

# Quantitative *In Silico* Prediction of the Rate of Protodeboronation by a Mechanistic Density Functional Theory-Aided Algorithm

Daniel S. Wigh, Matthieu Tissot, Patrick Pasau, Jonathan M. Goodman, and Alexei A. Lapkin\*

Cite This: *J. Phys. Chem. A* 2023, 127, 2628–2636

Read Online

ACCESS |



Metrics &amp; More

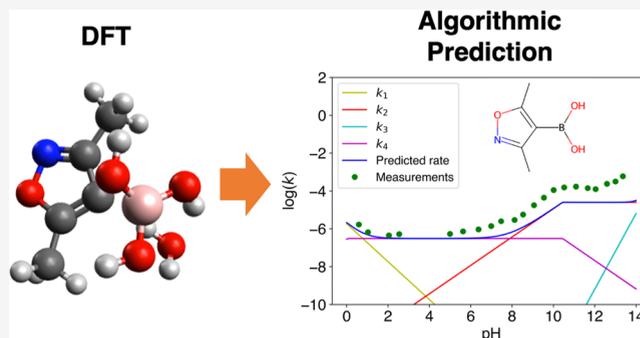


Article Recommendations



Supporting Information

**ABSTRACT:** Computational reaction prediction has become a ubiquitous task in chemistry due to the potential value accurate predictions can bring to chemists. Boronic acids are widely used in industry; however, understanding how to avoid the protodeboronation side reaction remains a challenge. We have developed an algorithm for *in silico* prediction of the rate of protodeboronation of boronic acids. A general mechanistic model devised through kinetic studies of protodeboronation was found in the literature and forms the foundation on which the algorithm presented in this work is built. Protodeboronation proceeds through 7 distinct pathways, though for any particular boronic acid, only a subset of mechanistic pathways are active. The rate of each active mechanistic pathway is linearly correlated with its characteristic energy difference, which in turn can be determined using Density Functional Theory. We validated the algorithm using leave-one-out cross-validation on a data set of 50 boronic acids and made a further 50 rate predictions on academically and industrially important boronic acids out of sample. We believe this work will provide great assistance to chemists performing reactions that feature boronic acids, such as Suzuki–Miyaura and Chan–Evans–Lam couplings.



## INTRODUCTION

Boronic acids (BAs) have long been of interest to the chemical sciences community<sup>1</sup> due to their irreplaceable role in communications,<sup>2</sup> materials,<sup>3–7</sup> and medicine.<sup>8</sup> BAs also feature as the nucleophilic coupling partner in a number of key organic reactions including the Suzuki–Miyaura reaction,<sup>9,10</sup> Chan–Evans–Lam coupling,<sup>11–13</sup> Liebeskind–Srogl coupling,<sup>14</sup> and oxidative Heck<sup>15</sup> and can also undergo addition with carbonyls, imines,<sup>16,17</sup> and enones.<sup>18,19</sup> A reaction will not work if one of the coupling partners degrades at a significantly faster rate than the rate of the intended reaction, and protodeboronation is one of the most significant, if not the most significant, degradation pathways/side reactions in the aforementioned named organic reactions. Being able to accurately predict the rate of protodeboronation would therefore be hugely beneficial to anyone working with BAs, as it can help with reaction planning.

Computational approaches to reaction planning and reaction prediction are becoming increasingly sophisticated, perhaps most notably at the intersection of machine learning and synthetic organic chemistry. However, there remains a rift between approaches in the big data regime, typically being more general but less accurate, and small data regime, with the opposite characteristics, though both approaches rely on effective molecular representation.<sup>20</sup> Machine learning architectures involving deep neural networks can work effectively with large data sets,<sup>21,22</sup> while small data approaches may use a

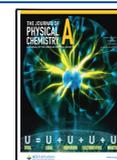
machine learning approach such as multitask learning,<sup>23</sup> an understanding of the kinetics,<sup>24,25</sup> and/or Density Functional Theory (DFT) calculations.<sup>26</sup> However, recent work suggests that more detailed descriptors do not necessarily lead to better predictions.<sup>27</sup> With just 469 data points, the algorithm presented in this work for protodeboronation prediction fits into the small-data regime, and it utilizes a mechanistic understanding of protodeboronation in combination with DFT calculations.

Kinetic studies are typically conducted to achieve total process understanding, which enables rapid identification of optimal reaction conditions for complex chemical systems.<sup>28</sup> The complexity of such systems may be simplified by “lumping” reaction paths (and therefore kinetic constants) together to identify observed reaction rates.<sup>29</sup> Kinetic studies typically involve running experiments at lab scale to determine the order of reaction with respect to each reactant, and estimate the kinetic constants given experimental data collected in a carefully designed campaign. Kinetic parameter

Received: November 24, 2022

Revised: February 27, 2023

Published: March 14, 2023



prediction (*in silico*) is less common than kinetic parameter estimation (given experimental data) since the prediction is inaccurate without a good understanding of the reaction system. In the case of protodeboronation, previously published work<sup>30–32</sup> allows the prediction of what the reaction pathways for a novel boronic acid will look like, which thereby provides context to the kinetic parameter prediction. This is not the case for many other systems, where kinetic data does not exist or where changing the reactants will lead to unpredictable changes in the reaction mechanism. To the best of our knowledge, this is the first purpose-built algorithm to be published in the literature capable of making *in silico* predictions of the rate of protodeboronation.

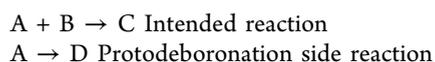
In the protodeboronation reaction the carbon–boron bond of a BA is broken and replaced with a carbon–hydrogen bond, as in Scheme 1. It is an irreversible reaction and will compete

### Scheme 1. General Scheme for Protodeboronation in Aqueous Solution

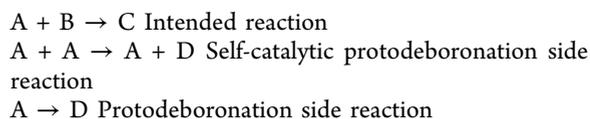


with the intended reaction. Protodeboronation is the most significant side reaction in Suzuki–Miyaura reactions, and since Suzuki–Miyaura reactions are often carried out in aqueous solution, understanding protodeboronation in aqueous solution is of significant interest. There are several different mechanistic pathways for protodeboronation in aqueous solution, most, if not all, of which have been outlined in previous literature, and a general mechanistic model was built to describe these simultaneous mechanistic pathways as a function of pH.<sup>30–32</sup>

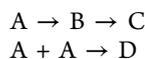
The general scheme for the Suzuki–Miyaura reaction in the presence of excess water is



In the special case where the self-catalytic protodeboronation mechanistic pathway is significant, the general scheme becomes



This is quite similar to the intensely studied Van de Vusse kinetic scheme:<sup>33,34</sup>



Van de Vusse kinetics explores how changes in kinetic constants affect the formation of a competing product and so change the yield of the desired product. This is a simplification of experimental chemistry, as Van de Vusse kinetics only looks at cases where there are competing reactions and low yield is due to the production of a different product rather than because the starting material is unreacted (which may be caused by slowness of the reaction, the degradation of the reagents in the reaction vessel, or other issues). The Van de Vusse kinetic scheme can nonetheless be adapted to the specific kinetic scheme of the Suzuki reaction to explore the intricate relationship between choice of BA, reaction

conditions, protodeboronation rate, and yield. While Van de Vusse kinetics deals with both sequential and parallel reaction pathways, the Suzuki coupling/protodeboronation system exclusively features parallel pathways. There may be one or many mechanistic pathways for protodeboronation happening simultaneously, and in this work this is dealt with by calculating the rate of each mechanistic pathway separately, before then adding up all contributions to arrive at the overall rate which can be observed in the lab. A linear model was set up for each mechanistic pathway, fitting the maximum observed rate of reaction against the results of quantum chemistry calculations carried out using DFT. DFT is a widely used method for calculating energies associated with chemical reactions<sup>35</sup> and generating relevant descriptors,<sup>36</sup> and it is also used beyond computational chemistry.

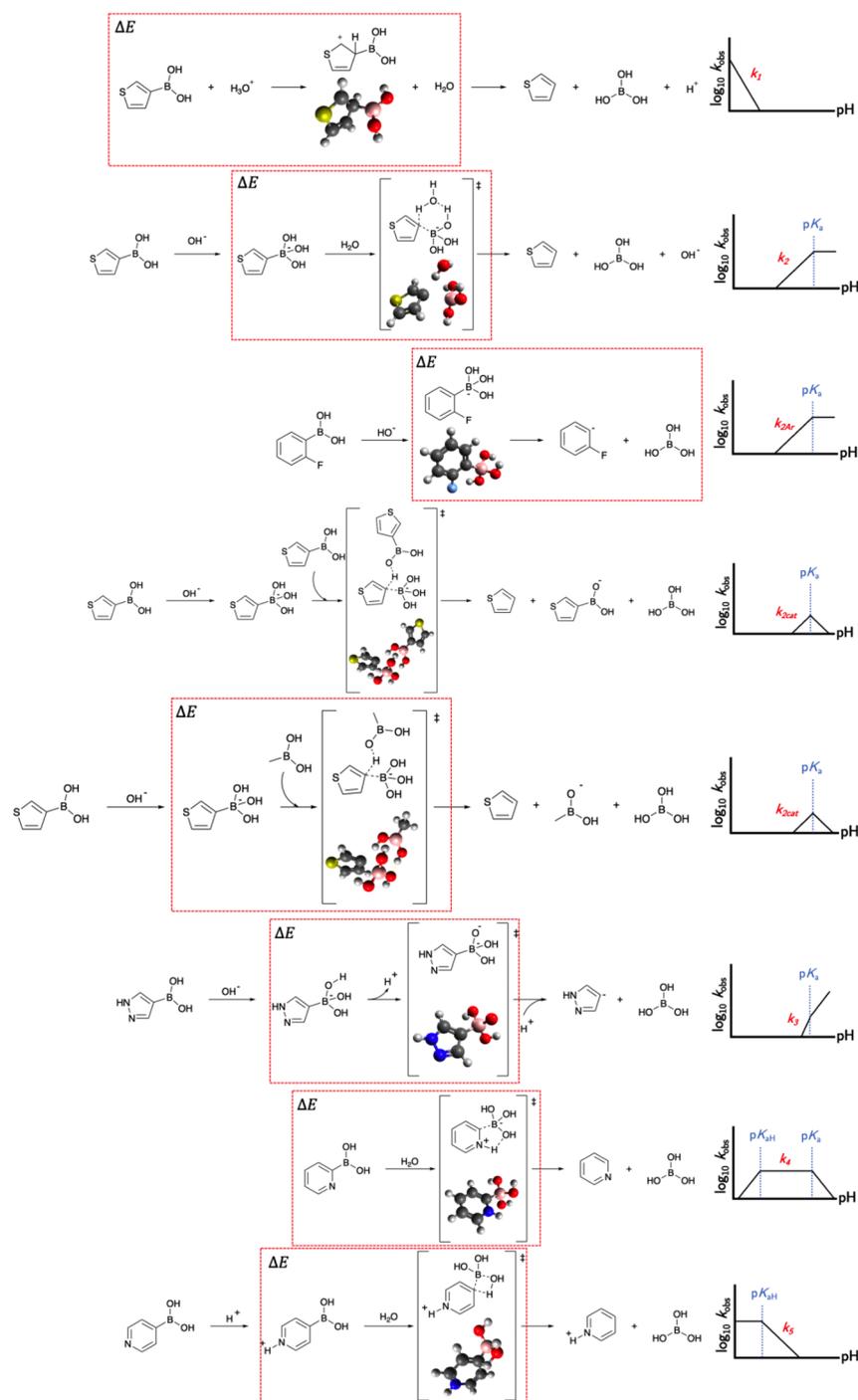
In this work, an algorithm combining the general mechanistic model for protodeboronation with DFT calculations fitted to experimental data was developed to predict the rate of aqueous protodeboronation as a function of pH for unseen BAs without the need for further wet lab experiments.

## METHODOLOGY

An algorithm is a sequence of steps to be performed in pursuit of some goal. In this work the goal is to predict the rate of protodeboronation for novel boronic acids using a small amount of experimental data. The data used is described below, with each algorithmic step then being presented in chronological order. An overview of how the quantum chemistry calculations were performed can be found at the end of the methodology section, and a full description can be found in the [Supporting Information](#).

**Data.** The data set used in this work comes from the Ph.D. thesis of Cox,<sup>30</sup> where he collected 484 data-points across 51 different BAs measuring rate of protodeboronation as a function of pH. An overview of the types of molecules considered in this study is visible in [Figure 2](#), with the full list of molecules available in the [Supporting Information](#). Most of the molecules were 5- or 6-membered ring aromatic BAs. In this work we used data on 50 of the 51 BAs, totalling 469 data points, excluding information only of the tosylate BA due to its bulky size which would complicate DFT calculations. The experiments were carried out in a 50/50 water/dioxane mixture between pH 0 and 14. The rate of reaction observed in the experiments,  $\log_{10}(k_{\text{obs}})$ , ranged from  $-9$  (approximately a 22 year half-life) to  $+2$  (approximately a 7 ms half-life). The half-life (in seconds) of a boronic acid can be calculated given the rate using the following formula:  $t_{1/2} = \ln(2)/k$  (valid for first-order reactions).

**General Mechanistic Model for Protodeboronation.** A general kinetic model for aqueous protodeboronation as a function of pH was developed by Cox et al.<sup>30</sup> and can be seen in [Figure 1](#). The general model arose from a combination of chemical insight and experimental data and was used to estimate the kinetic parameters of each mechanism. Traditional kinetics is an important tool to build an understanding of the underlying chemistry. However, it is not possible to extrapolate a traditional kinetics model to novel substrates without additional wet lab experiments. While Cox et al. were the first to publish a general mechanistic model for protodeboronation which is valid for a range of boronic acids across the whole aqueous pH spectrum, research into protodeboronation has been going on for decades. The mechanisms they propose are backed by previous litera-



**Figure 1.** Example of each of the mechanistic pathways of protodeboronation relevant to the boronic acids considered in this work.<sup>30</sup> A red box encloses the two stages used for calculation of  $\Delta E$  for each mechanism;  $k_{2cat}$  in the fourth row has no red box, since  $\Delta E$  was calculated using a simplified version of this mechanism (shown in the fifth row). A detailed description of each mechanistic pathway can be found in the [Supporting Information](#).

ture,<sup>10,37–44</sup> and coupling this with their experimental validation yields great credibility to the veracity of their proposed mechanistic pathways.

In this work,  $\log(k_{obs})$  refers to the overall protodeboronation rate of reaction, i.e., the observed/measured rate. Conversely,  $\log(k_n)$  (e.g.,  $\log(k_1)$ ) refers to the rate of reaction of a particular protodeboronation mechanism, e.g., mechanism 1, which cannot be measured directly. However, with knowledge of how the rate of each of the mechanisms changes

with pH, the pH at which  $\log(k_n) \approx \log(k_{obs})$  can be deduced. As an example, for a molecule with active mechanisms  $k_2$  and  $k_4$ ,  $\log(k_4) \approx \log(k_{obs})$  at  $\text{pH} = \text{p}K_{aH}$  (since the rate of  $k_2$  is negligible at this pH), and  $\log(k_2) \approx \log(k_{obs})$  at  $\text{pH} = 14$  (since the rate of  $k_4$  is negligible at this pH).

Each mechanism depends on a number of factors which in turn can be used to understand the shape of the  $\text{pH} - \log(k_{obs})$  curve. As an example, the  $k_4$  mechanism only occurs in the zwitterionic state; hence, only BAs with a basic site capable of

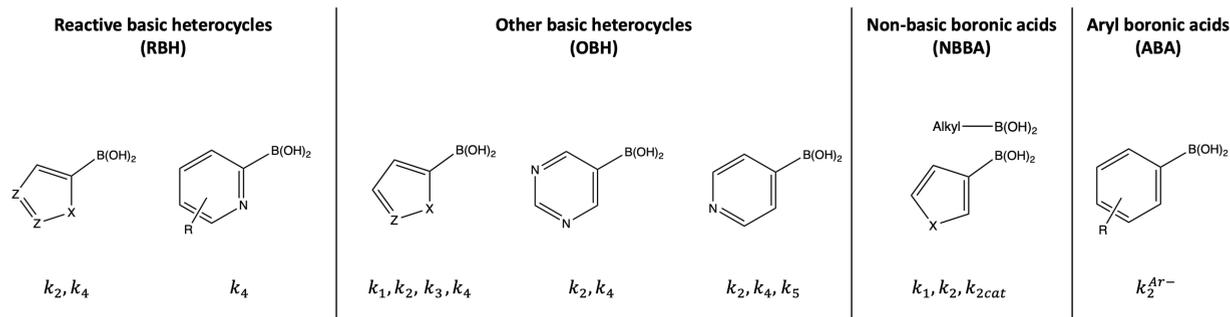


Figure 2. Heuristics for determining which mechanistic pathways are active for a particular boronic acid ( $Z = \text{basic nitrogen}$ ,  $X = \text{S, O}$ ).<sup>30</sup>

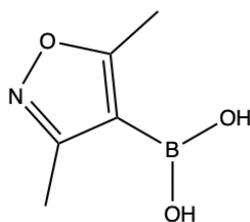
abstracting a proton can decompose through the  $k_4$  mechanism. The  $k_4$  mechanism is at its maximum rate when all of the substrate is in the zwitterionic form, i.e., when pH is between  $pK_a$  and  $pK_{aH}$ . The rate of the  $k_4$  mechanism linearly decreases as the pH moves either below  $pK_{aH}$  or above  $pK_a$ , because the proportion of substrate in zwitterionic form decreases.

**Determining Active Mechanisms.** The general mechanistic model for protodeboronation features 7 distinct mechanistic pathways:  $k_1$ ,  $k_2$ ,  $k_{2Ar}$ ,  $k_{2cat}$ ,  $k_3$ ,  $k_4$ , and  $k_5$ .  $k_{2Ar}$  refers to the mechanistic pathway for aryl boronic acids.  $k_{2cat}$  refers to the mechanistic pathway with a transition state featuring two reactant BAs coupled together during protodeboronation, so we believe autoprotodeboronation may be a more accurate labeling of this pathway than “autocatalysis” as Cox used, since the latter typically refers to catalysis by a reaction product, but we have nevertheless decided to keep Cox’s  $k_{2cat}$  labeling.

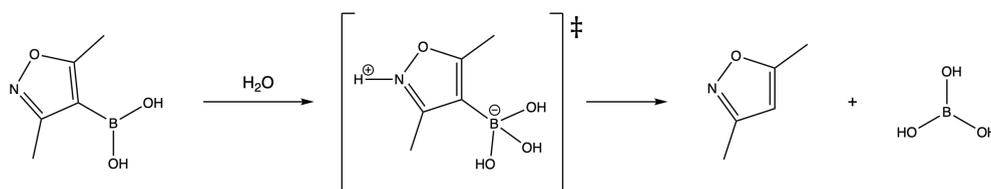
As previously described, not all 7 mechanistic pathways will be significant for any one BA, and a heuristic initially developed by Cox was used to determine which pathways are significant for each BA. These mechanism heuristics can be seen in Figure 2.

**Algorithmic Protodeboronation Prediction.** The journey of (3,5-dimethyl-1,2-oxazol-4-yl)boronic acid (DMOBA), shown in Scheme 2, through the algorithm will act as a case study to showcase how the model works.

Scheme 2. Structure of (3,5-Dimethyl-1,2-oxazol-4-yl)boronic Acid (DMOBA)



Scheme 3. Mechanistic Pathway  $k_4$  for (3,5-Dimethyl-1,2-oxazol-4-yl)boronic Acid



**1. Choice of Boronic Acid.** The chosen BA for which one would like to know the rate of protodeboronation in aqueous solution should fall within the scope of the model; i.e., the BA must be of type  $R-B(OH)_2$ , where the boron is directly bonded to a carbon atom. Predictive performance is better when the queried BA is similar to the BAs present in the data set, so the BA should ideally be small (fewer than 15 heavy atoms) and have the boron bonded to an aromatic carbon.

**2. Determine Active Mechanistic Pathways.** Using the heuristics presented in Figure 2, the significant mechanisms for the queried molecule can be determined; from this it is apparent that DMOBA is of type “Other Basic Heterocycle” (OBH) with active mechanisms  $k_1$ ,  $k_2$ ,  $k_3$ , and  $k_4$ . A more detailed written description of the heuristics used to determine which category of boronic acid a molecule belongs to can be found in the Supporting Information.

**3. Calculate  $\Delta E$  for Each Mechanistic Pathway.** An overview of the protodeboronation mechanisms can be seen in Figure 1, with the red box enclosing the structures used for the calculation of the characteristic energy difference ( $\Delta E$ ). For  $k_2$ ,  $k_{2cat}$ ,  $k_3$ ,  $k_4$ , and  $k_5$ , the characteristic energy difference is calculated as the difference in energy between the transition state and the reactant or preceding reaction intermediate, as per the Arrhenius equation. For  $k_1$  it is the difference in energy between the reaction intermediate and the reactant, while for  $k_{2Ar}$  it is between two reaction intermediates. As an example, the molecular structures used for the  $\Delta E$  calculations for the  $k_4$  mechanistic pathway for DMOBA can be seen in Scheme 3.

Naturally, the DFT energy calculations of the transition state was a transition state optimization, while calculations for reactants and reaction intermediates were a local energy minimum optimization. Further details for each mechanism can be found in the Quantum Chemistry Calculations section and in the Supporting Information.

**4. Linear Regression of  $\Delta E$  vs  $\log(k_n)$ .** Each of the 7 mechanisms has a particular pH (or pH range), where the rate of reaction is at its maximum. For DMOBA the  $k_1$ ,  $k_2$ ,  $k_3$ , and  $k_4$  mechanisms are active, and for these four mechanisms the maximum rate of reaction can be found at  $\text{pH} = 0$ ,  $\text{pH} > pK_a$ ,

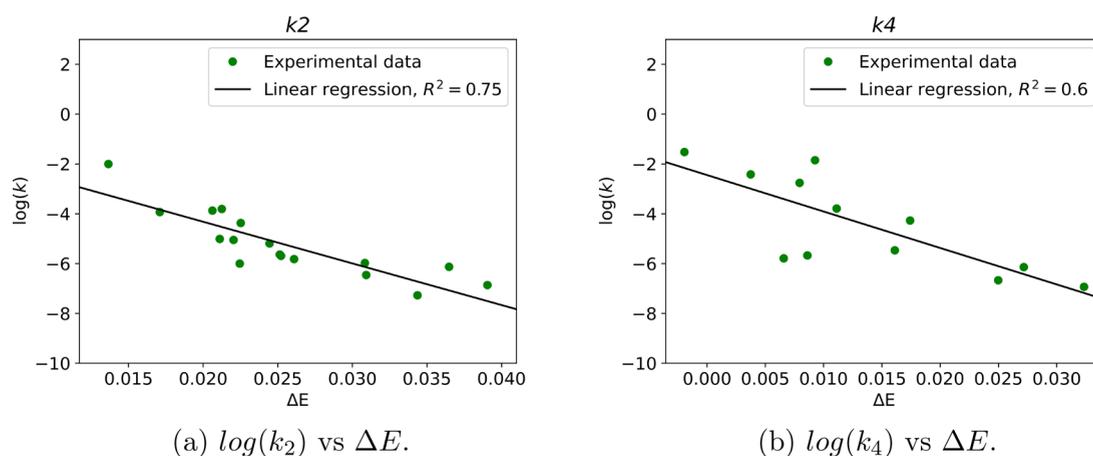


Figure 3. Linear regression of  $\log(k)$  vs  $\Delta E$  for the  $k_2$  and  $k_4$  mechanistic pathways.

pH = 14, and  $\text{p}K_{\text{aH}} < \text{pH} < \text{p}K_{\text{a}}$ , respectively (see also Figure 1).  $\text{p}K_{\text{a}}$  refers to the pH at which the acidic site on a molecule is 50% protonated, while  $\text{p}K_{\text{aH}}$  refers to the pH at which the basic site on a molecule is 50% protonated. While in general  $\text{p}K_{\text{a}} + \text{p}K_{\text{b}} = 14$  in water, the same is not generally true for  $\text{p}K_{\text{aH}}$ , i.e.,  $\text{p}K_{\text{a}} + \text{p}K_{\text{aH}} \neq 14$ .

The observed rate of reaction at  $\text{pH} = \text{p}K_{\text{a}}$  will be the result of constructive interference between  $k_2$  and  $k_4$ . However, we can deduce the rate of each mechanism by observing the reaction at the right pH:  $\log(k_{\text{obs}}) = \log(k_2)$  at  $\text{pH} = 14$  and  $\log(k_{\text{obs}}) = \log(k_4)$  at  $\text{pH} = \text{p}K_{\text{aH}}$ . Similar logic can be applied to each of the other mechanisms.

This analysis is repeated for all mechanisms of all other BAs, and we find that there is a linear relationship between the maximum observed rate of a mechanism and the associated  $\Delta E$  for each BA. This was already known to be true for the  $\log(k_{2\text{Ar}})$  mechanism,<sup>31</sup> and in this work we show that it is true for the other mechanisms as well. Linear regression for the  $k_2$  and  $k_4$  mechanisms can be seen in Figure 3. Using these linear relationships, the maximum rate of a mechanism for a novel BA can be predicted given the associated  $\Delta E$ , which can be calculated using DFT. Leave-one-out cross-validation (LOOCV) was used to assess the quality of the fit for each molecule and for the construction of the parity plot for Cox's molecules, while all data points were used for the prediction of protodeboronation rate for the novel molecules.

5. *Building a System of Linear Equations Using  $\text{p}K_{\text{a}}$ ,  $\text{p}K_{\text{aH}}$ , and Predicted  $\log(k_n)$ .* The rate curve as a function of pH can now be fully specified. For DMOBA the  $k_1$ ,  $k_2$ ,  $k_3$ , and  $k_4$  mechanisms are active. The rate curve for  $k_2$  is a linear increase in rate with increasing pH until  $\text{pH} = \text{p}K_{\text{a}}$ , after which the slope changes to 0. The rate function for  $k_2$  consists of 2 intersecting linear equations and can therefore be fully specified with 4 pieces of information:  $\text{p}K_{\text{a}}$  (measured experimentally), the maximum rate (calculated in the previous step), the slope when  $\text{pH} > \text{p}K_{\text{a}}$  (known to be 0), and the slope when  $\text{pH} < \text{p}K_{\text{a}}$  (determined by inspection of prior data). One can follow a similar argument to construct the rate curve for the other mechanisms, though of course the  $k_4$  mechanism consists of three intersecting linear equations rather than just two.

The slope of the linear increase/decrease in rate for each mechanism was determined by inspection of the data set and is reported below in units of  $\log(k_n)/\text{pH}$ :

1.  $k_1$ : -1
2.  $k_2$ : 0.75

3.  $k_{2\text{Ar}}$ : 0.75
4.  $k_{2\text{cat}}$ : 2, -2
5.  $k_3$ : 2
6.  $k_4$ : 0.75, -0.75
7.  $k_5$ : -0.75

With a fully specified system, we can now construct a predicted rate curve for each mechanism, as exemplified in Figure 4.

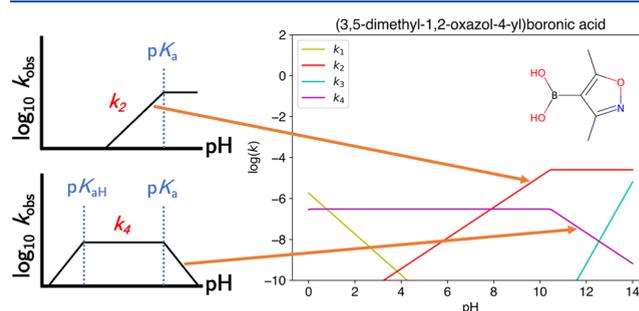


Figure 4. Constructing rate curves for each mechanism given the  $\text{p}K_{\text{a}}$ ,  $\text{p}K_{\text{aH}}$ , and slope of mechanistic deactivation with pH.

6. *Sum All Rate Curves.* The rate of any one mechanism cannot be measured directly, and the rate which is observed in an experiment is the combined rate of all active mechanisms. Thus, the final step in predicting the observed rate of protodeboronation for a BA is to sum over the rate of all active mechanisms. See also Figure 5 for an example of this. The experimental rate measurements span from +2 (half-life of roughly 7 ms) to -9 (half-life of roughly 22 years), and the predicted overall rate has been capped to fall within this region as well.

*Quantum Chemistry Calculations.* The transition state and ground state energies for all substrates across all mechanisms except mechanism  $k_{2\text{cat}}$  were calculated using DFT in Gaussian using M06L/6-311++G\*\* in water (using the keywords "SCRF = (Solvent = Water)"). The larger size of the  $k_{2\text{cat}}$  transition state presented convergence issues when optimizing with M06L/6-311++G\*\*, so instead an optimized geometry was found using B3LYP/6-31G(d), and this was followed by a single point energy calculation with M06L/6-311++G\*\*. M06L/6-311++G\*\* was the preferred computational approach due to its relatively high accuracy; meanwhile, the B3LYP/6-31G(d) computational approach is faster but less

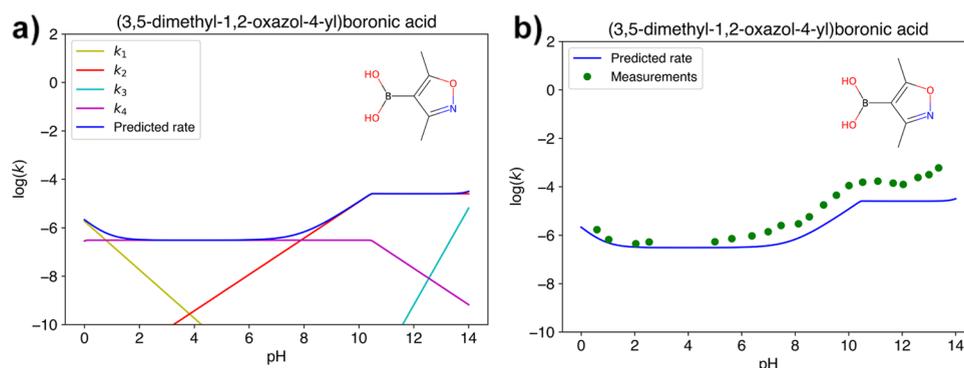


Figure 5. Predicted rate vs measured rate for DMOBA.

accurate (e.g., because it underbinds if no dispersion corrections are used). Local energy minimum calculations were initiated with the keyword “Opt”, while transition state calculations were initiated with the keywords “Opt=(TS, CalcFC, noeigentest)”. Full details how the DFT calculations were performed for each mechanism, including example input files and example mechanistic schemes, can be found in the [Supporting Information](#).

The  $k_{2cat}$  mechanism transition state, perhaps more accurately regarded as self-protodeboronation as opposed to autocatalytic, is a remarkably large structure featuring many atoms for large boronic acids, since the transition state for this mechanism involves the interaction between two boronic acid structures. Therefore, the  $k_{2cat}$  mechanism transition state needed to be simplified further, i.e., beyond just using a cheaper computational approach. The number of atoms involved in the transition state was reduced by replacing the boronic acid molecule which was not directly involved in bond breakage with a methyl group, as seen in [Figure 1](#).

## RESULTS AND DISCUSSION

Using the algorithm presented in this work enables the prediction of protodeboronation as a function of pH of BAs while only requiring knowledge of  $pK_a$ ,  $pK_{aH}$ , and energy difference calculations using DFT. It thus enables *in silico* prediction of the protodeboronation rate as a function of pH. There is a linear relationship between the rate of a particular mechanism,  $k_n$ , and the characteristic energy difference associated with that mechanism, and plots of these linear relationships can be found in the [Supporting Information](#).

Given the limited amount of available data, the performance of the algorithm was assessed using leave-one-out cross-validation (LOOCV). In LOOCV each molecule is iteratively held out to form the validation set, while the remaining molecules act as the training set. The algorithm was used to generate predictions for all molecules in a manner consistent with LOOCV, and the rate predictions were then plotted against rate measurements in a parity plot, as in [Figure 6](#).

We envision the ability to make *in silico* predictions of protodeboronation particularly useful to chemists running reactions featuring boronic acids, and we have used the algorithm to generate protodeboronation predictions for 50 novel boronic acids. Figures showing these predictions can be found in the [Supporting Information](#) and can be used “out of the box”. The associated code can be found on GitHub, should you want to generate predictions of your own. We also hope that this work may serve as inspiration as a method for predicting the rate of reaction when the mechanistic details

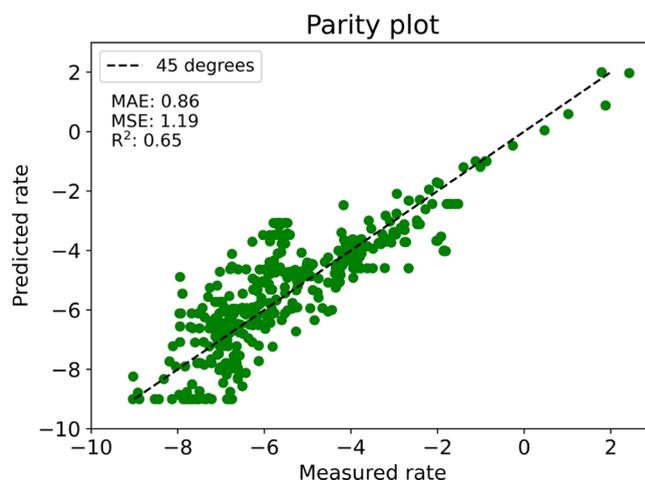


Figure 6. Parity plot showing the predictive accuracy of the algorithm. MAE: Mean Absolute Error. MSE: Mean Squared error.  $R^2$ : Coefficient of Determination.

about a reaction are known and can be simplified as a set of interacting linear equations.

**Predictions for Novel Boronic Acids.** We have generated predictions of protodeboronation rate vs pH for 50 novel BAs, the graphs of which can be found in the [Supporting Information](#). Of these 50, 38 BAs were selected from the top 100 most used boronic acids in the Reaxys database (after excluding boronic acids for which data already exists), while the remaining 12 were selected out of personal interest.

Thus, when a chemist is planning a reaction which involves a boronic acid, they will be able to look up the predicted rate of protodeboronation, and this may assist them in their reaction planning. Given how strongly the rate of protodeboronation depends on pH, it is important to know the pH of a reaction to understand how significant protodeboronation will be. Suzuki reactions are typically run under basic conditions, and prior work has shown that the pH can decrease by upward of 3 pH units over the course of the reaction.<sup>45</sup> This makes it difficult to predict the extent of protodeboronation during a Suzuki reaction, even when the protodeboronation rate as a function of pH is known exactly.

Applying this algorithm to understand literature data may be difficult for reasons beyond the pH varying over the course of a reaction. The vast majority of Suzuki reactions do not report any pH value. Our analysis revealed over 200 000 reactions on Reaxys labeled as a “Suzuki Coupling”; however, only 399 of these reported a pH value, and of these 399 reactions, many

reported a low pH value (below 7). Suzuki reactions are typically run in the presence of a base, and the reason for the majority of Suzuki reactions with a reported pH value being in the acidic region may be because the pH is only measured and reported when deviating from the norm.

**Model Explainability and Extrapolation.** The protodeboronation prediction algorithm relies on an accurate understanding of the active mechanisms, precise DFT calculations, and (measured)  $pK_a$  and  $pK_{aH}$  values (if applicable). Each of these aspects of the model contributes its own domain of applicability to the model, which can make the overall domain of applicability difficult to understand; however, as a guiding principle, the closer a novel molecule is to a molecule which already appears in the training data set, the more trustworthy the prediction. The rate predictions were capped to always fall between  $-9$  and  $+2$  ( $\approx 22$  years and 7 ms), since this was the range of rate values reported in the data, and also because this would cover a sufficiently large space for any practical considerations.

The highest accuracy predictions can be made for molecules which are active with the  $k_{2Ar}$  mechanism, since only one mechanism is active, it has the most data, as well as a near perfect correlation ( $R^2 = 0.97$ ) in the linear regression of  $\Delta E$  against  $\log(k)$  (see Figure 1c in the Supporting Information). The particular case of halogens attached to an aromatic ring containing a boronic acid was studied in detail by Cox,<sup>30</sup> where he observed that the more halogens attached to the ring, and the closer these halogens are to the boronic acid, the faster the rate of protodeboronation. The trend of increasing rate of protodeboronation as a halogen moves closer to the boronic acid can be observed by considering rate measurements and predictions for molecules 74–76 (see Section 4 of the Supporting Information). The fastest rate is observed with molecule 72, which has fluoride occupying all five positions on the ring (with the sixth position being the boronic acid). It is therefore reassuring to see the algorithm predicting high rates for molecules 102/112/148 and 103/111/149, as these are also boronic acids with halogens on the ring, with the fluorides having been replaced with chloride and bromide, respectively.

It is noteworthy that all the molecules dealt with in this work are boronic acids. It is unlikely that the algorithm would extend to molecules containing Bpins or other cousins of the boronic acids, as the mechanistic pathways are likely to be different. Each of the linear regression models of  $\Delta E$  against  $\log(k)$  (see Figure 1 in the Supporting Information) also provides guidance of the domain of applicability of the model. The  $k_3$  and  $k_5$  mechanisms have only two data points each, which is not enough to produce a convincing relationship. However, the reason for this lack of data is that molecules active with these mechanisms are quite uncommon, which in turn decreases the need for such predictions to be made. Furthermore, the range of energy difference ( $\Delta E$ ) observed for each mechanism may also indicate the domain of applicability for each mechanism. As an example, the range of  $\Delta E$  values observed for the  $k_{2cat}$  mechanism ranges from  $-0.0137$  to  $+0.0065$  hartree. The energy difference calculated for the  $k_{2cat}$  mechanism for molecule 134 is  $-0.0658$  hartree, which is quite far outside the range of values observed for the training molecules, and results in the rate prediction to reach the cap of  $\log(k) = +2$  ( $\approx 7$  ms) at  $pH = pK_a$ . The rate prediction for molecule 134 certainly looks interesting, so should the aforementioned reasons be grounds to dismiss the prediction or should it be trusted since the prediction conforms to chemical expect-

ations? (In other words, we know that electron density pulled away from the bond attached to the boronic acid will speed up protodeboronation; this is most notable for the fluorinated boronic acids, and in the case of molecule 134, the phenyl ring pulls electron density away from the conjugated double bond that the boronic acid is attached to.) Only lab experimentation would be able to settle this debate. Regardless, this molecule, as well as molecules 132 and 143–147, were included to explore how the algorithm behaves near the edge of its domain of applicability.

Overall, the algorithm is capable of making remarkably good predictions on the rate of protodeboronation, given the complexity of the many simultaneous mechanistic pathways. The coefficient of determination ( $R^2$ ) being 0.65 indicates that the algorithm is capable of explaining a majority of the variability in the data set, and with a mean absolute error of 0.86, one can expect predictions from the algorithm to typically be within 1  $\log(k)$  of the true rate. This can be valuable information for a chemist to help them understand whether the rate of protodeboronation for their boronic acid of interest under certain pH conditions will be on the order of magnitude of seconds, days, or years.

**Future Work.** Protodeboronation is a side reaction, and a significant challenge remains in understanding the rate of the intended reaction, be that the Suzuki reaction or any other. Ideally, the rate of the intended reaction should be significantly faster than the rate of the side reaction. This work allows for *in silico* prediction of the rate of protodeboronation. However, we do not know of any work which allows for *in silico* predictions of the rate of intended reactions.

The protodeboronation prediction algorithm relies on knowing  $pK_a$  and  $pK_{aH}$  of the queried BA. For the 50 BAs included in the initial model building, these values had been measured in a lab. However, finding experimentally measured  $pK_a$  and  $pK_{aH}$  values either in the literature or in a database such as DataWarrior<sup>46</sup> may not be possible for novel BAs. Using machine learning to predict  $pK_a$  and  $pK_{aH}$  is a potential solution for extrapolating from existing data. However, the uncertainty associated with machine learning models in chemistry can be significant, particularly when the domain of applicability is unclear (i.e., it is not known which molecules fall within the scope and beyond the scope of the model).  $pK_a$  and  $pK_{aH}$  varies greatly between solvents, and we were unable to find a ML model built to predict these values in a 50/50 water/dioxane mixture. Therefore, the  $pK_a$  and  $pK_{aH}$  values used for the 50 novel BAs were simply the mean values of Cox's molecules. Given the notable overlap in structure between the two sets of boronic acids, simply using the mean will likely still yield highly informative predictions.

This work sprung out from a general mechanistic model for protodeboronation, and we believe if general mechanistic models are developed for other reactions, it may become easier to create workflows similar to this for generating *in silico* rate predictions. We look forward to applying this methodology to other systems.

## CONCLUSIONS

Protodeboronation is a significant and unintended side reaction of numerous notable reactions, including the Suzuki–Miyaura reaction, and any work which can aid in our understanding of how to avoid it may provide much value to the chemistry community. In this work we present a 6 step algorithm for *in silico* prediction of the rate of protodeboro-

nation of boronic acids. The algorithm relies on mechanistic insight from previously published literature. Protodeboronation can occur along 7 distinct mechanistic pathways, each having its own characteristic energy difference. There is linear correlation between this energy difference and the reaction rate of the associated mechanistic pathway. Adding the rate of all mechanistic pathways finally yields the observed rate of reaction. Leave-one-out cross-validation revealed that this algorithmic approach provides predicted protodeboronation rates with high accuracy. The algorithm requires two inputs from the user: acidity information ( $pK_a$  and possibly  $pK_{aH}$ ) and energy difference calculations using Density Functional Theory (DFT) for each active mechanism. Data on  $pK_a$  and  $pK_{aH}$  are not available for all boronic acids, and predicting these values may introduce additional uncertainty into the algorithm. The need to perform quite a few DFT calculations for each novel boronic acid is a significant bottleneck for the widespread use of this algorithm, so to make this work more accessible, we have made protodeboronation predictions for 50 academically and industrially important boronic acids, which can be found in the [Supporting Information](#).

## ■ ASSOCIATED CONTENT

### Data Availability Statement

All the code and data associated with this project is freely available on <https://github.com/sustainable-processes/protodeboronation-prediction>, allowing anyone to generate a protodeboronation rate predictions for BAs that fall within the scope of the model, given appropriate DFT calculations and approximate values for  $pK_a$  and  $pK_{aH}$  (if applicable).

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpca.2c08250>.

A detailed explanation of each protodeboronation mechanism, protodeboronation predictions for 50 novel boronic acids, and optimized structures and energies for each relevant mechanism for all 100 molecules ([PDF](#))

## ■ AUTHOR INFORMATION

### Corresponding Author

Alexei A. Lapkin – Department of Chemical Engineering and Biotechnology, University of Cambridge, CB3 0AS Cambridge, U.K.; [orcid.org/0000-0001-7621-0889](https://orcid.org/0000-0001-7621-0889); Email: [aal35@cam.ac.uk](mailto:aal35@cam.ac.uk)

### Authors

Daniel S. Wigh – Department of Chemical Engineering and Biotechnology, University of Cambridge, CB3 0AS Cambridge, U.K.; [orcid.org/0000-0002-0494-643X](https://orcid.org/0000-0002-0494-643X)

Matthieu Tissot – UCB Biopharma SPRL, 1420 Braine l'Alleud, Belgium; [orcid.org/0000-0003-1164-6598](https://orcid.org/0000-0003-1164-6598)

Patrick Pasau – UCB Biopharma SPRL, 1420 Braine l'Alleud, Belgium; [orcid.org/0000-0002-1777-7911](https://orcid.org/0000-0002-1777-7911)

Jonathan M. Goodman – Yusuf Hamied Department of Chemistry, University of Cambridge, CB2 1EW Cambridge, U.K.; [orcid.org/0000-0002-8693-9136](https://orcid.org/0000-0002-8693-9136)

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jpca.2c08250>

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work is cofunded by UCB Pharma and Engineering and Physical Sciences Research Council via project EP/S024220/1 EPSRC Centre for Doctoral Training in Automated Chemical Synthesis Enabled by Digital Molecular Technologies.

## ■ REFERENCES

- (1) Hall, D. G. *Boronic Acids: Preparation and Applications in Organic Synthesis, Medicine and Materials*, 2nd ed.; Wiley-VCH: Weinheim, Germany, 2011.
- (2) Bull, S. D.; Davidson, M. G.; van den Elsen, J. M. H.; Fossey, J. S.; Jenkins, A. T. A.; Jiang, Y.-B.; Kubo, Y.; Marken, F.; Sakurai, K.; Zhao, J.; James, T. D. Exploiting the reversible covalent bonding of boronic acids: recognition, sensing, and assembly. *Acc. Chem. Res.* **2013**, *46*, 312–326.
- (3) Hirai, M.; Tanaka, N.; Sakai, M.; Yamaguchi, S. Structurally Constrained Boron-, Nitrogen-, Silicon-, and Phosphorus-Centered Polycyclic -Conjugated Systems. *Chem. Rev.* **2019**, *119*, 8291–8331.
- (4) von Grothuss, E.; John, A.; Kaese, T.; Wagner, M. Doping Polycyclic Aromatics with Boron for Superior Performance in Materials Science and Catalysis. *Asian Journal of Organic Chemistry* **2018**, *7*, 37–53.
- (5) Jäkle, F. Lewis acidic organoboron polymers. *Coord. Chem. Rev.* **2006**, *250*, 1107–1121.
- (6) Hudson, Z. M.; Wang, S. Impact of Donor-Acceptor Geometry and Metal Chelation on Photophysical Properties and Applications of Triarylboranes. *Acc. Chem. Res.* **2009**, *42*, 1584–1596.
- (7) Rao, Y.-L.; Amarne, H.; Wang, S. Photochromic four-coordinate N,C-chelate boron compounds. *Coord. Chem. Rev.* **2012**, *256*, 759–770.
- (8) Brooks, W. L. A.; Sumerlin, B. S. Synthesis and Applications of Boronic Acid-Containing Polymers: From Materials to Medicine. *Chem. Rev.* **2016**, *116*, 1375–1397.
- (9) Miyaura, N.; Yamada, K.; Suzuki, A. A new stereospecific cross-coupling by the palladium-catalyzed reaction of 1-alkenylboranes with 1-alkenyl or 1-alkynyl halides. *Tetrahedron Lett.* **1979**, *20*, 3437–3440.
- (10) Miyaura, N.; Suzuki, A. Palladium-Catalyzed Cross-Coupling Reactions of Organoboron Compounds. *Chem. Rev.* **1995**, *95*, 2457–2483.
- (11) Chan, D. M. T.; Monaco, K. L.; Wang, R.-P.; Winters, M. P. New N- and O-arylations with phenylboronic acids and cupric acetate. *Tetrahedron Lett.* **1998**, *39*, 2933–2936.
- (12) Evans, D. A.; Katz, J. L.; West, T. R. Synthesis of diaryl ethers through the copper-promoted arylation of phenols with arylboronic acids. An expedient synthesis of thyroxine. *Tetrahedron Lett.* **1998**, *39*, 2937–2940.
- (13) Lam, P. Y. S.; Clark, C. G.; Saubern, S.; Adams, J.; Winters, M. P.; Chan, D. M. T.; Combs, A. New aryl/heteroaryl C-N bond cross-coupling reactions via arylboronic acid/cupric acetate arylation. *Tetrahedron Lett.* **1998**, *39*, 2941–2944.
- (14) Liebeskind, L. S.; Srogl, J. Thiol Ester-Boronic Acid Coupling. A Mechanistically Unprecedented and General Ketone Synthesis. *J. Am. Chem. Soc.* **2000**, *122*, 11260–11261.
- (15) Cho, C. S.; Uemura, S. Palladium-catalyzed cross-coupling of aryl and alkenyl boronic acids with alkenes via oxidative addition of a carbon-boron bond to palladium(O). *J. Organomet. Chem.* **1994**, *465*, 85–92.
- (16) Sakai, M.; Ueda, M.; Miyaura, N. Rhodium-Catalyzed Addition of Organoboronic Acids to Aldehydes. *Angew. Chem., Int. Ed.* **1998**, *37*, 3279–3281.
- (17) Takezawa, A.; Yamaguchi, K.; Ohmura, T.; Yamamoto, Y.; Miyaura, N. Inter- and Intramolecular Additions of 1-Alkenylboronic Acids or Esters to Aldehydes and Ketones Catalyzed by Rhodium(I) Complexes in Basic, Aqueous Solutions. *Synlett* **2002**, *2002*, 1733–1735.
- (18) Sakai, M.; Hayashi, H.; Miyaura, N. Rhodium-Catalyzed Conjugate Addition of Aryl- or 1-Alkenylboronic Acids to Enones. *Organometallics* **1997**, *16*, 4229–4231.

- (19) Cho, C. S.; Motofusa, S.-i.; Ohe, K.; Uemura, S.; Shim, S. C. A New Catalytic Activity of Antimony(III) Chloride in Palladium(0)-Catalyzed Conjugate Addition of Aromatics to  $\alpha,\beta$ -Unsaturated Ketones and Aldehydes with Sodium Tetraphenylborate and Arylboronic Acids. *Journal of Organic Chemistry* **1995**, *60*, 883–888.
- (20) Wigh, D. S.; Goodman, J. M.; Lapkin, A. A. A review of molecular representation in the age of machine learning. *WIREs Computational Molecular Science* **2022**, *12*, No. e1603.
- (21) Gao, H.; Struble, T. J.; Coley, C. W.; Wang, Y.; Green, W. H.; Jensen, K. F. Using Machine Learning To Predict Suitable Conditions for Organic Reactions. *ACS Central Science* **2018**, *4*, 1465–1476.
- (22) Coley, C. W.; Green, W. H.; Jensen, K. F. Machine Learning in Computer-Aided Synthesis Planning. *Acc. Chem. Res.* **2018**, *51*, 1281–1289.
- (23) Felton, K.; Wigh, D.; Lapkin, A. Multi-task Bayesian Optimization of Chemical Reactions. *34th Conference on Neural Information Processing Systems; NeurIPS*, 2020.
- (24) Taylor, C. J.; Booth, M.; Manson, J. A.; Willis, M. J.; Clemens, G.; Taylor, B. A.; Chamberlain, T. W.; Bourne, R. A. Rapid, automated determination of reaction models and kinetic parameters. *Chemical Engineering Journal* **2021**, *413*, 127017.
- (25) Taylor, C. J.; Seki, H.; Dannheim, F. M.; Willis, M. J.; Clemens, G.; Taylor, B. A.; Chamberlain, T. W.; Bourne, R. A. An automated computational approach to kinetic model discrimination and parameter estimation. *React. Chem. Eng.* **2021**, *6*, 1404–1411.
- (26) Lawal, M. M.; Govender, T.; Maguire, G. E. M.; Honarparvar, B.; Kruger, H. G. Mechanistic investigation of the uncatalyzed esterification reaction of acetic acid and acid halides with methanol: a DFT study. *J. Mol. Model.* **2016**, *22*, 235.
- (27) Pomberger, A.; Pedrina McCarthy, A. A.; Khan, A.; Sung, S.; Taylor, C. J.; Gaunt, M. J.; Colwell, L.; Walz, D.; Lapkin, A. A. The effect of chemical representation on active machine learning towards closed-loop optimization. *React. Chem. Eng.* **2022**, *7*, 1368–1379.
- (28) Taylor, C. J.; Manson, J. A.; Clemens, G.; Taylor, B. A.; Chamberlain, T. W.; Bourne, R. A. Modern advancements in continuous-flow aided kinetic analysis. *Reaction Chemistry & Engineering* **2022**, *7*, 1037–1046.
- (29) Ancheyta-Juárez, J.; López-Isunza, F.; Aguilar-Rodríguez, E.; Moreno-Mayorga, J. C. A Strategy for Kinetic Parameter Estimation in the Fluid Catalytic Cracking Process. *Ind. Eng. Chem. Res.* **1997**, *36*, 5170–5174.
- (30) Cox, P. A.; Leach, A. G.; Campbell, A. D.; Lloyd-Jones, G. C. Protodeboronation of Heteroaromatic, Vinyl, and Cyclopropyl Boronic Acids: pH–Rate Profiles, Autocatalysis, and Disproportionation. *J. Am. Chem. Soc.* **2016**, *138*, 9145–9157.
- (31) Cox, P. A.; Reid, M.; Leach, A. G.; Campbell, A. D.; King, E. J.; Lloyd-Jones, G. C. Base-Catalyzed Aryl-B(OH)<sub>2</sub> Protodeboronation Revisited: From Concerted Proton Transfer to Liberation of a Transient Aryl Anion. *J. Am. Chem. Soc.* **2017**, *139*, 13156–13165.
- (32) Cox, P. Protodeboronation. *Ph.D. thesis*, The University of Edinburgh, 2016.
- (33) van de Vusse, J. G. Plug-flow type reactor versus tank reactor. *Chem. Eng. Sci.* **1964**, *19*, 994–996.
- (34) Riddlehoover, G. A.; Seagrave, R. C. Optimization of Van de Vusse Reaction Kinetics Using Semibatch Reactor Operation. *Industrial & Engineering Chemistry Fundamentals* **1973**, *12*, 444–447.
- (35) Sha, Y.; Yu, T. H.; Merinov, B. V.; Goddard, W. A. DFT Prediction of Oxygen Reduction Reaction on Palladium–Copper Alloy Surfaces. *ACS Catal.* **2014**, *4*, 1189–1197.
- (36) Frau, J.; Hernández-Haro, N.; Glossman-Mitnik, D. Computational prediction of the pK<sub>a</sub>s of small peptides through Conceptual DFT descriptors. *Chem. Phys. Lett.* **2017**, *671*, 138–141.
- (37) Ainley, A. D.; Challenger, F. CCLXXX.—Studies of the boron–carbon linkage. Part I. The oxidation and nitration of phenylboric acid. *Journal of the Chemical Society (Resumed)* **1930**, *0*, 2171–2180.
- (38) Kuivila, H. G.; Nahabedian, K. V. Electrophilic Displacement Reactions. X. General Acid Catalysis in the Protodeboronation of Areneboronic Acids 1–3. *J. Am. Chem. Soc.* **1961**, *83*, 2159–2163.
- (39) Kuivila, H. G.; Nahabedian, K. V. Electrophilic Displacement Reactions. XI. Solvent Isotope Effects in the Protodeboronation of Areneboronic Acids 1–3. *J. Am. Chem. Soc.* **1961**, *83*, 2164–2166.
- (40) Nahabedian, K. V.; Kuivila, H. G. Electrophilic Displacement Reactions. XII. Substituent Effects in the Protodeboronation of Areneboronic Acids 1–3. *J. Am. Chem. Soc.* **1961**, *83*, 2167–2174.
- (41) Kuivila, H. G.; Reuwer, J. F., Jr.; Mangravite, J. A. Electrophilic displacement reactions: xv. kinetics and mechanism of the base-catalyzed protodeboronation of areneboronic acids. *Can. J. Chem.* **1963**, *41*, 3081–3090.
- (42) Brown, R. D.; Buchanan, A. S.; Humffray, A. A. Protodeboronation of thiophenboronic acids. *Aust. J. Chem.* **1965**, *18*, 1521–1525.
- (43) Lozada, J.; Liu, Z.; Perrin, D. M. Base-Promoted Protodeboronation of 2,6-Disubstituted Arylboronic Acids. *Journal of Organic Chemistry* **2014**, *79*, 5365–5368.
- (44) Muir, C. W.; Vantourout, J. C.; Isidro-Llobet, A.; Macdonald, S. J. F.; Watson, A. J. B. One-Pot Homologation of Boronic Acids: A Platform for Diversity-Oriented Synthesis. *Org. Lett.* **2015**, *17*, 6030–6033.
- (45) Li, Z.; Gelbaum, C.; Heaner, W. L.; Fisk, J.; Jaganathan, A.; Holden, B.; Pollet, P.; Liotta, C. L. Palladium-Catalyzed Suzuki Reactions in Water with No Added Ligand: Effects of Reaction Scale, Temperature, pH of Aqueous Phase, and Substrate Structure. *Org. Process Res. Dev.* **2016**, *20*, 1489–1499.
- (46) Sander, T.; Freyss, J.; von Korff, M.; Rufener, C. DataWarrior: An Open-Source Program For Chemistry Aware Data Visualization And Analysis. *J. Chem. Inf. Model.* **2015**, *55*, 460–473.

## Recommended by ACS

### Reduction of CO<sub>2</sub> with Hydrated Electrons: Ab Initio Computational Studies for Finite-Size Cluster Models

Sebastian Pios and Wolfgang Domcke

APRIL 06, 2023

THE JOURNAL OF PHYSICAL CHEMISTRY A

READ 

### What the Heck?—Automated Regioselectivity Calculations of Palladium-Catalyzed Heck Reactions Using Quantum Chemistry

Nicolai Ree, Jan H. Jensen, *et al.*

DECEMBER 02, 2022

ACS OMEGA

READ 

### Divide-and-Conquer Linear-Scaling Quantum Chemical Computations

Hiromi Nakai, Yoshifumi Nishimura, *et al.*

JANUARY 11, 2023

THE JOURNAL OF PHYSICAL CHEMISTRY A

READ 

### Taming Keteniminium Reactivity by Steering Reaction Pathways: Computational Predictions and Experimental Validations

Mark A. Maskeri, K. N. Houk, *et al.*

DECEMBER 16, 2022

JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

READ 

Get More Suggestions >