-level context: Offline HTML viewer

We have also provided an offline HTML viewer (CCR_Viewer.html) that allows inspection of each catalyst. The viewer contains all catalysts in the dataset, along with molecular structures, descriptor assignments, and analysis results. The HTML file is self-contained, requires no internet connection, and can be opened in any standard web browser such as Google Chrome or Microsoft Edge.

Organization of the section-wise dataset

Following this guide, the Supplementary Information proceeds through section-wise layers.

- **S1–S2** define identity of catalyst, anchoring classes, and experimental context.
- **S3–S4** introduce mechanistic weighting and U, R descriptors.
- **S5** integrates U, R descriptors with evidence completeness.
- **S6** assigns experimental regime context.
- **S7** provides the final analysis including the quadrant definitions using fixed, dataset-level rules.
- **S8–S11** report quadrant-level enrichment, robustness of the landscape, forward validation and performance–interpretability divergence studies.
- **S12** lists the statistical analysis used throughout the manuscript.

Each section includes a CSV file, a definitions or rules document, a data integrity summary, and (where applicable) a verification script. The ESI and supplementary dataset provide a transparent view of the dataset. Readers are provided with output files of the scripts so that code execution remain optional.

S1. Identity and Anchoring Census

Purpose and scope

S1 dataset is used to define the list of catalysts throughout this study. It records the identity of each anchored molecular catalyst, a single reaction under which it is evaluated, and the anchoring class. All entries were compiled from our prior review of surface-immobilized base-metal molecular catalysts reported for hydrogen evolution (HER), oxygen evolution (OER), and their photoelectrochemical analogues (PEC-HER and PEC-OER).[1]

Definition of catalyst identity

Catalysts were named according to format "Metal–Index" (with optional letter suffixes) to maintain traceability from our parent review. Metal identity was inferred directly from the *Catalyst_ID* prefix. Out of 157 catalysts, 92 are cobalt-based, 39 are iron-based, and 26 are nickel-based (Figure S1a).

Assignment to a single reaction

Some catalysts were reported to operate under multiple reactions environments, such as both HER and OER or under both electrochemical and photoelectrochemical conditions.

When multiple reactions were reported for the same catalyst, a strict priority rule was applied following the order PEC-OER > PEC-HER > OER > HER.

This is because the photoelectrochemical reactions are much scarce in literature. At the same time, OER is inherently more difficult than HER because it involves a four-electron oxidation and high-energy intermediates, leading to larger kinetic barriers than the two-electron proton reduction in HER. This leads to 94 HER systems, 33 OER systems, 25 PEC-HER systems, and 5 PEC-OER (Figure S1b). The asymmetry reflects the distribution of the reported literature.

Description of anchoring and support

Anchoring_Type recorded the mode of attachment using the authors' terminology, whereas Support records the electrode or substrate material. These fields are taken directly from the source literature.

Anchoring-class categorization

Each catalyst was additionally assigned to one anchoring class that represented the dominant chemical nature of immobilization. The anchoring classes were classified into covalent, non-covalent, π–π, phosphonate/chelate, and core–shell/interfacial. Across the full dataset, the anchoring-class distribution revealed 56 covalent, 47 non-covalent, 32 π–π, 13 phosphonate/chelate, and 9 core–shell/interfacial systems. Anchoring strategies are not uniformly distributed across reaction types and are described in Figure S1b and S1c.

Data integrity

Following automated integrity checks, the S1 dataset was locked. All subsequent sections were cumulatively built on S1.

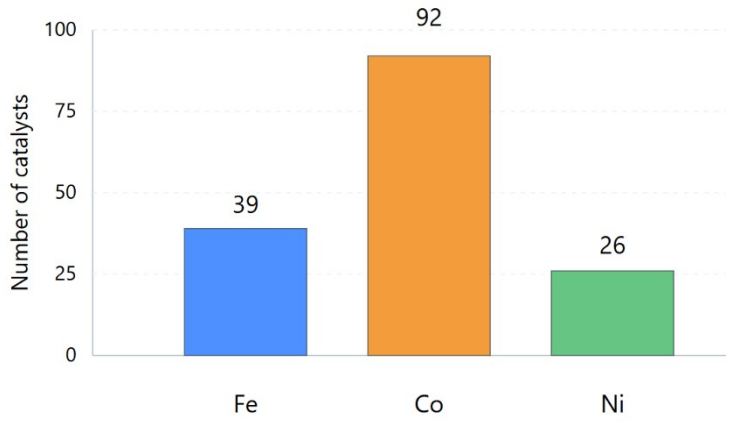


Figure S1a. Distribution of cobalt, iron, and nickel based anchored molecular electrocatalysts.

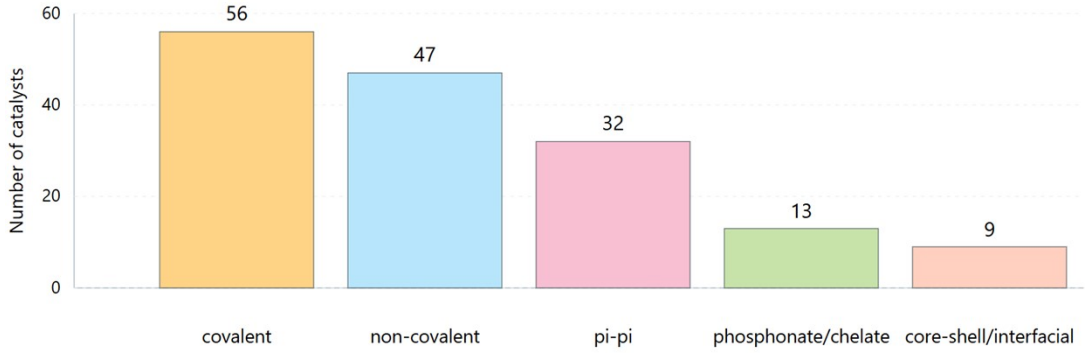


Figure S1b. Overall distribution of anchoring classes in the S1 dataset.

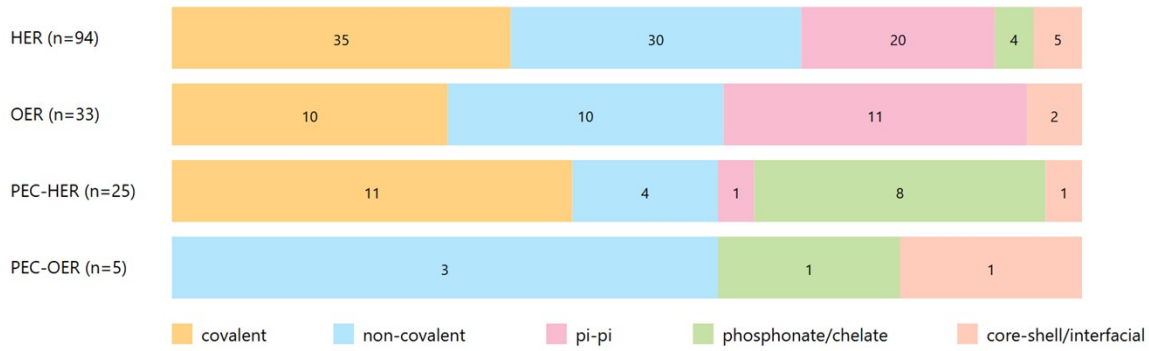


Figure S1c. Distribution of anchoring classes across HER, OER, PEC-HER, and PEC-OER reactions.

S2. Electrochemical and solvent medium reporting

Purpose and role of S2

Section S2 was used to document the electrochemical parameters and reaction media reported in the literature for each catalyst described in Section S1. Reporting gaps were retained because uneven reporting is itself a feature of the literature.

Scope and parameter definitions

S2 adds five reported parameters to the S1 and the values are transcribed as reported. No values were added unless values were reported explicitly.

• Overpotential at 10 mA cm^{-2} (η_{10}) was recorded in V. If η_{10} is only shown graphically in the reported without an extractable number, the field was left blank.

• Tafel slope was recorded as mV dec^{-1}, using the authors' primary reported slope.

• Stability duration was recorded in hours using by studying the chronoamperometric studies. If stability was only described qualitatively ("stable") without a duration, the field was left blank.

• Electrolyte was recorded from the literature directly such as "0.5 M H_2SO_4", "1.0 M KOH", "phosphate buffer".

• pH values were recorded only when explicitly reported. No values were inferred from electrolyte identity or composition.

Dataset coverage and missingness

Parameter	Reported	Missing
η_{10} (V)	59	98
Tafel (mV dec^{-1})	52	105
Stability (h)	104	53
Electrolyte	156	1
pH	126	31

Distributions of reported metrics

Parameter	Metal	n	Mean	Range
η_{10} (V)	Co	38	0.3807	0.081–0.59
	Fe	12	0.8146	0.406–1.123
	Ni	9	0.35	0.02–0.514
Tafel (mV dec^{-1})	Co	32	107.998	41.0–250.0
	Fe	14	174.8	60.0–326.2
	Ni	6	56.667	35.0–91.0
Stability (h)	Co	69	9.552	0.083–50
	Fe	21	11.333	1–36
	Ni	14	27.393	0.5–288
pH	Co	82	6.256	0–14
	Fe	26	7.765	1–13
	Ni	18	4.583	0–14

These summaries reflect reported values only and are not used here to rank catalysts or infer mechanism.

Electrolyte reporting

Across the dataset, 156 of 157 systems report an electrolyte descriptor, with 66 distinct electrolyte values in total (Figure S2c).

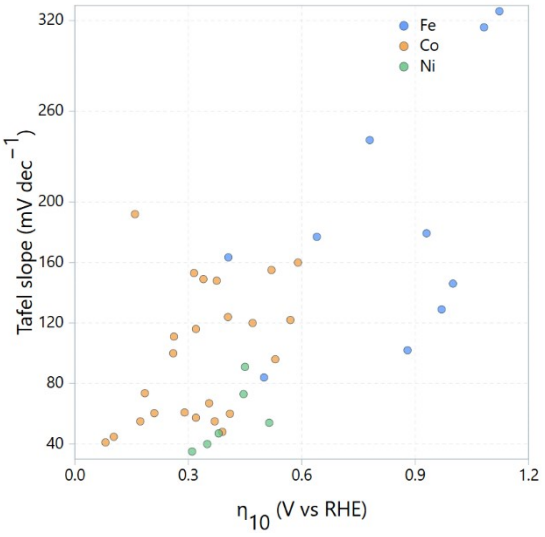


Figure S2a. η_{10} versus Tafel slope for anchored molecular catalysts.

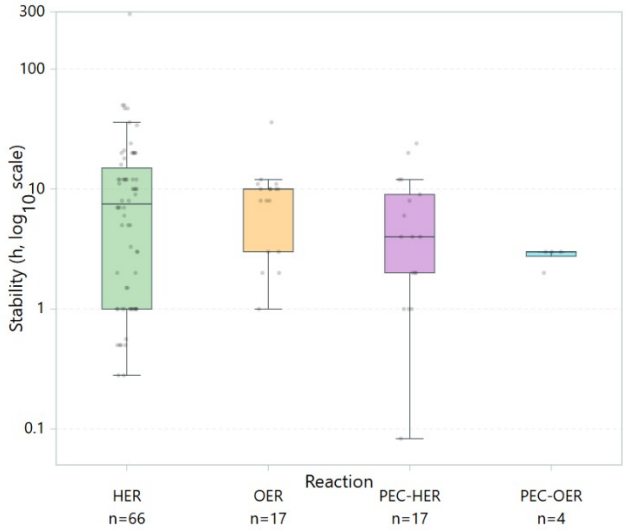


Figure S2b. Stability durations for anchored molecular catalysts, shown on a logarithmic time scale.

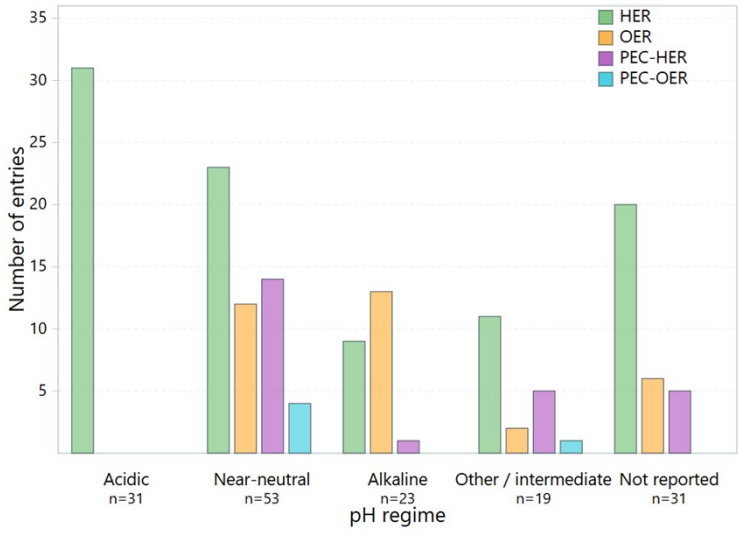


Figure S2c. pH distribution across reaction classes.

S3. Assignment of Mechanistic weights

Purpose and role of S3.

Section S3 was introduced to add anchoring-imposed mechanistic weights using predefined rules. We did not attempt to identify reaction mechanisms. Instead, we assigned continuous weight fractions for oxidative addition (OA), metal–ligand cooperativity (MLC), and proton-coupled electron transfer (PCET) to each catalyst using predefined that reflects anchoring chemistry and reaction environment at the interface.

Definition of mechanistic axes.

• OA (oxidation–addition–like character) suggests metal-centered redox activity and bond-making/bond-breaking dominated by the metal site.

• MLC (metal–ligand cooperativity) is associated with charge redistribution or bonding changes across the metal–ligand framework.

• PCET (proton-coupled electron transfer) is associated with proton and electron transfer as a dominant constraint on interfacial process.

We assume that the three mechanisms are not exclusive. Hence, each catalyst is associated with all three mechanistic contributions but are normalized such that OA + MLC + PCET = 1.

Structural seed as the primary determinant.

Anchoring Type	Description (condensed)	OA	MLC	PCET
Covalent anchoring	Rigid binding; restricted motion; stabilizes metal-centered redox	0.7	0.2	0.1
π–π anchoring	Delocalized electronic coupling; moderate flexibility; supports conjugation-driven PCET	0.15	0.25	0.6
Non-covalent anchoring	Minimal constraints; solvent-exposed; favors proton-coupled processes	0.1	0.3	0.6
Phosphonate / chelating anchoring	Defined but adjustable coordination; balances charge redistribution and proton handling	0.25	0.3	0.45
Core–shell / interfacial	Strong electronic coupling; ligand-mediated charge sharing dominates	0.3	0.5	0.2

Environmental modifiers

Small additive adjustments were added under specific conditions.

(i) η_{10} quartiles (for each reaction class):

• lower quartile (smaller η_{10}) → PCET +0.02

• upper quartile (larger η_{10}) → OA +0.02

• middle quartiles → MLC +0.02

(ii) Tafel slope quartiles (for each reaction class):

• lower quartile (smaller slope) → PCET +0.02
• upper quartile (larger slope) → OA +0.02
• middle quartiles → MLC +0.02
(iii) pH (only if reported):
• pH ≤ 3.0 → PCET +0.02
• 5.5 ≤ pH ≤ 8.5 → MLC +0.02
• pH ≥ 10.0 → OA +0.02
(iv) Electrolyte:
• PCET +0.02 if phosphate / borate / carbonate / sulfate present
• MLC +0.02 if PF_6^- / BF_4^- / ClO_4^- present
• OA +0.02 if halides (Cl, Br, I), NO_3^-, acetate, or mineral acids (HCl, HBr, HI) present

Photoelectrochemical adjustment

Photoelectrochemical systems were treated separately because charge delivery is mediated by photogenerated carriers.
• Reaction = PEC-HER → PCET +0.05
• Reaction = PEC-OER → OA +0.05
The photoelectrochemical adjustment (+0.05) is larger than other environmental modifiers (+0.02) because illumination changes the mode of charge delivery itself.

Normalization and dominant label assignment

After combining the baseline seeds and all applicable adjustments, values were normalized:
OA = OA_raw / OA_raw + MLC_raw + PCET_raw
MLC = MLC_raw / OA_raw + MLC_raw + PCET_raw
PCET = PCET_raw / OA_raw + MLC_raw + PCET_raw
ensuring OA + MLC + PCET = 1 for every catalyst. Dominant Mechanism (Mechanistic_Dominant) was then assigned strictly as *argmax*(OA, MLC, PCET).
Reaction vs Dominant Mechanism (Figure 3a):
• HER (n = 94): OA = 35, MLC = 5, PCET = 54
• OER (n = 33): OA = 10, MLC = 2, PCET = 21
• PEC-HER (n = 25): OA = 11, MLC = 1, PCET = 13
• PEC-OER (n = 5): OA = 0, MLC = 1, PCET = 4
Metal vs Dominant Mechanism (Figure 3b):
• Co (n = 92): OA = 27, MLC = 5, PCET = 60
• Fe (n = 39): OA = 19, MLC = 1, PCET = 19
• Ni (n = 26): OA = 10, MLC = 3, PCET = 13
Anchoring_Class vs Mechanistic_Dominant:
• covalent: OA-dominant across all entries
• non-covalent: PCET-dominant across all entries
• π–π: PCET-dominant across all entries
• phosphonate/chelate: PCET-dominant across all entries

• core–shell/interfacial: MLC-dominant across all entries

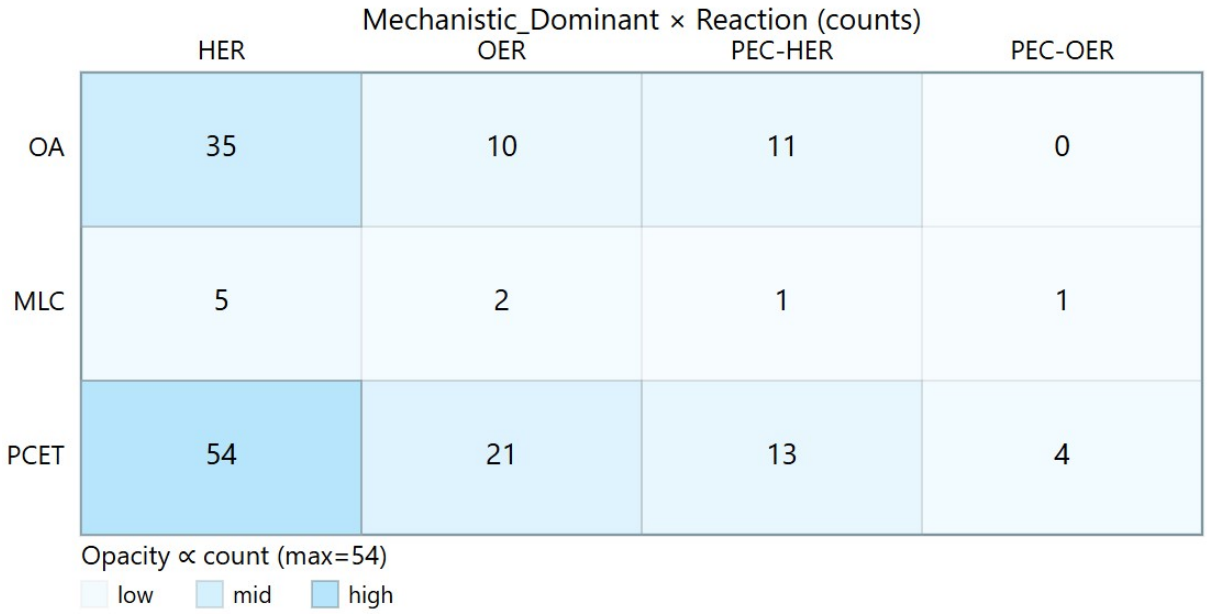


Figure S3a. Heatmap showing Dominant Mechanism vs reaction classes.

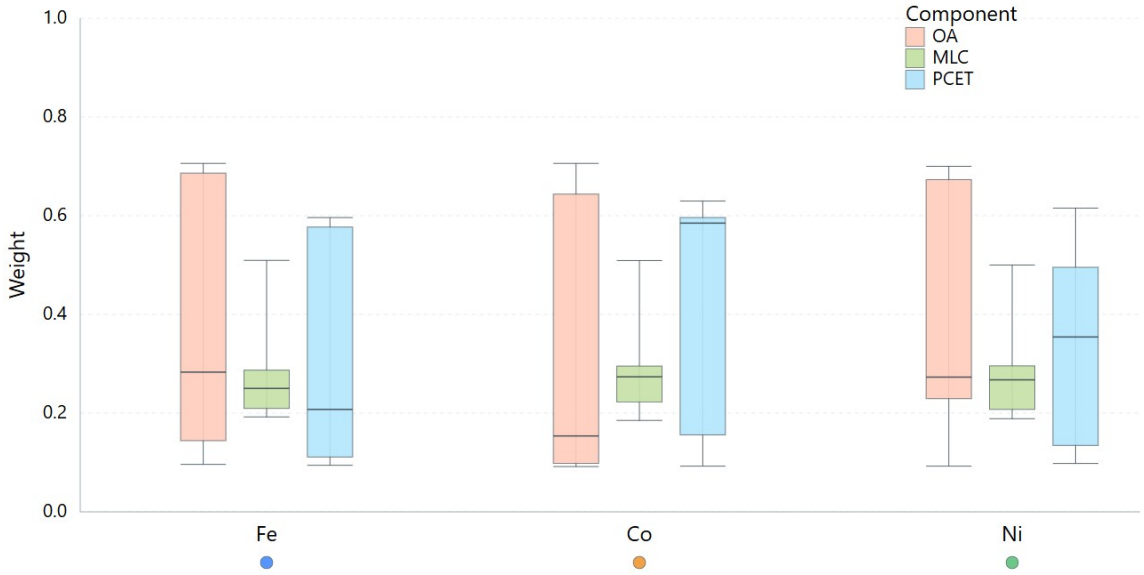


Figure S3b. OA, MLC, and PCET weights vs metal, with medians and spreads.

S4. Utility Descriptor Potentials and Rate and Stability Determining Limitations

Purpose

Section S4 was developed to explain why some anchored systems appeared more chemically interpretable, while others were associated with instability, transport issues, or interfacial constraints in the reported literature. Each system was treated as a balance between enabling features and limitations. Both exist together in anchored systems.

Origin of UDP and RSDL Values

All UDP and RSDL values were derived from source papers, as compiled in molecular explanation cards, and translated into bounded descriptor values using the fixed rules. These cards summarized the experimental and structural evidence reported in the literature, including anchoring chemistry, electrochemical behavior, spectroscopic signatures, and stability observations.

Meaning of the UDP Axes

U1: U1 described anchoring integrity. It reflected how securely a molecular catalyst remained attached to the support during operation. Covalent attachment, strong chelation, and well-defined interfacial and polymeric architectures supported higher values. At the same time loosely adsorbed films or drop-casted systems were associated support lower values.

U2: U2 described the proton environment surrounding the active site. Systems with internal proton relays, hydrogen-bond networks, and favorable electrolyte conditions scored higher. Systems that where anchoring motifs interfered with proton access, or used bulk solvent for proton delivery tended to score lower.

U3: U3 described electronic communication between the molecular catalyst and the support. Conjugated linkers, conductive substrates, and semiconductor interfaces supported higher values. Poor electronic coupling reduced this score. The axis reflected reported evidence for whether electrons could reasonably move through the interface and not related to how quickly catalysis occurred.

U4: U4 described persistence under operation. It reflected whether molecular features remained distinguishable during extended electrolysis or photoelectrochemical operation. Higher values corresponded to systems that retained molecular character, while lower values reflected degradation or restructuring.

Covalent and oxide-bound systems emphasized high anchoring integrity and coupling, while non-covalent and π–π systems emphasized environmental flexibility and electronic communication.

Meaning of the RSDL Axes

R1: R1 described the risk of deactivation. Leaching, detachment, nanoparticle formation, or rapid activity loss increased this score.

R2: R2 described electronic bottlenecks present in the system. Large charge-transfer resistances, high Tafel slopes, and misaligned energy levels contributed to higher values. This limitation reflected reported interfacial resistance.

R3: R3 described structural instability. Ligand oxidation, demetallation, corrosion, or film restructuring increased this score. This limitation was especially prominent under oxidative conditions.

R4: R4 described accessibility constraints. Thick films, dense packing, or restricted diffusion pathways increased this score.

Dataset-Level Structure of UDP

When averaged by anchoring type, clear patterns appeared

- core–shell systems showed high values across all UDP axes
- covalent systems have strong anchoring (U1) but lower proton environment (U2) and persistence (U4)

- non-covalent systems show moderate values across all axes
- phosphonate/chelating systems emphasize proton handling and electronic coupling
- π–π systems fall in intermediate positions

Dataset-Level Structure of RSDL

Limitations vary more with reaction type:

- HER systems show moderate limitations
- structural limitation (R3) is highest for HER and PEC-HER, lower for OER, lowest for PEC-OER

- photoelectrochemical systems show different patterns due to light and semiconductor effects

Combined Enablement and Limitation

Sum_U = U1 + U2 + U3 + U4 → total number of enabling features

Sum_R = R1 + R2 + R3 + R4 → total number of limitations

Sum_U summarized how many enabling features were represented and Sum_R captured the cumulative burden of rate- and stability-limiting constraints acting on a system. Systems with high Sum_U often revealed Sum_R, whereas systems with low Sum_R did not necessarily exhibit high Sum_U.

Dominant Pathways

For each catalyst, the highest UDP and RSDL value is taken as the dominant feature.

- covalent systems: anchoring (U1) often dominant; structural limits (R3) common
- non-covalent systems: electronic and environmental effects more dominant; deactivation and transport limits common

If multiple values were equal, the first in order was chosen (U1→U4; R1→R4).

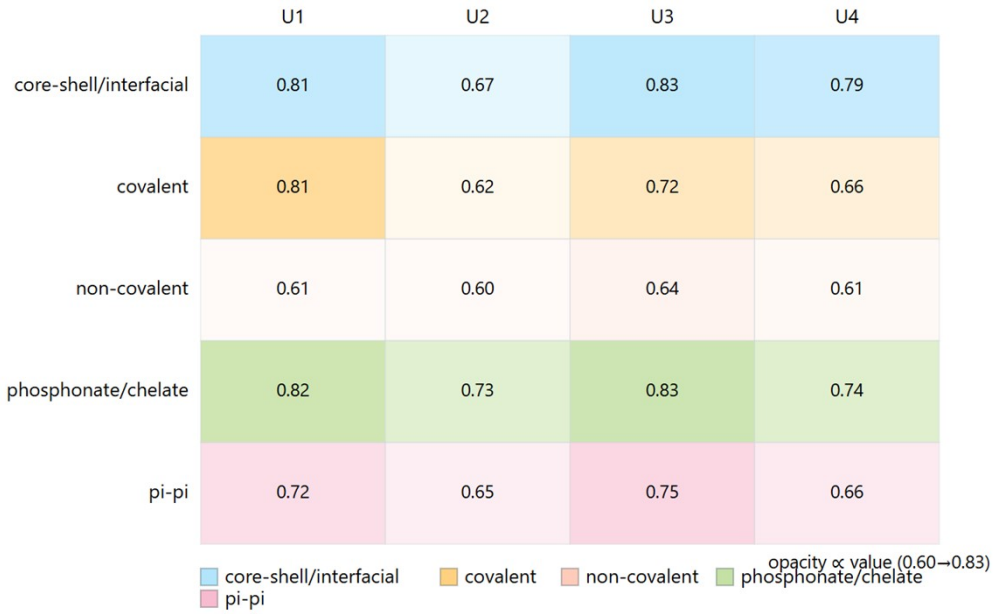


Figure S4a. U1–U4 values across anchoring motifs.

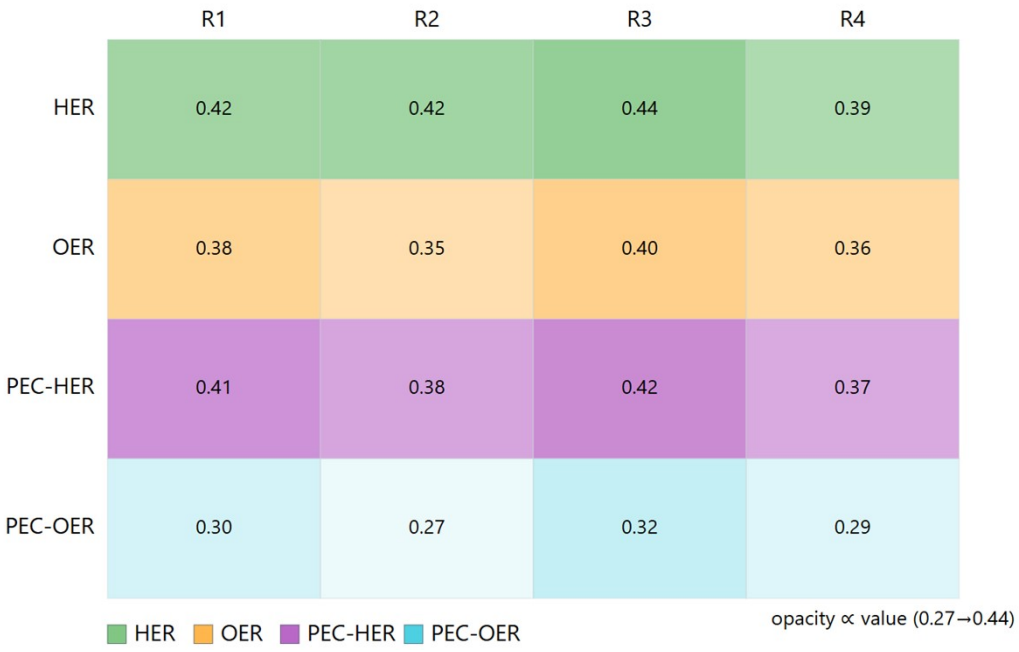


Figure S4b. R1–R4 values resolved by reaction types.

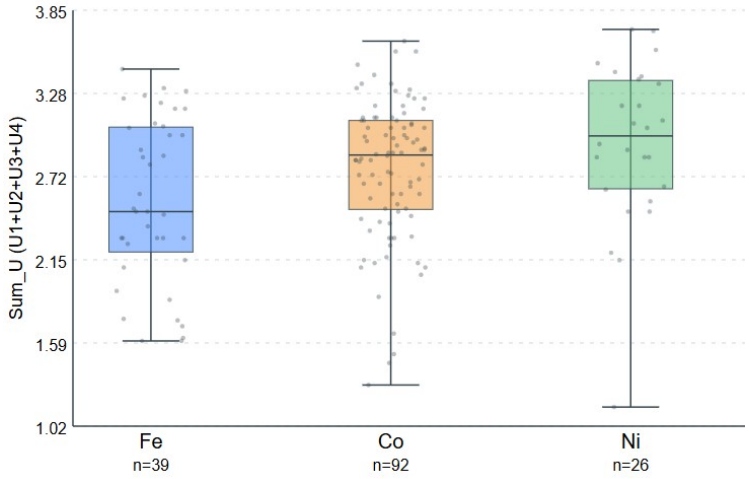


Figure S4c. Sum_U resolved by metal identities.

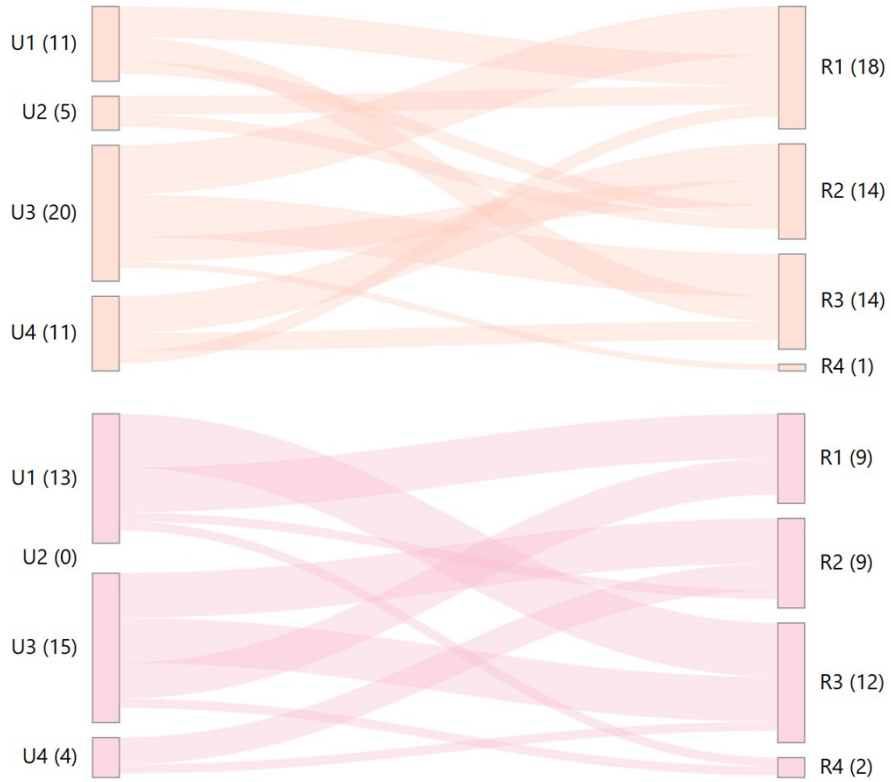


Figure S4d. Dominant UDP and RSDL axes across anchoring classes (covalent, yellow, n = 56; non-covalent, peach, n = 47; π–π, pink, n = 32).

S5. Integration of Design Logic and Evidence Completeness

Purpose

Earlier sections described the catalysts and their features. Those descriptors remained incomplete without accounting for how well they were supported by experimental evidence. S5 was used to check on how well these claims are supported by data. The role of S5 was to limit interpretation based on available evidence. Strong design ideas were not treated as reliable unless they are well supported.

Definition of evidence completeness

Evidence completeness in S5 was grouped into four orthogonal dimensions that captured different aspects of experimental support.

- **SC (structural):** how well the catalyst structure is confirmed
- **EC (electrochemical):** how much electrochemical data is reported
- **MC (mechanistic):** how much the reaction pathway is studied
- **AC (anchoring):** how clearly the immobilization is shown

Low evidence in any category did not mean poor chemistry. In many cases, it reflected differences in experimental details or publication norms rather than deficiencies in catalyst design.

Assignment logic for SC, EC, MC, and AC.

Each evidence score was assigned by reading the original papers. Scores were not binary. Intermediate values were used when evidence is partial but credible.
Example: if structure was confirmed, but surface binding was not directly shown → moderate SC
Example: if mechanism was suggested but not tested → moderate MC
All assignments follow fixed rules and were applied consistently. No adjustments were made later. Each score was based on specific statements, figures, or comparisons in the original reports. An independent reader can understand how the values were assigned.

Construction of composite evidence completeness (ECOMP).

Evidence from SC, EC, MC, and AC is combined into one value (ECOMP) using a geometric mean:

- low score in any one part reduces the overall value
- but it does not go to zero unless one part is zero

ECOMP does not measure performance. It only shows how evenly supported the evidence is.
A combined score is then defined:

- **DesignEvidenceScore = (Sum_U + Sum_R) × ECOMP**

Sum_U and Sum_R are combined with ECOMP to give a single score that reflects how well the design is supported by evidence.

- strong design + weak evidence → reduced score
- strong evidence cannot fix weak design

This score reflects balance, not optimization.

Dataset-level evidence completeness.

Across the dataset, some patterns appear:

- covalent systems tend to have higher structural and anchoring evidence
- core–shell systems show the highest structural scores
- electrochemical and mechanistic evidence vary widely across all types

Photoelectrochemical systems show different patterns, but PEC-OER has very few cases (n = 5), so conclusions are limited.

Reaction-resolved evidence and composite outcomes

When grouped by reaction, hydrogen evolution systems showed a relatively consistent spread of evidence, OER showed larger variation in reporting less uniform mechanistic documentation and photoelectrochemical systems showed uneven evidence because electrochemical, photophysical, and interfacial evidence were often reported unevenly.
Combined design–evidence score shows overlapping distributions across systems. This suggests strong conclusions require both design and sufficient evidence, not just high values alone

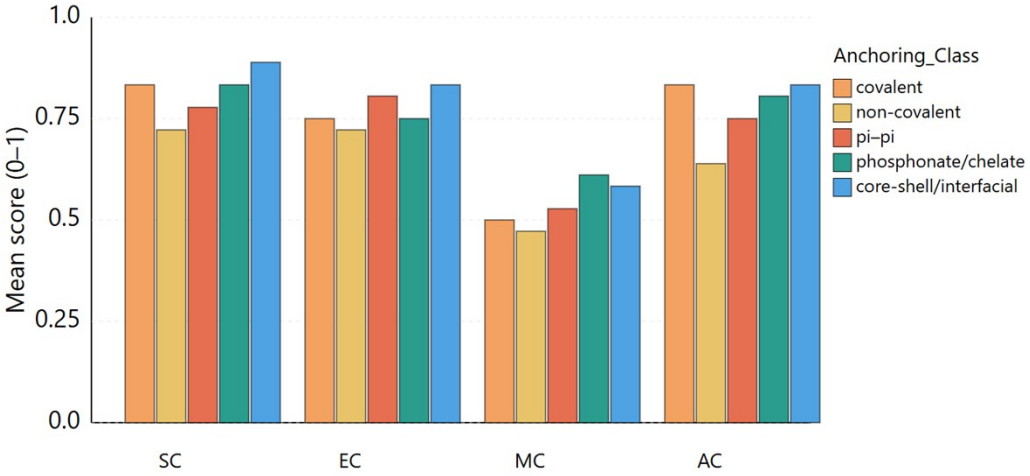


Figure S5a. Mean evidence completeness scores grouped by anchoring class.

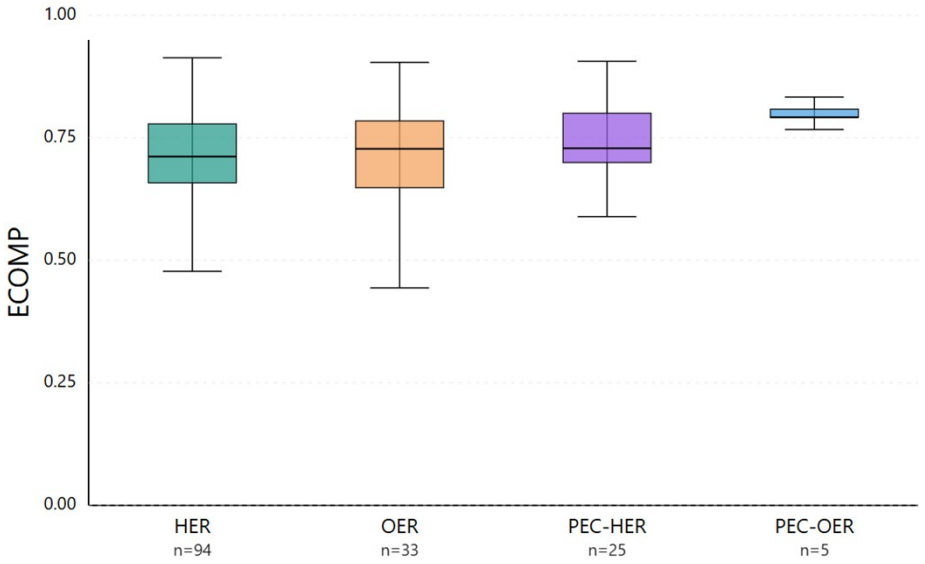


Figure S5b. Distributions of composite evidence completeness resolved by reaction class.

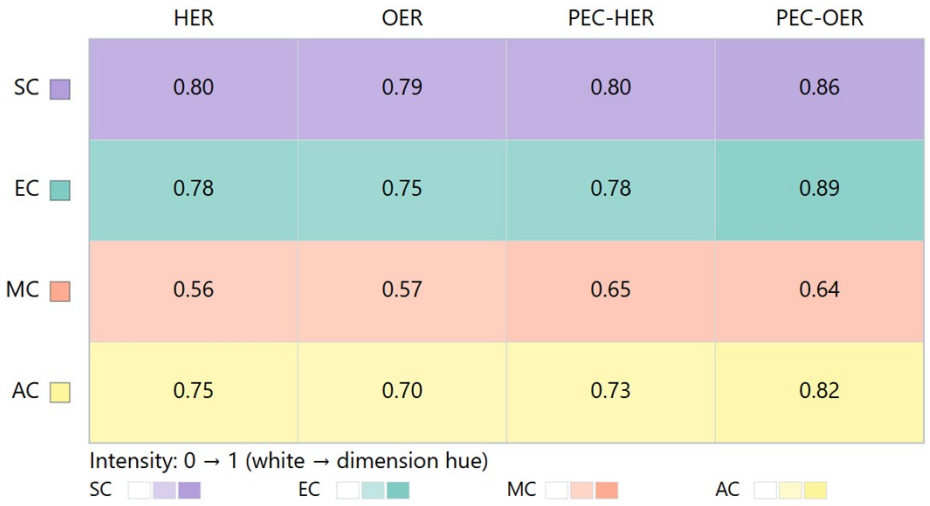


Figure S5c. Heatmap of average evidence completeness across reactions and evidence dimensions.

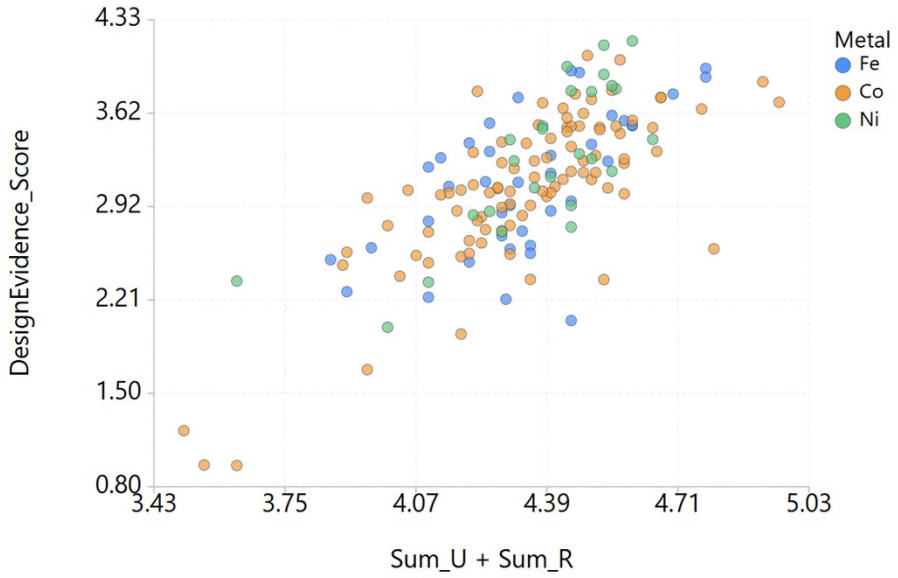


Figure S5d. Distribution of composite design–evidence scores plotted against Sum_U+Sum_R. Substantial overlap exists across reactions and anchoring classes.

S6. Experimental Regime Descriptors

Earlier sections were used for identity, anchoring, reaction type, and descriptor values. However, these values are incomplete without the experimental conditions. Electrocatalysis depends on pH, solvent, and illumination. S6 records this context. The purpose is to avoid invalid comparisons between systems studied under different conditions (for example HER vs PEC-HER). S6 describes the chemical environment of the reaction, not performance. Key differences include basic vs non-basic conditions, aqueous vs organic solvent, and presence of light. These factors change proton availability, charge transport, and interfacial behavior. As a result, these factors are treated as separated categories and not continuous variables. Each catalyst is assigned in a way such that it clearly showed how the experiment was done rather than how the catalyst had performed.

Assignment of Base as a descriptor

The Base descriptor indicates whether the system was studied in strongly basic aqueous conditions. It was assigned using reported pH values or clear use of hydroxide electrolytes (e.g., KOH, NaOH) from S2. In some cases, strong basicity was assigned based on clear chemical context even if pH was not reported. These were limited in number and are recorded in the supplementary dataset.

Assignment of Solvent

The solvent descriptor was used to separate aqueous systems from organic or mixed media. If any organic solvent was present, we classified them as non-aqueous, even if water was used as a co-solvent or pH was also reported. Mixed systems were grouped with organic media.

Assignment of Illumination.

The Illumination descriptor recorded whether the experiment was done under light. It separated photoelectrochemical systems from purely electrochemical ones. It is independent of pH, solvent, and

electrolyte, and does not indicate efficiency or performance. It only shows whether light was used in the experiment.

Across the dataset, the regime descriptors showed clear patterns. Most systems were studied in aqueous media, with only a small number in organic or mixed solvents. Strongly basic conditions were less common and appear across Fe, Co, and Ni systems with no apparent preference to a single metal.

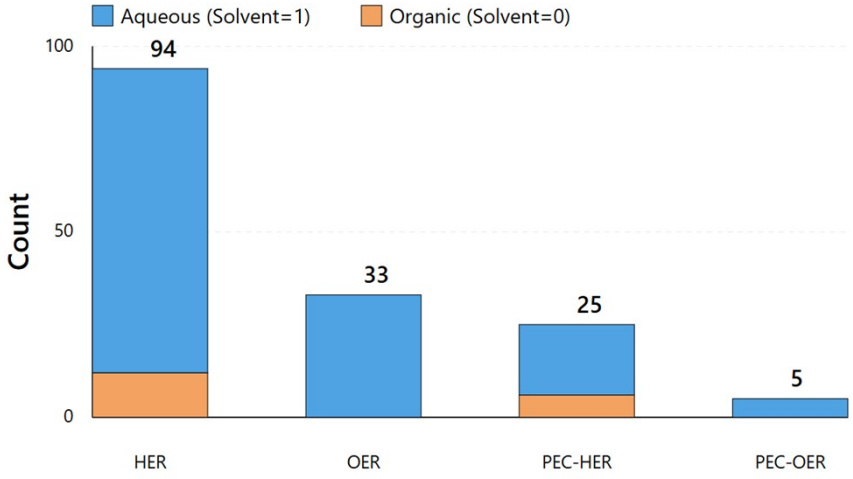


Figure S6a. Distribution of solvent across reaction classes.

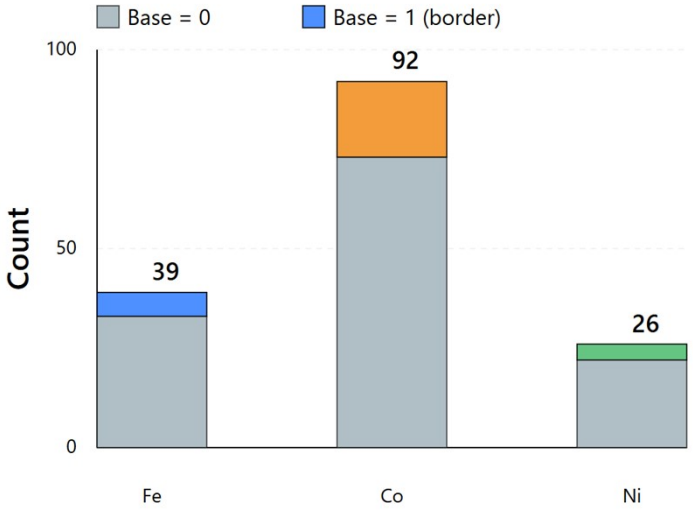


Figure S6b. Use of strongly basic aqueous conditions across metals.

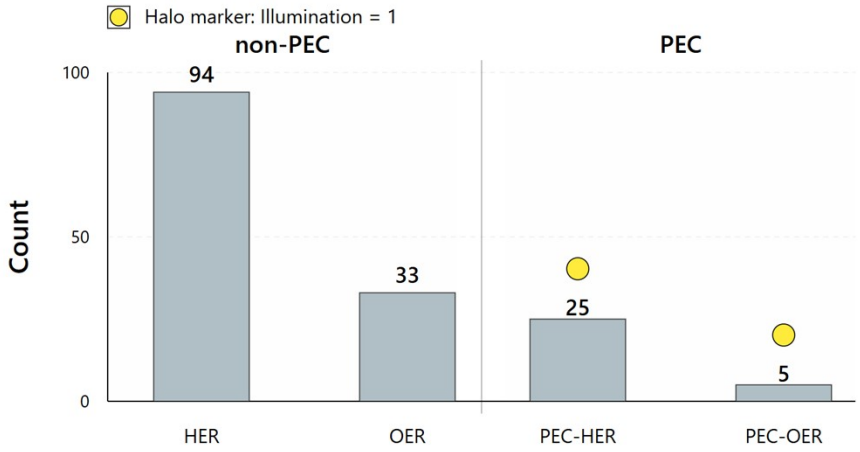


Figure S6c. Electrochemical systems vs photoelectrochemical systems.

Interpretive limits

As discussed, S6 provides context for the reactions. These 2 parameters ensure that comparisons are made between systems studied under similar conditions. All descriptors (Sum_U, Sum_R, ECOMP) are defined independently of these regimes. However, the dataset itself is unevenly distributed across these conditions.

Sum_U and Sum_R show a strong negative diagonal relationship (Pearson r = −0.8959; Spearman ρ = −0.8918). This pattern remains even if individual components are removed, so it is not driven by a single factor. Removal of R1 (Pearson r ≈ −0.8820; Spearman ρ ≈ −0.8801), and U1 (Pearson r ≈ −0.8964; Spearman ρ ≈ −0.8950) does not reduce the correlation. Both values come from the same underlying data and rules, so this pattern reflects how the dataset is built, and not a fundamental physical constraint. Anchoring effects are treated explicitly, while other effects are included indirectly through descriptors. As a result, interpretation applies to the full anchored system instead of an isolated molecule. Sum_U, Sum_R, and ECOMP are defined independently, but they often increase together because well-studied systems report more details. This reflects reporting depth, not true dependence. S6 defines the limits of interpretation. The results must be read within the experimental context, and observed patterns should not be overinterpreted beyond the dataset structure.

S7. Experimental Regime Organization of the Mechanistic Dataset

Purpose

Sections S1–S6 define all descriptors (identity, anchoring, environment, mechanism, limitations, evidence). Once these values were fixed, S7 was used to examine all these separate descriptors together across the dataset.

Variables used to construct the landscape

The two-dimensional plot is constructed using

$$\text{Sum\_U} = \sum_{i=1}^{4} U_i$$

$$\text{Sum\_R} = \sum_{i=1}^{4} R_i$$

where U1–U4 correspond to the Utility–Descriptor Potentials and R1–R4 correspond to the Rate- and Stability-Determining Limitations as discussed in S4. In other words, Sum_U counts how many enabling features are reported for a catalyst after anchoring and Sum_R counts how many limitations are reported.Neither Sum_U nor Sum_R includes catalytic activity, efficiency, selectivity, or performance outcome. They are derived solely from reported mechanistic and interfacial features.

In addition to these two axes, evidence completeness as discussed in S5 was combined to produce the raw order

Order_raw = (Sum_U + Sum_R) × ECOMP

The raw order was further used to derive a normalized ordering coordinate (Order Index),

Reference boundaries and quadrant definition

The dataset was divided using the median values of Sum_U and Sum_R:

- median(Sum_U) = 2.85
- median(Sum_R) = 1.58

These median values are just thresholds and carry no chemical meaning. They are only used to split the dataset into contrasting mechanistic utility and mechanistic penalty regions.

Values equal to the median were placed in the higher group. The quadrants were defined as:

Q1 (high utility, high penalty): Sum_U ≥ 2.85 and Sum_R ≥ 1.58

Q2 (high utility, low penalty): Sum_U ≥ 2.85 and Sum_R < 1.58

Q3 (low utility, high penalty): Sum_U < 2.85 and Sum_R ≥ 1.58

Q4 (low utility, low penalty): Sum_U < 2.85 and Sum_R < 1.58

Distribution of systems in the U–R space

Figure S7a plots all 157 catalysts using Sum_U (x-axis) and Sum_R (y-axis). Each point is one catalyst. Median lines at 2.85 (Sum_U) and 1.58 (Sum_R) divide the plot into four quadrants.

The distribution of systems across quadrants is non-uniform.

Q1 = 13, **Q2** = 68, **Q3** = 68, **Q4** = 8

Most reported systems fall within **Q2** and **Q3**, while **Q1** and **Q4** contain relatively few entries.

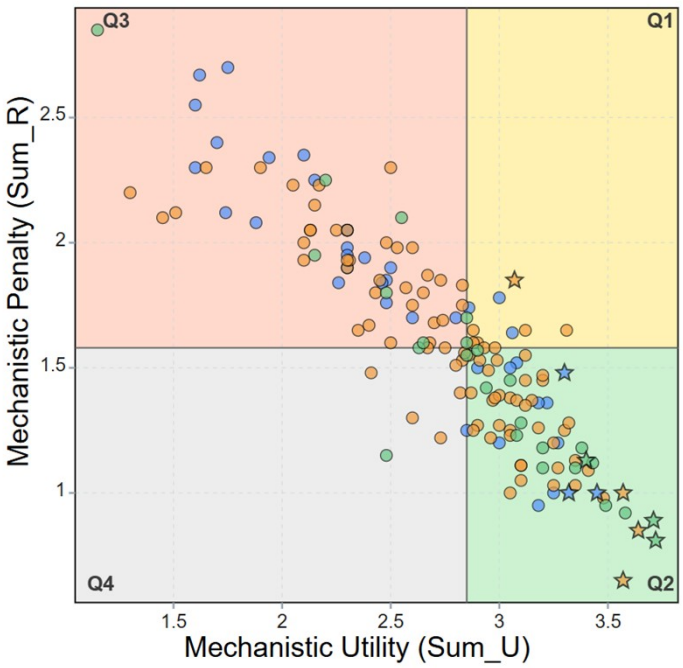


Figure S7a. Distribution of systems in the U–R space

Meaning of quadrant placement

Quadrant placement describes the balance between mechanistic utility (Sum_U) and articulated mechanistic penalty (Sum_R) relative to the dataset medians.

Systems in **Q1** are report multiple enabling features alongside multiple explicitly documented limitations.

Systems in **Q2** reflects fewer reported limitations relative to enabling features, rather than the absence of interfacial constraints.

Systems in **Q3** reported limitations exceed enabling features within the descriptor space.

Systems in **Q4** reflect sparse descriptor expression in both directions within the reported literature.

Reaction resolved quadrant placement

The reaction-resolved quadrant counts are displayed in Figure S7b

HER (n = 94): **Q1** = 9, **Q2** = 36, **Q3** = 46, **Q4** = 3

OER (n = 33): **Q1** = 0, **Q2** = 16, **Q3** = 13, **Q4** = 4

PEC-HER (n = 25): **Q1** = 4, **Q2** = 11, **Q3** = 9, **Q4** = 1

PEC-OER (n = 5): **Q1** = 0, **Q2** = 5, **Q3** = 0, **Q4** = 0

Metal resolved quadrant placement

The metal-resolved quadrant counts are displayed in Figure S7c

Cobalt (n = 92): **Q1** = 8, **Q2** = 39, **Q3** = 38, **Q4** = 7

Iron (n = 39): **Q1** = 3, **Q2** = 13, **Q3** = 23, **Q4** = 0

Nickel (n = 26): **Q1** = 2, **Q2** = 16, **Q3** = 7, **Q4** = 1

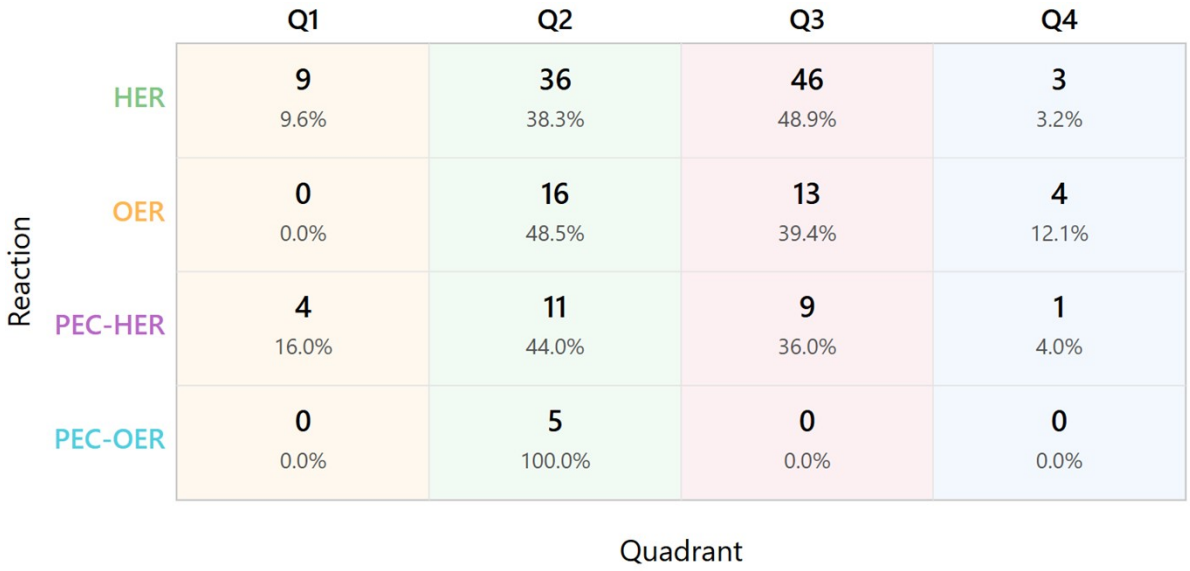


Figure S7b. Reaction resolved quadrant placement

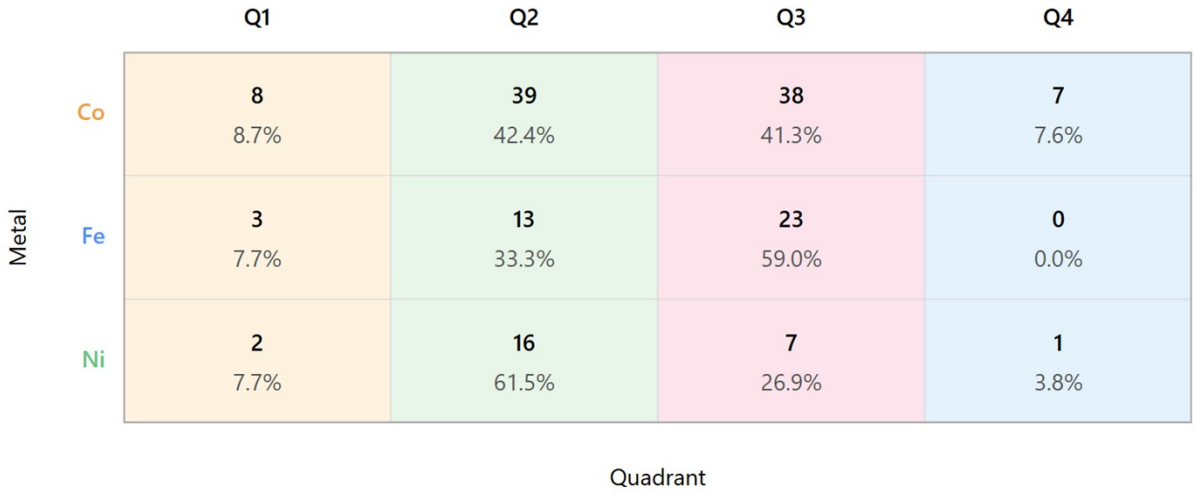


Figure S7c. Metal resolved quadrant placement

Order Index definition and distribution

Order Index is a single value that combines utility (Sum_U), limitation (Sum_R), and evidence (ECOMP).

Raw order is scaled between 0 and 1 using the lowest and highest values in the dataset:

- minimum = 0.9584
- maximum = 4.1696

So, the lowest system gets 0 and the highest gets 1.

The values are spread continuously, so fixed ranges are used to group them:

- **High:** ≥ 0.85 (25 systems)
- **Mid:** 0.70–0.85 (47 systems)
- **Low:** < 0.70 (85 systems)

These groups are only for organization and do not indicate performance differences.

Order Index is a derived, dataset-level quantity that summarizes descriptor resolution by combining mechanistic utility, mechanistic penalty, and evidence completeness into a single normalized scalar. It is constructed deterministically from frozen upstream values and does not encode catalytic activity, efficiency, or outcome.

Evidence completeness is quantified upstream in Section S5 as ECOMP. The raw order value is defined as the DesignEvidenceScore, given by:

DesignEvidenceScore = Order_raw = (Sum_U + Sum_R) × ECOMP

Order Index is obtained by linear min–max rescaling of Order_raw across the full frozen dataset (n = 157) according to:

Order Index(i) = (Order_raw(i) − min(Order_raw)) / (max(Order_raw) − min(Order_raw))

For the frozen dataset, the extrema are:

min(Order_raw) = 0.958441681

max(Order_raw) = 4.169606142

This normalization maps the lowest-order system in the dataset to Order Index = 0 and the highest-order system to Order Index = 1, with all other systems scaled linearly between these bounds.

Figure S7d shows the distribution of Order Index values as a histogram. Order Index spans a continuous range across the dataset under the present normalization, and fixed thresholds are therefore used to define coarse descriptive tiers rather than inferred clusters. For descriptive convenience, Order Index values were grouped into three coarse tiers using fixed thresholds:

High: Order Index ≥ 0.85 (n = 25)

Mid: 0.70 ≤ Order Index < 0.85 (n = 47)

Low: Order Index < 0.70 (n = 85)

These tiers are used only for organizational and visualization purposes and do not imply chemical discontinuities or performance-based ranking.

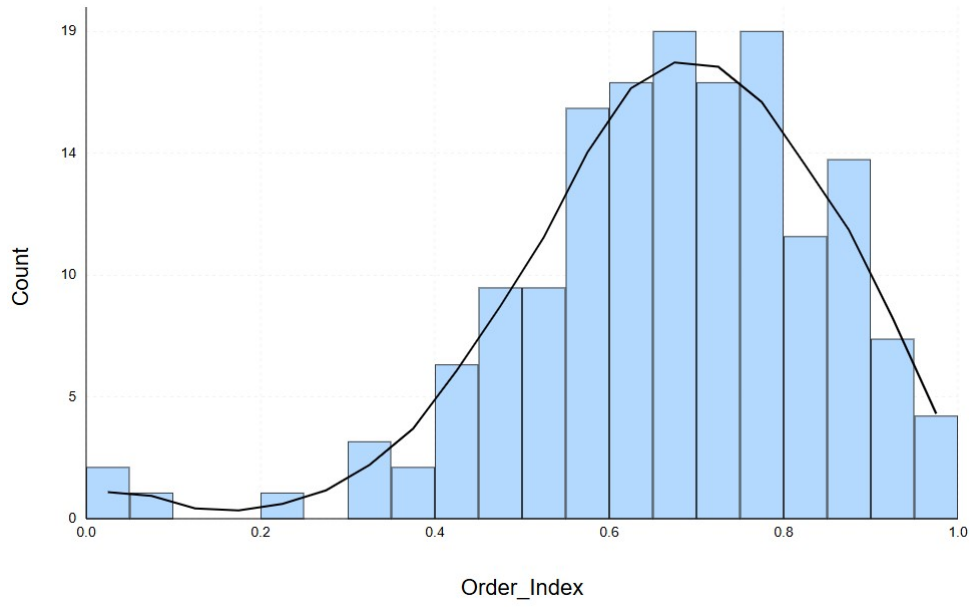


Figure S7d. Distribution of Order Index values across the dataset.

Relationship between Order Index and quadrant placement

Quadrants and Order Index are defined separately.

- Quadrants depend only on Sum_U and Sum_R
- Order Index depends on Sum_U, Sum_R, and ECOMP

Systems from all Order Index tiers appear in multiple quadrants. High-order and low-order systems are both spread across the dataset, including in high-utility regions.

This shows that quadrant position reflects balance (utility vs limitation), while Order Index reflects how well the system is described. The two are not linked by construction.

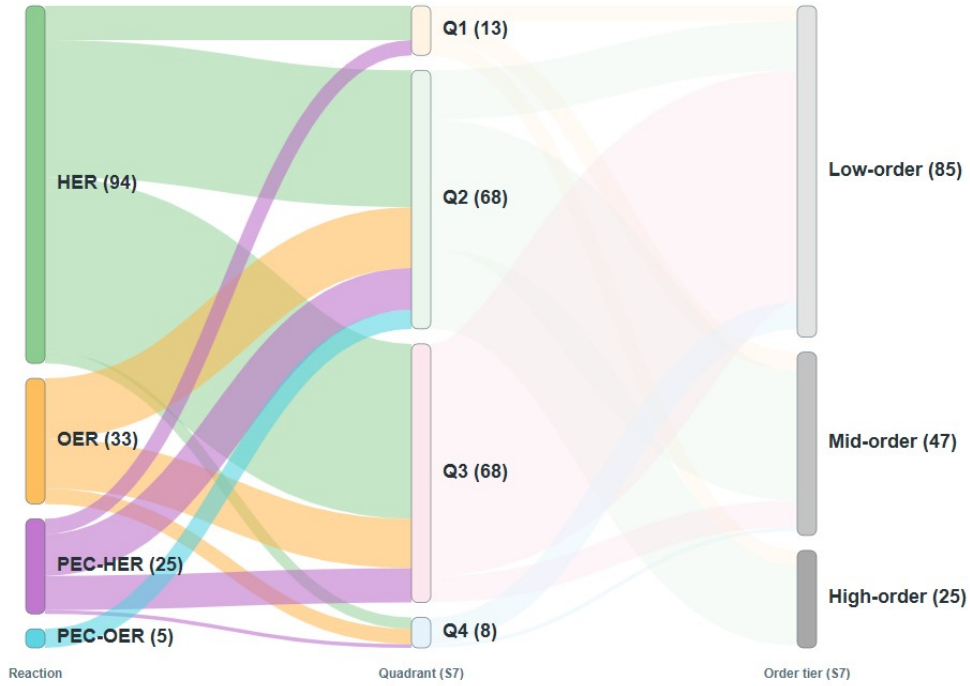


Figure S7e. Systems distributed across reaction type, quadrants, and Order Index tiers using a Sankey diagram. Box sizes show number of systems in each group and the link widths show counts between groups.

Manuscript case-study catalysts

Ten selected case-study systems are highlighted (Exemplar_Flag = 1 in S7.csv). Nine of these systems are in **Q2** quadrant and one in **Q1**. All the 10 catalysts belong to the High Order Index tier.

Scope and limits of interpretation

S7 organizes how utility, limitation, and evidence appear together in the dataset. It does not rank systems, predict performance, or guide design.

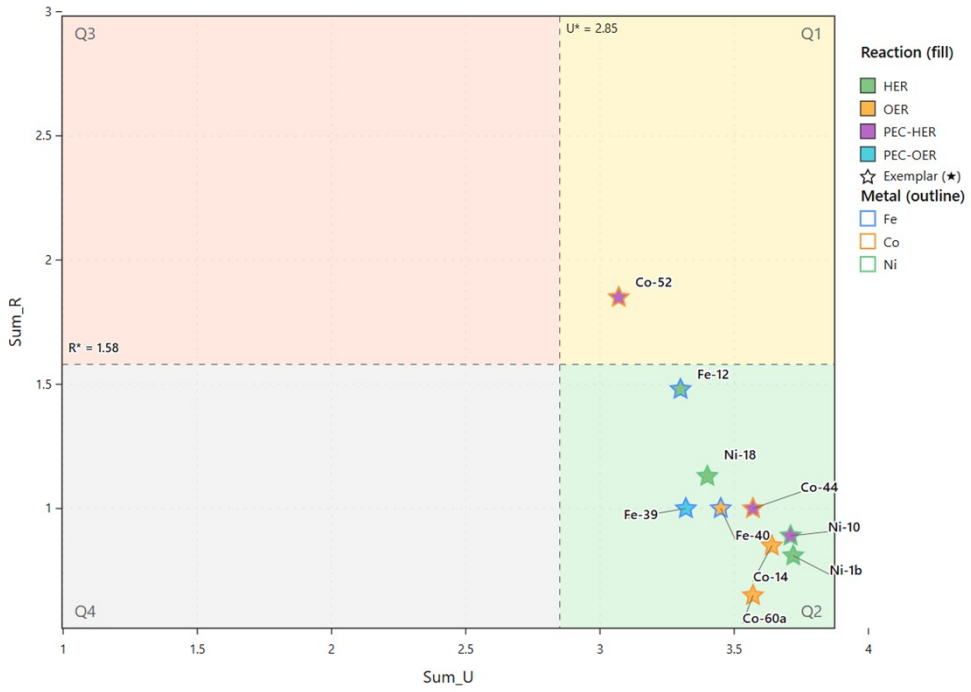


Figure S7f. Exemplar systems in U–R space.

S8. Quadrant-Level Enrichment and Ordering Statistics in the U–R Space

For any variable (reaction class, anchoring class, metal identity), quadrant-level enrichment was analyzed as a normalized fraction rather than individual counts. This is because population of each quadrant is not same.

For a category **c** and quadrant **q**, enrichment was calculated number of systems of type **c** in quadrant **q**, divided by the total number of systems of type **c** in the dataset.

This clearly establishes that enrichment reveals where a category is concentrated. Enrichment values were additionally normalized within each categorical family by dividing by the maximum enrichment observed across quadrants. Thus, the final values were in the range [0,1]. For reaction class, anchoring class, and metal identity, only the dominant enriched category per quadrant was reported. However, mechanistic channels are shown as complete distributions because their fractional weights already sum to unity for each system.

Reaction-class enrichment

- **Q1:** PEC-HER
  - Enrichment: 4/25 = 0.16
- **Q2:** PEC-OER
  - Enrichment: 5/5 = 1.00
- **Q3:** HER
  - Enrichment: 46/94 ≈ 0.489
- **Q4:** OER
  - Enrichment: 4/33 ≈ 0.121

Anchoring-class enrichment

- **Q1:** π–π anchoring
    - Enrichment: 5/32 ≈ 0.156
- **Q2:** Core–shell / interfacial anchoring
    - Enrichment: 7/9 ≈ 0.778
- **Q3:** Non-covalent anchoring
    - Enrichment: 31/47 ≈ 0.660
- **Q4:** π–π anchoring
    - Enrichment: 3/32 ≈ 0.094

Metal-identity enrichment by quadrant

- **Q1:** Co
    - Enrichment: 8/92 ≈ 0.087
- **Q2:** Ni
    - Enrichment: 16/26 ≈ 0.615
- **Q3:** Fe
    - Enrichment: 23/39 ≈ 0.590
- **Q4:** Co
    - Enrichment: 7/92 ≈ 0.076

Mechanistic channel aggregation

Each catalyst is assigned fractional contributions (OA, MLC, PCET) that add up to 1. For each quadrant, these values are averaged across all catalysts in that quadrant to get the mean contribution of each mechanism.

- **Q1 (n = 13):**
    - OA ≈ 0.371
    - MLC ≈ 0.258
    - PCET ≈ 0.371
    - Dominant channel: OA (tie resolved by predefined priority rules)
- **Q2 (n = 68):**
    - OA ≈ 0.380
    - MLC ≈ 0.270
    - PCET ≈ 0.350
    - Dominant channel: OA
- **Q3 (n = 68):**
    - OA ≈ 0.306
    - MLC ≈ 0.264
    - PCET ≈ 0.430
    - Dominant channel: PCET
- **Q4 (n = 8):**
    - OA ≈ 0.259

- MLC ≈ 0.263
- PCET ≈ 0.478
- Dominant channel: PCET

DesignEvidenceScore aggregation

For each quadrant, the DesignEvidenceScore (i.e., Order_raw) is averaged across all systems and normalized.

- **Q1:**
  - Mean = 3.38
  - Normalized = 0.85
- **Q2:**
  - Mean = 3.50
  - Normalized = 1.00
- **Q3:**
  - Mean = 2.70
  - Normalized = 0.00
- **Q4:**
  - Mean = 2.87
  - Normalized ≈ 0.21

S9. Robustness of Order Index

Purpose and rationale

The U–R geometry separates quadrant assignment from an orthogonal layer that classifies catalysts into High, Mid, and Low interpretability tiers. Quadrant membership depends only on fixed medians defined by Sum_U and Sum_R values. In contrast, the ordering layer additionally adds evidence completeness, yielding a continuous composite scalar (Order_raw) that could, in principle, be sensitive to numerical variability.

Perturbation protocol

To simulate uncertainty the raw order value for each catalyst was perturbed multiplicatively according to:

$$\text{Order\_raw}' = \text{Order\_raw} \times (1 + \delta)$$

where the perturbation factor δ was drawn independently for each catalyst from a uniform distribution: $\delta \sim \text{Uniform}(-\varepsilon, +\varepsilon)$

Three amplitudes were tested:

- ε = 0.05 (±5%)
- ε = 0.10 (±10%)
- ε = 0.15 (±15%)

For each ε value, 1000 independent Monte Carlo runs were performed. In each run:

1. All catalysts were perturbed simultaneously using independent random δ values.

2. The perturbed values were renormalized to produce a new Order Index.
3. Order tiers (High, Mid, Low) were reassigned using the same definitions as in the baseline analysis.
4. Tier assignments were compared against the unperturbed tier labels.

No other quantities such as U1–U4, R1–R4, evidence completeness scores, quadrant boundaries and medians were modified.

Quantification of tier transitions

For each perturbation run, tier-to-tier transitions were recorded for every catalyst. While six transitions are possible in principle (High→Mid, High→Low, Mid→High, Mid→Low, Low→Mid, Low→High), all transitions were recorded, with emphasis placed on those that could directly contradict/undermine interpretability ordering

- **Downward transitions:** High→Mid, High→Low, Mid→Low
- **Upward transitions:** Low→High

Results

The observed transition rates are summarized below.

ε = 0.05 (±5%)

- High → Mid: 4.52%
- High → Low: 0.00%
- Mid → Low: 5.01%
- Low → High: 0.00%

At small perturbation amplitudes, downward tier changes were limited to High→Mid and Mid→Low transitions. No direct High→Low or Low→High events observed.

ε = 0.10 (±10%)

- High → Mid: 7.96%
- High → Low: 0.21%
- Mid → Low: 11.78%
- Low → High: 0.00%

At moderate perturbation amplitudes, High-tier retention remained above 90%, and no Low-tier system transitioned directly into the High tier.

ε = 0.15 (±15%)

- High → Mid: 7.26%
- High → Low: 2.40%
- Mid → Low: 15.74%
- Low → High: 0.00%

Even under aggressive ±15% perturbation of the raw order scalar, approximately 90% of High-tier assignments were preserved, and no Low→High transitions were observed.

Interpretation

(i) **High-tier stability.** Across all amplitudes, High-tier systems show high retention. Even at ε = 0.15, less than 10% of High-tier assignments change, and direct High→Low transitions remain rare.
(ii) **Low-tier stability.** No Low-tier system transitions directly into the High tier at any amplitude that was tested.
(iii) **Instability near tier boundaries.** The largest transition rates involve Mid→Low changes. Thus, Mid-tier acts as a boundary buffer rather than a stable category.
(iv) **Decoupling from geometric structure.** Quadrant assignment depends only on utility–penalty medians. Hence, perturbation does not affect quadrant membership.

Implications for the Ordering Layer within the Frozen U–R Geometry

These results support the use of Order Index tiering as a secondary interpretive layer that complements the fixed quadrant structure.

S10. Forward Validation and Boundary-Distance Analysis

Purpose and scope

This section documents the forward projection of newly reported catalysts onto the frozen U–R landscape and quantifies their proximity to quadrant boundaries (Fig. S10a).[2-4] The analysis examines whether systems reported after construction of the frozen U–R geometry are consistent with the existing geometric structure, without modification of descriptors, medians, thresholds, or extraction rules.

Medians and quadrant boundaries	
Parameter	**Value**
Aggregate mechanistic utility median (U)	2.85
Aggregate mechanistic penalty median (R)	1.58
U boundary	Crossing U = 2.85 (left/right change)
R boundary	Crossing R = 1.58 (up/down change)

Raw placements of forward-projected catalysts

Eight catalysts (n = 8) were projected into U–R space using the same medians (U = 2.85, R = 1.58) using identical extraction rules. Seven of these systems were published after construction of the original dataset, and one system (**Co-91**) was omitted from the original review despite being available at the time.

Raw placement of forward-projected catalysts in U–R space					
Catalyst	**Reaction**	**Anchoring**	**Sum_U**	**Sum_R**	**Quadrant**
Fe-41	HER	covalent	2.9	1.8	**Q1**
Fe-42	HER	non-covalent	2.6	2.05	**Q3**
Fe-43	HER	non-covalent	2.6	2.14	**Q3**
Fe-44	HER	pi-pi	2.78	1.88	**Q3**
Fe-45	HER	pi-pi	2.78	1.86	**Q3**
Fe-46	HER	pi-pi	2.78	1.89	**Q3**

Fe-47	HER	pi-pi	2.75	1.97	**Q3**
Co-91	PEC-HER	interfacial	2.78	1.86	**Q3**

All forward-projected catalysts fall within the region occupied by the locked U–R space under the present dataset construction.

Boundary sensitivity analysis

To quantify forward placement stability, the minimum chemical shift required to alter quadrant assignment was computed for each system.

Definitions

- ΔU = Sum_U − 2.85
- ΔR = Sum_R − 1.58
- **MinShift** = min(|ΔU|, |ΔR|)

Boundary sensitivity of forward-projected catalysts			
Catalyst	**ΔU**	**ΔR**	**MinShift**
Fe-41	0.05	0.22	0.05
Fe-42	-0.25	0.47	0.25
Fe-43	-0.25	0.56	0.25
Fe-44	-0.07	0.3	0.07
Fe-45	-0.07	0.28	0.07
Fe-46	-0.07	0.31	0.07
Fe-47	-0.1	0.39	0.1
Co-91	-0.07	0.28	0.07

Condensed boundary-distance summary	
Catalyst	**Minimum chemical shift required to flip quadrant**
Fe-41	0.05
Fe-42	0.25
Fe-43	0.25
Fe-44	0.07
Fe-45	0.07
Fe-46	0.07
Fe-47	0.1
Co-91	0.07

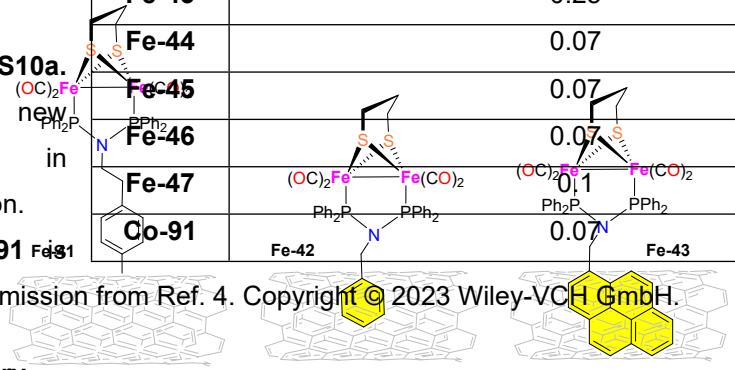
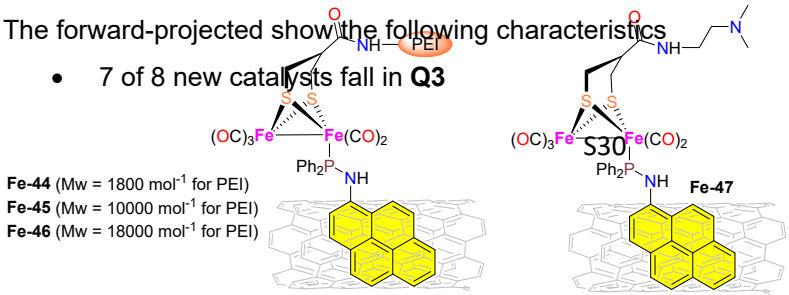


Figure S10a. Structure of eight new catalysts used in forward validation. The structure of **Co-91** is reproduced with permission from Ref. 4. Copyright © 2023 Wiley-VCH GmbH.

Summary

The forward-projected show the following characteristics

- 7 of 8 new catalysts fall in **Q3**



- 1 catalyst (**Fe-41**) lie in **Q1**
- 6 of 8 catalysts lie within 0.10 of a quadrant boundary
- Smallest boundary distance is 0.05
- Largest boundary distance is 0.25

These values describe geometric proximity only.

Relationship to the frozen U–R space

All forward placements reported here were obtained without modifying the medians, quadrant definitions, descriptor values, or extraction rules. The analysis shows consistency of newly reported systems with the existing U–R space without recalibrating or extending it.

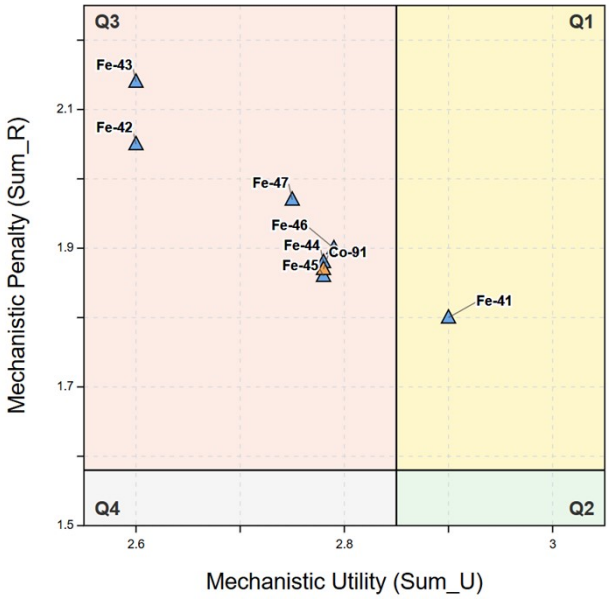


Figure S10b. Forward-projected catalysts mapped onto the U–R space shown in a zoomed view for clarity.

S11. Statistical analysis reported in the manuscript

This section reports the statistical analyses used in the manuscript, all carried out on the fixed S7 dataset. The analyses followed fixed rules and did not involve any parameter tuning.

Software and statistical conventions

All analyses were performed in Python using pandas, numpy, and scipy.stats. All tests were two-tailed and used a significance level of $\alpha = 0.05$.

Variable definitions

- *n*: number of catalysts included
- *p*: probability of observing a test statistic at least as extreme as that obtained under the null hypothesis

Dataset statistics

The dataset includes 157 catalysts. For the combined scores (Sum_U, Sum_R, and DesignEvidenceScore), we report the median as the main value, along with the interquartile range (IQR) and full range. This is because these scores may not follow a normal distribution.

- Sum_U: median = 2.8500; IQR = 2.4100–3.1000; range = 1.1500–3.7200
- Sum_R: median = 1.5800; IQR = 1.2600–1.9000; range = 0.6500–2.8500
- DesignEvidenceScore: median = 3.1414; IQR = 2.7742–3.5245; range = 0.9584–4.1696

Quadrant occupancy non-uniformity

We tested whether the catalysts are evenly distributed across the four quadrants (**Q1**–**Q4**) using a chi-square (χ^2) goodness-of-fit test. If the distribution were uniform, each quadrant would contain about 39.25 catalysts (157/4).

However, the counts obtained are **Q1** = 13, **Q2** = 68, **Q3** = 68, and **Q4** = 8.

The test result ($\chi^2(3) = 84.554$, $p = 3.235 \times 10^{-18}$) shows that the distribution is not uniform. Some quadrants are much more populated than others.

Diagonal structure of U–R space

The relationship between Sum_U and Sum_R was tested using Pearson correlation (n = 157).

The result shows a strong negative relationship ($r = -0.8959$), meaning that as Sum_U increases, Sum_R tends to decrease. This relationship explains about 80% of the variation ($R^2 = 0.8027$) and is highly statistically significant ($p = 1.711 \times 10^{-56}$). The 95% confidence interval for the correlation is −0.9230 to −0.8600, indicating that the strong negative relationship is consistent across the dataset.

Dependence between anchoring category and quadrant

We tested whether anchoring type and quadrant placement are related using a chi-square (χ^2) test (n = 157). The result ($\chi^2(12) = 27.787$, $p = 0.005943$) shows that the two are not independent, meaning anchoring type is associated with where catalysts fall in the quadrants. The effect size (Cramér's V = 0.2429) indicates a moderate strength of this relationship.

Differences in evidence completeness across quadrants

We tested whether DesignEvidenceScore differs across the four quadrants using the Kruskal–Wallis test (n = 157; **Q1** = 13, **Q2** = 68, **Q3** = 68, **Q4** = 8). The result ($H(3) = 85.915$, $p = 1.651 \times 10^{-18}$) shows clear differences between quadrants. The effect size ($\varepsilon^2 = 0.5419$) is large, meaning that evidence completeness varies strongly across the quadrants.

High-penalty occupancy

The number of catalysts occupying high-penalty quadrants (**Q1** and **Q3**) was 81 out of 157, corresponding to 51.59% of the dataset.

Association between reported η_{10} and aggregate penalty

For the subset of 59 catalysts where η_{10} is reported, its relationship with Sum_R was evaluated using Pearson correlation. The η_{10} values have a median of 0.4200 V, with an interquartile range of 0.3270–0.5300 V and an overall range of 0.0200–1.1230 V. The correlation coefficient is r = 0.5436 (p = 8.616

× 10^{-6}), indicating a positive association between η_{10} and Sum_R. The 95% confidence interval for r is 0.3340–0.7020.

Association between evidence completeness and U–R coordinates

For *DesignEvidenceScore* vs *Sum_U*:

- Pearson r = 0.8720, p = 6.017 × 10^{-50}, 95% CI = [0.8285, 0.9050]
- Spearman ρ = 0.8767, p = 3.958 × 10^{-51}

For *DesignEvidenceScore* vs *Sum_R*:

- Pearson r = −0.6657, p = 1.887 × 10^{-21}, 95% CI = [−0.7447, −0.5683]
- Spearman ρ = −0.6719, p = 5.855 × 10^{-22}

All values reported in this section were generated directly from S7.csv using scripts and correspond to the statistics cited in the main manuscript.

S12. Performance–Interpretability Divergence

We evaluated whether the commonly reported electrochemical parameter overpotential at 10 mA cm^{-2} (η_{10}), uniquely identifies systems that appear mechanistically interpretable within the U–R space. We only consider reports where η10_V values are reported (η_{10} subset; n = 59).

Association between η_{10} and aggregate penalty

- Pearson(η_{10}, Sum_R): r = 0.544, p = 8.6 × 10^{-6}
- Spearman(η_{10}, Sum_R): ρ = 0.503, p = 4.9 × 10^{-5}

This indicates that η_{10} and penalty are partially coupled but not diagnostically equivalent quantities. Systems with comparable η_{10} values can occupy distinct penalty–evidence regions in the U–R space.

Quintile-based activity

To avoid comparing very similar η_{10} values directly, the reported η_{10} (n = 59) were divided into five groups (quintiles). The best group contains the lowest 20% η_{10} values. For each group, we checked four things:

- how many fall in high-penalty quadrants (Q1 + Q3)
- how many are in the lowest evidence tier
- how many have both problems (high penalty + low evidence)
- how many are in Q2 but still have low evidence (evidence gap)

Across all 59 systems:

- 40.7% are in high-penalty quadrants
- 55.9% are in the lowest evidence tier
- 35.6% have both high penalty and low evidence
- 11.9% are in Q2 but still lack strong evidence

Even in the best-performing group (lowest η_{10}), 33.3% still fall into high-penalty or low-evidence categories. This shows that good η_{10} does not always mean low penalty or strong evidence, and this pattern comes directly from the dataset.

Misassignment risk metrics

We defined a few simple percentages to describe where systems fall:

- **M1:** 40.7% are in high-penalty quadrants (**Q1** + **Q3**)
- **M2:** 55.9% are in the lowest evidence tier
- **M3 (red zone):** 35.6% have both high penalty and low evidence
- **M4 (evidence gap):** 11.9% are in **Q2** but still have low evidence
- **M5 (screening failure):** 33.3% of the best η_{10} systems still have either high penalty or low evidence

These numbers show that overpotential values alone are not enough to identify systems that are low-penalty and well-supported.

Stratified sensitivity

When we look at different reaction types, the link between η_{10} and penalty stays moderate:

- HER: r = 0.585
- OER: r = 0.497

However, this relationship changes when grouped by basic conditions.

For strongly basic systems (Base = 1; n = 21), the correlation is very weak (r ≈ 0.022).

This shows that the connection between η_{10} and penalty is not consistent across all conditions.

S13. Circularity sensitivity check

This check was used to test whether U–R quadrant assignment and the η10–Sum_R divergence result depend on the S3 mechanistic-label columns.

Columns excluded from interpretation

• OA

• MLC

• PCET

• Mechanistic_Dominant

No-S3 quadrant recomputation

U–R quadrant assignment was recomputed using only Sum_U, Sum_R, and the fixed median boundaries.

• Stored and recomputed quadrant assignments matched for 157/157 catalysts.

• Q1 = 13

• Q2 = 68

• Q3 = 68

• Q4 = 8

η10–Sum_R divergence check

The η10–Sum_R analysis was repeated using the η10-reporting subset.

• η10-reporting subset: n = 59

• Pearson r = 0.5436, p = 8.62×10^{-6}

• Spearman ρ = 0.5031, p = 4.89×10^{-5}

Divergence summary

• High-penalty quadrants Q1 + Q3 = 24/59 (40.68%)
• Lowest evidence tier = 33/59 (55.93%)
• High-penalty + low-tier = 21/59 (35.59%)
• Q2 + low-tier = 7/59 (11.86%)
• Best-activity quintile with high penalty or low evidence = 4/12 (33.33%)

Interpretation

Removing the S3 mechanistic-label columns from the interpretation did not change U–R quadrant placement or the η10–Sum_R divergence result. The S3 mechanistic tendency labels therefore do not drive the U–R placement or the performance–interpretability divergence result.

References

1 T. Singh, S. Ghosh and A. Sarbajna, *Coord. Chem. Rev.*, 2026, **546**, 217071.

2 Y. Gao, Y.-L. Sun, Z. Guo, Y.-N. Liu, Y.-P. Qu and P.-H. Zhao, *Electrochim. Acta*, 2025, **529**, 146362.

3 P.-H. Zhao, Z. Guo, Y.-L. Sun, Y.-N. Liu, X.-B. Jing and L. Guo, *Electrochim. Acta*, 2026, **549**, 148024.

4 K. Tang, J.-Y. Shao and Y.-W. Zhong, *Chem. Eur. J.*, 2023, **29**, e202302663.