ACS Catalysis

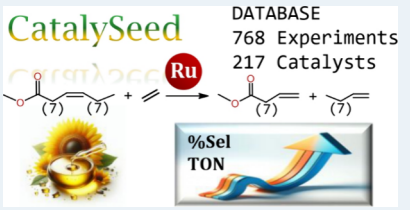
pubs.acs.org/acscatalysis
Research Article

CatalySeed: A Reaction Database for Ruthenium-Catalyzed Ethenolysis of Seed Oils with Applications in Machine Learning

Albert Poater,* Susana P. García-Abellán, Juan V. Alegre-Requena,* Bartosz Trzaskowski, and J. Pablo Martínez*

Cite This: ACS Catal. 2026, 16, 2160−2170
Read Online

ABSTRACT: Ethenolysis of unsaturated seed oils is an atom-efficient metathesis reaction that enables α-olefin production and fine chemical synthesis. By upcycling complex biobased molecules into value-added products, it supports circular chemical processes. In this study, we present a curated data set to support machine learning (ML) analysis of catalytic performance in the ethenolysis of seed oils. Through a detailed classification of 768 entries and 217 catalysts, along with the integration of the ROBERT ML framework, with the CatalySeed database we identify key electronic descriptors that correlate with experimental outcomes. Binary classification models for TON (threshold $\geq 0.75 \times 10^6$) and % selectivity (≥90%) achieved strong performance, suggesting that higher Ru partial charge tends to correlate with higher TON, while lower metal d-orbital character is generally associated with higher selectivity. These findings illustrate how this database, available through an open-access web server, enables ML to uncover predictive trends, supporting catalyst design strategies beyond conventional computational approaches for the transformation of renewable feedstocks.



KEYWORDS: *ethenolysis, oleate, ruthenium, machine learning, cheminformatics*

1. INTRODUCTION

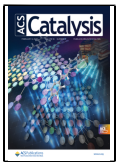
Biomass conversions sourced from renewable raw materials, specifically oils derived from plants, offer an eco-friendly and economically efficient alternative to mitigate dependence on crude oil and petrochemical products. The prospect of establishing sustainable processes is evident through the synthesis of value-added chemicals via the ethenolysis of seed oils.[1] However, while sustainable ethenolysis represents progress in the global energy transition, it introduces challenges related to metathesis reactions,[2] particularly in the production of fine chemicals from renewable feedstocks.[3] In this context, current biorefinery practices exhibit limitations with restricted applications.[4] Therefore, the synthesis of affordable, robust, and recoverable catalysts becomes increasingly critical for the practical implementation of these processes. The field lacks a unified and structured resource that consolidates the diverse experimental knowledge accumulated over decades, and machine learning (ML) offers a means to overcome this limitation.[5] In fact, it has become an indispensable tool for classifying data and making accurate predictions across scientific and technological domains.[6,7] The primary strength of ML lies in the ability to identify complex, nonlinear patterns within large and often noisy data sets, patterns that are difficult or impossible to detect using traditional statistical methods.[8] Beyond classification, the predictive capabilities of ML are transformative. Once trained on well-curated data sets, models can forecast properties or behaviors of unseen data, such as reaction outcomes, material performance, or catalytic activity.[9] This allows for rapid screening, reduced experimental workload, and data-driven decision-making.

2. DESCRIPTION OF THE DATABASE

We present CatalySeed, a new curated database developed to support a deeper understanding of the fundamental aspects of the ethenolysis of poly- and monounsaturated fatty acid esters catalyzed by ruthenium complexes. A dedicated web application of CatalySeed enables enhanced visualization and exploration of the data set.[10] For instance, Figure 1 summarizes the state-of-the-art in catalyst performance and was generated using the plotting utility accessible to users.

The database is structured as a relational table, illustrated in Table 1 with an example entry. Each of the 768 entries represents a unique catalytic experiment reported in 38 references, listed in reverse chronological order (with minor deviations) and matching the order used in the database.[11−48] Although group VI oxo complexes are also known to catalyze the ethenolysis of seed oils,[49,50] the database exclusively contains ruthenium-based catalysts to ensure data consistency and enable reliable ML modeling, while keeping an ambitious range of data compared to previous reports.[51] In this context, a notable effort,

Received: September 13, 2025
Revised: December 31, 2025
Accepted: January 6, 2026
Published: January 19, 2026



https://doi.org/10.1021/acscatal.5c06483
ACS Catal. 2026, 16, 2160−2170

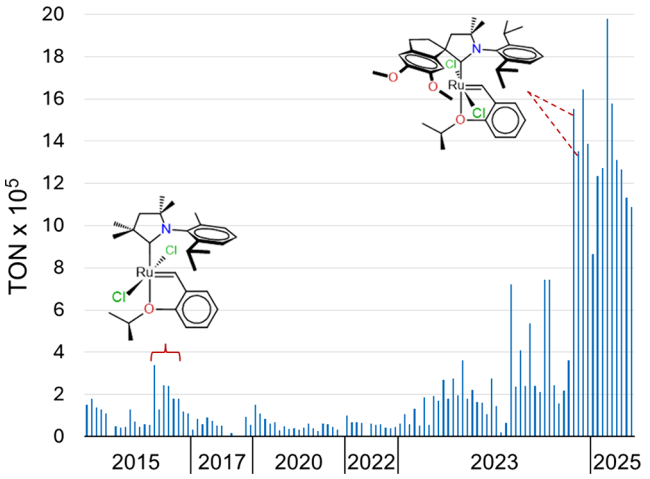


Figure 1. Evolution of the turnover number of the Ru-catalyzed ethenolysis of methyl oleate at low catalyst loads (≤10 ppm). Examples of early stage[35] and top-performing[11] catalysts are illustrated.

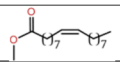
such as the work of Sigman and co-workers on N-heterocyclic carbene (NHC) Ru-based catalysts,[52,53] has provided valuable mechanistic insights into selectivity, yet broader data sets encompassing diverse substrates and catalyst classes have remained scarce.

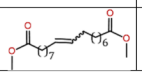
The information contained in CatalySeed is structured into 33 fields, each capturing a specific aspect of the catalytic experiments. The initial fields describe the reactants *Substrate 1* (ethylene; pressure in bar and purity in %) and *Substrate 2* (seed oil), followed by the series of compounds formed through productive and self-metathesis, as well as any minor or trace products. These compounds are stored as SMILES strings (Simplified Molecular Input Line Entry System), when available. Table 1 presents data for pure (99%) methyl oleate, although substrates can also consist of complex mixtures; for example, rapeseed (C16:0, C18:0/1/2/3) corresponds to methyl palmitate, methyl stearate, methyl oleate, methyl linoleate, and methyl linolenate, respectively.

Performance metrics, central to our ML analysis, include conversion (*Conv.* (%)), selectivity (*Sel.* (%)), yield (*Yield* (%)), and turnover number (*TON*). The field *Class* enables data classification derived from experimental conditions. The subsequent fields are associated with the experimental conditions under which the ethenolysis reactions were conducted. A free-text *Special conditions* field in the database records specific factors such as propenyl-benzenes as substrates,[21] product distillation,[22] the use of microwaves,[31] ionic liquids,[39] or a microreactor system.[44] The complete list of these cases, together with a general description of the curation procedure for each field, is provided in subsection B.5 of the Supporting Information to facilitate future expansion of the database.

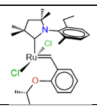
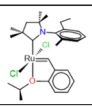
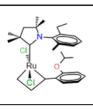
The fields that follow in the database describe the reaction environment. A dimeric Ru complex[25] may have been listed among the fields related to the reaction medium; however, since this case involves a single catalyst (four entries from one reference), its dimeric nature is instead specified in the *Special conditions* field. The reaction medium is defined by the use of solvents and/or solid supports such as silica.[17,34,38] In some instances, neat conditions are reported in the *Solvent* field, but the database records the initial catalyst solution, e.g., neat (catalyst added in toluene solution). Each entry in the database also includes the ruthenium complex as a SMILES string (*Catalyst*, 217 unique structures in total) and a short identifier (*code_name*). Subsections A.2 and A.5 of the Supporting Information provide relevant details for the generation of SMILES strings. Data traceability is ensured by the inclusion of the original label reported in each publication, recorded in the field *Label used in reference*. This field combines a structural identifier with a concise citation format (first author and year of publication), and it includes the digital object identifier (DOI). Furthermore, Table 2 provides a concise summary of the

Table 1. CatalySeed Database Structure with an Example Entry[a]

Entry	Subst. 1	P_C2H4 (bar)	Purity_C2H4 (%)	Subst. 2	Main comp.	Product 1	Product 2
487	C=C	10	99.95		C18:1 99%		

Entry	Homod. 1	Homod. 2	Trace prod.	Conv. (%)	Sel. (%)	Yield (%)	TON
487				37	86	32	110000

Entry	Class	Anal. method[b]	Reactor mode	Atmosphere	T (°C)	t (h)	Special cond.
487	S1	GC, IS: dodecane	batch reactor		40	3	

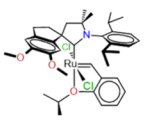
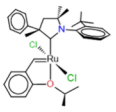
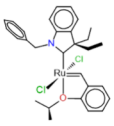
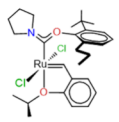
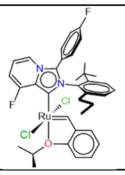
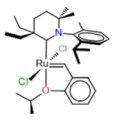
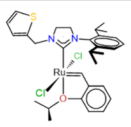
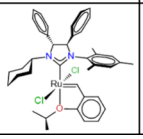
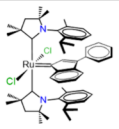
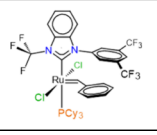
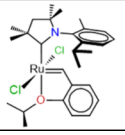
Entry	Sol. support	Catalyst phase	Solvent	[Cat.] (ppm)	Catalyst[d]	PreCat	MCB
487		Homogen.	neat[c]	3			

Entry	charge[e]	mult[e]	code_name	Label in ref.[f]
487	0	1	CAAC.09	9 in Marx2015

[a]Substrates, products, and ruthenium complexes are stored as SMILES strings in the database (e.g., Product 1, methyl 9-decenoate: COC(=O)CCCCCCCC=C),[10] while they are represented as images in this table for clarity. [b]GC: gas chromatography; IS: internal standard. [c]Neat conditions include an additional note: *neat* (*catalyst added in toluene solution*). [d]Base format utilized to derive SMILES of catalytic species. [e]Charge and multiplicity for the catalyst species listed within a given entry. [f]The digital object identifier (DOI) is also provided.

https://doi.org/10.1021/acscatal.5c06483
ACS Catal. 2026, 16, 2160−2170

Table 2. Summary of the Database Content, Including Structural Diversity (Showing Representative Structures) and Performance Range[a]

Catalyst codes	**sCAAC.01-19**	**CAAC.26-39**	**CAAC.40-42**	**AAOC.1-4, dAC**	**NHC.107-128**	**CAAC.43-46**
Entries range	1-125[11,12]	126-168[13−15]	169-179[16]	204-216[18]	217-257[19,20]	294-298[24]
Data classif.	S1, P1, O	S1, O	S1	S1	S1	S1
Feedstock[b]	**MO**, oils	**MO**, oils	**MO**	**MO**	**MO**	**MO**
Catalyst load[c]	0.5 ppm (63)	3 ppm (20)	3-100 ppm	50 ppm (6)	100 ppm (14)[d]	20 ppm (*all*)
Max. TON	2.6x10^6	7.4x10^5	6.1x10^4	1.0x10^5	1.1x10^5	1.7x10^3
Catalyst of max. TON						
Catalyst codes	**NHC.006-065**	**NHC.095-106**	**CAAC.01-08**	**NHC.090-094**	**CAAC.09-25**	**NHC.066-075**
Entries range	327-419[28]	420-438[29,30]	460-471[32]	472-479[33]	487-508[35]	702-725[45]
Data classif.	S1	S1	S1	S3	S1	S2
Feedstock[b]	**EO**	**EO**	**MO**	**EO**[e]	**MO**	**MO**
Catalyst load[c]	500 ppm (65)	500 ppm (15)	10 ppm (9)	100 ppm (6)	3 ppm (18)	100 ppm (13)
Max. TON	7.4x10^3	7.5x10^3	8.6x10^4	2.1x10^4	3.4x10^5	5.5x10^3
Catalyst of max. TON						

[a]The table is organized in reverse chronological order. Most frequent reaction conditions among the 768 entries: temperature = 40 °C (294 entries), reaction time = 6 h (252 entries), ethylene pressure = 10 bar (482 entries), and ethylene purity = 99.9% (242 entries). [b]Oil substrates (**MO** or **EO**: methyl or ethyl oleate). [c]The most frequent catalyst loading (number of entries) is shown. [d]A copper-based cocatalyst was used in several entries. [e]Solvated in toluene.

coverage of the data set. Complete information across all 33 fields and 768 entries is available in the CSV file provided in the Supporting Information.

Most fields in the database are self-explanatory, except for *Class*. This field was introduced to categorize the data according to variations in the practice of different experimental groups when reporting catalytic outcomes. This classification addresses the inherent challenge of comparing results reported by diverse research laboratories, where variations in outcome definitions can otherwise hinder consistency. It ensures that comparisons of reaction performance are unambiguous, a critical feature for ML investigations. On the other hand, minor residual variations in experimental execution such as differences in temperature control, reactant purity, or measurement precision are unavoidable. Nevertheless, reactions grouped within a given category remain comparable not by identical conditions, but by the consistency of outcome definitions, enabling trend-based analysis of catalyst behavior. In this context, the equations for substrate conversion and TON remain consistent through the database:

$$\%\ \text{Conv} = 100 - 100\frac{\text{final moles of } \mathbf{MO}}{\text{initial moles of } \mathbf{MO}} \quad (1)$$

$$\%\ \text{Yield} = \frac{(\%\ \text{Conv})(\%\ \text{Sel})}{100} \quad (2)$$

$$\text{TON} = \frac{(\%\ \text{Yield})(\text{initial moles of } \mathbf{MO})}{100(\text{moles of catalyst})} = \frac{10\,000(\%\ \text{Yield})}{\text{ppm of catalyst}} \quad (3)$$

In some cases, ethyl oleate was used but eq [1]−[3] are equivalent. The calculation of %Yield typically follows eq [2], although there are variations between references, along with differences in the equations used to determine selectivity. To address this variability, we define four classes. (i) Class S, where selectivity is reported with respect to all metathesis products; (ii) Class P1, where it is reported only for a single main product P1; (iii) Class U, where the equations used to calculate reaction outcomes are unspecified; and (iv) Class O, which encompasses complex oil substrates curated individually for each reference (a detailed reference-by-reference classification is provided in Section B of the Supporting Information). (i) Class S groups data suitable for ML analysis because it contains a larger number of comparable experimental outcomes. Nevertheless, the data in Class S required further subdivisions into classes S1, S2, and S3, which is justified as follows. For Class S1, fully represented in Table 2 and plotted in Figure 1 from the filtered data set, in addition to eq [1]−[3], selectivity was calculated as

$$\%\ \text{Sel} = 100\frac{\text{moles of P1} + \text{moles of P2}}{\text{moles P1} + \text{moles P2} + 2(\text{moles P1}' + \text{moles P2}')} \quad (4)$$

Derived from the ethenolysis of methyl oleate, P1 refers to *Product 1* in the database, methyl 9-decenoate, and P2 to *Product 2*, 1-decene. The corresponding self-metathesis products are denoted as P1′ (dimethyl 9-octadecene-1,18-dioate) and P2′ (9-octadecene), curated under the homodimer fields (see Table 1). The factor of 2 in the denominator accounts for the stoichiometry of self-metathesis, wherein two moles of substrate are required to form one mole of homodimer. In contrast, for Class S2,[41,42,44−46] selectivity was calculated as

https://doi.org/10.1021/acscatal.5c06483
ACS Catal. 2026, 16, 2160−2170

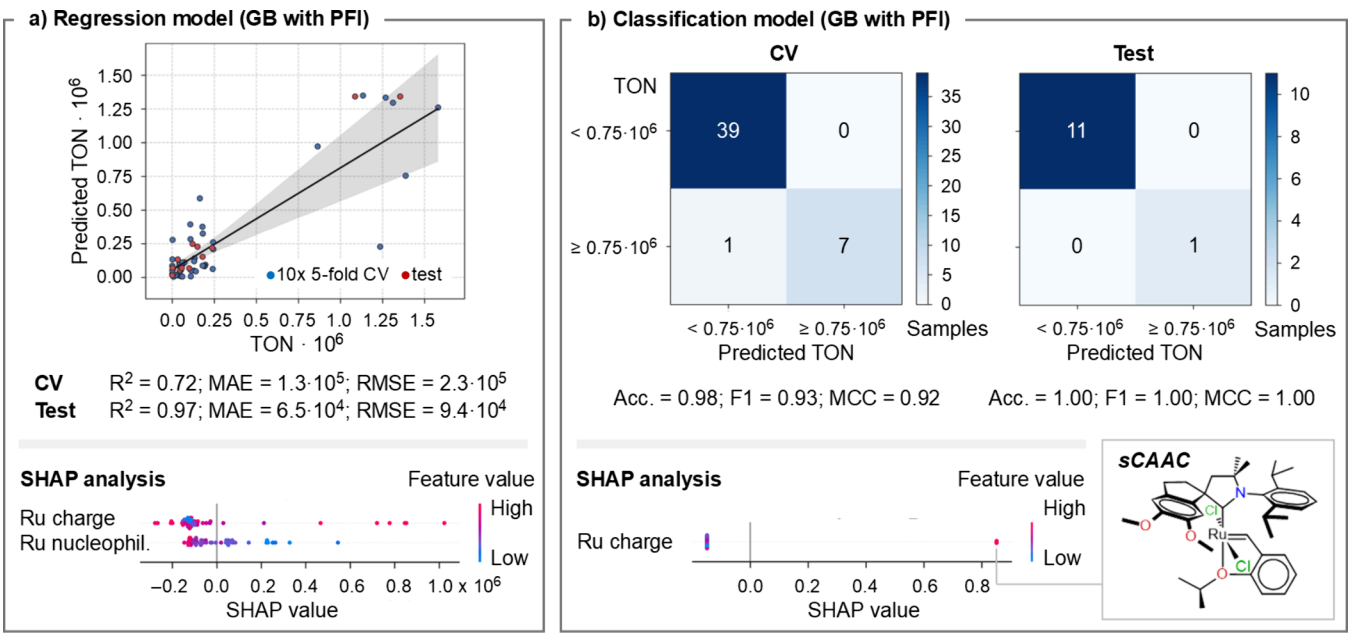


Figure 2. Gradient boosting (GB) models with permutation feature importance (PFI) filtering for CV and an 80:20 train:test split. (a) Regression scatter plot of predicted vs measured TON values. (b) Confusion matrices for the binary classification model illustrated for the prediction of the TON (threshold $\geq 0.75 \times 10^6$). SHAP summary plots are included for both models, illustrating feature contributions, along with an example of catalysts associated with the highest Ru partial charges. Both models were trained using repeated 10 × 5-fold CV on 47 training entries, with performance evaluated on a separate test set of 12 entries.

$$\%\ \mathrm{Sel} = 100\frac{\text{moles of P1} + \text{moles of P2}}{\text{moles P1} + \text{moles P2} + \text{moles P1}' + \text{moles P2}'} \quad (5)$$

thereby differing from eq [4]. Nevertheless, equivalent % Sel values are expected from eq [4] and [5] in highly selective reactions, as the self-metathesis terms become negligible. Class S3 comprises entries that use alternative definitions of selectivity not covered by eq [4] or [5].[33,34,39]

(ii) Class P1 includes entries reporting reaction outcomes specifically for the product methyl 9-decenoate.[11,17,22] Analogous eq [1] to [3] are applied; however, the numerator in eq [5] includes exclusively moles of P1. (iii) Class U (unspecified equations) contains data from four references;[25,26,47,48] and (iv) Class O comprises entries derived from canola, waste cooking, cashew nut, oleonitrile, or ricinus oils.[23,31,36−38,40,43] Although some values were not directly reported in the original publications (e.g., TON), they were derived using these well-defined arithmetic relationships consistent with the methodology used in each study. These additions enhance the completeness of the data set without altering any originally reported values. In fact, data extraction from each reference is detailed in Section B of the Supporting Information for transparency purposes. Additional studies on Ru-catalyzed ethenolysis of plant-derived oils and natural rubber are available in the literature;[54−58] however, their data could not be incorporated due to format incompatibility with the present database.

3. COMPUTATIONAL METHODS

To demonstrate the practical application of the CatalySeed database, we performed ML analyzes using ROBERT, a free and open-source software designed for automated ML workflows, tailored to supervised regression and classification problems in chemistry.[59−62] A key advantage of ROBERT is that it enables users to build and deploy models via a user-friendly interface, regardless of programming skill level, while ensuring high standards of reproducibility[63] and transparency, and achieving performance comparable to expert-developed models. The automated workflow begins with a CSV file containing the target values (*y*-values, e.g., TON) and SMILES strings.[64] The AQME (Automated Quantum Mechanical Environments)[65] program was employed to compute descriptors, which serve as *X*-values. AQME generates Boltzmann-weighted descriptors (by assigning populations to each conformer according to its relative energy) from SMILES using RDKit[66] and GFN2-xTB,[67,68] methods significantly faster than DFT, making them ideal for large databases. The resulting properties, including molecular electronic descriptors (e.g., HOMO, dipole moment, etc.) and Ru-centered atomic and steric descriptors (e.g., partial charge, buried volume, etc.), were used to capture features relevant to ML models.

To ensure consistency under comparable experimental conditions, the CatalySeed database was filtered to include 59 entries (59 catalysts) assigned to Class S1 and with catalyst loadings ≤ 20 ppm. This threshold reduces the influence of extreme catalyst quantities and reflects intrinsic reactivity rather than concentration-dependent effects. Although the reported reaction temperatures (25−60 °C) and times (3−6 h) vary among studies, these values correspond to the optimized conditions reported in each reference and were therefore retained as representative of the respective experimental protocols. The filtered data set used in this ML analysis is provided as a CSV file and illustrated in Table S1 (subsection A.1) in the Supporting Information. To enable RDKit-based conformer generation, CatalySeed and its filtered form include curated SMILES strings for two catalyst forms: *PreCat* (precatalyst) and *MCB* (metallacyclobutane). Additional details of the *Catalyst*, *PreCat*, and *MCB* fields are available in Table S2 in the Supporting Information (see also Table 1). Here, we focus on the *PreCat* structures, for which three conformers were generated using a custom conformer-generation script based on RDKit (Table S3 in the Supporting Information). Each geometry was visually inspected to ensure consistency with the ligand orientations reported in the corresponding crystal structures (see Table S4 in the Supporting Information). In asymmetric complexes, additional atropoisomeric arrangements could, in principle, occur due to rotation of the NHC or cyclic alkyl amino carbene (CAAC) ligand. However, the present data set includes exclusively the conformations experimentally observed in

https://doi.org/10.1021/acscatal.5c06483
ACS Catal. 2026, 16, 2160−2170

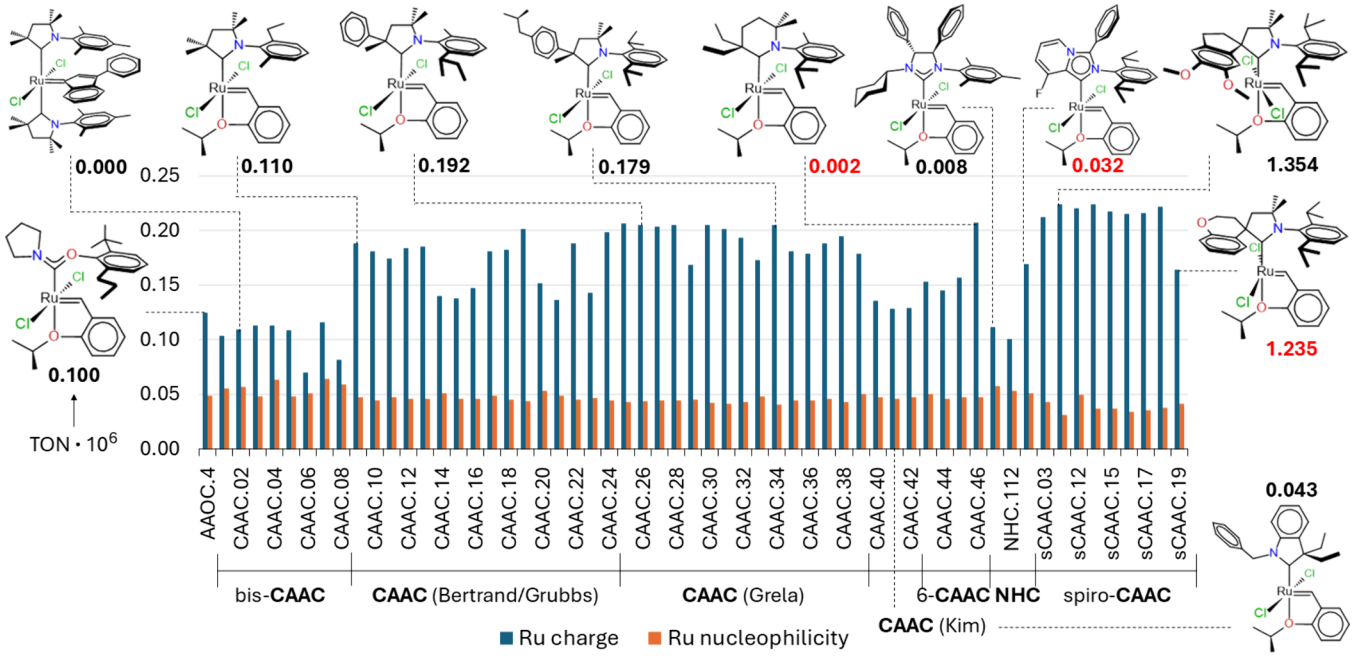


Figure 3. Ru partial charge (e) and Ru nucleophilicity across the catalysts included in the data set filtered under Class S1 (≤20 ppm) and used as input features in the GB models. Representative structures per group of catalysts are illustrated, with corresponding TON × 10^6 values (in red, deviations from the general tendency of higher Ru charge being associated with increased TON).

crystallographic data to maintain a uniform descriptor space for ML analysis. The conformers generated with this script were subsequently processed through AQME to perform geometry optimizations and descriptors generation using GFN2-xTB.

Subsequently, the AQME workflow compiled 17 atomic and 22 molecular descriptors into a CSV file available in the Supporting Information. These served as input for ROBERT, which performs analyses both with and without permutation feature importance (PFI) filtering to identify the most relevant descriptors for ML modeling. The metrics for the PFI-based ML simulations, which were comparable to or better than the non-PFI models, are discussed in this study; both sets of results are included in the ROBERT reports available in the Supporting Information. In this context, all models were evaluated using 10-times repeated 5-fold cross-validation (10 × 5-fold CV), which provides a robust framework for measuring model performance.

4. RESULTS AND DISCUSSION

We evaluated eight ML approaches that encompass both regression and binary classification tasks aimed at predicting %Conversion, %Selectivity, %Yield, and TON.[69] Across these eight approaches, ROBERT was used to screen six algorithms, including gradient boosting (GB)[70] and random forests (RF).[71] For each case, model performance was assessed through automated comparison of CV and test-set metrics (e.g., R^2, accuracy, etc.) to identify models with predictive value. In most tasks, the evaluated metrics indicated negligible predictive performance, and the corresponding models are not discussed further. Equivalent models based on descriptors from the *MCB* species (the intermediate generated in the initiation phase) were also tested, but no significant trends with the experimental targets were observed, thereby suggesting limited structure−property correlations for these structures. For *PreCat*, only three cases, TON (regression and classification) and %Selectivity (classification), yielded models with meaningful results, which are the focus of the present discussion.

4.1. Gradient-Boosting Models for TON

The regression model for TON, developed using GB[70] and selected as the best-performing algorithm through ROBERT's automated comparison of multiple ML models, achieved strong predictive accuracy, particularly for test data, with R^2 = 0.97 (Figure 2a). However, the distribution of predictions was skewed by 9 of 59 catalysts bearing spirocyclic alkyl amino carbenes (**sCAAC**) with exceptionally high TON values (>0.75 × 10^6), which account for only 15% of the entries in the filtered database. This imbalance likely led to a model that produced misleading CV and test results, showing relatively high R^2 values of 0.72 and 0.97, respectively, but lacking robustness.

The data imbalance prompted us to create a binary classification model based on GB (Figure 2b), in which we selected the threshold TON ≥ 0.75 × 10^6 to isolate the nine top-performing **sCAAC** catalysts. This type of model is effectively summarized by confusion matrices, where correct predictions appear on the diagonal. The performance metrics considered are accuracy (ratio of correct predictions to the total predictions), the F_1 score (harmonic mean of precision and recall), and the Matthews correlation coefficient (MCC), and in all cases the resulting model showed excellent performance for CV and test sets (accuracy, F_1 score, and MCC > 0.90).

The SHAP (SHapley Additive exPlanations) analysis provides insight into the GB models by ranking descriptors according to their influence on predicted TON. SHAP summary plots in Figure 2 highlight the Ru Mulliken atomic charge as the primary contributor in both TON models. Mulliken charges are well-known for their dependence on the computational method, which limits their robustness as absolute values, although their variations across catalyst families can still serve as inputs for ML simulations, thus supporting trend-based interpretation rather than precise quantitative predictions. In this context, a general correlation emerges in Figure 3: complexes with lower Ru charges tend to display lower TON values. At the left of the distribution, catalysts bearing acyclic-amino-oxy-carbene

https://doi.org/10.1021/acscatal.5c06483
ACS Catal. 2026, 16, 2160−2170

(**AAOC.4**)[18] and bis-CAAC ligands (**CAAC.01−08**)[32] exhibit the smallest charges (<0.13*e*) and correspondingly decreased TONs (<1 × 10^5). Moving across the **CAAC.09−25** series developed by Bertrand and Grubbs,[35] and **CAAC.26−39** reported by Grela et al.[13−15] modified with ibuprofen substituents or inverted CAAC orientations, the charges increase to 0.15−0.20*e* with TONs > 1 × 10^5. Figure 3 nevertheless reveals deviations from the general tendency, as some complexes exhibit Ru charges of ca. 0.15*e* yet reduced TON values, in certain cases reaching TON = 0 (e.g., **CAAC.20** with an adamantyl substituent). Another example is the six-membered CAAC catalyst **CAAC.46**[24] that, despite a higher Ru charge of 0.21*e*, achieves a TON value of only 1 705. In contrast, it is worth noting that the highest charges, 0.21−0.22*e*, observed in 8 of the 9 catalysts with spirocyclic scaffolds (**sCAAC** at the rightmost section of Figure 3), shift trend toward the highest TON values. In fact, we previously observed that catalysts bearing a spiro-CAAC ligand and exhibiting a more positively charged Ru center tend to correlate with higher TON values.[11] However, in other studies of olefin metathesis catalysts,[72−74] such correlations between Ru atom charges and catalytic performance were not consistently observed, indicating that tendencies may be context-dependent and should not be generalized across all types of olefin metathesis.

These trends allow a clearer interpretation of the SHAP diagrams. Complexes with relatively higher Ru charges (red points in Figure 2) are not uniformly associated with larger TON values, and therefore some appear in the negative SHAP region. By contrast, the relation between low Ru charge and low TON is consistently captured, with all corresponding blue points located in the negative SHAP region. The mixed behavior at high Ru charge is evident: while most **sCAAC** complexes combine increased charge with exceptionally high TONs, other families show high charges without comparable activity. This context-dependent effect may be modulated by additional factors such as Ru nucleophilicity, as considered in the regression model. Nonetheless, nucleophilicity shows weaker trends, with six of nine **sCAAC** complexes below 0.04 and the remaining examples distributed in the 0.04−0.07 range (see Figure 3). These observations highlight both the consistency and the limits of Ru-based descriptors, clarifying how they contribute to the ML analyses.

The selection of the population scheme is a critical factor influencing model reliability. Accordingly, in subsection A.6 xTB-derived charges are compared with those obtained at the ωB97XD/Def2-SVP (SDD for Ru) level.[75−82] Table S6 and Figure S1 show that xTB Mulliken and DFT Hirshfeld values are similar in magnitude, whereas DFT CM5 values follow the same order but are shifted upward by ca. 0.4*e*. Conversely, DFT Mulliken and NPA charges yield negative values, likely reflecting basis-set polarization effects in the former and the NBO partitioning of *d*-electron density in the latter. As shown in Table S7 and Figure S2, only CM5 charges reproduce the GB-PFI performance achieved with xTB Mulliken populations, thereby reflecting their methodological similarity and internal trend consistency. This outcome indicates that, while the GB model is sensitive to the quantum-mechanical method, some population schemes remain effective in extracting interpretable trends, with xTB providing a practical alternative to DFT.

In a previous computational analysis of Ru metathesis catalysts, electronic factors influencing catalytic efficiency were examined.[83] While not directly addressing ethenolysis, the authors found that charge distribution within the Ru coordination sphere acts as an integrated indicator of the donor−acceptor balance, whereby highly donating ligands stabilize active species through charge delocalization. Such observations offer a complementary electronic perspective on the Ru-based descriptors.

4.2. Random Forest Model for Selectivity

Next, we studied reaction selectivity using similar ML-driven strategies. For %Selectivity, regression was limited by an uneven data distribution, with a mean of 82% (standard deviation of 25%), and a large cluster near the upper end (90−98%, 32 of 59 entries) with few intermediate values. This pattern limits the ability to model a smooth continuous relationship. Consequently, a binary classification model using RF was investigated,[71] identified by ROBERT as the best-performing algorithm (Figure 4). The classification data set for %Selectivity is more balanced than its regression counterpart and the TON-based data sets, with 27 entries below and 32 entries above or equal to the 90% threshold. This improved balance allows for a more robust evaluation and reduces the risk of misleading or biased predictions. The 90% threshold serves as a practical benchmark for high selectivity, aligned with desired standards in catalyst performance.[28] The model exhibited good predictive power, achieving accuracies of 0.81 and 0.75, F1 scores of 0.82 and 0.80, and MCC values of 0.62 and 0.58 on the CV and test

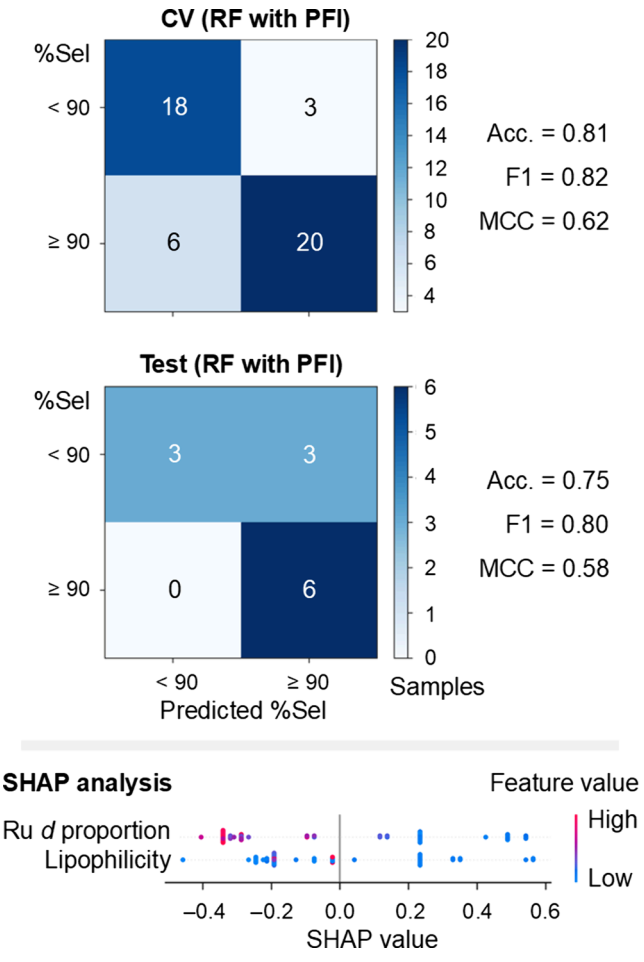


Figure 4. Confusion matrices for the random forest (PFI) binary classification model predicting % selectivity (threshold ≥90%), shown for CV and test sets (split 80:20), and SHAP summary plot illustrating feature contributions.

https://doi.org/10.1021/acscatal.5c06483
ACS Catal. 2026, 16, 2160−2170

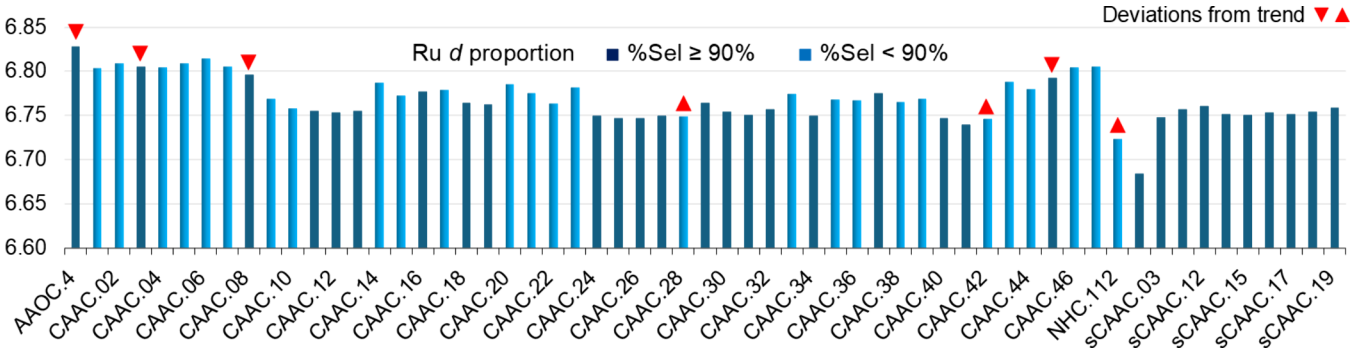


Figure 5. Ru *d*-orbital proportion across the catalyst groups used as input features in the random forest model.

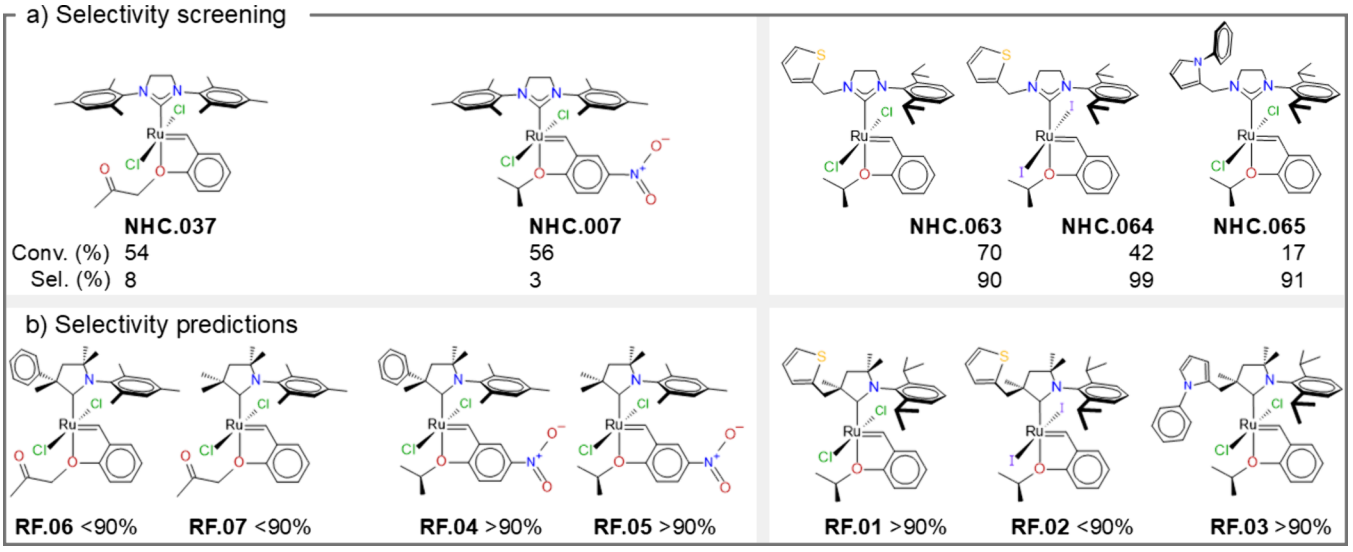


Figure 6. (a) Ru catalysts reported by Grela and co-workers,[28] together with their experimental conversion and selectivity values, illustrating the lowest and highest selectivity within the Hoveyda-type NHC series. (b) CAAC-based ligand modifications proposed in this work by analogy to these NHC trends. Selectivity predictions were obtained from the trained RF (PFI) classification model.

set, respectively. These results encourage the use of this model as a tool to assist in the design of highly selective catalysts, while keeping in mind the possibility of false positives (upper right cell of the test-set in Figure 4) and the need for experimental validation.

The descriptors for the classification model with %Selectivity as the target are the Ru *d*-orbital character, measured as the population partitioned to the *d*-shell, and molecular lipophilicity, measured by MolLogP (the logarithm of the octanol−water partition coefficient). High *d*-character (red points in Figure 4) appeared predominantly in the negative SHAP region indicating lower selectivity. However, the distribution of the *d*-character also highlights deviations from this trend (indicated by red symbols in Figure 5). In the left section of this plot, **AAOC.4** and the bis-CAAC complexes **CAAC.03** and **CAAC.08** combine high *d*-character with high selectivity. In contrast, the reduced *d*-character in **NHC.112** associated with low selectivity represents another deviation. The spirocyclic scaffold in **sCAAC** catalysts lowers the *d*-character, which aligns with their high selectivity. In the middle section, deviations from the general trend hinder direct interpretation, but the RF algorithm can still classify these cases by combining multiple decision boundaries derived from different subsets of the data.

For MolLogP, although a few red points also accumulate in the negative region of the SHAP plot (Figure 4), blue and red points are more interspersed, suggesting a less systematic effect compared to *d*-character. Therefore, this secondary feature provides only limited guidance for selectivity classification. In general, the SHAP patterns suggest that low lipophilicity and reduced *d*-character favor high selectivity predictions, but the influence of MolLogP appears less sharply defined.[69,84] Catalysts with lower lipophilicity might facilitate a more favorable partitioning of neat methyl oleate toward the metal center rather than the ligand environment, thereby enhancing productive metal-substrate interactions. Likewise, reduced *d*-character reflects a more electrophilic metal center, favoring selective olefin activation pathways.[85]

To the best of our knowledge, no computational studies have explicitly identified the Ru *d*-orbital contribution as a guiding factor for selectivity in olefin metathesis. Most mechanistic investigations have instead focused on the electronic influence of NHC and related carbene ligands. Cavallo and co-workers showed that aryl substituents on bis(aryl)-NHCs interact with Ru *d* orbitals, modulating metal−ligand back-donation.[86] Houk and Grubbs et al. further demonstrated that $d-\pi$ interactions between Ru and the alkylidene−olefin fragments stabilize the side-bound MCB responsible for Z-selectivity.[87] These studies highlight that variations in *d*-character provide a chemically

https://doi.org/10.1021/acscatal.5c06483
ACS Catal. 2026, 16, 2160−2170

meaningful measure of the metal−ligand electronic environment governing metathesis reactivity.[88] In parallel, de novo computational design strategies—based on evolutionary algorithms that iteratively generate, evaluate, and refine virtual catalyst structures—have also been applied to Ru complexes.[89] This study demonstrates the role of molecular descriptors within automated design approaches to estimate reactivity trends.

Finally, undesired double-bond migration may compete with productive metathesis thus affecting selectivity values. However, CAAC-bearing Ru catalysts have been shown to significantly suppress isomerization and related nonmetathetic side reactions, particularly in the absence of phosphine nucleophiles.[90] This behavior supports the interpretation that the selectivity trends observed in this study primarily reflect intrinsic catalytic effects rather than isomerization pathways.

4.3. RF-Model Guidance for Catalyst Design

To conclude, we illustrate how the RF model can assist ligand modification in Ru complexes. The catalytic performance of newly proposed structures cannot be inferred directly from their molecular structure, and modifications are thus typically guided by chemical judgment. In this example, the trained RF-PFI binary classification model with %Selectivity as the target is used to screen candidate ligands. The CatalySeed database provides the experimental reference needed to relate these model outputs to established structure−performance trends.

As a representative case, Grela and co-workers reported the ethenolysis of ethyl oleate using 65 Ru catalysts, predominantly Hoveyda-type NHC complexes, under open-air conditions without glovebox or Schlenk techniques.[28] These catalysts are included in the CatalySeed database (entries 327−419, summarized in Table 2) and are assigned to Class S1, since the selectivity values were derived according to eq [4]. The reaction conditions differ from those used to train the RF model, as they were chosen for practical application rather than controlled laboratory optimization. Nevertheless, the trends identified in this study enable identification of the ligands associated with the lowest and highest selectivity values within this series (Figure 6).

The SHAP analysis in Figure 4 highlights the relevance of the Ru *d*-proportion descriptor, as decreased *d*-character favors selectivity. However, this descriptor is only meaningful within the context of the RF model. Rather than inferring selectivity directly from a quantum-mechanical calculation of this single descriptor, the RF-PFI workflow evaluates it together with all other features to produce the prediction. The ROBERT code provides this evaluation in an automated workflow (details in subsection A.7 in the Supporting Information). Accordingly, Hoveyda-type NHC catalysts with <10% and >90% selectivity values were selected as reference points. Complex **NHC.037** represents the low-selectivity limit and is distinguished by an acetonoxy group in place of the conventional isopropoxy substituent on the benzylidene fragment. By analogy, complexes **RF.06** and **RF.07** (Figure 6b) were generated by introducing the acetonoxy substitution into a representative CAAC scaffold (e.g., **CAAC.21**). The RF−PFI model classifies these complexes as nonoptimally selective (<90%), in agreement with the performance of **NHC.037** (and **NHC.039**, bearing an *α*-methyl-acetonoxy group,%Sel = 12).

A complex bearing a nitro substituent also falls within the low-selectivity range. For example, **NHC.007**, which retains the typical NHC scaffold and 2-isopropoxy benzylidene fragment but carries a 5-NO_2 substitution, was reported to be non-selective. On the contrary, the CAAC-based complexes **RF.04** and **RF.05** are predicted by the RF−PFI model to be optimally selective (>90%). However, as discussed in the previous subsection, false positives are possible, and this prediction should therefore be interpreted with caution. Indeed, **NHC.048**, an analogous nitro-substituted dimesityl NHC complex, also exhibits negligible selectivity (5%).

Enhanced catalytic performance is observed for NHC complexes bearing a di-isopropylphenyl substituent. Furthermore, thiophen-2-ylmethyl and *N*-phenylpyrrolylmethyl substituents in the NHC scaffold (**NHC.063−065**) fall within the high-selectivity range. Guided by these trends, the phenyl group in **CAAC.25** was replaced to generate **RF.01** and **RF.03**, which are both predicted to retain high selectivity (>90%). However, high selectivity does not necessarily result in improved catalytic performance. For example, **NHC.065** exhibits decreased activity (%Conv. = 17), likely due to the flexibility of the *N*-phenylpyrrolyl substituent; as unprotected aryl groups are known to undergo ring-expansion decomposition with the Ru = CH_2 fragment.[91,92] Also, iodide substitution was reported to enhance selectivity;[28] however, replacing chloride with iodide in **RF.01** (yielding **RF.02**) led to a nonoptimal selectivity (<90%). This result indicates that iodide substitution remains uncertain in the CAAC context and may represent a false negative in the model. In general, these examples illustrate one possible rationale of the RF−PFI model as a tool for prioritizing candidate modifications for experimental validation. We therefore encourage the exploration of the CatalySeed database to identify alternative design strategies. Although the present models were trained predominantly on Hoveyda-type CAAC complexes, we are currently expanding the chemical space covered by the model to further increase its predictive robustness.

CONCLUSIONS

The CatalySeed database compiles 768 curated Ru-catalyzed ethenolysis reactions and supports reproducible benchmarking and ML analysis. Using the ROBERT platform, we evaluated three models: gradient boosting for TON (regression and classification) and random forest for %Selectivity (classification). All models demonstrated consistent predictive performance. Key descriptors—Ru atomic charge and *d*-orbital character—emerged as predictive features. These findings illustrate how ML can uncover chemical trends that complement traditional computational approaches. Overall, CatalySeed provides a versatile, data-rich resource for accelerating the design of efficient, selective, and recoverable Ru-based catalysts for green chemical processes, supporting the transition toward a more sustainable chemical industry.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscatal.5c06483.

Workflows for conformer generation, data set curation and classification, and illustrations of catalyst structures (PDF)
CatalySeed database (CSV)
Filtered database for ML simulations (CSV)
Descriptors and targets for ML simulations (CSV)
ROBERT reports (PDF)
Descriptors for ML predictions (CSV)

https://doi.org/10.1021/acscatal.5c06483
ACS Catal. 2026, 16, 2160−2170

xTB optimized geometries of conformers (XYZ)

AUTHOR INFORMATION

Corresponding Authors

Albert Poater − *Institut de Química Computacional i Catàlisi and Departament de Química, Universitat de Girona, 17003 Girona, Catalonia, Spain;* orcid.org/0000-0002-8997-2599; Email: albert.poater@udg.edu

Juan V. Alegre-Requena − *Departamento de Química Inorgánica, Instituto de Síntesis Química y Catálisis Homogénea (ISQCH), CSIC-Universidad de Zaragoza, 50009 Zaragoza, Spain;* orcid.org/0000-0002-0769-7168; Email: jv.alegre@csic.es

J. Pablo Martínez − *Centre of New Technologies, University of Warsaw, 02-097 Warszawa, Poland;* orcid.org/0000-0002-6589-790X; Email: p.martinez@cent.uw.edu.pl

Authors

Susana P. García-Abellán − *Departamento de Química Inorgánica, Instituto de Síntesis Química y Catálisis Homogénea (ISQCH), CSIC-Universidad de Zaragoza, 50009 Zaragoza, Spain;* orcid.org/0000-0002-3138-5527

Bartosz Trzaskowski − *Centre of New Technologies, University of Warsaw, 02-097 Warszawa, Poland;* orcid.org/0000-0003-2385-1476

Complete contact information is available at: https://pubs.acs.org/10.1021/acscatal.5c06483

Funding

The Spanish Ministerio de Ciencia, Innovación y Universidades (MCIN/AEI/10.13039/501100011033/FEDER, UE) for projects PID2024-155989NB-I00 (to A.P.), PID2022-140159NA-I00 (to J.V.A.-R.), the Generalitat de Catalunya for project 2021SGR623, Gobierno de Aragón-Fondo Social Europeo (Research Group E07_23R), and the Polish National Science Center through the SONATA-19 grant UMO-2023/51/D/ST4/01561 (to J.-P.M.).

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Spekreijse, J.; Sanders, J. P. M.; Bitter, J. H.; Scott, E. L. The Future of Ethenolysis in Biobased Chemistry. *ChemSusChem* **2017**, *10*, 470−482.

(2) Brotons-Rufes, A.; Posada-Pérez, S.; Martínez, J. P.; Nolan, S. P.; Poater, A. Challenges in Olefin Metathesis: Past, Present and Future. *Coord. Chem. Rev.* **2025**, *542*, 216827.

(3) Ivchenko, P. V.; Nifant'ev, I. E. The Chemistry of Oleates and Related Compounds in the 2020s. *Green Chem.* **2024**, *27*, 41−95.

(4) de Roo, S.; Einsiedler, F.; Mecking, S. Catalytic Biorefining of Natural Oils to Basic Olefinic Building Blocks of Proven Chemical Valorization Schemes. *Angew. Chem., Int. Ed.* **2023**, *62*, No. e202219222.

(5) Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse Molecular Design Using Machine Learning: Generative Models for Matter Engineering. *Science* **2018**, *361*, 360−365.

(6) Sanosa, N.; Dalmau, D.; Sampedro, D.; Alegre-Requena, J. V.; Funes-Ardoiz, I. Recent Advances of Machine Learning Applications in the Development of Experimental Homogeneous Catalysis. *Artif. Intell. Chem.* **2024**, *2*, 100068.

(7) Dalmau, D.; Alegre-Requena, J. V. Integrating Digital Chemistry within the Broader Chemistry Community. *Trends Chem.* **2024**, *6*, 459−469.

(8) Gromski, P. S.; Henson, A. B.; Granda, J. M.; Cronin, L. How to Explore Chemical Space Using Algorithms and Automation. *Nat. Rev. Chem.* **2019**, *3*, 119−128.

(9) Friederich, P.; Dos Passos Gomes, G.; De Bin, R.; Aspuru-Guzik, A.; Balcells, D. Machine Learning Dihydrogen Activation in the Chemical Space Surrounding Vaska's Complex. *Chem. Sci.* **2020**, *11*, 4584−4601.

(10) Martínez, J. P. *CatalySeed - A Database for Ethenolysis of Seed Oils*. University of Warsaw. https://catalyseed.cent.uw.edu.pl/ (accessed September 1, 2025).

(11) Gawin, R.; Tracz, A.; Krajczy, P.; Kozakiewicz-Piekarz, A.; Martínez, J. P.; Trzaskowski, B. Inhibition of the Decomposition Pathways of Ruthenium Olefin Metathesis Catalysts: Development of Highly Efficient Catalysts for Ethenolysis. *J. Am. Chem. Soc.* **2023**, *145*, 25010−25021.

(12) Afanaseva, A. V.; Vinogradov, A. A.; Vinogradov, A. A.; Minyaev, M. E.; Pyatakov, D. A.; Tavtorkin, A. N.; Bagrov, V. V.; Ivchenko, P. V.; Nifant'ev, I. E. The Impact of Ligand Structure and Reaction Temperature on Ethenolysis of Fatty Acid Methyl Esters Catalyzed by Spirocyclic Alkyl Amino Carbene Ru Complexes. *ChemSusChem* **2025**, *18*, No. e202402190.

(13) Sytniczuk, A.; Kajetanowicz, A.; Grela, K. "Inverted" Cyclic(Alkyl)(Amino)Carbene Ligands Allow Olefin Metathesis with Ethylene at Parts-per-Billion Catalyst Loading. *Chem. Catal.* **2023**, *3*, 100713.

(14) Sytniczuk, A.; Struzik, F.; Grela, K.; Kajetanowicz, A. A Tunable Family of CAAC-Ruthenium Olefin Metathesis Catalysts Modularly Derived from a Large-Scale Produced Ibuprofen Intermediate. *Chem. Sci.* **2023**, *14*, 10744−10755.

(15) Sytniczuk, A.; Struzik, F.; Purohit, V.; Grela, K.; Kajetanowicz, A. Aza-Claisen Rearrangement as a Key Step in Synthesis of Specialised Anilines Used in the Production of Efficient Ethenolysis Catalysts. *Catal. Sci. Technol.* **2023**, *13*, 3682−3688.

(16) Kim, M.; Kim, H.; Kim, S.; Hong, S.; Lee, E. Syntheses and Applications of Indol-2-Ylidene-Ligated Ruthenium-Based Olefin Metathesis Catalysts. *Organometallics* **2022**, *41*, 1905−1910.

(17) Aşkun, M.; Sagdic, K.; Inci, F.; Öztürk, B. Ö. Olefin Metathesis in Confined Spaces: The Encapsulation of Hoveyda-Grubbs Catalyst in Peanut, Square, and Capsule Shaped Hollow Silica Gels. *Catal. Sci. Technol.* **2022**, *12*, 6174−6183.

(18) Byun, S.; Park, D. A.; Kim, S.; Kim, S.; Ryu, J. Y.; Lee, J.; Hong, S. Highly Selective Ethenolysis with Acyclic-Aminooxycarbene Ruthenium Catalysts. *Inorg. Chem. Front.* **2022**, *9*, 323−331.

(19) Byun, S.; Park, S.; Choi, Y.; Ryu, J. Y.; Lee, J.; Choi, J.-H.; Hong, S. Highly Efficient Ethenolysis and Propenolysis of Methyl Oleate Catalyzed by Abnormal N-Heterocyclic Carbene Ruthenium Complexes in Combination with a Phosphine-Copper Cocatalyst. *ACS Catal.* **2020**, *10*, 10592−10601.

(20) Byun, S.; Seo, H.; Choi, J.-H.; Ryu, J. Y.; Lee, J.; Chung, W.-J.; Hong, S. Fluoro-Imidazopyridinylidene Ruthenium Catalysts for Cross Metathesis with Ethylene. *Organometallics* **2019**, *38*, 4121−4132.

(21) Avendaño Villarreal, J. A.; Granato, A. V.; Delolo, F. G.; dos Santos, E. N. Efficient Synthesis of Styrene Derivatives through Ethenolysis of Renewable Propenylbenzenes. *Mol. Catal.* **2021**, *509*, 111631.

(22) Kajetanowicz, A.; Chwalba, M.; Gawin, A.; Tracz, A.; Grela, K. Non-Glovebox Ethenolysis of Ethyl Oleate and FAME at Larger Scale Utilizing a Cyclic (Alkyl)(Amino)Carbene Ruthenium Catalyst. *Eur. J. Lipid Sci. Technol.* **2020**, *122*, 1900263.

(23) Pradhan, R. A.; Arshad, M.; Ullah, A. Solvent-Free Rapid Ethenolysis of Fatty Esters from Spent Hen and Other Lipidic Feedstock with High Turnover Numbers. *J. Ind. Eng. Chem.* **2020**, *84*, 42−45.

(24) Samkian, A. E.; Xu, Y.; Virgil, S. C.; Yoon, K.-Y.; Grubbs, R. H. Synthesis and Activity of Six-Membered Cyclic Alkyl Amino Carbene-

Ruthenium Olefin Metathesis Catalysts. *Organometallics* **2020**, *39*, 495−499.
(25) Nascimento, D. L.; Gawin, A.; Gawin, R.; Guńka, P. A.; Zachara, J.; Skowerski, K.; Fogg, D. E. Integrating Activity with Accessibility in Olefin Metathesis: An Unprecedentedly Reactive Ruthenium-Indenylidene Catalyst. *J. Am. Chem. Soc.* **2019**, *141*, 10626−10631.
(26) Zimmerer, J.; Pingen, D.; Hess, S. K.; Koengeter, T.; Mecking, S. Integrated Extraction and Catalytic Upgrading of Microalgae Lipids in Supercritical Carbon Dioxide. *Green Chem.* **2019**, *21*, 2428−2435.
(27) Małecki, P.; Gajda, K.; Gajda, R.; Woźniak, K.; Trzaskowski, B.; Kajetanowicz, A.; Grela, K. Specialized Ruthenium Olefin Metathesis Catalysts Bearing Bulky Unsymmetrical NHC Ligands: Computations, Synthesis, and Application. *ACS Catal.* **2019**, *9*, 587−598.
(28) Wyrębek, P.; Małecki, P.; Sytniczuk, A.; Kośnik, W.; Gawin, A.; Kostrzewa, J.; Kajetanowicz, A.; Grela, K. Looking for the Noncyclic-(Amino)(Alkyl)Carbene Ruthenium Catalyst for Ethenolysis of Ethyl Oleate: Selectivity Is on Target. *ACS Omega* **2018**, *3*, 18481−18488.
(29) Ambrosio, C.; Paradiso, V.; Costabile, C.; Bertolasi, V.; Caruso, T.; Grisi, F. Stable Ruthenium Olefin Metathesis Catalysts Bearing Symmetrical NHC Ligands with Primary and Secondary N-Alkyl Groups. *Dalton Trans.* **2018**, *47*, 6615−6627.
(30) Paradiso, V.; Bertolasi, V.; Costabile, C.; Caruso, T.; Dąbrowski, M.; Grela, K.; Grisi, F. Expanding the Family of Hoveyda-Grubbs Catalysts Containing Unsymmetrical NHC Ligands. *Organometallics* **2017**, *36*, 3692−3708.
(31) Ullah, A.; Arshad, M. Remarkably Efficient Microwave-Assisted Cross-Metathesis of Lipids under Solvent-Free Conditions. *ChemSusChem* **2017**, *10*, 2167−2174.
(32) Gawin, R.; Kozakiewicz, A.; Guńka, P. A.; Dąbrowski, P.; Skowerski, K. Bis(Cyclic Alkyl Amino Carbene) Ruthenium Complexes: A Versatile, Highly Efficient Tool for Olefin Metathesis. *Angew. Chem., Int. Ed.* **2017**, *56*, 981−986.
(33) Engl, P. S.; Fedorov, A.; Copéret, C.; Togni, A. N-Trifluoromethyl NHC Ligands Provide Selective Ruthenium Metathesis Catalysts. *Organometallics* **2016**, *35*, 887−893.
(34) Nieres, P. D.; Zelin, J.; Trasarti, A. F.; Apesteguía, C. R. Heterogeneous Catalysis for Valorisation of Vegetable Oils via Metathesis Reactions: Ethenolysis of Methyl Oleate. *Catal. Sci. Technol.* **2016**, *6*, 6561−6568.
(35) Marx, V. M.; Sullivan, A. H.; Melaimi, M.; Virgil, S. C.; Keitz, B. K.; Weinberger, D. S.; Bertrand, G.; Grubbs, R. H. Cyclic Alkyl Amino Carbene (CAAC) Ruthenium Complexes as Remarkably Active Catalysts for Ethenolysis. *Angew. Chem., Int. Ed.* **2015**, *54*, 1919−1923.
(36) Julis, J.; Bartlett, S. A.; Baader, S.; Beresford, N.; Routledge, E. J.; Cazin, C. S. J.; Cole-Hamilton, D. J. Selective Ethenolysis and Oestrogenicity of Compounds from Cashew Nut Shell Liquid. *Green Chem.* **2014**, *16*, 2846−2856.
(37) Baader, S.; Podsiadly, P. E.; Cole-Hamilton, D. J.; Goossen, L. J. Synthesis of Tsetse Fly Attractants from a Cashew Nut Shell Extract by Isomerising Metathesis. *Green Chem.* **2014**, *16*, 4885−4890.
(38) Shinde, T.; Varga, V.; Polášek, M.; Horáček, M.; Žilková, N.; Balcar, H. Metathesis of Cardanol over Ru Catalysts Supported on Mesoporous Molecular Sieve SBA-15. *Appl. Catal., A* **2014**, *478*, 138−145.
(39) Aydos, G. L. P.; Leal, B. C.; Perez-Lopez, O. W.; Dupont, J. Ionic-Tagged Catalytic Systems Applied to the Ethenolysis of Methyl Oleate. *Catal. Commun.* **2014**, *53*, 57−61.
(40) Bidange, J.; Dubois, J.-L.; Couturier, J.-L.; Fischmeister, C.; Bruneau, C. Ruthenium Catalyzed Ethenolysis of Renewable Oleonitrile. *Eur. J. Lipid Sci. Technol.* **2014**, *116*, 1583−1589.
(41) Zhang, J.; Song, S.; Wang, X.; Jiao, J.; Shi, M. Ruthenium-Catalyzed Olefin Metathesis Accelerated by the Steric Effect of the Backbone Substituent in Cyclic (Alkyl)(Amino) Carbenes. *Chem. Commun.* **2013**, *49*, 9491−9493.
(42) Miyazaki, H.; Herbert, M. B.; Liu, P.; Dong, X.; Xu, X.; Keitz, B. K.; Ung, T.; Mkrtumyan, G.; Houk, K. N.; Grubbs, R. H. Z-Selective Ethenolysis with a Ruthenium Metathesis Catalyst: Experiment and Theory. *J. Am. Chem. Soc.* **2013**, *135*, 5848−5858.
(43) Behr, A.; Krema, S.; Kämper, A. Ethenolysis of Ricinoleic Acid Methyl Ester - an Efficient Way to the Oleochemical Key Substance Methyl Dec-9-Enoate. *RSC Adv.* **2012**, *2*, 12775−12781.
(44) Park, C. P.; Van Wingerden, M. M.; Han, S.-Y.; Kim, D.-P.; Grubbs, R. H. Low Pressure Ethenolysis of Renewable Methyl Oleate in a Microchemical System. *Org. Lett.* **2011**, *13*, 2398−2401.
(45) Thomas, R. M.; Keitz, B. K.; Champagne, T. M.; Grubbs, R. H. Highly Selective Ruthenium Metathesis Catalysts for Ethenolysis. *J. Am. Chem. Soc.* **2011**, *133*, 7490−7496.
(46) Schrodi, Y.; Ung, T.; Vargas, A.; Mkrtumyan, G.; Lee, C. W.; Champagne, T. M.; Pederson, R. L.; Hong, S. H. Ruthenium Olefin Metathesis Catalysts for the Ethenolysis of Renewable Feedstocks. *CLEAN - Soil Air Water* **2008**, *36*, 669−673.
(47) Forman, G. S.; Bellabarba, R. M.; Tooze, R. P.; Slawin, A. M. Z.; Karch, R.; Winde, R. Metathesis of Renewable Unsaturated Fatty Acid Esters Catalysed by a Phoban-Indenylidene Ruthenium Catalyst. *J. Organomet. Chem.* **2006**, *691*, 5513−5516.
(48) Forman, G. S.; McConnell, A. E.; Hanton, M. J.; Slawin, A. M. Z.; Tooze, R. P.; van Rensburg, W. J.; Meyer, W. H.; Dwyer, C.; Kirk, M. M.; Serfontein, D. W. A Stable Ruthenium Catalyst for Productive Olefin Metathesis. *Organometallics* **2004**, *23*, 4824−4827.
(49) Rouge, P.; Szeto, K. C.; Bouhoute, Y.; Merle, N.; De Mallmann, A.; Delevoye, L.; Gauvin, R. M.; Taoufik, M. Ethenolysis of Renewable Methyl Oleate Catalyzed by Readily Accessible Supported Group VI Oxo Catalysts. *Organometallics* **2020**, *39*, 1105−1111.
(50) Tangyen, N.; Natongchai, W.; Del Gobbo, S.; D'Elia, V. Revisiting the Potential of Group VI Inorganic Precatalysts for the Ethenolysis of Fatty Acids through a Mechanochemical Approach. *ACS Omega* **2024**, *9*, 19712−19722.
(51) Martinez-Fernandez, M.; Yeamin, M. B.; Dalmau, D.; Carbó, J. J.; Poater, A.; Alegre-Requena, J. V. Data-Driven Analysis of Ni-Catalyzed Semihydrogenations of Alkynes. *Adv. Synth. Catal.* **2025**, *367*, No. e202401444.
(52) Engl, P. S.; Santiago, C. B.; Gordon, C. P.; Liao, W.-C.; Fedorov, A.; Copéret, C.; Sigman, M. S.; Togni, A. Exploiting and Understanding the Selectivity of Ru-N-Heterocyclic Carbene Metathesis Catalysts for the Ethenolysis of Cyclic Olefins to α,ω-Dienes. *J. Am. Chem. Soc.* **2017**, *139*, 13117−13125.
(53) Engl, P. S.; Santiago, C. B.; Gordon, C. P.; Liao, W.-C.; Fedorov, A.; Copéret, C.; Sigman, M. S.; Togni, A. Correction to "Exploiting and Understanding the Selectivity of Ru-N-Heterocyclic Carbene Metathesis Catalysts for the Ethenolysis of Cyclic Olefins to α,ω-Dienes.". *J. Am. Chem. Soc.* **2018**, *140*, 18227−18228.
(54) Yelchuri, V.; Srikanth, K.; Prasad, R. B. N.; Karuna, M. S. L. Olefin Metathesis of Fatty Acids and Vegetable Oils. *J. Chem. Sci.* **2019**, *131*, 39.
(55) Alexander, K. A.; Paulhus, E. A.; Lazarus, G. M. L.; Leadbeater, N. E. Exploring the Reactivity of a Ruthenium Complex in the Metathesis of Biorenewable Feedstocks to Generate Value-Added Chemicals. *J. Organomet. Chem.* **2016**, *812*, 74−80.
(56) Desnelli; Mujahidin, D.; Permana, Y.; Radiman, C. L. The Olefin Reaction between Crude Palm Oil Fatty Acid Methyl Ester (CPO FAME) and Ethylene Using Grubbs II Catalyst. *Procedia Chem.* **2015**, *17*, 44−48.
(57) Nickel, A.; Ung, T.; Mkrtumyan, G.; Uy, J.; Lee, C. W.; Stoianova, D.; Papazian, J.; Wei, W.-H.; Mallari, A.; Schrodi, Y.; Pederson, R. L. A Highly Efficient Olefin Metathesis Process for the Synthesis of Terminal Alkenes from Fatty Acid Esters. *Top. Catal.* **2012**, *55*, 518−523.
(58) Wolf, S.; Plenio, H. On the Ethenolysis of Natural Rubber and Squalene. *Green Chem.* **2011**, *13*, 2008−2012.
(59) Dalmau, D.; Alegre-Requena, J. V. ROBERT: Bridging the Gap between Machine Learning and Chemistry. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2024**, *14*, No. e1733.
(60) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.

https://doi.org/10.1021/acscatal.5c06483
ACS Catal. 2026, 16, 2160−2170

(61) Lundberg, S. M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J. M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nat. Mach. Intell.* **2020**, *2*, 56−67.
(62) Dalmau, D.; Sigman, M. S.; Alegre-Requena, J. V. Machine Learning Workflows beyond Linear Models in Low-Data Regimes. *Chem. Sci.* **2025**, *16*, 8555−8560.
(63) Hutson, M. Artificial Intelligence Faces Reproducibility Crisis. *Science* **2018**, *359*, 725−726.
(64) Janet, J. P.; Kulik, H. J. *Machine Learning in Chemistry*; ACS In Focus; American Chemical Society: Washington, DC, 2020.
(65) Alegre-Requena, J. V.; Sowndarya S. V, S.; Pérez-Soto, R.; Alturaifi, T. M.; Paton, R. S. AQME: Automated Quantum Mechanical Environments for Researchers and Educators. *WIREs Comput. Mol. Sci.* **2023**, *13*, No. e1663.
(66) Landrum, G. *RDKit: Open-Source Cheminformatics*. https://www.rdkit.org (accessed July 21, 2025).
(67) Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. Extended Tight-Binding Quantum Chemistry Methods. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2021**, *11*, No. e1493.
(68) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-XTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652−1671.
(69) Monreal-Corona, R.; Pla-Quintana, A.; Poater, A. Predictive Catalysis: A Valuable Step towards Machine Learning. *Trends Chem.* **2023**, *5*, 935−946.
(70) Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189−1232.
(71) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5−32.
(72) Jawiczuk, M.; Marczyk, A.; Młodzikowska-Pieńko, K.; Trzaskowski, B. Impact of the Carbene Derivative Charge on the Decomposition Rates of Hoveyda-Grubbs-like Metathesis Catalysts. *J. Phys. Chem. A* **2020**, *124*, 6158−6167.
(73) Trzaskowski, B.; Ostrowska, K. Structural Analogues of Hoveyda-Grubbs Catalysts Bearing the 1-Benzofuran Moiety or Isopropoxy-1-Benzofuran Derivatives as Olefin Metathesis Catalysts. *RSC Adv.* **2016**, *6*, 21423−21429.
(74) Trzaskowski, B.; Grela, K. Hoveyda-Grubbs Complexes with Boryl Anions Are Predicted to Be Fast Metathesis Catalysts. *Catal. Commun.* **2016**, *86*, 133−138.
(75) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16*, rev. C.01; Gaussian Inc.: Wallingford, CT, 2019.
(76) Becke, A. D. Density-Functional Thermochemistry. V. Systematic Optimization of Exchange-Correlation Functionals. *J. Chem. Phys.* **1997**, *107*, 8554.
(77) Wu, Q.; Yang, W. Empirical Correction to Density Functional Theory for van Der Waals Interactions. *J. Chem. Phys.* **2002**, *116*, 515.
(78) Grimme, S. Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27*, 1787−1799.
(79) Chai, J.-D.; Head-Gordon, M. Long-Range Corrected Hybrid Density Functionals with Damped Atom-Atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615−6620.
(80) Schäfer, A.; Horn, H.; Ahlrichs, R. Fully Optimized Contracted Gaussian Basis Sets for Atoms Li to Kr. *J. Chem. Phys.* **1992**, *97*, 2571−2577.
(81) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305.
(82) Andrae, D.; Häußermann, U.; Dolg, M.; Stoll, H.; Preuß, H. Energy-Adjusted Ab Initio Pseudopotentials for the Second and Third Row Transition Elements. *Theor. Chim. Acta* **1990**, *77*, 123−141.
(83) Occhipinti, G.; Bjørsvik, H.-R.; Jensen, V. R. Quantitative Structure-activity Relationships of Ruthenium Catalysts for Olefin Metathesis. *J. Am. Chem. Soc.* **2006**, *128*, 6952−6964.
(84) Escayola, S.; Bahri-Laleh, N.; Poater, A. %V_{Bur} Index and Steric Maps: From Predictive Catalysis to Machine Learning. *Chem. Soc. Rev.* **2024**, *53*, 853−882.
(85) Martínez, J. P.; Trzaskowski, B. Electrophilicity of Hoveyda-Grubbs Olefin Metathesis Catalysts as the Driving Force That Controls Initiation Rates. *ChemPhysChem* **2022**, *23*, No. e202200580.
(86) Credendino, R.; Falivene, L.; Cavallo, L. π-Face Donation from the Aromatic N-Substituent of N-Heterocyclic Carbene Ligands to Metal and Its Role in Catalysis. *J. Am. Chem. Soc.* **2012**, *134*, 8127−8135.
(87) Liu, P.; Xu, X.; Dong, X.; Keitz, B. K.; Herbert, M. B.; Grubbs, R. H.; Houk, K. N. Z-Selectivity in Olefin Metathesis with Chelated Ru Catalysts: Computational Studies of Mechanism and Selectivity. *J. Am. Chem. Soc.* **2012**, *134*, 1464−1467.
(88) Nelson, D. J.; Manzini, S.; Urbina-Blanco, C. A.; Nolan, S. P. Key Processes in Ruthenium-Catalysed Olefin Metathesis. *Chem. Commun.* **2014**, *50*, 10355−10375.
(89) Foscato, M.; Jensen, V. R. Automated in Silico Design of Homogeneous Catalysts. *ACS Catal.* **2020**, *10*, 2354−2377.
(90) Butilkov, D.; Frenklah, A.; Rozenberg, I.; Kozuch, S.; Lemcoff, N. G. Highly Selective Olefin Metathesis with CAAC-Containing Ruthenium Benzylidenes. *ACS Catal.* **2017**, *7*, 7634−7637.
(91) Jawiczuk, M.; Młodzikowska-Pieńko, K.; Osella, S.; Trzaskowski, B. Molecular Modeling of Mechanisms of Decomposition of Ruthenium Metathesis Catalysts by Acrylonitrile. *Organometallics* **2020**, *39*, 239−246.
(92) Poater, A.; Ragone, F.; Correa, A.; Cavallo, L. Exploring the Reactivity of Ru-Based Metathesis Catalysts with a π-Acid Ligand Trans to the Ru-Ylidene Bond. *J. Am. Chem. Soc.* **2009**, *131*, 9000−9006.

https://doi.org/10.1021/acscatal.5c06483
ACS Catal. 2026, 16, 2160−2170