

This is an author peer-reviewed manuscript version of the following article, accepted for publication in *Trends in Chemistry* and published online by Elsevier, in December 2023:

Monreal-Corona, R.; Pla-Quintana, A. i Poater, A. (2023). Predictive catalysis: a valuable step towards machine learning. *Trends In Chemistry*, 5 (12), p. 935-946. Available online at <https://doi.org/10.1016/j.trechm.2023.10.005>

The final published version of the article is available to subscribers, or by PDF purchase, at <https://doi.org/10.1016/j.trechm.2023.10.005>. © 2023 Elsevier Inc. All rights reserved.

© 2023 This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>



1 **Predictive Catalysis: a Valuable Step towards Machine**

2 **Learning**

3 Roger Monreal-Corona, Anna Pla-Quintana and Albert Poater*

4 Institut de Química Computacional i Catàlisi and Departament de Química, Universitat
5 de Girona, C/ Maria Aurèlia Capmany 69, 17003 Girona, Catalonia, Spain

6 albert.poater@udg.edu

7 8 **ABSTRACT**

9 As physical chemistry makes the transition to computational chemistry, a new growth in
10 the field is occurring. With the advent of predictive catalysis, computational chemistry is
11 becoming a key player in the optimization and development of catalytic processes.
12 Predictive catalysis refers to the use of computational and theoretical methods to predict
13 the properties and behavior of chemical systems, and more specifically, their catalytic
14 activity and selectivity. In this analysis, we take a look at what predictive catalysis has
15 done to date, and build a picture of how far it can go in the future, while also outlining
16 the challenges that need to be resolved to make it a powerful tool of general applicability.

17
18 **KEYWORDS:** predictive catalysis; DFT; computational chemistry; predictive
19 chemistry; generative chemistry; machine learning

20 21 **The advent of predictive catalysis**

22 Computational chemistry is a rapidly evolving field that uses computational methods to
23 study the structure and properties of molecules and solids [1]. Already in the 1960s, at
24 the dawn of computational chemistry, the development of theoretical methods to calculate
25 the electronic structure of molecules was used to understand and anticipate the reactivity
26 of chemical systems, laying the foundations of predictive chemistry. As computer power
27 and software improved over time, scientists were able to develop more sophisticated
28 methods for predicting the behavior of chemical systems, including the role of catalysts
29 in chemical reactions. In the late 20th century, advances in computational chemistry [2],
30 along with the growth of the field of catalysis [3], led to the development of predictive
31 catalysis as a distinct area of research.

32 The ultimate goal when developing a catalytic process is to grant access to new
33 compounds or discover new methodologies to synthesize known compounds in a more

34 efficient and sustainable way. Most advances are still made through trial-and-error
35 experimentation, although this is a significantly time consuming and expensive approach
36 to the problem. In contrast, predictive catalysis provides molecular-level insights into the
37 catalytic process and helps identify the key parameters that influence activity and
38 selectivity. This information is then used to predict the activity or selectivity of new
39 substrates or catalysts and thus reduce experimental effort.

40 But why stop at predictive catalysis and not dive into machine learning (ML)? ML is
41 already being used for reaction deployment [4,5], reaction development [6,7], and even
42 reaction discovery [8] following the classification recently proposed by Coley and
43 coworkers [9]. Although ML is helping the development of synthetic chemistry, more
44 accurate algorithms derived from more complete data sets are still needed. Additionally,
45 there is a need to identify faulty models of existing techniques, which will help push
46 predictive chemistry beyond current limits. Therefore, there is a lot of room for predictive
47 catalysis to expand, especially in basic research because reactions are rarely tested with
48 10 catalysts or more. This number is far from enough to build training, testing and
49 validation data sets, so a predictive catalysis approach that includes experimental testing
50 is much more appropriate in this scenario.

51 Indeed, predictive catalysis has become an important tool used in a wide range of
52 industrial and academic applications [10,11]. However, there are still a limited number of
53 articles that include appropriate specific experiments based on the calculations
54 performed. In other words, computational chemistry is very often used to predict, but still
55 in some cases the lack of synergy between calculations and experiments means that
56 experiments are not performed to corroborate the computational results. This only
57 represents a success in specific cases where no significant changes in the energy of the
58 rate determining step [12,13] or no clear trends in the electronic or geometric effect [14]
59 are observed, in which case the calculations indeed avoid inefficient and unnecessary
60 experiments. In all other cases it is a failure, since measures that could potentially inform
61 new or better results have not been attempted [15,16,17]. Although in some cases the
62 absence of experiments can only be explained because they have not yet aroused
63 sufficient interest [18,19], in other cases this comes down to the inability to efficiently
64 synthesize the starting materials or catalysts [84,20,21], or making computational
65 proposals that are too far from what is synthetically possible.

66 In this review we will focus our attention to works reported in the literature where
67 theoretical predictive results have ultimately been useful to discover new catalytic

68 reactivity as examples of predictive catalysis. Through the study of these works, we will
69 analyze important past studies and results, and determine where more focus is needed to
70 overcome the problems of predictive catalysis, and turn it into a truly useful step that fills
71 the gaps where machine learning cannot yet be applied.

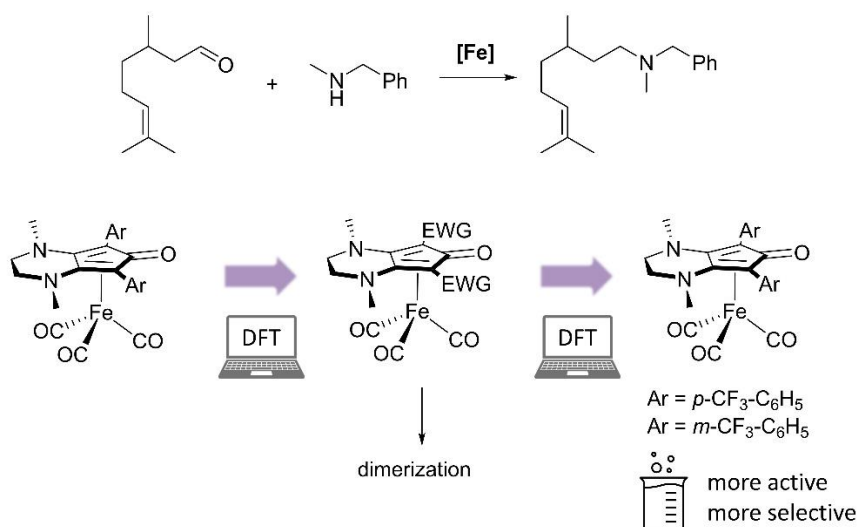
72

73 **A path towards optimization: predictive catalysis by DFT calculations**

74 Predictive catalysis by Density Functional Theory (DFT) calculations refers to the use of
75 computational methods to reveal the mechanism of a transformation and use the
76 mechanistic information to predict the outcome when varying the parameters. It involves
77 the use of DFT calculations to understand the interactions between the catalyst and the
78 reactants, and how these interactions influence reaction kinetics, product distribution, and
79 selectivity with the aim of using this knowledge to develop more efficient and sustainable
80 catalytic reactions. Below are three case studies that exemplify the use of this pathway
81 towards the optimization of catalytic processes.

82 A series of publications led by the Poater and Renaud teams focused on the improvement
83 of Knölker-type iron catalysts for a reductive amination reaction. In this work, DFT
84 calculations together with a detailed analysis of the chemical structure in terms of
85 geometry, fragment partial charges, effective oxidation states, and aromaticity were used
86 to understand the mechanism for the reductive amination of aliphatic carbonyl
87 compounds catalyzed by a Knölker-type iron catalyst (see Figure 1) [22]. With the
88 reaction mechanism in hand, different catalyst modifications were explored with the aim
89 of guiding the catalytic reactions towards milder conditions. It was found that the
90 presence of electron-withdrawing (EWG) substituents on the cyclopentadienone ring
91 would lower the activation barrier of the rate determining step (rds). In a follow-up work,
92 the synthesis of the catalysts that were predicted to show the highest activities was
93 experimentally attempted [23]. However, an undesired dimerization of the catalyst was
94 observed, leading the authors to change strategy. The reintroduction of the two phenyl
95 substituents present on the cyclopentadienone ring of the initial catalyst structure was
96 undertaken, and also the introduction of EWG substituents on these rings, instead of
97 inserting them directly on the cyclopentadienone ring. The new catalysts were
98 experimentally tested and showed a significant increase in catalytic activity on the
99 reductive amination of certain substrates. This is a clear example that predictive chemistry
100 with a high degree of synergy between calculations and experiments is extremely useful
101 for adapting the theoretically formulated hypothesis to experimentally feasible and useful

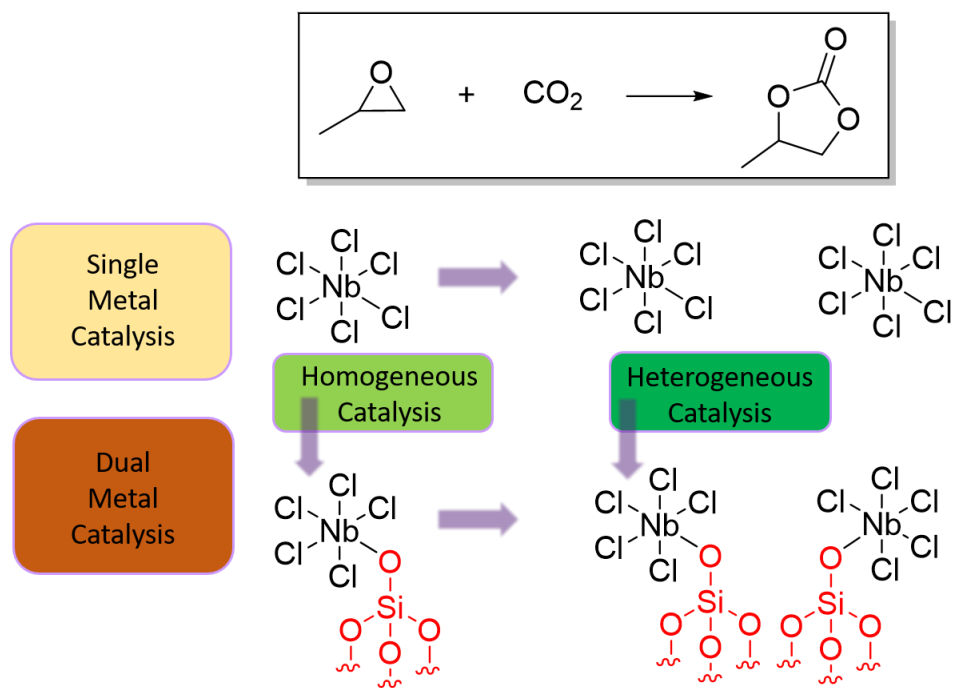
102 results. Along the same lines, several studies carried out by the scientific community
 103 demonstrate the use of predictive catalysis by DFT for the palladium-catalyzed
 104 hydrogenation of polyalphaolefins using palladium catalysts, with the aim of achieving
 105 new lubricants [24].
 106



107
 108 **Figure 1.** Two-step predictive catalysis approach for hydrogenation reactions with
 109 Knölker-type catalysts. A series of derivative Renaud catalysts with electron-withdrawing
 110 groups on the aryls of the cyclopentadienone improves the activity and selectivity [22,23].
 111

112 A second case study highlights the importance of elucidating the mechanism to propose
 113 a new catalyst system with better performance. In 2014 D'Elia et al. showed that the
 114 fixation of CO₂ by reaction with epoxides to form cyclic carbonates occurred not only
 115 with one NbCl₅ unit as the nucleophilicity activating agent on the epoxide oxygen atom,
 116 but that a second unit assists in the rate-determining step of the cyclization [25]. One year
 117 later, the authors decided to develop a heterogeneously catalyzed version of the reaction
 118 by immobilizing the NbCl₅ catalyst on silica. Based on the previous results that showed
 119 dual catalysis, as well as using the typical silica support SiO₂₋₇₀₀ the catalyst was also
 120 immobilized onto SiO₂₋₂₀₀, to bring the active centers closer together, obtaining
 121 remarkably higher efficiency as a result of vicinal niobium centers [26] (see Figure 2).
 122 An analogous change to a bimetallic mechanism was described in the hydrophenoxylation
 123 of alkynes catalyzed by a family of cationic NHC complexes [(NHC)Au-OH-
 124 Au(NHC)]⁺ [27]. Dual catalysis is a great way to trigger new reactivity to improve on

125 reported studies [28], although it should be noted that it is difficult to join two metal
 126 centers, especially in heterogeneous supports.



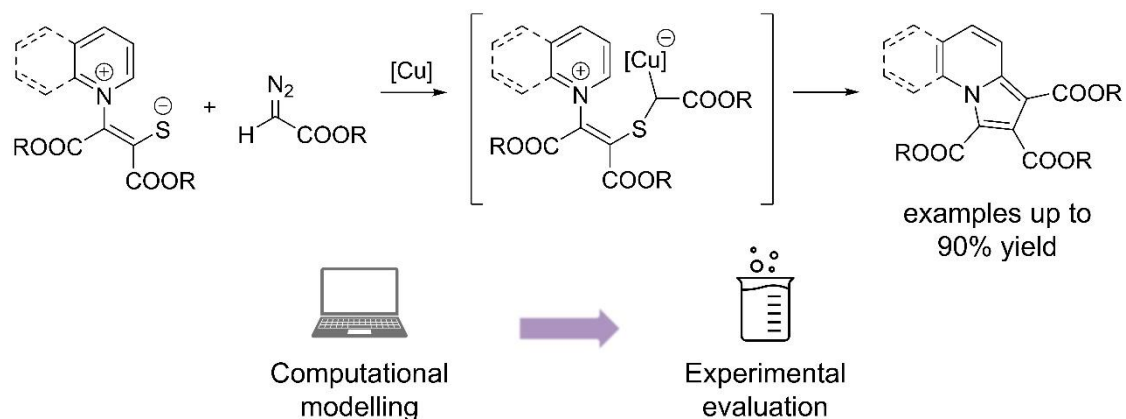
127

128 **Figure 2.** The evolution from homogeneous to heterogeneous catalysis in the fixation of
 129 CO₂ with epoxides together with dual metal catalysis that overcame the single atom
 130 catalysis by bending the CO₂. In the heterogeneous systems the replacement of SiO₂₋₇₀₀
 131 by SiO₂₋₂₀₀ improved the reactivity demonstrating that the heterogeneous system is more
 132 efficient and follows a dual metal catalysis mechanism as in homogeneous catalysis
 133 [25,26].

134

135 Finally, a third example shows how the use of predictive catalysis can accelerate the
 136 process of developing new reactivity by reducing the number of experiments performed.
 137 The study was focused on the synthesis of indolizines from the reaction of pyridinium
 138 1,4-zwitterionic thiolates (PZTs) and copper carbenes [29]. Before the study began, PZTs
 139 had been shown to provide access to a wide diversity of products, but the reactivity with
 140 metal carbenes remained unexplored. Thus, the first step consisted of evaluating the
 141 feasibility of the annulation reaction of PZTs and copper carbenes by means of theoretical
 142 calculations. A plausible mechanism was obtained with affordable reaction barriers at
 143 moderate temperature. With this information in hand, the rds was identified and different
 144 copper salts, PZTs and diazo derivatives were modelled to predict their performance
 145 toward the transformation of interest. Experiments were then performed to validate the
 146 accuracy of the predictions with regards to the efficiency of different copper salts (see

147 Figure 3), and with the optimized reaction conditions, the scope of the reaction was
148 evaluated.



149

150 **Figure 3.** Predictive catalysis work for the reaction of pyridinium 1,4-zwitterionic
151 thiolate and copper carbenes [29], leading to yields up to 90%.

152

153 It should be noted that, in this last example, the yields did not match with the predicted
154 rds energies for several of the different substrates being evaluated. This brings to light
155 one of the main problems of predictive catalysis: the oversight of alternative and
156 degradation pathways, in the previous case carbene dimerization. The advent of
157 automated reaction discovery programs although currently not sufficiently developed to
158 analyze all types of reactions, could be a powerful tool to solve this procedural flaw
159 [30,31,32]. Methods towards automatic search of reaction mechanisms generated by
160 Morokuma and coworkers [33,34] or conceptual DFT [39,35,36] should also help in this
161 task. In the meantime, thorough analysis of the results [37], and use of reactivity
162 knowledge to propose alternative or degradation pathways, remains the only solution.

163

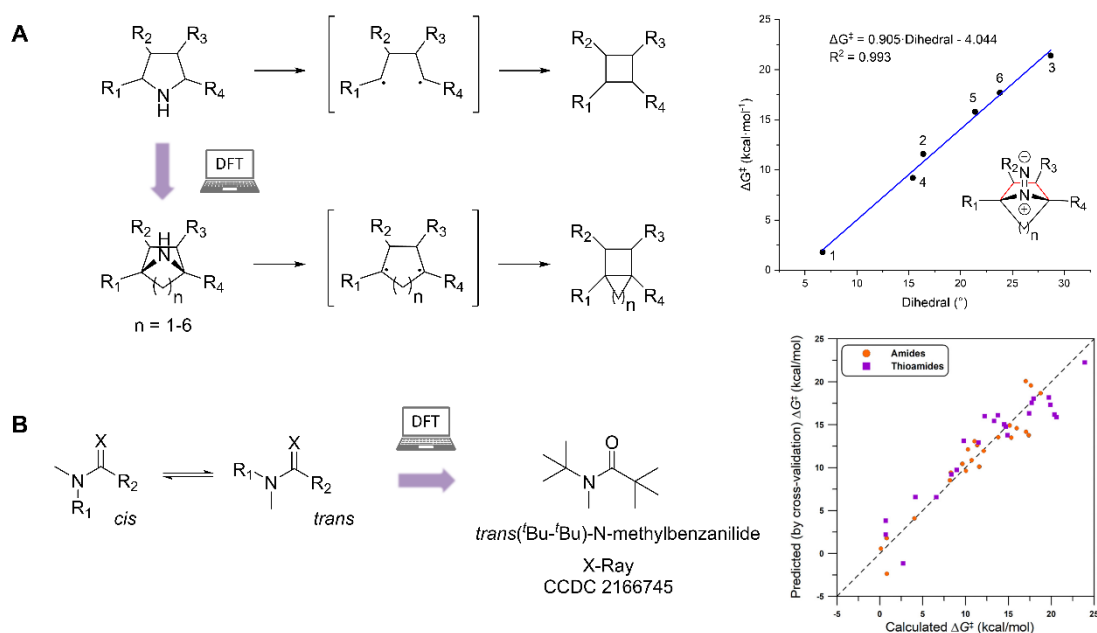
164 **Use of linear regression modelling to predict yield and enantioselectivity**

165 The correlations produced by equating molecular parameters with experimental outcomes
166 can be used, in a very intuitive way, to evaluate the origin of selectivity and generate new,
167 experimentally-testable hypotheses. This contrasts with ML, which, although an
168 extremely powerful methodology, produces results that are challenging to interpret
169 directly. There are multiple molecular parameters, some of them derived after quantitative
170 structure-activity relationships (QSAR) [38,39,40], and these can be obtained in different
171 ways, known as generative chemistry in the drug development arena [41,42]. One
172 example is the web server designed and created by Cavallo and collaborators to generate

173 steric maps [43,44,45] and steric indices such as the %V_{Bur} parameter [46]. These maps
174 and indices help to understand the pockets that the ligands create around a metal in a
175 catalyst to predict the reactivity that occurs in the first coordination sphere where the
176 metal center interacts with incoming substrates. Another example of a web server, this
177 time focused on the prediction of electronic parameters developed by Contreras, is
178 designed to calculate Non-covalent Interactions (NCIs) and generate NCI plots [47]. Non-
179 covalent interactions such as hydrogen bonding, van der Waals forces, and π - π
180 interactions, among others [48], play a crucial role in understanding molecular
181 interconnection. In addition, other web tools like the energetic span tool by Shaik and
182 Kozuch [49,50], and various tools by the Duarte [51,52] and Paton groups [53], with the
183 predictions of the Sterimol steric index, as well as the Kraken by Gensch, Sigman and
184 Aspuru-Guzik groups [54] provide a wealth of molecular parameters.

185 Pla-Quintana, Poater and coworkers, described in various case studies the potential of
186 these correlations to rationalize reaction yield and predict new reactivity. An initial study
187 dealt with photoredox chemistry, where the mechanism for the photocatalyzed reaction
188 of quinoxalinones with a hypervalent iodine(III) reagent to form triazoloquinoxalinones
189 was unveiled [55]. A linear relationship between the kinetic cost and the energy of the
190 LUMO and the electrophilicity of the key radical species was found. Multilinear
191 regressions were performed and the best agreement was obtained with three mentioned
192 descriptors. In addition, predictive chemistry gave the recipe for the electronic and steric
193 requirements of the substituents that led to a faster reaction. The same authors broadened
194 the definition of "predictive chemistry" by evaluating processes that do not involve the
195 use of a catalyst. A computational study of the functionalization of nitrous oxide for the
196 synthesis of triazolopyridine and triazoloquinoline scaffolds was reported [56]. After
197 identifying the rds of the transformation, the activation energy was correlated to the
198 experimental yields, allowing for direct prediction of the experimental performance from
199 kinetic calculations. Moreover, the feasibility of preparing triazolopyrazine,
200 triazolopyridazine, triazolopyridimidine and triazolotriazine scaffolds was evaluated by
201 applying the reported methodology to the diazine and triazine family of heterocycles,
202 demonstrating that the HOMO energy remains constant while the energy of the LUMO
203 decreases with increasing number of nitrogen atoms in the structure. Another work delved
204 into the stereoretentive formation of cyclobutanes from pyrrolidines [57]. The
205 computational study helped to rationalize the mechanism, as well as to propose the
206 structure of the species responsible for nitrogen transfer to generate the 1,1-diazene

207 reaction intermediate. In this case, the knowledge extracted from calculations was used
 208 to predict that the methodology could be amenable to the synthesis of [2]-ladderanes and
 209 bicyclic cyclobutanes, for which synthesis is not straightforward with the currently
 210 available methodologies (see Figure 4A). By studying the concerted decomposition of the
 211 1,1-diazene to the 1,4-biradical intermediate, the desirable geometrical parameters to
 212 obtain the desired product were ascertained, which is an important learning for future
 213 studies.
 214



215
 216 **Figure 4.** A) Predictive chemistry approach by DFT for the synthesis of [2]-ladderanes
 217 and bicyclic cyclobutanes [56] and B) computational search for *trans* (thio)amides leading
 218 to the synthesis of *trans*(*t*Bu-*t*Bu)-*N*-methylbenzanilide, fully characterized by X-Ray
 219 [57]..

220
 221 Another study confirmed the utility of predictive chemistry at the borders of organic and
 222 computational chemistry, unveiling the factors that lie behind the higher preference of the
 223 *cis* over the *trans* conformation in (thio)amides [58]. Through a guided predictive
 224 chemistry strategy, the effect of alkyl/aryl substituents on the carbon and nitrogen atoms
 225 on the stability of the isomers was assessed, substituents that favor the *trans* isomer
 226 identified, and the hypothesis checked by experimental synthesis of the best candidate,
 227 that had sterically demanding *t*-butyl substituents (see Figure 4B).

228 In another study, Bertrand and collaborators introduced cyclic (alkyl)(amino)carbenes
229 with a six-membered backbone (CAAC-6) as catalysts for the palladium-mediated α -
230 arylation of ketones with aryl chlorides [59], contrasting them with their five-membered
231 counterparts (CAAC-5). CAAC-6 exhibited increased %V_{Bur} and enhanced donor and
232 acceptor properties, supported by a visible spectrum $n \rightarrow \pi^*$ transition. Its unique
233 ambiphilic quality allowed it to insert intramolecularly into inactive C(sp³)-H bonds. As
234 ligands, CAAC-6 demonstrated superiority over CAAC-5 in palladium-mediated α -
235 arylation of ketones with aryl chlorides.

236 While predicting reaction yield tends to be difficult due to many on- and off-cycle events
237 that can substantially alter it, product selectivity is more reliable as it is determined by a
238 small number of elementary steps. Sigman and coworkers have applied multivariate linear
239 and polynomial regression analyses to optimize reaction selectivity by predicting catalyst,
240 ligand, and substrate effects. They elucidate substrate trends in enantioselective reaction
241 outcomes based on vibrational and Sterimol parameters [60,61] as quantitative
242 descriptors of mechanistic significance. In all the case studies, that span from peptide
243 catalyzed desymmetrization of bisphenols to rhodium-catalyzed asymmetric transfer
244 hydrogenation, the model validity was determined by assessing how well the model
245 predicts the reaction outcomes of new substrates. They then include the catalyst in their
246 studies to facilitate catalyst optimization with a data-intensive approach [62].
247 Identification of the interactions between substrate and catalyst in a BINOL-based
248 phosphoric acid catalyzed cross dehydrogenative coupling reaction, supported by
249 experimental evidence, afforded an efficient means of optimizing catalyst design with
250 reduced catalytic insight. Furthermore, the enantioselectivity was improved capitalizing
251 on the prediction that torsion of the triazole ring with respect to the phenyl leads to
252 enhanced enantioselectivity. The Sigman group is continuing to use data science
253 techniques to guide catalyst screening and explore the scope of the processes they develop
254 (for recent examples, see [63,64]).

255 Also in the field of data science, Paton and co-workers reported quantum chemical
256 calculations for organic radical species and associated closed-shell molecules [65], and
257 described how graph neural networks predictions can be incorporated into
258 mechanistically informed statistical models of chemical reactivity and selectivity
259 [66,67,68].

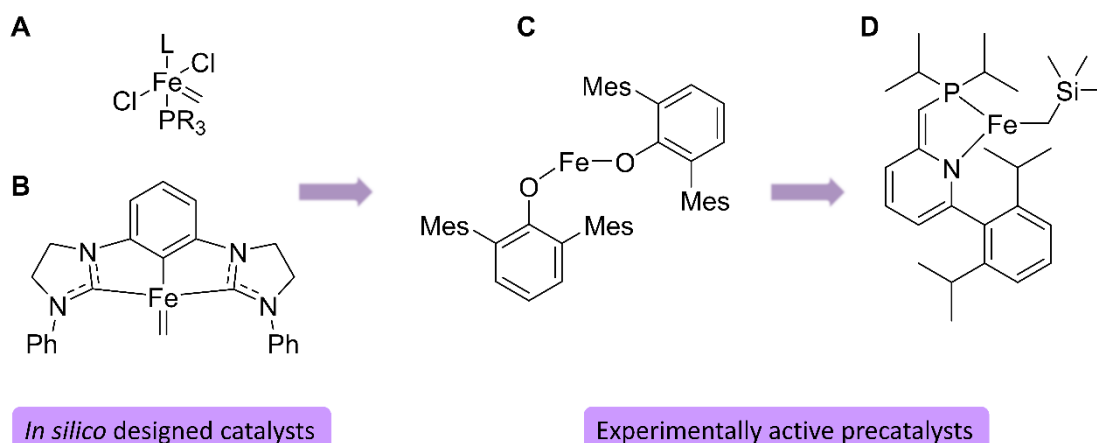
260

261 **Further applications of Predictive Catalysis**

262 **Metathesis reactions: new catalysts and tailored stereoselectivity**

263 There are many examples of the use of predictive catalysis in the well-studied olefin
264 metathesis reaction. Predictive catalysis was used to develop iron-based catalysts for this
265 purpose. Work by Poater et al. initially described the potential of Fe-based complexes as
266 olefin metathesis catalysts [69] and was followed by more computational efforts by
267 Solans-Monfort and coworkers [70]. In 2021 Bukhryakov and coworkers experimentally
268 reported that iron complexes were able to catalyze the ring-opening metathesis
269 polymerization of norbornene [71](see Figure 5), and in 2022 Milstein and coworkers
270 achieved a total of 880 TONs for this reaction [72]. It is important to point out that in this
271 case the initial theoretical works only served as a basis. They introduced the idea of
272 replacing the central metal in the catalyst, but it was necessary to experimentally exclude
273 the NHC ligands, present in the computational studies, from the iron catalyst to observe
274 reactivity. This was counterintuitive since NHC ligands are the most commonly used in
275 ruthenium olefin metathesis catalysts [73]. In fact, NHC ligands induced an
276 overstabilization of the intermediate metallacycle described in the beginning of the
277 reaction by Chauvin [74].

278



279

280 **Figure 5.** Evolution towards Fe-based olefin metathesis: In silico Fe-based (pre)catalysts
281 for olefin metathesis in silico designed, A: homologous to the 2nd Grubbs generation
282 catalyst (NHC = N-heterocyclic carbene) [75]; B: pincer based including the phosphine-
283 phosphonium ylide [69]; and experimentally tested, C: bis-alkoxide catalyst [71] and the
284 three-coordinate iron(II) catalysts for ring-opening metathesis polymerization of olefins
285 [72].

286

287 On the other hand, the Jensen group used predictive catalysis to improve the selectivity
288 of ruthenium-based metathesis catalysts. They developed, through iterative
289 computational prediction and experimental follow-up, new ruthenium complexes that are
290 able of catalyzing the metathesis homocoupling of terminal olefins to achieve *Z* alkenes
291 [76,77,78,79,80] and more recently *E* alkenes [81] in a highly selective manner. Since
292 stereoselective metathesis is achieved by controlling the stereochemistry in the rate-
293 determining step of the Chauvin mechanism [74], typically cycloreversion to release the
294 olefin product, the authors carried out a comprehensive study by modifying the steric
295 hindrance and electronics of the anionic ligands (from chloride to thiolate and isocyanate)
296 and neutral ligands (carbene, phosphines and pyridine) ligands. Although a good
297 correlation between predictions and experimental results was achieved, in certain cases
298 decomposition of the catalytic species leading to metathesis-active complexes caused the
299 predictions to fail.

300 Some works in this area have focused extensively on the computational studies, but have
301 not yet piqued interest experimentally even though they could be extremely valuable. For
302 example, there are in-silico studies of a highly sterically hindered second-generation
303 Grubbs catalyst with a C60-appended NHC ligand [82,83], and also of two novel bis-
304 ylidene ligands postulated to be active metathesis catalysts computationally, but their
305 activity and selectivity has not yet been experimentally tested [84].

306

307 ***Supra/macromolecular research: modification of carbon-based structures***

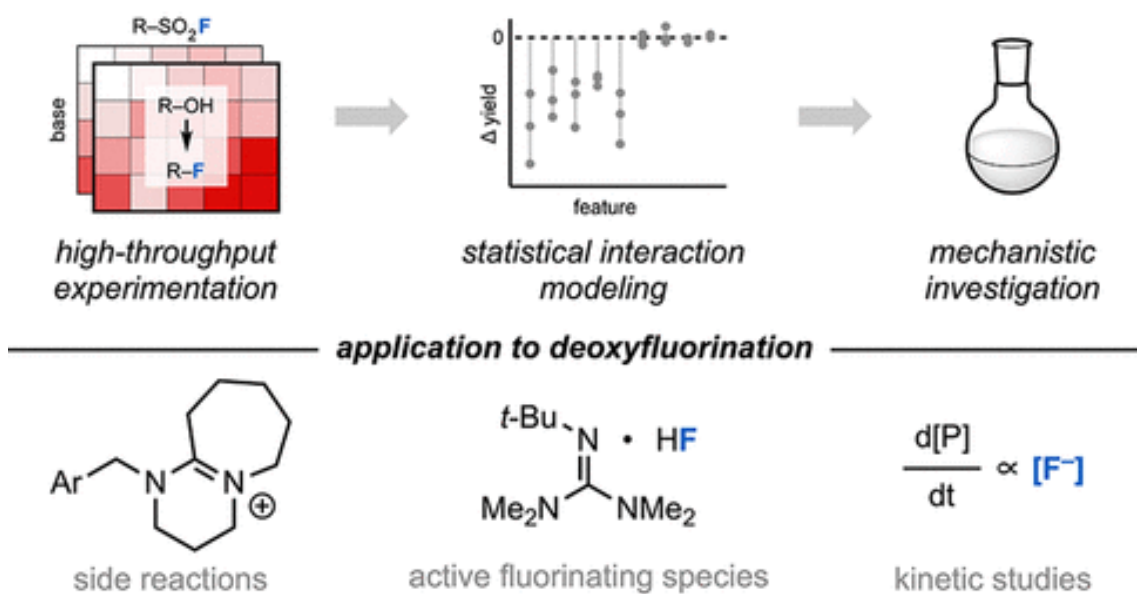
308 A second set of examples builds on the chemical modification of carbon-based materials.
309 The groups of Solà and Langa joined efforts to evaluate the feasibility of modifying
310 single-walled carbon nanotubes (SWCNTs) through a Pauson-Khand reaction (PKR)
311 [85]. An initial DFT study on different types of carbon nanotubes (zig-zag, armchair, and
312 chiral) revealed a new reaction pathway for the PKR with lower energy barriers,
313 alternative to the widely-accepted Magnus mechanism. The calculations estimated
314 reasonable energy barriers at room temperature and suggested that semiconducting
315 SWCNTs exhibit higher reactivity than metallic SWCNTs. Based on the theoretical
316 prediction, using the theoretically evaluated catalytic system, cyclopentenones were
317 successfully attached to SWCNTs and the resulting compounds characterized [85]. This
318 is a good example of the positive impact of putting together a collaborative theoretical-
319 experimental team for the successful development of new reactivity.

320 Other carbon nanostructures such as fullerenes can also be modified by transition-metal
321 catalyzed cyclization reactions. In this topic, Solà, Roglans and Lledó studied the
322 functionalization of C₆₀ through a [2+2+2] cycloaddition reaction. In an initial predictive
323 study, whether the reaction could be accomplished with rhodium in a catalytic fashion
324 was computationally evaluated [86]. The transformation had been previously reported,
325 but only with stoichiometric amounts of nickel or palladium and in a process with very
326 limited substrate scope. Once the reactivity was predicted as feasible, the same group
327 carried out the transformation in the laboratory; however, the reaction did not stop at the
328 [2+2+2] cycloaddition product, and further evolved through a formal [4+4]/retro [2+2+2]
329 rearrangement to open the fullerene cage. These functionalized fullerenes could be used
330 to enhance the open-circuit voltage of perovskite solar cells [87]. Key intermediates in
331 Pauson-Khand and [2+2+2] cycloaddition reactions are rhodacyclopentenes. Their
332 formation by oxidative cyclization has been computationally studied by Baik and co-
333 workers, and two mechanistic scenarios – closed-shell and open-shell – have been
334 recognized as leading to different reaction optimization and control strategies [88].
335 Predictive catalysis can also be applied to larger systems. In 2020, Liang et al. reported
336 the use of computational tools to study the adsorption and catalysis of MnN₄ embedded
337 in carbon structures such as graphene, graphene nanoribbons, graphene nanosheets,
338 carbon nanotubes and C₆₀ fullerenes for the absorption of toxic gases and their subsequent
339 transformation [89]. In the first part of their work, DFT calculations were conducted to
340 model the MnN₄-decorated carbon structures and evaluate the adsorption energy, charge
341 transfer, sensitivity, and Gibbs free energy. The absorption properties were found to be
342 highly dependent on the doped atoms, curvature, and the size of the substrate. Among all
343 the supports tested, the MnN₄-decorated graphene, graphene nanoribbons, and graphene
344 nanosheets were labelled as the more effective towards catalytic CO oxidation. In the
345 second part of the work, molecular dynamics (MD) simulations were performed to
346 determine the operating temperatures of gas sensors, valuable information necessary for
347 the design of new high-performance gas sensors. MD were used to speed up the
348 calculation in such large and computationally demanding systems.

349 ***Cross-coupling catalysts: filling the gap to ML***

350 Finally, a notable set of examples focuses on the development of efficient cross-coupling
351 catalysts. The Doyle group have worked intensively on the development of nickel
352 catalysts that are active in Suzuki-Miyaura cross-couplings. Based on the results of a
353 parametrization and modelling study demonstrating that remote steric hindrance was key

354 to activity, they developed a novel class of aryl alkyl phosphine ligands, named DinoPhos
 355 [90]. These new monodentate phosphine ligands, with both a large cone angle and small
 356 %V_{Bur} value [46], efficiently catalyze the Ni-catalyzed cross-coupling of acetals with aryl
 357 boroxines. In a follow-up study, the groups of Doyle and Sigman, in collaboration with
 358 Merck & Co, capitalized on the observation that preferential formation of monoligated
 359 nickel complexes was essential for reactivity, and that the formation of bisligated versus
 360 monoligated metal species correlated perfectly with the %V_{Bur(min)} value [91]. This
 361 correlation is valid not only for nickel complexes but also for palladium ones, which
 362 curiously require two ligands to show catalytic activity. A powerful predictive parameter
 363 of activity was thus determined. More recently, the Doyle group gave more insight into
 364 the comparison of monophosphine and bisphosphine precatalysts for nickel-catalyzed
 365 cross-couplings to provide guidelines for ligand selection when dealing with challenging
 366 substrates and future ligand design tailored to the mechanistic demands of this reactions
 367 [92]. They also reported the analysis of Buchwald-type phosphine structure-reactivity
 368 relationships in Ni-catalyzed cross-coupling reactions using %V_{Bur} threshold analysis
 369 alongside mechanistic organometallic investigations by DFT calculations, with the aim
 370 of serving as a guide to predict new catalytic reactivity of nickel complexes [93].
 371 Notably, the group also made a step forward in the field by using ML to predict the
 372 performance in Pd catalyzed C-N cross-coupling, using data obtained via high-throughput
 373 experimentation (HTE) in collaboration with Merck [94]. As highlighted in Figure 6, they
 374 developed a ML approach to model interaction effects in HTE datasets that were used to
 375 optimize the alcohol deoxyfluorination reaction [95].



376

377 **Figure 6.** Application of ML techniques to analyze HTE datasets is a significant
378 challenge in capturing interaction effects among reaction components. One bottleneck is
379 the presence of irrelevant features in the data. To address this issue, there is a proposal of
380 a two-step statistical modeling approach for HTE datasets. First, classical analysis of
381 variance to identify systematic effects affecting reaction yields throughout the
382 experiment, and second, chemistry-informed features to regress individual effects are
383 used. To validate this methodology a dataset involving alcohol deoxyfluorination was
384 used comprising 740 reactions [95].

385

386 The Schoenebeck group addressed the reductive elimination of ArCF_3 from palladium
387 complexes. They computationally designed a ligand that had a small bite angle in a
388 counterintuitive prediction based on the previous literature. Subsequent experimental
389 work demonstrated that the new ligand provides very high reactivity in the target reaction
390 [96]. Later the group also made a step forward to ML to provide insight into the
391 correlation between ligand and catalyst speciation in palladium complexes [97], and the
392 identification of suitable ligands that form trialkylphosphine-derived dihalogen-bridged
393 Ni(I) dimers with applications in site-selective catalysis [98].

394

395 **Concluding remarks**

396 To sum up, predictive catalysis uses computational simulations and theoretical models to
397 predict the behavior of catalysts and the reactions they catalyze. This approach leverages
398 advances in computational chemistry, materials science, and machine learning to create
399 detailed models of catalytic processes. The simulations can consider various factors that
400 affect catalysis, such as the electronic structure of the catalyst, the adsorption of reactants
401 and intermediates, the kinetics of reaction steps, and the thermodynamics of the reaction
402 system. But is predictive catalysis really at a point where calculations allow us to guide
403 experimental efforts? In our opinion, it is in the right path. It is extremely useful in
404 exploring reactivity and predicting if a reaction is feasible. Selectivity can also be
405 predicted with accuracy. However, predicting yields is still hampered by the difficulties
406 encountered in taking into account all the possible secondary reactivities. The increasing
407 use and importance of computational chemistry in mechanistic studies and its direct
408 correlation with experimental data are driving rapid advancements in the field.
409 Nonetheless, improved methodologies that automatically explore the various reactivities
410 would be needed to achieve an accurate picture of the interconnected reactions that often

411 occur in catalytic processes that are not fully optimized. In summary, the ease of obtaining
412 mechanistic data and parameters and the straightforward interpretation of the data
413 obtained make it a valuable tool for cost-effective experimentation. Therefore, we believe
414 it is a crucial step while we await the full development of machine learning potential.

415 (see Outstanding Questions)

416

417 ACKNOWLEDGEMENTS

418 A.P. and A.P.-Q. are Serra Húnter Fellows. A.P. received ICREA Academia Prize 2019.
419 We thank the Spanish Ministerio de Ciencia e Innovación for projects PID2021-
420 127423NB-I00 (to A.P.) and PID2020-113711GB-I00 (to A.P.-Q.) and the Generalitat de
421 Catalunya for project 2021SGR623. We thank the Spanish Ministerio de Universidades
422 for the predoctoral fellowship FPU20/00707 to R.M.-C.

423

424 REFERENCES

¹ Schaefer III, H.F. (1972) The electronic structure of atoms and molecules. Reading, Massachusetts: Addison-Wesley Publishing Co. p. 146. DOI: 476446781.

² Lipkowitz, K.B. and Boyd, D.B. (1990) "vol 1, preface". Reviews in Computational Chemistry. DOI: 10.1002/9780470125786. ISBN: 9780470125786.

³ Lindström, B. and Pettersson, L.J. (2003) A Brief History of Catalysis. CATTECH 7, 130–138.

⁴ Sanchez-Lengeling, B. (2018) Inverse molecular design using machine learning: Generative models for matter engineering. Science 361, 360–365.

⁵ Gromski, P.S. et al. (2019) How to explore chemical space using algorithms and automation. Nat. Rev. Chem. 3, 119–128.

⁶ Friederich, P. et al. (2020) Machine learning dihydrogen activation in the chemical space surrounding Vaska's complex. Chem. Sci. 11, 4584–4601.

⁷ Angello, N.H. et al. (2022) Closed-loop optimization of general reaction conditions for heteroaryl Suzuki-Miyaura coupling. Science 378, 399–405

⁸ Coley, C. W. et al. (2020) Autonomous Discovery in the Chemical Sciences Part II: Outlook. Angew. Chem. Int. Ed. 59, 23414-23436.

⁹ Tu, Z. et al. (2023) Predictive chemistry: machine learning for reaction deployment, reaction development, and reaction discovery. Chem. Sci. 14, 226–244.

¹⁰ Christensen, K.A. (2007) Chapter 11 - Design of Industrial Catalysts. Comput. Aided Chem. Eng. 23, 311–341.

¹¹ Sigman, M.S. et al. (2016) The Development of Multidimensional Analysis Tools for Asymmetric Catalysis and Beyond. Acc. Chem. Res. 49, 1292–1301.

¹² Cicolella, A. et al. (2022) Combining both Acceptorless Dehydrogenation and Borrowing Hydrogen Mechanisms in one System as described by DFT Calculations. Adv. Theory Simul. 5, 2100566.

¹³ Masdemont, J. et al. (2019) Mechanism of Coupling of Alcohols and Amines to Generate Aldimines and H₂ by a Pincer Manganese Catalyst. ACS Catal. 9, 1662–1669.

¹⁴ Escayola, S. et al. (2020) Mechanism of the Facile Nitrous Oxide fixation by Homogeneous Ruthenium Hydride Pincer Catalysts. Inorg. Chem. 59, 9374–9383.

¹⁵ Genç, A.E. et al. (2023) Transition metal-free synthesis of 2-aryl quinazolines via alcohol dehydrogenation. Mol. Catal. 541, 113119.

¹⁶ Barrales-Martínez, C. et al. (2023) Metal-free catalytic conversion of CO₂ into methanol: local electrophilicity as a tunable property in the design and performance of aniline-derived aminoborane-based FLPS. Inorg. Chem. Front. 10, 2344–2358.

¹⁷ González-Belman, O.F. et al. (2018) The role of the metal in the dual-metal catalysed hydrophenoxylation of diphenylacetylene. Catal. Sci. Technol. 8, 3638–3648.

- ¹⁸ Tomasini, M. et al. (2021) Towards Mild Conditions by Predictive Catalysis via Sterics in the Ru-Catalyzed Hydrogenation of Thioesters. *Mol. Catal.* 510, 111692.
- ¹⁹ Gimferrer, M. et al. (2018) Monitoring of the Phosphine Role in the Mechanism of Palladium-Catalyzed Benzosilole Formation from Aryloxyethynyl Silanes. *Organometallics* 37, 1456–1461.
- ²⁰ Escayola, S. et al. (2021) Chelation enforcing a dual gold configuration in the catalytic hydroxyphenoxylation of alkynes. *Appl. Organomet. Chem.* 35, e6362.
- ²¹ Ahmadi, M. et al. (2022) pH-Responsive Gelation in Metallo-Supramolecular Polymers Based on the Protic Pyridinedicarboxamide Ligand. *Chem. Mater.* 13, 6155–6169.
- ²² Gimferrer, M. et al. (2022) Knölker Iron Catalysts for Hydrogenation Revisited: A Nonspectator Solvent and Fine-Tuning. *Organometallics* 41, 1204–1215.
- ²³ Joly, N. et al. (2023) Enhancement of Knölker Iron Catalysts for Imine Hydrogenation by Predictive Catalysis: From Calculations to Selective Experiments. *Organometallics*, accepted
- ²⁴ Mehdizadeh, M. et al. (2022) Molecular modelling aided catalyst design for PAO oils hydrofinishing. *J. Mol. Liq.* 352, 118675.
- ²⁵ D'Elia, V. et al. (2014) Dynamics of the NbCl₅-catalyzed cycloaddition of propylene oxide and CO₂: assessing the dual role of the nucleophilic Co-catalysts. *Chem. Eur. J.* 20, 11870–11882.
- ²⁶ D'Elia, V. et al. (2015) Cooperative effect by monopodal surface niobium complexes enhancing cyclic carbonate production. *J. Am. Chem. Soc.* 137, 7728–7739.
- ²⁷ Gómez-Suárez, A. et al. (2016) On the Mechanism of the Digold(I)-Hydroxide-Catalysed Hydrophenoxylation of Alkynes. *Chem. Eur. J.* 22, 1125–1132.
- ²⁸ Arayachukiat, S. et al. (2017) Cycloaddition of CO₂ to challenging N-tosyl aziridines using a halogen-free niobium complex: Catalytic activity and mechanistic insights. *Mol. Catal.* 443, 280–285.
- ²⁹ Monreal-Corona, R. et al. (2023) Indolizine Synthesis through Annulation of Pyridinium 1,4-Thiolates and Copper Carbenes: A Predictive Catalysis Approach. *Adv. Synth. Catal.* 365, 760–766.
- ³⁰ Martínez-Núñez, E. (2015) An automated transition state search using classical trajectories initialized at multiple minima. *Phys. Chem. Chem. Phys.* 17, 14912–14921.
- ³¹ Martínez-Núñez, E. (2015) An automated method to find transition states using chemical dynamics simulations. *J. Comput. Chem.* 36, 222–234.
- ³² Martínez-Núñez, E. et al. (2021) AutoMeKin2021: An open-source program for automated reaction discovery. *J. Comput. Chem.* 42, 2036–2048.
- ³³ Maeda, S. et al. (2011) Finding reaction pathways of type A + B → X: Toward systematic prediction of reaction mechanisms. *J. Chem. Theory Comput.* 7, 2335–2345.
- ³⁴ Satoshi, M. et al. (2013) Systematic exploration of the mechanism of chemical reactions: The global reaction route mapping (GRRM) strategy using the ADDF and AFIR methods. *Phys. Chem. Chem. Phys.* 15, 3683–3701.
- ³⁵ Geerlings, P. et al. (2003) Conceptual density functional theory. *Chem. Rev.* 103, 1793–1873.
- ³⁶ Ayers, P.W. et al. (2005) Perturbative perspectives on the chemical reaction prediction problem. *Int. J. Quantum Chem.* 101, 520–534.
- ³⁷ Harvey, J.N. et al. (2019). Scope and Challenge of Computational Methods for Studying Mechanism and Reactivity in Homogeneous Catalysis. *ACS Catal.* 9, 6803–6813.
- ³⁸ Carbó-Dorca, R. and Besalú, E. (1998) Conceptual density functional theory. *J. Mol. Struct. THEOCHEM* 451, 11–23.
- ³⁹ Besalú, E. et al. (2002) Three-dimensional quantitative structure - Activity relationships from tuned molecular quantum similarity measures: Prediction of the corticosteroid-binding globulin binding affinity for a steroid family. *Acc. Chem. Res.* 35, 289–295.
- ⁴⁰ Poater, A. et al. (2010) Computational Methods to Predict the Reactivity of Nanoparticles Through Structure-Property Relationships. *Expert Opin. Drug Deliv.* 35, 295–305.
- ⁴¹ Bian, Y. and Xie, X.-Q. (2019) Generative chemistry: drug discovery with deep learning generative models. *J. Mol. Model.* 27, 71.
- ⁴² Bajorath, J. et al. (2021) Chemoinformatics and artificial intelligence colloquium: progress and challenges in developing bioactive compounds. *J. Cheminformatics* 14, 82.
- ⁴³ Falivene, L. et al. (2019) Towards the online computer-aided design of catalytic pockets. *Nat. Chem.* 11, 872–879.
- ⁴⁴ Falivene, L. et al. (2016) SambVca 2. A Web Tool for Analyzing Catalytic Pockets with Topographic Steric Maps. *Organometallics* 35, 2286–2293.
- ⁴⁵ Pump, E. et al. (2022) Regio, Stereo and Chemoselectivity of 2nd Generation Grubbs Ruthenium-Catalyzed Olefin Metathesis. *Catal. Today* 388-389, 394–402.
- ⁴⁶ Poater, A. et al. (2009) SambVca: A Web Application for the Calculation of Buried Volumes of N-Heterocyclic Carbene Ligands. *Eur. J. Inorg. Chem.* 1759–1766.

- ⁴⁷ Boto, R.A. et al. (2020) NCIPlot4: Fast, Robust, and Quantitative Analysis of Noncovalent Interactions. *J. Chem. Theory Comput.* 16, 4150–4158.
- ⁴⁸ Johnson, E.R. et al. (2010) Revealing Noncovalent Interactions, *J. Am. Chem. Soc.* 132, 6498–6506.
- ⁴⁹ Kozuch, S. and Shaik, S. (2010) Automatic analysis of computed catalytic cycles. *J. Comput. Chem.* 32, 978–985.
- ⁵⁰ Kozuch, S. and Shaik, S. (2010) How to conceptualize catalytic cycles? the energetic span model. *Acc. Chem. Res.* 44, 101–110.
- ⁵¹ Young, T.A. et al. (2020) cgbind: a Python module and web app for automated metallocage construction and host-guest characterization. *J. Chem. Inf. Model.* 60, 3546–3557.
- ⁵² Young, T.A. et al. (2020) audodE: automated calculation of reaction energy profiles – application to organic and organometallic reactions. *Angew. Chem. Int. Ed.* 133, 4312–5320.
- ⁵³ Brethomé, A.V. et al. (2019) Conformational Effects on Physical-Organic Descriptors: The Case of Sterimol Steric Parameters. *ACS Catal.* 9, 2313–2323.
- ⁵⁴ Gensch, T. et al. (2022) A comprehensive discovery platform for organophosphorus ligands for catalysis. *J. Am. Chem. Soc.* 144, 1205–1217.
- ⁵⁵ Monreal-Corona, R. et al. (2022) Photoredox catalysis leading to triazolo-quinoxalinones at room temperature: selectivity of the rate determining step. *Org. Biomol. Chem.* 20, 9330–9336.
- ⁵⁶ Monreal-Corona, R. et al. (2022) A Predictive Chemistry DFT Study of the N₂O Functionalization for the Preparation of Triazolopyridine and Triazoloquinoline Scaffolds. *Org. Chem. Front.* 9, 4347–4357.
- ⁵⁷ Monreal-Corona, R. et al. (2023) Stereoretentive formation of cyclobutanes from pyrrolidines: lessons learned from DFT studies of the reaction mechanism. *J. Org. Chem.* 88, 4619–4626.
- ⁵⁸ Tomasini, M. et al. (2022) A predictive journey towards trans-thioamides/amides. *Chem. Commun.* 58, 9950–9953.
- ⁵⁹ Weinstein, C.M. et al. (2018) Highly Ambiphilic Room Temperature Stable Six-Membered Cyclic (Alkyl)(amino)carbenes. *J. Am. Chem. Soc.* 140, 9255–9260.
- ⁶⁰ Milo, A. Et al. (2014) Interrogating selectivity in catalysis using molecular vibrations. *Nature* 507, 210–214.
- ⁶¹ Mess, E.N. et al. (2014) Designer substrate library for quantitative, predictive modeling of reaction performance. *Proc. Natl. Acad. Sci.* 111, 14698–14703.
- ⁶² Milo, A. et al. (2015) A data-intensive approach to mechanistic elucidation applied to chiral anion Catalysis. *Science* 347, 737–743.
- ⁶³ Ortiz, K.G. et al. (2023) Catalyst-controlled enantioselective and regiodivergent addition of aryl boron nucleophiles to N-alkyl nicotinate salts. *J. Am. Chem. Soc.* 145, 11781–11788.
- ⁶⁴ van Dijk, L. et al. (2023) Data science-enabled palladium-catalyzed enantioselective aryl-carbonylation of sulfonimidamides. *J. Am. Chem. Soc.*, DOI: 10.1021/jacs.3c06674.
- ⁶⁵ St. John, P.C. et al. (2020). Quantum chemical calculations for over 200,000 organic radical species and 40,000 associated closed-shell molecules. *Sci. Data* 7, 244.
- ⁶⁶ Guan, Y. et al. (2021) Real-time prediction of ¹H and ¹³C chemical shifts with DFT accuracy using a 3D graph neural network. *Chem. Sci.* 12, 12012–12026.
- ⁶⁷ Gallegos, L.C. et al. (2021) Importance of Engineered and Learned Molecular Representations in Predicting Organic Reactivity, Selectivity, and Chemical Properties. *Acc. Chem. Res.* 54, 827–836.
- ⁶⁸ Peng, Q. et al. (2016). Computing organic stereoselectivity-from concepts to quantitative calculations and predictions. *Chem. Soc. Rev.* 45, 6093–6107.
- ⁶⁹ Poater, A. et al. (2014) Comparing Ru and Fe-catalyzed olefin metathesis. *Dalton Trans.* 43, 11216–11220.
- ⁷⁰ Solans-Monfort, X. et al. (2016) Toward Olefin Metathesis with Iron Carbene Complexes: Benefits of Tridentate σ -Donating Ligands. *Organometallics* 35, 3914–3923.
- ⁷¹ Bukhryakov, K.V. et al. (2021) Stereospecific Ring-Opening Metathesis Polymerization of Norbornene Catalyzed by Iron Complexes. *Angew. Chem. Int. Ed.* 60, 2934–2938.
- ⁷² Takebayashi, S. et al. (2022) Iron-catalysed ring-opening metathesis polymerization of olefins and mechanistic studies. *Nat. Catal.* 5, 494–502.
- ⁷³ Vougioukalakis, G.C. and Grubbs, R.H. (2010) Ruthenium-based heterocyclic carbene-coordinated olefin metathesis catalysts. *Chem. Rev.* 109, 1746–1787.
- ⁷⁴ Hérisson, P.J.L. and Chauvin, Y. (1971) Catalyse de transformation des oléfines par les complexes du tungstène. II. Télomérisation des oléfines cycliques en présence d'oléfine cycliques. *Die Makromol. Chem.* 141, 161–176.
- ⁷⁵ Arnedo, L. et al. (2017) Olefin metathesis with Ru-based catalysts exchanging the typical N-heterocyclic carbenes by a phosphine-phosphonium ylide. *Catalysts* 7, 85.
- ⁷⁶ Jensen, V. R. et al. (2012) Novel Olefin Metathesis Catalysts. *Int. Patent Appl. WO* 2012032131.

-
- ⁷⁷ Occhipinti, G. et al. (2013) Simple and highly Z-selective ruthenium-based olefin metathesis catalyst. *J. Am. Chem. Soc.* 135, 3331–3334.
- ⁷⁸ Occhipinti, G. et al. (2014) Theory-assisted development of a robust and Z-selective olefin metathesis catalyst *Dalton Trans.* 43, 11106–11117.
- ⁷⁹ Smit, W. et al. (2016) Phosphine-Based Z-Selective Ruthenium Olefin Metathesis Catalysts. *Organometallics* 35, 1825–1837.
- ⁸⁰ Occhipinti, G. et al. (2017) Pyridine-Stabilized Fast-Initiating Ruthenium Monothiolate Catalysts for Z-Selective Olefin Metathesis. *Organometallics* 36, 3284–3292.
- ⁸¹ Reim, I. et al. (2022) Toward E-selective Olefin Metathesis: Computational Design and Experimental Realization of Ruthenium Thio-Indolate Catalysts. *Top. Catal.* 65, 448–461.
- ⁸² Martínez, J.P. et al. (2021) Predictive catalysis in olefin metathesis with Ru-based catalysts with annulated C₆₀ fullerenes in the N-heterocyclic carbenes. *Chem. Eur. J.* 27, 18074–18083.
- ⁸³ Martínez, J.P. et al. (2016) In Silico Olefin Metathesis with Ru-Based Catalysts Containing N-Heterocyclic Carbenes Bearing C₆₀ Fullerenes. *Chem. Eur. J.* 22, 6617–6623.
- ⁸⁴ Poater, A. et al. (2013) Exploring new generations of ruthenium olefin metathesis catalysts: the reactivity of a bis-ylidene ruthenium complex by DFT. *Dalton Trans.* 42, 7271–7275.
- ⁸⁵ Martínez, J.P. et al. (2018) Regioselectivity of the Pauson–Khand reaction in single-walled carbon nanotubes. *Nanoscale* 10, 15078–15089.
- ⁸⁶ Artigas, A. et al. (2017) A Computational study of the intermolecular [2+2+2] cycloaddition of acetylene and C₆₀ catalyzed by Wilkinson’s catalyst. *Chem. Eur. J.* 23, 15067–15072.
- ⁸⁷ Artigas A. et al. (2018) Expedient preparation of open-cage fullerenes by rhodium(I)-catalyzed [2+2+2] cycloaddition of diynes and C₆₀: an experimental and theoretical study. *Chem. Eur. J.* 24, 10653–10661.
- ⁸⁸ Park, Y. et al. (2016) Mechanism of Rh-Catalyzed Oxidative Cyclizations: Closed versus Open Shell Pathways. *Acc. Chem. Res.* 49, 1263–1270.
- ⁸⁹ Luo, M. et al. (2020) Theoretical study on the adsorption and predictive catalysis of MnN₄ embedded in carbon substrate for gas molecules, *Appl. Surf. Sci.* 525, 146480.
- ⁹⁰ Wu, K. and Doyle, A.G. (2017) Parameterization of phosphine ligands demonstrates enhancement of nickel catalysis via remote steric effects. *Nat. Chem.* 9, 779–784.
- ⁹¹ Newman-Stonebraker, S.H. et al. (2021) Univariate classification of phosphine ligation state and reactivity in cross-coupling catalysis. *Science* 374, 301–308.
- ⁹² Borowski, J.E. et al. (2023) Comparison of monophosphine and bisphosphine precatalysts for Ni-catalyzed Suzuki–Miyaura cross-coupling: understanding the role of the ligation state in catalysis. *ACS Catal.* 13, 7966–7977.
- ⁹³ Newman-Stonebraker, S.H. et al. (2022) Structure–reactivity relationships of Buchwald-type phosphines in nickel-catalyzed cross-couplings. *J. Am. Chem. Soc.* 144, 19635–19648.
- ⁹⁴ Ahneman, D.T. et al. (2018) Predicting reaction performance in C–N cross-coupling using machine learning. *Science* 360, 186–190.
- ⁹⁵ Żurański, A.M. et al. (2023) A Machine Learning Approach to Model Interaction Effects: Development and Application to Alcohol Deoxyfluorination. *J. Am. Chem. Soc.* 145, 7898–7909.
- ⁹⁶ Nielsen, M.C. et al. (2013) Computational ligand design for the reductive elimination of ArCF₃ from a small bite angle Pd^{II} Complex: remarkable effect of a perfluoroalkyl phosphine. *Angew. Chem. Int. Ed.* 53, 5903–5906.
- ⁹⁷ Hueffel, J.A. et al. (2021) Accelerated Dinuclear Palladium Catalyst Identification through Unsupervised Machine Learning. *Science* 374, 1134–1140.
- ⁹⁸ Karl, T.M. et al. (2023) Machine Learning-Guided Development of Trialkylphosphine Ni(I) Dimers and Applications in Site-Selective Catalysis. *J. Am. Chem. Soc.* 145, 15414–15424.