



A Journal of the Gesellschaft Deutscher Chemiker

Angewandte Chemie

GDCh

International Edition

www.angewandte.org

Accepted Article

Title: Electro-descriptors for the performance prediction of electro-organic synthesis

Authors: Yuxuan Chen, Bailin Tian, Zheng Cheng, Xiaoshan Li, Min Huang, Yuxia Sun, Shuai Liu, Xu Cheng, Shuhua Li, and Mengning Ding

This manuscript has been accepted after peer review and appears as an Accepted Article online prior to editing, proofing, and formal publication of the final Version of Record (VoR). This work is currently citable by using the Digital Object Identifier (DOI) given below. The VoR will be published online in Early View as soon as possible and may be different to this Accepted Article as a result of editing. Readers should obtain the VoR from the journal website shown below when it is published to ensure accuracy of information. The authors are responsible for the content of this Accepted Article.

To be cited as: *Angew. Chem. Int. Ed.* 10.1002/anie.202014072

Link to VoR: <https://doi.org/10.1002/anie.202014072>

RESEARCH ARTICLE

Electro-descriptors for the performance prediction of electro-organic synthesis

Yuxuan Chen^{[a]†}, Bailin Tian^{[a]†}, Zheng Cheng^{[a],[b]}, Xiaoshan Li^[a], Min Huang^[a], Yuxia Sun^[a], Shuai Liu^[c], Xu Cheng^[c], Shuhua Li^{*[a],[b]} and Mengning Ding^{*[a]}

[a] Yuxuan Chen, Bailin Tian, Zheng Cheng, Xiaoshan Li, Min Huang, Yuxia Sun, Prof. Shuhua Li, Prof. Mengning Ding
Key Laboratory of Mesoscopic Chemistry, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing 210023, China
E-mail: shuhua@nju.edu.cn and mding@nju.edu.cn

[b] Zheng Cheng, Prof. Shuhua Li
Institute of Theoretical and Computational Chemistry, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing 210023, China

[c] Shuai Liu, Prof. Xu Cheng
Jiangsu Key Laboratory of Advanced Organic Materials, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing, 210023, China

[†]These authors contribute equally to this work

Supporting information for this article is given via a link at the end of the document.

Abstract:

Electrochemical organic synthesis has attracted increasing attentions from both scientific and industrial communities as a sustainable and versatile synthetic platform. Quantitative assessment of the electro-organic reactions, including reaction thermodynamics, interfacial kinetics and coupled chemical processes, highlights the uniqueness of electro-synthesis and can lead to the development of analytical tool to guide their future design. Here we study the thermodynamics, electro-kinetics and performance (yield) of electro-organic reactions with different mechanisms, and conclude that electrochemical parameters such as onset potential, Tafel slope, and effective voltage can be utilized as multi electro-descriptors for the evaluation of reaction conditions and the prediction of corresponding reactivities (reaction yields). An “electro-descriptor-diagram” is generated, where reactive and non-reactive conditions/substances show distinct boundary. Successful predictions for reaction outcomes have been demonstrated using electro-descriptor diagram, or from Machine Learning (ML) algorithms with experimentally-derived electro-descriptors. This method represents a promising tool for the data-acquisition, mechanistic clarification, reaction prediction, and high-throughput screening of feasible substance and optimal conditions (solvents, electrolytes, additives and electrode materials) for general organic electro-synthesis.

Introduction

Since Faraday's pioneering work using current to drive nonspontaneous reactions^[1], electro-organic synthesis has emerged as a popular methodology offering advantages in synthetic sustainability and industrial safety, including mild reaction conditions, green chemical process, high selectivity, and fine control over reaction parameters^[2]. Numerous electro-organic methodologies have been developed^[3], enabling facile synthesis of building blocks of pharmaceutical molecules^[4], functional materials^[5] and chemical intermediates^[6]; and are therefore of critical significance to the chemical and pharmaceutical industry. However, typical development of a new synthetic methodology requires the optimization of reaction conditions (voltage, solvent, additives and temperature) and calls for high-volume screening in search of feasible substrates^[7], both are

time consuming and potentially expensive in the industrial scenarios. In addition, fewer efforts have been directed to the mechanistic aspects of electro-organic reactions, such as interfacial electron transfer kinetics^[5b, 8], coupling of homogenous chemical process and the potential development of heterogeneous electro-catalysts^[9], which could significantly benefit the development of novel synthetic approaches from the fundamental perspective.

Electro-organic research could further benefit from electrochemistry as an efficient tool for both quantitative analysis and mechanistic investigation. For example, with mature theories and experimental practice of Cyclic Voltammetry (CV)^[10], a comprehensive understanding of electro-kinetics^[11], including mass transportation, interfacial electron transfer and coupled chemical reaction^[12], can be acquired for complicated electro-organic systems to give insights into the synthetic pathway that helps to optimize the reaction. There are reported examples utilizing electrochemical analysis such as CV to determine the redox potentials of different functional groups and investigate the reaction mechanisms^[13], but it still calls for new strategy that further promote the correlated studies and optimization.

In this work, we have developed a novel analytical method based on voltammetry for the quantitative and systematic evaluation of electro-organic reactions, and more importantly to predict substance reactivities and the reaction yield under specific set of reaction conditions. We report three experimentally derived electro-descriptors from easy-to-obtain CV diagram of the electro-organic systems: onset potential, Tafel slope, and effective voltage, in correlation to the thermodynamic, electro-kinetic and experimental (empirical) aspects of the synthetic reaction. An “electro-descriptor diagram” is developed, in which all high-yield substrates and conditions form a reaction “hot zone” that can be clearly distinguished from non-reactive sets. This electro-descriptor approach successfully applies to three electro-organic reactions with different pathways, with clear identification of the reactive “hot zone”, enabling accurate prediction (through either electro-descriptor diagram or ML algorithm) of the reaction yield from unknown substances, electrode materials and reaction conditions. The electro-descriptors represent the first experimentally derived descriptor system for the performance prediction of the organic reactions, as an alternative to the *in silico* structure-related descriptors. Overall, this method represents a promising tool for the prediction and screening of reactive substrates and optimized reaction conditions.

RESEARCH ARTICLE

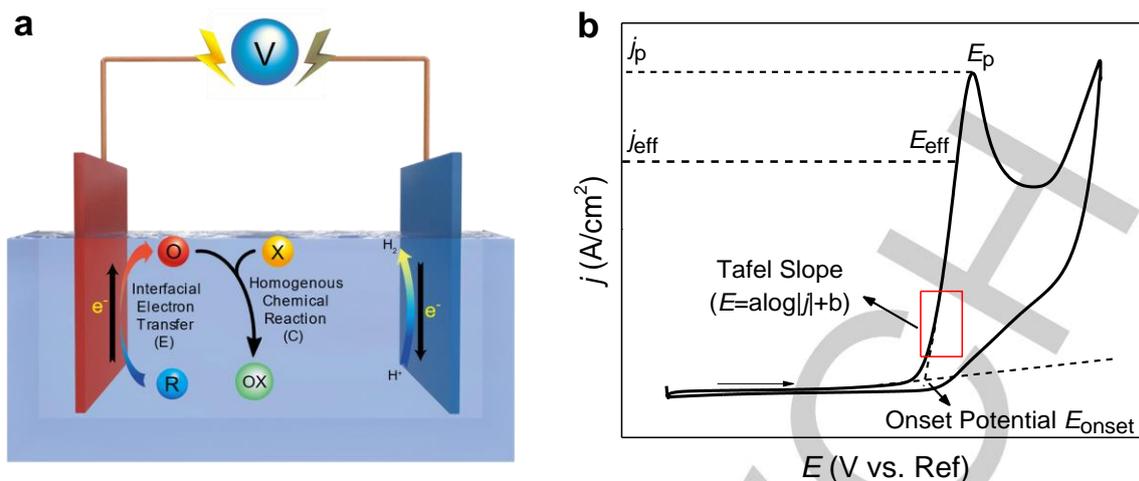
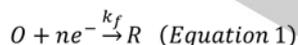


Figure 1 | Derivation of the electro-descriptors. **a.** Experimental set-up for an anodic electro-organic synthesis. R (substance in its reduced form) is first oxidized to O (oxidized form) on the anode (through interfacial electron transfer process, E), which then reacts with X (through homogenous chemical reaction, C) to form target molecule OX. **b.** Three electro-descriptors, onset potential, Tafel slope and effective voltage, in correspondence to the thermodynamic, electro-kinetic and operando (empirical reaction rate) perspectives of the electro-organic systems respectively, are identified from the CV diagram of different substrates and reaction conditions.

Results and Discussion

Electrochemical parameters from Cyclic Voltammogram. Figure 1b shows a representative CV curve of an anodic electro-synthetic system with typical reaction conditions. Three electro-descriptors, onset potential, Tafel slope and effective voltage are derived, representing three distinct yet correlated dimensions of the electrochemical system: thermodynamic aspect, electro-kinetics and empirical (*in operando*) parameters, respectively. For an anodic reaction, onset potential (E_{onset}) indicates the necessary potential to oxidize a substrate on the electrode surface: reaction with a higher onset potential is thermodynamically more unfavorable. Tafel slope serves as an electro-kinetic descriptor for the electron transfer on electrode. In an EC process (shown in Equation 1 and 2), Tafel slope measures the kinetics of the interfacial electron transfer (Equation 1).



A large Tafel slope is correlated with a relatively sluggish kinetic of the interfacial electron transfer. Finally, as typical bench research often followed the protocol of optimizing synthetic conditions for a “model reaction” and consequent exploring the scope of substrates, we define “effective voltage (current)” that can be extracted from actual electro-synthetic data for current (voltage) normalization to the model reaction. A higher effective voltage indicates that the electrolytic cell requires a higher driving force to reach the same *operando* reaction rate, which factors in empirical synthetic details.

Electro-descriptor analysis of an electro-organic system. We first studied the phosphorylation of tetrahydroisoquinoline^[15] (**1**) as a

proof-of-concept reaction as it represents a common synthetic methodology with typical reaction mechanism. Figure 2a demonstrates the electro-oxidation of **1** under controlled-voltage electrolysis (6.6 V cell voltage), following an EEC mechanism (E represents heterogeneous electron transfer on electrode and C represents a homogenous chemical reaction). To generate the electro-descriptors, CV tests were conducted on 29 sets of reaction conditions including 17 substrates and 12 entries with different solvents and additives for the same model substrate^[15h], as listed in Fig. 2b (also see Supplementary Table S1, S4 and Fig. S1). Voltammograms shown in Fig. 2c and 2d depict CV characteristics of typical reactive and non-reactive electro-organic systems. A clear distinction can be observed in onset potential (E_{onset}) of the reactive and nonreactive ones (Fig. 2c). Entry 1-2 has an E_{onset} (1.76 V vs. Ag/AgCl) that is much higher than Entry 1-4 (1.20 V vs. Ag/AgCl), which indicates that the oxidation of Entry 1-2 is thermodynamically unfavorable. From an electro-kinetic perspective, we also noticed that Tafel slopes of all the reactive systems demonstrate an optimal range from 0.4 V/dec to 0.7 V/dec, in distinct contrast to the non-reactive substrates (Fig. 2b), indicating the apparent existence of an ideal electro-kinetics to achieve desired products in electro-organic synthesis. A large Tafel slope observed in certain substrate suggests a sluggish kinetics that is unfavourable to its oxidation. On the other hand, a small Tafel slope showing fast oxidation kinetics, although favouring the production of R and its following reaction, is not always good as it might indicate the occurrence of side reactions (e.g., oxidation of aryl ring rather than amine) with distinct kinetics. From a quick assessment on these electrochemical parameters, the distinct features for optimal E_{onset} and Tafel slope indicate that one simple descriptor is not sufficient enough to evaluate the reactivity of an electro-organic system.

RESEARCH ARTICLE

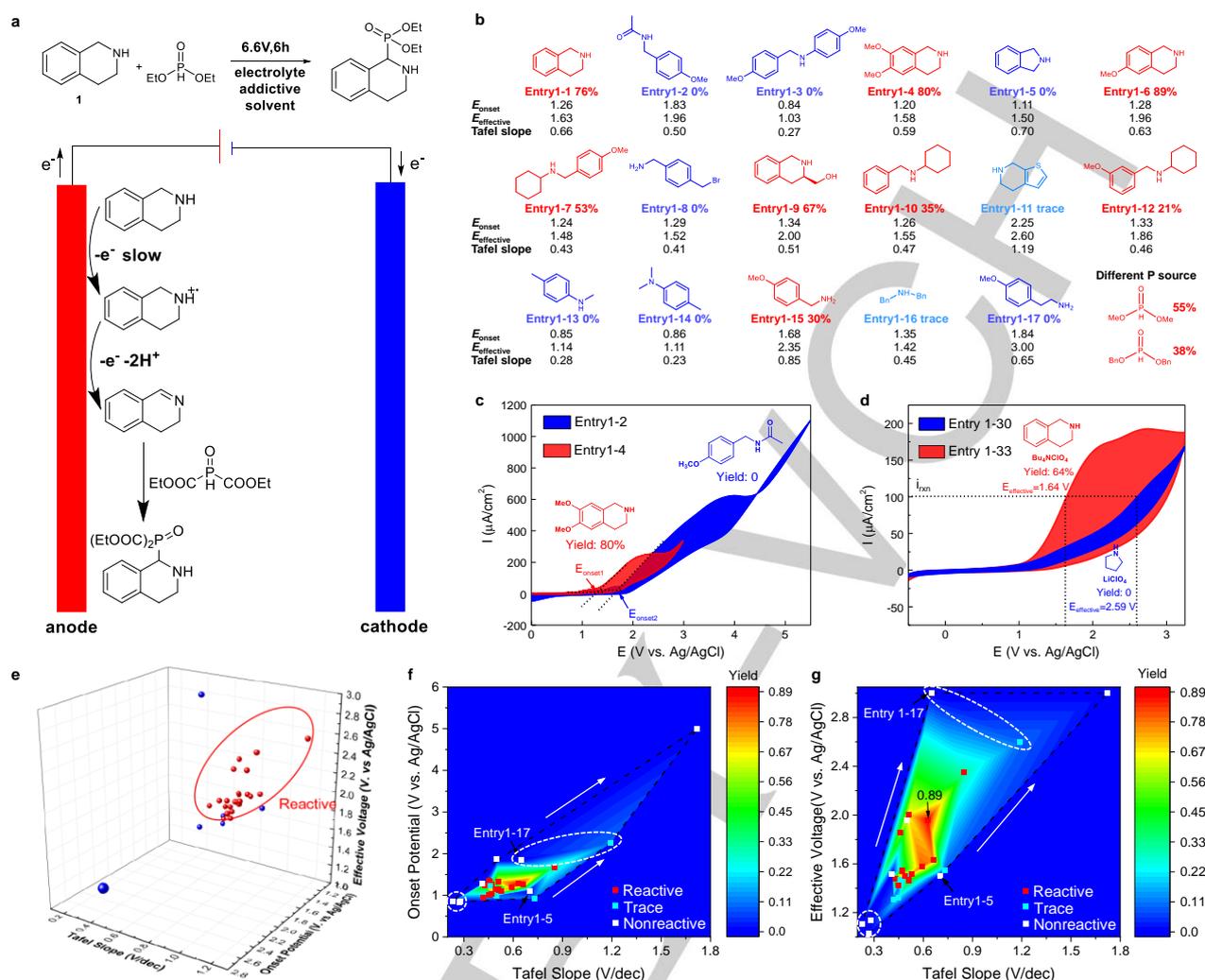


Figure 2] Electro-descriptor diagram of the EEC reaction. a. Schematic illustration of the electro-oxidation of tetrahydroisoquinoline (1). The quinoline substances went through a two-step electron transfer processes (EE) followed by the reaction with phosphates (C), and the first electron transfer is the rate determining step. **b.** The reaction yield and electro-descriptors of different substrates under same reaction condition optimized for model reaction with 1, different reaction conditions were also screened for model reaction (see Supplementary Table 1). **c.** CV of Entry 1-2 and Entry 1-4 demonstrating the importance of onset potential (E_{onset}): the nonreactive species have a higher onset potential. **d.** CV of Entry 1-30 and Entry 1-33 showing that despite close E_{onset} , the nonreactive entries (Entry 1-30) have a high effective voltage and a low reaction rate. **e.** The 3-D electro-descriptor diagram built with three electro-descriptors. A reactive zone is observed, with obvious boundary between reactive and nonreactive substances. **f, g.** 2-D electro-descriptor diagram, Onset potential—Tafel slope diagram (**f**) and Effective voltage—Tafel slope diagram (**g**) both show distinct reactive and non-reactive areas defined by reaction conditions with different yields, reaction yields are quantified by different color in the diagram (color bar on the right).

Effective voltage (or current) is defined as the voltage (current) needed to achieve the same *operando* current (voltage) as to the model reactions in a constant potential (current) scenario, which serves as a normalization to the model reaction. In some cases, effective voltage (see Fig. S2) leads to similar conclusions as onset potential, which is not surprising given that thermodynamically favorable reactions often show high overall charge transfer rate (*i.e.*, high faradaic current) under *operando* conditions. However, in some cases where E_{onset} is feeble to discriminate reactivity (see Fig. 2d), effective voltage could well distinguish the *operando* reaction rate difference. We later demonstrate that effective voltage offers independent evaluation to certain reactions where E_{onset} and Tafel slope are difficult to offer ideal distinction in the performance analysis,

while effective voltage (current) factors in actual experimental details and serves as an effective empirical descriptor.

Construction of multi electro-descriptor diagram. With all three electro-descriptors derived from the voltammogram of each electro-organic system (with specific substances and reaction conditions, see Fig. 2b and Fig. S4), they were then placed into the three coordinates system to provide a 3D “electro-descriptor diagram”. As shown in Fig. 2e, all reactive substances show a clear boundary with non-reactive ones, leading to a distinct reactive “hot zone”, indicating that for high yield reaction systems, these corresponding descriptors all fall into an optimal range. This apparent observation in the form of a diagram provides a convenient and potentially powerful tool to predict the

RESEARCH ARTICLE

reactivity of a target electro-organic reaction just by acquiring the voltammogram. Previously, single value of onset potential (or the apparent current increase) extracted from CV has been generally employed for reactivity evaluation in the electro-organic synthesis^{16f}. Our results reveal that data analysis with multi-dimensional electro-descriptors provide more accuracy for the reactivity prediction based on different fundamental perspectives. The 3D electro-descriptor diagram can further break down to two 2D diagrams, as shown in Figs. **2f** and **2g**, with additional color representation of reaction yield for more convenient visualization. Clear boundaries were first noted, defined by the most “out-bound”, typically “non-reactive” points in the diagram (shown as white points in Figs. **2f** and **2g**). The reaction “hot zone” can still be easily concluded from the 2D electro-descriptor diagram (Figs. **2f** and **2g**), where all the reactive substrates and conditions fall into an optimal 2D area. Reactive point that lies in the hot zone typically has a higher yield. The shape of reaction hot zone in Fig. **2f** indicates that the combination of two electro-descriptors (each represents kinetic and thermodynamic aspects), can be visually more effective to evaluate the reactivity of the system.

In addition, a linear correlation between Tafel slopes and onset potentials (as indicated by white arrows in Fig. **2f**) was observed. This correlation indicates an interesting principle in this electro-organic system: a more thermodynamically favorable reaction, to a certain extent, relates to a faster electro-kinetic process. Although this is not theoretically necessary in electrochemical synthesis, the possible underlying correlation between electro-kinetic and thermodynamic parameters, most likely appears in systems where the first charge transfer process on electrode surface is the rate-determine step, raises up a question whether their combination is sufficient to evaluate/predict electro-organic reactions with varying conditions. We later demonstrate that for different reaction mechanisms, thermodynamic, electro-kinetic, and overall reaction rate do not always have strict correlations. In Fig. **2f**, we did notice that the reactive points are crowded near the edge of the “hot zone”, which might bring uncertainty when used for prediction in unknown tests. Moreover, some nonreactive points (such as Entry 1-5 and Entry 1-17, labeled in Fig. **2f**) are located too close to the reactive ones, indicating the inefficiency using E_{onset} and Tafel slope to predict the reaction yield in certain “abnormal” cases.

For bench-top synthesis, it is relevant to measure and compare the actual reaction rate (*operando* current) to give the most accurate yield prediction. To this end, we introduce an empirical descriptor — effective voltage, which is derived from *operando* current normalization, to assess the rate of substrate's conversion (if no side reaction). Fig. **2d** demonstrates the effective voltage of two entries with similar E_{onset} , Entry 1-33 (reactive, 1.64 V vs. Ag/AgCl) and Entry 1-30 (nonreactive, 2.59 V vs. Ag/AgCl), revealing its role in

discriminating reactive and non-reactive systems. It can be easily rationalized that the reaction rate of non-reactive Entry 1-30 is slow. The introduction of effective voltage successfully addresses the abnormality issues with the E_{onset} —Tafel slope combination in this reaction system (Fig. **2f**). With Effective voltage—Tafel slope diagram (Fig. **2g**), a better defined reactive hot zone is observed for phosphorylation of unprotected secondary amine, and “non-reactive” points (such as Entry 1-5 & 1-17) can be better separated. Therefore, the combination of Tafel slope and effective voltage can better serve as the two electro-descriptor system for reactivity evaluation and prediction in this case.

Despite the clear separation of “non-reactive” substance/conditions (with zero yield of desired product) from “reactive” ones, it is also informative from a synthetic perspective to distinguish whether the “non-reactive” substance was electrochemically inert or was converted in a different pathway to give side products. To this end, crude product analysis was conducted for the typical “non-reactive” entries (see Supplementary Unreactive entry product analysis Section). Interestingly, the difference between “zero conversion” and “side reaction” can be reflected in the electro-descriptor diagram. For “0 yield” Entry 1-3, 1-13 and 1-14, product analysis show that the aryl rings in these substances were actually oxidized, rather than the intended oxidation of amine. As shown in Fig. **2f** and **2g** (for more detailed demonstrations in diagram, see Fig. **S6**), these batches appear in the bottom-left corner in electro-descriptor-diagrams, due to the significant small Tafel slopes. It is reasonable to conclude that such overly fast electro-kinetics indicate an alternative reaction pathway that leads to unintended side products. On the other hand, for “0 yield” Entry 1-11 and 1-17, most starting materials remain intact. Correspondingly, their position in the electro-descriptor diagram is located more to the top right corner, as a result of relatively large Tafel slope or effective voltage (see Fig. **S6**, effective voltages are more efficient to separate these points as their corresponding CV currents are small). For these two scenarios, their distinctive characteristics in the electro-descriptor diagram further demonstrate the application of this analytical approach to evaluate different reaction pathways for reactions without intended product. Additionally, for other cases such as Entry 1-2, 1-5 and 1-20, we also observed the oxidation of starting materials. These batches correspond to the cases where the electro-kinetics of side reactions is similar to the desired reactions, which require other electro-descriptors in addition to Tafel slope (onset potential for Entry 1-2 and effective voltage for Entry 1-5 and 1-20) for predictive accuracy. Overall, the product analysis of the “unreacted” trials in System 1 provides interesting information regarding the reaction pathways, with complementary perspectives that extend the analysis/prediction using this electro-descriptors system.

RESEARCH ARTICLE

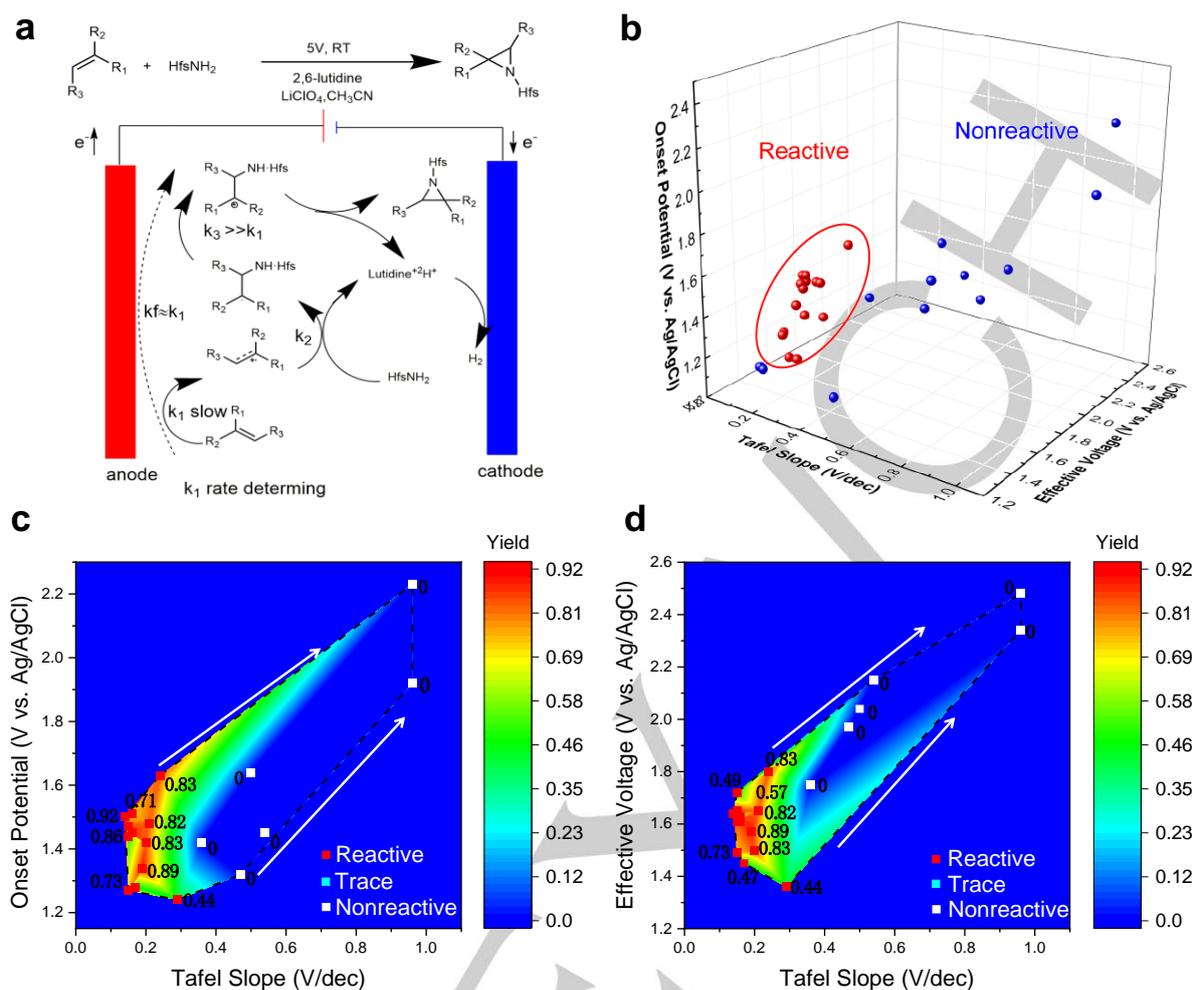


Figure 3| Electro-descriptor diagram of the ECEC reaction. **a.** A reported case of electrochemical oxidation and following aziridination of alkenes. The reaction undergoes an ECEC mechanism and the first electron transfer is the rate determining step. **b.** The electro-descriptor diagram is built using three electro-descriptors, showing a reactive hot zone with a distinct boundary. **c.** Onset potential—Tafel slope diagram shows a reactive zone where the high yield points are at the center of the reactive zone. The reactive substrates tend to have smaller Tafel slopes. **d.** Effective voltage—Tafel slope diagram also shows a reactive zone and high-yield points are close to the center. Nonreactive substrate generally has a higher effective voltage which indicates slow reaction rate in constant voltage electrolysis.

Electro-descriptor diagram for ECEC and catalytic EC reactions.

We further applied this electro-descriptor tool to investigate the electrochemical aziridination of alkene^[17], which is useful protocol to synthesize aziridines for pharmaceutical use^[18] and functional materials^[19]. The reaction undergoes an ECEC mechanism as shown in Figure 3a. CV tests were conducted at the reaction conditions reported in the reference^[5b], and three electro-descriptors were derived from CV diagrams. As shown in Fig. 3b, even more distinct reactive hot zone can be observed in its 3D electro-descriptor diagram. Fig 3c shows relatively more diverging correlation between the E_{onset} and Tafel slope, indicating that even though certain nonreactive points are thermodynamically favorable for the reaction, they demonstrate no reactivity (or poor selectivity) due to a sluggish electron-kinetics. In comparison, a relatively stronger correlation between Tafel slope and

effective voltage is observed in Fig. 3d. For similar reasons, we attribute the correlation to that the heterogeneous electron transfer is the rate determining step for this specific reaction. In the dual-descriptor diagram (Figs. 3c and 3d), it can also be concluded that Tafel slope along is sufficient enough to discriminate the reactivity of different substrates and conditions (in distinction to System 1). This result sheds further light on the mechanism of the electrochemical aziridination of alkene: the interfacial electron transfer is the rate determining step, the following chemical reaction and the second electron transfer proceeds at a much faster rate, so Tafel slope appears to be the dominating electro-kinetic descriptor in this reaction.

RESEARCH ARTICLE

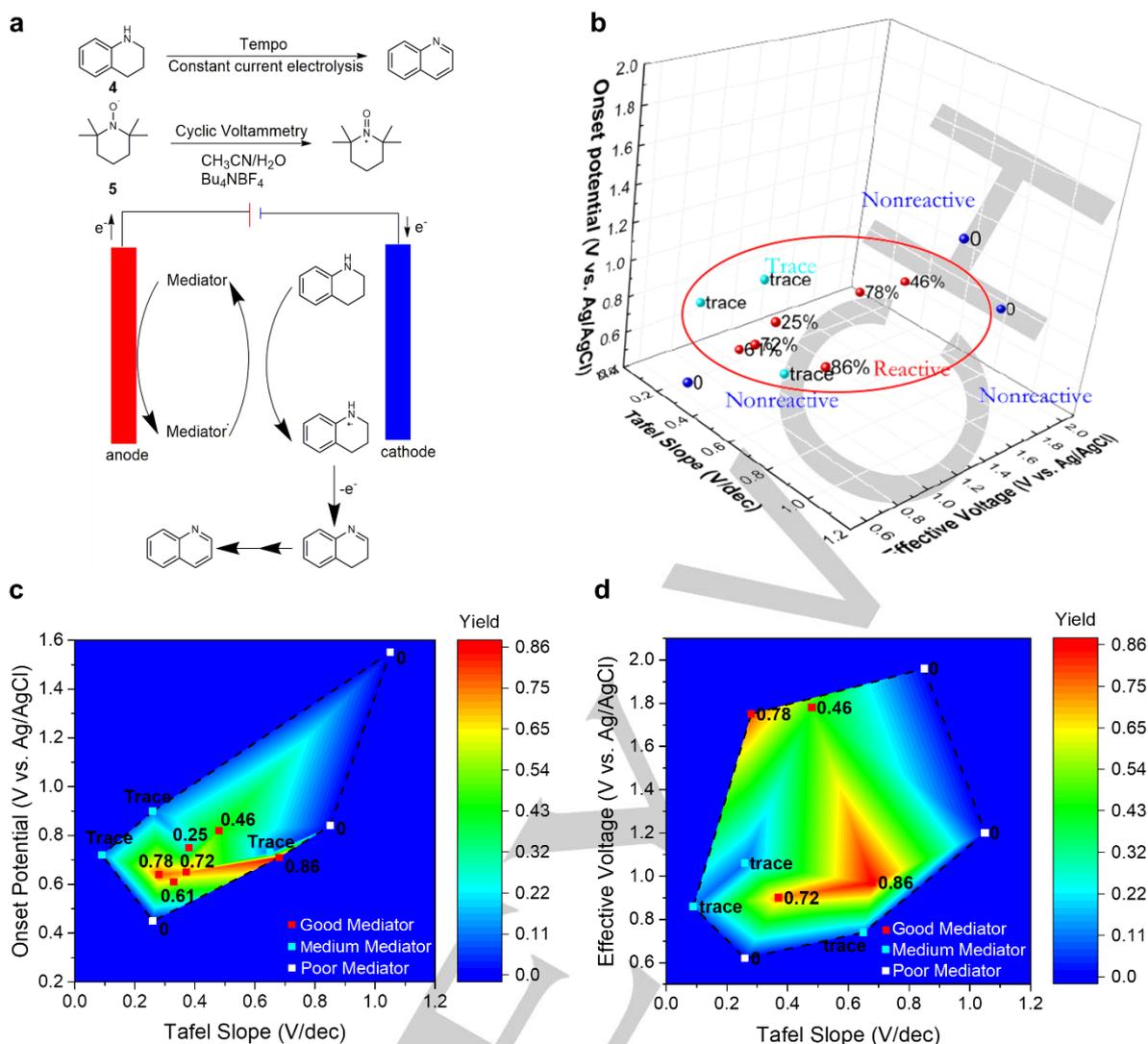


Figure 4| Electro-descriptor diagram of the EC reaction. a. Electrochemical dehydrogenation of N-Heterocycles utilizing organic electro-catalyst as reported. The reaction undergoes an EC mechanism: mediator **5** is oxidized on the electrode (E) and then diffuses back to the solution to dehydrate substance **4** (C). **b.** The electro-descriptor diagram of this EC reaction shows that efficient electro-catalysts form a reactive zone; medium electro-catalysts are close to the reactive zone and nonreactive points are far away from the reactive area. **c.** Onset potential—Tafel slope diagram shows that reactive points have a narrow distribution in onset potential. **d.** The distribution of reactive and non-reactive points shows less regularity in Effective voltage—Tafel slope diagram for this mediator mechanism.

In addition to the direct electrolysis, a large number of electro-organic synthesis have been conducted in an indirect fashion, with the addition of organic mediator to reduce the barrier for interfacial charge transfer^[20]. A better understanding of their electro-kinetics helps to sift out effective mediators under different conditions. Lei and co-workers reported an electrochemical dehydrogenation of N-Heterocycles utilizing TEMPO as electro-catalyst under constant current condition, following a typical EC mechanism (Figure 4a)^[21]. CV tests were run to 10 sets of electro-catalysts and conditions reported in the reference work^[21]. A hot zone is also observed in the 3D electro-descriptor diagram, with trace yield points at the periphery and nonreactive points in a distant area (Fig. 4b). A band-like reactive hot zone in E_{onset} —Tafel slope diagram (Fig. 4c), where all reactive points fall into a narrow range of onset potential but a wide range of Tafel slope, shows a significantly distinct reaction kinetics and mechanism (in comparison to Fig. 2f and Fig. 3c). It reveals that for the electro-

organic reactions catalyzed by a redox mediator, the thermodynamic effect dominates the overall progress. Once the first reactive barrier of the mediator is satisfied, the reaction proceeds well with different electrode kinetics. We also noticed that one point with a low effective voltage and low E_{onset} is nonreactive. An overly fast electro-kinetics might indicate side reactions such as destruction/coupling of mediator, leading to the loss of its catalytic activity. It should be noted that this electro-organic reaction is performed under constant current mode instead of constant voltage mode. While it does not significantly alter the result of E_{onset} —Tafel slope diagram, the distribution of reactive points is irregular in Effective voltage—Tafel slope diagram (Fig. 4d). In a galvanostatic electrolysis, the overall reaction rate is leveled to the same, thus effective voltage has no strong correlation with the reaction yield. The different observations from the electro-descriptor diagram call for a more systematic evaluation of the reactivity from both constant voltage and galvanostatic modes, when developing a new electro-organic methodology.

RESEARCH ARTICLE

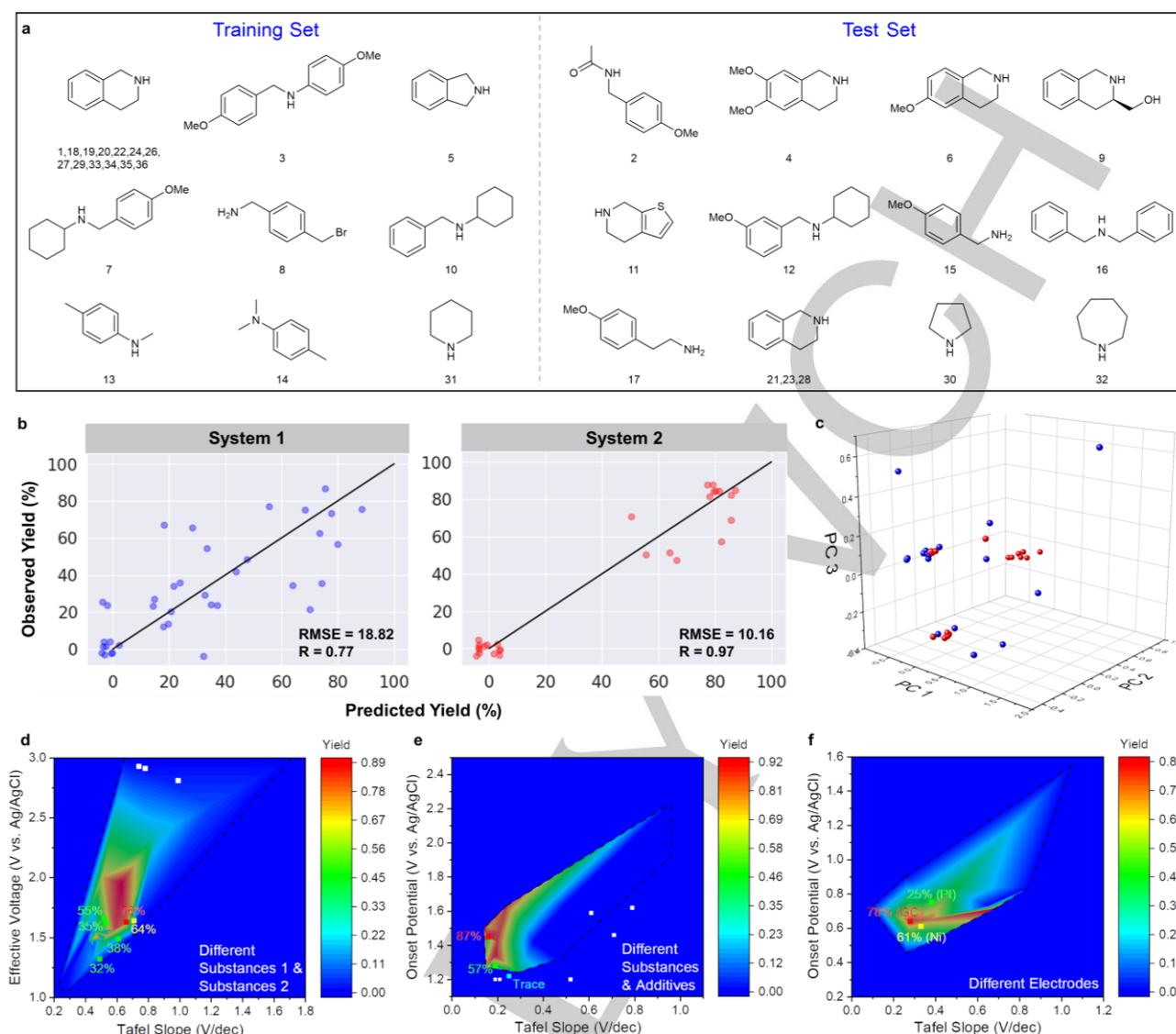


Figure 5 | Electro-descriptors for machine learning and visualized (diagram) prediction of the reaction performance. **a**. Machine learning (ML) analysis using decision tree algorithm to predict the yield of electro-phosphonylation of secondary amine (System 1). A training set of amines/conditions were randomly chosen to predict the performance of the test set. **b**. Observed versus predicted plots for System 1 (decision tree algorithm) and System 2 (k-nearest neighbor algorithm). Leave-one-out approach is applied to evaluate the performance of ML model, all predicted values are shown in the plot. As observed/predicted yields for non-reactive entries have the same value at (0, 0), uniform random noise with a size of 4.5 is added to the data (using a python library seaborn) to slightly separate them for better visualization. **c**. The activity diagram using three variables derived from Principle Component Analysis (PCA), in which red balls represent the reactive (yield $\geq 20\%$) entries while blue balls represent the nonreactive entries. **d**. Prediction of the reaction yields from different substances (both quinolines and phosphates in this bimolecular reaction) using pre-established electro-descriptors (effective voltage—Tafel slope) diagram for System 1. **e**. Prediction of the reaction yields from different substances and additives using pre-established electro-descriptors (onset potential—Tafel slope) diagram for System 2. **f**. Prediction of the reaction yields from different electrode materials using pre-established electro-descriptors (onset potential—Tafel slope) diagram for System 3. The actual yields (determined from bulk electro-synthesis) in **d-f** are listed on the side (white points represent zero yield reactions), which match well to the predictions from their positions in the electro-descriptor diagram. The most reactive points are listed for visual reference.

Machine learning and visualized prediction of synthetic performance using electro-descriptors. Machine learning (ML) methodologies are becoming effective in the design of synthetic route for various chemicals or in the prediction of product of a given reaction^[14b, 22]. However, due to the complex multidimensionality in an organic synthesis, it is more challenging to efficiently generate enough data for the employment of ML in reaction performance prediction^[23]. Doyle and co-workers has recently applied ML with atomic, molecular

and vibrational descriptors to enable yield predictions for Pd-catalyzed C-N cross-coupling^[23a]. The electro-descriptors defined in this work contain thermodynamic, kinetic and *operando* information, and therefore could serve as a group of natural descriptors for ML applications in electro-organic synthesis, with potential compatibility to the high-throughput analytical technologies. We have utilized different ML algorithms (see **Methods**) to predict the performance of reactions (yield) using electro-descriptors as input (combined with typical structural descriptors) in the phosphonylation of secondary

RESEARCH ARTICLE

amines (System 1) and aziridination of alkenes (System 2). CUR decomposition^[24] was used to construct uniformly-spaced training set, in order to guarantee that the training sets are evenly spread over the range of yields/conditions for more accurate ML prediction (see **Methods** for detail). In each case, 60% of data were used as training set to predict the yield in the remaining 40% test set, as shown in Figure **5a** (System 1). The predictive accuracies with different ML algorithms were evaluated and the k-nearest neighbor algorithm was found to provide better performance over random forest, gradient tree boosting, linear regression and decision tree algorithms. The best results are shown in Fig. **S8a** ($R=0.85$, $RMSE=17.08$) for System 1 and Fig. **S8b** ($R=0.98$, $RMSE=7.34$) for System 2. In order to avoid overfitting as the size of dataset is relatively small, the performances of ML models were further evaluated with the leave-one-out approach (see Table **S11** and corresponding discussions). The decision tree algorithm shows the best performance for System 1, whereas the k-nearest neighbor algorithm performs best for System 2. This difference is presumably due to the larger feature dimension in System 1 than in System 2, and k-nearest neighbor algorithm is more suitable with low dimensional situation. Moreover, both k-nearest neighbor and decision tree method are relatively insensitive to collinearity (see Table **S8** and **S9**). The best results are shown in Fig. **5b** ($R = 0.77$, $RMSE = 18.82$ for System 1 and $R = 0.97$, $RMSE = 10.16$ for System 2). The relative importance of the electro-descriptors was also evaluated with the random forest algorithm^[23a], results are shown in Fig. **S9** and Table **S10** (with corresponding discussions). Overall, these results indicate that our electro-descriptors can be successfully applied for ML prediction of the reaction performance in a given electro-organic synthesis.

To further confirm the correlations between electro-descriptors and the necessity of multi-descriptor system, Principal Component Analysis (PCA)^[25] was performed to the three electro-descriptors with additional electrochemical and structural characteristics used in ML model (see **Methods** and Table **S4-S7**). Principal component's composition reveals that all three electro-descriptors are crucial in discriminating reactivities of electro-organic reactions, and Pearson coefficient shows that there is no self-correlation among the reduced components. It is worth noticing that with successfully reduced correlation from PCA, the visualized separation of different reaction yields (Fig. **5c**) is less obvious compared to the electro-descriptor-diagram (Fig. **2e**). This result indicates the uniqueness of electro-descriptors in the evaluation of reactivities. Despite partial correlations, the thermodynamic, electro-kinetic, and *operando* aspect they each represents fundamentally determines the outcome of an electro-organic reaction under any given set of reaction conditions. Through accumulation of reaction data points, advanced computational methods can be potentially incorporated to help spot reaction hot zone and discriminate reactivity more efficiently through ML.

In principle, ML algorithms often require large volume of training sets for optimized precision, which is challenging for typical time-consuming bench-top reactions such as organic synthesis (recent advances in microreactors and high-throughput experimentation also offer tremendous potentials). With easy-to-obtain experimental electro-descriptors and the reactivity diagram, we further conducted visualized prediction using the electro-descriptor diagram as an alternative approach to the ML prediction. As shown in Fig. **5d**, with pre-established electro-descriptor diagram (Fig. **2g**), the reaction yield

from newly tested substances in System 1 can be accurately predicted according to their positions in the different reactivity zones in diagram. Excitingly, different quinolines and phosphates were both tested as unknown substances, and results showed that the electro-descriptor diagram can efficiently predict the performance from either reactant in a bimolecular reaction. In addition, this approach also works well in System 2 (used for yield prediction of new substances and new additives) and System 3 (used for yield prediction of new electrode materials) with pre-established electro-descriptor diagrams, results are shown in Fig. **5e** and **5f**. Other pre-established electro-descriptor diagrams of system 1-3 are listed in Fig. **S11** for reference. These successful yield predictions from different reaction variables demonstrate that the electro-descriptor diagram can serve as an alternative, effective and general tool for the prediction of outcomes in any electro-organic synthesis with given set of substances and reaction conditions.

Conclusion

In summary, we have identified and defined three electrochemical parameters, representing thermodynamic, kinetic and *operando* aspects, as multi electro-descriptors for electro-organic reactions. Electro-descriptor diagram with clear reactive hot zone has been developed, which can serve as an effective and powerful tool to evaluate and predict the outcome of a given electro-organic system. The electro-descriptor diagram applies to three electro-organic reactions with distinct mechanisms and provide different mechanistic insights. With electro-descriptors, successful predictions of reaction yields have been demonstrated with ML algorithms or direct visual prediction from electro-descriptor diagram. Such analysis and prediction could help avoid the vast expenses of time and resource using typical trial-and-error approach in the development of novel synthetic methods. We expect that with potential integration to the high-throughput experimentation and fast voltammetry to trace samplings, this electro-descriptor system will prove to be a universal and efficient tool for high-throughput screening of substances and optimization of conditions in future investigations.

Acknowledgements

The authors acknowledge the support by the Fundamental Research Funds for the Central Universities in China (020514380224), Natural Science Foundation of Jiangsu Province (BK20180321), and National Natural Science Foundation of China (21833002 and 21673110).

Keywords: electrochemistry • electro-descriptors • electroorganic synthesis • machine learning • reaction prediction

- [1] A. S. Lindsey, H. Jeskey, *Chem. Rev.* **1957**, *57*, 583-620.
[2] a) M. Yan, Y. Kawamata, P. S. Baran, *Chem. Rev.* **2017**, *117*, 13230-13319; b) E. J. Horn, B. R. Rosen, P. S. Baran, *ACS Cent. Sci.* **2016**, *2*, 302-308; c) R. Francke, R. D. Little, *Chem. Soc. Rev.* **2014**, *43*, 2492-2521; d) J. Yoshida, K. Kataoka, R. Horcajada, A. Nagaki, *Chem. Rev.* **2008**, *108*, 2265-2299.
[3] a) C. A. Paddon, M. Atobe, T. Fuchigami, P. He, P. Watts, S. J. Haswell, G. J. Pritchard, S. D. Bull, F. Marken, *J. Appl.*

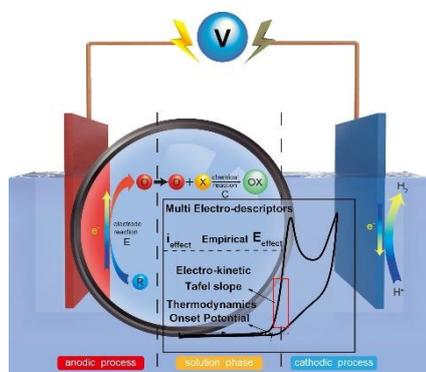
RESEARCH ARTICLE

- Electrochem.* **2006**, *36*, 617-634; b) J. C. Siu, N. Fu, S. Lin, *Acc. Chem. Res.* **2020**, *53*, 547-560; c) M. Elsherbini, T. Wirth, *Acc. Chem. Res.* **2019**, *52*, 3287-3296; d) K. Mitsudo, Y. Kurimoto, K. Yoshioka, S. Suga, *Chem. Rev.* **2018**, *118*, 5985-5999.
- [4] a) B. R. Rosen, E. W. Werner, A. G. O'Brien, P. S. Baran, *J. Am. Chem. Soc.* **2014**, *136*, 5571-5574; b) F. Xu, H. Long, J. Song, H.-C. Xu, *Angew. Chem. Int. Ed.* **2019**, *58*, 9017-9021.
- [5] a) I. C. Stewart, C. C. Lee, R. G. Bergman, F. D. Toste, *J. Am. Chem. Soc.* **2005**, *127*, 17616-17617; b) M. N. Jackson, S. Oh, C. J. Kaminsky, S. B. Chu, G. Zhang, J. T. Miller, Y. Surendranath, *J. Am. Chem. Soc.* **2018**, *140*, 1004-1010.
- [6] a) J. Li, L. He, X. Liu, X. Cheng, G. Li, *Angew. Chem. Int. Ed.* **2019**, *58*, 1759-1763; b) J. H. Simons, *J. Electrochem. Soc.* **1949**, *95*, 47-47.
- [7] A. R. Rosales, J. Wahlers, E. Limé, R. E. Meadows, K. W. Leslie, R. Savin, F. Bell, E. Hansen, P. Helquist, R. H. Munday, O. Wiest, P.-O. Norrby, *Nat. Catal.* **2018**, *2*, 41-45.
- [8] C. J. Kaminsky, J. Wright, Y. Surendranath, *ACS Catal.* **2019**, *9*, 3667-3671.
- [9] a) S. Li, W. Deng, S. Wang, P. Wang, D. An, Y. Li, Q. Zhang, Y. Wang, *ChemSusChem* **2018**, *11*, 1995-2028; b) W. Zhou, K. Cheng, J. Kang, C. Zhou, V. Subramanian, Q. Zhang, Y. Wang, *Chem. Soc. Rev.* **2019**, *48*, 3193-3228.
- [10] a) J. Heyrovsky, *Proc. Royal Soc. Lon.* **1923**, *102*, 628-640; b) J. Heyrovsky, *C. R. Hebd. Seances Acad. Sci.* **1924**, *179*, 1267-1268; c) R. S. Nicholson, *Anal. Chem.* **1965**, *37*, 1351-1355.
- [11] D. A. Buttry, M. D. Ward, *Chem. Rev.* **1992**, *92*, 1355-1379.
- [12] N. Elgrishi, K. J. Rountree, B. D. McCarthy, E. S. Rountree, T. T. Eisenhart, J. L. Dempsey, *J. Chem. Edu.* **2017**, *95*, 197-206.
- [13] a) F. Wang, S. S. Stahl, *Angew. Chem. Int. Ed.* **2019**, *58*, 6385-6390; b) C. Amatore, A. Jutand, G. Le Duc, *Chem. Eur. J.* **2011**, *17*, 2492-2503; c) C. Amatore, G. Le Duc, A. Jutand, *Chem. Eur. J.* **2013**, *19*, 10082-10093; d) D. P. Hickey, C. Sandford, Z. Rhodes, T. Gensch, L. R. Fries, M. S. Sigman, S. D. Minter, *J. Am. Chem. Soc.* **2019**, *141*, 1382-1392; e) Y. Kawamata, J. C. Vantourout, D. P. Hickey, P. Bai, L. Chen, Q. Hou, W. Qiao, K. Barman, M. A. Edwards, A. F. Garrido-Castro, J. N. deGruyter, H. Nakamura, K. Knouse, C. Qin, K. J. Clay, D. Bao, C. Li, J. T. Starr, C. Garcia-Irizarry, N. Sach, H. S. White, M. Neurock, S. D. Minter, P. S. Baran, *J. Am. Chem. Soc.* **2019**, *141*, 6392-6402; f) B. K. Peters, K. X. Rodriguez, S. H. Reisberg, S. B. Beil, D. P. Hickey, Y. Kawamata, M. Collins, J. Starr, L. Chen, S. Udyavara, K. Klunder, T. J. Gorey, S. L. Anderson, M. Neurock, S. D. Minter, P. S. Baran, *Science* **2019**, *363*, 838-845.
- [14] a) J. P. Reid, R. S. J. Proctor, M. S. Sigman, R. J. Phipps, *J. Am. Chem. Soc.* **2019**, *141*, 19178-19185; b) J. P. Reid, M. S. Sigman, *Nature* **2019**, *571*, 343-348; c) A. V. Brethomé, S. P. Fletcher, R. S. Paton, *ACS Catal.* **2019**, *9*, 2313-2323; d) S. G. Robinson, M. S. Sigman, *Acc. Chem. Res.* **2020**, *53*, 289-299.
- [15] a) G. M. Kosolapoff, *J. Am. Chem. Soc.* **1947**, *69*, 2112-2113; b) E. K. Fields, *J. Am. Chem. Soc.* **1952**, *74*, 1528-1531; c) R. Zhang, Y. Qin, L. Zhang, S. Luo, *J. Org. Chem.* **2019**, *84*, 2542-2555; d) J. Dhinesh Kumar, M. Lamani, K. Alagiri, K. R. Prabhu, *Org. Lett.* **2013**, *15*, 1092-1095; e) B. Lin, S. Shi, R. Lin, Y. Cui, M. Fang, G. Tang, Y. Zhao, *J. Org. Chem.* **2018**, *83*, 6754-6761; f) S. L. McDonald, Q. Wang, *Angew. Chem. Int. Ed.* **2014**, *53*, 1867-1871; g) M. Rueping, S. Zhu, R. M. Koenigs, *Chem. Comm.* **2011**, *47*, 8679-8681; h) M. Huang, J. Dai, X. Cheng, M. Ding, *Org. Lett.* **2019**, *21*, 7759-7762.
- [16] a) A. J. J. Lennox, J. E. Nutting, S. S. Stahl, *Chem. Sci.* **2018**, *9*, 356-361; b) M. Rafiee, F. Wang, D. P. Hruszkewycz, S. S. Stahl, *J. Am. Chem. Soc.* **2018**, *140*, 22-25; c) F. Wang, M. Rafiee, S. S. Stahl, *Angew. Chem. Int. Ed.* **2018**, *57*, 6686-6690; d) J. Li, W. Huang, J. Chen, L. He, X. Cheng, G. Li, *Angew. Chem. Int. Ed.* **2018**, *57*, 5695-5698.
- [17] a) J. Chen, W.-Q. Yan, C. M. Lam, C.-C. Zeng, L.-M. Hu, R. D. Little, *Org. Lett.* **2015**, *17*, 986-989; b) T. Siu, A. K. Yudin, *J. Am. Chem. Soc.* **2002**, *124*, 530-531; c) T. Siu, C. J. Picard, A. K. Yudin, *J. Org. Chem.* **2005**, *70*, 932-937; d) J. B. Sweeney, *Chem. Soc. Rev.* **2002**, *31*, 247-258; e) L. Degennaro, P. Trinchera, R. Luisi, *Chem. Rev.* **2014**, *114*, 7881-7929.
- [18] M. Miyazaki, K. Uoto, Y. Sugimoto, H. Naito, K. Yoshida, T. Okayama, H. Kawato, M. Miyazaki, M. Kitagawa, T. Seki, S. Fukutake, M. Aonuma, T. Soga, *Bioinorg. Med. Chem.* **2015**, *23*, 2360-2367.
- [19] I. C. Stewart, C. C. Lee, R. G. Bergman, T. F. Dean, *J. Am. Chem. Soc.* **2005**, *127*, 17616-17617.
- [20] a) J. E. Nutting, M. Rafiee, S. S. Stahl, *Chem. Rev.* **2018**, *118*, 4834-4885; b) E. Steckhan, *Angew. Chem. Int. Ed.* **1986**, *25*, 683-701; c) E. Steckhan, *Top. Curr. Chem.* **1987**, *142*, 1-69.
- [21] Y. Wu, H. Yi, A. Lei, *ACS Catal.* **2018**, *8*, 1192-1196.
- [22] a) H. Gao, T. J. Struble, C. W. Coley, Y. Wang, W. H. Green, K. F. Jensen, *ACS Cent. Sci.* **2018**, *4*, 1465-1476; b) R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, A. Aspuru-Guzik, *ACS Cent. Sci.* **2018**, *4*, 268-276; c) B. Liu, B. Ramsundar, P. Kawthekar, J. Shi, J. Gomes, Q. Luu Nguyen, S. Ho, J. Sloane, P. Wender, V. Pande, *ACS Cent. Sci.* **2017**, *3*, 1103-1113; d) B. Sanchez-Lengeling, A. Aspuru-Guzik, *Science* **2018**, *361*, 360; e) M. H. S. Segler, M. Preuss, M. P. Waller, *Nature* **2018**, *555*, 604-610; f) R. N. Straker, Q. Peng, A. Mekareeya, R. S. Paton, E. A. Anderson, *Nat. Commun.* **2016**, *7*, 10109; g) S. Szymkuć, E. P. Gajewska, T. Klucznik, K. Molga, P. Dittwald, M. Startek, M. Bajczyk, B. A. Grzybowski, *Angew. Chem. Int. Ed.* **2016**, *55*, 5904-5937; h) M. H. Todd, *Chem. Soc. Rev.* **2005**, *34*, 247-266; i) J. N. Wei, D. Duvenaud, A. Aspuru-Guzik, *ACS Cent. Sci.* **2016**, *2*, 725-732; j) A. F. Zahrt, J. J. Henle, B. T. Rose, Y. Wang, W. T. Darrow, S. E. Denmark, *Science* **2019**, *363*, eaau5631.
- [23] a) D. T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher, A. G. Doyle, *Science* **2018**, *360*, 186; b) F. Sandfort, F. Strieth-Kalthoff, M. Kühnemund, C. Beecks, F. Glorius, *Chem* **2020**, *6*, 1379-1390.
- [24] M. W. Mahoney, P. Drineas, *Proc. Nat. Acad. Sci.* **2009**, *106*, 697.
- [25] a) B. Moore, *IEEE Trans. Automat. Control* **1981**, *26*, 17-32; b) K. Pearson, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **1901**, *2*, 559-572; c) H. Zou, T. Hastie, R. Tibshirani, *Journal of Computational and Graphical Statistics* **2006**, *15*, 265-286.

RESEARCH ARTICLE

Entry for the Table of Contents

Insert graphic for Table of Contents here.



Insert text for Table of Contents here

For typical electro-organic synthetic systems, three electro-descriptors, onset potential (E_{onset}), Tafel slope and effective voltage ($E_{\text{effective}}$) are identified from CV diagram to characterize thermodynamic, electro-kinetic and empirical aspects of a given electro-organic reaction, providing an effective and powerful tool for machine learning or direct visual prediction of the reaction yield.